

# Fusion of Two-View Information: SVD Based Modeling for Computerized Classification of Breast Lesions on Mammograms

Rogério Daniel Dantas<sup>1</sup>, Marcelo Zanchetta do Nascimento<sup>2</sup>, Ricardo de Souza Jacomini<sup>2</sup>, Danilo César Pereira<sup>2</sup> and Rodrigo Pereira Ramos<sup>3</sup>

<sup>1</sup>*Universidade do Grande ABC*

<sup>2</sup>*Universidade Federal do ABC*

<sup>3</sup>*Universidade Federal do Vale do Sao Francisco  
Brazil*

## 1. Introduction

Over the past few years, the cancer has been one of the most responsible for the high number of deaths, and could become one of the main responsible for most deaths in the next decades. According to the World Health Organization, the number of deaths due to cancer, which was just 13% in 2008, is currently having a significant increase and one estimates that this number could reach approximately 12 million until 2030 (Tang et al., 2009).

Breast cancer is the second most common and leading cause of cancer death among women in Brazil. The National Cancer Institute (INCA) reports more than 50,000 new cases of the disease, with risk of 51 cases per 100,000 women. In the southeastern region of the country, this number is about 34% higher than the national average, with an estimate of 68 new cases per 100,000 women. For this reason, studies have shown that early detection is the key to improve breast cancer prognosis (Brasil, 2009).

Some works have shown that early detection of disease is a crucial factor for reducing mortality from breast cancer. Among the most important medical imaging methods (for example, MRI, ultrasonography and screen/film mammography), screen/film mammography is the method most easily accessible and has proved to be an effective aid for radiologists in the early detection of breast cancers and in the reduction of mortality rates. In this examination, four images are obtained, two corresponding to the right breast and two to the left breast of the projections cranio-caudal (CC) and medio-lateral oblique (MLO). The images acquisition improves visualization of breast tissue and increases the chances for detecting signs that characterize the presence of non-palpable lesions such as nodules, calcification, signs of bilateral asymmetry and architectural distortion (Rangayyan et al., 2007).

Retrospective evaluations of previous screening films of cancers, detected between screening rounds (interval cancers), show evidence of abnormality in 16% among 27% of cases. Missed cancers are due to many reasons: low disease prevalence, breast structure complexity, finding subtleties, and radiologist fatigue. To improve the accuracy of mammography, radiologists employed double reading of the same screening mammogram to increase the sensitivity rate

(Kinoshita et al., 2007). Second reading of the screening mammograms by a human reader can increase cancer detection rates (Thurfjell et al., 1994) (Warren & Duffy, 1995). Obviously, this procedure is too expensive, complex, and time consuming particularly in screening programs where a huge number of mammographic images have to be read (Mencattini et al., 2010).

The development of computerized systems as second readers represents an alternative. Computer-Aided Detection (CADe) and Computer-Aided Diagnosis (CADx) systems have been applied to mammographic images to assist radiologists on lesions analysis such as microcalcification, mass and architectural distortions. Algorithms for image processing together with artificial intelligence techniques, such as neural network and fuzzy logic, are used in order to enhance, segment, extract features and classify abnormalities (Jiang et al., 2007), (Balleyguier et al., 2007), (Doi et al., 1999). CADe schemes are systems that automatically detect suspicious lesions in mammograms, being used as a localization task. CADx systems extend the computer analysis to yield as output the characterization of a region or the estimated probability of lesion malignancy. The present chapter is focused on the classification task.

Although CADe and CADx systems have provided a large number of research and high rates of sensitivity, the majority of these works analyzes the MLO and CC views independently. In some situations, this system detects abnormalities in only one of the views. Radiologists believe that there is an inconsistency if a particular lesion is similar in both views and the system does not have the capability to find it. Studies have shown that these limitations have changed their impressions and radiologists are ignoring the results provided by these systems (Doi et al., 1999).

Several computer algorithms are used for identification of abnormalities in the breast by extracting 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 that are not easily interpreted by humans. This information can be obtained through a variety of image processing algorithms, calculated using a variety of statistical, structural and spectral techniques including co-occurrence matrices, fractal dimensions and multiresolution techniques. Using a different space by special data transform such as Fourier transform or wavelets transform could be helpful to separate a special data that contain specific characteristics. Multiresolution analysis allows for the preservation of an image according to certain level of resolution, i.e., it allows for the zooming in and out on the underlying texture structure. Therefore, the texture extraction is not affected by the size of the pixel neighborhood (Eltoukhy et al., 2010).

In this chapter, we present a method for extraction and selection of texture-related attributes and classification using the fusion of information from both CC and MLO visions. In the extraction stage, the wavelet transform method was applied to provide texture-related attributes for the considered images. Following, the singular value decomposition (SVD) technique was used to reduce the number of attributes. The application of analysis of variance (ANOVA) was also conducted for further reduction of the attributes after the SVD application. In the final step, we used the Random Forest and SVM classifiers to analyze mammogram lesions. The overall performance of the proposed method was evaluated by means of the area under the ROC curve.

## 2. Materials and methods

### 2.1 Data set

The database used in this work encompasses mammographic screen/film digitized images taken from the Digital Database for Screening Mammography (DDSM) (Balleyguier et al., 2007). 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.

For the evaluation of the algorithms, a data set was used comprising 480 mammographic images from 160 patients (being 80 with no lesions and 80 with malignant lesion) which were randomly selected from DDSM dataset. These lesions have different sizes, densities, and margins types. For each case, we used four mammograms, taken from the left and the right breasts, obtained in CC and MLO views. We selected images digitized with a Lumisys laser film scanner at 50 mm pixel size. Each image has a resolution of 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. The images used in the experiments were cuttings of size  $128 \times 128$  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 Figure 1. To obtain subimages 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 480 sub-images were acquired, and each of these cropped images was used for texture feature extraction and subsequent classification.

### 2.2 Multiresolution analysis

Image analysis on multiple scales allows image resolution to be modified in order to process as little data as possible. This is accomplished by selecting relevant details for a given task (Mallat, 1996). The basic idea is to represent the image on several resolutions and analyse them on frequency domain, through the application of function transforms. In this chapter, we analyse the performance of wavelet transforms on a multiresolution environment. The basics of wavelet transforms are presented as follows.

#### 2.2.1 Wavelet transforms

For a given function  $f(x) \in \mathbb{R}$ , the wavelet transform  $Wf(a, b) \in \mathbb{R}$  is obtained from the inner product of  $f(x)$  with a wavelet family, i.e.:

$$Wf(a, b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(x)dx \quad (1)$$

where

$$\psi_{a,b}(x) = a^{-1/2}\psi\left(\frac{x-b}{a}\right) \quad (2)$$

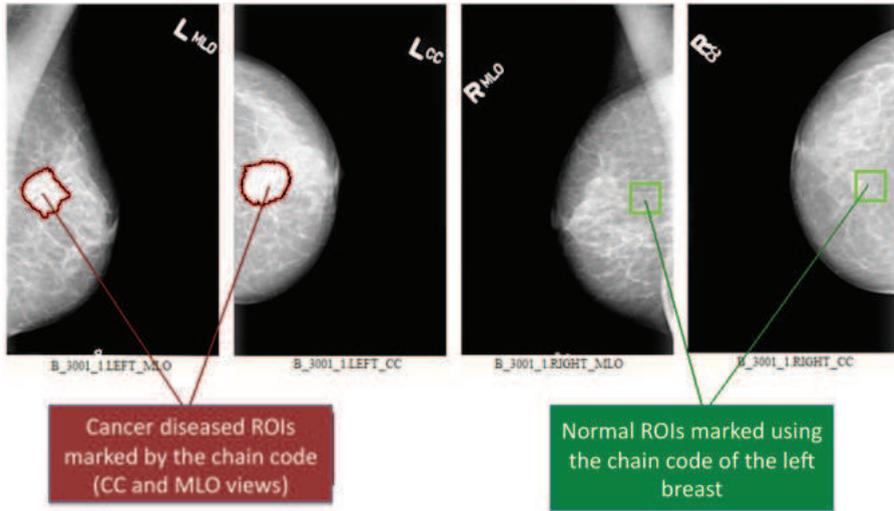


Fig. 1. Procedure applied for the selection of ROI's: images with malignant lesion and images no lesions.

are wavelets obtained from scaling and time shifting operations on a mother wavelet  $\psi(x)$ , with  $a$  and  $b$  being the scaling and translation factors, respectively. The mother wavelet  $\psi(x)$  is an integrable function with zero mean, i.e.:

$$\int_{-\infty}^{\infty} \psi(x)dx = 0 \tag{3}$$

and centered in the neighborhood of  $x = 0$ . Examples of wavelets can be found on reference (Mallat, 1998). The scale parameter  $a$  is referred to as the transform resolution level. When  $Wf(a, b)$  is known only for  $a < a_0$ , to recover  $f(x)$  it is necessary a complement of information corresponding to  $Wf(a, b)$  for  $a > a_0$ . This is obtained by introducing a scaling function  $\phi(x)$ , from which an approximation of  $f(x)$  at scale  $a$  is achieved:

$$Lf(a, b) = \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{a}} \phi\left(\frac{t-b}{a}\right) dx \tag{4}$$

In this way,  $Wf(a, b)$  and  $Lf(a, b)$  are defined as the detail and approximation coefficients of the wavelet transform, respectively.

To allow fast numerical implementations, it is common to impose that the scale and translation parameters vary only along discrete values, the most common being the dyadic wavelet decomposition, which is achieved when  $a = 2^j$  and  $b = k2^j$ , for integers  $j$  and  $k$ . Therefore, one can construct wavelet and scaling families given, respectively, by:

$$\psi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-2^j k}{2^j}\right) \tag{5}$$

$$\phi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{t-2^j k}{2^j}\right) \tag{6}$$

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

For discrete signals, the discrete wavelet transform (DWT) is obtained by the discretization of time as well as the translation and scale parameters. Mallat has proved that the dyadic DWT of a signal is equivalent to its decomposition through high-pass and low-pass filter banks, as many filters as is the desired resolution. The DWT achieves a multiresolution decomposition of a discrete signal  $f[n]$  on  $J$  octaves (resolutions) labeled  $j = 1, 2, \dots, J$  and given by:

$$f[n] = \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} a_{j,k} \tilde{g}[n - 2^j k] + \sum_{k \in \mathbb{Z}} d_{J,k} \tilde{h}[n - 2^J k] \tag{7}$$

The sequences  $\tilde{h}[n]$  and  $\tilde{g}[n]$  are low-pass and high-pass synthesis filters, respectively. They are called reconstruction filters.

The wavelet coefficients, namely the approximation coefficients  $a_{j,k}$  and the detail coefficients  $d_{j,k}$ , are obtained by the convolution operations:

$$a_{j,k} = \sum_n f[n] h_j[n - 2^j k] \tag{8}$$

$$d_{j,k} = \sum_n f[n] g_j[n - 2^j k] \tag{9}$$

where  $h_j[n]$  and  $g_j[n]$  are the analysis discrete filters, related to the approximation and detail coefficients, respectively. They are called decomposition filters. Considering two filters, a low-pass  $h[n]$  and a high-pass  $g[n]$ , the analysis filters can be obtained iteratively as:

$$g_1[n] = g[n] \tag{10}$$

$$h_1[n] = h[n] \tag{11}$$

$$g_{j+1} = \sum_k g_j[k] h[n - 2k] \tag{12}$$

$$h_{j+1} = \sum_k h_j[k] h[n - 2k] \tag{13}$$

At each  $j$ -th resolution, the wavelet coefficients are obtained by convolving the previous one with  $h_j[n]$  and  $g_j[n]$  and taking every other sample of the convolution (downsampling the result by a factor of 2), as illustrated by Figure 2. If the decomposition filter  $h[n]$  and the

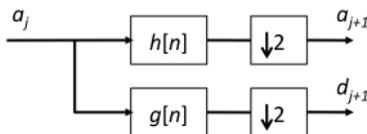


Fig. 2. DWT analysis filter bank.

reconstruction filter  $\tilde{h}[n]$  are considered to be equal, this defines conjugate mirror filters and  $h[n]$  and  $g[n]$  completely characterizes the dyadic DWT, with  $g[n] = (-1)^n h[N - 1 - n]$ , where  $N$  is the number of filter coefficients. In this case, as it can be seen, the filter  $h[n]$  plays an important role for wavelet transforms.

Different wavelet and scaling families can be constructed from the wavelet and scaling filters through:

$$\frac{1}{\sqrt{2}}\psi\left(\frac{t}{2}\right) = \sum_k g[k]\phi(t-k) \tag{14}$$

$$\frac{1}{\sqrt{2}}\phi\left(\frac{t}{2}\right) = \sum_k h[k]\phi(t-k) \tag{15}$$

As an example, the Daubechies wavelets are distinguished from each other by the number of vanishing moments. The Daubechies with 3 vanishing moments, referred to as Db3, was used in this work and has non-zero filter coefficients shown in Table 1.

$n$	$h[n]$
0	0.33267
1	0.80689
2	0.45988
3	-0.13501
4	-0.85441
5	0.35226

Table 1. Daubechies filter for a wavelet with 3 vanishing moments.

In what concerns images, which are bidimensional (2D) signals, the DWT can be computed as given by Figure 3, which shows an image decomposition from the  $j$ -th level to the  $(j + 1)$ -th level of the DWT. The coefficients  $LL_j$  represent the pixel values of the original image. Firstly, the image is passed through a pair of filters on each row, followed by a downsampling of 2. The results are used as inputs of two filter banks, which are applied at the columns of the image, followed by downsampling. Four sub-images are generated in this process: the approximation  $LL_{j+1}$ , which represents the original image with a smaller resolution, and the details  $LH_{j+1}, HL_{j+1}, HH_{j+1}$ , which represent the horizontal, vertical and diagonal directions, respectively.

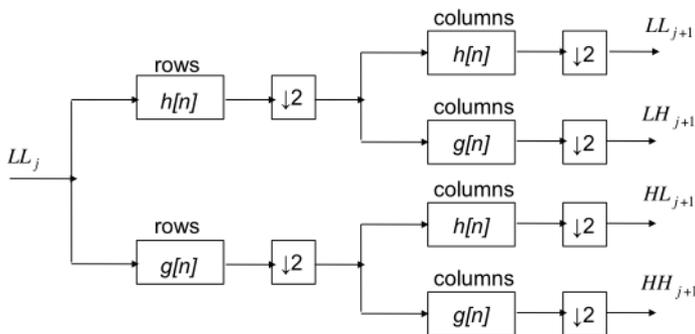


Fig. 3. Filter bank representing one DWT stage.

For each ROI, the 2D-DWT was applied using three different wavelet functions, Coiflet 5, Daubechies 3 e Symlet 4, with 2 resolution levels. The application of the first decomposition level yields the coefficient matrices  $LL_1, LH_1, HL_1$  and  $HH_1$ . The second decomposition level

applied in subband  $LL_1$  resulted in the coefficient matrices  $LL_2, LH_2, HL_2$  and  $HH_2$ , and so on. Once most of the information is contained in the detail coefficients, only the detail subimages are evaluated in this work.

### 2.2.2 Nonlinearity operator

Implementation of discrete wavelet transformation involves linear convolution of images with coefficients of filter meant for the wavelet basis function considered (Selvan & Ramakrishnan, 2007). However, linear convolution increases the size of subband images. This causes distortions at boundaries of the image, when the image is reconstructed.

To overcome this problem, we applied the method proposed by Selvan and Ramakrishnan to make the subband coefficients less sensitive to local variations (Selvan & Ramakrishnan, 2007). In this technique, each row and column of a  $Q \times Q$  image is periodically extended by  $R/2$  pixels on both sides, where  $R$  is the number of filter coefficients. That results in a  $(Q + R) \times (Q + R)$  image. Addition of any value lower than  $R/2$  will not yield the required core samples, after removing excess samples on the boundaries. An addition of any value higher than  $R/2$  will lead to more number of samples than required. Hence,  $R/2$  pixels are added on all four sides of the image. Convolution of the  $(Q + R) \times (Q + R)$  image, with filter of length  $R$ , yields  $((Q + R) + R - 1)$  samples in one dimension. Out of these samples, core  $(Q + 2R - 1 - 2(R - 1)) = (Q + 1)$  samples are considered for decimation, which results in  $[(Q + 1)/2]$  core samples after decimation. Hence, after one level of wavelet decomposition, a  $Q \times Q$  image yields four  $(Q/2) \times (Q/2)$  subbands, resulting in nonexpansive samples. In this case, the coefficient matrices were adjusted to the size  $64 \times 64$  for one decomposition level. For two decompositon level, this method yields subband matrices of size  $32 \times 32$ .

In order to make the subband coefficients less sensitive to local variations, nonlinearity and smoothing operators must be applied on wavelet transformation coefficients, before extracting the parameters. For nonlinearity and smoothing operations, the total energy of wavelet transformation coefficients in each sububand was calculated using the equation:

$$E_i = \frac{1}{PQ} \sum_{j=1}^P \sum_{k=1}^Q |w_i(j, k)|^2 \tag{16}$$

where  $E_i$  is the overall energy in the  $i$ -th subband,  $w_i(j, k)$  is the wavelet transformation coefficient at locations  $(j, k)$  in the  $i$ -th subband, and  $P$  and  $Q$  are the number of rows and columns of the  $i$ -th subband, respectively. Here, parameters  $P$  and  $Q$  have similar values.

In the present work, we also consider  $3 \times 3$  neighborhoods, compute their average, and normalize the average, as proposed by (Selvan & Ramakrishnan, 2007). At each location  $L_i(j, k)$ , the local energy was computed in a  $3 \times 3$  neighborhood using equation 17.

$$L_i(j, k) = \frac{1}{9} \sum_{u=0}^2 \sum_{v=0}^2 |w_i(j - 1 + u, k - 1 + v)|^2 \tag{17}$$

where the local energies  $L_i(j, k)$  were normalized by:

$$I_i(j, k) = \frac{L_i(j, k)}{E_i} \tag{18}$$

The central wavelet coefficient from the  $3 \times 3$  neighborhood was then replaced by the corresponding normalized energy.

Figure 4 represents the process of wavelet decomposition for each ROI to obtain the wavelet coefficient features.

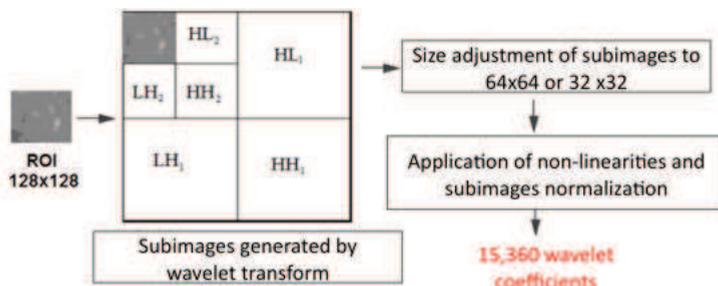


Fig. 4. Illustration Wavelet decomposition of a mammographic image.

This process produced 12,288 wavelet coefficients for the first decomposition ( $64 \times 64$  in 3 subimages) and 3,072 coefficients for the second decomposition ( $32 \times 32$  in 3 subimages). Gathering together all detail coefficients from the subbands resulted on a feature vector of 15,360 attributes.

### 2.3 Singular value decomposition

In this work, the number of wavelet coefficients is very large and the estimation of model parameters is computationally demanding. Singular value decomposition (SVD) based method is a very powerful mathematical tool which is mainly used here to reduce a large dataset to a dataset with significantly fewer values, still containing a large fraction of the variability present in the original data. SVD is extraordinarily useful and has many applications such as data analysis, signal processing, pattern recognition and image compression (Pedrini & Schwartz, 2008). It is a linear algebra tool that allows for the decomposition of a matrix into the product of three more simple matrices (Selvan & Ramakrishnan, 2007) (Ramakrishnan & Selvan, 2007a) (Ramakrishnan & Selvan, 2007b).

To explain the application of SVD, consider the matrix  $\mathbf{I}_i$  with size  $P \times Q$ , whose entries are the subband wavelet coefficients after the introduction of nonlinearity. The application of the SVD based method decomposes matrix  $\mathbf{I}_i$  into the product of three matrices given by:

$$\mathbf{I}_i = \mathbf{U}_i \mathbf{S}_i \mathbf{V}_i^T \quad (19)$$

where  $\mathbf{U}_i$ , with size  $P \times Q$ , and  $\mathbf{V}_i$ , with size  $Q \times Q$ , are orthogonal matrices whose columns are the eigenvectors of matrices  $\mathbf{I}_i \mathbf{I}_i^T$  and  $\mathbf{I}_i^T \mathbf{I}_i$ , respectively, and  $\mathbf{S}_i$ , with size  $Q \times Q$ , is a diagonal matrix whose non-zero entries are the singular values (square roots of the eigenvalues) of matrix  $\mathbf{I}_i \mathbf{I}_i^T$ , defined by:

$$\mathbf{S}_i = \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ & & \ddots & & \\ 0 & 0 & \cdots & 0 & \sigma_Q \end{bmatrix} \quad (20)$$

where  $\sigma_i$  are the non-zero singular values, with  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_Q$ . Once the SVD is unique for each matrix, the singular values completely represents the subband images.

A method of truncation of the lower singular values, which is equivalent to a filter based approach, was applied to matrix  $S_i$  for dimensionality reduction with images with noise. In (Selvan & Ramakrishnan, 2007), the authors have shown that the effect of noise is more intense on singular values with lower magnitudes. Therefore, the diagonal matrix can be truncated to a dimension  $K \times K$ , where  $K$  is given empirically by:

$$K = Q \frac{\sigma_1^2}{\sum_n \sigma_n^2} \tag{21}$$

where  $\sigma_n$  is the  $n$ -th singular value and  $\sigma_1$  is its highest value.

Since the wavelet transform in this work was considered for 2 resolution levels, there are subband images with sizes  $64 \times 64$  and  $32 \times 32$  pixels, what leads to a different number of truncated singular values (different  $K$ ) for each mammogram. Therefore, Equation 21 was used to get a value of  $K_r$  for each resolution level, and the overall  $K$  was obtained by averaging the number of truncated singular values obtained. Having defined the average value  $K$ , singular values were extracted from each of the eight wavelet subimages, resulting in a feature vector of  $6K$  elements representing texture characteristics of the original mammograms.

Figure 5 shows an example of the  $S_i$  diagonal matrix obtained from the subband  $LH_2$  of a Coiflet wavelet mother used for texture analysis. The values of the diagonal matrix was truncated by the average value  $K$ , which corresponds to the attributes in the region marked in green, representing the texture features.

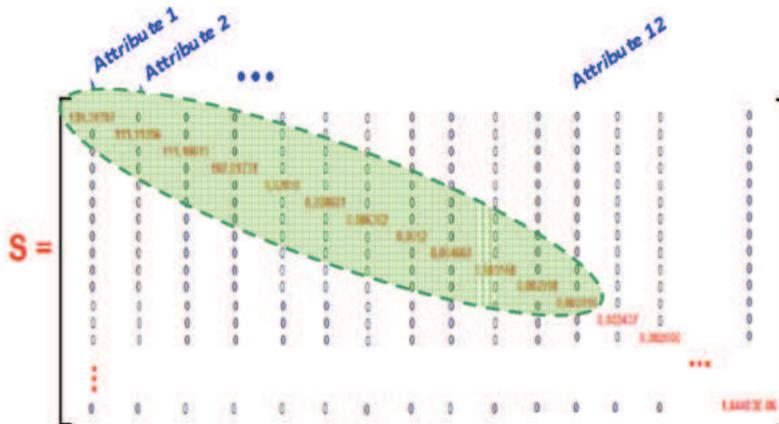


Fig. 5. Matrix  $S_i$  ( $32 \times 32$ ) of subband  $LH_2$  obtained by SVD from wavelet transform.

#### 2.4 Attributes reduction

In order to verify the relevance of the attributes obtained after the SVD procedure, we applied the technique of Analysis of Variance (ANOVA). This technique is a statistical model that compares the means of two or more experiments. This evaluation is performed to the extent

that the observed differences between the means are significant for the comparison of two estimates Vieira (2006). For our experiments, this comparison was carried out on the mass and normal mammographic images. Then, mean and standard deviation were calculated for sets of texture data of the masses and normal images. This set was used to evaluate if each attribute belonged to a particular class. However, when there are differences between the attribute and these classes, this information was eliminated because they did not suit as a reference in the definition of classes in the classification stage Pedrini & Schwartz (2008) Susomboon et al. (2008).

The one-way ANOVA method was firstly applied on the singular values attributes generated by the SVD procedure to determine the statistical significance of these values. In this method, the null hypothesis is that all attribute means are the same. The alternative hypothesis is that at least two of them are different. An F-test is applied to generate this test with a confidence interval of  $p = 5\%$ . The set with singular values was tested to evaluate if each attribute is different from each other. If any two groups are statistically the same, both are discarded since they do not contribute to the classification step.

In many cases, the alternative hypothesis is quite general and one needs some information about which pairs of means are significantly different, and which are not. This is done with a test called multiple comparison procedure. The multiple comparison method was applied recurrently to determine a minimum value for the number of reduced attributes. This step occurred with the change in the  $p$  value down to the value  $p = 1.10^{-16}$ .

The selection of sets for ANOVA evaluation and a graphic with the analysis of the attributes are illustrated in Figure 6

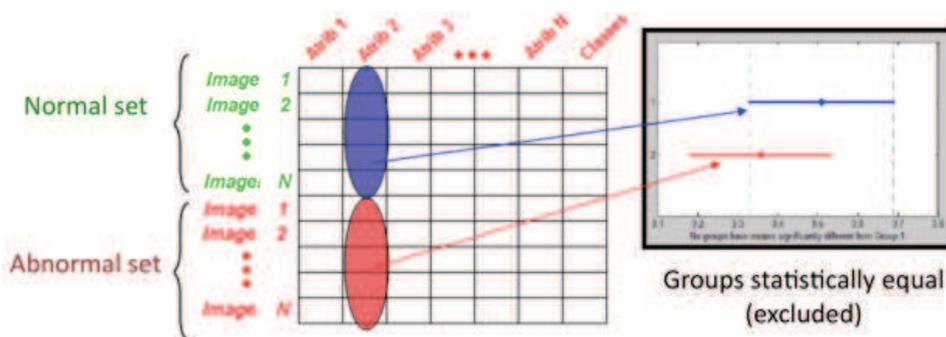


Fig. 6. Method of application of the ANOVA method on attribute 2.

## 2.5 Fusion of textural information

Several studies have demonstrated that the fusion of information extracted from two views, CC and MLO, allow for the reduction of false positive compared to the use of a single image Paquerault et al. (2002) Gupta et al. (2006).

For the process of fusion of information from two views, we applied the ANOVA technique to analyze the characteristics that are relevant in both images, i.e., information that was statistically different. The procedure resulted in an array of textural data with 50% of information from vision CC and 50% from MLO view. Only the common characteristics of the two views were used in the classification stage.

Figure 7 shows the block diagram of the algorithm to fusion information from the mammograms obtained in both CC and MLO view.

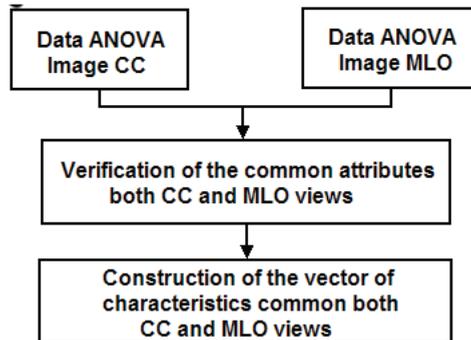


Fig. 7. Block diagram of the algorithm for obtaining the vector with the fusion of information from mammograms.

## 2.6 Classification stage

As a final stage of the computerized algorithm, after the information fusion step, a classification procedure was performed on the characteristic vectors outcome obtained from both views. In this work, we used the Random Forest and SVM classifiers, separately, to categorize the mammogram ROIs either as normal or abnormal tissues. These algorithms were implemented in WEKA (Waikato Environment for Knowledge Analysis) software (Vibha et al., 2006).

The Random Forest algorithm is built from a collection of classification trees. It is a concept of regression trees, bootstrap samples induced by a set of training data, using features selected in the random process of tree induction (Ramos & Nascimento, 2008).

The Support Vector Machine algorithm based on Sequential Minimal Optimization for implementation (SMO) is a machine learning technique based on statistical learning theory, which seeks to find a hyperplane with maximum separation between the classes, assuming that the data are linearly separable. If not, the SVM maps the data using a kernel function in a feature space of higher dimension, where the data becomes linearly separable (Osta et al., 2008).

To train and test the proposed computerized method, a cross validation procedure was performed on a dataset comprising 240 normal ROI (240 images) and 240 abnormal ROI (240 images). To obtain the performance for each classification we implemented a 10-fold cross validation procedure, in which the dataset were split into  $N$  parts.  $N - 1$  parts served as the training data to fit the classification model. The remaining part was used as the test data for the estimation of performance measures. Each of the  $N$  parts was used as test data in turn. The resulting  $N$  estimates were averaged to obtain the final expected value. Figure 8 presents the scheme of division into parts for the application of cross-validation procedure.

## 2.7 System evaluation

To give support to the performance evaluation of proposed method, we computed sensitivity (true positive rate), specificity (true negative rate) and accuracy for each of the confusion



Fig. 8. Division scheme for the application of training and test classification step.

matrices (Tsai & Lee, 2002). 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 2. The parameters employed for performance evaluation can be summarized

Class	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

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

in Table 3, 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. Overall accuracy is the probability that a diagnostic test is correctly performed.

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

Table 3. Definition of performance evaluation parameters.

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  $1 - \text{specificity}$ ), plotted on the  $x$ -axis. An ROC graph depicts relative trade-offs between benefits (TP) and costs (FP), 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.

### 3. Results

#### 3.1 Extraction and reduction from wavelet transform

In the proposed method, texture information was extracted by wavelet transform application and the SVD method was employed to reduce the number of coefficients before the classification stage. These methods were applied to each ROI obtained in the CC and MLO views. Table 4 shows the number of truncated singular values ( $K$  parameter) for the first

and second levels of each subband image of each view, as well the resulting number of attributes after SVD application. The resulting number of coefficients is obtained multiplying the second and third rows of the table by 3 and summing the result, since it was employed a two level decomposition that results in 3 subimages for each level. Once the wavelet

Mother wavelet	K (64×64)	K (32×32)	Number of Coefficients
<i>Coiflet 5</i>	23	12	105
<i>Daubechies 3</i>	26	14	120
<i>Symlets 4</i>	23	13	108

Table 4. Value of truncated K parameter and resulting number of coefficients after reduction for each subband image of each CC and MLO views.

transform is a natural filtering process derived from distinct discrete filters, the use of different wavelet mothers conducted to a different number of coefficients, reflected in the truncated K parameter, for each analyzed image. As seen in the previous table, in summary, the SVD method can effectively reduce the number of coefficients obtained from wavelet subimages. As stated before, each wavelet transformed ROI initially encompassed a total of 15,360 wavelet coefficients extracted from a two-level decomposition procedure. The application of the SVD method provided an average reduction of 99.3% on the number of attributes necessary for evaluation of texture information. Comparing the different wavelet mothers evaluated, it can be observed that the Coiflet 5 wavelet provided the smaller number of coefficients.

To reduce the number of attributes and provide a better separation of diseased and normal classes, the method of analysis of variance (ANOVA) with significance level of 5% was applied on the coefficients resulting from the SVD method. Table 5 shows the number of coefficients resulting from the application of ANOVA method with a significance level of 0.05.

Type of Wavelet Mother	CC	MLO
<i>Coiflet 5</i>	51	48
<i>Daubechies 3</i>	62	48
<i>Symlets 4</i>	56	43

Table 5. Number of wavelet coefficients after application of ANOVA method with  $p = 0.05$ .

The procedure to determine the significance was applied again to determine a minimum value in reducing the number of attributes. Table 6 presents the obtained results in relation to the number of coefficients after the application of ANOVA method with the limiting value set to  $p = 1.10^{-16}$  Applying the ANOVA method on the attributes resulting from the SVD method,

Type of Wavelet Mother	CC	MLO
<i>Coiflet 5</i>	7	14
<i>Daubechies 3</i>	5	11
<i>Symlets 4</i>	5	15

Table 6. Number of the wavelet coefficients after application of ANOVA method with  $p = 1 \times 10^{-16}$ .

it can be noticed that there is a significant reduction on the number of coefficients. These values are even more expressive when the ANOVA was applied recursively with the limiting significance level set to  $p = 1 \times 10^{-16}$ , which led to an average reduction of 81.34% compared to the first application of ANOVA.

### 3.2 Classification of CC and MLO views independently

Tables 7 and 8 show the mean values of sensitivity, specificity, accuracy and standard deviation ( $\sigma$ ) for the Random Forest and SVM classifiers, respectively, for CC e MLO views independent after the reduction process of wavelet coefficients. Sensitivity and specificity are the basic

	<i>Coiflet 5</i>		<i>Daubechies 3</i>		<i>Symlet 4</i>	
	CC	MLO	CC	MLO	CC	MLO
<b>Sensitivity</b>	74,17	76,04	75,83	77,50	74,44	77,78
<b>Specificity</b>	75,83	77,29	77,08	77,83	75,00	78,06
<b>Accuracy</b>	75,00	76,67	76,46	77,67	74,72	77,92
$\sigma$	16,20	15,18	14,83	14,65	15,89	14,54

Table 7. Measurements obtained with the Random Forest classifier for each view of mammogram.

	<i>Coiflet 5</i>		<i>Daubechies 3</i>		<i>Symlet 4</i>	
	CC	MLO	CC	MLO	CC	MLO
<b>Sensitivity</b>	70,00	69,38	67,92	71,17	67,78	71,11
<b>Specificity</b>	81,67	85,63	83,33	86,33	84,44	87,22
<b>Accuracy</b>	75,83	77,50	75,63	78,75	76,11	79,17
$\sigma$	9,58	10,63	10,58	10,56	10,83	10,32

Table 8. Measurements obtained with the SVM classifier for each view of mammogram.

measures of the precision for a given diagnostic test. Observing the data on Tables 7 and 8 we can see a great variation on the results for the different wavelets considered. From Table 7, concerning the Random Forest classifier, it can be noticed that the MLO view provided the best results for all three wavelet mothers used. Among the wavelet functions considered, there is little difference on the results (approximately 1%), and the best results were obtained with the application of the Symlet 4 wavelet mother.

A similar result can be observed for the SVM classifier, as it can be seen on Table 8 that shows better results for the MLO views. Comparing the values for the MLO and CC views, the first one resulted on average values 3% superior than the latter. These results show that an inconsistency can occur if a particular lesion is similar in both views and the system does not have the capability to find it. As stated in (Doi et al., 1999), studies have shown that these limitations have been considered mistrusted and radiologists tend to ignoring the results provided by these systems.

When comparing both classifiers, this inconsistency can also be noted, since there is a large variability of results. The Random Forest presented better results for false positive rates than the SVM classifier, but the latter, on average, performed better for the false negative tests. Considering the accuracy results (overall system performance), the SVM classifier presented the better results on average.

Those results are corroborated when considering the AUC as a measure of the system performance. Figure 9 shows the results of AUC for the Random Forest classifier applied to the CC and MLO views independently, after the reduction procedure, parameterized by the different wavelet functions. In Figure 10, the same results of AUC are shown when applying the SVM classifier. As it can be noticed, SVM performed better for almost all cases, the exception being the Daubechies 3 wavelet on the CC view.

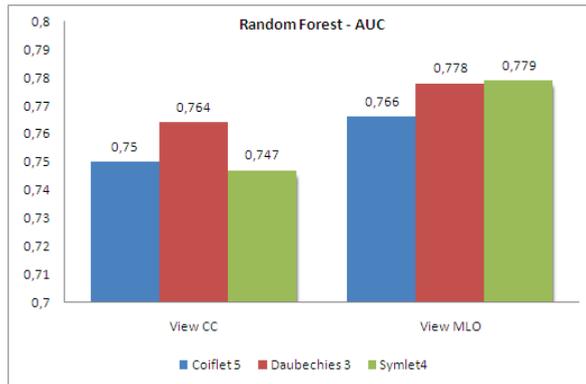


Fig. 9. Results of AUC after application of the reduction methods for different wavelets and Random Forest classifier.

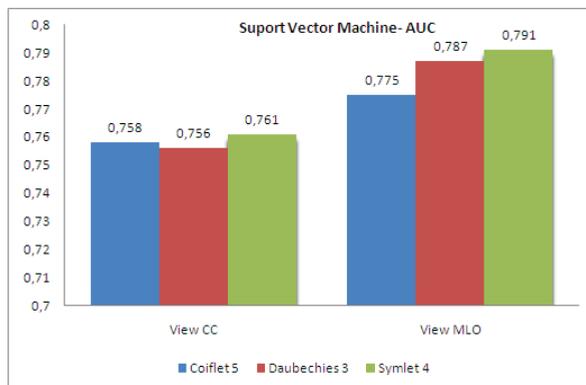


Fig. 10. Results of AUC after application of the reduction methods for different wavelets and SVM classifier.

### 3.3 Classification of two-view information fusion (CC and MLO)

Table 9 shows the number of texture data after application of the ANOVA technique in each view separately and the number of attributes after the two-view attributes fusion for each type of wavelet mother. Tables 10 and 11 show the mean values of sensitivity, specificity, accuracy

Type of Wavelet-Mother	CC	MLO	CC + MLO
<i>Coiflet 5</i>	7	14	14
<i>Daubechies 3</i>	5	11	10
<i>Symlets 4</i>	5	15	10

Table 9. Number of relevant textural information after application of the ANOVA technique with  $p = 1 \times 10^{-16}$ .

and standard deviation ( $\sigma$ ) for the process of two-view information fusion, using the Random Forest and SVM classifiers, respectively.

	<i>Coiflet 5</i>	<i>Daubechies 3</i>	<i>Symlet 4</i>
	CC + MLO	CC + MLO	CC + MLO
<b>Sensitivity</b>	82,50	83,33	82,50
<b>Specificity</b>	81,67	82,92	80,83
<b>Accuracy</b>	82,08	83,13	81,67
$\sigma$	12,43	11,74	13,20

Table 10. Results for two-view fusion obtained using the Random Forest classifier.

	<i>Coiflet 5</i>	<i>Daubechies 3</i>	<i>Symlet 4</i>
	CC + MLO	CC + MLO	CC + MLO
<b>Sensitivity</b>	75,83	77,08	75,00
<b>Specificity</b>	88,33	89,17	88,61
<b>Accuracy</b>	82,08	83,13	81,81
$\sigma$	10,77	9,89	9,94

Table 11. Results for two-view fusion obtained using the SVM classifier.

As seen on Tables 10 and 11, the results for the two classifiers using two-view CC and MLO information fusion were increased for all wavelet functions considered. As an example, for the Coiflet 5 wavelet, the fusion procedure led the sensitivity (fraction of the number of regions detected as positive against those that actually have a mass) to be increased about 5% compared to the result of the same wavelet for individual visions. The same occurred to the other two wavelets. Furthermore, the results also show an increase in the specificity

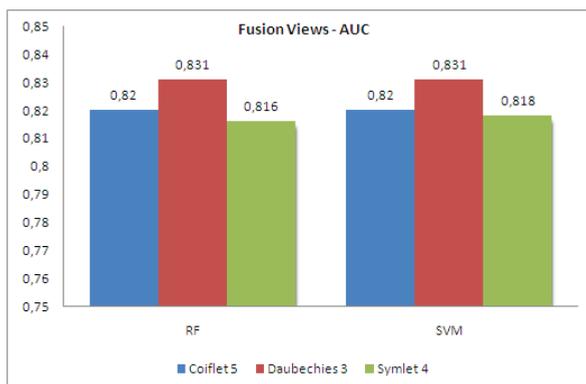


Fig. 11. Results of AUC after the two-view information fusion process for the Random Forest and SVM classifiers.

and accuracy parameters due to the two-view information fusion procedure. The results on the tables also show that the Daubechies 3 wavelet function have reached the best results compared to the other two applied wavelets. Since accuracy reflects the relation between sensitivity and specificity, that descriptor permitted to note the similar overall performance between the two classifiers. This result is also visible when evaluating the AUC performance of the classifiers. Figure 11 shows the results of AUC after CC and MLO information fusion for the three different types of wavelet mother and the Random Forest and SVM classifiers.

#### 4. Conclusion

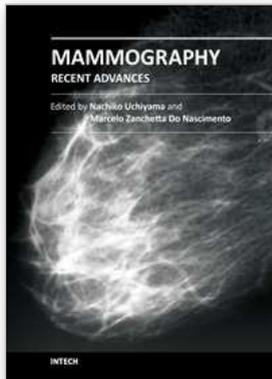
In this work, we develop a method for extraction and selection of attributes using the information fusion from both CC and MLO views for breast cancer classification of mammogram images. Comparing the results of Tables 7 with 10 and 8 with 11, we observe that the use of textural information, obtained from the fusion of the CC and MLO views, might raise the rates of the descriptors evaluated.

Experimental results showed that the three types of wavelet mother had superior performance over the area under the ROC curve for the SVM classifier. Performance evaluation was conducted in relation to the AUC and the results obtained showed that the best classification rates were obtained using the Daubechies wavelet, giving an AUC = 0.831 for the Support Vector Machine classifiers. Future experiments may use other types of feature descriptors to evaluate possible improvements in image classification.

#### 5. References

- Balleysguier, C., Ayadi, S., Nguyen, K. V., Vanel, D., Dromain, C. & Sigal, R. (2007). Birads classification in mammography., *Eur J Radiol* 61(2): 192–194.  
URL: <http://dx.doi.org/10.1016/j.ejrad.2006.08.033>
- Brasil, Ministerio da Saude, I. N. d. C. I. (2009). Estimativa 2010: incidencia de cancer no brasil.  
URL: <http://www.inca.gov.br/estimativa/2010/>
- Doi, K., MacMahon, H., Katsuragawa, S., Nishikawa, R. M. & Jiang, Y. (1999). Computer-aided diagnosis in radiology: potential and pitfalls., *Eur J Radiol* 31(2): 97–109.
- Eltoukhy, M. M., Faye, I. & Samir, B. B. (2010). Curvelet based feature extraction method for breast cancer diagnosis in digital mammogram, *Proc. Int Intelligent and Advanced Systems (ICIAS) Conf*, pp. 1–5.
- Gupta, S., Chyn, P. F. & Markey, M. K. (2006). Breast cancer cadx based on bi-rads descriptors from two mammographic views., *Med Phys* 33(6): 1810–7.
- Jiang, J., Yao, B. & Wason, A. (2007). A genetic algorithm design for microcalcification detection and classification in digital mammograms, *Computerized Medical Imaging and Graphics* 31(1): 49 – 61.  
URL: <http://www.sciencedirect.com/science/article/pii/S0895611106001029>
- Kinoshita, S., de Azevedo-Marques, P., Pereira, R., Rodrigues, J. & Rangayyan, R. (2007). Content-based retrieval of mammograms using visual features related to breast density patterns, *Journal of Digital Imaging* 20: 172–190. 10.1007/s10278-007-9004-0.  
URL: <http://dx.doi.org/10.1007/s10278-007-9004-0>
- Mallat, S. (1996). Wavelets for a vision, *Proceedings of the IEEE* 84(4): 604–614.
- Mallat, S. (1998). *A wavelet tour of signal processing*, Academic Press.
- Mencattini, A., Salmeri, M., Rabottino, G. & Salicone, S. (2010). Metrological characterization of a cadx system for the classification of breast masses in mammograms, *IEEE J IM* 59(11): 2792–2799.
- Osta, H., Qahwaji, R. & Ipson, S. (2008). Comparisons of feature selection methods using discrete wavelet transforms and support vector machines for mammogram images, *Systems, Signals and Devices, 2008. IEEE SSD 2008. 5th International Multi-Conference on*, pp. 1 –6.
- Paquerault, S., Petrick, N., Chan, H., Sahiner, B. & Helvie, M. (2002). Improvement of computerized mass detection on mammograms: Fusion of two-view information, *Medical Physics* 29: 238.
- Pedrin, H. & Schwartz, W. (2008). *Análise de imagens digitais: princípios, algoritmos e aplicações*, São Paulo: Thomson Learning .

- Ramakrishnan, S. & Selvan, S. (2007a). Image texture classification using wavelet based curve fitting and probabilistic neural network, *International Journal of Imaging Systems and Technology* 17(4): 266–275.
- Ramakrishnan, S. & Selvan, S. (2007b). Multiwavelets domain singular value features for image texture classification, *Journal of Zhejiang University-Science A* 8(4): 538–549.
- Ramos, R. & Nascimento, M. (2008). Comparação de extratores de características em imagens mamográficas., *XXI Congresso Brasileiro de Engenharia Biomédica* .
- Rangayyan, R. M., Ayres, F. J. & Desautels, J. L. (2007). A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs, *Journal of the Franklin Institute* 344(3-4): 312 – 348. Special Issue: Medical Applications of Signal Processing, Part I. URL: <http://www.sciencedirect.com/science/article/pii/S001600320600127X>
- Selvan, S. & Ramakrishnan, S. (2007). SVD-based modeling for image texture classification using wavelet transformation, *Image Processing, IEEE Transactions on* 16(11): 2688–2696.
- Susomboon, R., Raicu, D., Furst, J. & Johnson, T. (2008). A co-occurrence texture semi-invariance to direction, distance and patient size, *Proc. SPIE Medical Imaging*, Vol. 6914, Citeseer.
- Tang, J., Rangayyan, R. M., Xu, J., El Naqa, I. & Yang, Y. (2009). Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances, *IEEE Transactions on Information Technology in Biomedicine* 13(2): 236–251.
- Thurfjell, E. L., Lernevall, K. A. & Taube, A. A. (1994). Benefit of independent double reading in a population-based mammography screening program., *Radiology* 191(1): 241–244.
- Tsai, D. & Lee, Y. (2002). Fuzzy reasoning based computer-aided diagnosis for automated discrimination of myocardial heart disease from ultrasonic images., *Transactions* 84(12): 1431–1438.
- Vibha, L., Harshavardhan, G. M., Pranaw, K., Shenoy, P. D., Venugopal, K. R. & Patnaik, L. M. (2006). Classification of mammograms using decision trees, *Proceedings of the 10th International Database Engineering and Applications Symposium*, IEEE Computer Society, Washington, DC, USA, pp. 263–266. URL: <http://portal.acm.org/citation.cfm?id=1193213.1193866>
- Vieira, S. (2006). *Análise de variância: ANOVA*, Atlas.
- Warren, R. M. L. & Duffy, W. (1995). Comparison of single reading with double reading of mammograms, and change in effectiveness with experience, *Br J Radiol* 68(813): 958–962. URL: <http://bjr.birjournals.org/cgi/content/abstract/68/813/958>



## **Mammography - Recent Advances**

Edited by Dr. Nachiko Uchiyama

ISBN 978-953-51-0285-4

Hard cover, 418 pages

**Publisher** InTech

**Published online** 16, March, 2012

**Published in print edition** March, 2012

In this volume, the topics are constructed from a variety of contents: the bases of mammography systems, optimization of screening mammography with reference to evidence-based research, new technologies of image acquisition and its surrounding systems, and case reports with reference to up-to-date multimodality images of breast cancer. Mammography has been lagged in the transition to digital imaging systems because of the necessity of high resolution for diagnosis. However, in the past ten years, technical improvement has resolved the difficulties and boosted new diagnostic systems. We hope that the reader will learn the essentials of mammography and will be forward-looking for the new technologies. We want to express our sincere gratitude and appreciation to all the co-authors who have contributed their work to this volume.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Rogério Daniel Dantas, Marcelo Zanchetta do Nascimento, Ricardo de Souza Jacomini, Danilo César Pereira and Rodrigo Pereira Ramos (2012). Fusion of Two-View Information: SVD Based Modeling for Computerized Classification of Breast Lesions on Mammograms, Mammography - Recent Advances, Dr. Nachiko Uchiyama (Ed.), ISBN: 978-953-51-0285-4, InTech, Available from: <http://www.intechopen.com/books/mammography-recent-advances/fusion-of-two-view-information-svd-based-modeling-for-computerized-classification-of-breast-lesions>

# **INTECH**

open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.