



Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, LBP and SVM



Wener Borges de Sampaio^{a,*}, Aristófanés Corrêa Silva^a, Anselmo Cardoso de Paiva^a, Marcelo Gattass^b

^a Federal University of Maranhão - UFMA, Applied Computing Group - NCA Av. dos Portugueses, SN, Campus do Bacanga, Bacanga 65085-580 São Luís, MA, Brazil

^b Pontifical Catholic University of Rio de Janeiro - PUC-Rio R. São Vicente, 225, Gávea, 22453-900 Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Keywords:

Breast cancer
Computer-aided detection
Micro-genetic algorithm
Phylogenetic trees
Local binary patterns
Support vector machine

ABSTRACT

Breast cancer is the second commonest type of cancer in the world, and the commonest among women, corresponding to 22% of the new cases every year. This work presents a new computational methodology, which helps the specialists in the detection of breast masses based on the breast density. The proposed methodology is divided into stages with the objective of overcoming several difficulties associated with the detection of masses. In many of these stages, we brought contributions to the areas. The first stage is intended to detect the type of density of the breast, which can be either dense or non-dense. We proposed an adaptive algorithm capable of analyzing an image and telling if it is dense or non-dense. The first stage consists in the segmentation of the regions that look like masses. We propose a novel use of the micro-genetic algorithm to create a texture proximity mask and select the regions suspect of containing lesions. The next stage is the reduction of false positives, which were generated in the previous stage. To this end, we proposed two new approaches. The first reduction of false positives used DBSCAN and a proximity ranking of the textures extracted from the ROIs. In the second reduction of false positives, the resulting regions have their textures analyzed by the combination of Phylogenetic Trees, Local Binary Patterns and Support Vector Machines (SVM). A micro-genetic algorithm was used to choose the suspect regions that generate the best training models and maximize the classification of masses and non-masses used in the SVM. The best result produced a sensitivity of 92.99%, a rate of 0.15 false positives per image and an area under the FROC curve of 0.96 in the analysis of the non-dense breasts; and a sensitivity of 83.70%, a rate of 0.19 false positives per image and an area under the FROC curve of 0.85, in the analysis of the dense breasts.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Breast cancer is the most frequent type of cancer among the female population. It is also the type of cancer which most kills women (Parkin, Bray, Ferlay, & Pisani, 2005). The early diagnosis of this disease is the main form of fighting it.

A mammogram is a radiography of the breast, which allows the early detection of cancer, since it is capable of displaying lesions at their initial stage, having sizes in the range of millimeters. It is made through an appropriate X-ray device, called mammographer. The precision of the mammogram depends on several factors, such as size and location of the lesion, the density of the breast tissue and on the quality of the technical resources used. Besides, the task of carefully

interpreting a large number of cases demands time and an elevated degree of attention from the specialist physician.

According to Norman, et al. (Boyd et al., 2007), one of the factors that hinder the detection of masses by the specialists is the type of density of the breast, which can be dense (fibrous) or non-dense (fat). The fat tissue appears as a dark region in a mammogram. On the other hand, fibrous structures (including masses) appear as clearer regions of the mammogram. Due to these characteristics, it is more difficult for a radiologist to find lesions in dense breasts.

All these factors have motivated much research over the last decades, aiming at the development of computational systems to help the specialist physician in the task of interpreting radiological images. These Computer-Aided Detection (CAD)/Diagnosis (CADx) systems have gained more and more space in modern medicine, serving as a second information source for specialists and increasing the rates of correct detections in the identification of serious diseases, such as breast cancer (Fenton et al., 2007). However, most studies seen in the literature use the same techniques and configurations

* Corresponding author. Tel.: +559881607911.

E-mail addresses: wenersampaio@gmail.com, wenersampaio@yahoo.com.br (W.B. de Sampaio), ari@dee.ufma.br (A.C. Silva), paiva@deinf.ufma.br (A.C. de Paiva), mgattass@tegraf.puc-rio.br (M. Gattass).

both for dense and non-dense masses, whereas those techniques could be more appropriate for a specific kind of density.

The efficiency of CAD systems depend on the image processing techniques. The available literature brings acknowledged works that deal with the same problem approached by the methodology proposed herein, that is, the development of computational methods for aiding the specialist in the detection of lesions in mammograms.

The present bibliography contains a wide variety and combination of techniques intended to detect masses in mammograms. They normally employ techniques for describing the geometry and the texture of suspect regions.

For the geometric analysis, descriptors are used to find a shape pattern capable of representing and differentiating between masses and non-masses. Some examples of geometric descriptors are: area, perimeter, circularity, shape factor (Dong et al., 2015), eccentricity, circular density, circular disproportion and density (Sampaio, Moraes Diniz, Corrêa Silva, Cardoso de Paiva, & Gattass, 2011) active contour (Liu, Xu, Liu, & Feng, 2011), template matching (Nunes, Silva, & de Paiva, 2009), contourlet (Moayed, Azimifar, Boostani, & Katebi, 2010), generalized moment patterns (Deepak, Medathati, & Sivaswamy, 2012), density-weighted contrast enhancement (Petrick, Chan, Sahiner, & Wei, 1996), etc.

The vast majority of studies use some kind of texture analysis. These analysis usually employs statistical and geostatistical descriptors, diversity and richness-of-species indexes, with the objective of finding relations between the distribution of pixels belonging to masses and non-masses. Some examples of texture descriptors include: Local Binary Patterns (LBP) (Berbar, Reyad, & Hussain, 2012), Complete Local Binary Patterns (Liu et al., 2011), statistical fusion (Bajger, Ma, Williams, & Bottema, 2010), Gray Level Co-occurrence Matrix (GLCM) (Abdalla, Dress, & Zaki, 2011; Anitha & Peter, 2015; Tai, Chen, & Tsai, 2014), first-order and second-order statistical functions (Berbar et al., 2012; Jen & Yu, 2015), Gabor filters, image phase analysis, angular analysis of energy propagation, fractal analysis, Laws texture and Haralick descriptors (Banik, Rangayyan, & Desautels, 2011), gray-scale invariant ranklet texture (Masotti, Lanconelli, & Campanini, 2009), optical density transformation (Tai et al., 2014), pyramid decomposition (Lin, Chang, Yeh, Liu, & Yeh, 2014). Other descriptors are obtained by the Fourier and Wavelet transforms (Agrawal, Vatsa, & Singh, 2014; Kuo, Lin, Hsu, & Cheng, 2014; Lin et al., 2014), Phylogenetic Trees (Oliveira, Carvalho Filho, Silva, de Paiva, & Gattass, 2015), Vector Field Convolution (Dong et al., 2015).

The texture and geometry descriptors are, in general, used together with some machine learning technique, which will tell if the Region of Interest (ROI) analyzed belongs to a mass or to a non-mass. Some examples include the Support Vector Machine (SVM) (Agrawal et al., 2014; Berbar et al., 2012; Dong et al., 2015; Liu et al., 2011; Oliveira et al., 2015; Sampaio et al., 2011), Linear Discriminant Analysis (LDA) (Abdalla et al., 2011; Bajger et al., 2010), Artificial Neural Networks (ANN) (Abdalla et al., 2011; Banik et al., 2011; Lin et al., 2014) and k-Nearest Neighbors (KNN) (Berbar et al., 2012), Stepwise Discriminant Analysis (Tai et al., 2014), Particle Swarm Optimization (PSO) (Kuo et al., 2014).

Great efforts were joined to create image databases for open use by the scientific community. Among these databases, we highlight the Digital Database for Screening Mammography DDSM (Abdalla et al., 2011; Bajger et al., 2010; Berbar et al., 2012; Dong et al., 2015; Jen & Yu, 2015; Liu et al., 2011; Oliveira et al., 2015; Sampaio et al., 2011; Tai et al., 2014), and the Mammographic Image Analysis Society (MIAS) (Agrawal et al., 2014; Anitha & Peter, 2015; Deepak et al., 2012; Dong et al., 2015; Kuo et al., 2014; Moayed et al., 2010). Some studies use private databases with the same purpose (Bajger et al., 2010; Banik et al., 2011; Jen & Yu, 2015; Lin et al., 2014).

Analyzing the studies numbered above, we notice that most of them use the MIAS and DDSM databases. Nevertheless, none of them tries to use a methodology adapted to the density of the breast.

The cited works, in general, use supervised learning techniques to perform the classification of the ROIs. The SVM classifier, which got much attention, was also used in this work.

This work used two stages (segmentation and reduction of false positives using DBSCAN), which use information of the from the image under analysis, dismissing the use of a specific knowledge base for each stage.

The work by Oliveira et al. (2015) also extracts features using phylogenetic trees. However, it only takes into account the frequency of the gray levels of the ROIs under analysis. In our work, the phylogenetic trees use the spatial relation among gray levels, as well as their occurrences, thus performing both a local and a global analysis. We must stress that the work by Oliveira et al. (2015) did not apply the stages of segmentation and reduction of false positives, and for this reason had a superior performance, since classification errors in these stages were not computed.

Opposite the present work, none of the cited work use a methodology in an attempt to optimize their training models.

Table 17 summarizes the related works presented in this section.

This work presents a CAD methodology for helping the specialist physician in the task of detecting masses in mammographic images. The technique is adapted to the density of the breast. The first stage of the methodology removes artifacts outside the breast and makes the reduction of noise. The next stage classifies the density of the breast as dense or non-dense. The segmentation chooses the regions of the image that probably contain masses by means of a micro-genetic algorithm (μ GA). A first reduction of false positives (RFP) uses the Density Based Spatial Clustering of Applications with Noise (DBSCAN) and a proximity ranking of the textures extracted from the regions of interest (ROI). In the second RFP, the ROIs have their textures analyzed by the combination of Phylogenetic Trees, Local Binary Patterns and Support Vector Machines (SVM).

The next sections can be briefly described as follows. Section 2 presents the bibliographic review, containing the theoretical basis necessary for understanding the methodology. In Section 3 we present the five stages that compose the present work: image acquisition, density detection, pre-processing, segmentation of the regions of interest and reduction of false positives. In Section 4, we present and discuss the results. We also present case studies of the application of the proposed methodology. Section 5 concludes the work, presenting the main contributions and the efficiency of the methods used.

2. Theoretical basis

This section brings a quick review on Phylogenetic Trees, Local Binary Patterns (LBP), taxonomic diversity and distinction indexes and Micro-genetic Algorithm.

2.1. Phylogenetic trees and taxonomic indexes

Phylogenetic trees are used in Biology to describe the evolutionary relation among species. In these trees, the leaves represent the species and the internal nodes represent the common ancestors between the species. So, it is possible to relate the species under study from an evolutionary standpoint. The inclined cladogram is a graphic representation used to describe the phylogenetic relation between ancestor species (Hall & Greenstreet, 1998). These trees allow the extraction of indexes that relate the diversity, richness and parenthood between species.

Fig. 1 presents an example of ape phylogenetic tree, represented by an inclined cladogram, where one may notice that a chimpanzee has higher phylogenetic proximity to humans than a siamang. In this tree, the leaf nodes are the species analyzed, the internal nodes correspond to a common ancestor and the edges indicate the phylogenetic distance between both species.

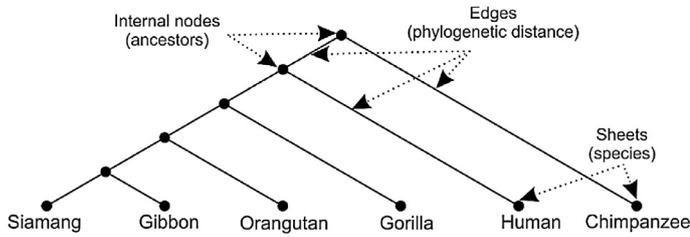


Fig. 1. Example of ape phylogenetic tree.

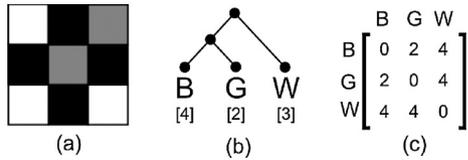


Fig. 2. Use of phylogenetic trees for analysis of texture. (a) Image with 3 gray levels. (b) Representation of a phylogenetic tree by inclined cladogram. (c) Representation of a phylogenetic tree by a matrix.

By means of phylogenetic trees, we can compute two taxonomic indexes that relate species belonging to a community. For these indexes, we must consider x_i and x_j as the number of individuals of the species i and j , respectively. The variable w_{ij} is the phylogenetic distance (weight) between the species i and j , that is, the number of edges that separate both species in the cladogram, and n is the total number of individuals in a certain area.

The first index is the taxonomic diversity (Δ), which considers the abundance of species and the taxonomic relation among them. So, its value expresses the mean taxonomic distance between any two individuals randomly chosen from the sample. The index (Δ) can be computed by Eq. (1).

$$\Delta = \frac{\sum \sum_{i < j} w_{ij} x_i x_j}{[n(n-1)/2]} \tag{1}$$

The second index is the taxonomic distinction (Δ^*), which represents the mean taxonomic distance between two individuals, provided that they belong to different species. It can be computed by means of Eq. (2).

$$\Delta^* = \frac{\sum \sum_{i < j} w_{ij} x_i x_j}{\sum \sum_{i < j} x_i x_j} \tag{2}$$

In image processing, we can use phylogenetic trees to describe a texture. In this case, the species are the gray levels. Fig. 2(a) contains an example of texture that has only three gray levels (black, gray and white). Fig. 2(b) contains a phylogenetic tree in the form of an inclined cladogram of the gray levels of Fig. 2(a), in which each subscript value corresponds to the number of individuals of the species. Fig. 2(c) contains the representation of the inclined cladogram of Fig. 2(b) in matrix form.

In this work, the phylogenetic trees are used to describe the texture of ROIs where all the species analyzed are LBP codes. The

objective is the classification of these ROIs into mass and non-mass by the SVM.

2.2. Local binary patterns

The Local Binary Patterns (LBP) turn any given texture, defined by a vicinity of radius R and formed by P neighbors, into a code by means of a thresholding (Ahonen, Hadid, & Pietikainen, 2006). Let I be an image and g_p the gray level of a pixel located in a circular region of P neighbors around the point $c = (y_c, x_c)$ within a distance R . The P pixels neighbor of the point c can be determined through Eqs. (3) and (4), in which $p = 0, 1, \dots, P - 1$ (Ahonen et al., 2006).

$$x_p = x + R \cos(2\pi p/P) \tag{3}$$

$$y_p = y - R \sin(2\pi p/P) \tag{4}$$

where y_p and x_p are the coordinates of the pixel p , neighbor of the point $c = (y_c, x_c)$, R is the radius and P is number of neighbors analyzed.

The binary sequence L_p is obtained through a thresholding of the elements P , where the threshold is the value of the central pixel p . Writing the thresholding function as $s(z)$, where z is the difference between the gray levels of the central pixel and a pixel of the vicinity, the thresholding function is defined as (Ahonen et al., 2006):

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} \tag{5}$$

The computation of the pattern is made through the summation of the thresholding of the vicinity, weighted by the square of the order of each neighbor. Eq. (6) defines the LBP operator (Ahonen et al., 2006).

$$LBP_{P,R}(y_c, x_c) = \sum_{p=0}^{P-1} s(g_p, g_c) 2^p \tag{6}$$

In this work, the LBP codes are used as the species analyzed by phylogenetic trees to describe the texture of segmented ROIs. The use of LBP is intended to provide the analysis of texture with information about the spatial relationship between neighbor pixels in a ROI (Fig. 3).

2.3. Micro-genetic algorithm

The Micro-genetic Algorithm (μ GA) is a kind of evolutionary algorithm whose main characteristic is the reduced size of the population in each generation (Tam, Cheng, & Lui, 2006). Since the population size is reduced, the genetic diversity tends to decrease rapidly, but when this occurs, the μ GA restarts the population, keeping just the elite unchanged. This feature allows the μ GA to avoid the early convergence and the production of sub-optimal results.

In the present work, the μ GA was used to create the best training model for the SVM classifier at the stages of breast density detection and training, as well as to create texture proximity masks in the segmentation.

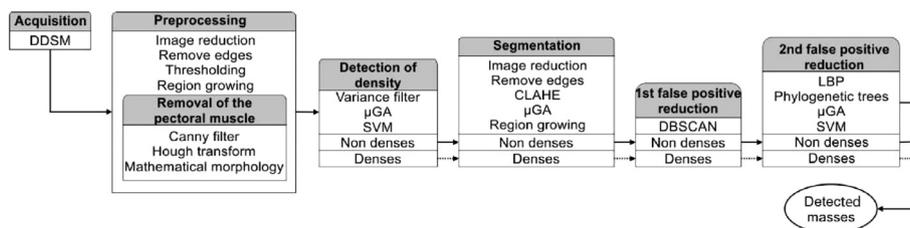


Fig. 3. Thorough representation of the methodology.

Table 1
Density definitions according to BI-RADS.

Density	Description	Note
1	Breast totally filled with fat.	Absence of fibrous tissue.
2	Presence of dispersed fibroglandular tissue	May hide a lesion in a mammogram.
3	The breast is heterogeneously dense.	May reduce the sensitivity of the exam.
4	The breast tissue is extremely dense.	Reduces the sensitivity of the exam.

3. Proposed methodology

The rest of this section describes each one of these stages in detail.

3.1. Acquisition

For the development of this methodology, we used the database of digitized mammograms called Digital Database for Screening Mammography (DDSM), available on the Internet. The DDSM is a public database containing 2620 cases, made freely available (Heath et al., 1998) through the efforts of some American institutions (Massachusetts General Hospital, Wake Forest University, and Washington University in St. Louis School of Medicine).

Each case contains two images of each breast in the craniocaudal (CC) and mediolateral oblique (MLO) projections, besides extra information about the exam (breast density, date of the study, patient's age, type of pathology, number of anomalies, etc.) and about the image (filename, film type, date of scanning, scanner type, sequence, pixels per line, bits per pixel, lesion location, etc.).

All the information contained in the DDSM was supplied by specialist physicians (Heath et al., 1998). The information about density of the breast is present in the DDSM, and can have values ranging from 1 to 4. These values can be interpreted using the definitions indicated by the BI-RADS, as described in Table 1.

In this work, we used 1727 images from the DDSM database, being 1049 images of non-dense breasts (values 1 and 2 in Table 1) and 678 images of dense breasts (values 3 and 4 in Table 1). These images must contain at least one mass lesion. Images containing other types of lesion are not part of this study.

3.2. Pre-processing

Images belonging to the DDSM database present noise and typical elements of a mammogram, which may interfere in the results of the proposed methodology. These undesired elements include identification marks of the patient or of the type of exam, pixels of the image background, and eventual noises produced by imperfections in the generation or scanning processes of the image. The objective of this pre-processing stage is to remove these unwanted elements. Example of such elements are shown in Fig. 4.

The first step of the pre-processing stage is a reduction of the original image size, resulting in an image with height of 1024 pixels. The same reduction proportion is made for the width. The objective of this reduction is to reduce the processing time, which is too high when one works with the image in its original size, that is, a mean height of 6000 pixels. This reduction does not imply in loss of information. The works by Sampaio et al. (2011), Nunes et al. (2009) are some examples in the literature which used image size reduction without loss of information or performance.

Next, all points within a distance of 30 pixels from the vertical border and 60 pixels from the horizontal borders are removed (Fig. 5(a)). This removal is intended to remove from the image the area between the borders of the film and the empty spaces without the film.

The background, with gray levels close to black, is removed with use of the thresholding algorithm. First, the image is divided in half,

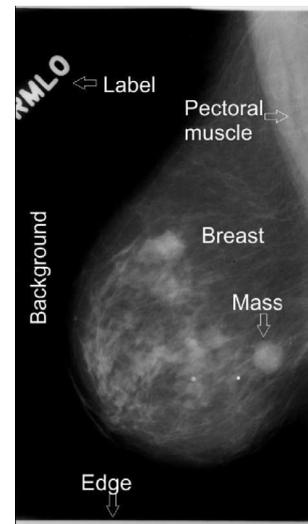


Fig. 4. Example of DDSM image with the typical elements found.

according to the white line in Fig. 5(b). Then, we compute the sum of all gray levels values of each half of the image. Since the gray levels close to black have values close to zero, then the size with the smallest sum contains the background portion, and is the opposite to the breast side.

The side containing the background is divided in half, according to the gray line in Fig. 5(c). With the slice closer to the border (and opposite to the breast), we apply the mean filter (Gonzalez, Woods, & Masters, 2009) for noise removal, and compute the mean (μ) and the standard deviation (σ) of the gray levels.

Let $L = \mu + \sigma$, if L is below 50, then we use L as threshold (Gonzalez et al., 2009). Otherwise, we use μ (values empirically determined). This thresholding serves to separate the background region, in such a way to preserve the contour of the breast.

With the thresholding, the more intense pixels, like the pixels of the breast and of the identification marks, are grouped into a cluster (white) and the less intense ones, like those of the background and noise, are grouped into another cluster (black). The group of pixels that correspond to the background is eliminated from the image with texture. The visual representation of the clusters created can be seen in Fig. 5(e).

In the resulting image, there are still some objects outside the breast, such as identification labels of the exam and some noise. These objects are removed by the region growing algorithm (Gonzalez et al., 2009), which separates the non-connected regions of the image (Fig. 5(f) and (g)). Only the region with the largest area, that is, the breast, is selected (Fig. 5(g)).

The MLO images usually contain the chest muscle, which is associated to high gray levels. This characteristic interferes in the frequency of the gray levels of the image. So, all MLO images were submitted to a new pre-processing for removal of the chest muscle, such as in the work by Sampaio et al. (2011).

At this point of the methodology, all the techniques were equally applied, both for dense and non-dense breasts. However, the proposed methodology will be divided into two processing stages.

The training and validation stage (Sections 3.3–3.5.2) uses a base with 80% of the 1049 non-dense images (839 images, with 888 masses) and 80% of the 678 dense breast images (543 images, with 562 masses) chosen in Section 3.1. Based on these images, we estimate empirically the parameters for the techniques used. Also based on these images, we estimate, empirically, the training models for optimization of the results of the classification in the stages of density detection and training/validation/test.

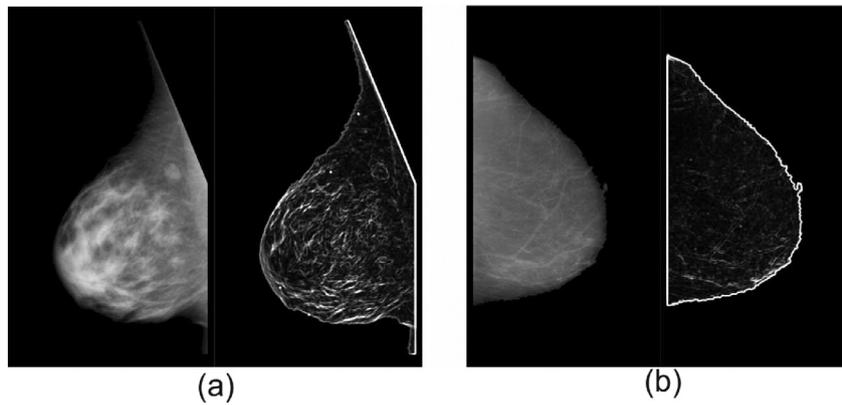


Fig. 7. Application of the variance filter. (a) Dense breast. (b) Non-dense breast.

Table 2
Number of vectors used to create the density training model.

Density	Total	Training (70%)	Validation (20%)	Not selected (10%)
Non-dense	839	587	168	84
Dense	543	380	109	54

3.3.3. Training and validation

The previous stage produces a set of feature vectors extracted from non-dense and dense breast images. These feature vectors will be used to determine an optimized training model by the μ GA. In order to create the (density) training model, it is necessary to divide the set of vectors in the proportion of 70% for training and 20% for validation. The remaining 10% are not selected. Table 2 contains the distribution of the number of vectors used to create the density training model.

Each chromosome is randomly formed by two sub-samples. The first sub-sample contains 70% of the feature vectors created, that is, they were extracted from 587 non-dense breast images and 380 dense breast images. The SVM will be trained with these vectors.

The second sub-sample contains 20% of the feature vectors, that is, they were extracted from 168 non-dense breast images and 109 dense breast images. These vectors will be classified by the SVM (validation).

The sub-sample formed by 10% of the feature vectors, which were not selected by the μ GA (84 vectors extracted from non-dense breast images and 54 vectors extracted from dense breast images), will be dismissed, since they degrade, have little or no contribution for the enhancement of the validation performance.

It is important to observe that the parameters used by the SVM, in this and in the other stages, were automatically determined according to (Chang & Lin, 2011).

The μ GA created to optimize the breast density detection has the following characteristics:

- Each generation has 10 chromosomes.
- The selection is made through the roulette wheel technique.
- The elite is formed by one chromosome.
- The crossover between two chromosomes A and B exchanges the features that they do not share. A vector can only be exchanged with another vector of the same sub-sample type (training with training, validation with validation). Each chromosome randomly changes between 1 and 5 of its vectors.
- The mutation of a chromosome A randomly replaces between 1 and 5 of its vectors with new ones which do not belong to A.
- In the cases of precocious convergence, the population is restarted, keeping only the elite unchanged. The early conver-

Table 3
Confusion matrix of the validation.

		Density	
		Non-dense	Dense
Classification	Non-dense	168	0
	Dense	0	109

gence is detected when 60% or more of the population has the same fitness as the fittest individual of the current generation and when the number of past generations is smaller than 50. These values were empirically defined, so that the initial number of repetitions is high, thus improving the scans in the feature space.

- The fitness of a chromosome is computed through the arithmetic mean of sensitivity, specificity and accuracy found in the classification of the validation sub-sample. The smaller the mean, the fittest will be the individual, and consequently better will be the correct detection of the dense or non-dense breasts.
- The stop criterion is reached when the fitness of the fittest individual is repeated for 100 consecutive generations.

According to (Ng, 2008) sensitivity, specificity and accuracy are statistics commonly used to indicate the efficiency of a certain test to indicate the presence or absence of a disease. Sensitivity evaluates how good a test is at indicating the presence of a disease. Specificity estimates the probability of the presence of a disease being dismissed. Accuracy measures the correctness of the identification or exclusion of a certain condition. Some examples of the application of the measurements in CAD systems are found in Sampaio et al. (2011) and Junior, de Paiva, Silva, and de Oliveira (2009).

Table 3 contains a confusion matrix of the validation, with the creation of the training model optimized by the μ GA. In this table, one can notice that the density training model achieved 100% sensitivity, specificity and accuracy.

The SVM uses a sub-sample of the training vectors to generate the separation surface between the classes under analysis. These vectors are known as support vectors.

A training using SVM suffers from overfitting when the number of support vectors used to generate the separation surface gets closer to the total number of vectors used in the training. In other words, the SVM loses its generalization capability. From the 967 feature vectors used in the training (Table 3), only 319, that is, 33% of them, were used as support vectors. So, we can state that there was no overfitting in the training.

The training model generated will be used in the detection of the breast density in the training set.

The next stages adjust the parameters of the techniques according to the density of the image under analysis. Table 4 contains the

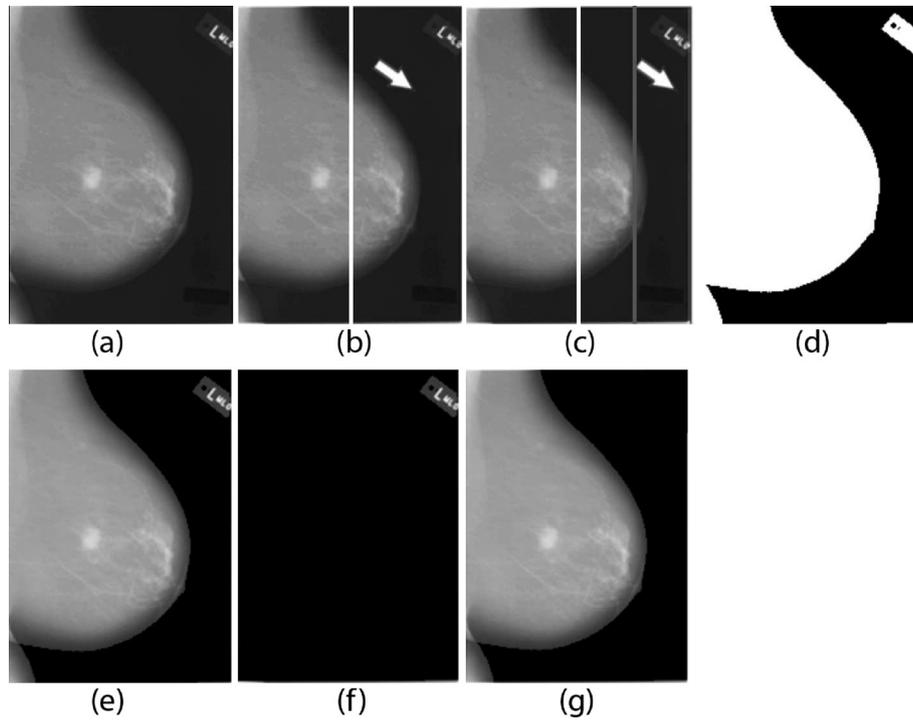


Fig. 5. Removal of background and noise. (a) Reduced image. (b) The arrow indicates the half of the image containing the background, the white line indicates the half of the image. (c) The gray line divides in half the part of the image containing the background. The arrow indicates the slice on which the threshold is computed. (d) Result of the thresholding. (e) Background removal. (f) and (g) Non-connected objects separated by the region growing algorithm. (g) Result of the pre-processing.

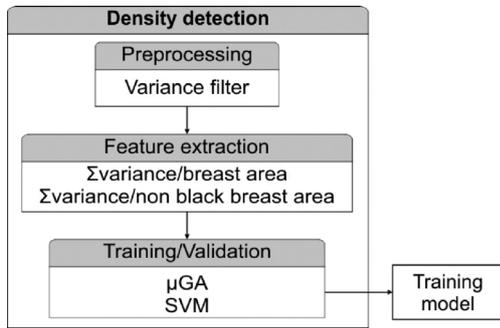


Fig. 6. Breast density detection scheme.

The test stage is described in Section 4.1. It is formed by 20% of the 1049 images of non-dense breasts (120 images, with 214 masses) and 20% of the images of dense breasts (135 images of dense breasts, with 135 masses) which did not participate in the training and validation stage.

3.3. Detection of breast density

The objective of this stage is to classify the breast either as dense or as non-dense. This is also a crucial stage, since it determined the use of distinct configurations for dense and non-dense breasts. Fig. 6 presents the stages of density detection.

3.3.1. Pre-processing

The pre-processing consists in the application of a variance filter (Gonzalez et al., 2009) of radius $r = 5$. The application of this filter is justified by the structural differences between dense and non-dense breasts, as observed in Fig. 7. After applying the variance filter, all of the computed values are scaled into the integer interval from 0 to 255.

Fig. 7(a) contains an example of dense breast and the application of the variance filter. It can be noticed that dense breasts have a large variation in the distribution of the gray tones. This implies directly in the higher number of structures highlighted by the variance filter.

Fig. 7(b) shows an example of non-dense breast and the application of the variance filter. The images of non-dense breasts have a more uniform distribution of gray tones, which implies in a smaller number of structures highlighted by the variance filter.

After the application of the variance filter, occurs the extraction of features.

3.3.2. Features extraction

Features extraction consists in taking two measurements that relate variance and the area of the images resulting from the variance filter.

Let v_{yx} be the scaled value of the variance for the pixels p_{yx} , belonging to the breast region, and A_m the area that contains only the breast. The first feature, internal variance ($V_{internal}$), consists in summing all the v_{yx} and dividing by A_m (Eq. (7)).

$$V_{internal} = \frac{\sum_y \sum_x v_{yx}}{A_m} \tag{7}$$

Let A_v be the area formed by the pixels of the breast whose values of v_{yx} are not zero. The second feature, mean variance (V_{mean}), consists in summing all the v_{yx} and dividing by A_v (Eq. (8)).

$$V_{mean} = \frac{\sum_y \sum_x v_{yx}}{A_v} \tag{8}$$

The division by the area (A_m or A_v) is justified by the existence of breasts of various sizes. So, a small and dense breast area (higher variance) could be confused with a big and dense breast (smaller variance) if we considered just the sum of the variances. From each image, we extract a features vector formed by these two features.

Table 4
Parameters used in segmentation and first false positives reduction.

Technique	Non-dense breasts	Dense breasts
Segmentation		
Mean filter (radius)	1 pixel	1 pixel
CLAHE (radius)	64 pixels	32 pixel
CLAHE (gray levels)	256	256
CLAHE (overlapping)	5 pixels	3 pixels
uGA (stop criterion)	30 repetitions of the elite	50 repetitions of the elite
uGA (mask size)	5 × 5	9 × 9
uGA (threshold)	110	90
1st false positive reduction		
Window size	32 × 32	32 × 32
DBSCAN (Eps)	5	3
DBSCAN (MinPts)	1	1

parameters used to segment ROIs and reduce false positives in dense and non-dense breasts.

3.4. Segmentation of the ROIs

This stage is indented to identify the regions of the breast with higher possibilities of being masses. In this work, the segmentation is done by the μ GA, in order to create the mass candidates.

Next we describe the starting steps of the segmentation of ROIs by means of the μ GA. These steps serve to prepare the image for the segmentation processes. They are:

- The height and the width of the analyzed image is reduced in half (height of 512 pixels). This occurs because the evolutionary methodologies, in general, work with several generations, so, this reduction reduces the total processing time necessary. In general, there is not much loss of information in images with 1024 pixels in height, and its copy reduced to 512 pixels in height did not led to great changes in the outcome of the segmentation, except for the time consumption.
- Application of the mean filter with radius $r = 1$ to reduce the noise in the image.
- A CLAHE contrast enhancement is applied (Ferguson, Arslan, Erdogan, & Parmley, 2008), in order to improve the visual distinction of the structures inside the breast.

The configurations of radius, size of the pixel overlapping adopted in CLAHE, stop criterion, proximity mask size and threshold used by the μ GA prevents the segmentation of masses with excess of extra pixels, in both dense and non-dense breasts.

After that, the μ GA is used to generate an optimized texture proximity mask. A chromosome is defined by a mask M (matrix) of size $N \times N$.

In order to compute the mask values, we first compute the mean (μ) and the standard deviation (σ) of the gray levels of all pixels of the resized image, not taking the background into account. Each element M_{ij} of the mask receives a value which is the sum of μ and a random value in the interval $[-\sigma, +\sigma]$. This serves to create a mask which is as close as possible to the texture of the breast under analysis, so we can say that M is a texture proximity mask.

To compute the fitness of a chromosome, we use all windows, without pixel overlapping, that do not contain background. Fig. 12 contains a similar example which uses 32×32 windows (to improve visualization).

For a better understanding, it is necessary to provide a better explanation of the concept of texture proximity. The texture proximity between a window centered at the pixel p_{yx} and the chromosome is obtained by the sum of the absolute values of the differences between all the pixels of the window and their counterparts in the chromosome.

The fitness of the chromosome is computed by the mean of the distances of all windows, regardless of the background. The smaller the fitness of a chromosome, smaller will be its mean distance with most part of the breast texture contained in the image, that is, the texture of the mask resembles the texture of the majority of the windows of the breast. The objective of the μ GA is to minimize this distance. Fig. 8 contains an example of the computation of the proximity between a chromosome and a 3×3 window.

Fig. 8(a) is the reduced image. Fig. 8(b) is the proximity mask (chromosome) created by the mean (103) and the standard deviation (3) computed in Fig. 8(a). Fig. 8(c) contains the absolute values of the differences between the corresponding elements of the window and of the proximity mask. Fig. 8(d) contains the sum of the values of Fig. 8(c). This value does not constitute the fitness of the chromosome yet, since it is just the distance between the window and the mask. The computation of the fitness must use the means of the distances between all the windows of the image.

The configurations of the μ GA used to create an optimized texture proximity mask are:

- Each generation has 10 chromosomes.
- The selection is done by the roulette wheel technique.
- The elite is formed by one chromosome.
- The crossover between two chromosomes A and B is produced by the arithmetic mean of the elements of the matrices of A and B.
- The mutation of a chromosome A chooses a line of the proximity matrix A and adds a random value between $[-5; +5]$ to each element of the chosen line. The final value of each element must be in the interval $[0, 255]$.
- In the case of precocious convergence, the population is restarted, keeping only the elite unchanged. The early convergence is detected when 60% or more of the population has the same fitness as

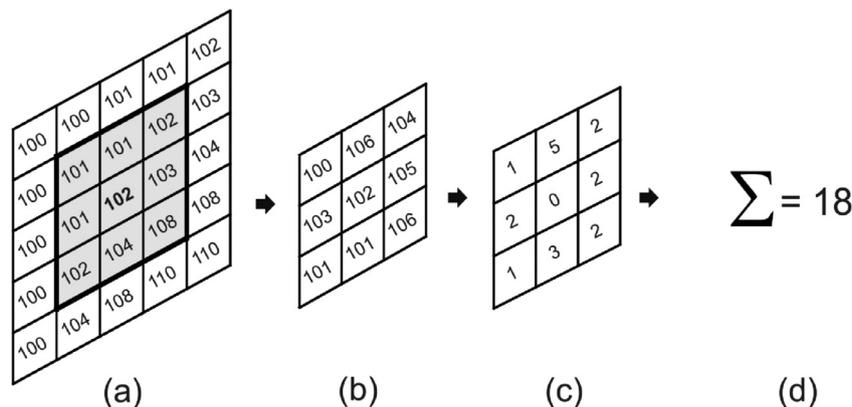


Fig. 8. Computation of the texture proximity. (a) Reduced image. (b) Proximity mask (chromosome). (c) Difference between (a) and (b). (d) Sum of the values of (c).

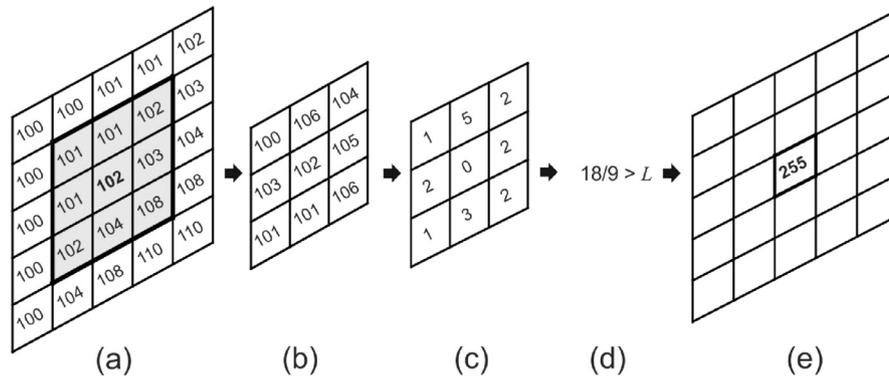


Fig. 9. Segmentation of a pixel. (a) Reduced image. (b) Proximity mask found by the μ GA. (c) Difference between (a) and (b). (d) Comparison with the threshold. (e) Segmented point.

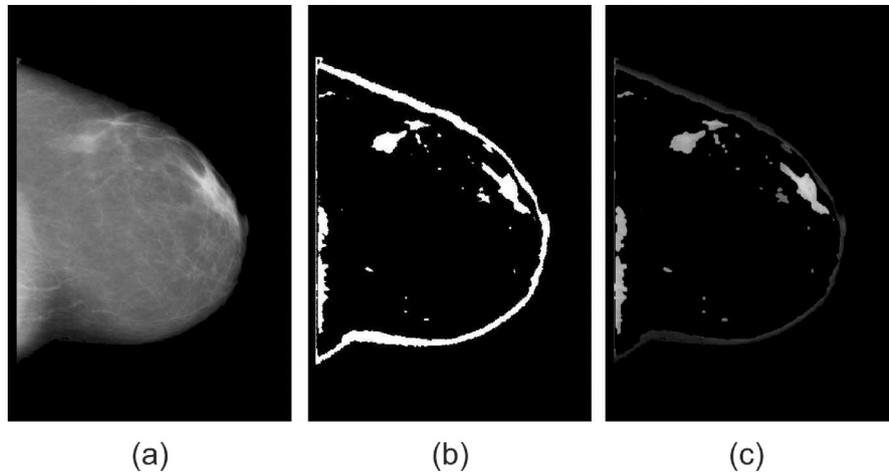


Fig. 10. Example of segmentation in a non-dense breast. (a) Input image. (b) Suspect regions. (c) Image resulting from the segmentation.

the fittest individual of the current generation and when the number of past generations is smaller than 50. These values were empirically defined, so that the initial number of repetitions is high, thus improving the scans in the feature space.

- O stop criterion is reached when the fitness of the fittest individual is repeated for 30 consecutive generations for non-dense breasts, and 50 consecutive generations for dense breasts.

With the texture proximity mask found by the μ GA, we carry out the segmentation itself.

The μ GA creates a proximity mask which resembles the majority of the texture of the breast (commonest texture, supposedly with no lesion). The regions having texture very different from the rest of the texture of breast are considered suspect of having masses. In other words, the segmentation consists in eliminating the areas of the breast whose textures are closer to the proximity mask, remaining only those with very different texture.

For all the p pixels of an image I , belonging to the breast, if the distance between the pixels of a window centered in p_{yx} and the proximity mask found is above a threshold L , then p_{yx} belongs to a region suspect of being a lesion (different from the majority of breast texture), and will get the value 255, otherwise will receive value 0 (zero). After this procedure, the size of the image is restored to the size before the segmentation, as well as the textures of the regions marked as suspect.

Fig. 9 presents an example of segmentation of the pixel highlighted in Fig. 9(a). Fig. 9(b) is the mask which represents the fittest chromosome found by the μ GA. Fig. 9(c) contains the differences between Fig. 9(a) and (b). Fig. 9(d) compares the mean of the sum of the

Table 5
Number of ROIs produced by the segmentation in the training and validation stage.

Type of density	Masses		Non-masses segmented		Proportion Non-masses/masses	
	Segmented		Lost			
	Total	%	Total	%		
Non-dense	865	97.41	23	2.59	9613	11.11
Dense	531	94.48	31	5.52	6933	13.06

values of Fig. 9(c) with a threshold. Fig. 9(e) contains the segmented pixel.

Fig. 10 contains an example of non-dense breast which was submitted to the segmentation process, in which Fig. 10(a) is the input image resulting from the pre-processing, Fig. 10(b) contains the regions marked as suspect and binarized, and Fig. 10(c) is the image resulting from the segmentation process, showing the suspect regions and their textures.

The final stage of the segmentation consists in isolating the suspect regions so that they can be individually analyzed in the next stage. Each ROI is separated by a region growing algorithm.

In this stage, we used 839 non-dense breasts images, with 888 masses; and 543 dense breasts images, with 562 masses. From the non-dense breasts, the segmentation extracted 865 ROIs from masses (97.41% of the total masses), and 9613 ROIs from non-masses. From the dense breasts, the segmentation extracted 531 ROIs from masses (94.48% of the total masses), and 6933 ROIs from non-masses. Table 5 contains the results of the segmentation for this stage.

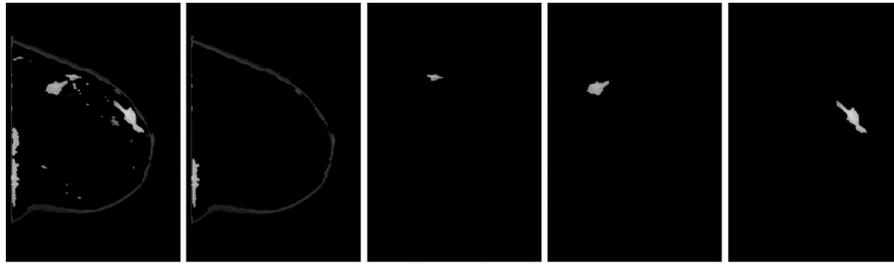


Fig. 11. Examples of ROIs separated by the region growing algorithm.

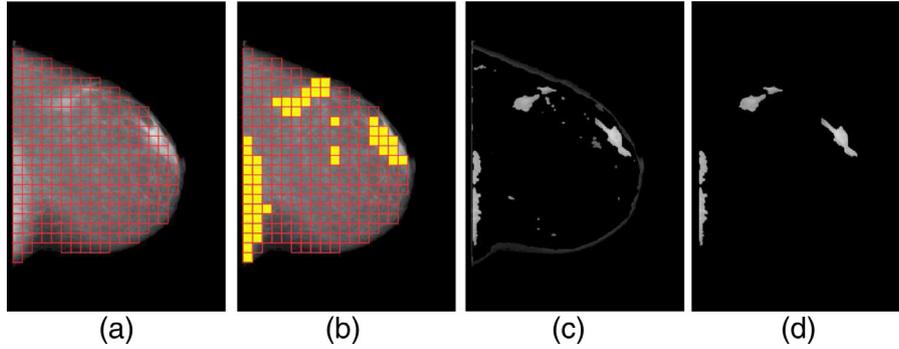


Fig. 12. Stages of the first reduction of false positives. (a) All the internal 32×32 windows in the breast region. (b) Windows removed from the windows list highlighted. (c) ROIs generated by the segmentation. (d) ROIs resulting from the reduction of false positives.

The segmentation achieved good results for both dense and non-dense breasts. However, the number of segmented non-masses is between 11.11 and 13.06 times bigger than the number of segmented masses. So, it is necessary to reduce the number of segmented non-masses. Fig. 11 contains examples of ROIs separated by the region growing algorithm.

3.5. Reduction of false positives

The objective of the segmentation is to generate ROIs. However, in some cases occurs a phenomenon known as super segmentation, which consists in the excessive segmentation of suspect regions. That is, besides marking areas that contain masses, a very large number of healthy regions is marked as lesions (false positives), which elevates the number of ROIs to be analyzed in the next stages. So, it is necessary to reduce the number of segmented false positives.

3.5.1. Reduction of false positives by DBSCAN

The first reduction of false positives consists in eliminating ROIs erroneously marked as lesion, in both dense and non-dense breasts.

For an image I , we create a list of all non-overlapped 32×32 -pixel windows that do not contain the background and that are on the region of the breast (Fig. 12(a)). We compute the mean μ_T (total mean) and the standard deviation σ_T (total standard deviation) of the gray levels of these windows.

The windows with mean above $\mu_T + \sigma_T$ are removed from the windows list (Fig. 12(b)), that is, windows that differ too much from the listed windows. This way, only the windows that have the gray levels of the majority of the windows in the internal area of the breast remain in the list.

With the windows that are not in the list yet, we compute μ_R , mean of the remaining windows. This mean will serve as distance metric for comparison of the segmented ROIs of I .

Let R_I be the set of ROIs of the image I (Fig. 12(c)), and r a ROI belonging to R_I . A ranking value is assigned to each ROI r of R_I as follows:

$$ranking_r = |\mu_R - \mu_r| \tag{9}$$

Table 6

Number of ROIs resulting from the reduction of false positives, using DBSCAN, in the training and validation stage.

Density	Masses				Non-masses		
	Previous stage	Selected	Lost		Previous stage	Remaining stage	
		Total	%	Total			%
Non-dense	865	843	94.93	45	5.07	9613	3197
Dense	531	522	92.88	40	7.12	6933	2099

where μ_r is the mean of the gray levels of the ROI r . This value informs the distance between the gray levels of a ROI and the gray levels of the image to which it belongs. The ROIs with higher ranking are the ones which most differ from the mean of the gray tones of the image, with a high chance of being lesions.

Next, we use DBSCAN to generate clusters of ROIs based on the analysis of the rankings. For non-dense breasts, the value of Eps is 5 and $MinPts$ is 1. For dense breasts, the value of Eps is 3 and $MinPts$ is 1. These are the best values empirically found.

After clustering, we analyze all the ROI groups formed. The group with the highest mean ranking is passed to the features extraction stage; the remaining ones are discarded (Fig. 12(d)).

Table 6 contains the results of the reduction of false positives of this stage.

The reduction false positives achieved a good performance in the training and validation stage, since the reduction of the non-masses in non-dense breasts was of 66.74% and 69.72% of non-masses in dense breasts.

The ROIs which were not discarded in the reduction false positives will have their texture features extracted in the second reduction of false positives.

3.5.2. Reduction of false positives by Phylogenetic Trees, LBP and SVM

The second reduction of false positives consists in analyzing the remaining ROIs by the combination of phylogenetic trees and LBP

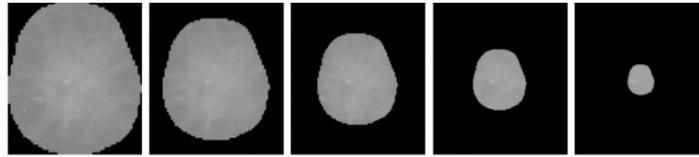


Fig. 15. Example of regions with the use of internal masks.



Fig. 16. Example of regions with the use of external masks.

The second sub-sample contains 20% of the feature vectors created, that is, were extracted from 178 segmented ROIs of non-dense breasts and 112 segmented ROIs of dense breasts. These vectors will be classified by the SVM (validation).

The sub-sample formed by 10% of the feature vectors which were not used will be dismissed, since its elements degrade or do not contribute to the enhancement of the performance of the validation.

The sub-sample of 10% are cases that contain noises or redundant information. So, they are ignored in the generation of an optimized training model. Figs. 17 and 18 illustrate cases of noise and redundancy.

Fig. 17(a) shows the original DDSM image. Fig. 17(b) shows the marking of a lesion. Fig. 17(c) contains the ROI used to generate a feature vector. This is an example of ROI that does not contribute to the performance of the training, since the texture information cannot be efficiently used to distinguish between mass and non-mass, due to the presence of noise.

Fig. 18(a) shows the original DDSM image. Fig. 18(b) shows the marking of a lesion. Fig. 18(c) shows the ROIs used to generate two feature vectors. This is an example of ROIs that have very similar textures, thus having redundant information.

Notice that the sub-sample of 10% does not prevent cases like these from being included in the optimized training model, that is,

the training/validation is not free from noises and ambiguities. Such cases are simply less likely to be selected. This rationale is applied in the same manner as in Section 3.3

The configurations of the μ GA used to optimize the reduction of false positives are:

- The crossover of two chromosomes A and B exchanges from 1 to 5 vectors between A and B. A vector can be exchanged only with another one of the same sub-sample type (training with training, validation with validation).
- The mutation of a chromosome A randomly replaces from 1 to 5 of its vectors with new ones which are not already in A.
- The population restart keeps only the elite.
- The fitness of a chromosome is computed by the simple arithmetic mean of sensitivity, specificity and accuracy, found in the classification of the validation sub-sample.
- The stop criterion is reached when the fitness of the fittest individual is repeated for 100 consecutive generations. These values were found empirically. In general, good results were achieved with 44–55 repetitions of the best fitness on average. So, in order to make sure that the feature space was well explored and that the classifier had not found a local sub-optimal, we double (almost) the number of repetitions of the best fitness. This is the same procedure applied in Section 3.3

The validation stage achieved good results. According to the Table 9, the analysis by circular sub-regions presented the worst performances, with sensitivity of 88.76%, specificity of 91.86% and accuracy of 91.21%. The analysis by internal masks produced the best result, with a sensitivity of 99.41%, specificity of 97.97% and accuracy of 98.27%.

The analysis by internal masks had the best result, with a sensitivity of 99.41%, specificity of 97.97% and accuracy of 98.27%.

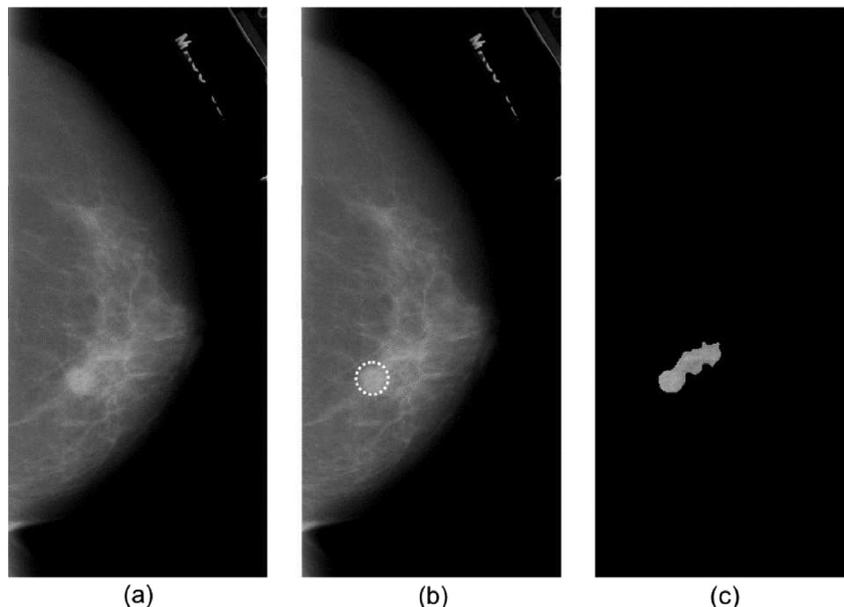


Fig. 17. Exemplo de ruído.

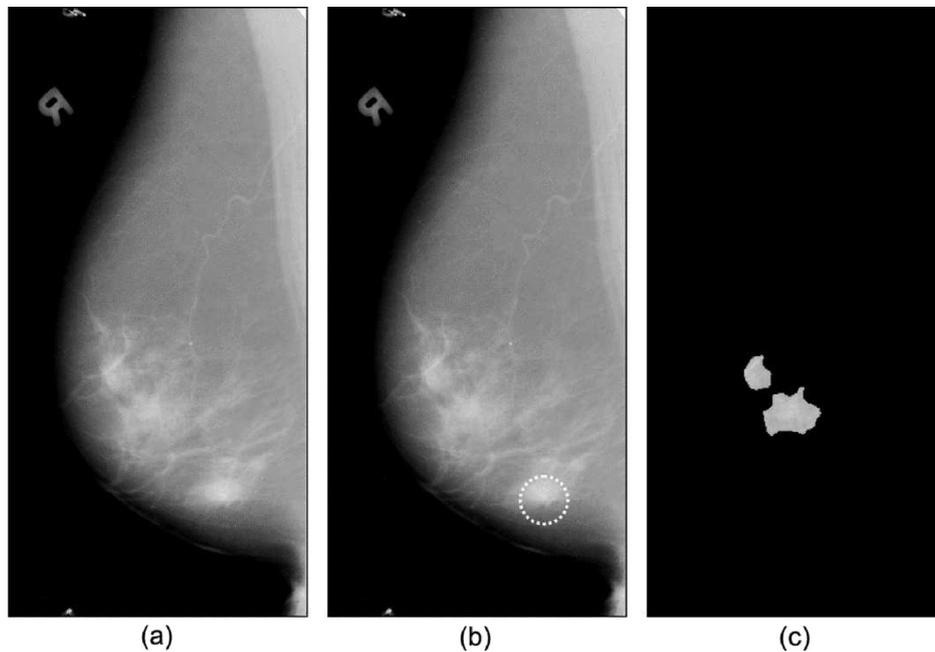


Fig. 18. Exemplo de redundância.

Table 9

Performance of the SVM classifier in the second reduction of false positives, with the texture features extracted from non-dense breasts

Sub-region	Masses		Non-masses		Performance		
	Success	Errors	Success	Errors	Sen. (%)	Spe. (%)	Acc. (%)
Circular	150	19	587	52	88.76	91.86	91.21
Circular crown	153	16	616	23	90.53	96.4	95.17
Internal mask	168	1	626	13	99.41	97.97	98.27
External mask	164	5	622	17	97.04	97.34	97.28

Table 10

Performance of the SVM classifier in the second reduction of false positives, with the texture features extracted from dense breasts.

Sub-region	Masses		Non-masses		Performance		
	Success	Errors	Success	Errors	Sen. (%)	Spe. (%)	Acc. (%)
Circular	93	11	383	36	89.42	91.41	91.01
Circular crown	97	7	378	41	93.27	90.21	90.82
Internal mask	101	3	399	20	97.12	95.23	95.60
External mask	100	4	388	31	96.15	92.60	93.31

Table 10 contains the result of the validation for each type of sub-region, using texture features extracted from dense breasts. The analysis by circular sub-regions had the worst results, with a sensitivity of 89.42%, specificity of 91.41% and accuracy of 91.01%. The analysis through internal masks had the best result, with a sensitivity of 97.12%, specificity of 95.23% and accuracy of 95.60%.

At the end of this stage, eight training models are created, one model for each of sub-region type of the ROI and density. The results achieved in this stage present a good indication that the μ GA managed to create good training models. However, these models must be tested with new cases.

4. Results and discussion

This section presents and discusses the final results, generated in the test stage and achieved according to the methodology described in Section 3.

Table 11

Confusion matrix of the test.

		Density	
		Non-dense	Dense
Classification	Non-dense	206	4
	Dense	2	133

4.1. Detection of masses

The best training models and parameters estimated in the training and validation stage will be tested with new cases. This new stage, called detection of masses, uses a base with 20% of the images that did not participate in the training and validation stage (Section 3.3), being formed by 210 non-dense breast images (with 214 masses) and 135 dense breast images (com 135 masses). Based on these images, we compute the general performance of the proposed methodology.

4.1.1. Acquisition and pre-processing

To the images used in the test stage, we applied the same pre-processing techniques described in Section 3.3.1 and, after that, classified them into non-dense or dense.

4.1.2. Detection of breast density

The objective of the stage of breast density detection is to classify the breast as non-dense or dense. It comprises three sub-stages: pre-processing, extraction of features and test. The sub-stages of pre-processing and extraction of features are identical to those described in Section 3.3.

After the extraction of features, the test sub-stage classifies the features extracted from the test images (set of 20%) into non-dense or dense, using the training model optimized by the μ GA.

The set of images of the test set is formed by 210 non-dense breast images and 135 dense breast images. The training model created by the μ GA is used by the SVM to classify the images into non-dense and dense. Table 11 contains the confusion matrix of the classification.

Looking at Table 11, one notices that the training model created achieved, in the classification of the density of the breast, 98.10% sensitivity, 98.52% specificity and 98.26% accuracy. As can be seen, the

Table 12
Number of ROIs produced in the segmentation, in the test stage.

Type of density	Masses				Non-masses Segmented	Proportion Non-masses/masses
	Segmented		Lost			
	Total	%	Total	%		
Non-dense	207	96.73	7	3.27	2031	9.81
Dense	127	94.07	8	5.93	1337	10.53

breast density detection stage had good results. This is an important fact, because the performances of the next stages will not be compromised.

All the images, correctly classified or not, will be used in the next stage. The use of all the images will indicate the global performance of the methodology, and not only the performance of each isolated stage.

The next stages adjust the parameters of the techniques employed according to the result of the classification of the density of the image under analysis.

4.1.3. Segmentation of the ROIs

This stage is intended to identify the regions of the breast with higher chances of containing masses. This is done by means of the μ GA. The procedures are identical to those described in Section 3.4, adapting the parameters according to the density, as in Table 4.

In this stage, we used 210 non-dense breast images, with 214 masses; and 135 dense breast images, com 135 masses. From the non-dense breasts, the segmentation extracted 207 ROIs containing masses (96.73% of the total of masses), and 2031 ROIs containing non-masses. From the dense breasts, the segmentation extracted 127 ROIs containing masses (94.07% of the total of masses), and 1337 ROIs containing non-masses. Table 12 contains the results of the segmentation of this stage.

The segmentation achieved good results, since more than 94% of the masses were segmented. However, the number of non-masses segmented is from 9.81 to 10.53 times bigger than the number of masses segmented. So, it is necessary to reduce the number of non-masses segmented.

4.1.4. Reduction of false positives using DBSCAN

The stage of reduction of false positives consists in eliminating the ROIs that do not have masses. The procedures are identical to those described in Section 3.5, having the parameters adapted according to Table 4.

For the non-dense breasts, from the 207 masses produced by the segmentation, after the reduction of false positives, there were still 203 masses left, which corresponds to 94.86% of the 214 total masses. From the 2031 non-masses produced by the segmentation, after the reduction of false positives, there were still 739 non-masses left.

For the dense breasts, from the 127 masses produced by the segmentation, after the reduction of false positives, there were 122 masses left, which corresponds to 90.37% of the 135 total masses. From the 1337 non-masses produced by the segmentation, after the reduction of false positives, 509 non-masses were left. Table 13 contains the results of the reduction of false positives of this stage.

The reduction of false positives using DBSCAN achieved a good performance in the test stage, since the reduction of non-masses in non-dense breasts was of 63.36%, and of 61.92% of the non-masses in dense breasts.

After the reduction of false positives using DBSCAN, the second reduction of false positives is started.

4.1.5. Reduction of false positives using Phylogenetic Trees, LBP and SVM

This second reduction of false positives uses texture features to extract descriptive measurements of the ROIs resulting from the first

Table 13
Number of ROIs resulting from the reduction of false positives using DBSCAN, in the test stage.

Type of density	Masses				Non-masses remaining	Proportion Non-masses/masses
	Selected		Lost			
	Total	%	Total	%		
Non-dense	203	94.86	11	5.14	739	3.64
Dense	112	90.37	13	9.36	509	4.17

Table 14
Performance of the SVM classifier in the second reduction of false positives, using the texture features extracted from non-dense breasts.

Sub-region	Masses		Non-masses		Performance		
	Success	Errors	Success	Errors	Sen. (%)	Spe. (%)	Acc. (%)
Circular	165	38	577	162	81.28	78.08	78.77
Circular crown	180	23	665	74	88.67	89.99	89.7
Internal mask	199	4	708	31	98.03	95.81	96.28
Internal mask	191	12	690	49	94.09	93.37	93.52

Table 15
Performance of the SVM classifier in the test stage, using the texture features extracted dense breasts.

Sub-region	Masses		Non-masses		Performance		
	Success	Errors	Success	Errors	Sen. (%)	Spe. (%)	Acc. (%)
Circular	93	29	412	97	76.23	80.94	80.03
Circular crown	96	26	455	54	78.69	89.39	87.32
Internal mask	113	9	483	26	92.62	94.89	94.45
External mask	112	10	468	41	91.80	91.94	91.91

reduction of false positives, and to form um set of feature vectors, by the combination of LBP and Phylogenetic Trees.

The procedures are identical to those described in Section 3.5.2, having the parameters adapted according to the density.

The optimized models created by the μ GA in the training and validation stage (Section 3.5.2) are used to test new feature vectors. This will indicate if the models created have a good generalization power for cases not trained.

At the end of the tests, we compute the values of sensitivity, specificity and accuracy. Table 14 contains the performance of the tests for each type of sub-region, using the feature vectors extracted from non-dense breasts. In this table, one can notice again that the analysis using circular sub-regions had the worst result, with 81.28% of sensitivity, 78.08% specificity and 78.77% accuracy. Again, the analysis using internal masks achieved the best performance, with 98.03% sensitivity, 95.81% specificity and 96.28% accuracy.

The Table 15 contains the test results for each type of sub-region, using the feature vectors extracted from dense breasts.

Analyzing the Table 15, it can be again noticed that the analysis using circular sub-regions presented the worst result, with 76.23% sensitivity, 80.94% specificity and 80.03% accuracy. Once more, the analysis using masks achieved the best performance, with 92.62 sensitivity, 94.89% specificity and 94.45% accuracy.

The results presented in Tables 14 and 15 are values which just indicate the performance in the second reduction of false positives. Analyzing the total of original images and masses for this stage, from the 214 masses in the starting 210 non-dense breast images, 199 were detected in the best case, in which the sub-regions are extracted by internal masks, resulting in a sensitivity of 92.99%, a rate of 0.15 false positives per image and an area under the FROC curve of 0.96.

Table 16
Areas under the FROC curves for non-dense and dense breasts.

	Non-dense	Dense
Circular	0.80	0.72
Circular crown	0.86	0.76
Internal mask	0.96	0.85
External mask	0.93	0.85

Analyzing the total of original images and masses for this stage, from the 135 masses in the starting 135 dense breast images, 113 were detected in the best case, in which the sub-regions are extracted by internal masks, resulting in a sensitivity of 83.70%, a rate of 0.19 false positives per image and an area under the FROC curve of 0.85.

Table 16 contains the areas under the FROC curves of the results obtained with non-dense and dense breasts.

Analyzing the results, the proposed methodology has, in general, a higher performance in the analyses of non-dense breasts. The analyses internal masks present superior performance compared to the other regional analyses applied.

Next, we present the cases of success and failures of the proposed methodology.

4.2. Case studies

In this section we examine the most important stage of the proposed methodology through some test cases. The objective is to ease

the understanding of the techniques used in each stage. For such, we present six cases.

4.2.1. First case: correct detection of the density

The first case, presented in Fig. 19, is an example of success in the detection of the density of non-dense and dense breasts.

As can be observed in Fig. 19(b), a application of the variance filter to Fig. 19(a) highlights just a few internal structures, which is expected for non-dense breasts. So, the values of $v_{internal}$ (Eq. (7)) and v_{mean} (Eq. (8)) are lower. In Fig. 19(d), the application of the variance filter to Fig. 19(c) highlights many internal structures, which is expected for dense breasts, so, $v_{internal}$ and v_{mean} are higher.

4.2.2. Second case: incorrect detection of the density

The second case, presented in Fig. 20, is an example of failure in the detection of the density of non-dense breasts and dense. As can be observed in Fig. 20(b), the application of the variance filter to Fig. 20(a) highlights many internal structures, which is not expected for non-dense breasts, so the values of $v_{internal}$ and v_{mean} are higher. In Fig. 20(d), the application of the variance filter to Fig. 20(c) highlights few internal structures, which does not match with dense breasts, so, $v_{internal}$ and v_{mean} have lower values.

4.2.3. Third case: success in the detection of masses

The third case is an example of success in the segmentation and in the reduction of false positives. Fig. 21(a) is the original image. A region marked in red in Fig. 21(b) contains a mass. This marking was retrieved from the DDSM. Fig. 21(c) contains the result of the segmentation. In it, the region containing the mass was correctly segmented.

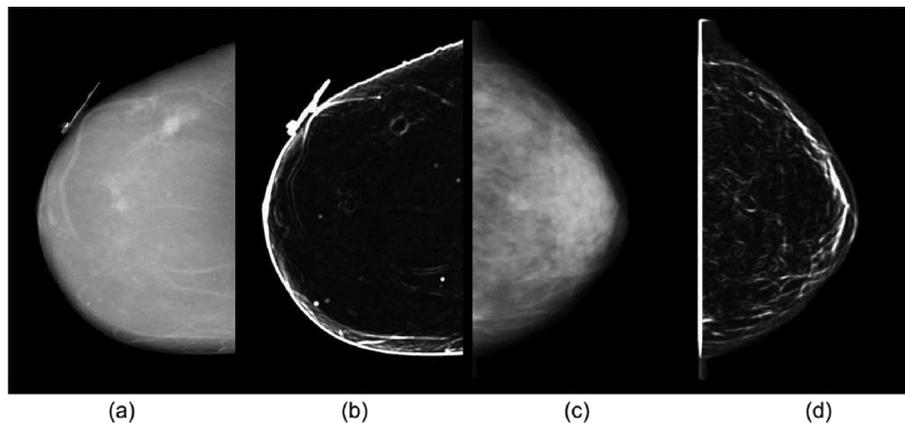


Fig. 19. Success in the detection of the density. (a) Non-dense breast (case A_1004). (b) Application of the variance filter to (a). (c) Dense breast (case A_1057). (d) Application of the variance filter to (c).

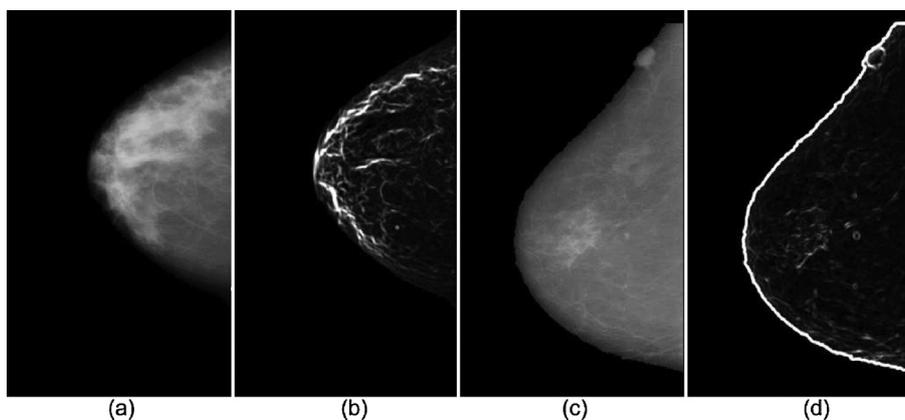


Fig. 20. Failure in the detection of the density. (a) Non-dense breast (case C_0095). (b) Application of the variance filter to (a). (c) Dense breast (case A_1581). (d) Application of the variance filter to (c).

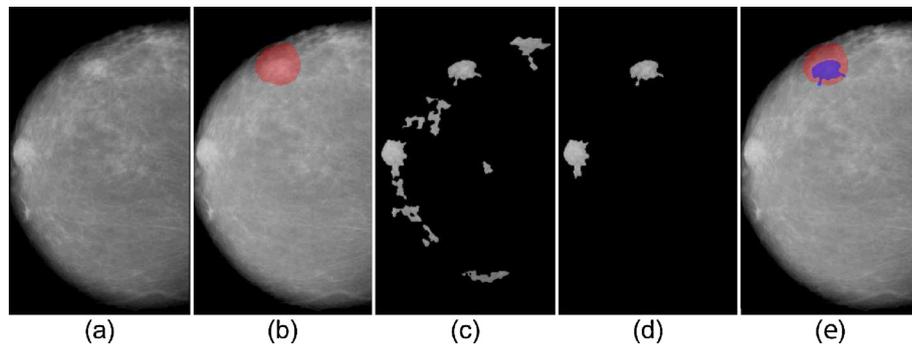


Fig. 21. Success in the detection of masses (case A_1498). (a) Pre-processed image. (b) The region in red indicates an area containing a mass. (c) Result of the segmentation. (d) ROIs resulting from the reduction of false positives. (e) In blue, the mass found by the methodology (false positive), and in red, the area pointed by the DDSM, containing the mass. In yellow, the segmented ROI which contains the mass, but was not selected (false negative).

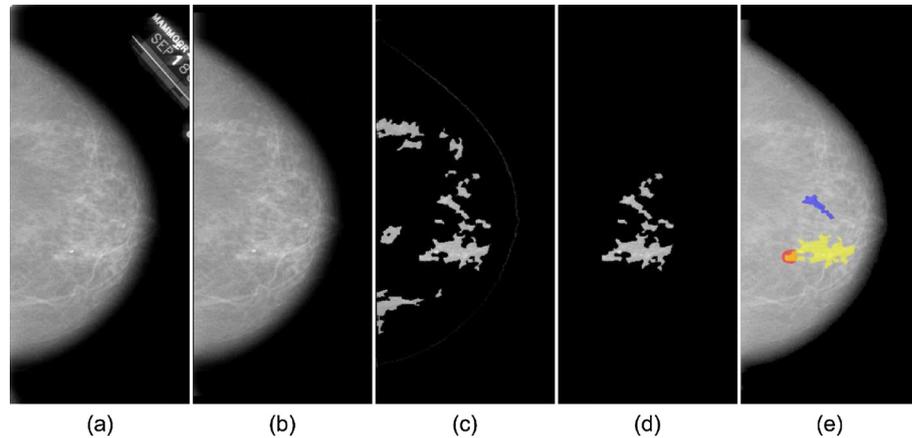


Fig. 22. Failure in the detection. (a) Original image (case A_1009). (b) Pre-processed image. (c) ROIs selected in the segmentation. (d) ROIs resulting from the reduction of false positives. (e) In blue, the mass found by the methodology (false positive), and in red, the area pointed by the DDSM, containing the mass. In yellow, the segmented ROI which contains the mass, but was not selected (false negative). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article)

Fig. 21(d) presents the result of the reduction of false positives using DBSCAN. And Fig. 21(e) contains the result of the reduction of false positives by phylogenetic trees, LBP and SVM.

4.2.4. Fourth case: failure in the detection of masses

The fourth case, presented in Fig. 22, is an example of failure in the detection. Fig. 22(a) presents the original image. Fig. 22(b) contains the result of the pre-processing. Next, the candidates selected in the segmentation (Fig. 22(c)). Fig. 22(d) contains the candidates selected in the reduction of false positives using DBSCAN and, finally, Fig. 22(e) presents the result of the reduction of false positives using phylogenetic trees, LBP and SVM. The region marked in blue is the region classified by the SVM as a mass, and in red is the region pointed by the DDSM.

Although the ROI on the mass was not removed in the reduction of false positives, SVM could not classify it correctly, discarding it and pointing as mass a region which actually corresponds to a non-mass. This case illustrates the occurrence of a false negative (mass classified as non-mass, in red) and a false positive (non-mass classified as mass, in blue). Among the possible causes for this failure, one is related to the fact that the segmented ROI has too many healthy areas of the breast, and this affected the extraction of features, and therefore, the SVM. The ROI erroneously marked as mass probably presented a texture pattern similar to a mass, influencing SVM.

4.2.5. Fifth case: segmentation failure

The fifth case, presented in Fig. 23, is an example of segmentation failure. The region marked in red in Fig. 23(b) contains a mass. This marking was retrieved from the DDSM. In Fig. 23(c), the region corresponding to a mass was not segmented. This failure can be explained

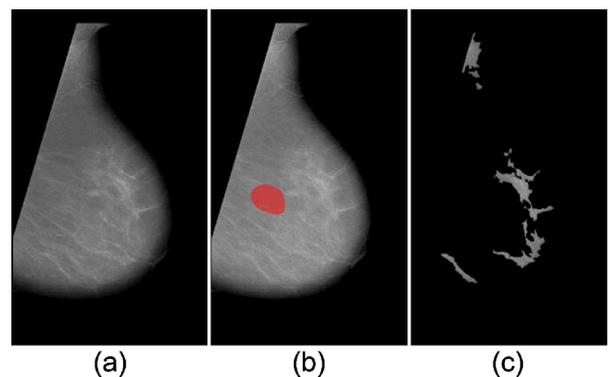


Fig. 23. Segmentation failure (case A_1494). (a) Pre-processed image. (b) The region in red indicates an area containing a mass. (c) Result of the segmentation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article)

by tones of the pixels of the lesion, which are very similar to healthy regions of the breast.

4.2.6. Sixth case: failure in the reduction of false positives

The sixth case, presented in Fig. 24, is an example of failure in the reduction of false positives. The region marked in red in Fig. 24(a) contains a mass. This marking was retrieved from the DDSM. In Fig. 24(b), the region corresponding to a lesion was also segmented. Fig. 24(c) is the result of the reduction of false positives using DBSCAN, in which ROI on the specialist's marking was removed.

This failure may have occurred because the ROI containing the mass had the mean of the gray tones below the majority of the other

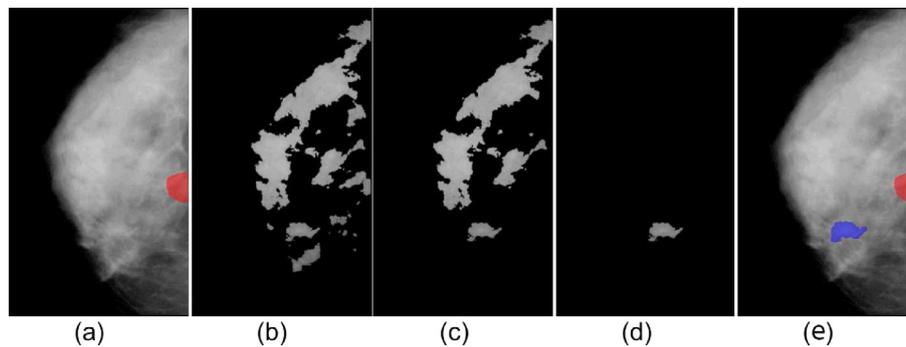


Fig. 24. Failure in the reduction of false positives (case A_1045). (a) Pre-processed image. (b) The region in red indicates an area containing a mass. (c) Result of the segmentation. (d) Result of the reduction of false positives. (e) In blue, the region erroneously detected, and in red, the mass indicated by the DDSM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article)

ROIs and, in this case, the DBSCAN considered it similar to the majority of the texture of the breast, that is, a healthy region.

Fig. 24(d) contains the result of the second reduction of false positives, and Fig. 24(e) contains the marking of the mass indicated by the DDSM, in red, and in blue, the marking erroneously indicated by the methodology. Even reducing the number of false positives segmented, the analysis of texture of the second reduction of false positives did not remove all the non-masses.

The next section discusses the main points of the proposed methodology.

4.3. Discussion

The proposed methodology presents a superior performance in the analysis of non-dense breasts. This may occur because, in dense breasts, segmented healthy structures are dense too, and this may affect the texture descriptors. Another factor is the number of structures segmented in dense breasts, which implies in the higher number of false positives that will be analyzed in the stages of reduction of false positives. This performance difference between the densities can be corrected by the use of distinct image processing and pattern recognition techniques for each density.

The combination of phylogenetic trees and LBP generates a texture descriptor capable of differentiating between masses and non-masses efficiently. This performance can be explained by the fact that the phylogenetic trees extract indexes that relate diversity, richness and the parenthood among the species under analysis. On the other hand, the LBP codes allow the analysis of the topological relations among the gray levels locally and in various directions.

The use of internal masks produce superior results, compared to the other analyses. This fact highlights the importance of texture information that follows the border of the ROI.

The large number of parameters empirically found is a negative point of the methodology. So, new studies must focus on finding these parameters automatically, when possible.

The μ AG can produce good training models, which leads to good results in tests, even with a long processing time.

From all the stages that form the proposed methodology, the ones with higher time consumption are those that create the training models, that is, estimation of the density-training models and estimation of the texture-training models.

The time necessary to create the density-training models is about 30 min. The texture-training models, in turn, take from 3 to 5 h, depending on the number of features analyzed and the number of vectors. The higher the number of features that form each model, longer the time needed to create them will be.

We stress that, after the training, the classification runs faster, taking about 10 min to classify the whole sample. The other stages of the methodology take less than one minute per image.

Table 17

Comparison of the performances of the methodologies presented in the related works section. The performance is measured by sensitivity (S), specificity (Sp), accuracy (Acc), area under the ROC curve, area under the FROC curve and false positives per image (FP/i).

Work	Base	S	Sp	Acc	ROC	FROC	FP/i
(Masotti et al., 2009)	DDSM	70	–	–	–	–	0.92
(Bajger et al., 2010)	Private	–	–	–	0.9	–	–
(Bajger et al., 2010)	DDSM	–	–	–	0.96	–	–
(Moayedi et al., 2010)	MIAS	–	–	91.52	–	–	–
(Abdalla et al., 2011)	DDSM	–	–	95.85	–	–	–
(Al Mutaz, Dress, & Zaki, 2011)	DDSM	91.67	84.17	–	–	–	–
(Banik et al., 2011)	Private	90	–	–	–	–	8.1
(Hu, Gao, & Li, 2011)	MIAS	91.3	–	–	–	–	0.71
(Liu et al., 2011)	DDSM	76.8	–	–	–	–	1.36
(Sampaio et al., 2011)	DDSM	80	–	–	–	–	0.84
(Deepak et al., 2012)	MIAS	100	97	98.9	0.98	–	–
(Berbar et al., 2012)	DDSM	–	–	98.63	–	–	–
(Tai et al., 2014)	DDSM	90.0	–	–	0.97	–	4.8
(Agrawal et al., 2014)	MIAS	–	–	–	0.89	–	–
(Lin et al., 2014)	Private	0.86	–	–	–	–	–
(Kuo et al., 2014)	MIAS	–	–	94.44	–	–	–
(Anitha & Peter, 2015)	MIAS	92.25	83.83	93.48	–	–	–
(Dong et al., 2015)	DDSM	94.78	91.76	93.24	–	–	0.95
(Jen & Yu, 2015)	MIAS	89.0	84.0	–	–	–	–
(Jen & Yu, 2015)	DDSM	86.0	84.0	–	–	–	–
(Oliveira et al., 2015)	MIAS	98.60	98.88	98.85	–	–	–
Our work (non-dense)	DDSM	92.99	–	–	–	0.96	0.15
Our work (dense)	DDSM	83.70	–	–	–	0.85	0.19

4.4. Comparison with other related works

This subsection brings a comparison between the results of the related works describe in Section 1 with the results achieved by the proposed methodology. The comparison is synthesized in Table 17, which lists the performances measured by sensitivity (S), specificity (Sp), accuracy (Acc), area under the ROC curve, area under the FROC curve and mean false positives per image (FP/i).

Compared to the works presented in the Table 17, the proposed methodology found, in general, a similar performances. As explained before, it is not possible to state which of the works has a superior performance.

5. Conclusion

This work presents a methodology for detection of masses in mammography by means of image processing techniques, pattern recognition and evolutionary algorithms which automatically adapt the density of each breast under analysis.

The proposed methodology uses images from DDSM, a public image database. However, other databases can be used, which will require only the adjustment of some parameters needed by the stages of the methodology. Besides the images, extra information is necessary, such as markings of the regions that contain masses and type of density of the breast under analysis.

The stage for detection of the density of the breast had a good performance, achieving 100% rates of sensitivity, specificity and accuracy in the validation of the training model and, in the test set, 98,10% sensitivity, 98,52% specificity and 98,26% accuracy. This indicates that μ GA managed to create good training models, which allowed a good generalization capability for new cases.

In the test set, the segmentation could separate 96,73% of the masses, with 2031 non-masses, in non-dense breasts; and 94.07% of the masses, with 1337 non-masses, in dense breasts. In the parameter estimation (training) set, the segmentation could separate 97.41% of the masses, with 9613 non-masses, in non-dense breasts; and 94.48% of the masses, with 9933 non-masses, in dense breasts.

These numbers are satisfactory, but the proportion of non-masses created is from 9.81 to 13.06 times bigger than the proportion of masses correctly segmented, which indicates the necessity of a stage for reducing the number of non-masses.

The reduction of false positives using DBSCAN also had a good performance, since it successfully reduced from 3.64 to 4.17 times the number of non-masses, without an excessive loss of masses.

We must emphasize that the combination of Phylogenetic Trees, LBP, division of the ROIs into sub-regions for extraction of texture features created good texture descriptors, which influenced the performance of the training models in the second reduction of false positives. This good performance is due to the contributions of the information about richness, diversity and phylogenetic relation among species, produced by the phylogenetic trees; local spatial texture relations produced by the LBP codes; and regional spatial relations produced by the division of the ROIs into sub-regions.

This work presents a series of contributions applied in several stages of the methodology, as follows.

Method for removal of the background and segmentation of the breast. This method allows the segmentation of the breast, in such a way that the line that separates the skin and the background of the image is kept, with a good performance. It also loses just a few pixels of the internal area of the breast, and also segmenting just a few pixels of the background, together with the breast. The efficiency of this method is explained by its capability to adapt to each image under analysis.

Method for detection of the density. This method presents a high performance at the detection of the density of the breast. This performance is achieved by the combination of the μ GA, good descriptors and the efficiency of the SVM.

Method for creation of the training models optimized by the μ GA. This method allows the creation of good training models for the SVM, optimizing its performance, by reducing noisy and redundant information.

Efficient method for describing texture based on the combination of Phylogenetic Trees, LBP and analysis in sub-regions. This technique relates an efficient information extractor, globally, to the phylogenetic information, and locally, to the LBP codes.

Adjustment of parameters according to the density classification of the breast. Allows the creation of a set of parameters applied to the specific density of each breast.

Method for segmentation of masses using the μ GA. Allows the estimation of the texture proximity window adapted to each image under analysis, in order to segment regions whose textures are not common in the image under analysis.

Method for reduction of false positives using DBSCAN. This technique allows efficiently removing non-masses, losing just a few masses, through a texture proximity ranking.

Acknowledgments

The authors acknowledge CAPES, CNPq and FAPEMA for financial support.

References

- Abdalla, A. M. M., Dress, S., & Zaki, N. (2011). Masses detection in digital mammogram by gray level reduction using texture coding method. *ijp*, 1, 1.
- Agrawal, P., Vatsa, M., & Singh, R. (2014). Saliency based mass detection from screening mammograms. *Signal Processing*, 99(0), 29–47. <http://dx.doi.org/10.1016/j.sigpro.2013.12.010>. <http://www.sciencedirect.com/science/article/pii/S0165168413004982>
- Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), 2037–2041. doi:10.1109/TPAMI.2006.244.
- Al Mutaz, M. A., Dress, S., & Zaki, N. (2011). Detection of masses in digital mammogram using second order statistics and artificial neural network. *International Journal of Computer Science & Information Technology (IJCSIT)*, 3(3), 176–186.
- Anitha, J., & Peter, J. D. (2015). Mammogram segmentation using maximal cell strength update in cellular automata. *Medical & biological engineering & computing*, 1–13.
- Bajger, M., Ma, F., Williams, S., & Bottema, M. (2010). Mammographic mass detection with statistical region merging. In *2010 international conference on digital image computing: Techniques and applications* (pp. 27–32). Institute of Electrical & Electronics Engineers (IEEE). doi:10.1109/DICTA.2010.14.
- Banik, S., Rangayyan, R. M., & Desautels, J. E. L. (2011). Detection of architectural distortion in prior mammograms. *IEEE Transactions on Medical Imaging*, 30(2), 279–294. doi:10.1109/TMI.2010.2076828.
- Berbar, M. A., Reyad, Y. A., & Hussain, M. (2012). Breast mass classification using statistical and local binary pattern features. In *Proceedings of the 2012 16th international conference on information visualisation* (pp. 486–490). Institute of Electrical & Electronics Engineers (IEEE). doi:10.1109/IV.2012.84.
- Boyd, N. F., Guo, H., Martin, L. J., Sun, L., Stone, J., Fishell, E., et al. (2007). Mammographic density and the risk and detection of breast cancer. *New England Journal of Medicine*, 356(3), 227–236. doi:10.1056/NEJMoa062790.
- Chang, C.-C., & Lin, C.-J. (2011). Libsvm. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 1–27. doi:10.1145/1961189.1961199.
- Deepak, K. S., Medathati, N. K., & Sivaswamy, J. (2012). Detection and discrimination of disease-related abnormalities based on learning normal cases. *Pattern Recognition*, 45(10), 3707–3716. doi:10.1016/j.patcog.2012.03.020.
- Dong, M., Lu, X., Ma, Y., Guo, Y., Ma, Y., & Wang, K. (2015). An efficient approach for automated mass segmentation and classification in mammograms. *Journal of digital imaging*, 1–13.
- Fenton, J. J., Taplin, S. H., Carney, P. A., Abraham, L., Sickles, E. A., D'Orsi, C., et al. (2007). Influence of computer-aided detection on performance of screening mammography. *New England Journal of Medicine*, 356(14), 1399–1409. doi:10.1056/NEJMoa066099.
- Ferguson, P. D., Arslan, T., Erdogan, A. T., & Parmley, A. (2008). Evaluation of contrast limited adaptive histogram equalization (clahe) enhancement on a fpga. In *Proceedings of the 2008 IEEE international soc conference* (pp. 119–122). Institute of Electrical & Electronics Engineers (IEEE). doi:10.1109/SOCC.2008.4641492.
- Gonzalez, R. C., Woods, R. E., & Masters, B. R. (2009). Digital image processing, third edition. *Journal of Biomedical Optics*, 14(2), 029901. doi:10.1117/1.3115362.
- Hall, S., & Greenstreet, S. (1998). Taxonomic distinctness and diversity measures: responses in marine fish communities. *Marine Ecology Progress Series*, 166, 227–229. doi:10.3354/meps166227.
- Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr, P., Moore, R., et al. (1998). Current status of the digital database for screening mammography. In *Proceedings of the Digital mammography* (pp. 457–460). Springer.
- Hu, K., Gao, X., & Li, F. (2011). Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms. *IEEE Transactions on Instrumentation and Measurement*, 60(2), 462–472. doi:10.1109/TIM.2010.2051060.
- Jen, C.-C., & Yu, S.-S. (2015). Automatic detection of abnormal mammograms in mammographic images. *Expert Systems with Applications*, 42(6), 3048–3055.
- Junior, G. B., de Paiva, A. C., Silva, A. C., & de Oliveira, A. C. M. (2009). Classification of breast tissues using moran's index and geary's coefficient as texture signatures and svm. *Computers in biology and medicine*, 39(12), 1063–1072. JID: 1250250; 2008/06/02 [received]; 2009/08/19 [revised]; 2009/08/24 [accepted]; 2009/10/01 [aheadofprint]; publish.
- Kuo, Y.-C., Lin, W.-C., Hsu, S.-C., & Cheng, A.-C. (2014). Mass detection in digital mammograms system based on pso algorithm. In *Proceedings of the international symposium on computer, consumer and control (is3c)*, 2014 (pp. 662–668). doi:10.1109/IS3C.2014.178.
- Lin, G.-S., Chang, Y.-C., Yeh, W.-C., Liu, K.-C., & Yeh, C.-H. (2014). A pyramid-based mass detection scheme based on texture analysis and neural classifier for digital mammograms. *Biomedical Engineering: Applications, Basis and Communications*, 26(04), 1440001. doi:10.4015/S1016237214400018.
- Liu, X., Xu, X., Liu, J., & Feng, Z. (2011). A new automatic method for mass detection in mammography with false positives reduction by supported vector machine. In *Proceedings of the 2011 4th international conference on biomedical engineering and informatics (bmei)* (pp. 33–37). Institute of Electrical & Electronics Engineers (IEEE). doi:10.1109/BMEI.2011.6098328.
- 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(2), 311. doi:10.1118/1.3049588.

- Moayed, F., Azimifar, Z., Boostani, R., & Katebi, S. (2010). Contourlet-based mammography mass classification using the svm family. *Computers in Biology and Medicine*, 40(4), 373–383. doi:10.1016/j.compbiomed.2009.12.006.
- Ng, E.-K. (2008). *Improved Sensitivity and Specificity of Breast Cancer Thermography*. Elsevier BV. doi:10.1016/B978-012374212-4.50048-1.
- Nunes, A. P., Silva, A. C., & de Paiva, A. C. (2009). Detection of masses in mammographic images using simpson's diversity index in circular regions and svm. In *Proceedings of the Machine learning and data mining in pattern recognition* (pp. 540–553). Springer.
- Oliveira, F. S. S. d., Carvalho Filho, A. O. d., Silva, A. C., de Paiva, A. C., & Gattass, M. (2015). Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and svm. *Computers in biology and medicine*, 57, 42–53.
- de Oliveira Martins, L., Silva, A. C., De Paiva, A. C., & Gattass, M. (2009). Detection of breast masses in mammogram images using growing neural gas algorithm and ripley's k function. *Journal of Signal Processing Systems*, 55(1-3), 77–90.
- Parkin, D. M., Bray, F., Ferlay, J., & Pisani, P. (2005). Global cancer statistics, 2002. *CA: A Cancer Journal for Clinicians*, 55(2), 74–108. doi:10.3322/canjclin.55.2.74.
- Petrick, N., Chan, H. P., Sahiner, B., & Wei, D. (1996). An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. *IEEE Transactions on Medical Imaging*, 15(1), 59–67. JID: 8310780; ppublish
- Sampaio, W., Moraes Diniz, E., Corrêa Silva, A., Cardoso de Paiva, A., & Gattass, M. (2011). Detection of masses in mammogram images using cnn, geostatistic functions and svm. *Computers in Biology and Medicine*, 41(8), 653–664.
- Tai, S.-C., Chen, Z.-S., & Tsai, W.-T. (2014). An automatic mass detection system in mammograms based on complex texture features. *IEEE Journal of Biomedical and Health Informatics*, 18(2), 618–627. doi:10.1109/JBHI.2013.2279097.
- Tam, V., Cheng, K.-Y., & Lui, K.-S. (2006). Using micro-genetic algorithms to improve localization in wireless sensor networks. *Journal of Communications*, 1(4). doi:10.4304/jcm.1.4.1-10.