Computer Methods and Programs in Biomedicine 156 (2018) 191–207

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks

João Otávio Bandeira Diniz[a,*], Pedro Henrique Bandeira Diniz[b], Thales Levi Azevedo Valente[b], Aristófanes Corrêa Silva[a], Anselmo Cardoso de Paiva[a], Marcelo Gattass[b]

[a] Applied Computing Group, Federal University of Maranhão – UFMA, NCA, Av. dos Portugueses, SN, Bacanga, São Luís, MA 65085-580, Brazil
[b] Pontifical Catholic University of Rio de Janeiro – PUC – Rio, R. São Vicente, 225, Gávea, Rio de Janeiro, RJ, 22453-900, Brazil

ARTICLE INFO

Article history:
Received 31 March 2017
Revised 13 December 2017
Accepted 10 January 2018

Keywords:
Bilateral asymmetry
Breast cancer
Computer-aided detection
Convolutional neural network
Deep learning
Similary indexes

ABSTRACT

Background and Objective: The processing of medical image is an important tool to assist in minimizing the degree of uncertainty of the specialist, while providing specialists with an additional source of detect and diagnosis information. Breast cancer is the most common type of cancer that affects the female population around the world. It is also the most deadly type of cancer among women. It is the second most common type of cancer among all others. The most common examination to diagnose breast cancer early is mammography. In the last decades, computational techniques have been developed with the purpose of automatically detecting structures that maybe associated with tumors in mammography examination. This work presents a computational methodology to automatically detection of mass regions in mammography by using a convolutional neural network.

Methods: The materials used in this work is the DDSM database. The method proposed consists of two phases: training phase and test phase. The training phase has 2 main steps: (1) create a model to classify breast tissue into dense and non-dense (2) create a model to classify regions of breast into mass and non-mass. The test phase has 7 step: (1) preprocessing; (2) registration; (3) segmentation; (4) first reduction of false positives; (5) preprocessing of regions segmented; (6) density tissue classification (7) second reduction of false positives where regions will be classified into mass and non-mass.

Results: The proposed method achieved 95.6% of accuracy in classify non-dense breasts tissue and 97,72% accuracy in classify dense breasts. To detect regions of mass in non-dense breast, the method achieved a sensitivity value of 91.5%, and specificity value of 90.7%, with 91% accuracy. To detect regions in dense breasts, our method achieved 90.4% of sensitivity and 96.4% of specificity, with accuracy of 94.8%.

Conclusions: According to the results achieved by CNN, we demonstrate the feasibility of using convolutional neural networks on medical image processing techniques for classification of breast tissue and mass detection.

1. Introduction

Breast cancer is the most common type of cancer that affects the female population around the world. It is also the most deadly type of cancer among women. It is the second most common type of cancer among all others and the fifth in causes of death. It should be noted that developing countries have the highest rate of cases [1].

On the other hand, it is known that breast cancer, when diagnosed early, has a high chance of being cured. Based on this scenario, world governments together with health professionals have mobilized various campaigns to alert the female population to the risks that cause this disease. Thus, it is emphasized that early diagnosis is the most effective way to combat this disease [1].

Currently, the most common examination to diagnose breast cancer early is mammography, a radiological exam that generates gray scale images of the breast. The medical specialist analyzes, and visually identifies where the lesion or lesions are located [2–5].

It is known that both breasts of the same patient tend to have a high degree of symmetry [6]. The medical specialist uses a comparison of the mammograms of these breasts to verify anomalies

* Corresponding author.
E-mail addresses: joao.obd@gmail.com (J.O. Bandeira Diniz), pedro_hbd@hotmail.com (P.H. Bandeira Diniz), selaht7@gmail.com (T.L. Azevedo Valente), aricsilva@gmail.com (A. Corrêa Silva), anselmo.c.paiva@gmail.com (A.C. de Paiva), mgattass@tecgraf.puc-rio.br (M. Gattass).

https://doi.org/10.1016/j.cmpb.2018.01.007
0169-2607/

based on the existing symmetry. This practice is commonly known as bilateral analysis. Therefore, it is mentioned that the use of computational techniques to make a spatial alignment (registration), and then the comparison of these mammograms, would allow for an evaluation of the degree of symmetry and determine the breasts as normal or non-normal [6,7].

It is also noteworthy that, in the last several decades, computational techniques have been developed for the purpose of automatically detecting structures that may be associated with tumors in mammography examinations, with the objective of improving the early detection rate of structures of interest associated with breast cancer [2,4,8–13].

All these factors have motivated several research studies over the last few decades, with the goal of developing computational systems to aid specialists in interpreting radiological images. These systems are known as Computer-Aided Detection (CAD) and Computer-Aided Diagnosis (CADx), and are already present in several diagnostic imaging centers, thus increasing the rates of accuracy in the early identification of serious diseases such as breast cancer [14].

There are few works in the literature that use different techniques for different types of tissue (dense and non-dense) [4,15], so appropriate techniques for each type of tissue would produce better results.

This work proposes a CAD methodology to assist the medical specialist in the task of detecting mass regions in mammographic images. This method uses a bilateral analysis of breast pairs based on breast density. In addition, image processing techniques, similarity indexes of biology, and convolutional neural networks are used to achieve this task.

The proposed methodology runs through all the requirements of a CAD system from the task of pre-processing the breast to the detection of mass regions. This methodology contributes the following: (a) the use of techniques commonly used in biology to segment suspicious regions between pairs of mammograms, (b) development of two techniques to reduce false positives by decreasing the number of segmented regions, (c) proposal for the use of a Convolutional Neural Network (CNN) capable of classifying the breasts between dense and non-dense, and (d) two other CNNs capable of detecting regions of masses in dense breasts and in non-dense breasts. We showed that convolutional neural networks can be well used in CAD systems.

The next sections are as follows. Section 2 presents a review of related work. In Section 3 presents the proposed methodology divided into two phases (training phase and test phase) with its sub-stages. Section 4 shows the results and discussions together with some cases of the application of the proposed methodology. Finally, Section 5 concludes the paper by examining the efficiency of the proposed method.

2. Related works

The literature indicates that the efficiency of CAD and CADx breast systems is closely related to the techniques that comprise them. Furthermore, a range of studies focusing on mammography mass detection can be seen in the literature, whether using the mammography pairs approach, based on breast density or even for patient monitoring over time. Current works considered relevant in the scientific field will be presented.

2.1. Mass detection using only a mammogram

Bajget et al. [16] presented an automatic method to detect masses in mammograms for the segmentation of regions of interest (ROIs) by statistical fusion and linear discriminant analysis (LDA). This methodology uses 36 images selected from a private mammographic image database and 48 images taken from the Digital Database for Screening Mammography (DDSM). For the classification of ROIs, the value of the area under the curve receiver operating characteristic (ROC), was 0.90 for the proprietary images and 0.96 for the DDSM images.

The methodology proposed by Hu et al. [17] uses two types of adaptive thresholding techniques, applied globally and locally for multi-resolution segmentation. Tests were performed on 170 mammographic images from the public database called Mammographic Image Analysis Society (miniMIAS). The results of the experiment reached 91.3% sensitivity, with a rate of 0.71 false positives per image (FPi).

In the work proposed by Sampaio et al. [4], a cellular neural network is proposed to segment the images and find ROIs. Then, shape and texture characteristics are extracted to describe them. ROIs are divided into two groups, mass and non-mass, and their characteristics are used to train a support vector machine (SVM). The result of the classification in the test images showed a sensitivity of 80% and 0.84 FPi.

Braz et al. [3] proposed a CAD system based on diversity indices (Shannon-Wiener, Simpson, J, Ed, Buzas-Gibson, Camargo, Hill, McIntosh, Total Diversity, Brillouin, and Berger-Parker), geostatistical indices (Global Getis, Local Ripley K, Joint-Counte, and Nearest Neighbor) and geometric indices extracted from concave geometries (Eccentricity, Circularity, Solidity, Orientation, Circular Disproportion, Circular Density, Square Density, Ring Density, and Quadratic Density) for mass detection using SVM. The results obtained in the miniMIAS database were 97.30% sensitivity, an area under the free response ROC curve (FROC) of 0.89, and a rate of 0.33 FPi. With the DDSM database, the results achieved were 91.63% sensitivity, an area under the FROC of 0.86, and a rate of 0.013 FPi.

Sampaio et al. [18] presented a methodology of mass detection adapted to the density of the breast. First, the type of tissue density (dense and non-dense) is detected, and then the proposed methodology follows a series of processes to perform mass detection. The work presents techniques to isolate the breast, discarding unnecessary regions. Subsequently, there is a segmentation stage where we obtain candidate regions for mass. This method also used two techniques of reduction of false positives. The first uses density-based spatial clustering of applications with noise (DBSCAN) and a texture proximity ranking extracted from the ROIs, and the second uses an SVM classifier to detect if the remaining ROIs are masses or non-masses. The best results for a non-dense breast produced a sensitivity of 94.02%, specificity of 82.28%, and accuracy of 84.08% with a rate of 0.85% FPi and area under the FROC of 1.13. In dense breasts, 89.13%, 88.61%, and 88.69% sensitivity, specificity and accuracy were obtained, respectively. Still in the dense breast, there was a rate of 0.71 FPi with an area under the FROC of 1.47.

2.2. Mass detection by bilateral analysis

In this section, we will present works that use a bilateral analysis approach, that is, a comparison between a patient's left and right image for mass detection.

Scutt et al. [6], in a first study, carried out a study on the asymmetry of the breasts. A relationship between asymmetry and the risk of cancer was initially observed. After comparing 250 cases of cancer patients and 250 cases of healthy patients occupying the same age group, the researchers concluded that the cancer group presented greater asymmetry in relation to the healthy group. The objective of the study was to obtain the first evidence that related the asymmetry between the breasts as a factor indicative of cancer risk. In the subsequent study, the researchers confirmed this relationship by evaluating 252 cases of normal women who did

not develop cancer and 252 cases of normal women who later developed cancer. It was observed that the group considered normal (that is, did not develop cancer) presented volume asymmetry with a mean of 52.99 ml, while the group that developed cancer had a mean of 63.17 ml. The study also came to the conclusion that for every 100 ml of volume asymmetry, the chance of developing cancer increased by 50%.

The work of Wu et al. [8] presented a method for reducing false positives in mass detection systems based on bilateral breast analysis. For each detected mass, ROIs were defined in both breasts, and from these regions were extracted characteristics of texture and morphology. Two types of characteristics were defined: unilateral (where the masses were found) and bilateral (a combination of the characteristics of the region with mass and the corresponding region in the other breast). Linear discriminant analysis was used to define the best unilateral and bilateral characteristics and to use them to classify true and false positives. The method was applied in 341 cases, obtaining a FPi of 0.35 as the best result.

Tzikopoulos et al. [15] demonstrate in their work a segmentation and classification based on a fully automated analysis bilateral scheme for mammograms in the miniMIAS database. Using techniques of image processing, segmentation, extraction of characteristics for categorization of breast density, and finally SVM in density classification, the study saw an accuracy of 85.7%. A classification is still made to detect asymmetry in the breasts, reaching a sensitivity of 84.47% in a set of tests of 322 mammograms.

Another methodology of bilateral analysis was presented by Wang et al. [19]. The work highlights that the asymmetry of bilateral mammographic density is a promising indicator in the evaluation of risk of a patient who has or develops breast cancer. The work was applied to a private database with 600 cases of exams, where there were 300 cases of cancer and 300 negative cases. Two CAD schemas are also presented, the first using a single image, detecting suspicious areas, and the second using the breast tissue of pairs of mammograms. The first scheme achieved a 74% sensitivity with 0.25% FPi. The second scheme achieved a 75% sensitivity. Tests were still made with the fusion of the two schemes, resulting in a sensitivity of 84%.

In the work proposed by Ereceira et al. [7], a methodology for mass detection is presented through the determination of suspicious regions between regions of pairs of left and right mammograms, and later classification of these regions. First, the breasts are isolated, and then the registration is used to align the breasts. Suspicious regions are detected by structural variations between corresponding regions, defined by a spatial descriptor of a dataset known as a cross-variogram. After the determination of suspicious regions in a pair of images, the variogram is used in each suspected region alone for mass and non-mass classification. This work presented a 96% accuracy as the best result.

In the work proposed by Sun et al. [20], a new bilateral comparison approach for the detection of breast cancer is reported. A private database of 180 cases is used, resulting in 180 bilateral mammograms. In the methodology, the breast is divided into five subregions according to the intensity of the breast area. Texture characteristics and morphological density characteristics of these subregions are extracted, and from this point a classification is made.

In Casti et al. [21], a method of bilateral analysis was proposed. The authors emphasize the importance of asymmetry detection analysis in the breasts. A bilateral convolution process is used to find anatomical marks. Texture characteristics are extracted using semivariogram descriptors and structural similarity indexes based on correlation in the wavelet domains. The tests are performed on 188 mammograms from the DDSM database and 94 from the miniMIAS database. This method also used three classifiers (LDA, Bayesian, and an artificial neural network of radial basis), and the best results had an accuracy of 94%.

The methodology presented in Kelder et al. [22] includes a bilateral approach based on the analysis of mammographic tissue finding ROIs, based on the asymmetry of pairs of left and right mammograms. Image registration is used to align the breasts, making the comparison of pairs of mammograms easier. Next, a Bayesian classifier is used. The tests are applied to 161 pairs of mammograms of the miniMIAS database. The results show an area under the FROC of 0.87 as the best result.

2.3. Detection of masses using convolutional neural network

This section presents some works that use convolutional neural networks in medical images.

In Arevalo et al. [23], the authors presented a CNN architecture for the classification of regions of the breast in lesions of the mass type. The authors highlight the fact that CNN architectures do not explicitly require the characteristic extraction step, which is a fundamental step in the task of image analysis. The authors apply tests to the Breast Cancer Digital Repository (BCDR) and cite that the experimental results showed that the adopted strategy surpasses the representation of the state of the art, presenting an area under the FROC of 0.86.

The work proposed by Dhungel et al. [24], authors present a novel approach for detecting masses in mammograms using a cascade of deep learning and random forest classifiers. The first stage classifier consists of a multi-scale deep belief network that selects suspicious regions to be further processed by a two-level cascade of deep convolutional neural networks. The regions that survive this deep learning analysis are then processed by a two-level cascade of random forest classifiers that use morphological and texture features extracted from regions selected along the cascade. Finally, regions that survive the cascade of random forest classifiers are combined using connected component analysis to produce state-of-the-art results. The final mass detection produced by this approach achieves the best results on DDSM of true positive rate of 0.75 at 4.8 false positive per image.

The work proposed by Suzuki et al. [11] discusses the difficulty of using CNN in medical imaging because of the problem of collecting a large amount of training data. To solve this problem, the methodology adopted a learning transfer strategy. The researchers also demonstrated the possibilities of CNN and the transfer learning strategy for the detection of masses in mammographic images. The study presents a CNN architecture consisting of 8 layers with weights, including 5 convoluted layers and 3 fully connected layers. This CNN was trained using about 1.2 million natural images to rank 1000 classes. Then, the last completely connected CNN layer was modified, and CNN was again trained using 1656 regions of interest in the mammographic image for two classes of classification: mass and normal. A detection test is performed on 198 mammographic images, including 99 mass images and 99 normal images. The experimental results showed that the sensitivity of mass detection was 89.9%, and the *FPi* was 19.2%.

In Kooi et al. [25], the authors provide a head-to-head comparison between a state-of-the art in mammography CAD system, relying on a manually designed feature set and a convolutional neural network, aiming for a system that can ultimately read mammograms independently. Both systems are trained on a large data set of around 45,000 images and results show the CNN outperforms the traditional CAD system at low sensitivity and performs comparable at high sensitivity. The use of CNN alone achieved an accuracy of 92,9%.

2.4. Considerations of related work

The presented works are relevant in the scientific environment, obtaining significant results in the task of mass detection. However, some peculiarities can be mentioned in these works:

- The studies that perform mass detection without using bilateral analysis use information from only one patient's breasts to detect the masses, and consequently a smaller number of images are used in these methodologies;
- By inserting bilateral information for the detection of masses, a patient is analyzed completely, detecting not only the masses but also asymmetric regions between pairs of mammograms;
- Works such as Ereceira et al. [7], achieved up to 100% sensitivity by adding bilateral information;
- It is known that breast density is a complicating factor in the task of mass detection, especially when the detection method takes into consideration texture, since the normal tissue texture in dense breasts is similar to mass texture in non-dense breasts;
- The methodology proposed by Sampaio et al. [18] shows the importance of treating breast density prior to detecting the mass;
- All segmentation methods (both texture and structure geometry to detect suspicious regions) generate several false positives in detection, hence the need for steps to eliminate regions that are non-masses;
- The task of detecting masses using classifier is not always trivial, since a good classifier depends directly on a good choice of characteristics;
- Convolutional neural networks have been growing significantly in the task of image classification, not only because they do not require an explicit feature extraction step, but also because they can classify several image domains.

In view of this, a methodology is proposed for the bilateral analysis of pairs of mammograms adapted to the density of the breast. Biological similarity indexes will be used to compare paired mammograms and to segment regions of interest. Then, false positive reduction techniques are used, eliminating regions that are non-masses and classifying the remaining regions into masses and non-masses from the use of convolutional neural networks. In addition, the methodology uses a CNN to classify the breasts by the density, producing a method adaptable to any breast tissue.

3. Materials and method

The proposed methodology is divided into two phases of training and testing. These phases are explained in detail after we explain the materials used in the methodology.

3.1. Materials

For the proposed methodology, a public image database was used. Known as the Digital Database for Screening Mammography (DDSM), it is widely used in the literature for validation of methods. This database contains more than 2500 exams purchased from Massachusetts General Hospital, Wake Forest University, and the University of Washington at St. Louis Medical School [26].

Each exam contains up to four images, two sides (left and right), and two projections [craniocaudal (CC) and mediolateral oblique (MLO)], as well as extra information about the exam (breast density, study date, patient age, type of pathology, number of anomalies, etc.) and about the image (file name, image type, scan date, scanner type, sequence, pixels per line, bits per pixel, lesion location, etc.). All information contained in the DDSM was provided by experts [26].

Breast density is an important factor in this work because different models will be created for different types of density. In DDSM, the breasts assume a value of 1 to 4 according to the BI-RADS, and the interpretation of these values is defined in Table 1

As in the work of Sampaio et al. [18], non-dense breasts were considered those with values 1 and 2, and dense images had values of 3 and 4, according to Table 1. In our method, we use the same criterion.

In this work, we used 1241 pairs of mammograms of DDSM, of which 502 were non-dense breasts and 739 were dense. These pairs of mammograms must contain at least one mass lesion. Images with another type of lesion were not considered in this study.

3.2. Training phase

In this phase, 80% of pairs of mammograms were used to create models capable of classifying breast into dense and non-dense and classifying segmented regions into regions of mass and non-mass. Fig. 1 illustrates the stages of this phase.

Initially, the density model will be presented to classify the breasts into dense and non-dense. Then, the classification models the mass regions and non-masses in dense and non-dense breasts.

3.2.1. Creation of the density model

At this stage, we choose to isolate the breast. For that, the preprocessing step was used. Then, regions are extracted from the breast, and a CNN is trained to classify new cases.

Preprocessing

The preprocessing step has the task of isolating the breast and discarding some peculiarities that the DDSM images have and are not necessary for the study, such as patient information, background noises, and slight deformations on the edges.

Initially, the original size of the mammograms is resized. Images from the DDSM have an average height of 6,000 pixels. In this work, the images were reduced to a height of 1,024 pixels by a proportional width to decrease the computational time. Some works in the literature adopted this practice and show that the resizing does not cause negative impacts on the results [4,7,9,10,18].

The proposed methodology adopts the technique of bilateral comparison of pairs of mammograms to verify asymmetric regions. In the literature, there are many works relating this type of technique to the existence of lesions, as presented in Section 2. To facilitate the registration step, avoiding unnecessary deformations, and also to facilitate the comparison in the segmentation step, it was chosen to mirror one of the breasts so that the two breasts of the same patient were locally close. This step can be seen in Fig. 2b.

To prepare the images for the other stages of the methodology, it is necessary to isolate the breast. The existence of undesired structures in the digitized mammograms (noise, borders and marks) may be disrupted and may not be relevant to the purpose of this methodology.

A methodology based on Sampaio et al. [18] was used. First, there is a removal of edges, where 30 pixels away from the side edges are removed. Then, a background removal is done. In this step, the pixel values are divided into two groups according to the intensity. The first is formed by the pixels of higher intensity and the second of the lower ones. Thus, the group of lower intensity has its values replaced by 0 (equivalent to black color).

Finally, to remove the external objects from the breast, a region growth algorithm is used. The seed of the region growth is positioned in the center of the half of the image whose sum of values of pixels is higher (that is, in the half where the breast is). The region growth stops when it reaches values of 0. The resulting image of region growth is then used as a mask over the original image, resulting in the end of the preprocessing step, and then the isolated breast. This step can be seen in Fig. 2c.

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.

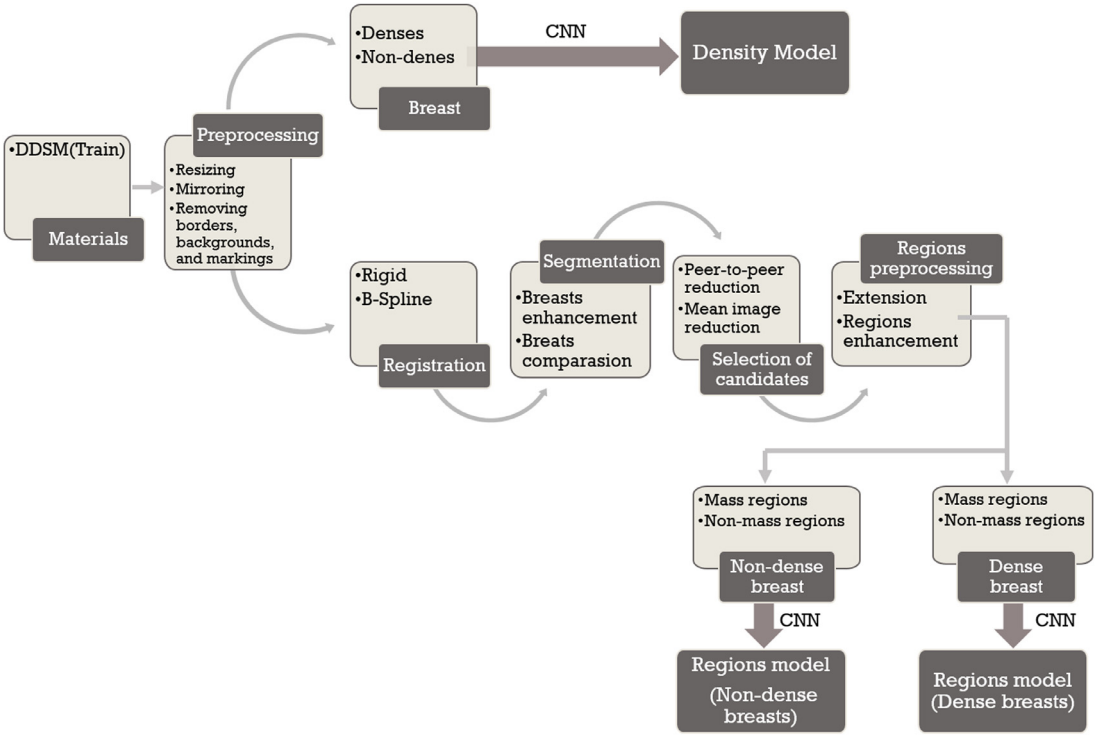

Fig. 1. Scheme of the training phase of the methodology.

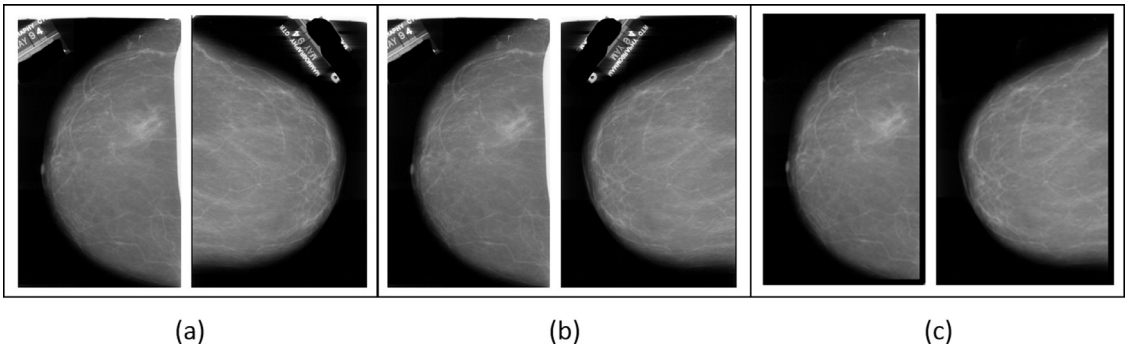

Fig. 2. Preprocessing step: (a) original images; (b) an image is mirrored; (c) both images removed edges, background and marks.

Density model

It was decided to create a model to classify breasts according to their density, since masses in non-dense breasts usually present a high density that can be confused with healthy tissue of the dense breasts, making it impossible to use a single model to classify masses in different types of breasts. For this task, a convolutional neural network was trained.

To generate a density model, windows of 128×128 pixels were extracted from dense and non-dense breasts of the DDSM for those windows to be trained by CNN. This size of the window was used because even with the division between dense and non-dense, the breasts present many internal variations of density, and with a larger window, more information will be presented to the CNN.

The convolutional neural networks are composed of 4 layers: convolutional layers, pooling layers, activation layers, and fully connected layers. The network extracts implicit features of visual patterns presented by the input and classifies patterns from the extracted features. The output of each layer serves as input to the next [27,28].

Knowing this, our architecture consists of two convolutional and two pooling layers interscaled, and finally a fully connected layer. The convolutional layers have trainable filters that are applied throughout the input image. Each filter detects a peculiar feature that occurs in any part of the input image. Eq. (1) describes the process of the convolution layer [27].

$$(h_k)_{ij} = (W_k * x)_{ij} + b_k \tag{1}$$

where k is the neuron index, h_k is the kth output feature maps, W_k is the kth filter, x is the input feature map, b_k is the bias term and "$*$" represents the two-dimensional spatial convolution.

Pooling layers are non-linear downsampling layers. In our work, we adopted a max-pooling function that yields maximum values over a neighborhood of feature maps. This results in a reduced-resolution feature map that is robust to variations in the feature locations (rotation and translation) of the previous layer.

Activation layers apply non-linear activations to input neurons. The most common activation functions are the hyperbolic tangent function, softmax function, sigmoid function, and rectified linear unit (ReLU). However, a CNN with ReLUs trains several times faster than other activation functions [29]. The ReLU function applies a simple threshold at zero and can be defined as shown in Eq. (2).

$$f(x) = max(0, x) \tag{2}$$

where x is the input to a neuron.

The last layer of the network is a fully connected layers that is responsible for classifying patterns presented to CNN, normally a multilayer perceptron (MLP). The network output is computed by

$$y_k = \sum_n W_{kn} x_n + b_k \tag{3}$$

where y_k is the kth output neuron, W_{kn} is the weight connecting x_n with y_k, x_n is the nth input neuron and b_k is the bias term of y_k.

The number of input neurons for the fully connected layer is defined by the number of pixels resulting from the layer prior to it.

Supervised training is carried out using a form of stochastic gradient descent (SGD) to minimize the discrepancy between the desired output and the current output of the network, based on some loss function.

Given this, our architecture works using the following scheme:

1. Input image;
2. Convolution layer with 20 filters of size 5×5;
3. Pooling layer with max-pooling function of size 2×2;
4. Convolution layer with 50 filters of size 5×5;
5. Pooling layer with max-pooling function of size 2×2;
6. Fully connected layer composed of an input layer with the number of pixels of the images prior to that layer, followed by an activation layer with ReLU functions, and an output layer with a softmax function giving the probability of the input to belong to either class (dense and non-dense).

The result of this step produces a density model that will be used in the test phase. The next step is to create models for classification between masses and non-masses, in both dense breast and non-dense breasts.

3.2.2. Creation of regions classification models

To create models that classify regions of the breast into masses and non-masses, some techniques that are necessary to select candidates will be later submitted to the CNN. These steps are described below.

Preprocessing and image registration

The preprocessing used in this step is the same as described in Section 3.2.1, for the same purpose of isolating the breast.

After preprocessing, the images went through the registration process. The registration of images is of great importance in several areas, especially in the medical area. The literature shows the importance of image registration in several works that use medical images [30,31]. This has been used in the segmentation, diagnosis, and temporal analysis of diseases [7,32–34].

Image registration consists of the following elements: two images, one fixed and one moving; a measure of similarity; a transformation function; an optimizer; and an interpolator [35,36].

The measure of similarity is a numerical measure that indicates how much the moving image is similar to a fixed image. The transform function is a function that maps the moving image to the still image. This transformation function is calculated by the optimizer by optimizing some measure of similarity. In this way, the registration can be seen as an optimization problem, in which an optimizer estimates the transformation function that best maps the fixed image of the mobile image, according to a value obtained by the measure of similarity.

Historically, image registration is divided into two groups: rigid and deformable (or non-rigid) registrations. Rigid registration considers that the moving image is an object that needs to be rotated and translated so that there is a satisfactory match with the fixed image. In this work, we used two types of registration: rigid registration and deformable registration B-Spline. For this, the moving image is defined as the one that was mirrored, and the fixed image the other. Information on these techniques was described by Brown et al. [35], Zitova et al. [36] and Shackleford et al. [33].

It is important to note that preprocessing, previously performed, removes unwanted elements from the image and thus enables better results in the registration, which takes into consideration only the breast. After the registration step, some deformations at the borders from the effects of translations and rotations can be seen at the borders of the image. To solve this problem, another step of removing the background, borders, and marks is used. An example of this step can be seen in Fig. 3.

With the breasts aligned, the next step is the segmentation, where regions of the left and right breasts will be compared in order to find asymmetries that may be indicative of masses.

Segmentation of asymmetric regions

The segmentation of images is the process of subdividing the images in order to find objects of interest. In medical images, the segmentation step plays a crucial role, most of the time searching to delimit a region of interest within the CAD/CADx system. In the medical area, there are several works on various purposes of segmentation depending on the mode of the images [37–39]. Currently, many techniques have been developed for segmentation in medical images. The literature demonstrates the importance of this step for the construction of diagnostic aid systems [40].

The segmentation step seeks to highlight regions of the breast with some asymmetry that may be masses. In this work, we propose the use of similarity indexes of biology [41] to highlight the asymmetries between pairs of mammograms. However, before the application of the indices was chosen to enhance internal structures of the breast in order to make a better comparison.

To enhance the internal structures, Contrast-Limited Adaptive Histogram Equalization (CLAHE) [42,43] was applied, followed by a mean filter [43] of kernel 3×3.

The image is divided into side by side windows of 13×13 pixels. This window size generated the best results and less mass loss after the segmentation step. Therefore, each pair of images of a patient will have the same number of windows, so it will be possible to calculate the similarity indexes of the communities. (In this context, each window is a community.)

Each window in one breast has its corresponding one in the other. Thus, the similarity between these two windows will be analyzed by using the similarity indexes of biology, which measure how much two communities are similar. In this work, we used the Jaccard, Anderberg, Czekanowsky, Kulcynski 2, and Ochiai [41] indexes. These indexes were used because they have a range of occurrence between 0 and 1, enabling us to measure whether the corresponding windows are similar, where 0 communities are to-

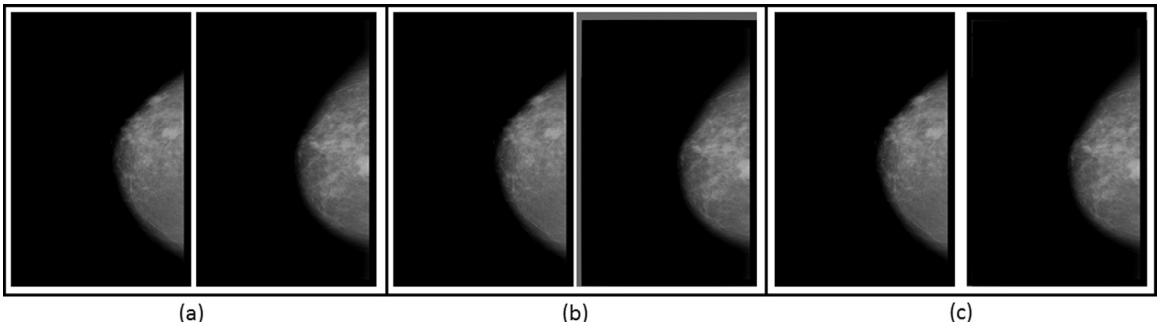

Fig. 3. Registration step: (a) images preprocessed; (b) image of right breast preprocessed and image of left breast after rigid and B-spline registration, (c) image of right breast preprocessed and image of left breast after removing background, borders, and marks.

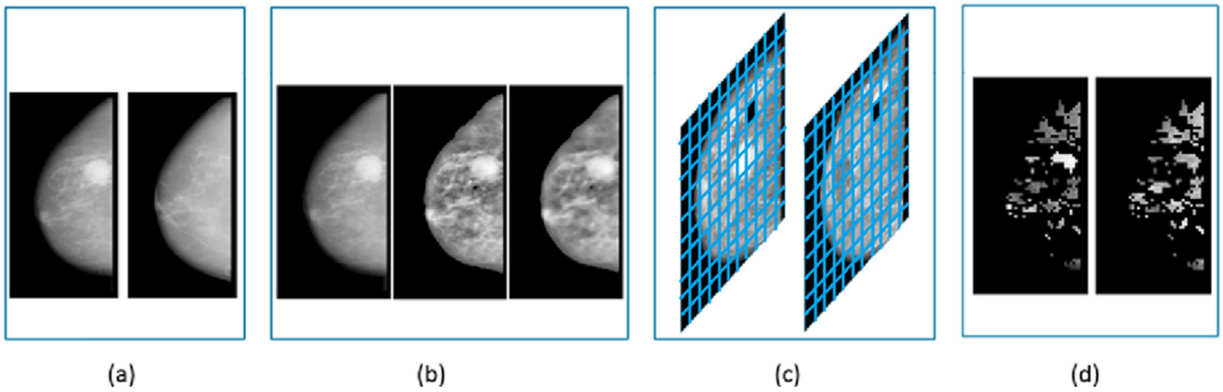

Fig. 4. Segmentation step: (a) input images, (b) images applied CLAHE and mean filter, (c) images divided by windows of 13×13, and (d) images with segmented regions by similarity indices.

Table 2
Similarity indexes [41].

Similarity indexes	Equations	Range of occurrence
Jaccard	$\frac{a}{a+b+c}$	[0, 1]
Anderberg	$\frac{a}{a+2(b+c)}$	[0, 1]
Czekanowsky	$\frac{2a}{2a+b+c}$	[0, 1]
Kulczynski 2	$\frac{a}{2}(\frac{1}{a+b}+\frac{1}{a+c})$	[0, 1]
Ochiai	$\frac{a}{\sqrt{(a+b)(a+c)}}$	[0, 1]

tally dissimilar, and where 1 is totally similar. The equations referring to these indexes are defined in Table 2.

According to Meyer et al. [41], given two groups (groups A and B) that wish to compare the similarity between them, the abovementioned indexes use 3 variables to calculate the similarity. However, in the context of image processing, and to measure the similarity between two corresponding windows (window 1 in the left breast and window 1 in the right breast), the variables are redefined to:

- a: pixels present in window 1 of the left breast and window 1 of the right breast;
- b: pixels present only in window 1 of the left breast and absent in window 1 of the right breast;
- c: pixels present only in window 1 of the right breast and absent in window 1 of the left breast.

For each pair of windows (one on the left breast and the one on the right breast), the five indexes are calculated. Windows in which all indices were equal to 0 (i.e. are pairs of completely dissimilar windows) will be maintained and the rest discarded, leaving only dissimilar regions in the images.

An illustration of the stages of the segmentation step is shown in Fig. 4.

With the segmented images, the next step is to filter these various regions to selected candidates to be submitted to the CNN.

Selection of candidates

The step of selecting of candidates attempts to eliminate excess regions that will not be useful for the creation of mass region classification models. For this, two techniques are defined: peer-to-peer reduction (PPR) and mean-per-image reduction (MIR). These two techniques will be explained below.

In PPR, the average of the pixel values for each window is first calculated, as described in Eq. (4).

$$M_{win} = \frac{1}{n^2}\sum_{i}^{n}\sum_{j}^{n} p_{ij} \tag{4}$$

where p_{ij} is the pixel value and n is the size of the windows. After calculating the average, the values M_{win} of the corresponding windows in the left and right breasts of the same patient are compared.

Denoting M_{winL} as the average value of any window in the left breast and M_{winR} the average value of a corresponding window in the right breast, we define Eq. (5).

$$\begin{cases} left \;\; windows \;\; discarted, \;\; if \;\; M_{winL} < M_{winR} \\ right \;\; windows \;\; discarted, \;\; if \;\; M_{winR} < M_{winL} \end{cases} \tag{5}$$

Now, in the MIR, no more pairs of mammograms will be used, but only one mammogram at a time. In this second approach, we first calculate the overall average of the image defined by Eq. (6).

$$M_g = \frac{1}{m}\sum_{i}^{m} M_{win,i} \tag{6}$$

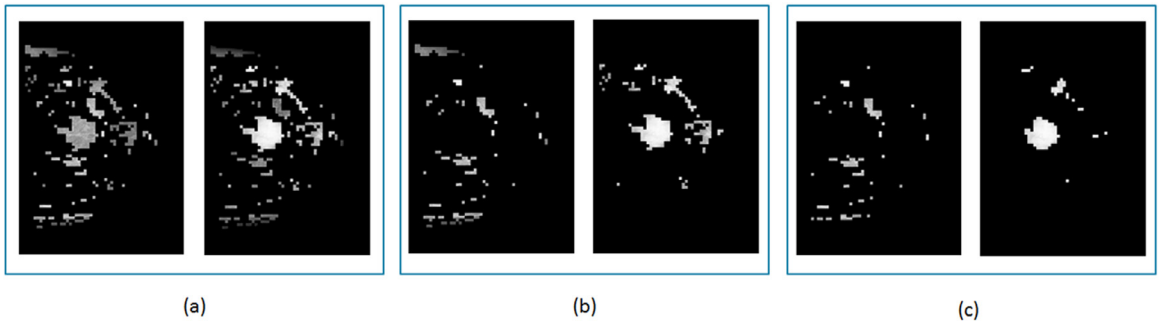

Fig. 5. Selection of candidate step: (a) segmented images as input, (b) images applied to PPR, and (c) images applied to MIR.

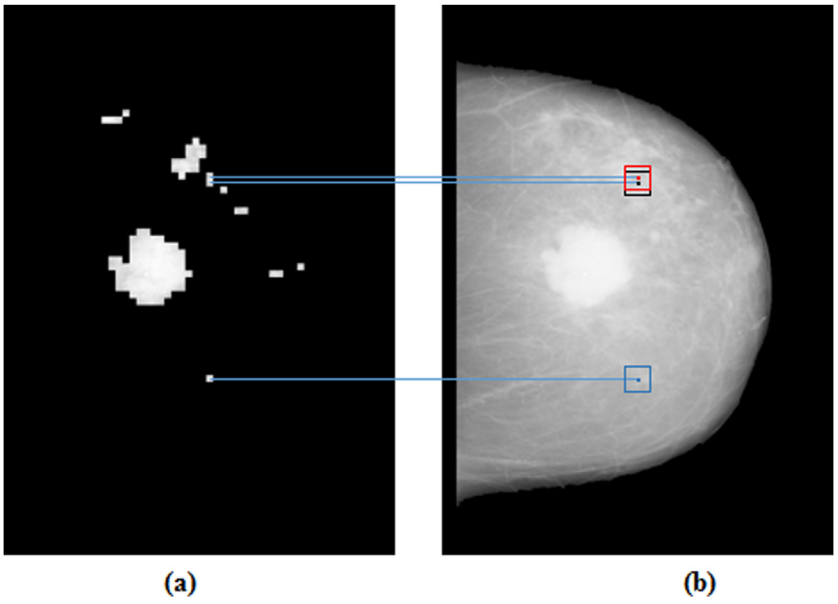

Fig. 6. Preprocessing of regions step: (a) dissimilar regions from the previous step, and (b) expanded regions.

where $M_{win,\,i}$ is the average of the window pixels i (Eq. (4)) and m is the total number of windows in the image. Thus, M_g is the sum of all averages divided by the number of windows of each mammography image.

Finally, all windows of the image are checked by comparing M_{win} of the window with the value of M_g, defining which windows will be discarded. This verification is presented in Eq. 7.

$$\begin{cases} window\ \ discarted, & if\ \ M_{win} < M_g \\ window\ \ maintained, & if\ \ M_{win} > M_g \end{cases} \tag{7}$$

At the end, there are regions with the highest averages according to the overall mean of the image. A result of this application can be seen in Fig. 5.

After the two techniques for candidate selection, the remaining regions will serve to create a classification model of mass and non-mass regions. However, before being subjected to classification, the regions will undergo a preprocessing process.

Preprocessing of regions

Because the convolutional neural network presents layers of convolution and pooling, using regions of 13×13 would generate few low-level features in the fully connected layer, knowing that the 13×13 sized windows regions will be expanded to 32×32. For this expansion, the central pixel of the 13×13 window is first denoted, and then a new window of 32×32 is defined with all neighboring pixels of the breast tissue to compose a 32×32 window. An illustration of this expansion can be seen in Fig. 6.

With the expanded regions, the next step is to apply a number of improvement techniques in order to highlight intrinsic characteristics of these regions. An approach based on Prakash et al. [44] was used. A method of highlighting and segmentation is proposed where noise removal and segmentation of texture region characteristics are prioritized. The proposed algorithm is divided into 10 steps, which were redefined to be applied in the regions extracted from the breasts. The steps are listed as follows:

1. Read the entry window. Apply a scale by doubling the image area (image interpolation);
2. Median filter in the image obtained in step 1;
3. Apply a sub-sample to the image obtained in step 2, reducing it to the size of the original image;
4. Use the CLAHE in the image from step 3;
5. A structuring element of a line is defined. This element is rotated between 0 to 180 with intervals of 15°. The bottom hat of the image from step 4 is obtained for each variation of the structuring element, generating 12 output images in this step;
6. Add up the 12 images from step 5, thus producing a high-quality image where local objects are identified more clearly;
7. The average filter is applied to the image obtained at 6. This will produce an image with noises, which will be interesting for the next steps;
8. The noise picture obtained in step 7 will be subtracted from that of step 6 (i.e., sum of the bottom hat minus the filter output from the mean) by removing the noises;

Fig. 7. Preprocessing of regions step: application of improvement techniques.

9. The CLAHE is applied in the image from step 8;
10. The last segmentation step is applied to a Laplacian of the Gaussian in the image resulting from step 9.

All filters used in this sub-step used 3×3 kernels. This kernel size was used because the input images were of very small size (32×32), and increasing the size of the kernel could not achieve the effect of highlighting the intrinsic regions.

The result of applying this algorithm is an image with segmented internal regions. An example of the application of this algorithm in a segmented breast region can be seen in Fig. 7 (step 5, where 12 images are generated, was not shown in the figure, since step 6 is the sum of those 12 images).

From the training images, we separate the regions of the masses and non-masses of both the dense and non-dense breasts, according to the marking file informed by the DDSM. Now, two region classification models will be created, one for dense breasts and one for non-dense breasts.

Models of regions in dense and non-dense breasts

After separating the masses and not-masses in both the dense breasts and the non-dense breasts, the images are submitted to a CNN where two classification models will be created. The CNN architecture used in this step is the same one used to create the density model composed of two convolution layers, two subsampling layers, and one fully connected layer.

At the end of this stage (and the training phase), there are three models: one for density classification, one for classification of masses and not-masses for dense breasts, and one for classification of masses and not-masses in non-dense breasts. Now, these models will be presented on a test base to classify the density and the segmented regions of the breasts.

3.3. Test phase

At this phase, the models created in Section 3.2 will be applied to the images of the separate test database of the DDSM. For this, some steps are required. These steps are illustrated in Fig. 8.

The pre-processing, registration, and segmentation steps are the same as those already presented in the training phase of the model. In the test phase, after the segmentation stage, the first reduction of false positives (1RFP) is done in order to eliminate regions that are not masses.

In 1RFP, the two techniques used in the selection-of-candidates step in the training phase are applied: peer-to-peer reduction and mean reduction by image. The next step after 1RFP is the preprocessing of regions. This step was also the same presented in the training of the model.

It is noteworthy that the steps of the test phase resemble those of the training phase of the model. This is necessary so that when tested, the models are robust both in the classification of the density and in the detection of the regions of masses.

In the next step, the breasts will be classified into dense and non-dense before classifying the regions of masses and non-masses.

3.3.1. Density classification

In this step, the model created in the training phase will classify the breasts as dense and non-dense. For this, a single region of each breast is extracted from the test base, and this region will be classified by the generated model, classifying the breast as dense or non-dense. The region is extracted from the center of the breast.

First, the breast width is found (red line in Fig. 9)). Then, half of that width is defined as the central pixel (red dot in Fig. 9), and a 128×128 window is extracted around it (black rectangle in Fig. 9).

As mentioned in the training, this window size was used because even with the division between dense and non-dense, the breasts present many internal variations of density, and with a larger window, more information will be presented to the CNN.

3.3.2. Second false positives reduction - 2RFP

With the test base classified and divided into dense and non-dense, the next phase is the second reduction of false positives where the regions of each type of breast will be classified, for which the models generated in the training phase are tested in the segmented regions of the breast.

Classification of non-dense breast regions

In this stage, the regions that were already segmented and pre-processed are classified by the model for non-dense breasts generated by the CNN in the training phase. The model will classify the regions of the test base in masses and non-masses.

Thus, at the end of this classification is the detection of the regions of mass. Therefore, metrics for validation of this model are calculated for validation of the methodology.

Classification of dense breast regions

Analogous to the classification of non-dense breasts, at this step, the regions that were already segmented and pre-processed are classified by the model for dense breasts generated by the CNN in the training phase. The model will classify the regions of the test base in masses and non-masses.

Therefore, at the end of this classification is the detection of the regions of mass in dense breasts. Finally, metrics for validation of this model are calculated to validate the methodology.

4. Results and discussion

In this section, we will present the models generated in the training phase of the proposed methodology and the results achieved in the test phase from these models.

First, the use of the DDSM to test the methodology is highlighted. It used 1241 pairs of mammograms, producing 2482 mammographic images. These images were divided into two datasets: a training base with 80% of that total and a 20% test base. This division was done in a totally random manner, where it was guaranteed that pairs of the same patient would be in the same base.

The training base was used in the training phase of the methodology proposed for model creation and went through the pipeline described in Section 3.2.

4.1. Training phase

Eighty percent of the 1241 pairs of mammograms were used to construct the training base, resulting in 402 pairs of non-dense

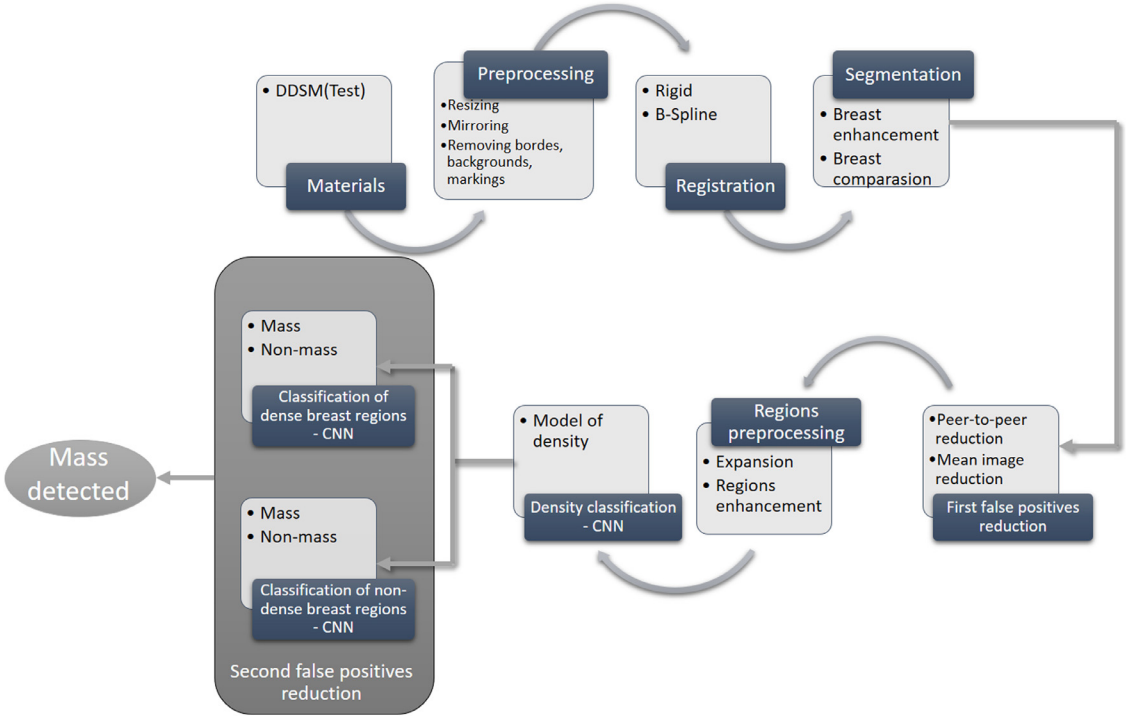

Fig. 8. Scheme of the test phase of the methodology.

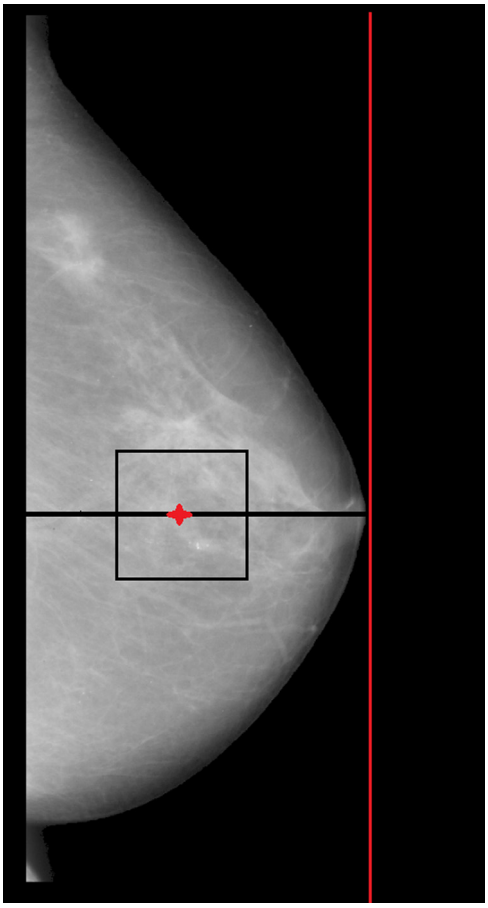

Fig. 9. Illustration of selection of region to be extracted from images of test base. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

mammograms and 593 pairs of mammograms of dense breasts. First, the density model was created, followed by the region classification model of both dense breasts and non-dense breasts.

4.1.1. Density model

As mentioned previously (Section 3.2.1), several 128×128 windows are extracted from the mammograms to create a model that can classify the breasts into dense and non-dense. There were several variations of extraction of the breast windows from the training base; however, the best result was when 5 regions of the non-dense breasts (4020 regions) and 3 regions of the dense (3558 regions) were extracted.

In the training of the density model, we used 100 epochs as a stopping criterion with a learning rate of 0.001, an input batch of 64 images, and 10% of the total windows for validation.

4.1.2. Model of region classification

To create region models, the images go through the processes of preprocessing, registration, segmentation, selection of candidates, and region preprocessing. After that, we obtain the regions that will produce the models capable of classifying masses and non-masses in dense and non-dense breasts.

In the non-dense breast region model, a training base composed of 402 pairs of non-dense breast images (after the above pipeline) produced 5416 regions of mass and 21,174 regions of non-mass. These regions were submitted to the CNN; however, variations were made in these regions in order to find a better model that could correctly classify the test base.

By making these variations in the regions, four training bases were created for model generation, as follows:

- Base 1: Original proportion, 5416 regions of mass and 21,174 regions of non-mass;
- Base 2: All regions of mass and proportional amount of non-mass regions, resulting in 5416 regions of mass and 5416 regions of non-mass;

- Base 3: The regions of mass are rotated at 90, 180, and 270° and use the original number of non-masses, resulting in 21,264 regions of mass and 21,174 non-mass;
- Base 4: The proportional amount of mass and non-mass regions used in base 2 are rotated at 90, 180 and 270°, resulting in 21,264 mass regions and 21,264 non-mass regions.

These bases were submitted to the CNN, which obeyed the criterion of 100 epochs, 0.001 learning rate, and 64 images in a batch, and used 10% of the total for validation. Thus, in the training phase, four models were produced from the four training bases cited for the classification of regions in non-dense breasts that will be tested in the test phase.

To create the dense breast region model, the 593 pairs of images of dense breasts produced 6347 regions of masses and 35,798 regions of non-mass after passing through the mentioned pipeline; however, variations were made in these regions so that it could find a better model. We also created 4 training bases for model generation, described below:

- Base 1: Original proportion, 6347 mass regions and 35,798 non-mass regions;
- Base 2: All regions of mass and proportional amount of non-mass regions, resulting in 6347 regions of mass and 6347 regions of non-mass;
- Base 3: The regions of mass are rotated at 90, 180 and 270° and use the original number of non-masses, resulting in 25,388 regions of mass and 35,798 of non-mass;
- Base 4: The proportional amount of mass and non-mass regions used in base 2 are rotated at 90, 180 and 270° , resulting in 25,388 regions of mass and 25,388 regions of non-mass.

The 4 bases were submitted to the CNN, which obeyed the criterion of 100 epochs, 0.001 learning rate, and 64 images in a batch, and used 10% of the total for validation. Four models were created from the 4 bases mentioned. These models will be applied in the test phase in order to find the model that obtains the best result.

4.2. Test phase

In this phase, 20% of mammograms from DDSM have gone through the steps described in Section 3.3. The 20% resulted in 246 pairs of mammograms where 100 pairs of mammograms were non-dense breasts containing 121 masses, and 146 pairs were dense breasts with 205 masses. It should be noted that the density, location, and quantity of masses were obtained from the tag files that accompany the images of the DDSM database cited in Section 3.1.

4.2.1. Preprocessing and image registration

Mammograms were submitted to the preprocessing process. First, they were resized. Then, one of the breasts was mirrored, and finally the breast was isolated with the process of removing edges, bottom, and marks.

All pairs of mammograms used in the work went through the registration process. Mammography that had not been mirrored in the preprocessing was treated as the fixed image, and the one that had been mirrored as the moving image. The moving image passes through the sequence of the two registers: rigid and B-spline.

4.2.2. Segmentation of asymmetric regions

Then, the mammography pairs were enhanced and their corresponding windows compared in search of regions considered asymmetric. In the segmentation step, there was a loss of 12 masses: 4 in the dense breasts (96.7% accuracy) and 8 in the non-dense breasts (96.1% accuracy). Fig. 10 shows an example of success in the segmentation phase.

In this case, it is noted that the similarity indexes used for segmentation were able to find the mass when comparing left and right breasts.

When there is a loss of mass in the segmentation, it usually occurs that the mass in one of the breasts when compared to the normal tissue of the other presents equal values of pixels, causing the indices to present values different from zero. This can occur either because of the image acquisition process (generating equal values in the mass position) or in small masses when the mass occupies a size close to the 13×13 window where the indexes are calculated.

At the end of the segmentation step, there are 31,657 regions of 13×13, where 5129 are regions of masses and 26,528 are regions of non-masses.

4.2.3. First reduction of false positives

In the first reduction of false positives (1RFP), the images from the previous step undergo two reduction techniques proposed in the methodology: par-a-pair reduction, mean-per-image reduction. The images in which there was loss of mass were not used at this step, so of the 246 pairs of mammograms, only 234 were used in this step.

The purpose of this step is to discard regions that are not masses. At the end of the 1RFP applied to the entire test base of the non-dense breasts, there was a loss of only 2 masses; when applied to dense breasts, there was a loss of 9 masses.

An example of the segmentation sequence along with 1RFP can be seen in Fig. 11.

It is observed in the result of 1RFP in Fig. 11, and that only with the segmentation steps and 1RFP, was it possible to identify the mass in the image (yellow region) according to the specialists marking (red marking) with only a false positive (green region).

At the end of this step, we discard pairs of mammograms in which it was not possible to find the masses, leaving only 223 pairs equivalent to 446 mammography images. Thus, there are 15,178 regions, where 4605 are regions of masses and 10,573 of non-masses. In the segmentation phase, there was a reduction of about 52% in the segmented regions, which shows the efficiency of the proposed technique for the reduction of false positives.

4.2.4. Preprocessing of regions

Now regions of 13×13 go through the region preprocessing process. Regions will be expanded and enhanced by a set of enhancement techniques.

Before these regions are classified into masses and non-masses, the breasts will be classified as dense and non-dense according to the model generated in the training phase.

This step is necessary to create two models to classify regions in both kinds of tissue. Using a single model to classify all regions in both dense and non-dense breasts would not be a trivial task because normal tissue in dense breasts resembles mass tissue in non-dense breasts.

4.2.5. Density classification

In this stage, the 446 mammography images are classified. First, a region of the breasts is extracted from the test base (Section 3.2.1), and is sorted by the density model produced in the training phase (Section 4.1.1).

According to information from the DDSM, 182 mammograms were non-dense, and 264 were dense. Thus, our model classified 180 breasts as non-dense (being able to correct classify 174 according to the specialist and classifying other 6 that were dense as not dense) and 266 as dense (where 258 were classified correctly and the others 8 were not).

Thus, our model reached a classification of the density of 95.60% of non-dense breasts and 97.72% correct in the dense

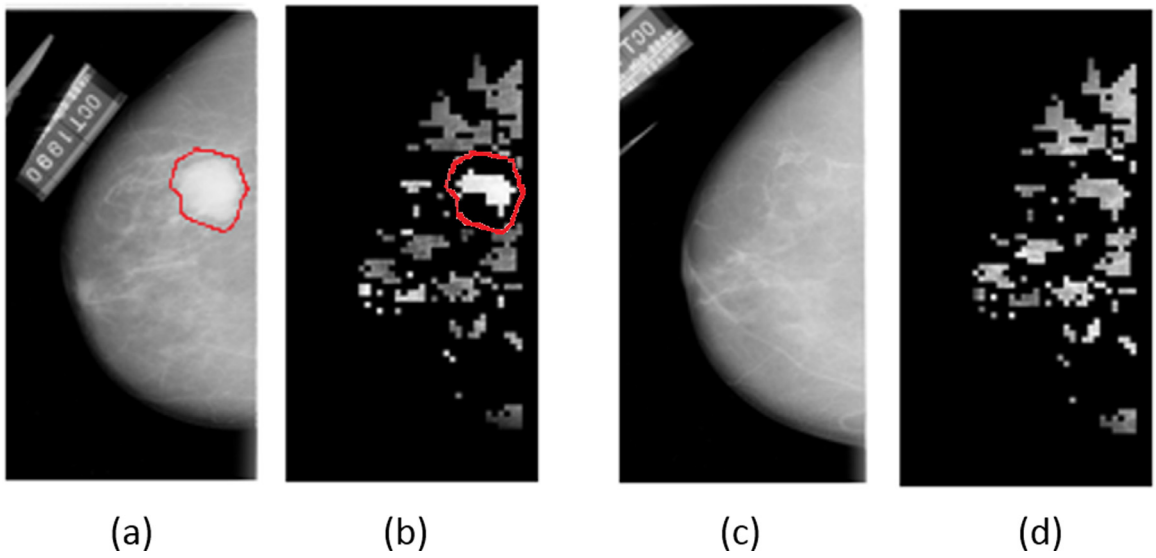

Fig. 10. Patient A_1006_1 from DDSM: (a) left breast with lesion, (b) left breast segmentation, (c) right breast, and (d) right breast segmentation.

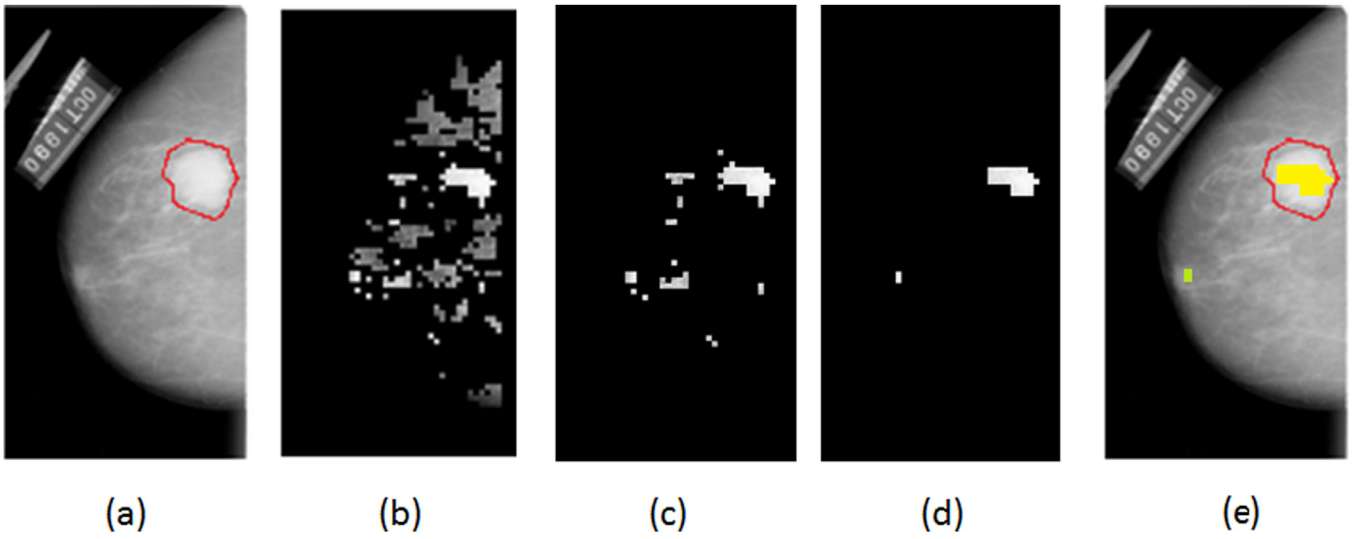

Fig. 11. Patient A_1006_1 from DDSM: (a) right breast with lesion, (b) right breast segmentation, (c) peer-to-peer reduction, (d) mean-peer-image reduction, and (e) mask of 1RFP result on original image.

breasts, generating a total accuracy mean of 96.86%. These results show how much the proposed methodology is able to classify dense and non-dense breasts, showing an efficiency of almost 97% in this task.

Even in the model classifying the different densities of the annotations yielded by the database, the methodology used the result of this classification to test the generated models to classify regions of masses in dense and non-dense breasts. Thus, the next step will be tested on the images classified by the density model, that is, 180 non-dense breasts and 266 dense breasts.

It should be noted that errors in density classification may negatively influence the region classification step because normal regions in dense breasts have high densities equivalent to regions of mass in non-dense breasts.

4.2.6. Second reduction of false positives

In the second reduction of false positives, the model generated for the classification of non-dense regions obtained in Section 4.1.2 will be applied to the 180 non-dense breasts classified by the density model. Likewise, the model generated for the classification of regions in dense breasts, will be applied to the 266 dense breasts classified by the density model.

Table 3
Results of models created in training phase applied to non-dense breasts.

Bases	Sensitivity	Specificity	Accuracy
1	91.56	90.73	91.04
2	90.38	47.36	62.75
3	84.41	72.87	76.99
4	88.98	54.65	66.93
Standard deviation	3.131	19.445	12.584

In the non-dense breast base of mammograms, 2643 were mass regions and 4748 were non-mass regions. Then, the models created by the training bases described in Section 4.1.2 were applied on a test basis, and the results of the validation metrics are presented in Table 3.

It can be observed that all variations of training bases presented a high sensitivity value. This shows that the models are able to detect the mass regions. However, bases 2, 3, and 4 did not present

Table 4
Results of models created in training phase applied to dense breasts.

Bases	Sensitivity	Specificity	Accuracy
1	90.36	96.35	94.84
2	96.07	47.68	59.87
3	94.18	51.25	62.07
4	91.43	85.28	86.83
Standard deviation	2.598	24.341	17.572

a satisfactory specificity. This implies that the models were not effective in classifying non-mass cases.

Even though the training bases were varied to find a better model, the best results resulted from the basis that used purely the extracted regions, obtaining a result of 91.5% sensitivity, 90.7% specificity, and 91.0% accuracy.

Another performance analysis of the model was the number of false positives per image (FPi), which resulted in 0.058 in the model created by base 1, showing that the methodology performed well and is quite promising.

The value of the FPi is very small because this metric is calculated by the number of regions. Thus, a specialists marking can have several regions, and since the number of total regions to be classified is very large, in the calculation of FPi it tends to low values.

With the results obtained by the model in this step, it can be observed that the methodology is able to generate a good classification in segmented regions in the non-dense breasts. In addition, the proposed method showed excellent results when dealing with false positives per image.

Now, the base of dense breast testing that is composed of 266 images, resulted in 1962 mass regions and 5825 non-mass regions. The models created and described in Section 4.1.2 were tested on this basis and the result of the validation metrics are presented in Table 4.

It was also observed in the case of non-dense breasts that all models applied to mammograms classified as dense had a high sensitivity value. The base-2 model was the most capable of detecting mass regions, generating a sensitivity of 96%.

However, when considering the values of specificity, it is noted that the models of base 1 and 4 were more effective in treating cases of non-masses. Notwithstanding, the best model was generated from base 1 (with no variation in regions), reaching a 90.4% sensitivity, 96.4% specificity, and 95.8% accuracy.

Analogous to that occurring in non-dense breasts, the false-positive rate per image stands out, resulting in 0.027, proving that the methodology is quite promising.

Similar to the results obtained in the non-dense breasts, it is observed that the methodology is able to generate a good classification in segmented regions in the images of dense mammograms. In addition, it is worth emphasizing that the classification model in dense breasts presented a better accuracy, showing that the methodology is more efficient in treating these types of patients.

The CNNs used in the methodology have the same architecture, both in density classification and in the detection of masses in dense and non-dense breasts. It is emphasized that this architecture and its parameters were used because in detection of masses, a 32×32 input image was used, and many layers of convolution and pooling would cause few pixels or none to reach the fully connected layer, making it impossible for the network to distinguish the classes between mass and non-mass.

On the other hand, if fewer convolution and pooling layers were used, fewer feature maps would be required, and consequently fewer features would be presented. The fully connected layer would generate few individual characteristics of each input image and proportionally more pixels since the image is larger because there would only be one layer of pooling, making the network slower and less discriminating.

In our approach, 13×13 regions are segmented and then classified. This patch of images does not have geometric or edge characteristics. In addition, these regions provide little information. We have only texture features, and using descriptors to extract those characteristics from them would be a difficult or impossible task. In this way, the use of the CNN becomes very important for our methodology. Therefore, because we use these images, we only need to present our image as an input from the network, and within its convolution and pooling layers the network can extract features automatically.

In addition, there is a need for a large number of input images in the input layer to create a good model; however, this limitation was supplied in our methodology by using a large number of images as input even in the density classification as the segmented regions.

However, as with all problems using a CNN, there are some negative points. The number of parameters that the network requires in the input layers, convolution, pooling, and in the fully connected layer are many. Thus, the study of these parameters sometimes requires a lot of time and experimentation.

By using only texture information and a larger number of images as input to the network, using a CNN to classify the breast tissue density and also using a CNN to detect mass, we demonstrate the feasibility of using convolutional neural networks on medical image processing techniques.

4.3. Review of results

The proposed methodology showed robustness in the task of detecting mass regions. All fundamental processes of a CAD system were included, going from the stage of pre-processing of the images to the classification of the mammary tissue and the detection of the regions of masses. Fig. 12 shows the fundamental results achieved in the test phase of the proposed method.

For the experiment and verification methodology, 1241 pairs of mammographic images covering dense and non-dense breasts were used. This compared to related work, and it shows a considerable number.

In the testing phase, the similarity indexes in the segmentation stage were able to segment asymmetric regions in the bilateral analysis, losing less than 4% of the masses in the 492 mammography images tested.

The two techniques proposed in 1RFP (peer-to-peer reduction and mean-by image reduction) also showed excellent results. At this step, only 11 masses were lost and reduced by more than 50% regions, resulting in an accuracy rate along with a segmentation of almost 93% of all masses of the test base.

In the density classification, an expressive result was obtained, with a mean of 97% of the cases. In the classification of regions, the best proposed models achieved a 91.0% accuracy in the non-dense breasts and 94.8% in the dense breasts. These numbers show how important the methodology is in covering the various types of breast tissue.

The specificity is higher in dense breasts owing to the fact that in the creation of the classification model, there are more non-mass regions (35,798 non-mass regions) compared to the number of regions in non-dense breasts (21,174 non-mass regions), making the network learn more characteristics of non-mass regions in the dense breasts.

Because the false positive numbers are smaller in dense mammograms, the accuracy becomes greater in this type of breast. However, considering a method that loses as few masses as possi-

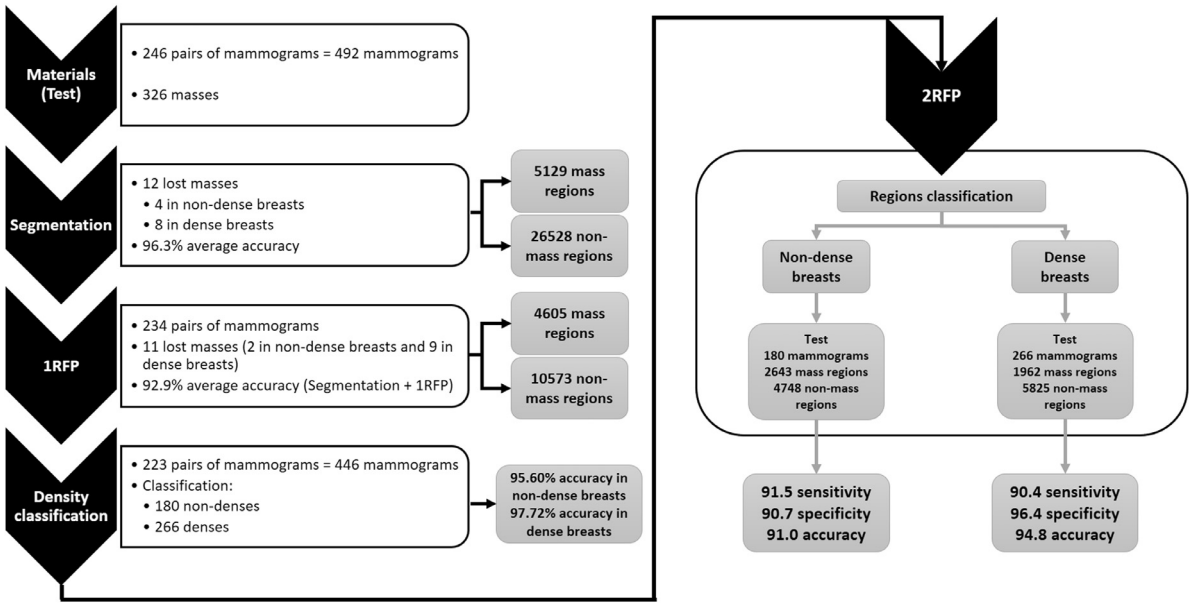

Fig. 12. Review of test results.

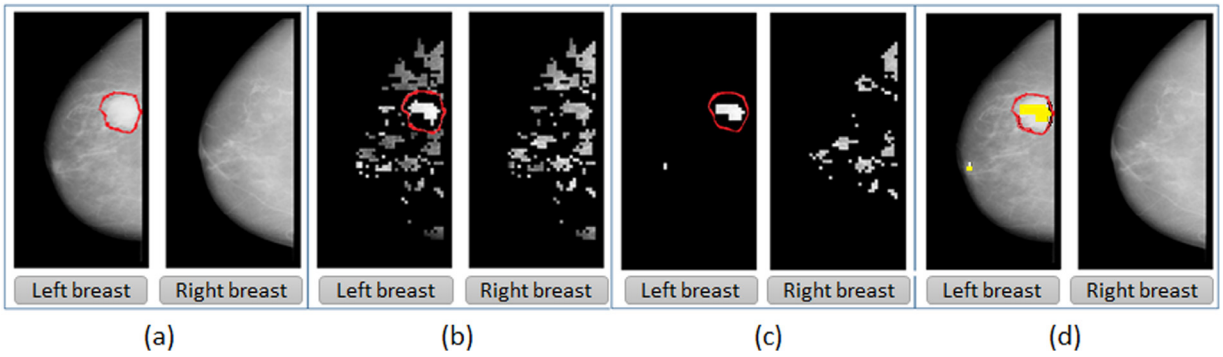

Fig. 13. Patient A_1006_1: (a) preprocessing and registration steps, (b) segmentation step, (c) 1RFP step, and (d) 2RFP step (mass detection).

ble, the non-dense model stands out, since the highest sensitivity is reached in this model.

4.4. Case studies

In order to demonstrate how the methodology was applied in the images of the DDSM, in this section will be shown specific cases, exemplifying the tests performed during the elaboration of the methodology.

4.4.1. First test case: success in non-dense breasts

The first example to be presented is the pairs of images of patient A_1006_1. In this case, at the end of the application of the entire methodology, it was possible to detect the mass regions according to the marking of the specialist.

Fig. 13a presents the results of the application of the sequence of the pre-processing and registration steps in this patient. The satisfactory results of these two steps can be observed, being able to align the patients breasts. Fig. 13b shows the segmentation step, where similarity indexes were able to delimit asymmetric regions. The red marking corresponds to the marking of the expert.

Mammograms bypass the 1RFP step (Fig. 13c). Then, the breasts are classified as dense and non-dense. In this example, the breasts were classified correctly as non-dense. Finally, their regions were classified as regions of mass and non-mass. Fig. 13d presents the final result of the regions of detected breasts. Yellow indicates regions considered as masses according to the model for classification of masses in non-dense breasts.

4.4.2. Second test case: error in non-dense breasts

In the second case, an error is shown in the segmentation step. However, before reaching the segmentation stage, the breasts go through the pre-processing and registration process, as shown in Fig. 14a. Then, the segmentation step fails, as shown in Fig. 14b.

It is observed that there was an error in the segmentation step. It is possible to emphasize the fact that the mass is very small. This can happen for two probable reasons: the mass is smaller or equal to the window 13×13, and when comparing with the other breast of using windows side by side. Because the mass is so small, part of it may have been in a 13×13 window, and another part in another window. Thus, as there was a loss of mass, this mammogram does not follow the other stages of false-positive reductions.

4.4.3. Third test case: success in dense breasts

The third example shows pairs of mammograms from patient A_1155_1 from DDSM. In this case, at the end of any methodology, it was also possible to detect the mass regions. An example of the pairs after the pre-processing and registration step can be seen in Fig. 15a. Fig. 15b illustrates the segmentation step of the asymmetric regions. The red marking corresponds to the marking of the specialist.

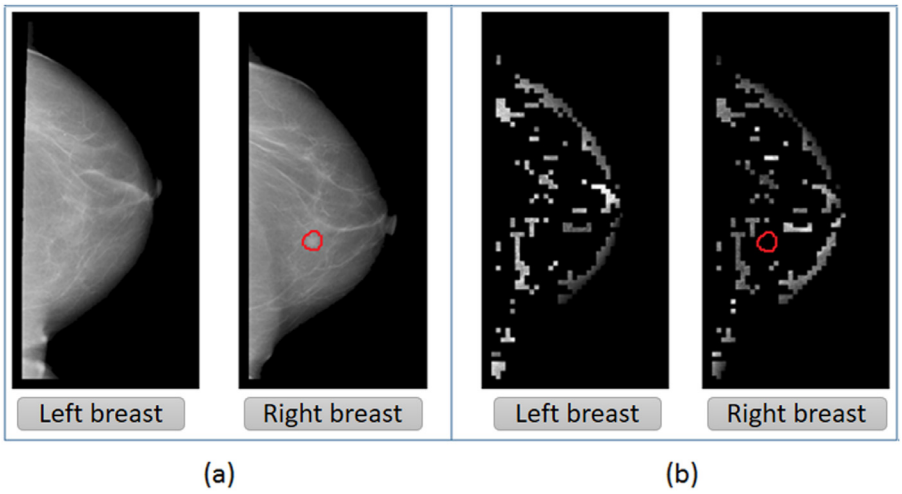

Fig. 14. Patient A_1017_1: (a) preprocessing and registration steps and (b) segmentation step.

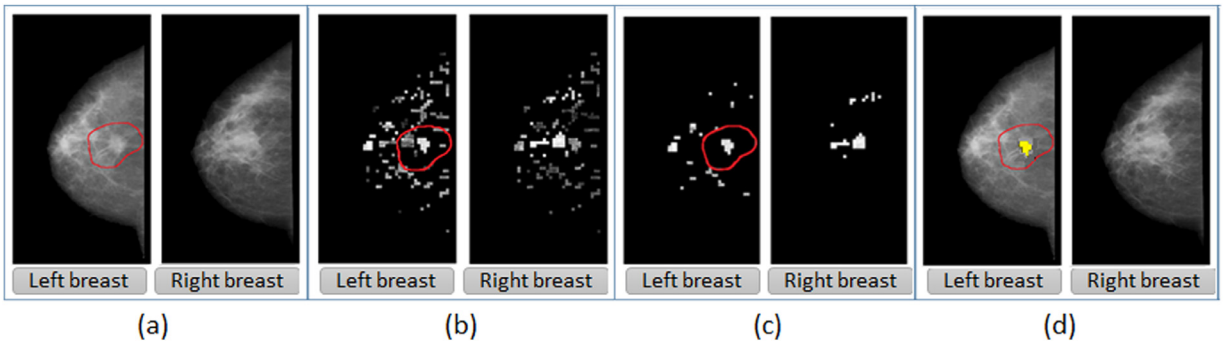

Fig. 15. Patient A_1155_1: (a) preprocessing and registration steps, (b) segmentation step, (c) 1RFP step, and (d) 2RFP step (mass detection).

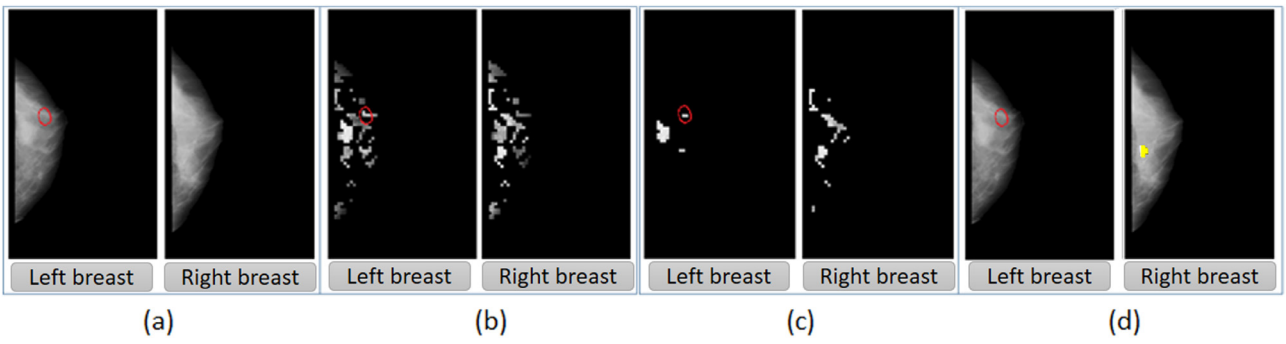

Fig. 16. Patient C_3003_1: (a) preprocessing and registration steps, (b) segmentation step, (c) 1RFP step, and (d) 2RFP step (mass detection).

In Fig. 15c, the results of the application of 1RFP are shown, and Fig. 15d shows the final result in the detection of regions of mass in patient A_1155_1. Yellow indicates regions considered as masses by the model.

4.4.4. Fourth test case: error in dense breasts

This last example presents a case of error in dense breasts. Patient C_3003_1 passed by the pre-processing step, the breasts were first isolated, and then one of them passed by the registration step in order to align one breast to another (Fig. 16a). Then the segmentation step was executed (Fig. 16b).

Mammograms then underwent the first reduction of false positives (Fig. 16c). Finally, the regions were classified into masses and non-masses (Fig. 16d where the yellow indicates regions classified into masses). It is noticed that in this case, even classifying some regions as regions of mass, the model was not able to detect the mass according to the marking of the expert. As was quoted in the first study case, since the mass problem is very small, an error may have occurred in the classification of the region as mass and non-mass.

4.5. Comparison of the methodology with related works

In this section we present the results of the proposed methodology compared to the related works described in Section 2. This comparison is expressed in Table 5, where you find the performance metrics.

Sm is the sample size, *Sen* is the Sensitivity, *Spec* is the Specificity, *Acc* is the Accuracy, *FPi* is the number of false positives per image, and *FROC* the area under the FROC.

It is observed that the proposed methodology presents expressive values in relation to the related works. Comparing the values

Table 5
Comparison of the methodology with the related works.

Work	Base	Sm	Sen.	Spec.	Acc.	FPi	FROC
Mass detection using only a mammogram							
Bajger et al. [16]	Private	36	–	–	–	–	0.90
	DDSM	48					0.96
Hu et al. [17]	MIAS	170	91.30	–	–	0.71	–
Sampaio et al. [4]	DDSM	623	80.00	–	–	0.84	–
Braz et al. [3]	MIAS	74	97.30	–	–	0.33	0.89
	DDSM	621	91.63	–	–	0.01	0.86
Sampaio et al. [18]	DDSM Non-dense	1049	94.02	82.28	84.08	0.85	1.13
	DDSM Dense	678	89.13	88.61	88.69	0.71	1.47
Mass detection by bilateral analysis							
Wu et al. [8]	Private	682	–	–	–	–	0.35
Tzikoupoulos et al. [15]	MIAS	–	84.47	–	–	–	–
Wang et al. [19]	Private	600	84.00	–	–	–	–
Ereceira et al. [7]	DDSM	620	100.00	95.34	96.00	–	–
Sun et al. [20]	Private	360	–	–	–	–	0.76
Casti et al. [21]	MIAS	94	–	–	–	–	94.00
	DDSM	188	–	–		–	–
Kelder et al. [22]	MIAS	322	–	–	–	–	0.87
Detection of masses using convolutional neural network							
Arevalo et al. [23]	BCDR	–	–	–	–	–	0.86
Suzuki et al. [11]	Private	198	89.90	–	–	–	–
Proposed method	DDSM Non-dense	1004	91.56	90.73	91.04	0.06	–
	DDSM Dense	1482	90.36	96.35	94.84	0.03	–

of false positives per image, only Braz et al. (2014) [3] presents smaller values, with a 0.01 *FPi*. Considering only the accuracy of the model for dense breasts with a result of 94.84%, only the methodology proposed by Ereceira et al. (2013) [7] gives better results with a 96% accuracy.

However, it is worth remembering that it is not possible to make a faithful comparison because the works do not use the same images of the proposed methodology. However, it is observed that this methodology used more images among the cited works (2,486 mammograms in total). This shows how widespread the method is.

It is also highlighted that the proposed methodology classifies the type of breast density and detects the masses. The only work presented that has this characteristic is that of Sampaio et al. [18], which uses evolutionary algorithms to adapt the model parameters to breast density. The proposed methodology was able to achieve better performance in the density classification than the cited work.

Thus, it is concluded that the proposed methodology occupies a prominent place in the literature, being able to reach its goal in the task of mass detection.

5. Conclusion

This work presented a methodology for the detection of masses in pairs of mammography images adapted to the density of breast tissue using image processing techniques and similarity index techniques from biology and convolutional neural networks.

The DDSM database was used to detect suspicious regions (asymmetric regions between pairs). The use of similarity indexes of biology was proposed for the task of detecting these regions in the database mammography pairs. In general, the use of these indices proved to be very promising since there was less than a 4% loss of the masses. However, there is a need to use false-positive reduction techniques in the task of discarding the excess of regions that are non-masses. In this task, two false-positive reduction techniques were proposed (peer-to-peer reduction and mean reduction by image), which showed a high efficiency.

Also proposed is the use of convolutional neural networks for both the classification task of the density of breast tissue, and for the final detection of the masses. In density classification, the use of a CNN was quite effective. Even with a simple architecture, it was possible to hit about 97% of breast density. It is noteworthy that there are few studies in the literature that classify breast density prior to mass detection, which demonstrates the promise of the proposed method.

The detection of regions of masses and non-masses based on density previously classified also presents significant results compared to those of related works. It is noteworthy that with the use of convolutional neural networks, an explicit stage of feature extraction is abstracted, which is a fundamental step in image processing.

The CNNs used in the methodology have the same architecture, both in density classification and in the detection of masses in dense and non-dense breasts. It is emphasized that this architecture and its parameters were used because in the detection of masses, a 32×32 input image was used, and many layers of convolution and pooling cause few pixels to reach the fully connected layer. This makes it difficult for the network to distinguish the classes between mass and non-mass. However, as future work, an analysis is proposed to improve the architecture of these CNNs in order to improve the results in the steps in which they are used.

We also highlight the large number of empirical parameters used in CNNs, which may become a negative factor of the methodology. Therefore, it is interesting to study and develop new techniques (such as the use of evolutionary algorithms) in order to find these parameters automatically.

Even so, the methodology presented was very promising. It is possible to affirm that the proposed methodology can assist the expert in the task of mass detection, serving as an aid mechanism in the task of detecting masses and asymmetric regions.

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012, Int. J. Cancer 136 (5) (2015) E359–E386.

[2] M.L. Giger, Computer-aided diagnosis of breast lesions in medical images, Comput. Sci. Eng. 2 (5) (2000) 39–45.

[3] G.B. Junior, S.V. da Rocha, M. Gattass, A.C. Silva, A.C. de Paiva, A mass classification using spatial diversity approaches in mammography images for false positive reduction, Expert Syst. Appl. 40 (18) (2013) 7534–7543.

[4] W.B. Sampaio, E.M. Diniz, A.C. Silva, A.C. De Paiva, M. Gattass, Detection of masses in mammogram images using cnn, geostatistic functions and svm, Comput. Biol. Med. 41 (8) (2011) 653–664.

[5] A.J. Méndez, P.G. Tahoces, M.J. Lado, M. Souto, J. Correa, J.J. Vidal, Automatic detection of breast border and nipple in digital mammograms, Comput. Methods Programs Biomed. 49 (3) (1996) 253–262.

[6] D. Scutt, G.A. Lancaster, J.T. Manning, Breast asymmetry and predisposition to breast cancer, Breast Cancer Res. 8 (2) (2006) R14.

[7] D.R. Ericeira, A.C. Silva, A.C. De Paiva, M. Gattass, Detection of masses based on asymmetric regions of digital bilateral mammograms using spatial description with variogram and cross-variogram functions, Comput. Biol. Med. 43 (8) (2013) 987–999.

[8] Y.-T. Wu, J. Wei, L.M. Hadjiiski, B. Sahiner, C. Zhou, J. Ge, J. Shi, Y. Zhang, H.-P. Chan, Bilateral analysis based false positive reduction for computer-aided mass detection, Med Phys 34 (8) (2007) 3334–3344.

[9] A.P. Nunes, A.C. Silva, A.C. de Paiva, Detection of masses in mammographic images Using Simpsons diversity index in circular regions and SVM, in: Machine Learning and Data Mining in Pattern Recognition, Springer, 2009, pp. 540–553.

[10] O.P.S. Neto, O. Carvalho, W. Sampaio, A. Correa, A. Paiva, Automatic segmentation of masses in digital mammograms using particle swarm optimization and graph clustering, in: Proceedings of the 2015 International Conference on Systems, Signals and Image Processing (IWSSIP), IEEE, 2015, pp. 109–112.

[11] S. Suzuki, X. Zhang, N. Homma, K. Ichiji, N. Sugita, Y. Kawasumi, T. Ishibashi, M. Yoshizawa, Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis, in: Proceedings of the 2016 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), IEEE, 2016, pp. 1382–1386.

[12] D. Blokh, E. Afrimzon, I. Stambler, E. Korech, Y. Shafran, N. Zurgil, M. Deutsch, Breast cancer detection by michaelis–menten constants via linear programming, Comput. Methods Programs Biomed. 85 (3) (2007) 210–213.

[13] M. Karnan, K. Thangavel, Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications, Comput. Methods Programs Biomed. 87 (1) (2007) 12–20.

[14] J.J. Fenton, S.H. Taplin, P.A. Carney, L. Abraham, E.A. Sickles, C. D'Orsi, E.A. Berns, G. Cutter, R.E. Hendrick, W.E. Barlow, et al., Influence of computer-aided detection on performance of screening mammography, N top N. Engl. J. Med. 356 (14) (2007) 1399–1409.

[15] S.D. Tzikopoulos, M.E. Mavroforakis, H.V. Georgiou, N. Dimitropoulos, S. Theodoridis, A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry, Comput. Methods Programs Biomed. 102 (1) (2011) 47–63.

[16] M. Bajger, F. Ma, S. Williams, M. Bottema, Mammographic mass detection with statistical region merging, in: Proceedings of the 2010 International Conference on Digital Image Computing: Techniques and Applications (DICTA), IEEE, 2010, pp. 27–32.

[17] K. Hu, X. Gao, F. Li, Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms, IEEE Trans. Instrum. Meas. 60 (2) (2011) 462–472.

[18] W.B. de Sampaio, A.C. Silva, A.C. de Paiva, M. Gattass, Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, lbp and svm, Expert Syst. Appl. 42 (22) (2015) 8911–8928.

[19] X. Wang, L. Li, W. Xu, W. Liu, D. Lederman, B. Zheng, Improving performance of computer-aided detection of masses by incorporating bilateral mammographic density asymmetry: an assessment, Acad. Radiol. 19 (3) (2012) 303–310.

[20] W. Sun, T.-L.B. Tseng, B. Zheng, J. Zhang, W. Qian, A new breast cancer risk analysis approach using features extracted from multiple sub-regions on bilateral mammograms, in: SPIE Medical Imaging, International Society for Optics and Photonics, 2015, pp. 941422–941428.

[21] P. Casti, A. Mencattini, M. Salmeri, R.M. Rangayyan, Analysis of structural similarity in mammograms for detection of bilateral asymmetry, IEEE Trans. Med. Imaging 34 (2) (2015) 662–671.

[22] A. Kelder, Y. Zigel, D. Lederman, B. Zheng, A new computer-aided detection scheme based on assessment of local bilateral mammographic feature asymmetry-a preliminary evaluation, in: Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, pp. 6394–6397.

[23] J. Arevalo, F.A. González, R. Ramos-Pollán, J.L. Oliveira, M.A.G. Lopez, Convolutional neural networks for mammography mass lesion classification, in: Proceedings of the37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, pp. 797–800.

[24] N. Dhungel, G. Carneiro, A.P. Bradley, Automated mass detection in mammograms using cascaded deep learning and random forests, in: Proceedings of the 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA), IEEE, 2015, pp. 1–8.

[25] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C.I. Sánchez, R. Mann, A. den Heeten, N. Karssemeijer, Large scale deep learning for computer aided detection of mammographic lesions, Med. Image Anal. 35 (2017) 303–312.

[26] M. Heath, K. Bowyer, D. Kopans, R. Moore, W.P. Kegelmeyer, The digital database for screening mammography, in: Proceedings of the 5th International Workshop on Digital Mammography, Citeseer, 2000, pp. 212–218.

[27] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[28] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[29] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing systems, 2012, pp. 1097–1105.

[30] J.M.R. Tavares, Analysis of biomedical images based on automated methods of image registration, in: Proceedings of the International Symposium on Visual Computing, Springer, 2014, pp. 21–30.

[31] F.P. Oliveira, J.M.R. Tavares, Medical image registration: a review, Comput. Methods Biomech. Biomed. Engin. 17 (2) (2014) 73–93.

[32] K. Deguchi, Registration techniques for partially covered image sequence, in: Proceedings of the 8th International Confernce on Pattern Recognition, 1986, pp. 1186–1189.

[33] J.A. Shackleford, N. Kandasamy, G. Sharp, On developing b-spline registration algorithms for multi-core processors, Phys. Med. Biol. 55 (21) (2010) 6329.

[34] A.A. Goshtasby, 2-D And 3-D image registration: For medical, remote sensing, and industrial applications, John Wiley & Sons, 2005.

[35] L.G. Brown, A survey of image registration techniques, ACM Comput. Surv. 24 (4) (1992) 325–376.

[36] B. Zitova, J. Flusser, Image registration methods: a survey, Image Vis. Comput. 21 (11) (2003) 977–1000.

[37] Z. Ma, J.M.R. Tavares, R.N. Jorge, T. Mascarenhas, A review of algorithms for medical image segmentation and their applications to the female pelvic cavity, Comput. Methods Biomech. Biomed. Eng. 13 (2) (2010) 235–246.

[38] Z. Ma, J.M.R. Tavares, et al., A review of the quantification and classification of pigmented skin lesions: from dedicated to hand-held devices, J. Med. Syst. 39 (11) (2015) 177.

[39] D.S. Jodas, A.S. Pereira, J.M.R. Tavares, A review of computational methods applied for identification and quantification of atherosclerotic plaques in images, Expert Syst. Appl. 46 (2016) 1–14.

[40] Z. Ma, J.M.R. da Silva Tavares, R.M.N.J. Jorge, A review on the current segmentation algorithms for medical images, IMAGAPP 2009 (2009).

[41] A.d.S. Meyer, A.A.F. Garcia, A.P.d. Souza, C.L.d. Souza Jr, Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (zea mays l), Genet. Mol. Biol. 27 (1) (2004) 83–91.

[42] P.D. Ferguson, T. Arslan, A.T. Erdogan, A. Parmley, Evaluation of contrast limited adaptive histogram equalization (clahe) enhancement on a fpga, in: Proceedings of the 2008 IEEE International SOC Conference, IEEE, 2008, pp. 119–122.

[43] C. Rafael Gonzalez, R. Woods, Digital Image Processing, Pearson Education, 2002.

[44] M.J. Prakash, V.V. Kumar, A.V. Babu, Morphology based technique for texture enhancement and segmentation, Signal Image Process. 4 (1) (2013) 49.