CrossMark

Automatic mass detection in mammography images using particle swarm optimization and functional diversity indexes

Otilio Paulo S. Neto[1,2] (iD) · **Aristófanes C. Silva[2]** ·
Anselmo C. Paiva[2] · **Marcelo Gattass[3]**

Abstract This paper proposes a computational method to assist in detection of masses in dense and non-dense breasts on mammography images. The proposed methodology is divided into six steps. In summary, the first step consist of the images acquisition that was obtained from the Digital Database for Screening Mammography (DDSM). In the second step, a preprocessing is performed in order to remove noises and enhance the images. In the third step, the segmentation is performed to find the regions of interest (ROIs) that are candidates for masses using Particle Swarm Optimization (PSO). The fourth step consists in the first false positives reduction based on reduction by distance and Graph Clustering. The fifth step is the second false positive reduction based on texture features using functional diversity indexes. Finally, in the sixth step, the support vector machine (SVM) is used to classify ROIs in whether mass or non-mass. The best results were found in case of dense breast tissue, resulting in a sensitivity of 97.52%, specificity of 92.28%, accuracy of 94.82%, false positives rate per image of 0.38 and free-curve receiver operating characteristic of 0.98.

Keywords Breast cancer · Particle swarm optimization · Graph clustering · Functional diversity indexes

1 Introduction

With the large increase of breast cancer in the world, it is increasingly necessary an early diagnosis in search of cure [15]. This increase is due to several factors such as feeding, stress, cigarette, drinks, and so on [19]. In a survey conducted by the National Cancer

✉ Otilio Paulo S. Neto
  otilio.paulo@ifpi.edu.br

[1] Federal Institute of Piauí, Teresina, Brazil

[2] Federal University of Maranhão, São Luis, Brazil

[3] Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil

⌂ Springer

Institute of Brazil, breast cancer is the second most common in the world and the most common among women [15]. On the other hand, it is known that when breast cancer is diagnosed early, there is a great chance of cure [14].

The most common mechanism to diagnose breast cancer is a mammography, which is a radiological exam that generates an image in breast grayscale. The radiologist analyzes and identifies visually where to find the lesion [8]. With the use of mammography, a reduction was observed in the mortality rate of this disease [27].

However, the evaluation of mammography images is subjective, requiring great experience of the radiologist. In recent decades, computational techniques have been developed with the purpose of automatically detecting structures that may be associated with tumors in mammography, improving the early detection rate [8].

Automatic mass detection motivates the development of systems assisted by computer (*Computer-Aided Detection* - CAD), due to the focus of investigation that the issue refers to. It is estimated that the sensitivity, the CAD systems ability to detect true positives, varies on mean 88% to 92% [2]. Notwithstanding, higher rates can be found in the literature sensitivity as 99% [1, 7]. Systems that have sensitivity values close to the aforementioned mean, can be considered sufficient to support the radiologist.

In literature, there are several methods that perform breast segmentation, considering both the oblique lateral medium and craniocaudal (called ipsilateral) incidences to detect mass based in the differences between them [23, 30, 32]. A similar principle is applied to methodologies that use the bilateral vision [16, 29, 31, 33].

In this paper, we propose an automatic method for detecting masses in mammography using the algorithms of particle swarm optimization (PSO) [17], graph clustering [26] and the functional diversity of indexes (FD) [22]. The methodology highlights the following contributions to the computer science field: *i)* the creation of an automatic method for mass segmentation, combining PSO with the Otsu algorithm; *ii)* the use of graph clustering and functional diversity indexes in false positive reductions.

This paper includes 4 more sections, which are structured as follows: The Section 2 presents the related works. In Section 3, will be described all the development steps of this research, starting by the acquisition of digital Database image for Screening Mammography (DDSM), followed by the preprocessing of segmentation using the PSO, the reduction of false positives, the features extraction, and finally the classification using support vector machine. In Section 4, the results apropos of the discussions and cases of success and failure of the methodology are presented. Finally, the Section 5 presents the inferred conclusions regarding the proposed methodology.

2 Related works

The available literature brings works that deal with the problem of mass detection in digital mammography images, which is the purpose of this work. Some of the works are highlighted below.

In the work of Hu et al. [13] combining adaptive limia-authorizations globally and place for segmentation in multi-resolution. The work of Liu et al. [18] presents a system for automatic detection of masses in digital mammograms. This system combines two techniques, the multiple concentric layers and the narrow strip region based on active contours, targeting the regions suspected of containing lesions. For the extraction of the ROIs texture features, the complete local binary standard (CLBP) was used to be classified by support vector machine (SVM).

The work proposed by Sampaio et al. [24] uses cellular neural networks to segment mammography images and generate the ROIs. It combines shape features (eccentricity, circularity, circular density disproportion circular and density) and texture features (Ripley's K function, indexes of Moran and Geary) to describe the ROIs. The extracted features were classified as mass and non-mass using SVM.

The method of Dong et al. [5] for classification of ROIs, obtained from DDSM base, uses chain code to indicate the ROIs. Its internal structure is enhanced by rough set. The convolution vector fields are used to extract 32 features of the ROIs. These features are used in the training and classification step, where the performance of classifiers Random Forest, SVM, genetic SVM, PSO, PSO-SVM and decision trees are compared. The best performance of the method was using the classifier Random Forest.

The proposed work of Braz [3] refers to detection of masses of regions in digitized mammography, using a methodology that involves aspects of the need to find suspicious regions and describes them in a discriminatory way. This study aims to evaluate the extraction features with the diversity in approaches and geostatistical analysis in order to obtain a classification of suspicious regions using SVM as classifier. With the results found in this paper, we highlight the high sensitivity and low mean rate of false positives when using concave geometry to extract features.

The work of Sampaio et al. [25] presents a computational method to aid in the detection of masses based on the density of the breast. In the segmentation step, they used a micro Genetic Algorithm to create a texture proximity mask and select regions suspected of containing lesion. Next was carried out two-step reduction of false positives. The first uses the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and a proximity ranking extracted texture of the breast regions of interest. In the second the resulting regions have their textures and forms analyzed by the combination of phylogenetic trees and geometric descriptors, Local Binary Pattern and SVM. A Micro Genetic Algorithm was used to choose the suspicious regions that generate the best training models and maximize the classification masses and non-masses used in SVM.

The aforementioned works commonly use segmentation techniques and features extraction based on shape and texture analysis to detect mass in mammograms. Therefore, the related works do not use an optimization process during the segmentation step and the features extraction is based on whole ROIs. Thus, we propose a automatic mass detection methodology using the PSO to optimize the regions during the segmentation process and a new feature extractor using functional diversity index in the texture analysis of the sub-regions of candidates to mass and non-mass. In the following sections the proposed methodology of this work is presented

3 Proposed methodology

In this section the study of the steps in the proposed methodology of this study is presented in detail as shown in Fig. 1.

3.1 Acquisition images

The images database used in the tests was the Digital Database for Screening Mammography (DDSM) [12], which is a public database containing over 2,000 cases, provided free of charge on the Internet. Each database case has four breast images (projections Craniocaudal - CC and Medium Lateral Oblique - MLO), besides information about the examination
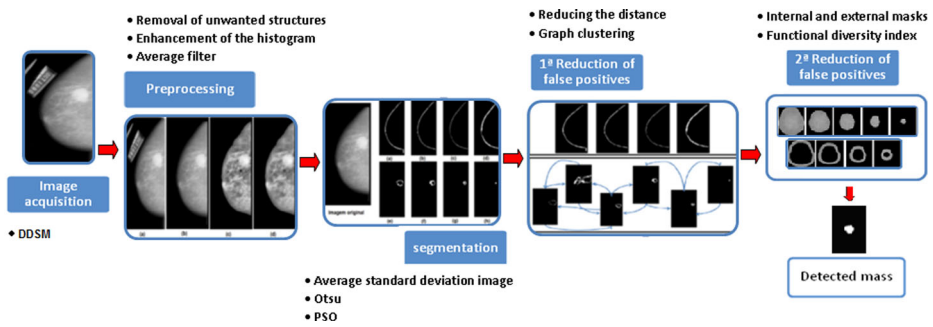
Fig. 1 Methodology steps

and the image data. All information contained in DDSM were provided by radiologists [12].

For the realization of this study, we used 621 images with the existence of at least one lesion as criterion of inclusion. These were the same used in the work of [3] and [24], so that more accurate comparison could be done.

3.2 Preprocessing

Prior to segmentation in mammograms it is important to highlight the existence of structures that are any unwanted segmentation methodology in digital mammography, such as noises, borders, markings and pectoral muscle, originated in the acquisition of the images.

To remove these structures, we used a methodology developed by Sampaio et al. [24]. After removal of these structures, a local enhancement technique based on the histogram is applied, the Contrast Limited Adaptive Histogram Equalization (CLAHE) and then applies the mean filter.

The CLAHE is a local enhancement contrast technique that changes the shade of gray from a pixel through the analysis of your neighborhood. The CLAHE avoids the increased contrast in noise based on an adaptive histogram equalization. In short, each pixel is changed based on the histogram neighborhood around where the transformation function is obtained through the function of cumulative distribution of pixel values in the neighborhood [9, 34].

The mean filter [10] was used in the proposed methodology to reduce existing noises in digital mammography, in order to enhance the internal structures of mammography [10], as shown in Fig. 2. Finalized the realization of the filters, the images served as the entrance to the breast segmentation, detailed in sections that follow.

3.3 Segmentation of mammography images

After the preprocessing step, the images were submitted to the segmentation process, following the steps described below.

3.3.1 Standard deviation mean image

The mean standard deviation of the image (DPMI) is the value of the mean standard deviation of all the windows of image, which provides the basis for comparison between the standard deviations of each generated cluster. In this process, the image is divided into
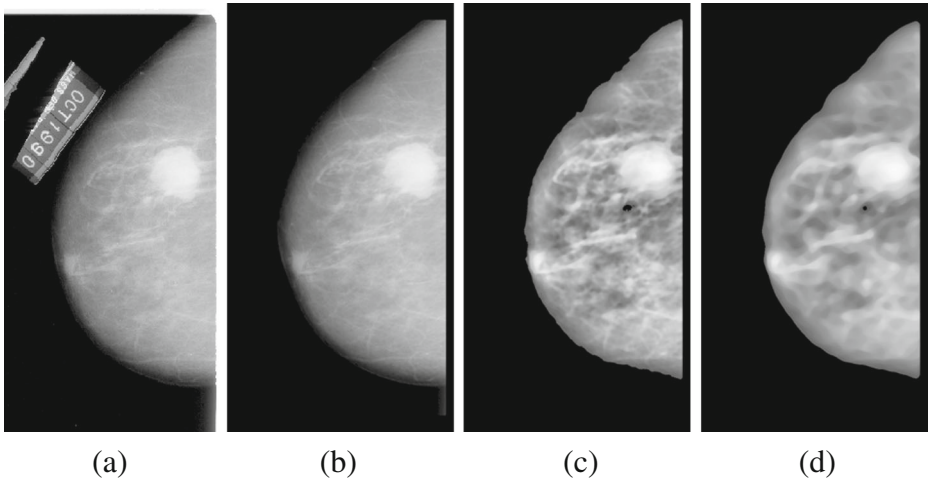
Fig. 2 A result from the steps of improvement: **a** Original image; **b** No image edges and markings; **c** Local Highlight (CLAHE); **d** Mean Filter

12×12 window (Fig. 3). This window size was chosen empirically for presenting the best results. Then the standard deviation is calculated for each window (δ_j), sum up all the patterns and deviations divided by the number of clusters, according to (1) and (2):

$$dpmi = \frac{1}{M} \sum_{j=1}^{M} \delta_j \tag{1}$$

$$\delta_j = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (x_i - \mu_j)^2} \tag{2}$$

where M is the number of windows found, δ_j is the standard deviation of each window, N represents the number of pixels x_i of the window and μ_j is the mean of window elements.

3.3.2 Otsu algorithm

The Otsu algorithm [21] is performed to find the first threshold image, dividing it into two clusters (Fig. 4a). Then, it is calculated with the (3), standard deviation of each cluster generated.

$$\delta^2 = \frac{1}{N} \sum_{j=1}^{N} (y_i - m_{ij})^2 \tag{3}$$

If $\delta^2 >$ DPMI, the centroid g of each cluster is calculated as the new threshold with the (4). The aforementioned cluster is divided into two from this new threshold (Fig. 4b). The standard deviation of each cluster is Once more again calculated. If ($\delta^2 >$ DPMI), then this previous step is performed recursively until δ^2 is not greater than DPMI.

$$g = \frac{1}{N} \sum_{j=1}^{N} I(w_i) \tag{4}$$

Fig. 3 Breast image divided into
12×12 windows
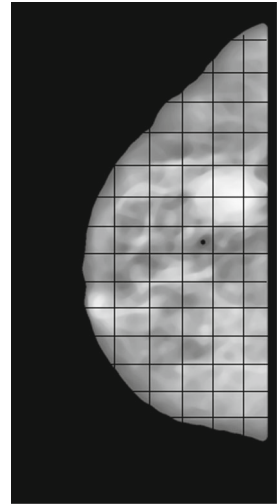
where, I is the intensity of the pixel w_i and N is the number of pixels in the group. The found thresholds will be used as centroids for the positions x_i vector initial particle of particle swarm (5). Figure 4 shows the assembled cluster from the vector x_i, thresholds generated by the Otsu algorithm.

$$x_i = (m_{i1}m_{i2}, ..., m_{ij}m_{iNc}), \qquad (5)$$

The Particle Swarm Optimization (PSO) receives the vector x_i and optimizes its values, seeking to find the best thresholds and consequently the best of the clusters. N_c is the number of clusters. The result of this process is explained in Fig. 5 and detailed in the next session.

3.3.3 Particle swarm optimization (PSO)

According to Merwe and Engelbrecht [20] PSO is an algorithm based on the social behavior of a flock of birds in motion. The authors say that given a problem, the PSO maintains a population of particles where each particle represents a potential solution to the problem and is associated with a position in a space of multidimensional search.
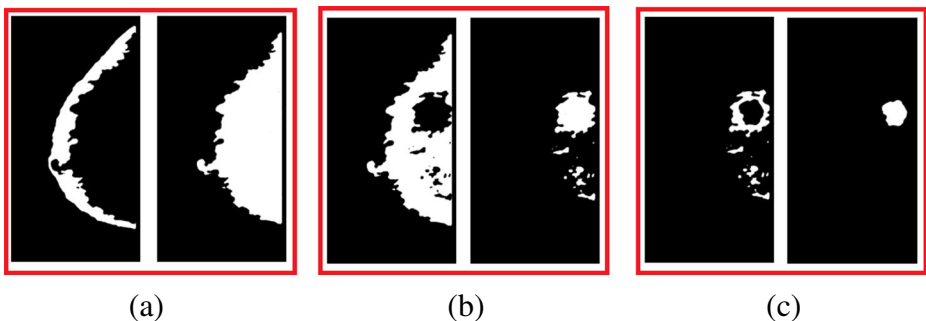
Fig. 4 Result of clusters generated by Otsu algorithm: **a** Generated first cluster; **b** Second cluster generated from above; **c** Third cluster of generated from the second
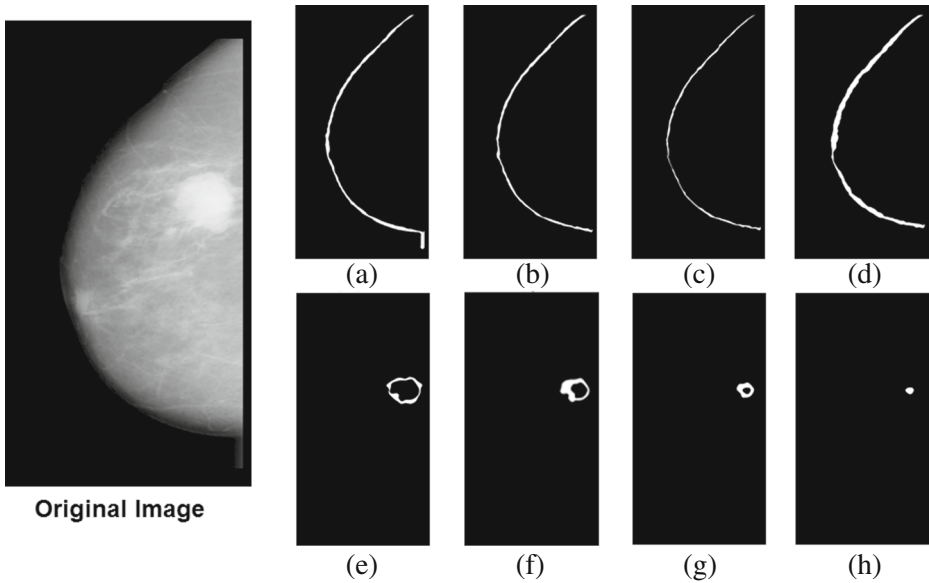
Fig. 5 Results of the segmentation process: original left image, the other pictures are generated clusters

In this work the PSO is used to optimize the values of the initial vector x_i of the first particle swarm generated through the Otsu algorithm (Section 3.3.2).

For each particle, it generates a random positions vector according to the quantity of particles, (6).

$$x_i = rand(0, 1).(ub - lb) + lb, \qquad (6)$$

where, ub and lb represent the minimum and maximum pixels values found in the image, respectively.

For each element of the vector x_i of each particle, it uses the velocity value given by (7).

$$v_i = wv_i + (c_1.rand(0, 1).(pbest - x_i)) + (c_2.rand(0, 1).(gbest - x_i)), \qquad (7)$$

where *pbest* represents the local best value x_i and *gbest* represents the global best value of all particles. Then, updates the value of x_i, as described in (8).

$$x_i = x_i + v_i \qquad (8)$$

After all elements of the position vector x_i are updated, new values of *pbest* and *gbest* are found. In the possession of the vector of positions x_i, containing all the centroids of the particle, it is calculated the smallest Euclidean distance of all the pixels of the image with respect to all the centroids of the vector of positions x_i. Each pixel will be grouped to the centroid that, given (9), returns the smallest value.

$$d_{min} = \sqrt{(z_p - x_i)^2} \qquad (9)$$

where, d_{min} is the smallest distance from the pixel z_p in relation to the vector centroids' position x_i.

Finally, it is calculated the aptitude value (fitness) of the particle (10) of the updated values of the positions vector. The smaller the found value, the better it is aptitude (fitness) of the particle

$$\delta^2 = (\sum_{j=1}^{N_c} [\frac{\sum_{\forall z_p \in C_{ij}} d_{min}}{|C_{ij}|}])/N_c \qquad (10)$$

where $|C_{ij}|$ is the amount of cluster elements and N_c is the amount of clusters found.

After these steps, the process is repeated until the number of iterations is reached, as stated in the values of the initial parameters. At the end of the iterations, the best vector positions x_i will represent the centroids set for generation of clusters, hence the creation of grouped images. Figure 5 illustrates the result.

3.4 Reduction of false positive

In this step, the region growth was initially applied in order to isolate the ROI. Then, two false positive reduction are executed. The first reduction is performed using an area filter, called in this work of reduction by distance and graph clustering [26]. In the second, we applied a texture descriptor using functional diversity indexes to classify the ROIs in whether mass or non-mass. These reductions are detailed in the following sections.

3.4.1 First reduction of false positives

In this step, we describe how the reductions were made. In search of the best results, we apply two techniques to reduce false positives, detailed below.

- *Reduction by distance*: it is the euclidean distance d between the first point (x_1, y_1) until the last point (x_2, y_2) of the image region. This process removes the regions that have the euclidean distance greater than 55%, discarding these ROIs. The percentage of 55% was chosen in empirically, which showed the best results. In Fig. 5 are shown some examples of these ROIs.
- *Reduction with Graph Clustering*(GC): it is the process of grouping the vertices of the graph in clusters leading for the structure of the edges of the graph. In this work, the GC is used to join the neighboring ROIs, building the graph from the union of these ROIs. For this, we have adopted some definitions: the neighborhood was defined in 3x3, the graph G will be build from the ROI of an image, verifying getting your neighborhood with all existing original cluster ROIs. Finally, the graph G will be directed.

After these definitions, we analyze the neighborhood of the current ROI, if any pixel of this ROI has in the minimum two neighboring pixels in another ROI, we can say that these ROIs are adjacent, and therefore, there must be an edge in G connecting the corresponding vertices.

Once the process above is ended, nodes of the graph G that have more than two links must be removed. All nodes that have nothing or at the most two links must remain, because nodes will be origin of candidates ROIs. Each nodes resulting in graph G of the process is a ROI (Fig. 6a).

From a node any graph G is calculated by its circular form factor value (FFC) (11), This way:

- If *FFC* is less than 10%: the node will be discarded and another node will be chosen, and the process is repeated.
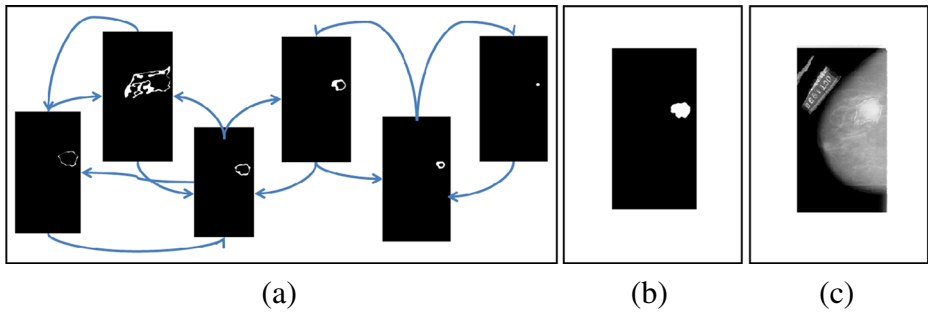
Fig. 6 Graph Clustering Result: **a** Graph generated by neighboring ROIs; **b** Final generated image from union of neighboring ROIs; **c** Original image with the marking radiologist

– If *FFC* is greater than 10%: its adjacencies are verified, making the joints of nodes. After each union, the *FFC* is recalculated. If *FFC* is greater than 10%, the union is valid and will be verified (if any) the next adjacency. If the *FFC*, after each union, results in a value less than 10%, this union will not be valid, and the node that was attached will be discarded.

$$FFC = \frac{4.\pi.A}{P^2}, \tag{11}$$

where *A* and *P* correspond to the perimeter and area, respectively, of each ROI. We adopted the percentage of 10% for the *FFC*, because of the tests, presenting the best results.

Finally, there were only ROIs that did not have any links left and the resulting ROI from the union of neighboring ROIs (Fig. 6b).

3.4.2 Second reduction of false positives

After the first reduction of false positives, the second reduction is applied in each region that will be individually characterized as mass or non-mass using the functional diversity indexes [22].

At first, it is necessary to prepare the sub-regions of each ROI for the functional diversity indexes to be extracted. Thus, the analysis of texture will be more appropriate because different regions of each ROI has been analyzed. In this study, we generated five internal masks and four external masks, as a form of representation from different regions of the same object, allowing to extract more features.

The generation of internal masks starts with a seed in the center of the ROI. This seed grows up to a certain limit as a percentage, preserving all the pixels of the original image, yielding the five internal masks. We can better visualize this process in Fig. 7.

The generation of external masks, the principle is the difference of two internal, consecutive same center. This is done to obtain details of other breast regions. Figure 8 shows an example of the external masks.

After the generation of internal and external masks, the features vectors are extracted using the functional diversity indexes. The study of phylogenetic diversity is a topic that has been increasing in recent years in various fields of ecology. Thus, suggesting that the concept is gaining great importance. Because of the potential relationship between the functional diversity and the operation and maintenance of the processes of community's [22], it is important to define precisely the concept of functional diversity.
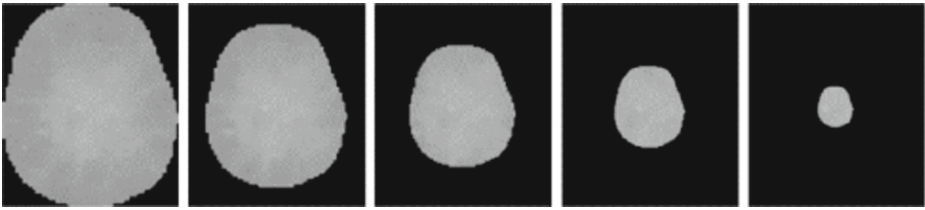
Fig. 7 Example of internal masks. From *left to right*: original ROI, ROI 80%, ROI 60%, ROI 40% e ROI 20%. Source: [25]

Tilman [28] defines functional diversity (FD) as the value and variation of species and their features that influence the functioning of communities. To further facilitate understanding, Table 1 shows the nomenclature regarding the biology and the methodology used. These definitions were adopted in this work, along with the creation of the dendrogram and functional diversity measures described below:

– Creation of Dendrogram from the pixel belonging to an ROI. In construction, we used the Otsu algorithm to separate groups of pixels with similarities in grayscale, forming communities. Each node represents separated groups the Otsu algorithm. To illustrate better, see the numbers 2 and 5 in Fig. 9, they belong to the same community, and represent the pixel values of ROI. The other numbers belonging to other nodes, form the other community.

– Features extraction using abundant functional diversity index (FADa), was removed from the ROI through the (12), taking into consideration the abundance of the amount of pixels from the same species in the dendrogram. S represents the number of dendrogram species.

$$FADa = \sum_{i=1}^{S-1} \sum_{j=i+j}^{S} d_{ij} a_i a_j, \tag{12}$$

To calculate the distance d_{ij} between the pairs of species (pixels) from ROI, we used the (13), where the value of X in that equation is the position of the species in the dendrogram. The abundance a for this case is the amount of each pixel with the same value as (14).

$$d_{ij} = \frac{1}{n} \sum_{k=1}^{n} (X_{ik} - X_{jk})^2, \tag{13}$$

Fig. 8 Example of external masks. From *left to right*: Difference of the original ROI 80%, ROI difference with 80% to 60%, the difference ROI 60% to 40%, ROI difference of 40% to 20%. Source: [25]

Table 1 Illustration of nomenclatures metholody in relation to biology

Biology	Proposal
Community	Sub-Region of interest (candidates) in mammography
Species	Pixel region of interest
Individual	Pixel value
Abundance: Number of individuals from a referred species	Number of pixels with the same value in a region of interest

$$\sum_{i=1}^{S} a_i = 1 \tag{14}$$

One can exemplifies how to calculate $FADa$, by analyzing the data from the Fig. 9. One can notice that there are 5 species of pixels (2, 5, 8, 9, 10), each one with its quantity, that is, $2 = 1, 5 = 2, 8 = 3, 9 = 1$ and $10 = 2$. Summing up the quantities of each species results the value 9, then the abundance a_i is the individual abundance of each specy and the sum of all the abundances of each species is equal to 1, that is:

$$\sum_{i=1}^{S} a_i = \frac{1}{9} + \frac{2}{9} + \frac{3}{9} + \frac{1}{9} + \frac{2}{9} = 1.$$

If the species 2 and 8 in the dendrogram and taken into consideration, one will perceive that the distance between them will be $d_{ij} = 2$, since this distance is calculated between species positions in the dendrogram, species 2 and 8 being the first and third positions respectively ($X_{ik} = 1$ and $X_{jk} = 3$). Thereby, considering the species chosen leads to:

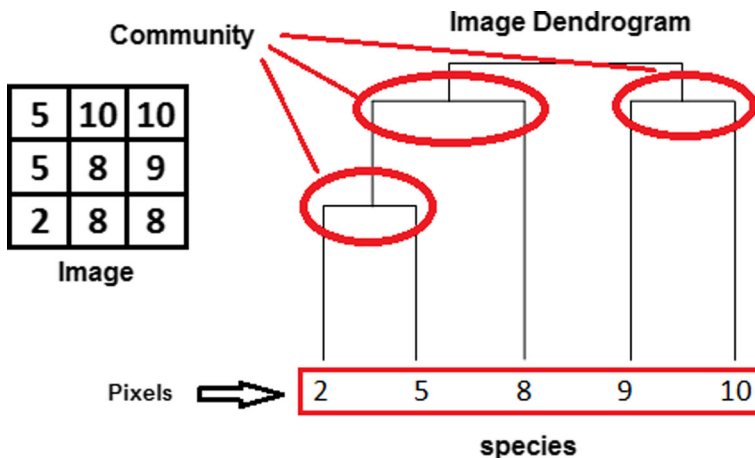$$FADa = 2 * \frac{1}{9} * \frac{3}{9} = 0.074074.$$

Fig. 9 Example of a dendrogram representing species (pixel) and groups (Community pixels)

– Features extraction using abundant functional diversity index of species (FADe), was removed from the ROI by (15), taking into consideration the abundance amount of pixel from the same species in the dendrogram.

$$FADe = \sum_{i=1}^{S-1} \sum_{j=i+j}^{S} d_{ij} e_i e_j \tag{15}$$

The distance from d_{ij} between the positions of the species (pixel value) from ROI is calculated by (16), considering Y the value of the species (pixel value) in the dendrogram. The abundance e, in this case is the amount of pixels with the same value (17).

$$d_{ij} = \frac{1}{n} \sum_{k=1}^{n} (Y_{ik} - Y_{jk})^2, \tag{16}$$

$$\sum_{i=1}^{S} e_i = 1 \tag{17}$$

To illustrate how to calculate $FADe$, one will analyze in the same way as in the previous example the data from Fig. 9. It can be noticed that there are 5 species of pixels (2, 5, 8, 9, 10), each with its quantity, that is, $2 = 1$, $5 = 2$, $8 = 3$, $9 = 1$ and $10 = 2$. The sum of the pixel quantities of each species is equal to 9, then e_i is the individual abundance of each species and the sum of all abundances of each species equals 1, that is:

$$\sum_{i=1}^{S} e_i = \frac{1}{9} + \frac{2}{9} + \frac{3}{9} + \frac{1}{9} + \frac{2}{9} = 1$$

Considering the same species 2 and 8 from the dendrogram, it can be perceived that the distance between them will be $d_{ij} = 18$, because in this case the distance will be calculated between the pixel values of each species ($Y_{ik} = 2$ and $Y_{jk} = 8$). Therefore, considering the species in question leads to:

$$FADe = 18 * \frac{1}{9} * \frac{3}{9} = 0.666666.$$

– Features extraction using abundant functional diversity index of pixel (FADp), is removed from the ROI using (18), taking into consideration the abundance as the total sum of the pixel values from the same species in the dendrogram.

$$FADp = \sum_{i=1}^{S-1} \sum_{j=i+j}^{S} d_{ij} p_i p_j, \tag{18}$$

The d_{ij} (19) is the distance between the positions of species (pixel value) from ROI, and Z the value species (pixel value) in the dendrogram. The abundance p, in this case is the sum total from the same value pixel values as (20).

$$d_{ij} = \frac{1}{n} \sum_{k=1}^{n} (Z_{ik} - Z_{jk})^2, \tag{19}$$

$$\sum_{i=1}^{S} p_i = 1 \tag{20}$$

To illustrate how to calculate $FADp$, one will analyze in the same way as of the the previous example the data from the Fig. 9. Note that there are also 5 species of pixels (2, 5, 8, 9, 10), each with its quantity, that is, $2 = 1, 5 = 2, 8 = 3, 9 = 1$ and $10 = 2$. In this case, the total abundance is the sum of the pixel quantities of the species, that is, $2 \Rightarrow 2*1 = 2, 5 \Rightarrow 5*2 = 10, 8 \Rightarrow 8*3 = 24, 9 \Rightarrow 9*1 = 9$ and $10 \Rightarrow 10*2 = 20$. Sum up results the value 65. p_i is the individual abundance of each species and the sum of all abundances of each species equals 1, that is:

$$\sum_{i=1}^{S} p_i = \frac{2}{65} + \frac{10}{65} + \frac{24}{65} + \frac{9}{65} + \frac{20}{65} = 1.$$

Considering the same species 2 and 8 from the dendrogram, one will obtain the distance between them $d_{ij} = 18$, because in this case the distance will be calculated between the pixel values of each species ($Z_{ik} = 2$ e $Z_{jk} = 8$). Therefore, taking into account these species, one obtains:

$$FADp = 18 * \frac{2}{65} * \frac{24}{65} = 0.204497.$$

In the training phase, each region is labeled by mass or non-mass, according to the marking tip. Pattern recognition was based on the texture using the SVM to classify regions. During the training and testing of segmented regions, it generates a feature vector with each label belonging to its class (mass or non-mass). To perform recognition, the first normalizes values of the variables for a better convergence of the SVM. Then, the following steps are performed:

1. The features of basis is balanced through the synthetic minority oversampling technique (SMOTE) [4].
2. It is based on the separation from training and testing randomly, with the proportions: 20%/80%; 60%/40%; 40%/60% e 80%/20%. these proportions can generate a more robust model.
3. The base is trained and tested five (5) times by the classifier;
4. Averaged execution of 5 (training/test) methodology for validation.

After the realization from the methodology classification step results the proposal is found and can be viewed in the following section.

4 Results and discussion

At this step, we used 388 non-dense mammography and 233 dense mammograms acquired from DDSM totaling 621 mammograms. These were selected on the criterion of having at least one mass lesion, according to the specification of the radiologist in the overlay file. Another criterion was the fact that the work of [24] and [3] used the same images, so you can make a better comparison of results.

For the execution of the PSO, some parameters must be initialized, and in this work, after several tests, the best values are shown in Table 2. To validate the results of the methodology are used the measures of sensitivity, specificity, accuracy, mean rate of false positives per image (FP/i) and free-response receiver operating characteristic curve (FROC) [3].

The test performed in this study were divided in 3 different ways, taking into consideration the density of the breasts. The first presents the results in non-dense breasts. The second in dense breasts and the third in dense breasts and non-dense together.

Table 2 Parameters used in the implementation of PSO

Parameter	Value	Description
k	10	Minimal amount of swarm
Dim	2*k	function from dimension of aptitude (*fitness*)
L	255	Number of gray levels
SwarmSize	3*Dim	Size of the swarm
lb	0	Aptitude lower limit of each particle
ub	L	Aptitude upper limit of Each particle
Iterations	k*Dim	Number of Iterations
Vmax	L	Maximum velocity of each particle
Vi	0	Initial velocity rate of particle
w	0.729	Inertia particle constant
c_1	2.05	Cognitive learning Constants
c_2	2.05	Social learning Constants

There were 3 tests performed using functional diversity indexes (Functional Diversity -FD), internal and external masks and SMOTE. Through these descriptors, the features extracted vectors of breasts were formed as follows: *i*) FD Index without masks; *ii*) FD Index with masks; *iii*) FD Index with masks + SMOTE. The results can be viewed in the following sections.

By way of comparison of results of the methodology, was used in all tests Haralick descriptors [11], classic approach to literature. These descriptors have a description of the texture based on statistical second order, from the calculation of the matrix of co-occurrence, consisting of counting how many different combinations of gray levels occur in an image in a given direction. To obtain such matrices, we consider the variation of the distance and direction to be followed between neighboring pixels. Generally they are used four directions: 0^0, 45^0, 90^0 e 135^0 More details can be seen in [6, 11].

4.1 Results from the non-dense breasts

For this result used 388 non-dense breasts, which corresponds to 430 masses, due to some breasts possessing more than one mass. The steps of the method and test results are shown in Fig. 10. In the following sections are all detailed steps.

4.1.1 Segmentation non-dense breasts

At this step we used the Otsu algorithm and the PSO. The first was used to find the initial thresholds, yielding the threshold vector. The second is used of this vector as the initial particle, optimizing for best value thresholds and consequently more homogeneous regions from breast.

After the segmentation process, one can see in Fig. 10 that the 388 mammograms were generated 61.556 candidates regions mass and non-mass. These regions served input the first reduction of false positives, described below.

4.1.2 First reduction of false positives in non-dense breasts

Prior to the first reduction of false positives (RFP) there were 61,556 ROIs acquired in the segmentation step. After the reduction process (Section 3.4), the amount ROIs of candidates
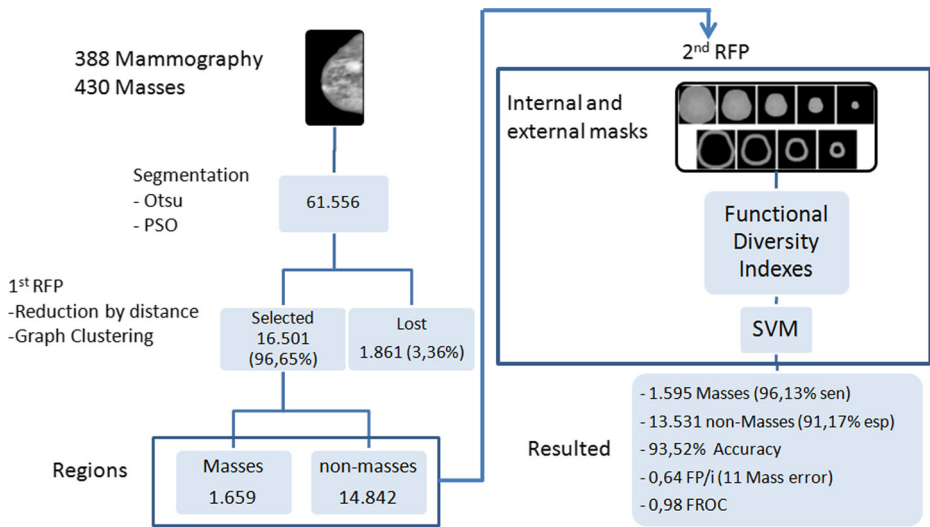
Fig. 10 Methodology results in non-dense breasts

was reduced to 16,501. Of these, 1,659 were considered mass and 14,842 non-masses. This was only possible by means of the reduction process by the distance and Graph Clustering, both detailed in Section 3.4.1.

Analyzing the data in Fig. 10, it is verified that there were 3,36% lost in the candidate regions. It is observed that such false positive reduction performed well with a 96,65% success rate.

4.1.3 Second reduction of false positives non-dense breasts

After the first reduction, the second reduction of false positives begins. For this, the internal and external masks were used, the texture analysis with levels functional diversity (FD), the SMOTE for balancing the base and the SVM to classify the ROIs candidates mass or non-mass.

All the tests were performed five times and at the end the arithmetic mean of each was calculated. Table 3 are shown the tests of the best result in the methodology. Looking at that table, we can infer that in all tests the sensitivity, specificity and accuracy behaved satisfactory in all cases. Besides, even in the worst case (20/80), the figures showed that the technique is promising. We can say also that the 80/20 test, showed significant values, showing that the methodology is effective in the detection of masses in mammography images.

Table 3 Result of the best classification's methodology test in non-dense breasts

Train/Test	Sensitivity	Specificity	Accuracy
20/80	84.46	79.67	81.93
40/60	91.48	85.50	88.33
60/40	94.73	89.10	91.76
80/20	96.13	91.17	93.52

Table 4 Results of experiments on classification of non-dense breasts

Tests	Sensitivity	Specificity	Accuracy
Haralick	60.00	70.58	70.12
FD index without masks	93.62	64.55	65.15
FD index with masks	67.86	73.46	73.22
FD index with masks + SMOTE	96.13	91.17	93.52

Table 4 shows the performance of tests for different experiments, using the features vectors extracted from the non-dense breasts from DDSM. The first experiment conducted used the Haralick texture descriptors [11], such as: contrast; second angular momentum; energy; homogeneity; entropy; correlation; dissimilarity; maximum entropy and inverse variance. It is noted that the results were not satisfactory.

Other experiments were carried out by making a combination of functional diversity indexes, the internal and external masks and SMOTE. It is noticed that the best result was a combination of indexes of functional diversity with masks along with the SMOTE. Getting the best performance mean with 96.13% sensitivity, 91.17% specificity and 93.52% accuracy.

Another analysis performed to evaluate the performance of the methodology was the FROC curve, which showed the value of 0.98 and 0.64 of FP/i. This result lead us to the conclusion that the methodology performed well.

4.2 Results with dense breasts

At this step we used 233 dense breasts, corresponding to 247 masses. The stages of this step can be seen in Fig. 11 and the details described in the following sections.
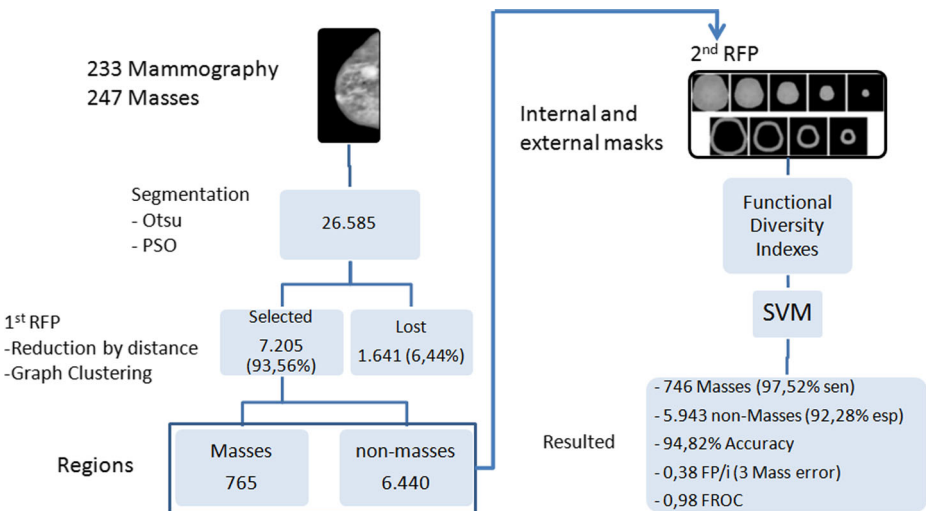

Fig. 11 Methodology results in dense breasts

Table 5 Result of the best methodology test classification in dense breasts

Train/Test	Sensitivity	Specificity	Accuracy
20/80	86.87	79.63	83.15
40/60	93.68	86.91	90.23
60/40	96.50	89.81	93.03
80/20	97.52	92.28	94.82

4.2.1 Segmentation breasts dense

In this point, the Otsu algorithm was applied to find the initial vector of thresholds and PSO for optimizing the values of this vector. Based on these thresholds, the breast regions were obtained.

During the process of segmentation, were generated 26,585 candidates regions the mass and non-mass of 233 mammography. In Fig. 11 one can see that these values are input for the first step of false positive reduction, described in the following section.

4.2.2 First reduction false positive breast dense

At this step, it was also used to reduction by the distance to remove unwanted clusters and Graph Clustering to unite neighboring ROIs. After 26,585 ROIs pre-candidates, acquired in targeting, passed the first reduction of false positives (RFP), there were only 7,205, representing a decrease of 72,90%. Resulting in 765 masses and 6,440 non-masses.

Figure 11 shows the results obtained with aforementioned reduction, where the loss of 6,44% during the process shows that the method is efficient and promising.

4.2.3 Second reduction of false positives dense breasts

In this second reduction of false positives, one used the internal and external masks, Sections 7 and 8, the texture analysis with the contents of FD was repeated, and finally, the SVM to classify the candidate ROIs mass or non-mass.

In dense breasts, tests were also conducted five (5) times and at the end the arithmetic mean of each of them is calculated. Using the features of vectors extracted from the breast, it is observed that the methodology presented a satisfactory performance, reaching 97.52% of sensitivity, 92.28% of specificity and 94.82% of accuracy. The result of the improved test methodology in dense breasts can be seen in Table 5.

Different types of experiments were performed. Initially Haralick texture descriptors were used, where the results were not satisfactory. Then FD indexes combinations were

Table 6 Results of experiments performed in the classification of dense breasts

Tests	Sensitivity	Specificity	Accuracy
Haralick	67.86	72.22	72.04
FD index without masks	89.47	72.90	73.32
FD index with masks	96.36	71.46	72.06
FD index with masks + SMOTE	97.52	92.28	94.82

performed, internal and external masks, and SMOTE. The best results found by the methodology was the combination from texture analysis using the levels of FD with the masks and SMOTE. In Table 6 results of experiments for dense breast are shown.

Finally, we analyzed the area under the curve FROC to evaluate the performance of the methodology, obtaining the value of 0.98 and 0.38 of FP/i. This result led us to the conclusion that the methodology performed satisfactory.

4.3 Results with the dense breasts and non-dense breasts

A test was performed with all the dense and non-dense breasts, making a total of 621 breasts, corresponding to 677 mass. Figure 12 shows the steps that were performed in this step. In the following sections the whole process is detailed.

4.3.1 Segmentation in dense and non-dense breast

This segmentation, the Otsu algorithm was also used to find the initial thresholds and the PSO to optimize these thresholds. It was possible to find more homogeneous regions.

After the segmentation process, from the 621 mammography that were generated 61,556 candidates regions were mass and non-mass. These regions served input the reduction of false positives, as shown in Fig. 12.

4.3.2 First reduction of false positives in dense and non-dense breasts

In this reduction, we used 88,141 ROIs pre-candidates arising from the segmentation. After the process of the first false positives reduction (RFP), there were only 23,706, corresponding to a decrease of 73.10%. Remaining at the end 2,424 masses and 21,282 non-masses.

In Fig. 12 one can see that after this reduction process, only 4,18% mass were lost, which shows the efficiency of the method.
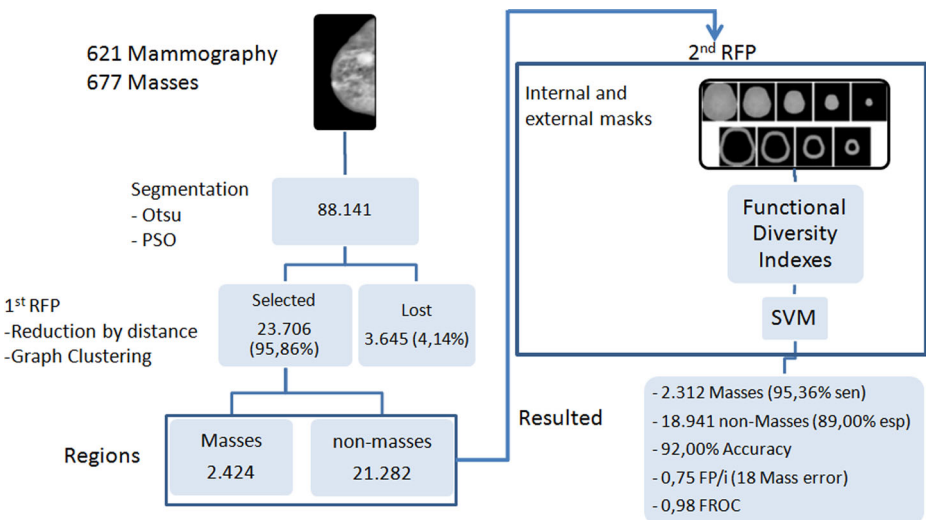

Fig. 12 Methodology results in dense and non-dense breasts

Table 7 Result of the best classification's methodology test in dense and non-dense breasts

Train/Test	Sensitivity	Specificity	Accuracy
20/80	83.39	78.18	80.66
40/60	90.61	83.75	87.03
60/40	93.36	87.08	90.08
80/20	95.36	89.00	92.00

4.3.3 Second reduction of false positives dense and non-dense breasts

Once the previous reduction step is completed, a second step of false positive reduction is performed. In this step, the same techniques presented in the previous sections for texture analysis and submitted to the SVM were used to classify the candidate ROIs in mass and non-mass.

The tests complied the same criteria employed in dense breasts and non-dense. They were performed five (5) times and at the end, the arithmetic mean is calculated for each of them. In Table 7 the best results found in the methodology are shown. Analyzing the data we can see that in the tests, the sensitivity metrics, specificity and accuracy of the test 80/20, presented the best values and right next to the previous tests (dense and non-dense breasts).

In Table 8, we present the results of experiments conducted in this work. Therefore, verified once again, that the proposed method had satisfactory values, proving to be efficient.

To finalize the methodology of the tests with all the dense and non dense breasts, we used the same metrics of evaluation from previous tests, and the obtained values were 95.36% of sensitivity, 89.00% of specificity and 92.00 of accuracy.

Completing the analysis using the FROC curve to evaluate the performance of the methodology, the obtained value was 0.98. Another analysis was the number of false positives per image with the value of 0.75, showing that the method is promising.

4.4 Study of cases

For a better understanding of the methodology, in this section will be presented some specific cases in order to exemplify the tests performed during the research.

4.4.1 First case: successful mass detection in non-dense breast

The first successful case shown is the image A_1006_1.LEFT_CC, which from the beginning of the process to the end, showed good results in every step of the methodology. Figure 13

Table 8 Results of experiments performed in the classification of dense and non-dense breasts

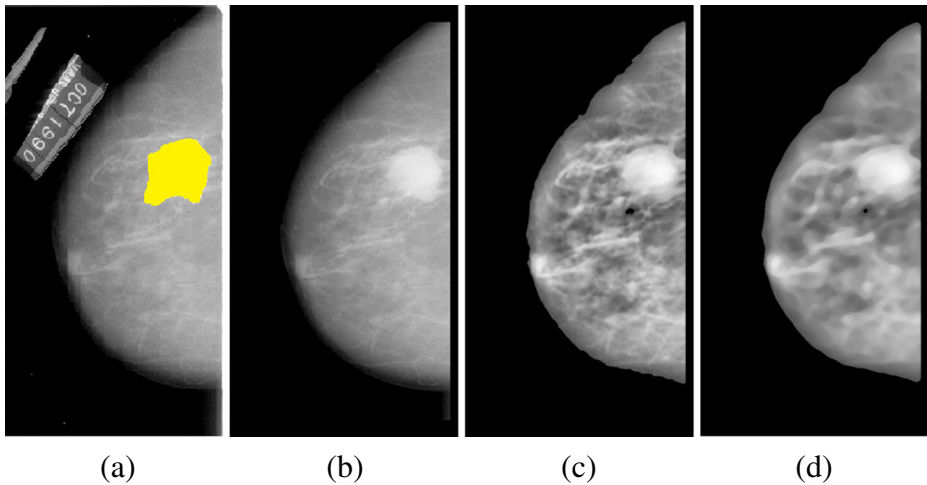| Tests	Sensitivity	Specificity	Accuracy
Haralick	60.00	68.89	70.04
Index of FD without Masks	90.41	63.91	64.32
Index of FD with Masks	66.36	70.36	70.16
Index of FD with Masks + SMOTE	95.36	89.00	92.00

Fig. 13 Image preprocessing A_1006_1.LEFT_CC: **a** Original image with the marking radiologist in yellow; **b** No image edges and markings; **c** Enhancement CLAHE; **d** Mean Filter

shows the process that begins with the removal of unwanted structures from the original image (a) shortly after removal of these structures (b) is applied to local enhancement technique based on CLAHE histogram (c), and the filter mean (d).

By carrying out the preprocessing, the image of Fig. 13d is subjected to segmentation step to be extracted from candidate regions. At this step, 132 ROIs were generated, but with the reduction of false positives, the masses were found. Figure 14 shows the final result of the methodology.

4.4.2 Second case: failure detection of non-dense breast

In the second case, the detection mass in a non-dense breast presented an error during the segmentation process. Figure 15 shows two images, the first being the original image with
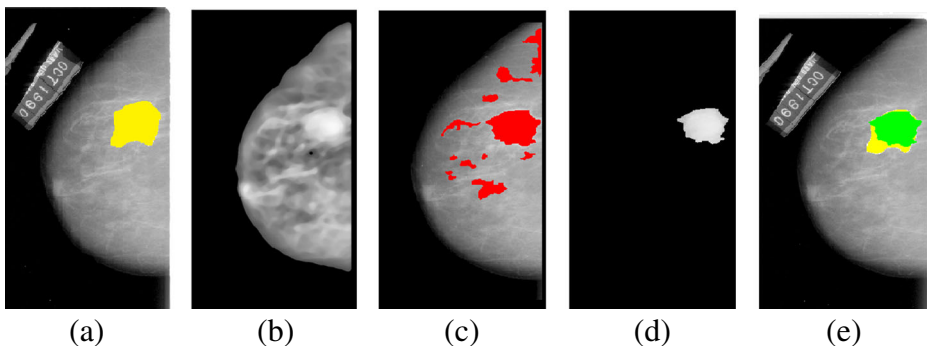

Fig. 14 Methodology Results applied to image A_1006_1.LEFT_CC: **a** Original image with the marking radiologist in yellow; **b** Pre-processed image; **c** Image with segmented regions in red; **d** Image targeted mass; **e** Marking the image of radiologist in yellow and green segmented mass
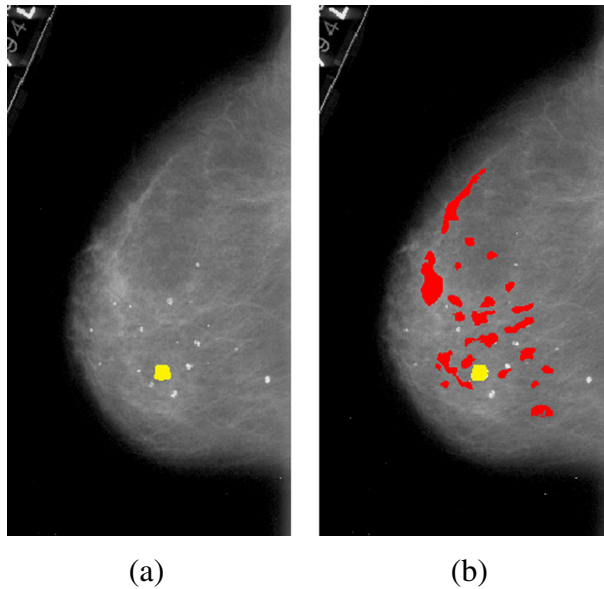
(a)   (b)

Fig. 15 Methodology Results applied to image A_1512_1.LEFT_CC: **a** Original image with the marking radiologist in yellow; **b** Image with the result of the methodology in red put on the original image with the marking radiologist in yellow

the localization of the lesion made by the radiologist (a), and second, the image with the regions targeted by the methodology, along with the marking.

Looking at the image of the Fig. 15, once can notice that the lesion is very small and difficult to detect. Therefore, for this case, the method failed to detect the mass, since the beginning of the segmentation.

4.4.3 Third case: success in detecting mass in dense breast

The third success story is the image A_1036_1.LEFT_CC, referring to a dense breast, which also acted in a satisfactory manner during the steps of the method. Figure 16 shows the original image with the marking tip (a), the ROIs candidates generated by segmentation in red (b), the mass detected by the method (c) and an image showing the original with the marking and mass in green color overlapping.

Looking at the region marked with the region of the detected mass, it can be said that the methodology found virtually the entire structure of the lesion in question; it is occupying almost all the highlighted region.

4.4.4 fourth case: failure detecting mass in dense breast

To illustrate a failure case of the methodology, presented in Fig. 17 the result image A_1512_1.LEFT_CC, the mass was lost from the step of segmentation (c) because no mass candidate region was found. Analyzing carefully, one can notice in the original image without marking to the (a), the region and difficult visibility.
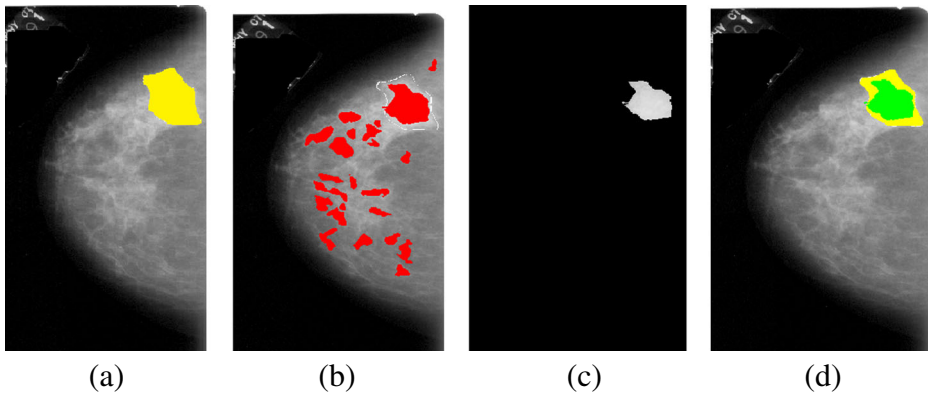
Fig. 16 Methodology Results applied to image A_1036_1.LEFT_C: **a** Original image with the marking radiologist in yellow; **b** Image with segmented regions in red; **c** Mass of the image detected by the method **d** Image with the result of the methodology green overlapping the original image with the yellow marking radiologist

Looking at the image with the marking in yellow (b), and looking at the same region in the original image (a) one can actually see that the region would be less likely to arrange as a lesion, but even so, the methodology found regions near the mass (c).

4.5 Methodology comparison with related work

In this section we make the comparison with the results of related works, described above, with the values obtained from the proposed methodology. Such a comparative is summarized in Table 9, which contains the metrics of performance measures through sensitivity
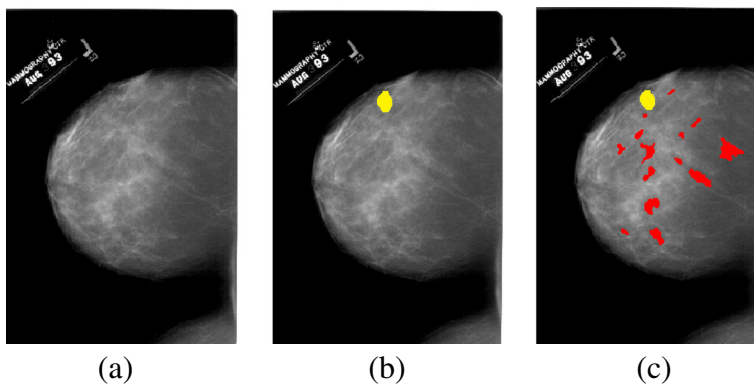

Fig. 17 Methodology Results applied to image A_1512_1.LEFT_CC: **a** Original image; **b** Original image with the marking radiologist in yellow (**c**) image with the result of the methodology in red and marking radiologist in yellow

Table 9 Comparison of the results obtained from the proposed method with the values of the related works

Works	Base	QTD	Sen.	Esp.	Acc	ROC	FP/i	FROC
IMAGECHECKER	Privative	—	88.00	—	—	—	—	—
HU et al.	MIAS	170	91.30	—	—	—	0.71	—
LIU et al.	DDSM	125	76.80	—	—	—	1.36	—
SAMPAIO et al.	DDSM	623	80.00	—	—	—	0.84	—
AL MUTAZ et al.	DDSM	120	91.67	84.17	—	—	—	—
AL MUTAZ et al.	DDSM	60	—	—	95.85	—	—	—
BRAZ	MIAS	74	97.30	—	—	—	0.33	0.89
	DDSM	621	91.63	—	—	—	0.01	0.86
DONG et al.	DDSM	200	94.78	91.76	93.24	0.95	—	—
MARTINS et al.	DDSM	600	98.60	98.85	98.88	—	—	—
SAMPAIO et al. (non-Dense)	DDSM	1049	94.02	82.28	84.08	—	0.85	1.13
SAMPAIO et al. (Dense)	DDSM	678	89.13	88.61	88.69	—	0.71	1.47
Methodology[1]	DDSM	388	96.13	91.17	93.52	—	0.64	0.98
Methodology[2]	DDSM	233	97.52	92.28	94.82	—	0.38	0.98
Methodology[3]	DDSM	621	95.36	89.00	92.00	—	0.75	0.98

[1]Non-Dense breast [2]Denses breast [3]Denses and non-denses breast

Performance comparison of metrics are measured by: Sensitivity (Sen.); Specificity (Esp.); Accuracy (Acu.); ROC (ROC) curve; area under the curve FROC (FROC); and false positives per image (FP/i)

(Sen.), specificity (Esp.), accuracy (Acu.), Area under the ROC curve, the media false positives per image (FP/i), area under the curve FROC and the sample size (QTD).

One can observe in Table 9 that the proposed methodology has significant values comparing with the related works. For a successful comparison, it is necessary that the images and the sample base are the same. Based on this precept, we compared the present method with the work of Braz [3] and Sampaio et al. [24], because we used the same base of images and the same amount of sample, and the sensitivity value was higher than of the related works, reaching a sensitivity rate of 97.52%.

5 Conclusion

In view of the arguments presented, it is clear that the proposed method achieved its goal of detecting masses automatically (with the precision 97.52%) digital mammography using particle swarm optimization (PSO) and functional diversity index.

In the segmentation step, performed with the Weight, which is one of the main techniques used in this methodology, we were capable of finding more homogeneous regions. On the other hand, it brought up the problem of generating many false positives. For this reason, two steps of false positive reduction were performed, highlighting in particular the Graph Clustering because it was by means of that we managed to unite neighboring ROIs and put away a significant part of the non-masses, without the loss undertook the research. It was

obtained in the best case a rate of 73.19% of reduction, and a 96,65% accuracy rate for non-dense breasts.

The best results of this study were obtained using dense breasts with functional diversity indexes along with masks (in-tender and external) and SMOTE. The values were 97.52% sensitivity, 92.28% of specificity, 94.82% accuracy, 0.38 false positives per image and 0.98 area under the curve FROC. Finally, the proposed methodology can assist in mass detection, providing the radiologist second opinion in the early detection of breast cancer.

References

1. Bose JSC, Karnan M, Sivakumar R (2010) Detection of masses in digital mammograms. Int J Comput Netw Secur 2(2):78
2. Boyle P, Levin B et al (2008) World cancer report 2008. IARC Press International Agency for Research on Cancer
3. Braz JG (2014) Detecção de regiões de massas em mamografias usando índices de diversidade, geoestatística e geometria côncava. Ph.D. thesis, Universidade Federal do maranhão, Programa de pós-graduação em Engenharia de Eletricidade são Luis - MA
4. Chawla NV, Bowyer K, Hall LO, Kegelmeyer WP (2002) Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research pp 321–357
5. 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 pp 1–13
6. Eleyan A, Demirel H (2011) Co-occurrence matrix and its statistical features as a new approach for face recognition. Turk J Electr Eng Comput Sci 19(1):97–107
7. Gao X, Wang Y, Li X, Tao D (2010) On combining morphological component analysis and concentric morphology model for mammographic mass detection. IEEE Trans Inf Technol Biomed 14(2):266–273
8. Giger ML (2000) Computer-aided diagnosis of breast lesions in medical images. Comput Sci Eng 2(5):39–45
9. Gonzales R, Woods R (2010) Processamento Digital de Imagens. 3a. ed. Pearson Prentice Hall, São Paulo
10. Gonzalez RC, Woods RE (2007) Digital image processing. 3a ed. Prentice Hall Upper Saddle River, NJ
11. Haralick RM, Shanmugam K et al (1973) Textural features for image classification. IEEE Transactions on systems, man, and cybernetics (6) pp 610–621
12. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer P (2000) The digital database for screening mammography. In: Proceedings of the 5th international workshop on digital mammography, pp. 212–218. Citeseer
13. Hu K, Gao X, Li F (2011) Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms. IEEE Trans Instrum Meas 60(2):462–472
14. INCA dSJAG (2015) Instituto Nacional de câncer josé Alencar Gomes da Silva. Disponível em: Acessado em: 05/11/2015. www2.inca.gov.br
15. INCA INdC (2014) Estimativas de incidência de câncer no brasil. rio de janeiro 2014
16. Ke L, Mu NKY (2010) Mass computer-aided diagnosis method in mammogram based on texture features. In: IEEE. 3Rd international conference on biomedical engineering and informatics (BMEI). yantai, China, 2010. pp. v. 1, p. 354–357
17. Kennedy J (2010) Particle swarm optimization. In: Encyclopedia of machine learning, pp 760–766. Springer
18. Liu X, Xu X, Liu J, Feng Z (2011) A new automatic method for mass detection inmammography with false positives reduction by supported vector machine. In: Biomedical engineering and informatics (BMEI), 2011 4th international conference on, vol. 1, pp 33–37. IEEE

19. Mcpherson K, Steel C, Dixon J (2000) Breast cancer—epidemiology, risk factors, and genetics. Bmj 321(7261):624–628

20. Merwe DVD, Engelbrecht AP (2003) Data clustering using particle swarm optimization. In: Evolutionary computation, 2003. CEC'03. The 2003 congress on, vol. 1, pp 215–220. IEEE

21. Otsu N (1975) A threshold selection method from gray-level histograms. Automatica 11(285-296):23–27

22. Petch EYOL, Gaston KJ (2006) Functional diversity: back to basics and looking forward. Ecol Lett 9(6):741–758

23. Qian W, Song D, Lei MRS, Eikman E (2007) Computeraided mass detection based on ipsilateral multiview mammograms. Acad Radiol 14(5):530–538

24. Sampaio WB, Diniz EM, Silva AC, DE Paiva AC, Gattass M (2011) Detection of masses in mammogram images using cnn, geostatistic functions and svm. Comput Biol Med 41(8):653–664

25. Sampaio WBd, Silva AC, DE Paiva AC, Gattass M (2015) Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, lbp and svm. Expert Syst Appl 42(22):8911–8928

26. Schaeffer SE (2007) Graph clustering. Comput Sci Rev 1(1):27–64. doi:10.1016/j.cosrev.2007.05.001

27. Society AAC (2013) Learn about breast cancer

28. Tilman D (2001) Functional diversity. Enc Biodiversity 3(1):109–120

29. Tzikopoulos S, Mavroforakis M, Georgiou H, Dimitropoulos N, Theodoridis S (2011) A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. Acad Radiol Elsevier 102(1):47–63

30. van Engeland S, Karssemeijer N (2007) Combining two mammographic projections in a computer aided mass detection method. Med Phys 34(3):898–905

31. Wang X, Li L, Xu W, Liu W, Lederman D, Zheng B (2012) Improving performance of computer-aided detection of masses by incorporating bilateral mammographic density asymmetry: an assessment. Comput Methods Prog Biomed Elsevier 19(3):303–310

32. Wei J, Chan H, Zhou C, Wu Y, Sahiner B, Hadjiiski L, Roubidoux MA, Helvie MA (2011) Computer-aided detection of breast masses: Four-view strategy for screening mammography. Med Phys 38:1867

33. Wu Y, Wei J, Hadjiiski L, Sahiner B, Zhou C, Ge J, Shi J, Zhang Y, Chan H (2007) Bilateral analysis based false positive reduction for computer-aided mass detection. Med Phys NIH Public Access 34(8):3334

34. Zuiderveld K (1994) Contrast limited adaptive histogram equalization. In: Graphics gems IV, pp 474–485. Academic Press Professional, Inc

Otilio Paulo S. Neto PhD in Electrical Engineering with emphasis on Image Processing and Computer Vision, master's at Electric Engineering from Universidade Federal do Maranho (2016) and graduation at Bacharel em Sistemas de Informao from Faculdade Integral Diferencial (2006). Currently Professor at the Instituto Federal do Piau (IFPI) - Brazil, working with image processing and computer vision in the discovery of breast and lung cancer.

Aristófanes Corra Silva received a PhD degree in In- formatics from Pontifical Catholic University of Rio de Janeiro - Brazil in 2004. Currently he is a Professor at the Federal University of Maranho (UFMA), Brazil. He teaches image processing, pattern recognition and programming language. His research interests include image processing, image understanding, medical image processing, machine vision, artificial intelligence, pattern recognition and, machine learning.

Anselmo C. Paiva received BSc in civil engineering from Maranho State Univeristy -Brazil in 1990, a MSc in civil engineering-Structures and a PhD in Informatics from Pontiphical Catholic University of Rio de Janeiro - Brazil in 1993 and 2002. Currently he is a Professor at the Fed- eral University of Maranho (UFMA), Brazil.

Marcelo Gattass took his Ph.D. in 1982 from Cornell University and is a full professor at PUC-Rios Computer Science Department. He currently teaches Computer Graphics and 3D Computer Vision for graduate students and Data Structures for undergraduates. He is also the Director of Tecgraf/PUC-Rio Computer Graphics Technology Laboratory, where he supervises several industry-cooperation projects in the areas of 3D modeling and visualization, geographic information systems, user interfaces, and web-based applications.