

Automatic Detection of Masses in Mammograms Using Quality Threshold Clustering, Correlogram Function, and SVM

Jobberth de Nazaré Silva · Antonio Oseas de Carvalho Filho ·
Aristófanés Corrêa Silva · Anselmo Cardoso de Paiva ·
Marcelo Gattass

Published online: 3 October 2014
© Society for Imaging Informatics in Medicine 2014

Abstract Breast cancer is the second most common type of cancer in the world. Several computer-aided detection and diagnosis systems have been used to assist health experts and to indicate suspect areas that would be difficult to perceive by the human eye; this approach has aided in the detection and diagnosis of cancer. The present work proposes a method for the automatic detection of masses in digital mammograms by using quality threshold (QT), a correlogram function, and the support vector machine (SVM). This methodology comprises the following steps: The first step is to perform preprocessing with a low-pass filter, which increases the scale of the contrast, and the next step is to use an enhancement to the wavelet transform with a linear function. After the preprocessing is segmentation using QT; then, we perform post-processing, which involves the selection of the best mass candidates. This step is performed

by analyzing the shape descriptors through the SVM. For the stage that involves the extraction of texture features, we used Haralick descriptors and a correlogram function. In the classification stage, the SVM was again used for training, validation, and final test. The results were as follows: sensitivity 92.31 %, specificity 82.2 %, accuracy 83.53 %, mean rate of false positives per image 1.12, and area under the receiver operating characteristic (ROC) curve 0.8033. Breast cancer is notable for presenting the highest mortality rate in addition to one of the smallest survival rates after diagnosis. An early diagnosis means a considerable increase in the survival chance of the patients. The methodology proposed herein contributes to the early diagnosis and survival rate and, thus, proves to be a useful tool for specialists who attempt to anticipate the detection of masses.

Keywords Medical image · Breast cancer · Computer-aided detection · Mass detection · Quality threshold

J. de Nazaré Silva · A. O. de Carvalho Filho · A. C. Silva (✉) ·
A. C. de Paiva
Applied Computing Group - NCA/UFMA, Federal University
of Maranhão, Av. dos Portugueses, SN, Campus do Bacanga,
Bacanga, São Luís, MA 65085-580, Brazil
e-mail: ari@dee.ufma.br

J. de Nazaré Silva
e-mail: jobberth1@yahoo.com.br

A. O. de Carvalho Filho
e-mail: antoniooseas@gmail.com

A. C. de Paiva
e-mail: paiva@deinf.ufma.br

M. Gattass
Department of Computer Science, Pontifical Catholic University
of Rio de Janeiro - PUC-Rio, R. Marquês de São Vicente, 225,
Gávea, Rio de Janeiro, RJ 22453-900, Brazil
e-mail: mgattass@tecgraf.puc-rio.br

Introduction

Breast cancer is the most frequent type of cancer in the female population [32]. According to Parkin et al. [25], it is also the type of cancer that kills women most often. Early diagnosis of this disease is the main form of fighting it.

A mammogram is radiography of the breast, which allows the early detection of breast cancer because it is capable of displaying lesions at their initial stages, having sizes in the range of millimeters. It is obtained through an appropriate X-ray device, called the mammographer. The precision of the mammogram depends on several factors, such as the size and location of the lesion, the density of the breast tissue and the quality

of the technical resources used. In addition, the task of carefully interpreting a large number of cases demands time and an elevated degree of attention from a specialist physician.

In recent decades, computational techniques have been developed with the purpose of automatically detecting structures that might be associated with lesions in the mammography exams, with the objective of improving the rate of early detection of structures that are connected with breast cancer [8]. These schemes are known as computer-aided detection (CAD) and computer-aided diagnosis (CADx) systems, and they are already present in many image-based diagnosing centers, mainly in developed countries such as the USA and some countries in Europe [10, 30]. CAD and CADx systems provide a second opinion, helping the radiologist in the interpretation of results that, in many cases, become difficult due to distortions that are suffered by the image during the acquisition process.

In this study, we propose a new methodology for the automatic detection of masses in digital mammograms, i.e., a CAD system. The procedure consists of preprocessing for image enhancement with a low-pass filter, an increase in the contrast scale, and a wavelet transform. Afterward, the clustering is performed by means of a quality threshold (QT), after which occurs post-processing, whose purpose is the reduction of false positives based on shape and texture measurements. These measurements will be used in the classification that is performed by a support vector machine (SVM).

The main contribution of this study is the detection of regions that are suspicious for having breast cancer based on digital mammograms using QT for the segmentation of regions of interest and a correlogram function for the extraction of texture from the mammograms. The use of both techniques contributes to CAD methodologies and to studies in this area. Furthermore, the relevant results achieved in this work allow the methodology to be used by specialist physicians in the analysis of mammograms because this approach does not suffer from the human variability that results from individual perceptions and decision making.

This paper is organized as follows: In the “[Related Works](#)” section, related studies are presented. The “[Materials and Methods](#)” section presents all of the steps of the automatic segmentation of masses, explaining in detail the use of techniques such as the QT algorithm and the extraction of shape and texture features from the mass candidates, which compose the proposed methodology. In the “[Results and Discussion](#)” section, we analyze all of the results that were achieved with the application of the methodology. Finally, in the “[Conclusion](#)” section, we present the final remarks about this study.

Related Works

In the literature, there are acknowledged studies that concern the same problem that is addressed by the methodology proposed herein, i.e., a methodology that helps specialists in the detection of masses in digital mammograms

Oliveira Martins et al. [24] present a methodology that uses Moran’s indexes and Geary’s coefficients as measurements to be extracted from suspect regions of the mammograms, with the objective of discriminating them into normal and abnormal categories. Both of these techniques have proven to be very promising. They achieved an accuracy of 96.04 % and an area under the receiver operating characteristic curve (AUC ROC) of 0.946 with the Geary’s coefficient. With the Moran’s index, they achieved an accuracy of 99.39 % and an AUC ROC of 1 in the classification of tissues in mammograms as either normal or abnormal. The methodology also achieved an accuracy of 88.31 % and an AUC ROC of 0.804 with Moran’s index in the classification of tissues as either benign or malignant.

Oliveira Martins et al. [23] introduced a methodology that uses a growing neural gas (GNG) to segment the lesion candidates, and the SVM was pooled with Ripley’s K-function for the detection of masses. In that methodology, they used 997 images from the Digital Database for Screening Mammography (DDSM), with 436 images used for testing and 561 used in the evaluation of the process of detecting the masses. The results were 89.30 % sensitivity, with 0.93 false positives per image and 0.02 false negatives per image.

Oliveira Martins et al. [19] also introduced a methodology for the detection of masses in digital mammograms using the K-means algorithm for the segmentation of the images and a co-occurrence matrix to describe the textures of the segmented structures. The classification of these structures is performed by the SVM, which classifies them into masses and non-masses. This methodology achieved an 85 % accuracy.

Nunes et al. [21] proposed a methodology for the detection of masses using the K-means algorithm and template matching for the segmentation of suspect regions. This research used 650 images from the DDSM database for tests. The segmentation of regions of interest found 92.77 % of the masses and also selected 2076 non-masses. Afterward, geometry and texture features were extracted from each of these regions, with texture being described through Simpson’s diversity index to reduce the number of false positives. Finally, the data were submitted to an SVM for the classification of suspect regions into mass and non-mass categories. The results were a mean accuracy of 83.94 %, sensitivity of 83.24 %, specificity of 84.14 %, 0.55 false positives per image, and 0.17 false negatives per image.

Bajger et al. [4] presented an automatic method for the detection of masses in mammograms that used the segmentation of the region of interest by statistical fusion and linear discriminant analysis. The performance of this methodology was evaluated with 36 images selected from a proprietary mammogram database and with 48 images from the DDSM database. The AUC for the classification of each region was 0.90 for the proprietary images and 0.96 for the DDSM database.

Liu et al. [18] presented a system for the automatic detection of masses in mammographic images. This system combined the techniques of multiple concentric layers, narrow range regions with active contours, full binary local patterns, and SVMs. The resulting method was evaluated with a set of 231 images that contained 245 masses. From those images, 125 (containing 133 masses) were used to train the SVM. The remaining images were used for performance testing. The investigators achieved a rate of 1.36 false positives per image and a sensitivity of 76.8 %.

Abdalla et al. [2] detected masses in digital mammograms by means of second-order statistics. The extraction of features from the segmented region of interest was performed by means of gray level co-occurrence matrices (GLCM). The results showed that the GLCM had a sensitivity and specificity of 91.67 and 84.17 %, respectively.

Abdalla et al. [1] propose a methodology for the automatic classification of masses in digital mammograms. This methodology uses Haralick's descriptors derived from the gray-level co-occurrence matrix for the extraction of texture features. The methodology achieved an accuracy of 95.85 %.

A new approach was presented by Mini [20] for the classification of digital mammograms into normal and abnormal classes for the detection of breast cancer. First, the structures in the mammogram produced by normal glandular tissue of varying density are eliminated by using a wavelet transform and a probabilistic neural network for classification. The results show a sensitivity of 58 % and an accuracy of 60 % for normal cases and a sensitivity of 91 % and an accuracy of 94.2 % for abnormal cases.

AbuBaker [3] proposed a method for the detection of masses through wavelet transform decomposition and SVM. The Mammographic Image Analysis Society (MIAS) image database was used. The results generated a rate of 0.05 false positives per image and an accuracy of 94 %.

Hussain et al. [13] proposed the representation of textural features of mammograms using the Weber local descriptor (WLD). The WLD was used to represent the statistical properties of the masses and to reduce the rate of false positives and false negatives per image. The WLD builds statistical data on protuberant micro-patterns together with the gradient orientation of the present pixel. The SVM was employed for the classification of suspect regions into mass and

non-mass categories. The accuracy of the proposed system was 98.8 % using DDSM.

The Gabor filter banks, which allow the manipulation of several parameters such as frequency, orientation, eccentricity and symmetry, were used by Hussain et al. [14] to extract the local spatial textures from the properties of the masses that are present in the mammograms, at different scales and in different directions. By using this feature-extraction resource and SVM with cross-validation, an accuracy of 99.5 % was achieved.

de Carvalho et al. [6] presented an application of the QT algorithm for the segmentation of structures that resemble a lung nodule. That work showed that the QT was able to detect up to 95 % of the lung nodules in the exam database that was used. In addition, the easy setup and use were the main advantages of the algorithm. On the other hand, it was shown that the processing time demanded by the QT is too high.

The studies described above prove that the methodologies for the computer-based detection of masses in mammograms are prominent. Another important point is the use of the public image database, DDSM, which deserves attention due to the number of studies that have used it. It can be noted as well that the use of SVMs as classifiers presents excellent performance with regard to the generalization of the results.

In general, we can enumerate certain problems that have been found in some of the studies cited above to exemplify noise that was left in the stage of the mammogram background removal, a large number of mass-candidate regions after segmentation, and a high rate of false positives.

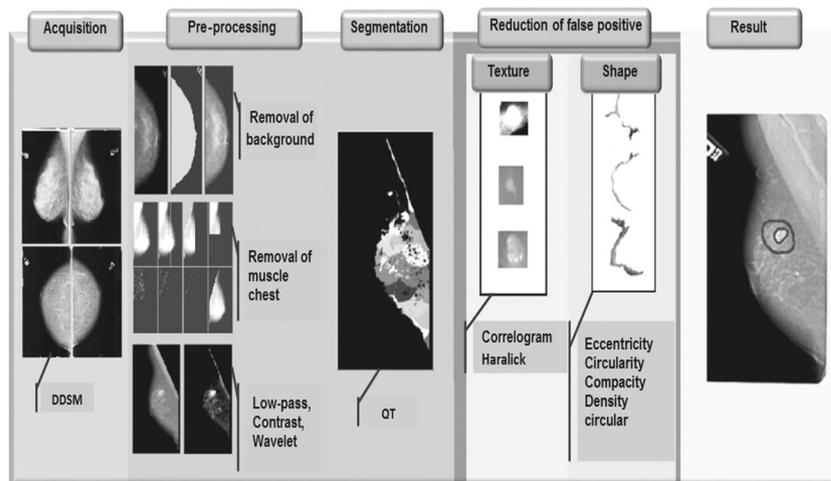
Materials and Methods

In this section, we explain the techniques that were used in this work for the detection of masses in mammographic images. The methodology comprises the following stages: image acquisition, preprocessing, segmentation of regions of interest, and reduction of false positives. This scheme can be observed in Fig. 1.

Image Acquisition

The images used in this work were obtained from the DDSM digital image database, which contains 2620 cases acquired in the Massachusetts General Hospital, Wake Forest University, and Washington University School of Medicine in St. Louis [12]. Each case has four images (left and right breast in craniocaudal and mediolateral view). The DDSM database contains descriptions of lesions in the mammograms according to the American College of

Fig. 1 Stages of the proposed methodology



Radiology in addition to information about the images themselves (e.g., the type of film and digitizer used, number of pixels, and number of bits per pixel).

In this study, we used 599 images from the DDSM, all of them randomly chosen, among which 517 images present at least one mass and the remaining 82 are normal; in other words, they have no mass.

Preprocessing

The preprocessing stage involves image resizing, noise reduction, and the segmentation of the breasts. Each pair of images to be used in the methodology has its dimension reduced by a factor of 4 to reduce the processing time. We ran tests with the image in its original size (approximately 5000×5000 pixels), with reduction factors of 2, 3, 4, and 5. For the original image, the processing time was too high. The reductions with factors from 2 to 4 had very close results. This finding can be explained by the fact that there are no masses that have very small dimensions in the DDSM. If there were such masses, the reduction by a factor of 4 would make the mass less visible in the image. This problem occurred when the image was reduced using a factor of 5, which hindered the correct identification of some of the masses. For these reasons, we chose the reduction by a factor of 4.

The images underwent a noise reduction process. In this study, the noise reduction was performed through the application of the 5×5 median filter to the whole image. We ran tests in which the kernel of the median filter was size 3×3 , but in many images, the noise was not completely removed. Thus, we tested the filter with size 5×5 and obtained excellent results. The noise was removed in virtually all of the images. For this reason, we chose this size.

Mammogram images usually present a series of elements (called artifacts) that lie on the background region and contain information about the exam (e.g., labels, data about

the film used). These artifacts could interfere in the image segmentation and, therefore, must be preferably removed. The segmentation of the breast region is intended to remove artifacts and reduce the regions scanned by the algorithms, increasing the speed and precision of the operations that will be performed on the images. In this study, we used a method that was proposed in [27].

Selective Contrast Increase

In this stage, we attempted to make the ROIs more perceptible by applying a selective contrast increase. Through an empirical analysis of the several mammograms used in this study, we realized that the masses have minimum and maximum gray-level values that are between 70 and 235. The pixels in this interval were scaled to gray-level values in the range of 0 to 255. The other pixels were mapped to zero. By taking this step, we emphasize the regions that will

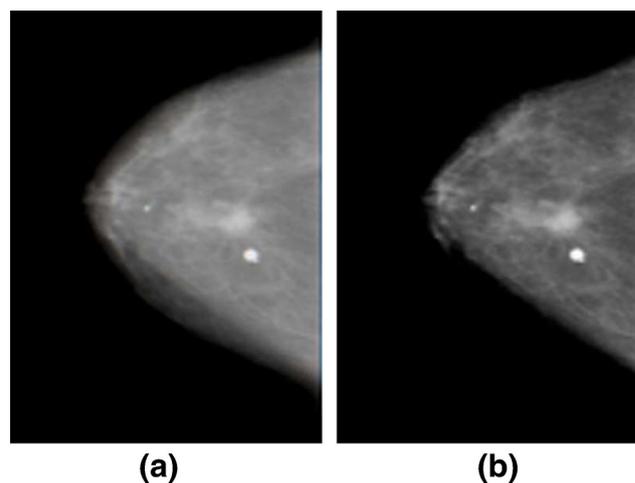


Fig. 2 Application of selective contrast increase: **a** reduced and improved image and **b** image with selective contrast increase

eventually be classified as masses. In Fig. 2, we can see the increase in the contrast applied to the image processing in the previous stage.

Enhancement with Wavelet Transform

For the finalization of the preprocessing stage, we applied an enhancement with the Daubechies 4 wavelet transform (DWTB4) [31] to make the mass-candidate regions more evident. After decomposing the image into its coefficient groups, we applied a linear function (1) to each coefficient to map them to other coefficients and enhance the image after the application of the reverse DWTB4. After some empirical tests, the best value for K in the linear equation was $\sqrt{2}$. The result of the application of enhancement with DWTB4 to the image with selective contrast increase can be seen in Fig. 3.

$$Wout_{i,j} = K \times Win_{i,j} \tag{1}$$

where i and j are the indexes that indicate the position of the coefficients in the sub-bands, $Win_{i,j}$ are the wavelet components of the original image, $Wout_{i,j}$ are the enhanced DWTB4 coefficients, and K is the parameter responsible for the adjustment of the method (which is in the dependence of the decomposition level, in this case, the gain itself). The value of this parameter decreases to a rate of $K^{0.5}$ with the wavelet decomposition level.

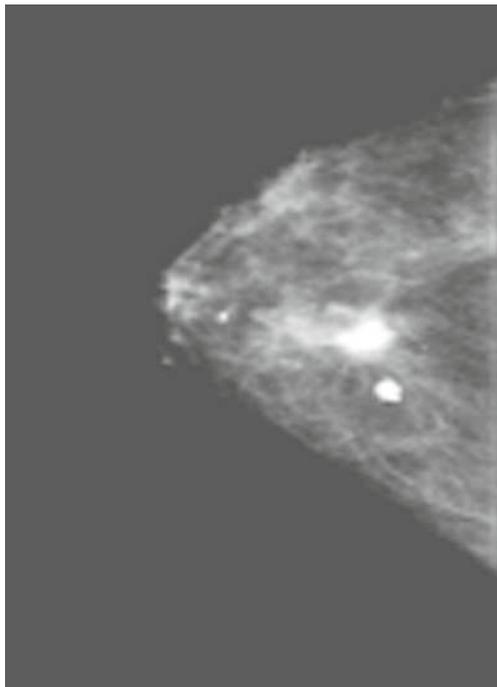


Fig. 3 Mammogram enhanced with DWTB4

Segmentation

To perform segmentation of the regions of interest, we used the QT clustering algorithm [15, 17]. The QT algorithm was first proposed for clustering genes of the human chromosome. Its operation does not depend on the pre-determination of the number of clusters, always returning the global minimum of the solution.

This model results in a more precise clustering because a suitable number of clusters is allocated for each case. The computational cost of the model, on the other hand, is significantly higher than the K-means and C-means models [33].

QT requires only one parameter for its execution, which is called the *quality threshold* or *diameter threshold*. This parameter is responsible for delimiting the area of each cluster. We will call this threshold the *maximum diameter*.

In our implementation, in addition to the maximum diameter of the original algorithm, we will use a minimum diameter. This minimum diameter has the goal of avoiding the formation of clusters that have too-small diameters because we found that by means of observations in various mammograms of the DDSM, diameters that were smaller did not correspond to masses. In this way, we minimize the prejudicial effects (increased processing time, poorly formed shapes, and a small area for computing the texture) that are caused by structures with diameters of below 15 in the subsequent steps of the methodology. We defined empirically the values 15 and 100 for the minimum and maximum diameter, respectively. At the formation of the cluster, to determine whether a randomly chosen pixel belongs to a candidate cluster or not, we consider the following criterion: if the value that is computed through Eq. 2, between the random pixel and the reference pixels of this cluster is at most the maximum diameter, then the pixel belongs to a candidate cluster. This process is called the formation of candidate clusters.

$$C = \sqrt{(i - i') + (j - j') + (\text{pixel} - \text{pixel}')^2} \tag{2}$$

where i and j are the indexes of a reference pixel of the candidate cluster, i' and j' are the indexes of the pixel that was randomly chosen, and pixel and pixel' are the gray levels of the reference pixels of the candidate clusters and of the randomly chosen pixels, respectively.

In this way, the candidate that possesses the highest number of points associated with it will be chosen as the true cluster. Thus, all of the points that belong to this candidate will be removed from the main sample, and the process is repeated until all of the points are located in their ideal clusters.

In Fig. 4, we have an example of a mammogram that was segmented using the QT algorithm. This figure illustrates in five stages the process of the formation of clusters.

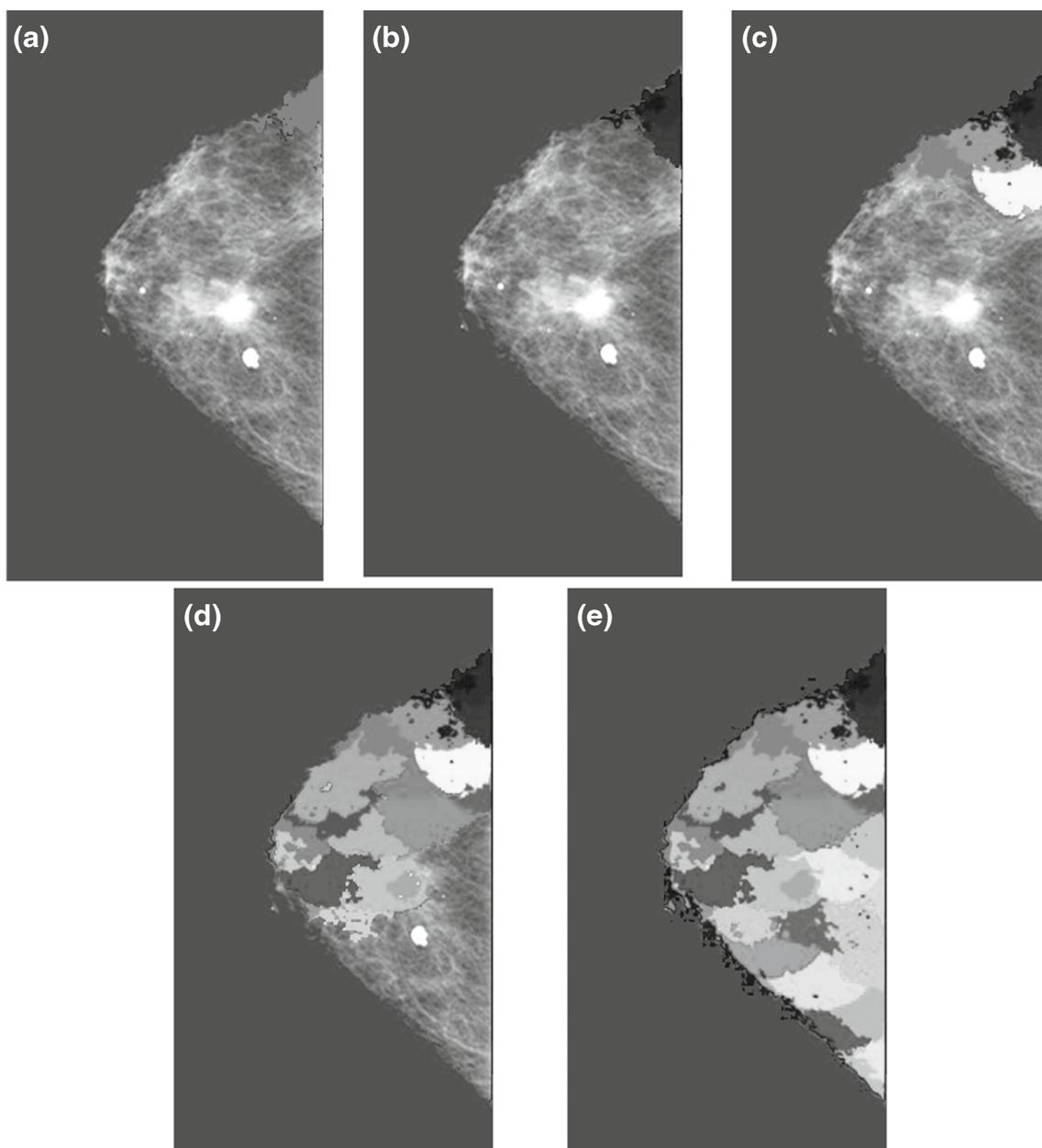


Fig. 4 Process of formation of clusters by the QT: **a** cluster 1, **b** clusters 1 and 2, **c** clusters 1 to 5, **d** clusters 1 to 18, and **e** all clusters

Reduction of False Positives

The result of the previous stages yields several clusters that are not masses; these are considered to be false positives. This stage of the methodology is intended to reduce, as much as possible, the number of false positives by using shape and texture descriptors combined with the SVM [9].

The SVM is a powerful state-of-the-art algorithm that has a strong theoretical foundation based on the Vapnik-Chervonenkis theory. The SVM has strong regularization properties. Regularization refers to the generalization of the model to new data. This attribute was the main

reason for choosing this classifier in our work. The accuracy of an SVM model is highly dependent on the selection of kernel parameters such as C and λ for a radial basis function (RBF). We used the LibSVM software [7] to estimate both of these parameters. The values estimated for C and λ were, respectively, 512 and 0.00712343 for the first false-positive reduction by using shape descriptors, and 512 and 0.00832342 for a second false-positive reduction with texture descriptors.

The segmentation performed by the QT uses spatial values to delimit the clusters. Thus, these clusters can produce the following cases:

1. Cluster with high texture variation (characteristic of a mass)
2. Cluster with low texture variation (characteristic of a non-mass)
3. Cluster with round shape (characteristic of a mass)
4. Cluster with irregular shape (characteristic of a non-mass)
5. Cluster with high texture variation (characteristic of a mass), but with irregular shape (characteristic of a non-mass)
6. Cluster with low texture variation (characteristic of a non-mass), but with a round shape (characteristic of a mass)

During the tests, we observed that the production of the cluster of the cases 4 and 5 had very expressive numbers. This fact jeopardized the performance of the SVM in the reduction of false positives when the texture and shape descriptors were combined because the classifier was unable to determine a pattern for mass and non-mass categories.

To solve this problem, we adopted the strategy of reducing the false positives into two stages. In the first stage, we used the shape descriptors and the SVM, and in the second stage, we used texture descriptors and the SVM. Thus, we made a significant improvement in the reduction of false positives.

Shape Descriptors

After the clustering stage with QT, we observed that some mass-candidate clusters presented “elongated,” “twisted,” or “hollow” shapes, as seen in Fig. 5, not matching the expected shape of a mass, which is closer to “circular” [26].



Fig. 5 Clusters with inappropriate shapes

Such clusters were eliminated after the analysis of the eccentricity, circularity, compactness [22], and circular density shape descriptors [29] and were classified by the SVM algorithm. Thus, several clusters with these characteristics were eliminated, which resulted in a significant reduction in false positives for the subsequent processing stages.

Texture Descriptors

The next stage in the methodology is the extraction of features, which is intended to cause a deeper reduction in false positives. For the extraction of texture features, we used Haralick’s descriptors based on the co-occurrence matrix and the correlogram function.

Haralick’s Descriptors

Haralick et al. [11] proposed a method for the extraction of textural features based on the spatial relation that exists between the gray levels of the image, computing so-called co-occurrence matrices (GLCM) and, based on these matrices, determining 14 statistical parameters that describe texture. Due to its simplicity and efficiency, this method became very popular in the classification of textures. According to [5], among the 14 statistics originally proposed, only 6 have a higher relevance: the second angular momentum, contrast, correlation, entropy, variance, and homogeneity. These six features are used here in the present study.

First, we compute the GLCM for each cluster. In the computation of these matrices, we adopted distance three; in other words, the reference pixel is related to three neighboring subsequent pixels in a certain direction. This distance was adopted because after some empirical tests with it, Haralick’s texture measurements that were extracted were those that best characterized the clusters, which led to better classification. The four directions that were adopted were 0° , 45° , 90° , and 135° , with respect to the horizontal plane. Thus, for each distance/direction combination, there is a GLCM; thus, we obtained 12 co-occurrence matrices for each cluster. In turn, for each co-occurrence matrix, we computed the six descriptors that are cited above, which led to a total of 72 features per cluster.

Correlogram Function

In addition to Haralick’s descriptors, the correlogram function (correlogram) was used in the extraction of features from the clusters. The correlogram is the normalized version of the covariance function. The correlation coefficients range from -1 to 1 . The correlation is expected to be higher for the units that are close to each other (correlation = 1 for

distance zero), and it tends to zero as the distance between the units increases [28].

The correlation is defined by

$$\rho(h) = \frac{C(h)}{\sigma_{-h}\sigma_{+h}} \quad (3)$$

where σ_{-h} is the standard deviation of the values of the vector origins, and σ_{+h} is the standard deviation of the values of the vector targets. These parameters are defined by the following:

$$\sigma_{-h} = \sqrt{\frac{1}{N(h)} \sum_{i=1}^{N(h)} x_i^2 - m_{-h}^2} \quad (4)$$

$$\sigma_{+h} = \sqrt{\frac{1}{N(h)} \sum_{i=1}^{N(h)} y_i^2 - m_{+h}^2} \quad (5)$$

$C(h)$ is defined by

$$C(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} x_i y_i - m_{-h} m_{+h} \quad (6)$$

where x_i e y_i are the starting and target pixels, respectively, h is the distance between the pixel at the origin and the pixel

at the extremity, $N(h)$ is the number of pairs within the distance h , m_{-h} is the mean of the values of the pixels at the origin, and m_{+h} is the mean of the target pixels.

Based on the diameter of the clusters found in the segmentation stage, for as many as 100 pixels, we assigned values of h that ranged from 1 to 85 pixels to this function. Thus, with this configuration of the function, we started having texture measurements that were capable of providing a good specification of the clusters because they result from the relationship among almost all of the pixels of the segmented regions.

Results and Discussion

In this section, we present the results that were achieved with the methodology that was developed. We analyze the behavior of the stages described in the “Materials and Methods” section. The results are examined at two distinct moments: the first moment is the training and validation of the SVM model and the second moment concerns the test of the process of the detection of masses. In this test, new mammograms are used, which are verified in their measurements of true positive (TP), false

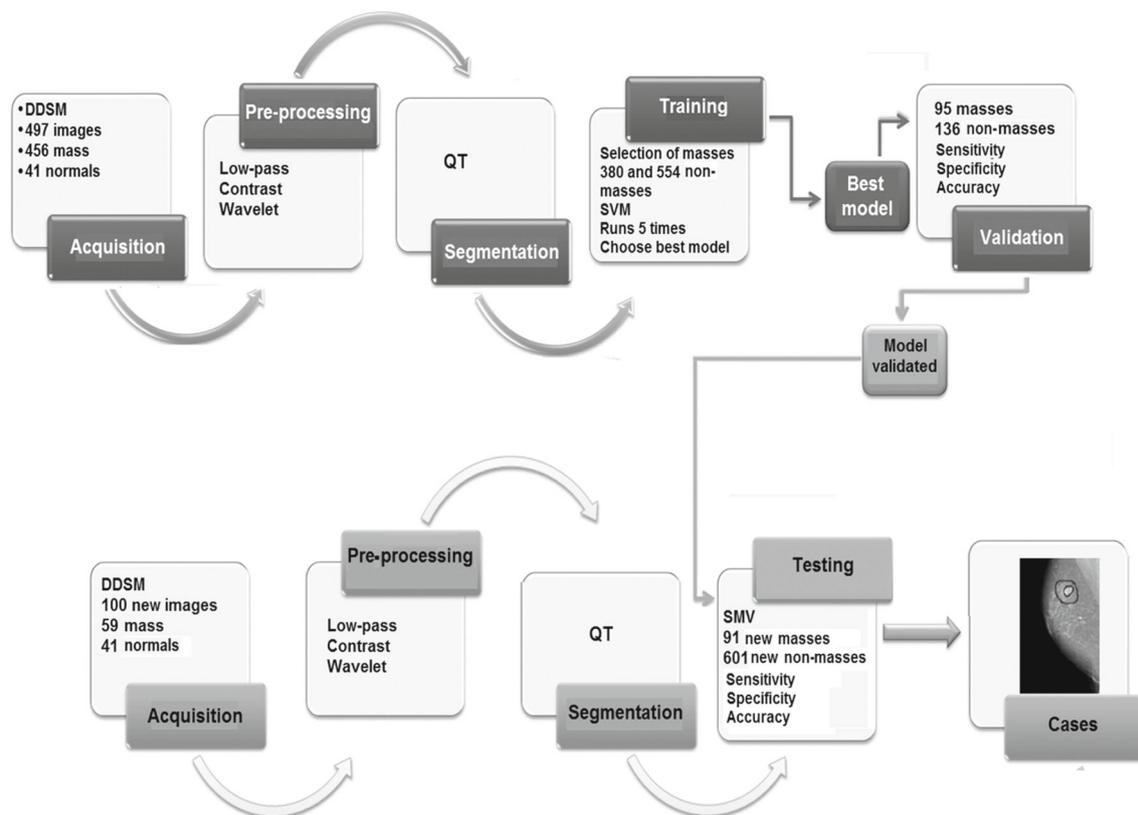


Fig. 6 Results analysis flow

positive (FP), true negative (TN), false negative (FN), sensitivity (Se), specificity (Sp), accuracy (Ac), false positives rate per image (FPI), ROC, and free-response ROC curve (FROC) [16]. We also study five cases of the detection of masses and make a comparison with the results obtained by the methods referred to in the “Related Works” section.

To analyze and validate the proposed methodology, we used 599 mammograms, among which 517 presented at least one mass and 82 were normal (without masses). From these 599 mammograms, 499 (458 with masses and 41 normal) were used for the training and validation of the SVM model, and the remaining 100 (59 with masses and 41 normal) were used to examine the process of finding masses in the mammograms.

An overall scheme with the flow of the analysis of the results can be seen in Fig. 6.

Training and Validation of the SVM Model

In this stage, we used 499 mammogram images (458 with masses and 41 normal). In 33 (6 %) images, no masses were segmented. Thus, from the total of 458 images that contain masses, only 425 will be used for the next stages, i.e., training and validation.

For training the SVM, the sample was defined in the proportion of one mass for each two non-masses. This division comes from the fact that there is a larger number of non-mass clusters. Thus, from all of the clusters found in the 425 mammograms, we randomly chose 1396 mass clusters and 2792 non-mass clusters, totaling 4188 clusters.

Selection of the Best Mass Candidates Based on Shape Features

From the 4188 initial clusters, we chose those that best represent the mass and non-mass classes according to their shape features, to train and validate the SVM model with shape descriptors for selecting the best mass candidates. Thus, we chose 1167 out of the 4188 clusters that best represent the masses and non-masses. These clusters were distributed according to Table 1.

With this configuration, the validation of the SVM model achieved the results that are observed in Table 2. Having this model generated and validated, from the 4188 clusters, only 2033 (704 mass clusters and 1329 non-mass ones) were

Table 1 Configuration of the mass and non-mass clusters according to the shape measurements for training and validation of the SVM model

Stages	Mass clusters	Non-mass clusters
Train	380 (80 %)	554 (80 %)
Validation	95 (20 %)	138 (20 %)

Table 2 Results of the validation of the SVM model according to the shape descriptors

TP	FP	TN	FN	Se (%)	Sp (%)	Ac (%)
81	19	116	17	82.65	85.92	84.54

selected as the best mass candidates; then, we proceeded to the next stage.

hypertargetsec:6

Selection of the Best Mass Candidates Based on Texture Features

The clusters that resulted from the previous stages have their texture features extracted to be used in the new training and validation of the generated SVM models. The new base is formed from 2033 clusters that result from post-processing, where 704 are mass clusters and 1329 are non-mass clusters.

The distribution of clusters for training the classifier and for validation of the SVM model was performed in three different manners: first, we used 30 % (610 clusters, composed of 211 masses and 399 non-masses) for training the classifier and 70 % (1423 clusters, composed of 493 masses and 930 non-masses) for validation of the SVM model; then, we adopted 50 % (1016 clusters, composed of 352 masses and 664 non-masses) for training the classifier and 50 % (1017 clusters, composed of 352 masses and 665 non-masses) to validate the SVM model; and finally, we assigned 70 % (1423 clusters, composed of 493 masses and 930 non-masses) for training the classifier and 30 % (610 clusters, 211 masses and 399 non-masses) for validation of the SVM model.

Each configuration was randomly executed five times, and then, we obtained the values for sensitivity, specificity, and accuracy. The highest means serve as a basis for choosing the best configuration. Next, using the best configuration, we chose the best model. In our work, the configuration (70/30) presented the highest mean for the sensitivity (72.49 %), specificity (85.13 %), and accuracy (81.08 %). Then, for this configuration, we checked which of the five classifications presented the highest values for sensitivity, specificity, and accuracy. As seen in Table 3, the best model achieved 73.96, 86.81, and 82.76 % for the sensitivity, specificity, and accuracy, respectively.

Test Model

In this section, we check the overall efficiency of the proposed methodology. For this step, we used 100 new mammograms, from which 59 have at least one mass and the remaining 41 are normal (present no masses).

Table 3 Results achieved with the configurations for training and validation of the SVM model with texture measurements (Haralick and correlogram)

Training/Validation	TP	FP	TN	FN	Se (%)	Sp (%)	Ac (%)
30/70	342	132	786	163	67.72	85.62	79.27
	346	135	800	142	71	85.56	80.53
	339	114	808	162	67.66	87.64	80.6
	357	144	791	131	73.16	84.6	80.67
	353	152	781	137	72.04	83.71	80
			Mean	70.31	85.42	80.21	
50/50	258	127	556	75	77.78	81	80.12
	255	99	559	103	71.23	84.95	80.12
	233	93	582	108	68.33	86.22	80.22
	244	86	590	96	71.76	87.28	82.08
	248	108	554	106	70.06	83.69	79
			Mean	71.83	84.62	80.3	
70/30	154	58	332	65	70.32	85.13	80.8
	167	84	314	44	79.15	78.89	79.14
	155	54	344	56	73.46	86.43	82.26
	139	46	351	73	65.57	88.41	80.46
	142	55	362	50	73.96	86.81	82.76
			Mean	72.49	85.13	81.08	

The clustering process through QT segmented the masses in 54 of the 59 mammograms that contained masses. This finding means that there was an accuracy of 91.52 % in the segmentation of the masses. This segmentation also generated 192 mass clusters and 1214 non-mass clusters.

The results achieved by the proposed method for the final test using the best SVM model generated in the training and validation stage (“[Selection of the Best Mass Candidates Based on the Shape Features](#)” section) can be observed in Table 4.

Case Studies

In this stage, we checked all of the steps of the proposed methodology in three cases that have different degrees of complexity. In the first case, the methodology was successful in the detection of the mass. In the second case, the methodology detected the mass and other regions that are

Table 4 Test results

Masses	Non-masses	TP	FP	TN	FN	Se (%)	Sp (%)	Ac (%)	FPI	FROC
91	601	84	107	494	7	92.31	82.2	83.53	1.12	0.8033

non-masses. Finally, in the third case, a normal mammogram was analyzed, and the mammogram was successful, not detecting any mass.

First Case

The steps for the detection of masses in mammograms can be seen in Fig. 7. The mammogram used in this case was the A_1306_1.LEFT_MLO, from the benign volume 02 in the DDSM database. Figure 7 displays the lesion, which is marked by the specialist with a red contour. In the first step of preprocessing, the undesired objects are removed for processing. These objects are not part of the breast itself and include side edges, chest muscle, and labels placed by the specialist. The result is seen in Fig. 7b.

At the second moment, still in preprocessing, the mammogram undergoes an improvement of the internal structures by the combined action of a low-pass filter, a selective contrast increase, and an enhancement with WT. The result can be seen in Fig. 7c. In the next stage, the mammogram is segmented (Fig. 7d) through the QT clustering algorithm using maximum and minimum diameters of 100 and 15 pixels, respectively.

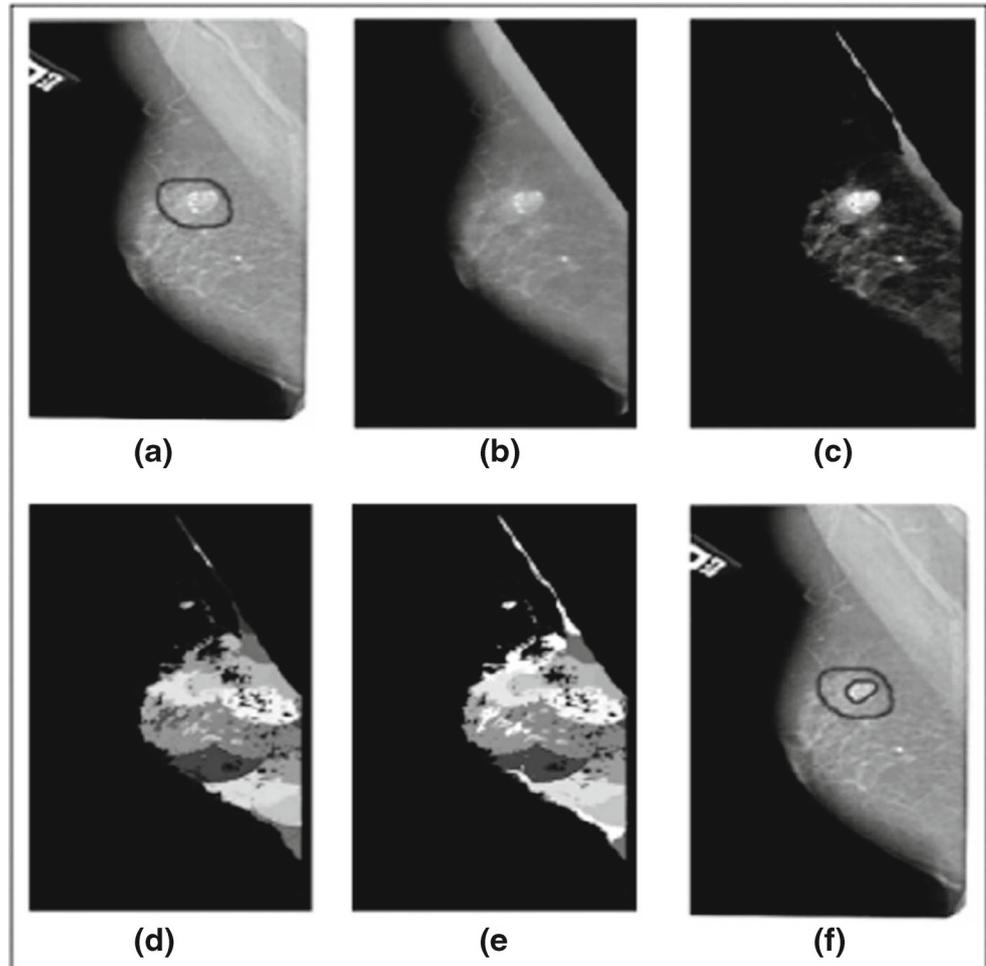
Observing the mammogram resulting from the segmentation with QT, we note that the shape of certain clusters is not compatible with the shape of the masses because they are elongated or too distorted. Thus, in the next stage, these clusters will be eliminated, which causes the selection of the best mass candidates through the analysis of geometric features proposed by the SVM model. Figure 7e shows the eliminated clusters according to their shapes (in white).

In this way, only clusters that are judged as masses in the previous stage have their texture features extracted by Haralick’s descriptors and by the correlogram function. Using the SVM model, the clusters that are considered to be masses are indicated, which ends the detection process. In Fig. 7f, we see the successful detection with the mass circled inside the radiologist’s indication.

Second Case

The mammogram used in the analysis was the B_3387_1.RIGHT_CC, from the cancer volume 15, in the DDSM database, which is represented in Fig. 8a, where the mass is identified by a specialist with a circle. Again, in the preprocessing stage, the image background and unwanted objects are removed (Fig. 8b), and we enhance the internal structures of the breast through the same techniques used for the first case. The result can be seen in Fig. 8c. For segmentation, QT was used again, with the same parameters cited earlier (Fig. 8d). The post-processing with reduction of mass candidates also occurred through the analysis of geometric features by the SVM classifier. The result can be

Fig. 7 First case: **a** mammogram with lesion marked by the specialist, **b** mammogram after background removal, **c** enhanced mammogram, **d** mammogram segmented by QT, **e** mammogram with removed clusters painted in white, and **f** mammogram with correct detection of the mass



seen in Fig. 8e, where the clusters that are not considered to be masses by this stage are shown in white.

The mass clusters of the previous stage have their texture features extracted by Haralick's descriptors and the correlogram function, to be classified by the SVM. The result of this classification indicates the clusters that most resemble a mass. In Fig. 8f, we show mass clusters tangled by a convex lace.

In this case, we note that the methodology could detect the mass that was indicated by the specialist, but it mistakenly determined another structure to be a mass. In this situation, we observe that the region that was erroneously detected as a mass passed the stage of reduction of mass candidates (post-processing) because it has a "round" shape that is typical of masses. Thus, the failure is not in this stage. In addition to having an acceptable shape, the structure has gray levels that are close to those of real masses detected by the methodology. Thus, the error occurred in the classification because the SVM model used by the classifier was unable to see that this cluster was not a mass.

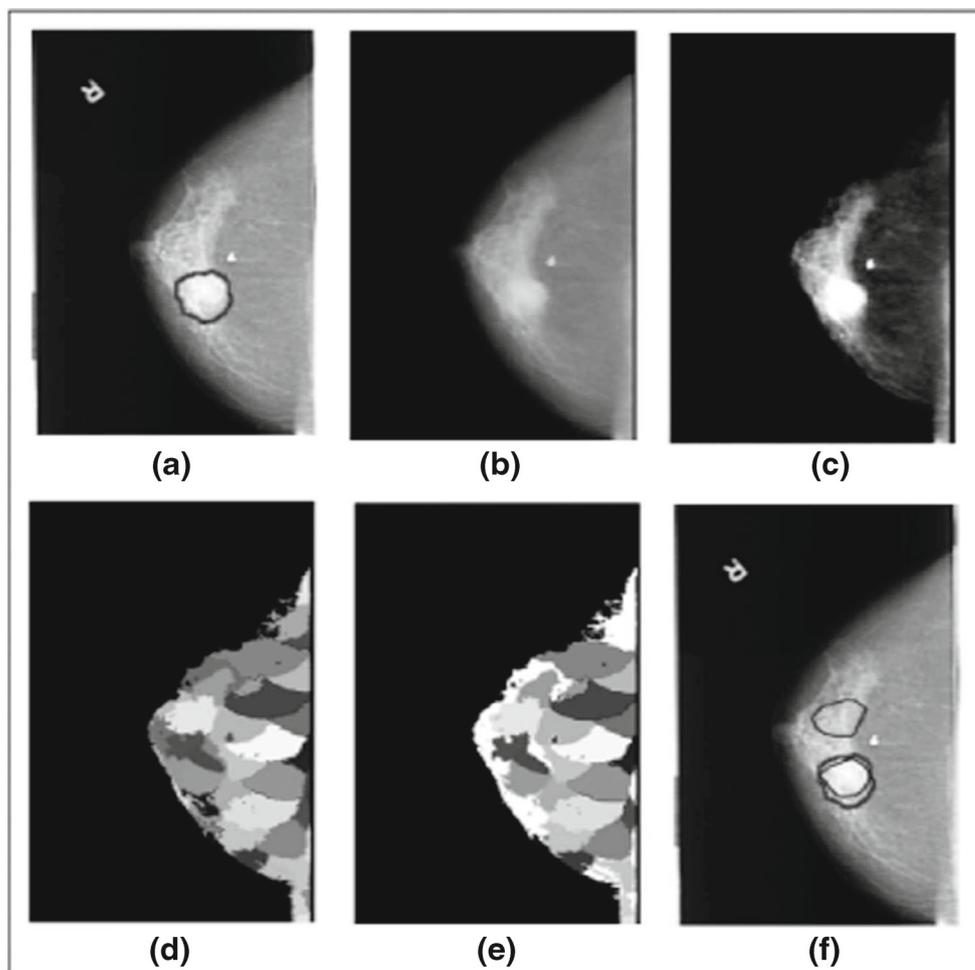
Third Case

In the third case, the methodology had good performance and did not detect any masses, as expected, because we used an image of a normal breast. The mammogram used for this case was A_0255_1.RIGHT_CC, from normal volume 02, in the DDSM database, which is seen circled in Fig. 9a, where a lesion is indicated by a specialist.

Again, by means of the preprocessing, the image passes through a process of removal of undesired objects (Fig. 9b) and an enhancement of the structures inside the breast (Fig. 9c) using the already mentioned techniques. The segmentation is performed by the QT algorithm again (Fig. 9d). The best mass-candidate clusters were selected by using shape measurements together with the SVM classifier. The clusters eliminated in this stage are painted in white in Fig. 9e.

The clusters that result from the previous stage have their texture features extracted by Haralick's descriptors and a correlogram function. Then, the features are classified by

Fig. 8 Second case: **a** mammogram with lesion marked by the specialist, **b** mammogram with background removed, **c** enhanced mammogram, **d** mammogram segmented by QT, **e** mammogram with removed clusters painted in white, and **f** mammogram with correct detection of mass and another one mistakenly detected



the SVM. The result of the detection can be seen in Fig. 9f, where no mass was detected in the original DDSM image, as expected for this case.

Results Comparison with Other Related Studies

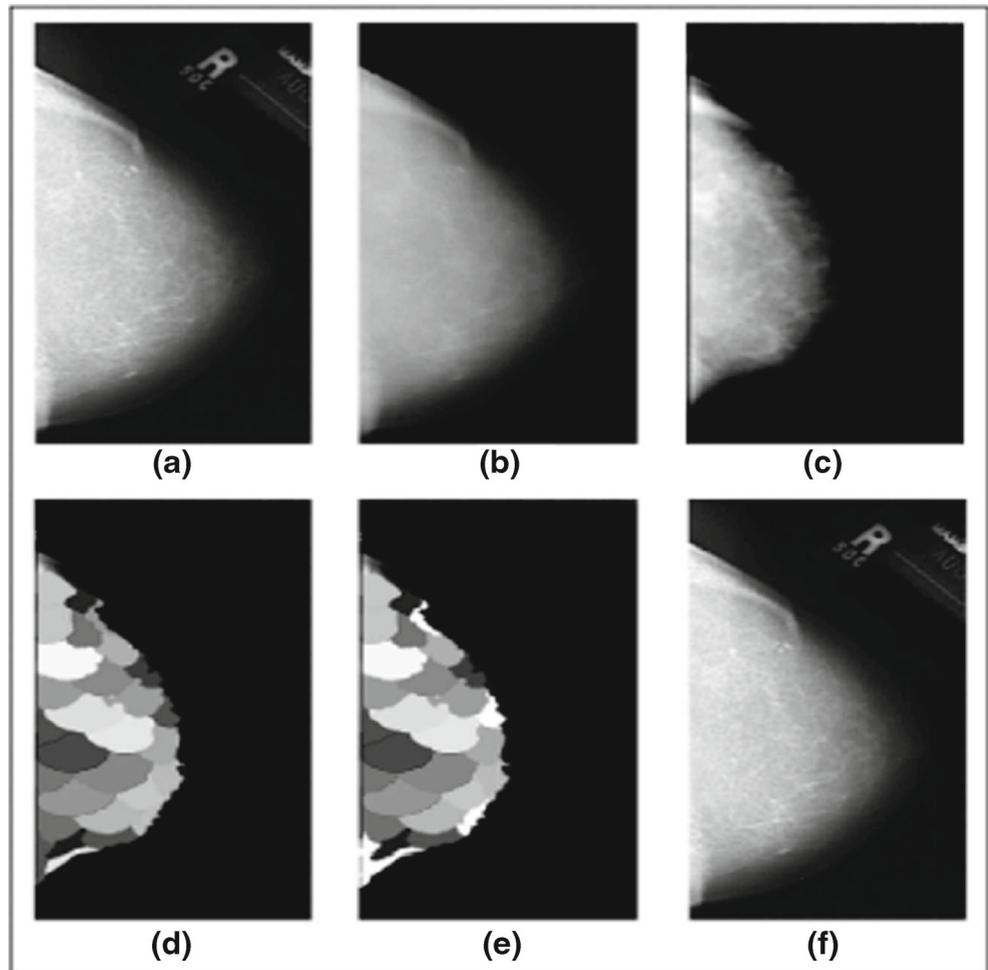
In Table 5, we make a comparison between the proposed methodology and the studies listed in the “[Related Works](#)” section to show the quality of this work. We must emphasize that for a reliable comparison, it would be necessary to use the same database, the same resolution, and the same amount of data for training, validation of the SVM, and testing.

Some of the methodologies cited in Table 5 use sensitivity as a metric for the detection performance, and the proposed methodology achieved good sensitivity, 92.31 %; it can be observed that this outcome is the highest sensitivity among the listed studies. This metric is important for measuring how capable the system is in detecting masses.

Only a few methodologies used specificity as a metric for validating the performance, which represents an obstacle in knowing the ability of such methodologies in correctly detecting non-masses. In the proposed methodology, the specificity achieved was 82.2 %, which can be considered to be acceptable.

With respect to the accuracy of 83.53 %, although it was one of the smallest accuracies, outperforming only that obtained by [20], this outcome is considered to be satisfactory if we consider that the image database used in the methodology contained only 59 images with masses and 41 normal images. Thus, with the large number of normal images in the test, it is acceptable to have a large number of non-mass clusters, which causes the TPI to increase. This result causes the diminution of the specificity and, consequently, accuracy. Still concerning the number of normal images used in the tests, a rate of FPI of 1.12 is considered to be compatible with the other studies.

Fig. 9 Third case: **a** Normal image, **b** image with background removed, **c** enhanced image, **d** image segmented by QT, **e** image with removed clusters painted in white, and **f** normal image with no mass detected



None of the other methodologies used the area under the FROC curve, which achieved 0.8033 in the present work.

Discussion

We evaluated the proposed methodology by applying it to a set of exams from the DDSM database. The experimental results led to the following conclusions:

1. The main contribution of this methodology is the QT algorithm, which has the goal of performing the segmentation of structures that are similar to masses. We are not aware of any other use of this technique for this application. In most of the cases, the QT could segment the mass without altering its shape or joining other structures of the breast. In addition to those factors, the QT is fast and easy to use, which makes the use of the methodology easy.
2. Another contribution, the use of a correlogram function combined with other texture and shape measurements, led to a significant reduction in the number of false

positives. We believe that this index had good performance for discriminating mass from non-mass categories.

3. In some of the cases, the methodology failed. Analyzing those cases, we note the following situations: (a) Classification between mass and non-mass based on texture—In this case, the values of the gray levels are very close to the values for the non-mass. As a result, the SVM failed. Most of the cases correspond to exams that were classified in DDSM as dense breasts—(b) Classification between mass and non-mass based on shape: In this case, the shape of the mass is very similar to the non-mass. We believe that at the moment of the segmentation, the QT combined close structures (with the same intensity and inside the same maximum diameter) and altered the shape of the mass; (c) Incorrect segmentation—In this case, the mass was combined with other structures of the breast. This problem occurred mainly due to masses and other structures that have very close pixel values and reduced sizes.

Table 5 Comparison between the results of proposed methodology and those of the methodologies listed in the “Related Works” section

Work	Database	Se (%)	Sp (%)	Ac (%)	ROC	FPI	FROC
Oliveira Martins et al. Oliveira-Martins:2007fk	DDSM	–	–	99.39	1	–	–
Oliveira Martins et al. [23]	DDSM	89.30	–	–	–	0.93	–
de Oliveira Martins et al. [19]	DDSM	–	–	85	–	–	–
Nunes et al. [21]	DDSM	83.24	84.14	83.94	–	0.55	–
Bajger et al. [4]	Proprietary	–	–	–	0.9	–	–
Liu et al. [18]	DDSM	76.8	–	–	–	1.36	–
Abdalla et al. [2]	DDSM	91.67	84.17	–	–	–	–
Abdalla et al. [1]	DDSM	–	–	95.85	–	–	–
Mini [20]	MIAS	58	–	60	–	–	–
AbuBaker [3]	MIAS	–	–	94	–	0.05	–
Hussain et al. [13]	DDSM	–	–	98.8	–	–	–
Hussain et al. [14]	DDSM	–	–	99.5	–	–	–
<i>Proposed method</i>	DDSM	92.31	82.2	83.53	–	1.12	0.8033

- The time spent for the whole method appears to be promising because it takes an average of 0.3 s to process one image. We highlight the performance of the QT algorithm, which performs segmentation in a relatively short time, considering the amount of information to be processed. This run time could vary (increase or decrease) depending on the threshold.
- One of the difficulties encountered in the process of segmentation with QT is the choice of the maximum diameter to form the cluster because only after various tests that we found the diameter that best suits the various sizes of the masses.
- With respect to the related studies (Table 5), we had the best sensitivity percentage. Overall, these results show that the methodology finds a large number of masses in exams. However, we had a low accuracy and a high FPI rate. It is important to stress that our testing base is much larger than that of the mentioned studies, and for this reason, those indexes are not very good. However, in general, the methodology had good performance.
- Although we consider the results of the methodology to be acceptable, some of the aspects can be improved, which enables more suggestive results. One of these aspects concerns the segmentation performed by the QT, where the choice of the diameter that is used to form the cluster is sometimes unsuitable; sometimes, it is too small and does not cover the whole mass, and sometimes, it is too large and forms clusters that cover the mass and other lesions. An improvement would be to develop a dynamic diameter that would, somehow, adapt to the size of the segmented region or to use some other property of the image as a clustering parameter.
- Another aspect that could be improved concerns post-processing with shape measures together with SVM to reduce false positives because we note that some “distorted” clusters end up passing to the next stage instead of being eliminated. Edge descriptors, such as curves and deformation energy, could be used together with shape descriptors.

Conclusion

In the present paper, we have developed a methodology for the automatic detection of breast masses. The results achieved in this study point to an acceptable performance of the developed methodology. A configuration of 1423 clusters for training the SVM (70 %) and 610 clusters for validation (30 %) led to the most suitable SVM model for testing new exams. In the testing stage with 100 new exams, 59 with masses and 41 normal, we achieved a sensitivity of 92.31 %, a specificity of 82.2 %, an FPI rate of 1.12, and a FROC of 0.8033.

Based on the results, we believe that such measures provide significant support for a more detailed clinical investigation, and the outcomes were very encouraging when the nodules were classified with the support vector machine, nearest mean classifier, and linear classifier based on a normal density. Nevertheless, it is necessary to perform tests with a larger database and more complex cases to obtain a more precise behavior pattern.

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

References

1. Abdalla, AMM, Dress, S, Zaki, N: Detection of masses in digital mammogram using second order statistics and artificial neural network. *Int J Comput Sci Inf Technol (IJCSIT)* 3(3):176–185, 2011
2. Abdalla, AMM, Dress, S, Zaki, N: Masses detection in digital mammogram by gray level reduction using texture coding method. *Int J Comput Appl* 29(4):19–23, 2011 Published by Foundation of Computer Science, New York, USA.
3. AbuBaker, A: Mass lesion detection using wavelet decomposition transform and support vector machine. *Int J Comput Sci Inf Technol (IJCSIT)* 4(2):33–46, 2012
4. Bajger M, Ma F, Williams S, Bottema, MJ: Mammographic mass detection with statistical region merging. *DICTA*, 2010, pp 27–32
5. Baraldi A, Parmiggiani F.: An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Trans Geosci Remote Sens* 33(2):293–304, 1995 doi:10.1109/36.377929.
6. de Carvalho Filho AO, de Sampaio WB, Silva AC, de Paiva AC, Nunes RA, Gattass M: Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. *Artif Intell Med* 60(3):165–177, 2014 doi:10.1016/j.artmed.2013.11.002, <http://www.sciencedirect.com/science/article/pii/S0933365713001541>.
7. Chang CC, Lin CJ: LIBSVM—a library for support vector machines 2009. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
8. Dengler J, Behrens S, Desaga JF: Segmentation of microcalcifications in mammograms: DAGM-Symposium, 1991, pp 380–385.
9. Duda RO, Hart PE: *Pattern Classification and Scene Analysis*. New York: Wiley, 1973
10. Fandos-Morera A, Prats-Esteve M, Tura-Soteras JM, Traveria-Cros A: Breast tumors: composition of microcalcifications. *Radiology* 169(2):325–327, 1988 <http://radiology.rsna.org/content/169/2/325.full.pdf+html>.
11. Haralick R, Shanmugam K, Dinstein I: Textural features for image classification. *SMC* 3(6):610–621, 1973
12. Heath M, Bowyer K, Kopans D: *Current Status of the Digital Database for Screening Mammography: Digital Mammography* (Kluwer Academic), 1998, pp 457–460.
13. Hussain M, Khan S, Ghulam M, Bebis G: Mass detection in digital mammograms using optimized Gabor filter bank: *ISVC* (2), 2012, pp 82–91.
14. Hussain M, Khan S, Muhammad G, Bebis G: A comparison of different Gabor features for mass classification in mammography: *SITIS*, 2012, pp 142–148.
15. Jiang D, Tang C, Zhang A: Cluster analysis for gene expression data: a survey. *IEEE Trans Knowl Data Eng* 16:1370–1386, 2004 doi:10.1109/TKDE.2004.68.
16. Khurd P, Liu B, Gindi G: Ideal AFROC and FROC observers. *IEEE Trans. Med. Imaging* 29(2):375–386, 2010.
17. Heyer LJ, Kruglyak S, Yooshep S: Exploring expression data: identification and analysis of coexpressed genes. *Genome Research* 9:1106–1115, 1999
18. Liu X, Xu X, Liu J, Feng Z: A new automatic method for mass detection in mammography with false positives reduction by supported vector machine. In: Ding Y, Peng Y, Shi R, Hao K, Wang, L. Eds. *BMEI: IEEE*, 2011, pp 33–37
19. Martins LO, Junior GB, Silva AC, de Paiva AC, Gattass, M: Detection of masses in digital mammograms using k-means and support vector machine. *Electronic Letters on Computer Vision and Image Analysis* 8(2):39–50, 2009 <http://elcvia.cvc.uab.es/article/view/216/235>.
20. Mini MG: Neural network based classification of digitized mammograms. In: *Proceedings of the Second Kuwait Conference on e-Services and e-Systems*. New York, NY, USA: KCESS '11 ACM, 2011, pp 2:1–2:5 doi:10.1145/2107556.2107558.
21. Nunes AP, Silva AC, de Paiva AC: Detection of masses in mammographic images using Simpson's diversity index in circular regions and SVM: *MLDM*, 2009, pp 540–553.
22. Nunes AP, Silva AC, Paiva ACD: Detection of masses in mammographic images using geometry, Simpson's diversity index and SVM. *Int J Signal Imaging Syst Eng* 3(1):43–51, 2010 doi:10.1504/IJSISE.2010.034631.
23. Oliveira Martins L, Silva A, de Paiva A, Gattass M: Detection of breast masses in mammogram images using growing neural gas algorithm and Ripley's k function. *J Signal Process Syst* 55(1-3):77–90, 2009 doi:10.1007/s11265-008-0209-3.
24. Oliveira Martins L, Silva EC, Silva A, Paiva A, Gattass M: Classification of breast masses in mammogram images using Ripley's k function and support vector machine. In: Perner P. Ed. *Machine Learning and Data Mining in Pattern Recognition, Lecture Notes in Computer Science: Springer, Berlin*. Vol. 4571, 2007, pp 784–794 doi:10.1007/978-3-540-73499-4_59.
25. Parkin DM, Bray F, Ferlay J, Pisani P: Global cancer statistics, 2002. *CA Cancer J Clin* 55(2):74–108, 2005 doi:10.3322/canjclin.55.2.74.
26. Rangayyan RM, Nguyen TM, Ayres FJ, Nandi AK: Effect of pixel resolution on texture features of breast masses in mammograms. *J. Digital Imaging* 23(5):547–553, 2010
27. Sampaio WB, Diniz EM, Silva AC, de Paiva AC, Gattass M: Detection of masses in mammogram images using CNN, geostatistic functions and SVM. *Comput Biol Med* 41(8):653–664, 2011 doi:10.1016/j.combiomed.2011.05.017, <http://www.sciencedirect.com/science/article/pii/S0010482511001132>.
28. Silva A, Carvalho P, Gattass M: Analysis of spatial variability using geostatistical functions for diagnosis of lung nodule in computerized tomography images. *Pattern Anal Applic* 7(3):227–234, 2004
29. da Silva Sousa JRF, Silva AC, de Paiva AC, Nunes RA: Methodology for automatic detection of lung nodules in computerized tomography images. *Comput Methods Prog Biomed* 98(1):1–4, 2010
30. Taylor PM, Champness J, Given-Wilson RM, Potts HWW, Johnston K: An evaluation of the impact of computer-based prompts on screen readers' interpretation of mammograms. *Br J Radiol* 77(913):21–27, 2004 doi:10.1259/bjr/34203805, <http://bjr.birjournals.org/content/77/913/21.full.pdf+html>.
31. Vidaurrazaga M, Diago LA, Cruz A: Contrast enhancement with wavelet transform in radiological images. In: *Engineering in Medicine and Biology Society*, 2000. In: *Proceedings of the 22nd Annual International Conference of the IEEE*, 2000, vol. 3, pp. 1760–1763 doi:10.1109/IEMBS.2000.900425.
32. WHO: The global burden of disease: 2004 update (2013). <http://www.who.int/en/>. Accessed August 2013.
33. Xu, R, Wunsch DI: Survey of clustering algorithms. *IEEE Trans Neural Netw* 16(3):645–678, 2005 doi:10.1109/TNN.2005.845141.