CAD System for Breast US Images with Speckle Noise Reduction and Bio-inspired Segmentation

Paulo S. S. Rodrigues and Guilherme A. Wachs Lopes
Computer Science Department
Centro Universitário FEI
São Bernardo do Campo, SP, Brazil
Email: {psergio,gwachs}@fei.edu.br

Gilson A. Giraldi
COMAC
National Laboratory for Scientific Computing
Petrópolis, RJ, Brazil
Email: gilson@lncc.br

Celia A. Z. Barcelos, Luciana Vieira, and Denise Guliato
Federal University of Uberlandia, UFU, FMC
Uberlandia, Minas Gerais, Brazil
Email:celiazb@ufu.br, lucianacomp@gmail.com,
guliato@ufu.br

Bikesh Kumar Singh
Department of Biomedical Engineering
National Institute of Technology Raipur
Raipur, India
Email: bsingh.bme@nitrr.ac.in

***Abstract*—Ultrasound (US) images are highly susceptible to speckle-like noise which makes imperative to use specific techniques for image smoothing. However, this process can lead to undesirable side effects such as the degradation of the real contour of the region of interest (ROI). In such context, this paper presents a new methodology for computer aided diagnosis (CAD) systems whose heart is the combination of a method for speckle noise reduction, with histogram equalization and a technique for image segmentation that uses the bio-inspired firefly algorithm and Bayesian model. The segmentation approach and the equalization are applied in two distinct stages: globally and locally. The global application produces an initial coarse estimate of the ROI, and the local application defines this region more precisely. In the classification step we carried out experiments which show that the combination of features computed both within and below the lesion strongly influences the final accuracy. We show that the gray-scale distribution and statistical moments within the lesion together with gray-scale distribution and contrast of the region below the lesion is the combination that produces the better classification results. Experiments in a database of 250 US images of breast anomalies (100 benign and 150 malignant) show that the proposed methodology reaches performance of 95%.**

I. Introduction

Breast cancer is the second leading cause of death, due to cancer diseases, in women. With the improvement of technology, computer aided diagnosis (CAD) systems, based on ultrasound (US) images, have contributed to early diagnosis. Generally speaking, these systems are composed of several stages that take as input an image of a specific region and provides as output a diagnostic prediction. In general, conventional CAD systems can be split into four main parts: preprocessing, segmentation, feature extraction and classification. CAD approaches based on deep neural networks have the advantage of reducing the burden of feature selection by learning a set of transformation functions that computes image features directly from the data [1].

All the mentioned stages are affected by speckle noise, very common in US images. This noise type arises due to the reverberation of the sound signal on the embedded surface and is composed of very small light and dark spot regions throughout the image. To reduce this problem, in the first stage, US images have been pre-processed using low-pass filters based on Gaussian functions, although they degrade the contrast [2]. One way to recover the contrast after noise elimination is through the use of histogram equalization which can be applied globally or locally [3]. However, in order to apply equalization in a specific region, it is necessary to know this region beforehand. Thus, it is usual to apply histogram equalization only globally. Still in the first stage, we find multiresolution strategies based on pyramid and wavelets [2], [4], fuzzy logic approaches [5], thresholding and mathematical morphology [2], [6].

Speckle noise interferes especially in segmentation which results in splitting of the region of interest (ROI) or its merge with the background. The ROI may be heterogeneous or almost homogeneous, depending on the type of lesion and its evolution. For example, malignant tumors tend to have several layers with different textures and widths. Hence, even with specific methods for reducing speckle noise, the final segmentation result may still be not adequate. The combination between multiresolution techniques, active contour models, like level sets, or graph cuts, is a way to increase the robustness [2], [7].

In all the mentioned works, we have noticed that global segmentation methods may not fully reveal the whole information embedded in a lesion's neighborhood. In order to address this challenge, in this paper we propose a method that applies image processing in two stages: initially in a global basis, as it is done in most of the projects studied in the literature, to get an estimate of the ROI through noise reduction, equalization, bio-inspired firefly algorithm and a Bayesian model. Then, these operations, except noise reduction, are applied again but now as local processes. We show that this iterative strategy defines the ROI more precisely. Finally, the lesion's features

are extracted and used as training data of a support vector machine (SVM) classifier with cross-validation technique.

The proposed method consists of nine steps: (i) Preprocessing with speckle noise reduction [8]; (ii) Histogram equalization; (iii) Global firefly segmentation [9]; (iv) Coarse segmentation of the ROI through Bayesian method [10]; (v) Local histogram equalization; (vi) Local firefly segmentation; (vii) Final Bayesian ROI definition; (viii) Feature extraction; (ix) SVM classification. This methodology is the main contribution of this paper. As other contributions, we analyze seven traditional features, extracted from both the lesion and its surrounding regions [11]. We investigate the contribution of each feature on classification performance to reduce the computational complexity by rejecting irrelevant features. It is found that gray-scale distribution and statistical moments within the lesion together with gray-scale distribution and contrast of the region below the lesion give the combination that achieved the best overall performance, which is competitive with the state-of-the-art in the area that is composed by deep learning strategies (see section II).

The results obtained on a base of 250 images of breast cancer, manually annotated by experts, indicate the efficiency of the proposed model that reaches a maximum performance of 95%, measured as the area under the ROC curve, constructed with the SVM classifier.

The remainder of the paper is organized as following. Section II describes related works. Section III reviews some background for our methodology. The proposed CAD system is then detailed in section IV. Next, section V describes the computational experiments. Then, in the final part of the paper, we discuss relevant findings in section VI and present conclusions/future works in section VII.

II. Related Works

In this section, we analyze related methods, supported by recent surveys [12], [13], highlighting their strengths, weaknesses and significant observations. The work [3] employed Markov Random Field (MRF) for detecting breast lesions in US images, feature extraction from segmented images followed by classification using fuzzy SVM. The technique achieved overall accuracy of 94.25% but using a small database comprising of 36 malignant and 51 benign cases.

Liu et al. [5] reported an automatic segmentation and a classification method that uses local texture features of breast US images. The image was divided into small lattices and local texture information was used for classifying image lattices using SVM. In [6] it is presented an automated cancer detection method for women with dense breast. Initially, it is applied adaptive thresholding and k-means clustering to detect tumor candidate regions. Then, 18 features are extracted from tumor candidates for characterization using Markov-Chain Monte Carlo Bayesian neural network. The method achieved median detection sensitivity of around 50%.

The topographic watershed technique was applied in [14] for segmentation and abnormality extraction in breast US images. The approach determined the probability of a segmented object being tumor region using features related to morphology, texture and intensity. In [2] we find a combination of Gaussian smoothing, histogram equalization, mean shift filtering and graph cuts for detecting breast lesions. This method achieved a true positive rate of 91.7% and false positive rate of 11.9% respectively. However, the data size used for evaluation was also limited in this study (31 malignant and 38 benign tumors).

In [15], Zhou et al. employed thresholding and disk expansion method for segmentation of breast lesions in US. Their technique extracted half contour features for discriminating sample groups with overall classification accuracy of 74%. In [4], the authors proposed multiresolution analysis for detecting breast tumors in US images. They calculated the undecimated discrete wavelet transform of the image followed by feature extraction and dimensionality reduction using principal component analysis. Lastly, they used Fuzzy c-means clustering for segmentation. This technique reported a Dice coefficient of 0.8595 in a test database with 150 breast ultrasound images, including 60 malignant and 90 benign cases. The work [7] transformed the image into neutrosophic domain followed by computation of neutrosophic similarity score (NSS), that is used as measure of belongingness to the true tumor. A level set method is finally used to segment the tumor using the NSS value in a database with 66 samples.

These reviewed approaches have some limitations: (1) The data size used in most of the studies is limited. Thus, it is difficult to evaluate the generalization ability of these techniques; (2) Except [15], none of the reported studies included images with artifacts such as acoustic shadowing and acoustic enhancement. (3) None of these studies evaluated the impact of pre-processing techniques such as noise reduction, enhancement etc. on the performance of segmentation and classification techniques; (4) It is also found that most of the reported techniques use only features extracted from region inside the lesion.

More recently, literature results have demonstrated the potential of deep learning methods to perform automatic US image analysis tasks such as lesion classification [13]. In this line, the work [16] reports a neural networks accuracy of 0.96, measured as the area under the ROC curve on a training set with 445 samples, using an image analysis software (ViDi Suite v. 2.0; ViDi Systems Inc, Villaz-Saint-Pierre, Switzerland). However, the performance on the validation set (192 samples) was 0.84. In [17] a biopsy-proven benchmarking dataset was built from 5151 cases containing a total of 7408 US breast images, with 4254 benign and 3154 malignant lesions. The developed method includes histogram equalization, image cropping and margin augmentation. The GoogLeNet convolutional neural network was trained using the database to differentiate benign and malignant tumors achieving area under the ROC curve of over 0.9 for training data with and without augmentation.

Our study has attempted to fill the research gaps mentioned in items (1)-(4) above related to traditional (non deep learning)

CAD systems. Regarding the deep learning based techniques, the goal is to get a CAD approach with proven generalization capability without the need of a huge training data sets, like the one used in [17]. The next sections describe how we achieve such goals.

III. Technical Background

Let $I : \Omega \subset \mathbb{R}^2 \rightarrow [0,1]$ the observed image which corresponds to another image $u : \Omega \subset \mathbb{R}^2 \rightarrow [0,1]$ corrupted by speckle noise. The variational approach presented in [8] for image smoothing considers the functional:

$$F(u,I) = gJ(u) + \lambda(1-g)F_{fidel}(u,I), \quad u \in C^2(\Omega), \tag{1}$$

where $C^2(\Omega)$ is a space of two times continuously differentiable functions, $J : C^2(\Omega) \rightarrow \mathbb{R}$ is a smoothing term, $F_{fidel} : C^2(\Omega) \times C^1(\Omega) \rightarrow \mathbb{R}$ is the fidelity term, which establishes a compromise between the observed image I and the smoothed one u, and λ is a constant. The function g is a smooth non-increasing function, such that $0 \leq g \leq 1.0$ and $g(s) \longrightarrow 0$ when $s \rightarrow \infty$.

The procedure proposed in [8] tackles the speckle noise reduction by solving the problem:

$$\min_{u \in C^2(\Omega)} F(u,I). \tag{2}$$

The functionals J and F_{fidel} are defined, respectively, by:

$$J(u) = \int_\Omega \|\nabla u(x,y)\|^2 \, dx \; dy,$$

$$F_{fidel}(u,I) = \int_\Omega \frac{(I-u)^2}{u} dx \; dy,$$

where ∇u is the gradient operator and $\|\cdot\|$ means norm-2. The usual choice for the function g in expression (1), imported from the anisotropic diffusion literature, is given by:

$$g(\|\nabla I(x,y)\|) = \frac{1}{1 + k\|\nabla G_\sigma I(x,y)\|^2}, \tag{3}$$

where G_σ is a Gaussian function with variance σ.

The numerical solution for problem (2), is obtained through an iterative scheme yielded from the associated Euler-Lagrange equations and traditional finite difference techniques [8].

Regarding the firefly algorithm presented in [9], its background is the non-extensive Tsallis entropy (S_q), that is the generalization of Shannon's entropy (S_1), given, respectively by:

$$S_q = \frac{1 - \sum_{i=0}^{L-1} p_i^q}{q-1}, \qquad S_1 = -\sum_{i=0}^{L-1} p_i \log(p_i), \tag{4}$$

with p_i being the relative frequency of intensities $i = 0, 1, 2, ..., L-1$, and L is the number of gray-levels of the image field.

We can show that Tsallis entropy reduces to the Shannon's one when q tends to 1. The main characteristic observed in the Tsallis statistic is the use of a real parameter q, called the non-extensive parameter. One property of Tsallis entropy, called pseudo-additivity, is formulated as:

$$S_q(A \oplus B) = S_q(A) + S_q(B) + (1-q)S_q(A)S_q(B), \tag{5}$$

where $S_q(A \oplus B)$ is the global entropy of the system $A \oplus B$ and $S_q(A)$ and $S_q(B)$ are the individual entropies for subsystems A and B, respectively.

The focused firefly algorithm [9] uses an objective function given by expression (5). The general idea behind the firefly algorithm for multi-segmentation of images is to associate to each firefly a possible solution to the problem of partitioning the gray-scale distribution of the image (histogram). Thus, considering a swarm of fireflies with random guesses, the best solution is associated with firefly brightness and it is found through an attractiveness process between them. The kernel of the algorithm is its evaluation function Z, which depends on the current problem. Each firefly is considered a d-dimensional variable, where each dimension is a distinct threshold, partitioning the histogram space. The d-level firefly solution is a set of d thresholds that defines $d+1$ regions, possibly disconnected, in the image. In the firefly implementation described in [9], the setup includes the following parameters: absorption coefficient (γ); step of motion (α); attractivity factor (β). The histogram equalization applied uses a traditional implementation of Matlab.

IV. Proposed Methodology

As described in the introduction, the methodology proposed in this paper contains nine steps, three of them (equalization, bio-inspired firefly algorithm and Bayesian model) are applied twice at different times: globally and locally. The whole pipeline is: (i) Pre-processing with speckle noise reduction [8]; (ii) Histogram equalization; (iii) Global firefly segmentation [9]; (iv) Coarse segmentation of the ROI through Bayesian method [10]; (v) Local histogram equalization; (vi) Local firefly segmentation; (vii) Final Bayesian ROI definition; (viii) Feature extraction; (ix) SVM classification.

Each of these nine steps is detailed below. The presentation is divided into two major blocks: ROI segmentation and SVM classification.

A. ROI Segmentation

As previously stated, the US images are severely affected by the speckle noise as observed in the Figure 1.(a). Because of this, in the first step we apply the noise reduction method formulated through equation (2). By using the function in expression (3), the smooth process becomes stronger than fidelity in regions containing weaker edges and in parts of the image with stronger edges the opposite occurs. Figure 1.(b) shows the solution u of problem (2) where the image I is the one in Figure 1.(a).

Despite of the desired image smoothing, we also notice in Figure 1.(b) the side effect of contrast reduction which in turn influences the ROI extraction. This unwanted effect is mitigated through the use of global histogram equalization process. The result can be visualized in Figure 1.(c), which

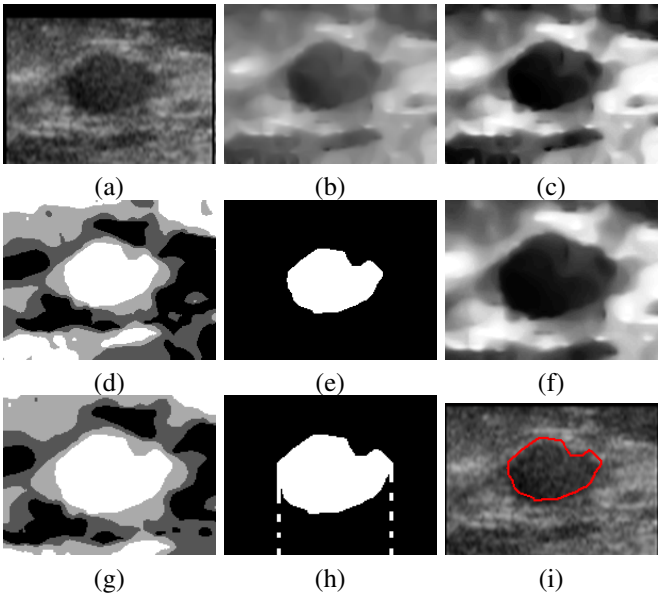

Fig. 1. (a) Original Image. (b) Speckle noise reduction. (c) Global histogram equalization. (d) Global firefly segmentation. (e) Mask from the Bayesian model. (f) Local histogram equalization inside window around the mask shown in (e). (g) Local firefly segmentation. (h) Final ROI and boundary of region bellow it. (i) Boundary of ROI over the original image.

shows improvement of the global contrast although local details may be lost because all images regions influence one's each others. Notwithstanding, we can apply the segmentation approach, the firefly algorithm (section III), to get a coarse ROI segmentation, as shown in Figure 1.(d).

The decision to use multi-segmentation instead of the traditional binary segmentation is motivated by the fact that the US of breast images have ROIs in a broad intensity range that can not be segmented through only one threshold. For example, images of malignant lesions are frequently heterogeneous, containing lighter shades around the center and darker shades on the boundaries. Thus, when labeling the lesions interior, the use of more than one threshold makes easier the extraction and analysis of the whole lesion. Thus, a process of multi-segmentation is more suitable. However, this process may generate multiple regions which must be subsequently evaluated in order to achieve which of them is the ROI. In fact, in Figure 1.(d), the multi-segmentation firefly algorithm generates regions with white intensity, light gray, dark gray and black tones.

In the proposed methodology, all regions labeled white as output of segmentation step are called candidate regions and are processed in the next stage by estimating the joint probability of size and location. In this research, an important fact is observed for each database image W. Each candidate region R_c, generated in the segmentation, occurs near the central point WC of the image W, which in turn is a cropping window defined by a specialist at clinical inspection time. The size of W must be enough to include the lesion and a certain amount of neighboring pixels. Thus, it is expected that the lesion occupies most part of the central area inside the W when compared to all other candidate regions around. Therefore, we define $P(R_c|A,C)$ as the probability of a candidate region R_c to be a ROI, while its area A is observed with the centroid C, generating the following expression [10]:

$$P(R_c|A,C) = 1 - ((1 - P(R_c|A)) \times (1 - P(R_c|C))), \quad (6)$$

where the conditional probability $P(R_c|A)$ is defined as the ratio between the observed area A and the area of W and the term $P(R_c|C)$ is defined as the ratio between the Euclidean distance between C and WC and half of the diagonal of the image W. Finally, we define the coarse ROI within the image W as the one that maximizes the observed $P(R_c|A,C)$ among all candidate regions.

Figure 1.(e) shows the result of such ROI computed with Figure 1.(d) as input. This ROI estimation is used as a mask to a further new local processing. This strategy is used to allow only a region around the lesion to influence the contrast enhancement, highlighting local details that could not be noticed under global equalization and segmentation. The Figure 1.(f) shows the final effect of such histogram equalization, re-scaled to the original patch size, applied inside a bounding box of the region shown in Figure 1.(e).

Then, the bounding box of the mask shown in 1.(e) can also be used to locally apply the firefly segmentation process over the image in Figure 1.(f) , given the result of Figure 1.(g). As before, the output of the local segmentation processing is a set of candidate regions, but with more accuracy. Each one of these regions is then submitted to the same joint probability model described by equation (6). Finally, the region most likely to represent the final ROI is taken, which ends step (vii). Figure 1.(h) shows this final ROI and the boundary (dashed lines) of the region bellow it that will be used to compute image features outside the ROI. The boundary of the ROI is pictured in Figure 1.(i) in order to allow a visual evaluation of the final result.

B. SVM Classification

In the step (viii), we use the ROI achieved in step (vii) as a mask over the corresponding original image to extract the seven lesion features reported in Table I.

TABLE I
LABELS FOR THE SEVEN BREAST LESION FEATURES USED IN OUR EXPERIMENTS.

Label	Feature
F1	Circularity
F2	Gray-scale distribution bellow lesion
F3	Gray-scale distribution within lesion
F4	Contrast below lesion
F5	Contrast within lesion
F6	Statistical moments below lesion
F7	Statistical moments within lesion

To compute the circularity we take a set of uniformly distributed points in the mask boundary and compute the 10-bins distribution of their distance respect to the ROI centroid. We apply the Shannon's entropy, computed by S_1 in equation (4), to get a measure of the contrast of a target region.

The statistical moments associated to a ROI are computed by:

$$\mu_n = \sum_{i=0}^{L-1} (i-m)^n \, p_i, \tag{7}$$

where m is the mean intensity inside the ROI and $n = 0, 1, .., 5$ (analogous for the region bellow the ROI, whose mask is represented in Figure 1.h)

The circularity means how close the shape is to a circle. Once benign lesions generally have more circular areas compared with the malignant ones, this quantity can be a good discriminant. The feature F2 is directly linked to acoustic shadow and acoustic enhancement. The former is a visual characteristic common in benign areas while the latter occurs in malignant lesions.

The features F4 and F5 link the contrast with the homogeneity of the intensity patterns of the lesion. Malignant lesions tend to be less homogeneous than benign ones indicating less organization, or, more formally, a histogram closer to the uniform distribution than benign tumors. Therefore, as higher entropy is then more heterogeneous is the lesion region and, consequently, the chance to be a malignant lesion is higher. Features F6 and F7 are histogram features, computed by equation (7).

The feature vectors generated are used as input for a SVM classifier that finally defines the class to which each lesion belongs to, in this case, malignant or benign. The SVM model is the non-separable one [18], available in the Matlab, with linear kernel and relaxation parameter $C = 0.8$.

V. Experimental Results

This section shows the experimental results regarding the methodology presented in section IV. Specifically, we want to demonstrate the impact of each stage of the proposed pipeline in the final SVM classifier and the sensitivity of the methodology under combinations of the seven lesion features described in Table I.

The database used is available in [19] and it consists of 250 gray level (8-bits) US images with low pixel resolution, being 150 malignant lesions and 100 benign. Each image is a cropping window made by experts around the breast lesion. The high resolution US images were acquired by the Voluson 730 (General Electric, USA) scanner with a Voluson small part transducer S-VNW5-10.

Firstly, the performance of speckle noise reduction approach, presented in section III, is evaluated. Figures 2.(a),(c) show clearly the phenomenon of acoustic shadow (darker area below central lesion) and reinforcement (brighter area below lesion), respectively. In the second row, this phenomenon is highlighted after the application of speckle noise reduction strategy. These results of noise reduction are achieved with 300 iterations of the numerical scheme developed to solve problem (2) (see [8], for details). When the global histogram equalization is applied at every image of the second row of Figure 2, the results can be seen in Figures 2.(g)-(i). The histogram equalization improves the contrast inside regions with narrow range of gray-scale patterns.

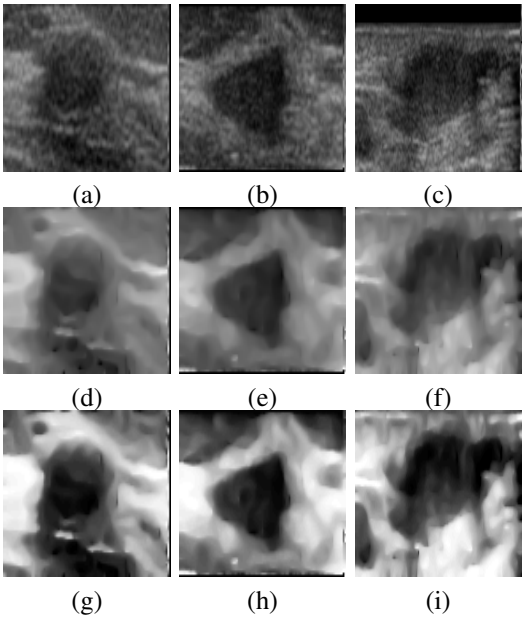

Fig. 2. (a)-(c) Example of three US input images. (d)-(f) Speckle noise reduction outputs. (g)-(i) Results after histogram equalization of smoothed images.

The firefly segmentation results, for two threshold levels, can be evaluated in Figure 3. The images in Figures 3.(a)-(c) show the results with the original images shown in Figures 2.(a)-(c), respectively, as inputs. The achieved thresholds are shown below each image. The second row (Figures 3.(d)-(f)) pictures analogous results when both noise reduction and histogram equalization were applied before segmentation, showing the improvement in the bio-inspired segmentation due to the pre-processing steps.

The firefly segmentation was running with the same parametrization proposed in [9]: $\beta = 1.0$, $\gamma = 1.0$, $\alpha = 0.97$, $q = 0.35$, using 50 fireflies and 100 iterations. With the multi-segmented images as input, each ROI is defined by the Bayesian model formulated by expression (6). Figures 3.(g)-(i) show the obtained ROIs.

The influence of denoising and histogram equalization on the ROI extraction can be visually analyzed in the following example. Figure 4-(a) shows an image from the database. This image was processed using the proposed methodology but in three different ways: (A) Without speckle noise reduction (Figures 4.(a)-(c)-(e)-(i)); (B) Removing the global histogram equalization and noise reduction (Figures 4.(a)-(f)-(j)); (C) Using all the nine steps (Figures 4.(a)-(b)-(d)-(g)-(k)). The parameters for firefly heuristic were the same as those used to obtain the results of Figure 3.

A visual inspection shows that there may be greater impact on the results after the inclusion of each stage, especially over histogram equalization with speckle noise reduction as pre-processing. Notably, without noise reduction, the segmentation process produced non-smooth boundary and wrong ROI topol-

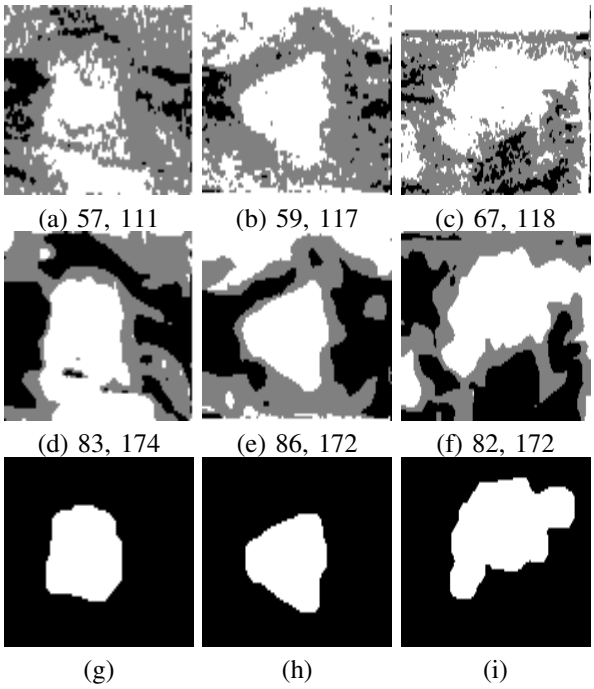

Fig. 3. Three examples of two level segmentation outputs with obtained thresholds. (a)-(c) Segmentations without pre-processing. (d)-(f) Results after speckle noise reduction followed by histogram equalization. (g)-(i) The final ROIs achieved through the maximization of expression (6) over images (d)-(f).

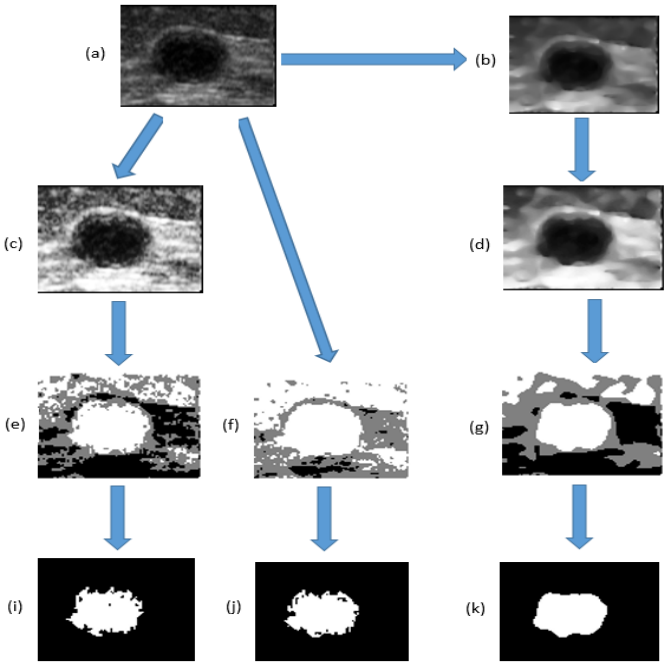

Fig. 4. (a)-(c)-(e)-(i) Segmentation without noise reduction. (a)-(f)-(j) ROI extraction without global histogram equalization and noise reduction. (a)-(b)-(d)-(g)-(k) Segmentation with the whole pipeline.

ogy, as observed by comparing Figures 4-(i)-(j) with Figure 4-(k).

When the final result produces no region of interest, or the final region can not be considered a lesion region (such as a noise or a small region around the ROI), we call such result as a "degenerate case". These cases degrade the accuracy of SVM classifier. In order to reduce the number of degenerate cases, steps (v), (vi) and (vii) are each one a repetition of steps (ii), (iii) and (iv), respectively, but now as local processing. The strategy here is to increase the impact of segmentation and equalization operations, giving more accuracy to Bayesian classifier in the selection of the final ROI.

To quantify the influence of global operations, a general quantitative result was achieved when 250 images from the database were processed. When it was not applied any speckle noise reduction neither histogram equalization (see example in Figure 4-(j)) we noted $25/250$ (10.0%) degenerate cases. However, when it was included global histogram equalization (see example in Figure 4-(i)), the total degenerated cases decreased to $5/250(2.0\%)$, giving a clear improvement of 80% compared to the case without any pre-processing. On the other hand, when we did not apply the global histogram equalization but only the speckle noise reduction we noted $12/250(4.8\%)$ degenerate cases, an improvement of 5.2% considering only original image. Finally, when we applied the speckle noise reduction combined with global histogram equalization, we observed no degenerated cases (Figure 4-(k) is an example). These results show that the speckle noise reduction combined with histogram equalization is effective to minimize the number of degenerated cases in the output of the Bayesian model for ROI extraction.

The next tests were carried out with four different configuration scenarios: (i) Original images without any pre-processing (called scenario 1, labeled $SC1$); (ii) global histogram equalization alone (called scenario 2, labeled $SC2$); (iii) speckle noise reduction alone (called scenario 3, labeled $SC3$), and (iv) speckle noise reduction and histogram equalization together (called scenario 4, labeled $SC4$).

The ROI obtained in step (vii) of the methodology is used as a mask over the original image to compute the features of Table I in the step (viii). Those features were combined one by one, two by two, $\cdots$, six by six, and all of them together generating 127 possible features combinations. Thus, 127 feature spaces are yielded and each of them was tested under all four described scenarios SC1, SC2, SC3, and SC4. So, for each one of the 250 images from the database we calculate 127 feature vectors. Therefore, we build 4×127 databases of feature vectors, each one composed by 250 feature arrays.

In the final step of the proposed methodology, each database is then submitted to a classification process with a SVM implementation made in Matlab. In this paper, following related works [16], we use 5-fold cross-validation technique with 200 vectors for training and 50 for tests. So, we train the SVM using 4-folds and use the remaining one to compute the true positive and false positive rates. We repeat this procedure

a number of times to generate the whole ROC curve [20].

Table II shows the comparative performance results, computed as A_z area under the SVM ROC curve computed for the four proposed scenarios. The first column reports the feature space dimension, the other columns provide the best classification performance and the corresponding feature space for each scenario.

TABLE II
BEST PERFORMANCE, COMPUTED AS A_z AREA UNDER THE SVM ROC CURVE, AND CORRESPONDING FEATURE SPACE FOR EACH SCENARIO. IN BLACK, THE MAXIMUM AREA/FEATURE SPACE FOR EACH LINE.

Dim.	SC1	SC2	SC3	SC4
1	0.85 2	0.91 3	0.88 2 *or* 3	**0.93** **3**
2	0.910 24	0.915 35	0.912 36	**0.942** **36**
3	0.903 124	0.912 345	0.926 234	**0.948** **356**
4	0.904 1245	0.912 3456 *or* 3567	0.925 1236	**0.950** **2347**
5	0.910 12345	0.905 34567	0.930 12345	**0.948** **23457**
6	0.903 123457	0.892 123457	0.925 123456	**0.948** **134567**

For instance, the second line of Table II, with Dim. equal to 1, shows that the best performance (0.93) when only one feature is used in the proposed CAD system is achieved with the $F3$ feature (gray-scale distribution within lesion) used in scenario $SC4$.

The third line of Table II shows the results when the seven features are combined in pairs. In this experiment, the best overall performance (0.942) was achieved in $SC4$ scenario with the combination of features $F3$ (gray-scale distribution within lesion) and $F6$ (invariant moments bellow lesion), which is higher than the best previous achieved, suggesting that the combination of two characteristics improves the proposed method.

The fourth line of Table II reports the result for three-dimensional feature spaces generated by the combination of the 7 features taken 3 at a time. In this case, the scenario $SC4$ with the features $F3F5F6$ gives the best performance of 0.948.

The results for experiments using four-dimensional features spaces are shown in fifth line of Table II. Like the already reported tests, the scenario $SC4$ achieves the best performance. The best feature space in this case is $F2F3F4F7$ and the obtained accuracy (0.950) is higher than in the previous experiments.

Table II shows in the sixth line the experimental results when the feature vectors are combinations of five characteristics. In this experiment, again the $SC4$ scenario reached the best global performance (0.948), with the combination of $F2F3F4F5F7$ features.

In turn, seventh line of Table II shows the comparative performance results when the features are combined six by six. In this experiment, the best performance among all was again the $SC4$ scenario, which achieved the best overall performance (0.948), with the combination $F1F3F4F5F6F7$.

When all seven characteristics ($F1F2F3F4F5F6F7$) are used together, the best overall performance (0.941) is also for the $SC4$ scenario, followed by scenarios $SC3$, $SC2$ and $SC1$, in this order.

VI. DISCUSSION

The computational experiments consider four different configuration scenarios: $SC1$, $SC2$, $SC3$ and $SC4$. Clearly, the $SC4$ was the one with the best overall performance among all, demonstrating that the speckle noise reduction process combined with the histogram equalization algorithm improved the proposed CAD system.

Regarding to firefly segmentation algorithm, we note in Figure 4 that this algorithm is strongly affected by noise reduction and the equalization stages reducing the number of degenerate cases from 25 in $SC1$ scenario to zero in $SC4$ scenario. This difference of degenerate cases in favor of $SC4$ scenario allows a better assessment of the ROIs and is responsible for the SVM better performance in the latter stage of the proposed CAD system. The Figures 4.(e)-(f) show that without both noise reduction and equalization the firefly tends to generate over-segmentation yielding many regions.

The firefly+Bayesian model efficiency can be measured by the SVM classification rate. In fact, the performance of 95% is a strong indication of the segmentation efficiency.

On the other hand, the results reported in Table II allow to compare the efficiency of each feature alone. From the second line it can be seen that the best one is feature $F3$ in scenario $SC4$. The feature $F2$ achieves the second place with superior performance in $SC1$ and $SC3$. This indicates that the gray-scale distribution inside the region lesion was the most discriminant feature followed by the gray-scale distribution below the lesion. However, by comparing the second line with the other ones in Table II, we can notice that, in general, the use of only a single feature gives the worst overall performance.

At the other extreme, the combination of features that led to best overall performance (0.95, Table II) was $F2F3F4F7$ in $SC4$ scenario. Therefore, the use of features F1, F5 and F6 does not add discriminant information since it was not verified an increase in the system efficiency.

This work addresses the research gaps (1)-(4) mentioned in section II for non deep learning CAD systems. In fact, respect to gap (1), we shall highlight that our database has 250 US breast cancer images that is a number considerably larger than the mentioned in [3], [4], [7]. Also, it includes some cases with acoustic shadow and acoustic reinforcement (Figure 2.(a),(c)) to demonstrate the capabilities of our segmentation method against such artifacts (gap (2)). Moreover, we evaluated the consequences of our pre-processing techniques on the overall system performance (see Figure 4 and Table II), addressing gap (3). We have observed that the mentioned artifacts are also suggestive of lesion characteristics. So, we introduce features F2, F4, and F6 to take into account this fact. The obtained

results show that such strategy effectively helps efficiency of the method which addresses gap (4).

In the proposed pipeline there is no assumptions related to probabilistic models like in [3], or the need of large training datasets like in deep learning approaches [13]. Different from our technique, the method proposed in [14], that attained maximum sensitivity of 100%, may not be suited to images corrupted by acoustic shadow and acoustic enhancement, since these characteristics may distort the local texture of the lesions. Regarding the work [2], though computationally faster, this technique is sensitive to blurring resulting in unclear tumor contour. The application of the noise reduction method proposed in [8] and histogram equalization avoids blurring problems inside our methodology. Although each tested image is a cropping window holding the ROI we can notice in the Figures 1.(a) and 2.(a)-(c) that they present other tissues besides the ROI. So, the task is more than to delineate the ROI boundary because the methodology should extract the lesion from its surrounding background (see section VII).

Up to the best of our knowledge, the maximum deep learning performance reported in the literature (see the penultimate paragraph of section II) for breast US classification is 0.96 [16]. Our methodology gets a performance of 0.95 which is close to 0.96, without problems regarding transfer learning and data augmentation that usually deep learning methods undergo when only a small data set is available for training (see [21], section 2.2). Hence, we believe that we have achieved our goal of getting an efficient CAD approach without the need of huge training data sets.

VII. Conclusions and Future Works

This paper proposes a CAD system for automatic diagnosis of US images of breast lesions and evaluates its performance using a test database with 100 images of benign and 150 of malignant lesions. The average area under the ROC curve is used to measure the SVM classifier performance. The results highlighted the importance of all stages of proposed pipeline behind the CAD system. In particular, the speckle noise reduction followed by a histogram equalization strategy allows the generation of better lesion boundaries in the subsequent stages (Figure 4).

To prove the efficiency of the method, the SVM classifier is used with 250 US images represented by seven lesion features (Table I) showing that the combination of $F2F3F4F7$ features in $SC4$ scenario gives the best classification result of 95.00%. In further works, we could apply the course segmentation method proposed in [5] to compute the cropping window for a high resolution US image. Moreover, with the goal of comparison, we intend to employ the proposed method on datasets used elsewhere as well as to apply related methods on the same dataset used in this paper.

References

[1] J. S. Choi, B. K. Han, E. S. Ko, J. M. Bae, E. Y. Ko, S. H. Song, M. R. Kwon, J. H. Shin, and S. Y. Hahn, "Effect of a deep learning framework-based computer-aided diagnosis system on the diagnostic performance of radiologists in differentiating between malignant and benign masses on breast ultrasonography," *Korean journal of radiology*, vol. 20, no. 5, pp. 749–758, 2019.

[2] Z. Zhou, W. Wu, S. Wu, P. H. Tsui, C. C. Lin, L. Zhang, and T. Wang, "Semi-automatic breast ultrasound image segmentation based on mean shift and graph cuts," *Ultrasonic Imaging*, pp. 1–21, 2014.

[3] X. Shi, Cheng, H. D., L. Hu, W. Ju, and J. Tian, "Detection and classification of masses in breast ultrasound images," *Digital Signal Processing*, no. 3, pp. 824–836, 2010.

[4] K. M. Prabusankarlal, T. P., and R. Manavalan, "Segmentation of breast lesions in ultrasound images through multiresolution analysis using undecimated discrete wavelet transform," *Ultrasonic Imaging*, pp. 1–19, 2015.

[5] L. B., H. D. Cheng, J. Huang, J. Tian, X. Tang, and J. Liu, "Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images," *Pattern Recognition*, no. 1, pp. 280–298, 2010.

[6] K. Drukker, C. A. Sennett, and M. L. Giger, "Computerized detection of breast cancer on automated breast ultrasound imaging of women with dense breast," *Medical Physics*, no. 1, pp. 1–19, 2014.

[7] Y. Guo, A. Sengur, and J.-W. Tian, "A novel breast ultrasound image segmentation algorithm based on neutrosophic similarity score and level set," *Computer Methods and Programs in Biomedicine*, vol. 123, pp. 43 – 53, 2016.

[8] C. Barcelos and L. Vieira, "Ultrasound speckle noise reduction via an adaptive edge-controlled variational method," in *Proceedings of International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2014.

[9] P. Rodrigues, G. Wachs-Lopes, G. Giraldi, H. Erdmann, and M. Ribeiro, "Improving a firefly meta-heuristic for multilevel image segmentation using tsallis entropy," *Pattern Analysis and Application*, vol. 18, no. 1, pp. 1–22, 2015.

[10] B. Ribeiro-Neto, I. Silva, and R. Muntz, "Bayesian network models for IR," in *Soft Computing in Information Retrieval Techniques and Applications*, F. Crestani and G. Pasi, Eds. Springer Verlag, 2000, pp. 259–291.

[11] P. S. Rodrigues, G. A. Giraldi, M. Provenzano, M. D. Faria, R. Chang, and J. S. Suri, "A new methodology based on q-entropy for breast lesion classification in 3-d ultrasound images," in *2006 Int. Conf. of the IEEE Engin. in Medicine and Biology Society*, Aug 2006, pp. 1048–1051.

[12] M. Xian, Y. Zhang, H.-D. Cheng, F. Xu, B. Zhang, and J. Ding, "Automatic breast ultrasound image segmentation: A survey," *Pattern Recognition*, vol. 79, pp. 340–355, 2018.

[13] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261 – 275, 2019.

[14] C. M. Lo, R. T. Chen, Y. C. Chang, Y. W. Yang, M. J. Hung, C. S. Huang, and R. F. Chang, "Multi-dimensional tumor detection in automated whole breast ultrasound using topographic watershed," *IEEE Trans. on Medical Imaging*, no. 7, pp. 1503–1511, 2014.

[15] Z. Zhou, S. Wu, K. J. Chang, W. R. Chen, Y. S. Chen, W. H. Kuo, C. C. Lin, and P. H. Tsui, "Classification of benign and malignant breast tumors in ultrasound images with posterior acoustic shadowing using half-contour features," *J. Med. Biol. Eng.*, pp. 178–187, 2015.

[16] A. S. Becker, M. P. Mueller, E. Stoffel, M. Marcon, S. Ghafoor, and A. Boss, "Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study." *The British journal of radiology*, vol. 91 1083, pp. 1–8, 2017.

[17] S. Han, H.-K. Kang, J.-Y. Jeong, M.-H. Park, W. Kim, W.-C. Bang, and Y.-K. Seong, "A deep learning framework for supporting the classification of breast lesions in ultrasound images," *Physics in Medicine & Biology*, vol. 62, no. 19, pp. 7714–7728, sep 2017.

[18] V. N. Vapnik, *Statistical Learning Theory*. John Wiley & Sons, INC., 1998.

[19] Database, "Download from Dropbox url," https://goo.gl/ld5dFO.

[20] T. Fawcett, "An introduction to roc analysis," *Pattern Recogn. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[21] A. Garcia-Garcia, S. Orts, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. G. Rodríguez, "A survey on deep learning techniques for image and video semantic segmentation," *Appl. Soft Comput.*, vol. 70, pp. 41–65, 2018.