Research Article

Application of Artificial Neural Network Models in Segmentation and Classification of Nodules in Breast Ultrasound Digital Images

Karem D. Marcomini, 1 Antonio A. O. Carneiro, 2 and Homero Schiabel 1

1 Department of Electrical and Computer Engineering, University of São Paulo, 400 Avenida Trabalhador São-Carlense, 13566-590 São Carlos, SP, Brazil
2 Department of Physics, University of São Paulo, 3900 Avenida Bandeirantes, 14040-901 Ribeirão Preto, SP, Brazil

Correspondence should be addressed to Karem D. Marcomini; karem.dm@usp.br

Received 1 March 2016; Accepted 17 May 2016

Academic Editor: Sonal Jain

Copyright © 2016 Karem D. Marcomini et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This research presents a methodology for the automatic detection and characterization of breast sonographic findings. We performed the tests in ultrasound images obtained from breast phantoms made of tissue mimicking material. When the results were considerable, we applied the same techniques to clinical examinations. The process was started employing preprocessing (Wiener filter, equalization, and median filter) to minimize noise. Then, five segmentation techniques were investigated to determine the most concise representation of the lesion contour, enabling us to consider the neural network SOM as the most relevant. After the delimitation of the object, the most expressive features were defined to the morphological description of the finding, generating the input data to the neural Multilayer Perceptron (MLP) classifier. The accuracy achieved during training with simulated images was 94.2%, producing an AUC of 0.92. To evaluating the data generalization, the classification was performed with a group of unknown images to the system, both to simulators and to clinical trials, resulting in an accuracy of 90% and 81%, respectively. The proposed classifier proved to be an important tool for the diagnosis in breast ultrasound.

1. Introduction

According to the American Cancer Society, more than 178,000 women are affected by breast cancer every year; international statistics reports estimated 1,152,161 new cases annually. This form of the disease is the leading killer of women between 40 and 55 years old and is the second leading cause of death overall in women [1]. Due to this, screening techniques allowing early detection and diagnosis have been studied in order to increase the chances of survival using less aggressive treatment [2, 3].

Among the screening techniques currently available, mammography is the most often used, considered as the gold standard to breast tumor detection. However, this procedure is less effective when investigating dense breasts due to relatively high false negative rates [1]. Moreover, the number of unnecessary biopsies is very large and can lead to changes in the parenchyma making it difficult to read subsequent mammographic images [2].

In recent years, ultrasonography has proven a valuable technique used as an adjunct to conventional mammography for the detection and classification of breast lesions [1]. This procedure has been used to obtain additional diagnostic information, in order to reduce the number of unnecessary biopsies and assist with more accuracy the diagnosis of simple cysts (around 96–100% of efficacy when both of these techniques are used together) [4]. An additional advantage of ultrasound is that it does not use ionizing radiation and therefore is useful especially for younger patients who tend to have dense breasts [5].

The detection of abnormalities in medical images is a procedure prone to errors, even for qualified radiologists, due to the subjectivity in defining boundaries, overlap between benign and malignant characteristics, and the presence of...
artifacts that may confuse the diagnosis [6]. In order to increase diagnostic accuracy and minimize such errors, computational tools have been developed to provide a second opinion for the specialist and assist in early detection of breast cancer. In this context, this work aims to develop a tool to aid the diagnosis based on the automatic detection of lesions in ultrasound images and the consequent classification of such finding as clinically suspicious (malignant) or not (benign) considering an analysis of their morphological characteristics.

2. Materials and Methods

The database used in this research consisted of two distinct sets of breast ultrasound images. The first corresponds to images from breast phantoms and the second from conventional clinical examinations.

2.1. Database. The phantom images were acquired from tests performed by the Group of Innovation in Medical Instrumentation and Ultrasound (GIIMUS) from University of São Paulo, Brazil. The phantoms used in this research were BB-1 model (breast biopsy phantom, ATS Laboratories) and models previously developed by Vieira et al. [7]. All the phantoms were made of an acoustically tissue mimicking material and have a shape similar to the breast of an adult woman. Figure 1 illustrates few examples of these phantoms.

All these phantoms were submitted to ultrasound beams from a GE Logic-Book XP portable device, operating in the frequency range of 1–10 MHz. A total of 144 phantom images in B-mode were acquired as those illustrated in Figure 2.

Two medical centers of imaging diagnosis at São Carlos, SP, Brazil, provided clinical ultrasound images. Four devices were used in this process of obtaining these images, Siemens G50, Medison X8, Toshiba Nemio 30, and General Electric Logiq P5, considering a broadband linear transducer of 7.5–10 MHz frequency range. A total of 123 images were acquired during imaging routine procedures. Figure 3 shows some examples of B-mode ultrasound images acquired with such different equipment.

3. Images Processing

For each image, an experienced radiologist performed the analysis in order to detect lesions with suspicious appearance and then selected the regions of interest (ROIs). These ROIs had rectangular shape and included the lesion and surrounding tissue. Therefore, such a procedure resulted in 173 ROIs corresponding to actual clinical ultrasound images and 144 from phantoms images.

In order to remove noise and to smooth the image components, ROIs have been preprocessed by a Wiener filter, followed by the contrast enhancement (image equalization) and median filter.

Based on the variability in segmentation techniques, the efficacy was evaluated taking into account some techniques applied in order to highlight the lesion: active contour [14], region growing [15], fuzzy $c$-means [16], $k$-means [17], and SOM neural network [18, 19].

A postprocessing technique was then applied aiming to improve the segmentation quality since many pixels were verified to be disconnected from the actual lesion after the use of some of the techniques mentioned above. As a consequence, this effect has produced a more spiculated and noisier appearance than the nodule actually had. In addition, internal valleys were identified. Therefore, artifacts disconnected from the object of interest were eliminated and the internal valleys have joined the region [20].

4. Features Extraction and Selection

Feature extraction in digital images is a critical step for identifying objects. In most cases, the use of more than one measure is required in order to decide to which class the pattern belongs. The most common is to extract from each sample several measures and then represent them through a vector, which will serve as an input to the classifier [21].

In general, benign tumors correspond to softer shapes and malignant tumors tend to have irregular edges [10]. Thus, based on the shape of the lesion as previously reported [22], 24 morphological features were extracted from each one.
Figure 2: Ultrasound images obtained from exposures of the phantoms shown in Figure 1.

Figure 3: Breast clinical B-mode image acquired from (a) General Electric Logiq P5, (b) Medison X8, (c) Siemens G50, and (d) Toshiba Nemio 30.
However, due to the large number of features considered, Gaussian distribution curves [21] were used regarding this features set optimization, so that it could accurately describe the identified object. In this procedure, each descriptor is normalized and Gaussian curves are generated based on the distribution of values presented for each class (or just two for this work’s purposes). The analysis of these curves is performed by visual inspection of some details. The most important is the level of curves intersection: the smaller its occurrence, the higher the probability of such feature representing such a category. Furthermore, the distribution range of the values in the abscissa axis must be checked in order to determine the optimum distance between the classes. Figure 4 shows some examples of these curves.

5. Classification

Multilayer Perceptron (MLP) neural network is a tool frequently used in differentiating between benign and malignant lesions. Its topology consists of sensory units which include the input layer, one or more hidden layers (also known as intermediary), and an output layer [23]. The learning process is supervised; that is, the desired outputs are required. A supervised learning algorithm analyzes, through comparative
actions between inputs and the desired output, the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unknown instances.

After performing comparisons in the learning method (the backpropagation), the synaptic weights are adjusted continuously reaching for convergence. In this step, the discrepancy between the responses produced by the network and the desired signal is evaluated. The network adjusts the values of the synaptic weights and this process is finished only when the error assumes an acceptable value [23].

The cross-validation method is used to evaluate data generalization. This procedure was performed through random partitioning of the dataset into two subsets: training and test [23]. The training was accomplished only in phantom images due to the low number of malignant cases in actual clinical exams. Thus, 144 ROIs from phantoms images were used for classification, 72 corresponding to benign and 72 to malignant images. From these 144 images, 70% were designated for training and 30% for validation.

6. Results and Discussion

The first computational test was performed on a set of 80 ROIs from phantom images and 50 from actual clinic ones. After the ROI selection, the following preprocessing techniques were applied: Wiener filter, equalization, and median filter. Figures 5 and 6 show some examples of this preprocessing effect on phantom and clinical images.
In the second step, some segmentation techniques (as those previously mentioned in Section 3) were applied in order to precisely delineate the lesions with smooth and regular edges. The active contour was the only technique that did not use the preprocessing. However, it was necessary to apply opening and closing morphological operators after segmentation to smooth the segmented lesion edge [22]. Manual segmentation provided by an experienced radiologist as well as some examples of the effects of segmentation techniques application and postprocessing on phantom images can be seen in Figure 7.

In order to confirm the validation of these procedures also to actual clinical images, those same techniques were applied and similarly evaluated. Some corresponding results are shown in Figure 8.

Due to the visual subjectivity in evaluation to find the most accurate detector, measures were determined to quantify the distance between the edge automatically defined and that manually delineated by an experienced radiologist. Hence, ten measures were evaluated, according to the descriptions in [24–26]. The calculated values for the phantoms images dataset are shown in Table 1 while Table 2 reports those values but they were calculated for the actual breast images dataset.

Based on these data, the advantage of segmentation by active contours and by the SOM neural network can be noted in comparison to the others. In addition, when comparing both methods with the delimitation by the radiologist, they reported greater accuracy and low error rates. Thus, both detectors were tested individually to gather the best classification results with phantoms as well as actual clinical images.

The first step was the extraction of 24 morphological features [22] only for the 144 phantoms images (72 corresponding to benign signals and 72 to malignant ones). Then, the most relevant features were selected by means of Gaussian distribution curves. Among the 24 curves produced, those

---

**Figure 6**: Preprocessing application to ROIs from actual breast ultrasound images: (a) original image; (b) after Wiener filtering; (c) with equalization; and (d) after being submitted to a median filter.
Figure 7: Results from tests with phantom images: (a) original ROI; segmentations by (b) active contour, (c) region growing, (d) fuzzy \( c \text{-means}, (e) k \text{-means}, (f) \text{SOM}, \) and (g) the radiologist delimitation.

Figure 8: Results from tests with actual images: (a) original ROI; segmentations by (b) active contour, (c) region growing, (d) fuzzy \( c \text{-means}, (e) k \text{-means}, (f) \text{SOM}, \) and (g) delineation by the radiologist.

not evidencing appropriate visual results, that is, those with fully or partially overlapping areas, were discarded. Just 8 Gaussian distributions provided good partition for both the active contour segmentation and SOM, as shown in Table 3.

Each of the 8 selected features by the Gaussian distribution curves was individually introduced to MLP, but the results achieved were not significant. Thus, tests were performed with all possible combinations.

The network topological configuration was continuously adjusted during the training process. The amount of neurons in the single hidden layer varied from 1 to 9, and the learning rate was constantly adjusted between 0.1 and 0.9. During this
Table 1: Evaluation metrics applied to phantoms images regarding the five segmentation techniques applied in the tests.

<table>
<thead>
<tr>
<th>Method</th>
<th>AOM (%)</th>
<th>AUM (%)</th>
<th>AVM (%)</th>
<th>CM (%)</th>
<th>CP (%)</th>
<th>CR (%)</th>
<th>Q (%)</th>
<th>A (%)</th>
<th>Err. (%)</th>
<th>FPR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active contour</td>
<td>83.07</td>
<td>15.44</td>
<td>1.90</td>
<td>84.55</td>
<td>98.10</td>
<td>83.07</td>
<td>94.73</td>
<td>5.27</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Region growing</td>
<td>75.02</td>
<td>22.50</td>
<td>3.23</td>
<td>83.09</td>
<td>77.49</td>
<td>96.76</td>
<td>75.02</td>
<td>91.51</td>
<td>8.48</td>
<td>1.06</td>
</tr>
<tr>
<td>Fuzzy c-means</td>
<td>72.38</td>
<td>19.75</td>
<td>18.51</td>
<td>78.03</td>
<td>80.24</td>
<td>81.48</td>
<td>72.38</td>
<td>87.81</td>
<td>12.18</td>
<td>8.02</td>
</tr>
<tr>
<td>k-means</td>
<td>60.73</td>
<td>9.72</td>
<td>36.78</td>
<td>71.41</td>
<td>90.27</td>
<td>63.21</td>
<td>60.73</td>
<td>77.63</td>
<td>22.36</td>
<td>27.69</td>
</tr>
<tr>
<td>SOM</td>
<td>82.50</td>
<td>13.17</td>
<td>5.17</td>
<td>88.05</td>
<td>86.82</td>
<td>94.82</td>
<td>82.50</td>
<td>94.32</td>
<td>5.67</td>
<td>2.02</td>
</tr>
<tr>
<td>Required value</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Evaluation metrics applied in actual breast images regarding the five segmentation techniques applied in the tests.

<table>
<thead>
<tr>
<th>Method</th>
<th>AOM (%)</th>
<th>AUM (%)</th>
<th>AVM (%)</th>
<th>CM (%)</th>
<th>CP (%)</th>
<th>CR (%)</th>
<th>Q (%)</th>
<th>A (%)</th>
<th>Err. (%)</th>
<th>FPR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active contour</td>
<td>81.69</td>
<td>8.57</td>
<td>11.53</td>
<td>87.20</td>
<td>91.43</td>
<td>81.69</td>
<td>95.83</td>
<td>4.17</td>
<td>3.05</td>
<td></td>
</tr>
<tr>
<td>Region growing</td>
<td>70.85</td>
<td>19.42</td>
<td>25.55</td>
<td>93.22</td>
<td>67.13</td>
<td>92.07</td>
<td>62.32</td>
<td>89.79</td>
<td>10.21</td>
<td>2.19</td>
</tr>
<tr>
<td>Fuzzy c-means</td>
<td>59.86</td>
<td>24.51</td>
<td>33.63</td>
<td>71.96</td>
<td>77.41</td>
<td>72.84</td>
<td>57.36</td>
<td>85.19</td>
<td>14.81</td>
<td>10.87</td>
</tr>
<tr>
<td>k-means</td>
<td>38.92</td>
<td>30.74</td>
<td>57.02</td>
<td>42.26</td>
<td>87.65</td>
<td>45.65</td>
<td>41.61</td>
<td>67.29</td>
<td>32.71</td>
<td>37.89</td>
</tr>
<tr>
<td>SOM</td>
<td>75.25</td>
<td>19.75</td>
<td>24.09</td>
<td>93.97</td>
<td>67.46</td>
<td>96.77</td>
<td>65.15</td>
<td>91.33</td>
<td>8.66</td>
<td>0.62</td>
</tr>
<tr>
<td>Required value</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: The most relevant descriptors selected by Gaussian distribution curves for each segmentation method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Most relevant features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active contour</td>
<td>Perimeter [8], Compactness [9], Circularity [8], Convexity [9], Form factor [10], Area ratio [11], Residue [11], Solidity [9]</td>
</tr>
<tr>
<td>SOM</td>
<td>Perimeter [8], Compactness [9], Circularity [8], Convexity [9], Form factor [10], Area ratio [11], Rectangularity [11], Solidity [9]</td>
</tr>
</tbody>
</table>

Table 4: Topological configuration of MLP.

<table>
<thead>
<tr>
<th>MLP</th>
<th>Active contour</th>
<th>SOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
<td>3 (form factor, area ratio, and solidity)</td>
<td>3 (form factor, area ratio, and solidity)</td>
</tr>
<tr>
<td>Number of intermediate layers</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Number of neurons in the hidden layer</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Number of outputs</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 9 illustrates the respective Gaussian distribution curves used for determining each descriptor selected by their overlapping analysis.

After completing the classification procedure, we obtained the values of true positive (TP), false positive (FP), false negative (FN), and true negative (TN), as shown in Table 5.

In Figure 10, ROC curves regarding the classification of phantom images with MLP are shown as for segmentation by SOM network (a) as well as by active contour (b). The values calculated for the areas under the curve (AUC) are displayed in their respective graphics.

The performance of this classifier was compared to others previously described by correlate literature with similar purposes. Table 6 shows the data previously presented by each one of those studies in comparison to our results mainly in terms of accuracy, sensitivity, and specificity rates.

Both classification proposals, with the contour segmented by SOM (case 1) and active contour (case 2), have yielded high sensitivity and specificity in breast lesions classification, similar to the works considered in Table 6. The same accuracy was achieved for both cases and it was higher than all the results presented by such works. In our study, the sensitivity was higher in case 1, while the specificity was higher in case 2.

Additionally, it is important to stress that in our study the training and validation procedures were performed with phantom images, due to the small set of clinical images corresponding to the class “malignant.” These phantom images had structures of interest with a relatively regular shape and in some cases they were easily segmented. This aided in achieving high accuracy in both cases (for benign and for malignant simulated structures). In order to evaluate whether the classifier is able to generalize and reach similar results when applied to actual clinical images, tests with the second dataset were performed, taking into account the fact that it was trained with phantom images which have
Figure 9: The best results for extracted descriptors from Gaussian curves distribution. (a) Segmentation by active contour; (b) detection by the SOM network.
slightly different intrinsic characteristics. With this purpose, an experienced radiologist determined 173 ROIs which were automatically segmented by both techniques (i.e., active contour and SOM). Then, measures of solidity, area ratio, and form factor were extracted from each ROI. Finally, these data were used by the neural network classifier with the topological configuration providing the best result for each detector (shown in Table 4). The results obtained in this classification are given in Table 7.

The high FP rates are largely related to morphological differences between phantom images and actual ones, mainly the simulated structures in the phantoms under test with more rounded shapes. As a consequence, clinical images with elongated shape tumors were erroneously classified. Figure 11 illustrates an example of such feature.

Even with these morphological differences between the types of images, the classifier achieved good data generalization, reaching 100% of sensitivity and 78% of specificity when using segmentation by SOM. In the classification after segmentation by active contour, the sensitivity rate decreased significantly (only 63%) though specificity was almost the same: 79%. This difference is evident when the area under the curve is calculated; corresponding results are shown in Figure 12.

7. Conclusions

The overlap of benign and malignant characteristics in interpreting ultrasound images turns the process subjective and tends to complicate the diagnosis of breast lesions. For this reason, CAD schemes have emerged to improve the analysis of the radiologist by means of computerized characterization.

Some flaws however often arise in many CAD schemes when evaluating images from breast ultrasound acquisitions mainly due to speckle noise influence on the lesions boundaries definition. This is the reason of applying preprocessing techniques before the segmentation procedure.
Table 7: Classification of actual clinical images.

<table>
<thead>
<tr>
<th>Detector</th>
<th>Classes</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Benign</td>
<td>Malignant</td>
<td>TP</td>
<td>FN</td>
<td>FP</td>
</tr>
<tr>
<td>Active contour</td>
<td>150</td>
<td>23</td>
<td>15</td>
<td>8</td>
<td>31</td>
</tr>
<tr>
<td>SOM</td>
<td>23</td>
<td>0</td>
<td>23</td>
<td>33</td>
<td>117</td>
</tr>
</tbody>
</table>

Figure 11: Example of a morphological difference between (a) phantom image and (b) actual clinical exam.

Figure 12: ROC curves obtained from the classification with actual clinical images. (a) Segmentation by SOM network; (b) segmentation by active contours.

Testing five detectors techniques and measuring how close they were relative to the ground truth, only the SOM network and the active contour yielded significant accuracy rates.

A differential of active contour technique is not requiring the preprocessing and the use of morphological operators to smooth edges. Nevertheless, this final smoothing caused some changes in lesion limits, influencing the classification step. Moreover, its algorithmic complexity requires many numerical operations and iterations until convergence of data has been reached. Consequently, this leads to a high computational cost, making the processing relatively slow; about 30 seconds for each ROI.

The segmentation by SOM network on the other hand produced smoother contours and faster outcome results allowing a better understanding of the morphological differences between benign and malignant lesions.

Based on such a result, the classification by MLP was performed for both detectors. After extensive tests and topological changes, the classification taking breast phantom images with detection by active contour was more than 95% accurate, a rate higher than that for detection by SOM network (94.24%). However, the index of the classification with detection by active contour decreased significantly when applied to actual clinical images, which has registered an accuracy of 77.5%. In contrast, the classification accuracy
rate with detection by SOM was almost 81%. Therefore, the detection using the neural SOM network allowed better accuracy and data generalization when new images were introduced to the classifier.

Another important feature to be stressed is the contribution that breast phantoms ultrasound images have made to this investigation, since they constituted a manageable dataset with known materials and morphological characteristics. Furthermore, as large variations related to the simulated lesion morphology can be provided, the dataset becomes useful for the network training before the tests with actual US images, especially with the new guidelines and recommendations for clinical use of ultrasound elastography.

The system described here has an intuitive and easy interface, providing fast and accurate responses to the specialist. The tool is fully automated but allows, if necessary, user intervention to improve the segmentation process, due to manual change in the maximum and minimum parameters in the training. It is considered that the system has produced good results, acting as an important tool for the aid in diagnosis in breast ultrasound.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

Thanks are due to FAPESP (Proc. 2010/14194-5) for the financial support; to Focus Medical Clinic and the Integrated Center for Imaging Diagnosis at Santa Casa of São Carlos (Brazil) for providing the US clinical images; and to Dr. Luciana B. Verçosa, M.D., for her contribution to experimental tests.

References


