

# Breast Cancer Diagnosis Based on a Suitable Combination of Deformable Models and Artificial Neural Networks Techniques

Yosvany López<sup>1</sup>, Andra Novoa<sup>1</sup>, Miguel A. Guevara<sup>1</sup>, and Augusto Silva<sup>2</sup>

<sup>1</sup>Center for Advanced Computer Sciences Technologies, Ciego de Ávila University, Cuba

<sup>2</sup>IEETA, Aveiro University, Portugal

{yosvany, andra}@informatica.unica.cu,

{mangel, asilva}@ieeta.pt

**Abstract.** According to the World Health Organization (WHO) breast cancer is the most common cancer suffered by women in the world, which during the last two decades has increased the women mortality in developing countries. Mammography is the best method used for screening; it is a test producing no inconvenience and with small diagnostic doubts of breast cancer since the preclinical phase. For this reason, unfailing Computer-Aided Diagnosis systems for automated detection/classification of abnormalities are very useful and helpful to medical personnel. In this work is proposed a novel method that combines deformable models and Artificial Neural Networks among others techniques to diagnose diverse mammography abnormalities (calcifications, well-defined / circumscribed masses, spiculated masses, ill-defined masses, architectural distortions and asymmetries) as benign or malignant. The proposed algorithm was validated on the Mammographic Image Analysis Society (MiniMIAS) database in a dataset formed by 100 mammography images, which were selected randomly.

**Keywords:** Breast cancer, mammography images, deformable models, artificial neural networks.

## 1 Introduction

According to the World Health Organization (WHO) breast cancer is the most common cancer suffered by women in the world with 1151298 cases in 2002. During the last two decades this disease has increased the women mortality in developing countries having, up to date, 31% of cases. Breast cancer is one of the major causes of death in women aged 35 – 64 years in Latin America.

Mammography is the best method used for screening; it is a test producing no patients inconvenience and with small diagnostic doubts of breast cancer since the pre-clinical phase. For this reason, unfailing Computer-Aided Diagnosis (CAD) systems for automated detection/classification of abnormalities are very useful and helpful, providing a valuable “second opinion” to medical personnel [1, 2].

Several methods have been reported to detect the presence of abnormalities in digital mammograms. Basically is possible to classify the developed methods in two classes: methods for enhancing mammography abnormalities details (only to be classified by specialized medical personnel) and automatic or semiautomatic CAD methods (in which is produced an automatic or semiautomatic abnormalities classification). Some of CAD methods reported are: methods for masses detection using multiple circular path convolution neural networks models [3], the combination of Artificial Neural Networks (ANN) and Wavelet Transform (WT) for detecting microcalcifications [4], CAD algorithms to identify breast nodule malignancy combining multiple sonographic features and ANN classifiers [5], learning contextual relationships in mammograms based on hierarchical pyramid ANN [6], neural-genetic algorithms for feature selection to classify microcalcifications [7], and methods for discrimination and classification of mammograms in benign, malignant and normal tissues using independent component analysis and ANN [8] among others techniques.

Methods mentioned before represent important approximations (with interesting results) to improve (in major or minor degree) the mammography image analysis process, but in general (at present) reported techniques are paying attention only to classify some specific class of masses or calcifications and not put together the complete classification problem. Wide-ranging techniques including the possibility to classify calcifications and diverse classes of masses are still a not solved problem.

In this work we propose a new method that combines deformable models and ANN among others techniques to diagnose a wide range of mammography abnormalities classes (calcifications, well-defined/circumscribed masses, spiculated masses, ill-defined masses, architectural distortions and asymmetries) as benign or malignant tissues. The proposed algorithm was validated on the Mammographic Image Analysis Society (MiniMIAS) database, with a dataset formed by 100 images selected randomly, of which 60 images were used in the training set and the rest for method evaluation.

The remainder paper is organized as follows: section 2 describes the technical aspect related with the new proposed method. Section 3 outlines the achieved results. Conclusions are presented in section 4, where also are included some ideas for future work.

## 2 Proposed Method

The developed method includes five steps: region of interest (ROI) selection, adjust image intensities, segmentation, features extraction and abnormalities classification. An algorithm prototype was implemented in MATLAB (version 7.0) to test the proposed method on the MiniMIAS database. The MiniMIAS (MiniMammographic) is a reduce version of Mammographic Image Analysis Society (MIAS) database with a resolution of 200 microns by pixel and clipped/padded so that every image size is equal to 1024 x 1024 pixels. We selected randomly a representative dataset that included examples of diverse classes of abnormalities: calcifications, well-defined/circumscribed masses, spiculated masses, ill-defined masses, architectural distortions and asymmetries. The dataset was composed by 100 images, of which 60 images were used in the training set and the rest for method evaluation.

## 2.1 ROI Selection

ROI Selection is a fundamental step because the selected region represents the mammography image part used as input for the developed algorithm. This process allows that the user select for processing only regions suspicious to contain abnormalities (see Fig. 1b).

## 2.2 Adjust Image Intensities

One of the most common defects of photographic or electronic images is poor contrast resulting from a reduced, and perhaps nonlinear, image amplitude range. Image contrast can be often improved by amplitude rescaling of each pixel. Adjust image intensities is a conventional contrast enhancement technique. To enhance the mammography images contrast we map the grayscale intensity values of input mammography images to new values such that 1% of data is saturated at low and high intensities to produce a new image in which the contrast is increased (see Fig. 1c).

## 2.3 Segmentation

Segmentation is considered an important step in any image analysis process because correct objects segmentation can reduce drastically the time and the computational cost of any image analysis process. Deformable models (snakes) have been used successfully in many and diverse image segmentation tasks. Our method used a novel variant of deformable models: the live wire techniques to produce faster and precise abnormalities segmentation of selected ROIs (see Fig. 1d).

Live wire (or intelligent scissors) is an interactive boundary tracing technique, considered as a competing technique to snakes. This technique allows (with minimal user interaction) to exercise control over the segmentation process. Live wire have two essential components: a local cost function that assigns lower cost to image features of interest (edges) and an expansion process to form optimal boundaries for objects of interest, based on the cost function and seed points provided by the user. Boundary finding in live wire can be formulated as a directed graph search for an optimal (minimum cost) path using Dijkstra's algorithm in the underlying graph model. A more profound live wire technique definition, with examples of applications to medical images are presented in [9],[10],[11] and [12].

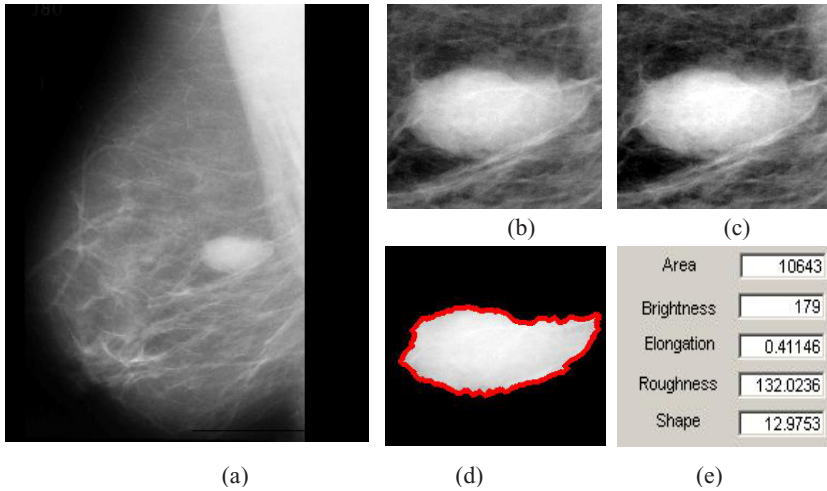
We apply the live wire technique to produce the first curve contour approximation  $E$  (edge points) of the abnormality present in the selected ROI (see Fig. 1d). Due to  $E$  be not a continuous curve we interpolate  $E$  using a spline function to produce a continuous curve: a better abnormality approximation. Final  $E$  contour is used after to compute the morphometric feature vector of the abnormality under study.

## 2.4 Feature Vector

Diverse sets of morphometric features were evaluated and tested [13], but based on statistical evaluations we selected a set of five features to include in the feature vector (related with shape and size), which were sufficient to classify correctly the abnormalities under study. These features are: object area, brightness, object shape,

roughness and elongation. For computing these features we take as input the edge pixels that belong to the closed curve (contour) represented by  $E$ . Mathematic formulation used for computing the features was the following:

- $O$  set of pixels that belong to the abnormality segmented
- $E \subset O$  edge pixels (subset of  $O$ )
- $Area = |O|$
- $Perimeter = length(E)$
- $Elongation = \frac{diam}{DIAM}$ , where  $diam$  and  $DIAM$  represent the minimum and maximum diameters
- $Roughness = \frac{Perimeter^2}{4 \times \pi \times Area}$
- $Shape = \frac{Perimeter \times (diam + DIAM)}{8 \times (Area)}$
- $Brightness = mean(O)$  Intensity levels average of pixels that belong to the selected abnormality



**Fig. 1.** a) Original Image, b) ROI selected, c) Adjusted intensities of (b), d) Segmented abnormality, e) Feature vector computed from (d)

### 2.5 Abnormalities Classification

Semiautomatic or automatic abnormalities classification has the aim to offer one “second opinion” (diagnostic) to medical personnel about abnormalities present on the mammography images. We designed and tested different ANN models to classify

abnormalities under study, but better results were obtained with Feedforward Backpropagation (FFBP) and Learning Vector Quantization (LVQ) ANN models. These neural networks models were trained with 60 vectors, which contain the computed features described before in 2.4 (including benign and malignant examples) representative of diverse abnormalities classes: calcifications, well-defined/ circumscribed masses, spiculated masses, ill-defined masses, architectural distortion and asymmetries.

### 2.5.1 Feedforward Backpropagation Neural Network

The Feedforward Backpropagation (FFBP) is one of the more studied neural network by the scientific community and the most common used in many medical applications. Morphologically, the FFBP is formed by a set of organized neurons in layers: hidden and output layers. Network architecture is determined by the number of neurons in the hidden layers.

The learning process of a FFBP network is characterized to be supervised, the network parameters (known as weights) are estimated from a group (pairs) of training patterns composed for input and output patterns  $\{(x^t, y^t)\} t = 1..n$ .

The backpropagation algorithm [14] is a generalization of the delta rule proposed by Widrow-Hoff [15]. The term “backpropagation” refers to the form in that the error gradient function is calculated for the FFBP network. Therefore, the network adjust takes place as a result of the estimation of weights parameters. The learning involves an adjustment of the weights comparing the desired output with the network answer so that the error is minimized.

Our FFBP network was designed with three layers: two hidden layers with 14 and 8 neurons respectively and an output layer with 12 neurons. Each neuron from output layer represent one (benign or malignant) abnormality class (calcifications, well-defined/circumscribed masses, spiculated masses, ill-defined masses, architectural distortions and asymmetries). The logarithmic sigmoid (*logsig*) was the transfer function used on the three layers. Mathematical definition of this function is the following:  $\text{logsig}(n) = 1 / (1 + \exp(-n))$ .

A MATLAB network implementation can be observed below in Fig.2.

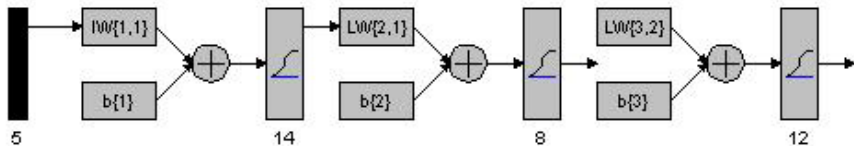


Fig. 2. Feedforward Backpropagation Neural Network

### 2.5.2 Learning Vector Quantization

Learning Vector Quantization (LVQ) is a precursor of the well-known self-organizing maps (also called Kohonen feature maps) and like them it can be seen as a special kind of ANN[16]. Both types of networks represent a set of reference vectors, the positions of which are optimized with a given dataset. A neural network for learning

vector quantization consists of two layers: an input layer and an output layer. It represents a set of reference vectors, the coordinates of which are the weights of the connections leading from the input neurons to an output neuron. Hence, one may also say that each output neuron corresponds to one reference vector. The learning method of learning vector quantization is often called competition learning, because it works as follows: for each training pattern the reference vector that is closest to it is determined. The corresponding output neuron is also called the winner neuron. The weights of the connections to this neuron - and this neuron only: the winner takes all - are then adapted. The direction of the adaption depends on whether the class of the training pattern and the class assigned to the reference vector coincide or not. If they coincide, the reference vector is moved closer to the training pattern, otherwise it is moved farther away. This movement of the reference vector is controlled by a parameter called the learning rate. It states as a fraction of the distance to the training pattern how far the reference vector is moved. Usually the learning rate is decreased in the course of time, so that initial changes are larger than changes made in later epochs of the training process. Learning may be terminated when the positions of the reference vectors do hardly change anymore.

We utilized an LVQ network formed by a competitive layer (input layer) of 16 neurons and a linear layer (output layer) of 12 neurons. Each neuron from linear layer represent one (benign or malignant) abnormality class (calcifications, well-defined/circumscribed masses, spiculated masses, ill-defined masses, architectural distortions and asymmetries), that we consider the adequate design to classify the abnormality classes under study.

A MATLAB network implementation can be observed in Fig.3.

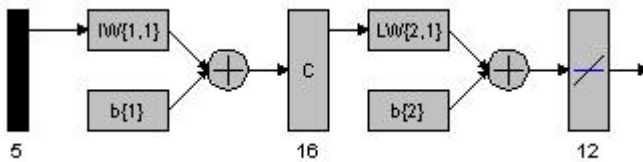


Fig. 3. Learning Vector Quantization Neural Network

### 3 Results and Discussions

Proposed method was evaluated in the MiniMIAS (MiniMammographic) database. MiniMIAS is a reduce version of MIAS database, with a resolution of 200 microns by pixel and clipped/padded so that every image size is equal to 1024 x 1024 pixels. MiniMIAS have included an information file with several details as: image reference number, type of background tissues, class of abnormality present, severity of abnormality (benign or malignant), the x, y coordinates of the abnormality center and an approximated radius (in pixels) of a circle, which enclose the abnormality. This information file was used to select correct ROI surrounding the abnormalities.

We selected randomly a representative dataset (including examples of all abnormalities classes: calcifications, well-defined/circumscribed masses, spiculated masses, ill-defined masses, architectural distortions and asymmetries) formed by 100 images, of which 60 images were used in the training set and 40 for method evaluation.

FFBP and LVQ developed networks models were then trained with a matrix formed by the 60 vectors representing the selected morphometric features (abnormality area, brightness, roughness, shape and elongation). These vectors were computed from abnormalities presents on the images ROIs belonging to the training set.

Classification results are expressed in terms of three parameters: True Positive (**TP**), False Positive (**FP**) and False Negative (**FN**). A **TP** is obtained when a mammogram abnormality is classified into the correct (benign or malignant) class. When a benign mammogram abnormality is incorrectly classified into another benign class or into a malignant class, it is defined as a **FP**. A **FN** is obtained when a malignant mammogram abnormality is incorrectly classified into another malignant class or into a benign class.

Table 1 resume obtained results with the application of both (FFBP and LVQ) developed neural networks models to the simulation set formed by 40 abnormalities vectors (not includes in the training set). Calcifications, other ill-defined masses and well-defined/circumscribed masses were the classes with major incidence of true positives. Spiculated masses and architectural distortion were the classes with minor incidence of true positives classifications.

Summarizing was observed that FFBP model had a better performance in relation with the LVQ model. FFBP model obtained a classification score of 97.5%, 39 true positives, only 1 false positive and 0 false negative versus LVQ model with a classification score of 72.5%, 29 true positives, 5 false positives and 6 false negatives.

**Table 1.** Classification score

Anormalities	No. ROIs	Neural Networks							
		FFBP				LVQ			
		TP	FP	FN	(%)	TP	FP	FN	(%)
Calcifications	7	7	-	-	100	7	-	-	100
Well-defined/ Circumscribed Masses	7	7	-	-	100	6	1	-	85.7
Spiculated Masses	7	7	-	-	100	3	2	2	42.9
Other Ill-defined Masses	7	7	-	-	100	7	-	-	100
Architectural Dis- tortion	6	5	1	-	83.3	2	2	2	33.3
Asymmetry Masses	6	6	-	-	100	4	-	2	66.6
<b>Total</b>	40	39	1	-	<b>97.5</b>	29	5	6	<b>72.5</b>

## 4 Conclusions

We have developed a new method based on the suitable combination of deformable models and ANN capable to classify correctly (as benign or malignant) six different

types of breast cancer abnormalities on mammography images. Also was demonstrated that a simple set of five morphological features (feature vector) computed from segmented abnormality edges are sufficient to obtain good (true positives) classification results. A full prototype was implemented in MATLAB (version 7.0) to test the algorithm performance, which was verified successfully with two different neural networks models: FFBP (97.5%) and LVQ (72.5%) true positives in a dataset of 40 vectors not include in the networks training set.

Our future work will be focused to enrich the feature vector with new features and to improve the networks models that allow better abnormalities classification.

## Acknowledgments

This project was carried out in the Center for Advanced Computer Sciences Technologies at Ciego de Avila University, Cuba. The authors thanks the Cuban Academic of Sciences for their financial support as well as the Radiology Department at the Moron Hospital, Cuba for help us in the evaluation of method developed.

## References

1. Nishikawa, S.R., Papaioannou, J., Onsis, R., Haldemann Heusler, R.A., Giger, M.L., et al.: Performance of a prototype clinical intelligent mammography workstation. In: Digital mammography, Elsevier, Amsterdam (1996)
2. Roehrig, J.D.T., Hasegawa, A., Hunt, B., Marshall, J., Romsdahl, H., et al.: Clinical results with R2 ImageChecker system. In: Digital mammography, Kluwer academic Publishers, Dordrecht (1998)
3. Shih-Chung, B.L.H.L., Wang, Y., Kinnard, L., Freedman, M.T.: Multiple Circular Path Convolution Neural Network System for Detection of Mammographic Masses. *IEEE Transaction on Medical Imaging* 150–158 (2002)
4. Sepehr M.H.Jamarani, B.G.R.-r., Behnam, C.H.: A Novel Method for Breast Cancer Prognosis Using Wavelet Packet Based Neural Network. In: *IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, China (2005)
5. Segyeong Joo, Y.S.Y., Moon, W.K., Kim, H.C.: Computer-Aided Diagnosis of Solid Breast Nodules: Use of an Artificial Neural Network Based on Multiple Sonographic Features. *IEEE Transaction on Medical Imaging* 1292–1300 (2004)
6. Paul Sajda, C.S., Pearson, J.: Learning Contextual Relationships in Mammograms Using a Hierarchical Pyramid Neural Network. *IEEE Transaction on Medical Imaging* 239–250 (2002)
7. Ping Zhang, B.V., Kumar, K.: Neural vs. statistical classifier in conjunction with genetic algorithm based feature selection. *Pattern Recognition Letters* 1–11 (2004)
8. Lúcio, F.A., Campos, A.C.S., Barros, A.K.: Diagnosis of Breast Cancer in Digital Mammograms Using Independent Component Analysis and Neural Networks. In: Sanfeliu, A., Cortés, M.L. (eds.) *CIARP 2005. LNCS*, vol. 3773, Springer, Heidelberg (2005)
9. Artur Chodorowski, U.M., Langille, M., Hamarneh, G.: Color Lesion Boundary Detection Using Live Wire. *Medical Imaging* (2005)
10. Jianming Liang, T.M., Terzopoulos, D.: United Snakes. *Medical Image Analysis* p. 215–233 (2006)



11. Ghassan Hamarneh, J.H.: DTMRI: Segmentation using DT-Snakes and DT-Livewire. In: IEEE International Symposium on Signal Processing and Information Technology (2006)
12. Kelvin Poon, G.H.: Rafeef Abugharbieh, Segmentation of Complex Objects with Non-Spherical Topologies from Volumetric Medical Images using 3D Livewire (2006)
13. Karsten Rodenacker, E.B.: A feature set for cytometry on digitized microscopic images. *Analytical Cellular Pathology* 1–36 (2003)
14. Hinton, G.E., William, D.E.R.R.J.: Learning internal representation by back-propagation errors. *Nature* 323, 533–536 (1986)
15. Widrow, B.: M.E.H. Adaptive switching circuits. in IRE WESCON Convention Record, New York IRE (1960)
16. Kohonen, T. (ed.): *Self-Organization and Associative Memory*, 2nd edn. Springer, Berlin (1987)