Automatic Breast Cancer Detection Methodology Using Artificial Neural Networks

MALIK BRAIK
Information Technology Department
Al-Balqa' Applied University
JORDAN
m_fjo@yahoo.com

ALAA SHETA
Computers and Systems Department
Electronics Research Institute (ERI)
EGYPT
asheta66@gmail.com

SULTAN AL-JAHDALI
Computers and Information Systems
Taif University
SAUDI ARABIA
aljahdali@tu.edu.sa

Abstract: - Artificial Neural Network (ANN) techniques are increasingly being applied in many areas of medical fields for the analysis of complex data. In this paper, a new morphological approach based ANN is developed for an efficient detection of breast cancer from mammograms. The developed scheme consists of three steps: 1) a preprocessing step using traditional image enhancement techniques, 2) a proposed morphological segmentation algorithm and, 3) ANN using some extracted features is being applied for possible detection of abnormal cancer cells. The feasibility of the proposed approach was explored on 32 commonly virulent images provided by the Medical City Hospital in Jordan. The obtained results are encouraging; performing well in segmenting breast medical images and often generates a favorable detection of breast cancer from mammograms.

Key-words: - Extracted features, Artificial Neural Networks, Morphological Operations, Mammogram Images.

1 Introduction

Medical imaging technology plays a vital role in diagnosis of breast cancer disease, and can help in evaluating a wide range of therapies. Effective prevention is not potential, but at least, an early detection can reduce the chance of cancers from becoming obstinate. Medical imaging has been used recently in many applications including: surgery simulations, surgical planning, automated classification of blood cells, and detection of tumors [1]. Breast cancer is the most common cancer among women, and women have the highest chance of survival if the cancer could be detected at the early stages due to the fact, that an early diagnosis is more curable than in a later stage [2], [3]. Ultrasound and magnetic resonance imaging are not early detection screening tools, but primarily are used as diagnostic tools [4]. Available tools and information must be integrated to make an early detection of cancer as a practice; this can provide a real-time health data to the doctor, and it will provide a very robust health information system [5]. There have been many recognized studies and literature surveys on developing breast cancer disease; also, many segmentation algorithms have been developed during the past decades for breast cancer detection [6]. However, accomplishing breast cancer detection is currently a challenging problem and much research is being undertaken to improve the current methods. In conclusion, the aim of this work is twofold: 1) identifying the region

ISBN: 978-960-474-260-8

of interests in breast tumors using some basic morphological operations; 2) distinguishing between "cancer" and "non-cancer" of breast mammogram images using Artificial Neural Networks (ANNs).

2 Proposed Methodology

Detecting tumor locations are likely to give better performance, and can provide clues to treat the cancers in its early stages. Visual interpretation of cancer manually is both inefficient and relatively difficult; also, it requires the expertise of trained radiologists. A closer inspection of mammogram images reveals several difficulties for possible cancer detection. Those difficulties include global appearance and the variations in the recording procedure [7]. The proposed scheme is capable for detecting possible tumor locations of breast in mammograms based on morphological operations and ANNs. The developed processing scheme has three main parts:

1. **Enhancement stage**: This stage consists of contrast stretching and histogram equalization.
2. **Segmentation stage**: This stage has being applied using some basic morphological operations for locating the virulent regions in the breast image.
3. **Detection stage**: A detection criterion based on ANNs is applied for distinguishing cancerous from non-cancerous images.

The general outline and details of the proposed detection algorithm are presented in the following sections.

3 Enhancement Stage

Medical images suffer from a wide variety of distortion in various image processing applications. So, this stage is aimed to realizing an improvement in the quality of a given medical image. In this paper, the enhancement stage is accomplished by contrast stretching and histogram equalization in succession [7].

3.1 Contrast Stretching

The simplest method to increase the contrast of the image is to adjust the histogram so that there is a greater separation between foreground and background gray level distributions. Applying contrast enhancement filters improve the readability of low contrast areas in the image. Also, they will destroy areas of the image where the intensity of the pixels is outside the range of intensities being enhanced [7].

3.2 Histogram Equalization

It is beneficial in histogram equalization to produce output images that are easily analyzed by the human eye. It modifies the image such that its histogram has a desired shape. This is useful in stretching the low contrast levels with narrow histograms. The uniform distribution of the histogram is considered as a probability distribution, it achieves the maximum entropy, which contains the most information. Therefore, redistribute the gray levels to obtain a uniform histogram as possible, and then the image information should be maximized [7].

4 Segmentation Stage

The segmentation stage is primarily used to accurately segment the masses and distinguish malignant from benign tumors for the breast images. Thus it provides the following goals:

1. Specifying the locations of suspicious areas to assist radiologists during the diagnosis.
2. Classification the abnormalities of the breast into benign or malignant.
3. Spotting salient regions in mammograms such that salient regions correspond to distinctive areas that may include the breast boundary, the pectoral muscle, candidate masses and some other dense tissue regions.

4.1 Morphological Operations

Morphological operations are a way of extracting image elements. Dilation and erosion are the two basic morphological operations. Dilation is usually used to smooth boundaries of the regions or bridge very small gaps between neighboring regions. Erosion has an opposite effect of dilation, such that it shrinks the objects uniformly. With A and S as sets in Z^2, dilation of A by S is defined as given in (1) [7].

ISBN: 978-960-474-260-8

$$A \oplus S = \{z | (\hat{S})_z \cap A \neq \Phi \qquad (1)$$

Also, erosion of A by S is defined as given in (2) [7].

$$A \ominus S = \{z | (S)_z \subseteq A \qquad (2)$$

Other morphological operations like closing and opening were also introduced to assist removing some tiny objects from the image. The segmentation methodology can be concisely summarized into a group of steps as follows:

Step 1: The Region Of Interests (ROIs) is selected for the purpose of segmentation. ROIs is then identified and labeled. ROIs could be tumor or any other relevant matter.
Step 2: After specifying the ROIs, all small undesirable objects are reduced or completely removed.
Step 3: After eliminating all tiny objects, the image is smoothed using multidimensional filter.
Step 4: Dilation and erosion are applied using flat linear structuring element in each specified neighborhood for the selected ROIs in Step 3.
Step 5: Suppressing light structures that are connected to image border to reduce the overall intensity of the ROIs.
Step 6: Reconstructing the final segmented image from the previous two steps.

5 Detection Stage

Neural network is chosen as a classification tool due to its well known technique as a successful classifier for many applications [8], [9]. Dataset for training and evaluating consists of 32 mammogram images. The output segmented images are split into two parts; the training segmented images which are used to train the NN model, whilst a testing segmented images are used to verify the accuracy of the trained NN model. The training images consists of 50% of the total output mammogram segmented images with the testing images consisting of the remaining 50%. Each dataset had two classified images, examples of the two classes, i.e. cancer and no cancer, must be provided to train the classification model. The network was trained using the feed-forward back propagation network to minimize the Mean Square Error (MSE) function [10]. All the output segmented images were fed into the ANN, then, the segmented object is recognized if it has a cancer or not. ANN is trained several times for possible detection of breast cancer cells.

6 Experimental Results

6.1 Stage 1: Enhancement

For the purpose of proffer the results of the proposed cancer detection algorithm; a sample malignant image was discussed in this research. The results of applying contrast stretching process and histogram equalization to a breast cancer image are shown in Fig. 1. It can be seen from Fig. 1 that the dynamic range of the gray levels in the image being enhanced is increased. As such, the malignant regions appear visibly in the enhanced images. In addition, the overall intensity of the mammogram image is increased, such that, the histogram of the enhanced image is transformed to one that approximates a uniform distribution as shown in the histogram of the equalized image.

6.2 Stage 2: Segmentation

The proposed segmentation procedure is applied for the whole set of tested images. The results of the segmentation procedure on a malignant image are shown in Fig. 2.

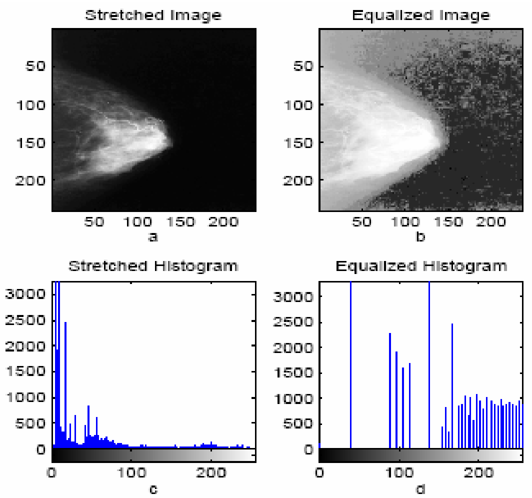

Fig. 1 Enhancement results: a) Enhanced image using contrast stretching, b) Enhanced image using histogram equalization, c) Histogram of the stretched image, d) Histogram of the equalized image.

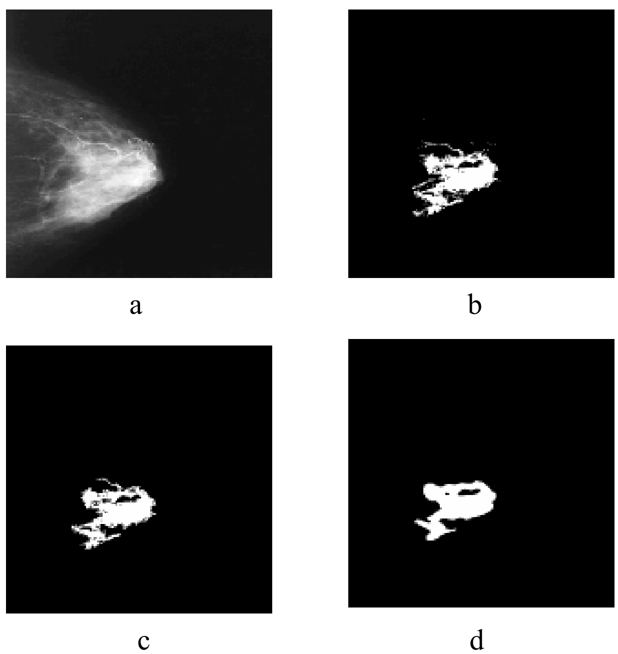

Fig. 2: Results of the segmentation procedure: a) The input mammogram image, b) The suspicious regions in the mammogram image, c) Removing all small objects from the image, d) The final segmented image

The regions of interest are selected by executing step 2 in the segmentation procedure as shown in Fig. 2 (b). Some small objects are produced after executing step 2, moreover, the mammogram images involve several dark regions and noise, which may trick the segmentation process, and these small undesirable objects are reduced in step 3 as shown in Fig. 2 (c). The morphological approach eliminates the remnant dark regions in the image using the proposed filters, namely: the circular averaging filter, erosion filter, and the dilation operator. The dilation operator is applied to bridge the small gaps using the flat structuring element. As a result, tiny protrusions are eliminated from the image. Through the segmentation method; the binary image contains only 0's that represents the background and 1's that represents the object. Fig. 3 shows the binary segmented image.

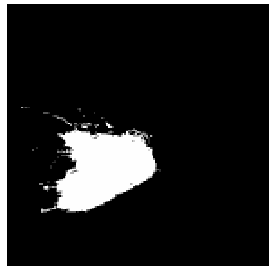

Fig. 3 The output of binary image segmentation

It can be observed from Fig. 3 that some tiny protuberances are not completely removed from the segmented image; therefore, other steps are inescapable to strengthen the segmentation result. The final segmented image resulted from the last steps of the segmentation procedure is shown Fig. 2 (d). Consequently, the morphological operations perform very well in extracting the image elements and can express the image details in a specific way. Hence, it can produce the best segmentation results of the mammograms including only the upnormal areas, as a prior task to posterior classification.

6.3 Stage 3: ANN for Cancer Detection

The proposed detection method based on ANN exhibits a high accuracy of correct classification. An example of a class containing suspicious regions classified by ANNs as breast cancer is shown in Fig. 4.

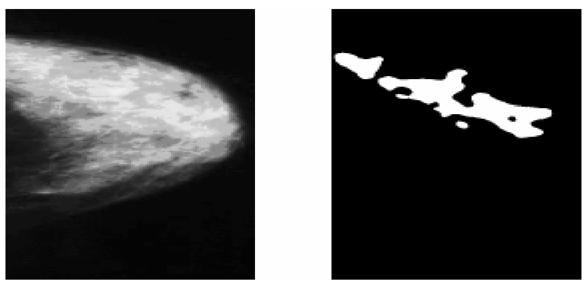

Fig. 4 Testing case: A mammogram image containing Brest cancer classified by ANN

In this study, ANN is utilized a three - layer feed-forward network with back-propagation algorithm to model breast cancer mammography images. The first layer has 1 neuron; the second layer has 10 neurons, whilst the third layer has only 1 neuron. The convergence rate of classifying the breast cancer based ANN is recognized using MSE. The MSE value of 0.0001 is used to measure the prediction accuracy; hence, the convergence rate achieved a high accuracy with 15 iterations. The convergence curve of the detection process is shown in Fig. 5. In Fig. 5 the MSE value is converged to an optimum value, thus the presented method may be used as an effective prompting tool to assist radiologist in diagnosis of breast cancer, then, ANN is a powerful tool to model complex medical data which may capable of high classification accuracies.

7 Conclusions and Future Works

In this paper an image-processing-based approach is proposed for breast cancer segmentation and detection. The proposed technique is composed of three main steps; in the first two steps the images at hand are enhanced and segmented using traditional enhancement and a morphological approach, respectively, in the last step the segmented objects were passed through pre-trained neural networks. This type of approach can help the diagnostics as a useful view, and can significantly support an accurate detection of breast cancer disease.

ISBN: 978-960-474-260-8

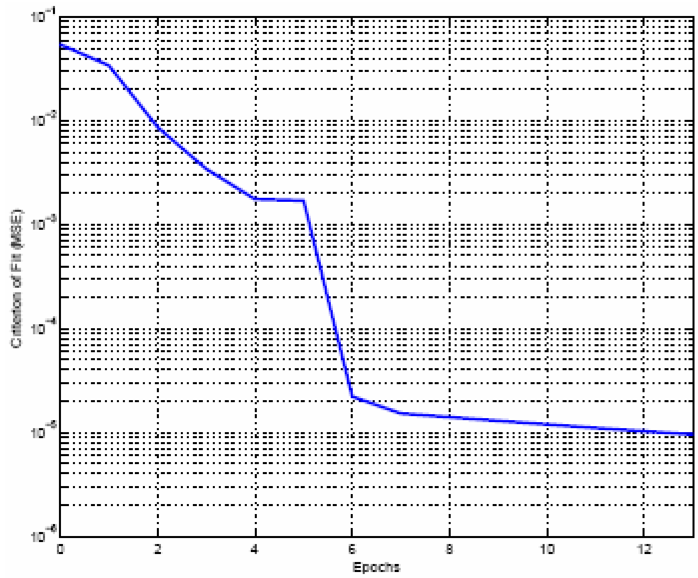

Fig. 5 Convergence curve of ANNs

References

[1] M. Alata, M. Molhim, and A. Ramini, “Optimizing of fuzzy c-means clustering algorithm using GA,” in *Proceedings of World Academy Of Science, Engineering, and Technology*, vol. 29, pp. 224– 229, 2008.

[2] N. M. Sudharsan, Ng E. Y.-K., and Teh S. L., “Surface temperature distribution of a breast with/without tumor,” *International Journal of Computer Methods in Biomechanics and Biomedical Engineering*, vol. 2, 1998.

[3] K. A. Barron, “Modeling and uncertainty in breast cancer decision making,” Master’s thesis, The Pennsylvania State University, the Graduate School, Industrial Engineering, 2006.

[4] C. M. Clark, H. O. Lawrence, goldgof B. Dimtry, C. P. Laurence, V. P. Rebert, and S. S. Martin, “Mri segmentation using fuzzy clustering techniques,” in *IEEE Trans. on Engineering on medical and biology*, pp. 730–742, 1994.

[5] M. Hadhoud1, M. Amin, and W. Dabbour, “Detection of breast cancer tumor algorithm using mathematical morphology and wavelet analysis,” in *Proceedings of ICGST, GVIP 05 Conference*, vol. 29, pp. 19– 21, 2005.

[6] F. X. Wu, “Genetic weighted k-means algorithm for clustering largescale gene expression data.,” *BMC bioinformatics*, vol. 9 Suppl 6, 2008.

[7] R. Gonzalez and R. Woods, *Digital Image Processing*. Prentice Hall, 3rd Edition, 2008.

[8] R. Pydipati, “Evaluation of classifiers for automatic disease detection in citrus leaves using machine vision,” Master’s thesis, University Of Florida, 2004.

[9] E. Alba and J. F. Chicano, “Training neural networks with genetic algorithms hybrid algorithms,” *Departamento de Lenguajes y Ciencias de la Computacion University of Malaga, Spain, Ministry of Science and Technology and FEDER under contract TIC2002-04498-C05-02 (the TRACER project, http://tracer.lcc.uma.es)*, 2002.

[10] N. Chiras, C. Evans, and D. Rees, “Non-linear GAS turbine modeling using feedforward neural networks,” *Proceedings of ASME TURBO EXPO June 3-6, Amsterdam, The Netherlands GT-30035, University of Glamorgan, publisher of Electronics, Pontypridd, CF37 1DL, Wales,UK*, 2002.