

Vol. 5 No. 4 (Apr 2012) ISSN: 0974- 6846

Solving mammography problems of breast cancer detection using artificial neural networks and image processing techniques

^{1*}Moh'd Rasoul A. Al-Hadidi, ²Mohammed Y. Al-Gawagzeh and ³ Bayan A. Alsaaidah

Computer Engineering Department, Engineering College, Al-Balqa Applied University Al-Salt 1911, Jordan mohammed_hadidi@bau.edu.jo*, mohammed_gogazeh@yahoo.com, eng_bayan85@yahoo.com

Abstract

In this paper, we propose a complementary technique of breast cancer diagnosis that covers five stages of breast cancer detection based on mammography, which solves many of the problems found otherwise. We also show a very large area where many methods and techniques can be successfully merged in order to obtain a useful result for human use. These include scaling of the image, removing small objects, smoothing, extracting features, ROI extraction and many image processing techniques. Besides, neural networks are used here to train the system to detect cancer according to the dataset. This combination of multiple techniques can solve problems of the breast cancer detection with a high degree of accuracy. Examples and comparisons are given to illustrate and prove this method.

Keywords: Breast cancer, Benign, Mammography, Image processing, Neural networks.

Introduction

Breast cancer affects a large number women population. There are many factors that cause the breast cancer; one of these is the X-Ray (Dubey et al., 2010). In Jordan, the breast cancer is the most common cancer which affects women and it has been increasing over the years. From 1996 to 2007, it is increased from 448 to 817 cases so 369 new cases have been recorded (Boston, 2009). However, in spite of the ever increasing cases of breast cancer over the past years, breast cancer mortality rate has declined amongst women of all ages (Bray et al., 2004). There is the self-exam. A woman can do on monthly bases and this can be considered as early detection of the disease. But at present, mammography test is performed by specialized doctor and is the most effective method for early detection of breast cancer (Linda et al., 2000; Kulkarni et al., 2010). The main goal of mammography test is the early detection of through the detection breast cancer of characteristic masses and micro calcifications which is considered as an important sign of breast cancer (Cristina et al., 2006).

Mammography is used as a breast examination where the patient is undressed to the waist and stand in the front of the X-ray machine. Then each breast is compressed between two plates to take two images of each breast by using a brief X-ray pulse (Robin & Mason, 2002).

There are other exams to detect breast cancer, such as: Clinical Breast Exam (CBE), Computer Aided Detection (CAD), and the blood test.

Artificial Neural Networks (ANN) "are nonlinear information processing devices, built from interconnected elementary processing devices called neurons inspired by the biological nervous systems" (Deepa *et al.*, 2011). It is an effective method to solve a complex problems such as the data mining problem (Usha Rani, 2010), Research article "Mammography and the NP-problem (Asadi et al., 2011).

Inspiring the biological neural networks, Artificial Neural Network consists of compositions of single, non linear processing units which are called neurons that are organized in an inter-connected graph by a weighted connections, each connection has a parameter (weight) (Rojas,1996). These weights are updated until the neural network reached the target output to become a trained neural network.

The use of image processing is a physical process used to convert an image signal of the breast into a physical image. As it is known, this image signal is a digital signal or analog signal, and the actual output can be either an actual physical image or the characteristics of this image. In diverse spectrum of human activities there are a large number of applications of image processing, from remotely sensed scene interpretation to biomedical image interpretation (Acharya & Ray, 2005).

According to relevant scientific researches, breast cancer is a wide field in which scientists can prepare their research papers and give new solutions to breast cancer detection problems and they may find a more effective method to detect the breast cancer at an earlier stage.

Most of the presented methods and techniques concerned the breast cancer detection covers a specific stage of the breast cancer. This study shows results that cover five main stages (stage 0, stage I, stage II, stage III, and stage IV). Moreover, the proposed algorithm is a complementary technique that gives a satisfied result with less processing time and high accuracy comparing with many complex algorithms which were presented by other studies.

Firstly, this paper determines area of interest which performs the breast in the mammogram photo using sequence of functions: Adaptive Histogram Equalization to improve contrast in image, threshold



to make object easier for identification, image object boundaries to detect undesirable small objects to be removed, and morphological closing for smoothing the desired region also makes edge detection easier. Then we obtain important features from ROI which are needed for the neural network operation to be trained. After training the neural network, it should be able to define the state of breast to be normal or has a benign tumor or one of the stages of breast cancer.

There are many research studies have been conducted earlier on breast cancer detection, the following introduces some of them.

In 2010, Raman et al. (2010) presented a classification system for the analysis of mammographic tumor by using machine learning techniques. It provides only the first stage of mammogram mass segmentation result. Abdrabou and Salem (2010) presented a model which effectively classifies breast cancer tumors as either malignant or benign, named as a benign/malignant breast cancer classification model based on a combination of ontology and case-based reasoning. This classification system makes use of clinical data. Two case-based reasoning (CBR) object-oriented frameworks based on ontology are used jCOLIBRI and myCBR. They build a breast cancer diagnostic prototype. And they make an examination during prototyping to examine the use and functionality of the two focused frameworks.

Adusei et al., (2010) designed a database system of mammographic images to make the dealing with these images by retrieving or analyzing intelligently and efficiently. They designed the architecture for the CAD/CADx system and then focused primarily on the design and modeling of the mammography database. In their design, they combined two standards, the Breast Imaging Reporting And Data System (BI-RADS) by American College of Radiology and the Facility Oncology Registry Data Standards (FORDS) by the Commission on Cancer standards to reach a completed system.

Patel and Sinha (2010) present a new method for medical image enhancement which is based on the concept of fractal derivatives and selecting image processing techniques such as segmentation of an image with self similar properties. This concept is based on two criteria: self-similarity and fractional dimensionality. They tested the different alternatives for the algorithm. They found that the outcomes of the experiments were significantly different from each other. That was helpful for the segmentation of a mammogram image and it may be used for developing an expert system to detect the breast cancer at an early stage.

Singh and Mohapatra (2011) showed the result of applying image processing threshold, edge based and watershed segmentation on the mammogram

Vol. 5 No. 4 (Apr 2012)

ISSN: 0974-6846

breast cancer image. It was a novel approach to identify the presence of breast cancer mass and calcification in mammograms using image processing functions, K-means and Fuzzy C-Means clustering for clear identification of clusters. By combining Kmeans and C-means they successfully detected the breast cancer area in raw mammograms images and also classified the total cancer affected area. This system is a real-time system and a visual detection of the breast images.

Gupta et al., (2011) used a data mining techniques to diagnose and prognose the breast cancer. They used the data mining techniques to enhance the breast cancer diagnosis and prognosis, where the breast cancer diagnosis and prognosis are two of the classification problems.

Maitra *et al.* (2011) presented newly a Binary Homogeneity Enhancement Algorithm (BHEA). They detected the abnormal masses by anatomical segmentation of Breast Region Of Interest (ROI). They used a Medio-Lateral Oblique (MLO) view of the Pectoral Muscle mammograms, the Detection Algorithm (PMDA), the Anatomical Segmentation of Breast ROI (ASB) algorithm, and the Seeded Region Growing Algorithm (SRGA). These algorithms are a fully autonomous, and are able to isolate many types of abnormalities.

Experimental and methodology

The algorithm is implemented through MATLAB code executed in an .m file. This file opens a by which all of the graphical user interface algorithm's features can be tested through varying testing of different images, and by the ability to change the algorithm's variables through edit boxes and sliders. An auto button allows the program to automatically set these variables in accordance with the chosen image.

The graphical user interface is simple in that it only requires the user to choose a mammogram image to be processed. The graphical user interface goes through the steps and finally gives a result of the percentage of the cancer stage present in the picture. Graphical user interface

The graphical user interface was created to provide the outmost user friendliness and ease of use. All input to the program and its main algorithm is performed through the GUI, and all the results are produced through it as well. In this GUI, the picture which will be detected is selected, and there are many factors and values (such as: the scaling factor, bandwidth threshold, initial index, and the initial row) of this picture helped to apply the processing that is explained later in this paper in more details and the processed figures are shown in this interface. Fig.1 shows a main view of the GUI. In the GUI there are many buttons:



Vol. 5 No. 4 (Apr 2012)

ISSN: 0974- 6846

Fig.1. GUI parts



1. These are the slider values that change the implementation algorithm's various variables and parameters. They can be changed manually or automatically through (3).

2. Pressing this button starts the algorithm implementation and ends with the algorithm's result.

3. This button changes the slider and edit boxes in (1) to their optimal values in accordance with the image chosen.

4. This is the directory and name of the image chosen to be run in the implemented algorithm.

5. These plots show the algorithm being implemented step by step and end up with the breast cancer detection result.

Proposed algorithm design

The methods contained in the algorithm can be appropriately summarized in Fig.2 and explained in the following sections (3.3) and (3.4):

Fig. 2. Proposed algorithm



Image processing

Image Reading: In MATLAB environment a binary and gray-scale images is represented by one 2-dimensional array, a color image is represented by a 3-dimensional array (one 2-dimensional array for each of the color planes or color channels red, green and blue), (Fig.3) (Thaler ጲ Hochreutener, 2008).

The origin is in the left of the image and its size is defined by two parameters: width (number of columns of the array), height (number of rows of the array). The xand y-coordinates are chosen where the z-axis

points to the front. A single point in the image is called a pixel (Gonzalez *et al.,* 2004). Research article "Mammograph Fig. 3. Scaled image vs RGB image as array in atlab (Thaler & Hochreutener, 2008)



Image Scaling: Image scaling is an important process for several situations in both image processing and image analysis (Gonzalez et al., 2004). In this system image scaling is used in order to reduce computing time.

In the proposed algorithm the image is taken and scaled according to the values specified in the graphical user interfaces edit box. This is done as the first step in the algorithm as a measure to reduce CPU overload and time required to compute the analyzing later procedures. This is done by applying a special command in MATLAB (Fig.4). The image is scaled by 0.1 and the result as follows:



• Image Dimensions: 226x180 pixels

Image Size: 40680 bytes

• There was a reduction in size but without loss of features.

Adaptive histogram equalization (AHE): is a technique in computer image processing to improve contrast in images. It computes many histograms, each corresponds to a different part of the image, then uses them to redistribute the lightness values of the image. The ordinary histogram equalization process uses only a single histogram for an entire image (Gonzalez *et al.*, 2004).

Adaptive histogram equalization is considered as an image enhancement technique to improve an image's local contrast, and to bring out more details in the image. AHE technique can also produce a significant noise. A generalization of adaptive histogram equalization called contrast limited adaptive histogram equalization (Vij &



Vol. 5 No. 4 (Apr 2012)

ISSN: 0974-6846

Fig. 5. Histogram before histogram equalization operation



Fig. 6. Histogram after histogram equalization operation



Fig.7. Image after B/W conversion



Fig. 8. Object detection



Fig. 9. Image after removing small objects



Fig. 10. Image after smoothing



Fig.11. Image before edge detection



Fig.12. Edge detection for the upper part of breast





Vol. 5 No. 4 (Apr 2012) ISSN: 0974- 6846

Singh, 2009). Fig.5 and 6 show the image before and after histogram equalization process.

Converting to black and white: This process is done on a grayscale image by assigning each pixel in the image as true value or false value, based on a certain threshold (Gonzalez et al., 2004) (Fig.7).

Image Flipping: Flipping process of the image occurs when we want to operate on one profile image. In the proposed algorithm, all subsequent image processing operations are taken to be done on image from the left profile of the human breast.

Image objects boundaries and area: This function is performed to detect the object through the boundaries of the image and extracting the properties of each object which is detected as an area. The image contains objects which are identified by the connected white pixels, or the connected true values in the image matrix (Fig.8).

Removing Small Objects: It works with one main object and filter out all the remaining objects. This is done through deleting any object that has a size below the size of the largest object in the image (after sorting) (Fig. 9). Deleting an object essentially makes all of its pixels in the image matrix given a value of false.

Smoothing (morphological closing): This is done through creating a structuring element and then performing morphological closing using it. Morphological closing was performed in order to smooth out the edges and to make sure that no irregularities occur at the edge of the image. This is critical in preparation for the step to occur later by which edge detection is accomplished. Fig.10 illustrates smoothing operation affect.

Édge Detection: The edges of the image assign characteristics of the object boundaries, and are useful for segmentation, registration, and identification of objects in a scene (Vasudevan *et al.*, 2010). This is calculated by taking a row number from the graphical user interface (ideally half the image length) and then finding the first true value across that row as being the column value. By using this method we have row-column pair that defines a point on the image that lies on the edge.

In the proposed algorithm, the point on the edge is taken as to be nearest as possible to the middle of the region of interest. Then, two traces are created, for each direction (up/down). The number of tracing point is taken to be large enough so as to make sure it covers the region of interest. This number will be cut after the tracing procedure is complete, through the steps described next. Starting conditions

Condition 1: Direction is not reversed for more than 20 consecutive times while being reversed less than 6 times. Condition 2: At least a point in trace is taken from a graphical user interface edit button, which is dependent on the image size. The starting index is the maximum of these two conditions. *Stopping conditions*

Condition 1: trace x value remained constant at least a number of times. It is taken from an edit box in the graphical user interface, which is dependent on the image size. Condition 2: trace x values decreased. The stopping index is the minimum of these two conditions. Fig.11 & 12 show edge detection operation.

Extracting the region of interest

After the region of interest is decided through edge detection tracing, the edge detection result will be a mask by which the original image is run through. By this, everything outside the mask will be deleted, and only the pixels inside the mask will remain. Fig.13, 14, 15 and 16 show steps to extract ROI.

Feature extraction: Once the region of interest has been obtained, the three features from which the neural network operates need to be extracted. The three features are: Spread, Intensity, and Lymph.

Each of these features is a rough percentage. The spread percentage identifies how large a tumor is. Intensity identifies how strong a tumor appears to be, while lymph identifies how large the tumor possibly reached the lymph nodes. Tumors are assumed to be the brightest objects inside the finally obtained region of interest, using the previously discussed object boundary tracing and property extraction of the objects. Fig. 17 and 18 show the tumor identification and the three features extracted. Once these three features are obtained, they are used as input to the neural networks.

Artificial neural networks

The algorithm uses a feed-forward back propagation network with three hidden layers, consisting of 50, 10, 50 neurons for the first, second, and third layer respectively. During training, the features are extracted from the images in which the diagnosis is known. From there, each stage of cancer was linked through the neural network training with its three features. After training is over, the trained networks are stored to be used in the algorithm. Whenever an image is taken as input in the algorithm, it is simulated with the trained net-works and from the results a percentage can be given to which diagnosis should be taken from the mammogram.

Matlab is a good programming toolbox package provides functional software environment for creating neural network. The main goal of this package is to provide users with a set of integrated tools neural networks to create models of biological and simulate them easily, without the need of extensive coding.



2525

Fig. 13. Edge detection for the lower part of breast



Fig. 14. Original Image



Fig. 15. ROI recognition



Fig.16. Extracted ROI





Fig. 18. Tumor Identification and extracted features



19: ANN structure



Fig. 20 Succeeded ROI extraction



Fig. 21.Succeeded ROI extraction



Fig.22. Succeeded ROI extraction 4.2 neural network results





Fig. 23. Succeeded ANN





Neural Network Results		
Benign		-0.71315 ^{Minimize}
Normal	:	C.080977
Stage O	3	9.2528e-008
Stage I	:	0.37322
Stage II	•	-1.4185e-0C9
Stage III	:	-5.3944e-006
Stage IV	:	0.84523



Vol. 5 No. 4 (Apr 2012) ISS

012) ISSN: 0974- 6846

Neural network creation is the last step in the proposed algorithm, There is a neural network for each stage. The code takes the desired picture and runs it through the neural networks one by one, each time resulting in a score. For each network we can summarize the most important operations.

Creation network

The function and its own parameters below are used to create and define our neural network:

net = newfit(inputs,targets,numHiddenNeurons):

The arguments of this function is arranged as follows:

1. The inputs of neural network where it contains three important features (intensity, spread, lymph) extracted from the image. 2) Stated target for each stage performed by one dimensional binary array (just one element of this array has value'1' and other elements are assigned to zero that's to separate the desired target from other ones). 3) The number of neurons in the three hidden layers fixed to 50 for the first layer ,10 for the second one, 50 for the third (Fig.19).

Training and Testing stages: The function and its own parameters below are used to train our neural network:net = train(net,inputs,targets);the function parameters are:

1. net: the neural network which created previously.

2. Inputs: inputs of the created neural network as defined before.

3. Targets: stated target the neural network.

Results and analysis

Image processing discussion

As a way to test the image processing capabilities of the proposed algorithm, a loop was created, where the algorithm automatically takes all images in a pre-set folder and runs the algorithm's steps on them, one by one, all whilst recording the results for the operator. Out of a test batch of 125 images, the algorithm successfully extracted 121 regions of interest without human intervention, resulting in a 96.8% success rate without any human intervention. It is worth notina that any human intervention is due to illtaken mammograms and problems in the mammogram itself. In fact, all of the four pictures with which the algorithm failed to decide a region of interest to, the mammogram had a foreign object in the picture out of place. Foreign objects are taken into account in the proposed algorithm, but some forms are too illogical and out of place for the proposed algorithm to deal with. Fig.20, 21 and 22 show the results of one of the succeeded results. Table 1 shows the real results and percentages for each stage which we get from the implementation of neural network. Where: Ti: Total number of images. Ei: No. of Images which exceed. Ps: Percentage for each situation.



Table 1 Results percentages

Table. T Results percentages				
Situation	Ti	Ei	Ps	
Normal	43	34	79.07%	
Benign	40	30	75.0%	
Stage 0	3	0	0.0%	
Stage I	12	5	41.67%	
Stage II	4	2	50.0%	
Stage III	8	3	37.5%	
Stage IV	11	8	72.73%	
T I				

The total number is 125 but we exclude 4 images for which the algorithm failed to decide a region of interest. The Table 1 shows that stage 0 records have the worst result because a few number of images which we get for them, followed by stage III. Also the situations of Normal and Benign records have the best results due to more images available for them. Stage I and Stage II record similar results due to convergence in the case. Stage IV records a great result because of its different characteristics from any other stage.

Table 2. Success percentage

	1 0	
With Stage 0	Without Stage 0	
67.77%	69.49%	

Table 2 shows the final percentage for the neural network with and without stage 0. Table 3 shows the final percentage for the neural network with and without stage III. Table 4 shows the results with and without Stage 0 and Stage III. Despite a rather small sample size of input images, coupled with ambiguous organization of pictures taken from the physician, the results showed great promise. The Fig..23 and 24 are the representative of many tested examples and should be diagnosed as stage IV.

Table. 3	Success	Percentage
----------	---------	------------

With Stage III	Without Stage III
67.77%	69.91%

Conclusion

A novel technique was presented in this paper. It incorporates neural networks in conjunction with advanced image processing procedure as a method by which breast cancer diagnosis was performed based on mammogram pictures obtained. The proposed algorithm showed great success in identifying the region of interest and correctly segmenting all of the input test images. It provides with the proposed algorithm a very high level of robustness. The results of the neural networks gave better accuracy for each breast cancer stage. The resulting diagnosis showed great promise for being an invaluable and dependable tool for the diagnosis of breast cancer. Different methods were seamlessly joined together and meshed in a highly technical algorithm which can be considered efficient and very



easy to use. Thus, our work shows a very large area of methods and Techniques can be successfully merged in order to obtain a useful result for human use.

Future work

proposed The next step might take the algorithm as one module of a larger algorithm that continues the work presented here. As with all optimization methods, given a larger database to work with will result in a higher accuracy range for the neural network results, and further work on the image processing coding might give efficiency boosts and a shorter CPU time for the calculations. Also an X-ray camera with USB port can be added to the system which enables the female patient to capture the image of her breast by herself then send it to a computer which has shared folder, and by simple code we can make a connection between shared folder and this algorithm. Once the image arrives to the shared folder the algorithm will run and give the medical report with no need to go to a doctor.

Acknowledgement

We would like to thank Dr.William Ma'yaah, consultant radiologist from Al-Basheer Hospital,Jordan who helped so much to do this work.

References

- Abdrabou E and Salem A (2010) A breast cancer classifier based on a combination of case-based reasoning and ontology approach. Proc. Int. Multiconf. Com sci. Inform. Technol, ISBN 978-83-60810-27-9 ISSN 1896-7094.
- Acharya T and Ray AK (2005) Image processing: principles and applications. Wiley-Interscience, Hoboken NJ, ISBN 0471719986.
- Adusei I, Kuljaca O and Agyepong K (2010) Intelligent mammography database management system for a computer- aided breast cancer detection and diagnosis. *Int. J. Managing. Inform. Technol.* 2 (2), 1-13.
- 4. Asadi A, Naserasadi A and Asadi Z (2011) A new hybrid algorithm for traveler salesman problem based on genetic algorithms and artificial neural net- works. *Int. J. Compu. Appl.* 24(5), 6-9.
- 5. Bray F, McCarron P and Parkin D (2004) The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Res.* 6, 229-239.
- Cristina L, Ponomaryov V and Luis J (2006) Detection of micro calcifications in digital mammogram images using wavelet transform. Electronics, Robotics and Automotive Mechanics Conf. (CERMA'06), 2, 58-61.
- 7. Deepa S and Aruna Devi B (2011) A survey on artificial intelligence approaches for medical image classification. *Indian. J. Sci. Technol.* 4(11), 1583-1595.



ISSN: 0974-6846

2528

- 8. Dubey R, Hanmandlu M and Gupta S (2010) Level set detected masses in digital mammograms. *Indian. J.Sci. Technol.* 3 (1), 9-13.
- Gonzalez R, Woods R and Eddins S (2004) Digital images processing. 2nd edition, Prentice-Hall, Inc. NJ, 07458, ISBN 0-13-008519-7.
- Gupta S, Kumar D and Sharma A (2011) Data mining classification techniques applied for breast cancer diagnosis and prognosis. *IJCSE*. 2, 188-195.
- 11. Kulkarni D, Bhagyashree S and Udupi G (2010) Texture analysis of mammographic images. *Int. J. Com. Appl.* 5, 12-17.
- Linda J. Warren Burhenne, Susan A. Wood, Carl J. D'Orsi, Stephen A. Feig, Daniel B. Kopans, Kathryn F. O'Shaughnessy, Edward A. Sickles, Laszlo Tabar, Carl J. Vyborny, and Ronald A. Castellino (2000) Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiol.* 215, 554-562.
- Maitra I, Nag S and Bandyopadhyay S (2011) Automated digital mammogram segmentation for detection of abnormal masses using binary homogeneity enhancement algorithm. *Indian.* J. Com. Sci. Engg (IJCSE), 2 (3), 416-427.
- 14. Patel B and Sinha G (2010) Early detection of breast cancer using self similar fractal method. *Int. J. Com. Appl.* 10 (4), 39-44.
- 15. Raman V, Sumari P and Rajeswari M (2010) A theoretical methodology and prototype implementation for detection segmentation classification of digital mammogram tumor by machine learning and problem solving approach. *Int. J. Com. Sci. Issues.* 7 (5), 38-44.
- Robin J and Mason P (2002) Handbook of pharmacy healthcare (diseases and patient advice). *Pharmaceutical Press*, London, UK. ISBN 0-85369-507-5.
- 17. Rojas R (1996) Neural networks: a systematic Introduction. Springer-Verlag.
- Singh N and Mohapatra A (2011) Breast cancer mass detection in mammograms using k-means and fuzzy c-means clustering. *Int. J. Com. Appl.* 22 (2), 15-21.
- 19. Thaler M and Hochreutener H (2008) Image processing basics using MATLAB.
- Usha Rani K (2010) Parallel approach for diagnosis of breast cancer using neural network technique. *Int. J. Com. Appl.* (09758887), 10 (3), 1-5.
- Vasudevan K, Dharmendra T, Sivaraman R and Karthick S (2010) Automotive image processing technique using cannys edge detector. *IJEST*, 2 (7), 2632-2644.
- 22. Vij K and Singh Y (2009) Enhancement of images using histogram processing techniques. *Int. J. Comp. Tech. Appl.* 2 (2), 309-313.