Neural Network Analysis of Breast Cancer from Mammographic Evaluation

P. Abdolmaleki PhD[1].
M. Guiti MD[2].
M. Tahmasebi MSc[3].

Background/Objective: Mammographic differentiation of benign lesions from malignancies is a difficult task. We developed an artificial neural network (ANN) as a diagnostic aid in mammography using radiographic features as input.

Materials & Methods: A three-layered ANN was used to differentiate malignant from benign findings in a group of patients with proven breast lesions on the basis of morphological data extracted from conventional mammograms. Our database included 122 patient records on 14 qualitative variables. The database was randomly divided into training and validation samples including 82 and 40 patient records, respectively, to construct the ANN and validate its performance. Sensitivity, specificity, accuracy and receiver operating characteristic curve (ROC) analysis for this method and the radiologist were compared.

Results: Our results showed that the neural network model was able to correctly classify 30 out of 40 cases presented in the validation sample. Comparing the output with that of the radiologist, showed a reasonable diagnostic accuracy (75%), a moderate specificity (64%) and a relatively high sensitivity (89%).

Conclusion: A diagnostic aid was developed that accurately differentiates malignant from benign pattern using radiological features extracted from mammograms.

Keywords: breast neoplasms, computer, neural network

Introduction

Breast cancer is the most common cause of cancer deaths among women. Mammographic screening can reduce the mortality from breast cancer by as much as 20–30%, by detecting nonpalpable, noninvasive and early invasive tumors.[1-3] Early detection at the initial stage of breast cancer (i.e. DCIS)—in which the malignant process is confined within ducts and lobules, and its removal by surgical excision may cure the patient—could be performed using mammography findings.[1-3] Identification of DCIS is predominantly done based on the mammographic finding of clustered microcalcifications and characteristics of the mass.

DCIS consists up to 20% of mammographically detected cancers.[1-3] Accurate differential analysis between malignant and benign masses is very important, because it affects the patient management and choice of treatment. However, analysis is difficult in 40–50% of the cases where the features of the calcifications and their cluster are classified as indeterminate or equivocal. It may be the main reason for the obtained low positive predictive value (PPV): approximately 35% or less of women who undergo biopsy for a histopathologic of mammography of breast cancer diagnosis of breast cancer are found to have malignancies.[4] One goal of applying of computer-aided diagnosis (CAD) to mammography is to reduce the false-positive rate; avoiding benign biopsies which spares women the unnecessary discomfort, anxiety, and expense.

During the past decade, artificial neural network (ANN) has been intensively applied to radiological assessments to predict the biopsy outcome in breast

1. Assistant Professor, Department of Biophysics, Tarbiat Modarres University, Tehran, Iran.
2- Assistant Professor, Department of Radiology, Medical Imaging Center, Imam Khomeini Hospital, Tehran University of medical sciences, Tehran, Iran.
3. Department of Biophysics, Tarbiat Modarres University, Tehran, Iran.

Corresponding Author:
Parviz Abdolmaleki
Address:
Department of Biophysics, Faculty of Science, Tarbiat Modarres University, Tehran, Iran.
P.O. Box 14115/175
Tel: 009821- 88009730
Fax: 009821- 88011001
E-mail: parviz@modares.ac.ir

Received October 24, 2005;
Accepted after revision April 13, 2006.

Spring 2006; 3: 155-162

cancer. It involves processing various radiological features subjectively and/or objectively extracted from different modalities on the basis of defined criteria.[5-10] ANN is a computer algorithm capable of learning important relationships from a set of data, and applying this knowledge to evaluate new cases, as an expert system. ANN has two basic elements: processing elements and weighted connections. A collection of processing elements is defined as different layers including an input layer, one or more hidden layers, and an output layer. The connection weights store the information in the form of weight matrices.[11] The neural network learning procedure determines, in turn, the value of the connection weights.[12]

In this study, we intended to establish a neural network model to work as a tool for the radiologist to predict the biopsy outcome using data extracted from conventional mammography. The performance of the established model is then compared with that of the radiologist using common statistical indices including accuracy, sensitivity, specificity and receiver operating characteristic curve (ROC) analysis.

Materials and Methods

Our goal was to apply the neural network analysis to the data collected in a study designed to predict the malignancy of breast cancer on the basis of features extracted from conventional mammograms, using defined criteria. Our study group consisted of 122 consecutive patients (age range: 23-80 years; mean age: 49.7 years) with histopathlogical reports of their biopsies available. The patient group consisted of 51 malignant lesions and 71 benign entities. Most of the malignant cases were invasive carcinomas (n=32) with the majority being ductal carinomas (n=24); while most of the benign lesions were fibrocystic disease (n=39). Twelve (24%) of the malignant lesions were noninvasive.

Table 1 summarizes the distribution of lesions, histopathologically.

Data acquisition

Imaging was performed at the Imaging Center, Imam Khomeini Hospital, during 2000 to 2003.

Hook wire localization of microcalcifications under mammographic guidance was used in all cases. Two patients had three clusters, and nine patients had two clusters of microcalcifications on mammograms of the same breast, and seven patients had one cluster in each breast. The remaining 104 patients had only one cluster of microcalcifications in one breast. All lesions were histologically confirmed after biopsy or surgical excision. An expert radiologist (MG) read the mammogram images (Figure 1) and graded findings: mass size, shape, margins, density, asymmetric density, parenchymal distortion; calcification size, shape, number, density, distribution; general impression based on the microcalcification data and associated features. The presence of associated features was ranked on a scale of 0–4 with the increasing likelihood of malignancy for higher scores. In the case of more than one associated feature, the one with the highest rank was considered. The findings were ranked using a 2–5 scales with increasing likelihood of malignancy. Table 2 shows all the parameters (except age) in our database, which represented the subjective features extracted by the radiologist.

Neural Network Model

A feed-forward back-propagation artificial neural network (BP-ANN) can learn a function of mapping inputs to outputs by being trained with cases of input–output pairs.[11–13] The neural network which was

Table 1. Distribution of histopathologic lesions.

Histopathologic diagnosis	No. of lesions
Malignant	51
Intraductal carcinoma	19
Ductal carcinoma in situ (DCIS)	12
DCIS with microinvasion	7
Invasive carcinoma	32
Ductal with negative axilla	16
Ductal with invaded lymph nodes	8
Medullary	5
Lobular	2
Papillary	1
Benign	71
Fibrocytic disease	39
Proliferative	16
Nonproliferative	23
Mammary duct ectasia	3
Fibroadenoma	12
Papillomatosis	10
Fat necrosis	7

employed in this study had three layers. The first layer consisted of 14 input elements that each corresponded to data extracted from reading mammograms. The second layer, the hidden layer, had 8 nodes. Finally, the output layer had 1 element (Figure 2) that assigned 1 to malignant and 0 to benign lesions. Table 3 summarizes the neural network parameters.

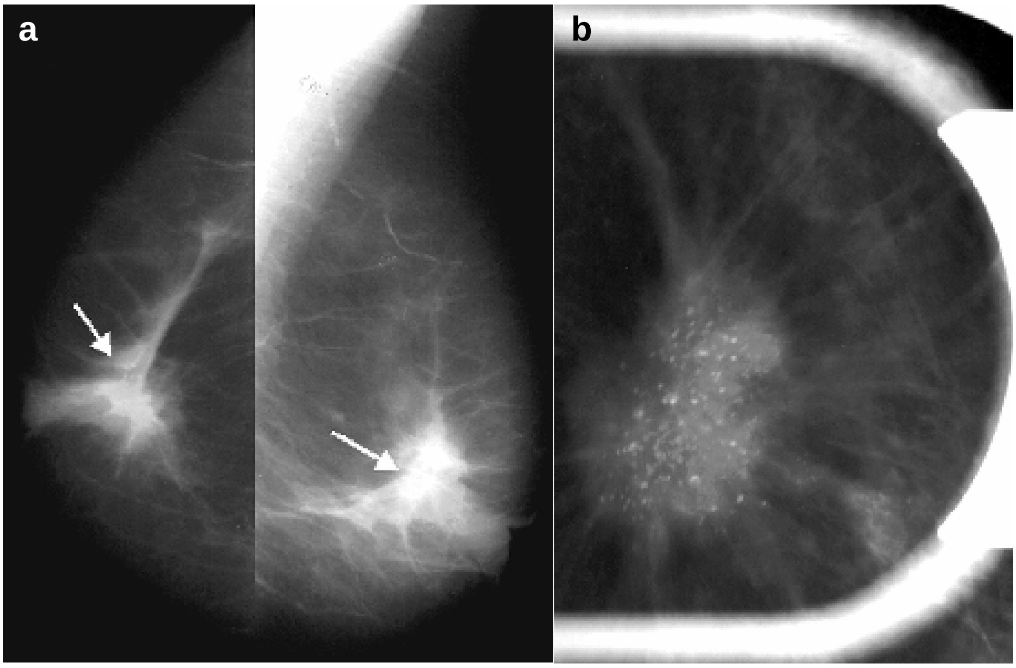

Fig 1. Mammogram displaying (a) mass (b) microcalcification

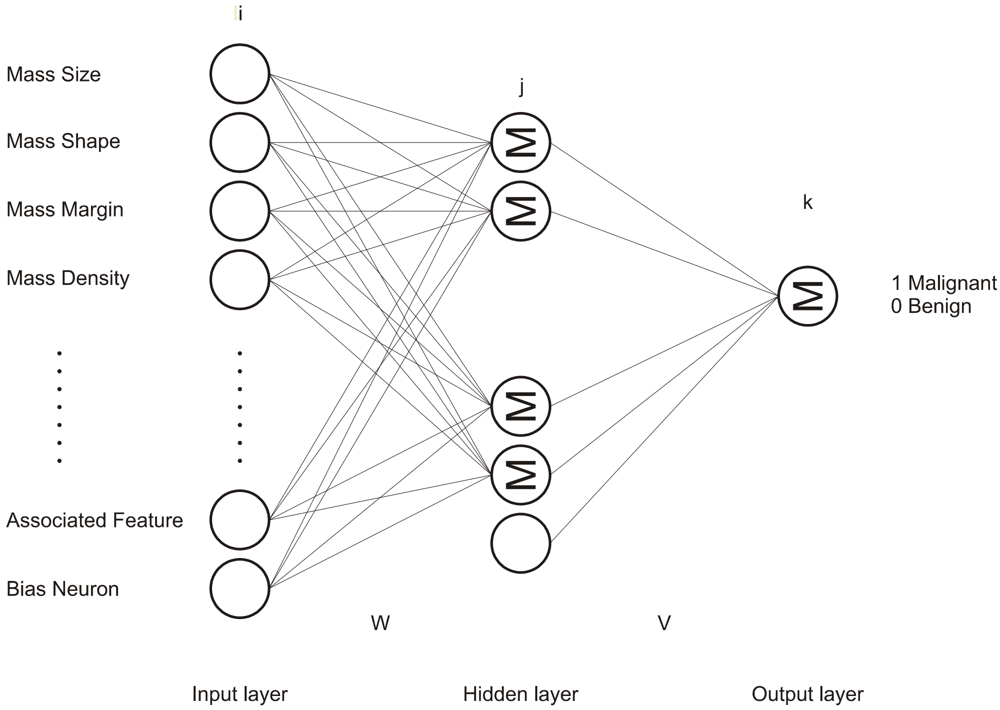

Fig 2. Schematic diagram showing the topology of a three-layer, feed-forward neural network with fourtheen input nodes, eight hidden nodes, and one output node that assinged 0 to benign and 1 to malignant tumors. Each node is connected to all nodes in the next layer through links with a weight matrix element. W and V refer to the weight matrices, and i, j, and k refer to the three layers. Bias neurons that only accept constant input (1) are placed in input and hidden layers in order to provide a non-zero output in the case of all zero inputs.

Generally, a neural network application learns in the same manner that a radiology resident does. The resident is shown an image (as input), and is initially told the right answer (output) by his or her attending radiologist. After the resident is presented with a large number of input images and output diagnoses, learning occurs and he or she gradually becomes a proficient radiologist. Similarly, the network is presented with a large number of cases (both input and output) and a learning algorithm is invoked. The learning algorithm gradually manipulates the network coefficients, changing the network connection weights, until the network becomes capable of responding with the correct output for a given input. Once the network is trained, in principle, the network would be proficient at a specified task. Hence, facing a problem never seen before by the neural network (as input), it can generate a suggestion or diagnosis.[11-13]

Finally, after the network had been trained perfectly in every simulation, the testing set was pre-

Table 2. Coding of the evaluated parameters from mammography of 122 patients, which used as input into the ANN during the training and validation procedures.

Radiol. Features	Findings	code
Mass Size	No mass	0
	Mass	size(mm)
Mass shape	No mass	0
	Circumscribed	1
	Lobulated	2
	Irregulare	3
	Spiculated	4
Mass margin	No mass	0
	Well-defined	1
	Obscured	2
	ill-defined	3
Mass density	No mass	0
	Low	1
	Mixed	2
	Medium	3
	High	4
Asymmetric density	Absent	0
	Difuse	1
	Segmental	2
	Focal	3
Parenchymal distortion	Absent	0
	Present	1
Calcification size	No	0
	Macro	1
	Micro-mono	2
	Micro Irregular	3
Calcification shape	No	0
	Rounded	1
	Punctuated	2
	Granular	3
	Rod-shape	4
	Y-shaped Mixed	5
Calcification number	No	0
	0-5	1
	5-10	2
	10-30	3
	>30	4
Calcification density	No	0
	Mono	1
	Irregular	2
Calcification distribution	No	0
	Scattered	1
	Group	2
	Cluster	3
	Mixed	4
Impression of radiologist based on the microcalcifications data	Benign	0
	Possibly benign	1
	Indeterminate	2
	Possibly malignant	3
	Malignant	4
Associated features	No	0
	Nipple retraction	1
	Skin thickening	2
	Skin retraction	3
	Axillary adenopathy	4

put vectors in the range of 0–1. Our network was trained perfectly over 200,000 iterations in each learning process within 30 minutes on a personal computer (Pentium 2.8 MHz, IBM compatible machine). The software used to construct the neural network was programed locally in MATLAB programming language.[14]

Practically, to establish the ANN, which could improve radiologist's performance in differentiating malignant from benign lesions, mammograms of a group of patients with hystopathologically proven breast lesions were used. Then, the above-mentioned features were obtained. Using a database of 122 cases, we randomly selected two-thirds (n=82) patients (including 49 benign and 33 malignant cases whose group identity was known) to compose the training sample. To prepare the validation sample, the rest of the (n=40) patients (22 benign and 18 malignant cases) were used.

Performance evaluation

To evaluate the performance of ANN and that of the radiologist, the receiver operating characteristic (ROC) analysis was used. We applied the ROCFIT software for Apple Macintosh based on the Charles E. Metz algorithm.[15] The ROC curve is a plot of sensitivity versus 1–(specificity). The area under the ROC curve is the most commonly used accuracy index. For a perfect classification accuracy, the ROC curve reaches sensitivity of 1.0 at a constant specificity of 1.0.[16]

After ANN had been perfectly established, the validation samples were presented to the network giving a posterior probability in the range of 0–1. Taking into consideration the posterior probability of malignancy, the diagnostic performance of this approach was estimated. In this regard, the true positive and the false-positive fractions were determined. These data were then used to plot the ROC curves. Ultimately, the area under the ROC curve (A_z) was used to compare the performance of the neural network method, with the expert radiologist's who participated in the testing (validating) procedure.

To evaluate the performance of the observer, the participating expert radiologist (MG) was asked to record her findings into one of the five categories, with increasing likelihood of malignancy: 1= benign,

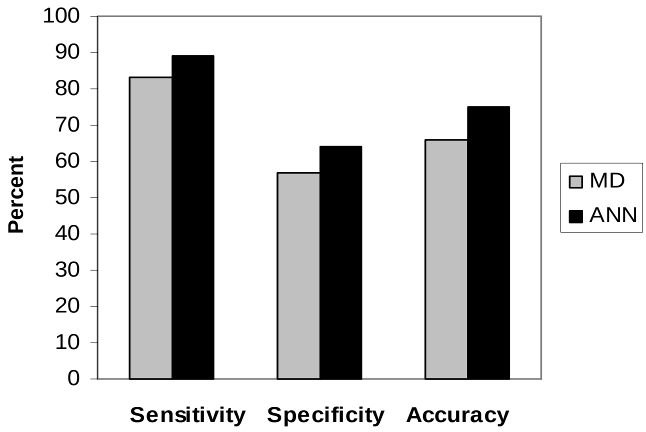

Fig 3. Comparative histogram of statistical indexes obtained for the expert radiologist (n=71) and the neural network (n=40).

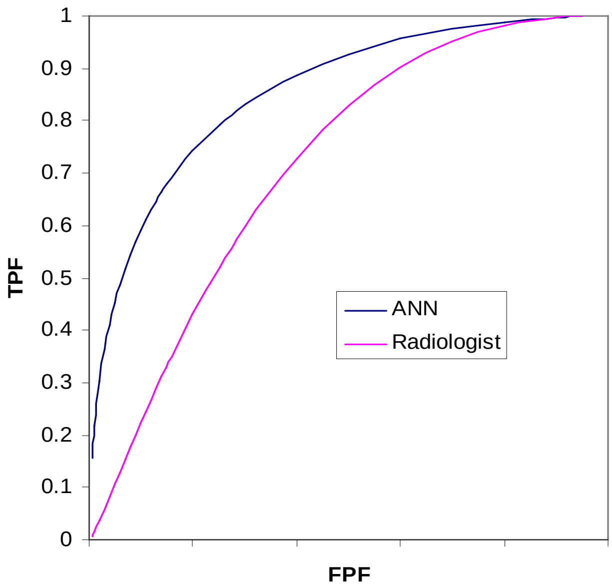

Fig 4. ROC curves for the expert radiologist and the neural network.

2= probably benign, 3= indeterminate, 4= probably malignant, 5= malignant. Similarly, to evaluate the performance of the neural network, the network output was classified into five categories: output in the range of (0–0.2) = benign, (0.2–0.4) = probably benign, (0.4–0.6) = equivocal, (0.6–0.8) = probably malignant, and output in range of (0.8–1) = malignant.

Results

The histological findings of biopsies were malignant in 51 cases (42%) and benign in 71 cases (58%). The most common malignant lesion was invasive ductal carcinoma and DCIS, while the most common benign lesions were fibrocystic disease (Table 1).

In our database, we had only 65 cases (53%) with tumor mass in which the tumor size ranged from 10 to 80 mm with a mean size of 45 mm. The remaining 57 cases (47%) had no mass while they had microcalcifications.

Radiologist's performance

The experienced radiologist read the images and classified them into benign and malignant groups using a five-scale category with increasing likelihood of malignancy. She could not reach a final diagnosis in 51 cases (41%) and simply classified them as indeterminate or equivocal.

The statistical results of sensitivity, specificity and accuracy obtained from the remaining cases (n=71) in which the radiologist could reach a final decision were 83%, 62% and 66%, respectively.

Artificial neural network

The output of the neural network on validation samples (n=40) showed correct classification in 16 of 18 patients with malignant breast cancers and 14 of 22 with a benign entity. The respective results of sensitivity, specificity and accuracy of 89%, 64%, and 75% obtained for the ANN were comparable to the results obtained for the expert radiologist, which were 84%, 57%, and 66%, respectively (Figure 3).

Analyzing the radiologist's impression showed that she made a wrong decision in 4 out of 24 cases of malignancy—indicating a false positive fraction of %17. Similarly she made a wrong decision in 18 out of 47 cases with a benign entity—a false negative fraction of 38%. Output of the ANN showed that it made a wrong suggestion in 8 out of 22 cases of malignancy—36% false positive, while its suggestion in benign cases was wrong in 2 out of 18 cases—a false negative of 11%.

Regardless of inequality in the size of the database, the ANN outperformed the radiologist by decreasing the false negative fraction from 38% to 11%. In contrast, the radiologist was sharp enough to outperform the ANN in terms of introducing a lower false positive fraction (17%) compared with that obtained from the ANN (36%).

We also applied ROC analysis as a measure of the discriminating ability of the ANN. Using the best results of the ANN and the radiologist, the ROC analysis was performed (Figure 4).

The obtained areas under the ROC curves (Az) are

Table 3. Neural network parameters

Index	Description
Number of input process elements	14, each corresponded to the evaluated featers, normalized from 0–1 for input.
Number of hidden process elements	8, optimized by trial and error.
Number of output process elements	1, range 0-1, trained to give 0 for benign and 1 for malignant.
Activation function	Sigmoid, 0 minimum, +1 maximum, gain 1.0.
Training algorithm	Back propagation, generalized delta learning rule.
Learning rate	0.3, optimized by trial and errors.
Momentum coefficient	0.9
Connected output to input	No
Initial connection weights	Randomly selected, range -0.5–0.5, standard deviation 0.55.
Structure	3 layers, feed-forward, fully connected.
Display mode	Real number.
Training time	Average 30 minutes.

presented in Table 4.

Discussion

Breast cancer remains the most frequently diagnosed and the second most lethal cancer for women in the world. Until an effective preventive measure becomes widely available, early detection followed by effective treatment is the only recourse for reducing breast cancer mortality. The important role that mammography is playing in breast cancer detection can be attributed largely to the technical improvements and dedication of radiologists to breast imaging. However, despite remarkable advances, it is known that these procedures are unable to detect all breast cancers due to problems associated with mammogram interpretation. These problems can be decomposed into two sub-problems. The first deals with the detection and localization of the regions of interest (ROIs), which include suspicious lesions. The second, and more difficult sub-problem, is the categorization of the identified lesions as malignant or benign. The differentiation is difficult due to the presence of considered overlap between benign and malignant patterns.[1-3, 17]

In this study, we designed a model based on the neural network analysis to increase the radiologist's ability to differentiate benign lesions from malignant tumors. The ability of the ANN was evaluated in a group of 122 patients with proven breast lesions; to investigate whether the ANN can improve the specificity while keeping high sensitivity. We hope to decrease the number of cases sent for biopsy; especially in a significant fraction of patients who are going under the biopsy procedure for apparently benign lesions.

Using the guidelines for features selection from the available literature, the parameters were evaluated by a participating radiologist with a high level of expertise. The extracted data were then presented to the established neural network. The average output of the ANN yielded a reasonable sensitivity (89%) and moderate accuracy (75%) comparable to the one obtained by the radiologist (84% and 66%). This finding demonstrates a consistent high sensitivity with a moderate specificity for the ANN in differentiating between benign and malignant breast tumors. The moderate-to-low specificity values obtained for the ANN may appear as a limitation of the ANN. However, this might be related to the overlap between

Table 4. Comparative performance of the ANN (n=40) and the radiologist (n=71)

Parameter	Expert radiologist	Neural network model
Sensitivity (%)	83	89
Specificity (%)	62	64
Accuracy (%)	69	75
False positive fraction	4 of 24	8 of 22
False negative fraction	18 of 47	2 of 18
Misclassified rate (%)	30	25
Area under the ROC curve	0.7151±0.0669	0.8536±0.0628

benign and malignant patterns. [17,18] The higher level of ANN performance may be related to the unique ability of the neural network in making associations between too many nonlinear and dependent parameters by addressing them as proportional weights. These weights, which are adjusted by the training procedure, are important for the neural network because they datermine the importance of each input element for internal calculation in the testing procedure. This provides little help for the radiologist who wants to clarify the relative prognostic importance of each feature. It means, although the ANN may work as an excellent predictor of malignancy, it may not be able to explain which findings are more relevant in reaching the diagnosis. This can be pointed out as another limitation for the ANN.

Review of the previous studies suggests that the accuracy, sensitivity and specificity of every diagnostic procedure are strongly dependent on the distribution of the benign and malignant patterns among their group of selected patients. Therefore, the obtained data by the ANN and the participated radiologist may not show their exact performance. To justify this point, we used ROC analysis for comparisons. By introducing a relative ROC area (Az) of 0.8536 for the ANN compared to 0.7151 obtained by the radiologist, the ROC analysis supported and enforced our results (Figure 4).

In a comprehensive view, we think that applying the ANN would be helpful to radiologist in differentiating malignant from benign tumors, to enhance diagnostic accuracy and consistency, and to decrease the rate of biopsy. Furthermore, using the ability of the ANN in making combinations between various subjective and objective features of different weights, tumor classification can be more objective, automated and probably more consistent. This would be better-highlighted if we trained the ANN by an experienced radiologist and use it for other radiologists with less experience. Therefore, an ANN trained by a highly experienced radiologist can be used as a back-up system to support the less experienced radiologist, especially in rural areas where they cannot reach a second reader for consensus. This application would be frequent for breast cancer, in which the radiologists label many cases as indeterminate.

In conclusion, we have established an ANN classifier which is able to predict the probability of malignancy of breast cancer using features extracted from the conventional mammogram. This model can be optimized to work better by using a combination of radiological findings from mammography, histomorphological parameters and clinical data in a large population of patients.

Acknowledgment

The authors thank the members of the Breast Imaging sections at Tehran Medical University Center. We also acknowledge Majid Roohandeh, MS, for his helpful assistance.

References

1. Coakley KS, Quintarelli F, van Doorn T, Hirst C. Classification of equivocal mammograms through digital analysis. The Breast 1994; 3: 222–226
2. Cronin KA, Yu B, Krapcho M, Miglioretti DL, Fay MP, Izmirlian et al. Modeling the dissemination of mammography in the United States. Cancer Causes Control 2005; 16: 701-712
3. Parker J, Dance DR, Davies DH, Yeoman LJ, Michell MJ, Humphreys S. Classification of DCIS by image analysis of calcification from digital mammograms. Br J Radiol 1995; 68: 150–159
4. Lev-Toaff AS, Feig SA, Saitas VL, Finkel GC, Schwartz GF. Stability of malignant breast microcalcifications. Radiology 1994; 192: 153–156
5. Markopoulos C, Kouskos E, Koufopoulos K, Kyriakou V, Gogas J. Use of artificial neural networks (computer analysis) in the diagnosis of microcalcifications on mammography. Eur J Radiol 2001; 39: 60–65
6. Abdolmaleki P, Buadu LD, Naderimanesh H. Feature extraction and classification of breast cancer on dynamic magnetic resonance imaging appearance using artificial neural network. Cancer letters 2001; 171: 183-191
7. Zhanga L, Sankara R, Qianb W: Advances in microcalcification clusters detection in mammography. Computers in Biology and Medicine 2002; 32: 515–528
8. Mavroforakis M, Georgiou H, Dimitropoulos N, Cavouras D,Theodoridis S: Significance analysis of qualitative mammographic features, using linear classifiers, neural networks and support vector machines. Eur J Radiol. 2005; 54: 80-89
9. Burnside ES, Rubin DL, Shachter RD: Using a Bayesian network to predict the probability and type of breast cancer represented by microcalcifications on mammography. Medinfo. 2004; 11: 13-17
10. Bocchi L, Coppini G, Nori J, Valli G: Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks. Med Eng Phys. 2004; 26: 303-312
11. Bishop CM: Neural networks for pattern recognition. Oxford: Oxford University Press; 1996
12. Wasserman PD. Neural computing: theory and practice. New York: Vannortrand reinhold Press; 1989: 10-42
13. Rumelhart DE, McClelland JL. Parallel distributed processing: explorations in the microstructures of cognition. Cambridge: MIT Press; 1986
14. MATLAB user's guide. Natick. Massachusetts: MathWorks Inc;1992
15. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. Invest. Radiol 1989; 24(3): 234-245

16. Metz CE. ROC methodology in radiologic imaging. Invest Radiol 1986 21: 720-733
17. Sickles EA. Breast calcifications: mammographic evaluation. Radiology 1986; 160: 289–93
18. Bushbinder SS, Leichter IS, Bamberger PN, Novac B, Lederman R, Fields S et al. Analysis of clustered microcalcifications by using single numeric classifier extracted from mammographic digital images. Acad Radiol 1998; 5: 779–784