Computer-Aided Diagnosis of Breast Cancer Using Artificial Neural Networks: Comparison of Backpropagation and Genetic Algorithms

Yuan-Hsiang Chang, Bin Zheng, Xiao-Hui Wang, Walter F. Good
Department of Radiology, University of Pittsburgh, Pittsburgh, PA 15261-0001
Schang@radserv.arad.upmc.edu

Abstract

The authors investigated computer-aided diagnosis (CAD) schemes to determine the probability for the presence of breast cancer using artificial neural networks (ANNs) that were trained by a Backpropagation (BP) algorithm or by a Genetic Algorithm (GA). A clinical database of 418 previously verified patient cases was employed and randomly partitioned into two independent sets for CAD training and testing. During training, the BP and the GA were independently applied to optimize, or to evolve the inter-connecting weights of the ANNs. Both the BP-trained and the GA-trained CAD performances were then compared using receiver-operating characteristics (ROC) analysis. In the training set, the BP-trained and the GA-trained CAD schemes yielded the areas under ROC curves of 0.91 and 0.93, respectively. In the testing set, both the BP-trained and the GA-trained ANNs yielded the areas under ROC curves of approximately 0.83. These results demonstrated that the GA performed slightly better, although not significantly, than BP for the training of the CAD schemes.

Introduction

Breast cancer is one of the major causes of death in women [1,2]. Although screening mammography has been demonstrated to effectively detect early breast cancer [3], differentiating between malignant and benign cases is still challenging for radiologists. Computer-aided diagnosis (CAD) schemes for breast cancer have been investigated with the hope to provide "second opinion" to radiologists so that their diagnostic accuracy and efficiency can be significantly improved [4-7]. To date, many CAD schemes have been investigated to identify suspicious regions for specific breast abnormalities (e.g., microcalcifications and masses) on mammograms [8-12]. Such CAD schemes tend to focus on mammographic findings, while radiologists tend to integrate not only mammographic findings but also patient-related information (e.g., patient's medical history, and physical findings) in breast cancer diagnosis.

A number of researchers have investigated various approaches to integrate mammographic findings, as well as patient-related information, in developing CAD schemes. Artificial neural networks (ANNs) were most commonly used and have been demonstrated to be potentially useful for solving medical problems and decision making [13-16]. For breast cancer diagnosis, Lo JY *et al.* investigated an ANN on the basis of 8 mammographic findings and patient age [17]. Baker JA *et al.* investigated an ANN using the Breast Imaging Recording and Data System (BI-RADS) with 10 lesion descriptors and 8 inputs from patient's medical history [18]. In addition to ANN applications, Kahn *et al.* explored a Bayesian network (MammoNet) that integrated 5 patient-history features, 2 physical findings, and 15 mammographic findings to determine the probability of breast cancer malignancy [19].

Genetic algorithms (GAs) have been widely used in science as adaptive algorithms for solving practical problems such as optimization and machine learning [20-22]. Because GAs can be used to solve problems with different search spaces and parameters, GAs have also been widely used in biology and medical applications [23-26]. For breast cancer diagnosis, Sahiner *et al.* investigated a new approach that included a genetic algorithm for image feature selection, and a linear discriminant classifier or a backpropagation neural network in the task of differentiating regions of interest (ROIs) on mammograms as either mass or normal tissue [27]. They concluded that GAs provide versatility in the design of linear or nonlinear classifiers without a trade-off in the effectiveness of the selected features.

In this study, we explored an ANN using 13 features (5 patient-history features, 4 physical findings, and 4 mammographic findings) to determine the probability for the presence of breast cancer. First, a clinical database of 418 cases was collected and randomly partitioned into two independent sets for training and testing of the ANN. Then, the ANN was trained using either a conventional backpropagation algorithm (BP) or a genetic algorithm

0-7803-5529-6/99/$10.00 ©1999 IEEE

(GA). Finally, both the BP-trained and the GA-trained ANNs were evaluated and compared using the receiver-operating characteristics (ROC) analysis [28,29]. Preliminary results are presented.

Materials and Methods

A clinical database of 418 patient cases was employed, which had been previously collected and verified by radiologists with pathology reports. All the 418 cases were acquired during breast examinations at the University of Pittsburgh Medical Center (UPMC) and its affiliate hospitals. Within this database, 92 cases were previously diagnosed as "positive" for breast cancer, while 326 cases were previously diagnosed as "negative". "Negative" cases were verified by radiologists with at least two-year follow-up negative mammogram readings. Within each case, 5 patient-history features, 4 physical findings, and 4 mammographic findings were collected and used as inputs to the CAD scheme to determine the probability for the presence of breast cancer. Table 1 summarizes definitions of all the features (and findings) and their states.

Figure 1 is a schematic diagram of the CAD algorithm for breast cancer diagnosis, which uses ANNs that were trained by backpropagation or genetic algorithms. All input nodes (besides microcalcifications and masses) to the ANN were assigned as either 1 (Yes or Present) or 0 (No or Absent). Regardless of the number of suspicious regions as depicted on mammograms, input nodes for microcalcifications or masses to the ANN were assigned as 1 if present with level of concerns 4-5; 0.5 if present with level of concerns 1–3; or 0 if absent. Here, the levels of concerns were subjectively rated by radiologists, where level 5 was the most suspicious and level 1 was the least suspicious for malignancy of the microcalcifications or masses. Target output node of the ANN was assigned as 1 to indicate the presence of breast cancer, or 0 its absence.

In this study, the ANN was a multi-layer, feed-forward neural network with inter-connecting weights. The following logistic activation function was used for each of the processing units in the ANN:

$$O_{pj} = \frac{1}{1+e^{-(\sum_i W_{ji} O_{pi} + \theta_j)}}$$

where O_{pj} is the *j-th* unit of the output pattern produced by the input pattern p, W_{ji} is the weight from the *i-th* to *j-th* units and θ_j is the threshold of the *j-th* unit. The error measure between the target and actual output valued summed over the output units is defined to be:

$$E_p = \frac{1}{2}\sum_j (t_{pj} - O_{pj})^2$$

Table 1. Definitions of features (findings) and their states.

Category	Node Description	State Description
Diagnosis	Breast cancer	Present, Absent
Patient History	Family member has breast cancer?	Yes, No
	Have ever been pregnant?	Yes, No
	Have gone through menopause?	Yes, No
	Taking female hormone?	Yes, No
	Drinking alcohol or smoking?	Yes, No
Physical Findings	Skin change (Retraction)?	Yes, No
	Nipple Discharge	Yes, No
	Pain	Yes, No
	Have Lump(s)	Yes, No
Mammographic Findings	Architecture Distortion	Present, Absent
	Microcalcifications	Level of concerns: 4 – 5 Level of concerns: 1 – 3 Absent
	Masses	Level of concerns: 4 – 5 Level of concerns: 1 – 3 Absent
	Asymmetry	Present, Absent

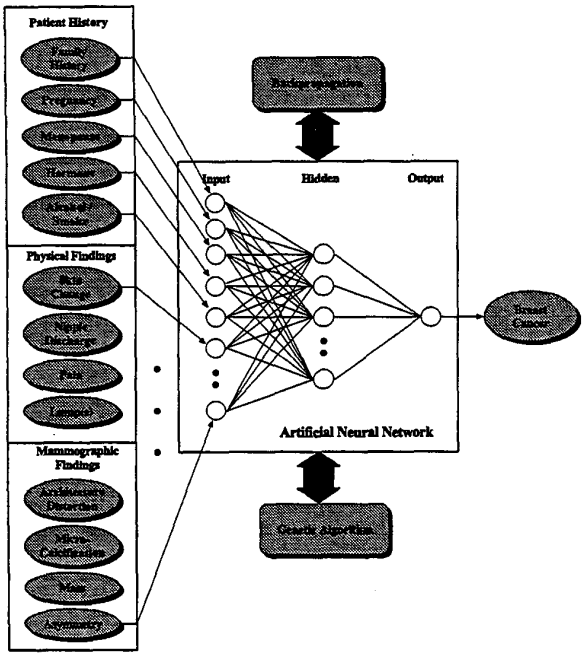

Figure 1. A schematic diagram of computer-aided diagnosis (CAD) for breast cancer using artificial neural networks trained by backpropagation or genetic algorithms.

while $E = \sum E_p$ is the overall error measure calculated by summing all error measures in the training set. In this preliminary study, a three-layer ANN including 13 input units, 4 hidden units, and 1 output unit was empirically chosen. A value ranging from 0 to 1, for the output unit, was used to indicate the probability for the presence of breast cancer. Both the backpropagation and the genetic algorithm were independently used to optimize, or to evolve, the inter-connecting weights in the ANN. These are described as follows.

A. Backpropagation Algorithm

Backpropagation algorithms have been commonly used for the training of ANNs [30,31]. In this process, weights between the processing units are iteratively adjusted so that the overall error measure E is minimized. This can be implemented by:

$$\Delta W_{ji}(n+1) = \eta(\delta_{pj} O_{pi}) + \alpha \Delta W_{ji}(n)$$

where η is the learning rate, n is the number of iterations, α is the momentum, and δ_{pj} is the error signal. Although the backpropagation algorithm can be used to provide a solution to the problem of training ANNs, the algorithm may converge to a local minimum that results in a sub-optimal solution.

B. Genetic Algorithm

Genetic algorithms can be used to determine the inter-connecting weights of the ANN (i.e., to evolve weights in a fixed-structured, three-layer ANN). Similar to the genetic algorithm as described by David Montana and Lawrence Davis [20,21], we applied a genetic algorithm as follows:

Step 1. Let a chromosome be a "vector" of all the inter-connecting weights of the ANN. Initialize a population of chromosomes (i.e., weight vectors) with each weight being between –1.0 and + 1.0.

Step 2. Evaluate the fitness of each chromosome in the population. In this study, maximum fitness was equivalent to minimum overall error measure E in the training set. Then, apply "Roulette Wheel Parent Selection" to choose parent chromosomes for mating [20].

Step 3. Apply the crossover operation by taking two parent chromosomes (Parent 1 and 2) to produce two offspring chromosomes (Child 1 and 2). First, copy all the weights to the output unit of Child 1 from Parent 1; and Child 2 from Parent 2, respectively. Then, copy all the weights to the odd and even hidden units of Child 1 from those of Parent 1 and 2, respectively. Alternatively, copy all the weights to the odd and even hidden units of Child 2 from those of Parent 2 and 1, respectively.

Step 4. Apply the mutation operation by randomly selecting a non-input unit and, for each incoming weight to the unit, add a random value between –1.0 and + 1.0 to the weight.

Step 5. Delete members of parent chromosomes to make room for offspring chromosomes. Evaluate the offspring chromosomes and insert them into the population.

Step 6. Increase generation by one. Repeat step 2 through 5 until a specific generation has been reached.

GAs were applied to evolve the inter-connecting weights in the ANN, although GAs may only yield near-optimal solutions because of the large search space (multi-dimensional error surface).

C. Training and Testing

To assess whether the ANN can learn from known cases and correctly classify unknown cases, a clinical database with 92 positive and 326 negative cases was used and randomly partitioned into two independent sets for training

and testing, each with 46 positive and 163 negative cases. The training set was used to determine the optimal set of inter-connecting weights of the ANN, while the test set was used to evaluate the performance of the trained ANN.

During training of the network, the BP required approximately two ANN evaluations (i.e., one forward propagation and one backward error propagation) for each iteration, while the GA required only one ANN evaluation (i.e., forward propagation) for each generation and each chromosome. Therefore, the parameters for the BP and the GA were chosen so that the total number of ANN evaluations were identical for comparison (i.e., total number of iterations $n \times 2$ in BP = total number of populations × total number of generations in GA). Preliminary results from the BP-trained ANN and the GA-trained ANN are presented here.

Results

Figure 2 and 3 show the ROC curves for the BP-trained and the GA-trained ANNs, respectively. The parameters for BP training were chosen as follows: the learning rate $\eta = 0.01$, the momentum $\alpha = 0.8$, and the number of iterations $n = 2500$. The parameters for GA training were chosen as follows: the total number of population = 20 and the total number of generations = 250. Notice that the total number of ANN evaluations for both the BP and the GA were equivalent (i.e., 5000 ANN evaluations). The BP-trained ANN yielded the areas under the ROC curves (Az) of 0.91 ± 0.018 in the training set, and 0.83 ± 0.028 in the test set. The GA-trained ANN yielded the areas under the ROC curves (Az) of 0.93 ± 0.016 in the training set, and 0.83 ± 0.021 in the test set.

Figure 4 shows the areas under ROC curves for both the BP-trained and the GA-trained ANNs at selected numbers of ANN evaluations (i.e., multiples of 1000). As expected, the areas under ROC curves for the training set increased as the number of ANN evaluations increased. But, the areas under the ROC curves for the test set decreased as the number of ANN evaluations increased. With less than about 3000 ANN evaluations, the ANN may be "under-trained"; while with more than about 7000 ANN evaluations, the ANN may be "over-trained". The optimal performance of the CAD scheme was achieved at about 5000 ANN evaluations.

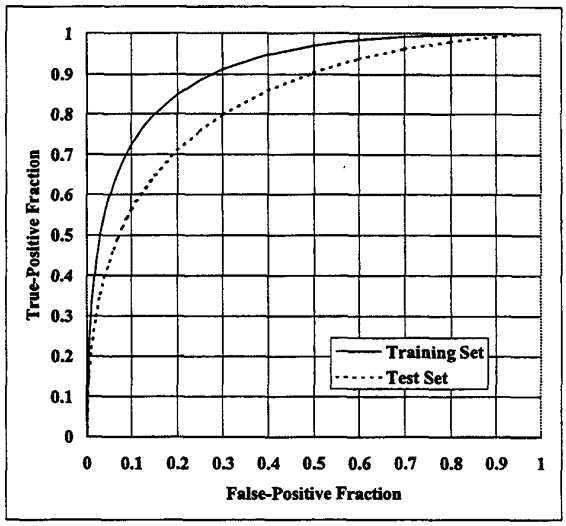

Figure 2. ROC analysis of the BP-trained ANN for both the training and the test sets, each with 46 positive and 163 negative cases.

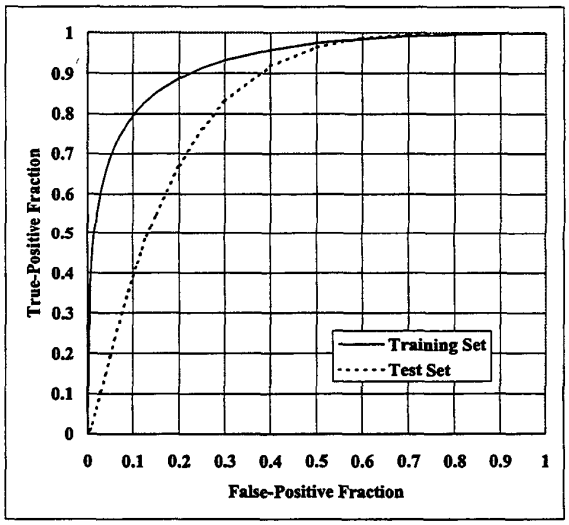

Figure 3. ROC analysis of the GA-trained ANN for both the training and the test sets, each with 46 positive and 163 negative cases.

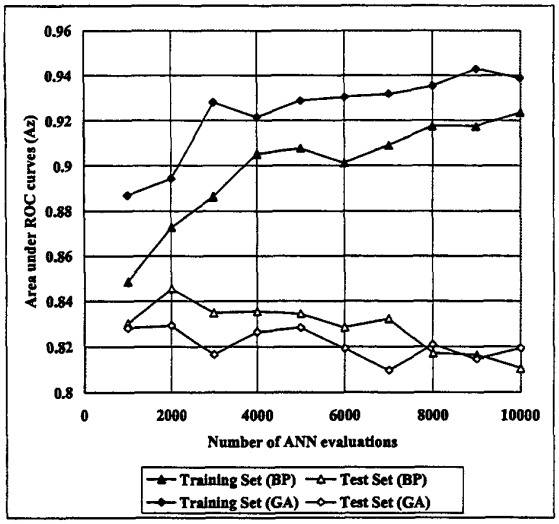

Figure 4. Results of areas under ROC curves (Az) for the BP-trained and the GA-trained ANNs at selected number of ANN evaluations (multiples of 1000). Both the training set and the test set included 46 positive and 163 negative cases.

Discussion

We have investigated CAD schemes using ANNs to determine the probability for the presence of breast cancer based on 13 features (i.e., 5 patient-history features, 4 physical findings, and 4 mammographic findings). Although other features may be available through patients' questionnaires (e.g., age, or previous biopsy), and/or other findings (e.g., glandular or fatty breast tissue, halo sign, or tumor locations), these 13 features were empirically chosen as inputs to the CAD scheme. Although it is not obvious how the subjective ratings (i.e., levels of concerns) by radiologists contributed to the overall CAD of breast cancer, because of the complicated inter-connections within the ANN, preliminary results are encouraging.

In comparison to the conventional BP training algorithm, the GA was shown to provide some benefit in evolving the inter-connecting weights for the ANNs. Although the GA-trained ANN didn't outperform the BP-trained ANN at all numbers of ANN evaluations in the test set, the GA-trained ANN was found to converge faster than the BP-trained ANN in the training set. We believe that the GAs may have advantages over conventional BP training techniques depending on the specific problem being addressed. For example, the GA can be adopted to minimize the overall error measure or to maximize the area under the ROC curve as required in clinical situations; while traditionally the BP algorithm has been limited to minimizing the overall error measure with respect to the weights of the ANN.

Even with the number of training cases larger than the number of inter-connecting weights, the ANN may be "over-trained". Although such an over-trained ANN can perform better in the training set, the ANN may not necessarily perform more robustly in the test set. The ANN used in this study included a total of 61 weights (52 weights from input to hidden layers, 4 weights from hidden to output layers, and 5 thresholds). And, the number of cases used to train the ANN was 209 (46 positive and 163 negative cases). Because cases used to train and test a CAD scheme can significantly affect the test results [32], a large variety of both positive and negative cases must be evaluated before the CAD scheme can be widely used. In addition, the adequacy of the size of both the training set and the test set must be evaluated to assure the scheme's consistency and generalizability [33]. The use of either BP or GA training, however, may not assure the robustness of the CAD in this regard.

Future efforts to improve the performance of CAD for breast cancer may include the following: 1) *Feature selection*: An optimal set of features from a variety of patient information and findings must be chosen as inputs to the CAD scheme. The use of GAs is very promising in feature selections [34]; 2) *Structure of the network*: The use of different architectures for the network may emphasize independent features while de-emphasize correlated features. For example, it may be desired to investigate a combined network with three independent sub-ANNs that are optimized on the basis of patient-history features, physical findings, and mammographic findings. Then, the combined ANN with the three sub-ANNs may be easily analyzed for human reasoning and understanding of breast cancer, although such an issue is beyond the scope of this preliminary study.

Acknowledgements

The authors are grateful to Charles E. Metz, PhD, for the LABROC programs for ROC analysis. This study was supported in part by the NIH Grants CA77850 and CA79587.

References

[1] Boring CC, Squires TS, Tong T. Cancer statistics 1993. CA: *Cancer J. Clinicians* 1993; 43:7-26.

[2] Wingo PA, Tong T, Bolden S. Cancer statistics 1995. CA: *Cancer J. Clinicians* 1995; 45:8-30.

[3] Feig SA. Decreased breast cancer mortality through mammographic screening: Results of clinical trials. *Radiology* 1988; 167:659-665.

[4] Chan HP, Doi K, Vyborny CJ, et al. Improvement in radiologists' detection of clustered microcalcifications on mammograms: The potential of computer-aided diagnosis. *Invest Radiol* 1990; 25:1102-1110.

[5] Kegelmeyer WP, Pruneda JM, Bourland PD, Hillis A, Riggs MW, Nipper ML. Computer-aided mammographic screening for spiculated lesions. *Radiology* 1994; 191:331-337.

[6] Vyborny CJ, Giger ML. Computer vision and artificial intelligence in mammography.*AJR* 1994;162:699-708.

[7] Karssemeijer N, Hendriks JH. Computer-assisted reading of mammograms. *Eur Radiol* 1997; 7:743-748.

[8] Chan HP, Doi K, Vyborny CJ, Lam KL, Schmidt RA. Computer-aided detection of microcalcifications in mammograms: Methodology and preliminary clinical study. *Invest Radiol* 1988; 23:664-671.

[9] Lefebvre F, Benali H. A fractal approach to the segmentation of microcalcifications in digital mammograms.*Med Phys* 1995; 22:381-390.

[10] Zheng B, Chang YH, Gur D. Computerized detection of masses in digitized mammograms using single image segmentation and a multi-layer topographic feature analysis. *Acad Radiol* 1995; 2:959-966.

[11] Petrick N, Chan HP, Wei D, Sahiner B, Helvie MA, Adler DD. Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification.*Med Phys* 1996;23:1685-1696.

[12] Chang YH, Zheng B, Gur D. Computerized Identification of suspicious regions for masses in digitized mammograms. *Invest Radiol* 1996; 31:146-153.

[13] Piraino DW, Amartur SC, Richmond BJ, Schils JP, Thome JM, Belhobek GH, Schlucter MD. Application of an artificial neural network in radiographic diagnosis.*J Digital Imaging* 1991; 4:226-232.

[14] Wu Y, Doi K, Metz CE, Asada N, Giger ML. Simulation studies of data classification by artificial neural networks: Potential applications in medical imaging and decision making. *J Digital Imaging* 1993; 6:117-125.

[15] Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE. Artificial neural networks in mammography: application to decision making in diagnosis of breast cancer. *Radiology* 1993; 187:81-87.

[16] Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Harrell FE Jr, Marks JR, Winchester DP, Bostwick DG. Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer* 1997; 79:857-862

[17] Lo JY, Baker JA, Kornguth PJ, Floyd CE Jr. Computer-aided diagnosis of breast cancer: artificial neural network approach for optimized merging of mammographic features. *Acad Radiol* 1995; 2:841-850.

[18] Baker JA, Kornguth PJ, Lo JY, Williford ME, Floyd CE Jr. Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology* 1995; 196:817-822.

[19] Kahn CE Jr., Roberts LM, Shaffer KA, Haddawy P. Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Comput Biol Med* 1997; 27:19-29.

[20] Davis L (Editor). *Handbook of genetic algorithms*. (Van Nostrand Reinhold, New York, 1991).

[21] Mitchell M. *An introduction to genetic algorithms*. (The MIT Press, Cambridge, Massachusetts, 1997).

[22] Luger GF, Stubblefield WA. *Artificial intelligence*, Third Edition. (Addison Wesley, Reading, Massachusetts, 1998).

[23] Levin M. Use of genetic algorithms to solve biomedical problems. *MD Comput* 1995; 12:193-199.

[24] Park LJ, Park CH, Park C, Lee T. Application of genetic algorithms to parameter estimation of bioprocesses. *Med Biol Eng Comput* 1997; 35:47-49.

[25] Baumgart-Schmitt R, Herrmann WM, Eilers R, Bes F. On the use of neural network techniques to analyze sleep EEG data. First communication: application of evolutionary and genetic algorithms to reduce the feature space and to develop classification rules. *Neuropsychobiology* 1997; 36:194-210.

[26] Baumgart-Schmitt R, Herrmann WM, Eilers R. On the use of neural network techniques to analyze sleep EEG data. Third communication: robustification of the classificator by applying an algorithm obtained from 9 different networks. *Neuropsychobiology* 1998; 37:49-58.

[27] Sahiner B, Chan HP, Wei D, Petrick N, Helvie MA, Adler DD, Goodsitt MM. Image feature selection by a genetic algorithm: application to classification of mass and normal breast tissue. *Med Phys* 1996; 23:1671-1684.

[28] Swets JA. ROC analysis applied to the evaluation of medical imaging techniques.*Invest Radiol* 1979; 14:109-121.

[29] Metz CE. ROC methodology in radiologic imaging. *Invest Radiol* 1986;21:720-733.

[30] Hecht-Nielsen R. *Neurocomputing*. (Addison Wesley, Reading, New York, 1998).

[31] Rumelhart DE, McClelland JL, the PDP Research Group. *Parallel distributed processing*, Volume 1 (The MIT Press, Cambridge, Massachusetts, 1986).

[32] Nishikawa RM, Giger ML, Doi K, Metz CE, Yin FF, Vyborny CJ, Schmidt RA. Effect of case selection on the performance of computer-aided detection schemes. *Med Phys* 1994; 21:265-269.

[33] Zheng B, Chang YH, Good WF, Gur D. Adequacy testing of training set sample sizes in the development of a computer-assisted diagnosis scheme.*Acad Radiol* 1997; 4:497-502.

[34] Zheng B, Chang YH, Wang XH, Good WF, Gur D. Feature selection for computerized mass detection in digitized mammograms using a genetic algorithm.*Acad Radiol* 1999, in press.