Comparative Analysis of Logistic Regression and Artificial Neural Network for Computer-Aided Diagnosis of Breast Masses[1]

Jae H. Song, MS, Santosh S. Venkatesh, PhD, Emily A. Conant, MD, Peter H. Arger, MD, Chandra M. Sehgal, PhD

Rationale and Objective. To compare logistic regression and artificial neural network for computer-aided diagnosis on breast sonograms.

Materials and Methods. Ultrasound images of 24 malignant and 30 benign masses were analyzed quantitatively for margin sharpness, margin echogenicity, and angular variation in margin. These features and age of patients were used with two pattern classifiers, logistic regression, and an artificial neural network to differentiate between malignant and benign masses. The performance of two methods was compared by receiver operating characteristic (ROC) analysis.

Results. The area under the ROC curve *Az* (±SD) of the logistic regression analysis was 0.853 ± 0.059 with 95% confidence limit (0.760–0.950). The area under the ROC curve of the artificial neural network analysis was 0.856 ± 0.058 with 95% confidence limit (0.734–0.936). Although both the logistic regression and the artificial neural network had the same area under the ROC curve, the shapes of two curves were different. At 95% sensitivity, the artificial neural network had 76.5% specificity, whereas logistic regression had 64.7% specificity.

Conclusion. There was no difference in performance between logistic regression and the artificial neural network as measured by the area under the ROC curve. However, at a fixed 95% sensitivity, the artificial neural network had higher (12%) specificity compared with logistic regression value.

Key Words. Breast cancer; breast ultrasound; sonography; logistic regression; artificial neural network; cancer diagnosis.

Breast cancer is one of the most common cancers in women (1). Sonography is now commonly used in combination with other modalities for imaging breasts. Although ultrasound can diagnose simple cysts in the breast with an accuracy of 96%–100%, its use for unequivocal differentiation between solid benign and malignant masses has proven to be more difficult. Despite considerable efforts toward improving imaging techniques, including sonography, the final confirmation of whether a solid breast lesion is malignant or benign is still made by biopsy.

Interpreting ultrasound breast images is almost exclusively done by qualitative assessment of lesion features within an image. Variations in human perception and the lack of standard definitions of the features observed in the images contribute to the variability in diagnosis. There is growing interest in using computer-based "intelligent" algorithms to recognize and classify breast masses as malignant or benign (2–10). The expectation is that such systems will help improve observers' diagnostic performance and reduce the number of biopsies without increasing the number of missed cancers.

Acad Radiol 2005; 12:487–495

[1] From the Department of Radiology (E.A.C., P.H.A., C.M.S.) and Department of Electrical Engineering (J.H.S., S.S.V.), Pennsylvania Medical Center, 3400 Spruce Street, Philadelphia, PA 19104. Received September 17, 2004; revised and accepted December 18, 2004. Supported in part by NIH grants CA85424 and CA87526. **Address correspondence to** C.M.S. e-mail: sehgalc@uphs.upenn.edu

doi:10.1016/j.acra.2004.12.016

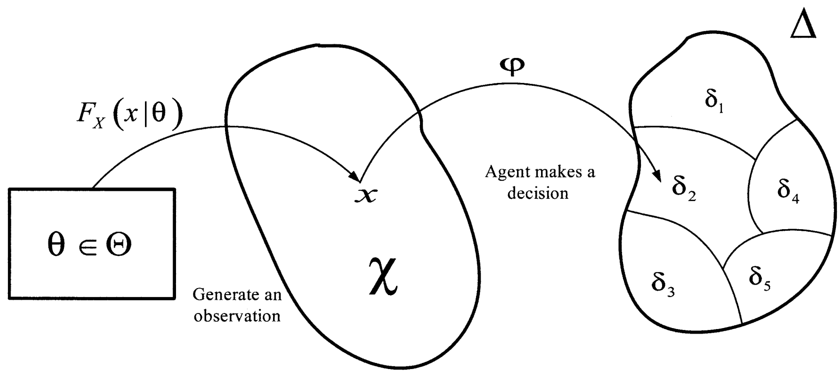

Figure 1. Diagram showing the mapping behavior of artificial neural networks. The sample space is mapped to the decision space according to the decision rule φ. This diagram is adapted from (37).

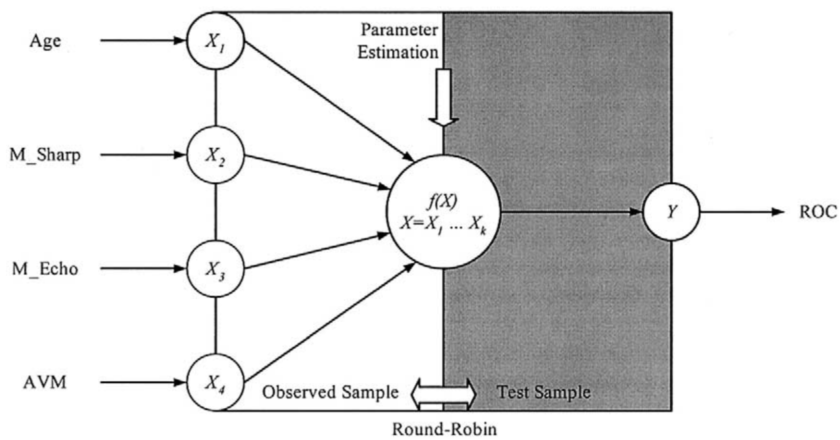

Figure 2. Diagram showing the logistic regression procedure. The parameters α and β_i are used to estimate probability of malignancy (P_m) of the test sample.

In recent years, considerable attention has been given to using computer-based methods to classify medical data according to different diseases and outcomes. The general approach has been to develop computer algorithms that learn decision characteristics for data classification and then use them to classify future patients with unknown disease states or therapy outcomes. Several quantitative models ranging from simple linear discriminant analysis to more complex logistic or Cox regression analysis and neural networks have been proposed. Unfortunately, there is no well-accepted theory to guide the choice of model based on the complexity and the nature of the diagnostic task (11). A frequently used approach is to pick a single quantitative model that yields satisfactory results. There is now an increasing effort to compare different quantitative models for specific data classification tasks (12–24). Sargent performed meta-analysis comparing artificial neural networks (ANN) with regression models (12) in 28 studies and found that in 36% of cases, ANN, in 14% regression methods, performed better and in the remaining 50% of cases both modes had similar performance. In another study, Dreiseitl and Ohno-Machado surveyed 72 articles comparing ANN with logistic regression (LR) (13). They observed that although ANN and LR had better discriminating power in 18% and 1% of cases, respectively, in 42% of cases, there was no difference between the two models. The remaining 39% of the articles did not have adequate statistical testing and they were not considered for analysis. In summary, there is no universal approach for selecting one model over the others for data classification; each task must be evaluated on a case-by-case basis.

Recently, we proposed the use of logistic regression to classify breast masses (25). Although LR can be used to predict the probability of malignancy, there is also considerable interest in using artificial neural networks for the same purpose (2,26). ANN is based on finding an opti-

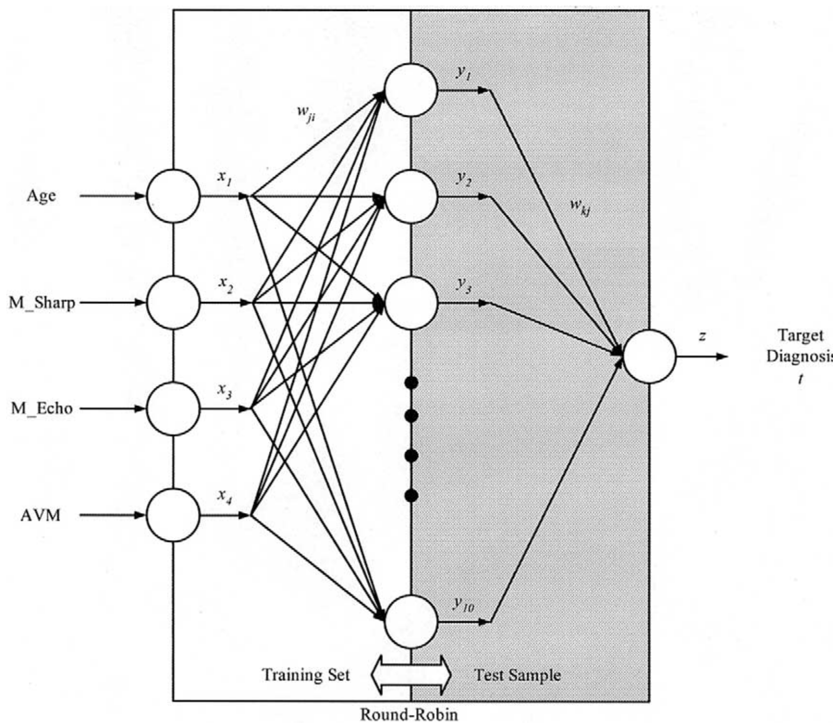

Figure 3. Diagram showing the artificial neural network procedure. The network is trained with the training set to set up the weight connections appropriately, and the output of the test sample is calculated by applying those weight values.

mal path from the sample space to decision space (Fig. 1). The choices of nature are elements θ of the parameter space È, and the decision space Δ contains the possible decisions δ. The decision rule φ provides the relation between the observation and the decisions, mapping the samples to the decision regions. In this study, we used a supervised learning paradigm for ANN. This involves feeding unique input samples (image features) and the matching responses (outcomes) to let the network learn from the examples and compose a map that interrelates inputs to outputs through a complex set of interconnecting pathways or operations. The input-output map is the major characteristic of ANN, which in turn is used to predict the outcome when presented with a new input. Unlike logistic regression, which fits the data to a descriptive function, in ANN the input data is transformed on each layer, changing its dimensional space to define the rule to get to the decision region. Thus the two approaches are inherently different, raising the question if one approach has better diagnostic performance than the other.

To investigate this question, we compared the performance of LR and ANN as pattern classifiers for sonographic features. Both the methods were tested using the same set of images and image features.

MATERIALS AND METHODS

From a library of images of suspicious masses, 54 cases (24 malignant and 30 benign) were identified that had excision or core biopsy and ultrasound imaging in both radial and antiradial planes. The studies were approved by the institutional review committee.

Seven features representing the shape and the margin characteristics of the mass (margin sharpness, margin echogenicity, angular continuity, tissue attenuation, mass attenuation, and excess attenuation) and patients' ages were measured quantitatively from sonograms (25). The features were analyzed by LR and ANN. The computer programs for feature extraction, LR and ANN, were developed using the IDL platform (Research System Inc., Denver, CO).

Feature Analysis

Arithmetic mean and standard deviation of each feature was derived for the malignant and benign groups. A two-tailed Student *t*-test of unequal variance was used to determine the statistical significance of the difference between the two groups. Four of seven extracted features showed a significant difference between the benign and

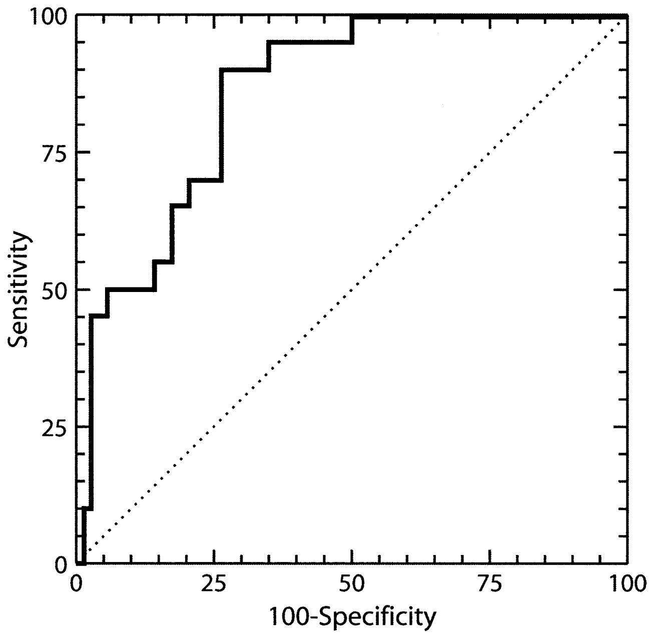

Figure 4. Graph showing receiver operating characteristics (ROC) of the logistic regression analysis. The area under the ROC curve, A_z, was 0.853 ± 0.059 with 95% confidence limit (0.76–0.95); the curve shows 64.7% specificity at 95% sensitivity.

malignant groups. Only these four features were used as inputs for the LR and ANN analyses. These features, described earlier, included patients' age (age) and three ultrasound image features: margin sharpness (M-sharp), margin echogenicity (M-echo), and angular variance in margin (AVM). M-sharp represents the diffuseness of the border of a mass, M-echo represents the difference of brightness (measured in difference in mean grayscale level) between the inside and outside of a mass at the margin, and AVM quantifies the inhomogeneity in margin brightness with angle. The same images were used to evaluate the performance of both LR and ANN.

Logistic Regression

From a pool of n samples, n-1 were fitted to a LR model defined by $\log \text{it}(P_m) = \alpha + \sum_{i=1}^{4} \beta_i X_i$, where P_m is the probability of malignancy, and α and β_i are unknown parameters that determine the shape of the logistic curve as described in Fig. 2. The unknown parameters α and β_i were estimated using a maximum-likelihood approach. Using the estimated parameters $\hat{\alpha}$ and $\hat{\beta}_i$, the probability of malignancy for the test sample (nth sample) was calculated. This procedure was repeated using round-robin substitution until the probability of malignancy for every sample in the data was determined.

ANN

A multilayer perceptron model of ANN using sigmoid function was constructed. The network consisted of an input layer, a hidden layer, and an output layer. The input layer contained 4 units (neurons) corresponding to four input features; the hidden layer contained 10 units transforming the input features from 4-dimensional to 10-dimensional space. Finally, the output layer had only one neuron, representing two possible diagnostic states: malignant or benign. The architecture of the ANN is depicted in Figure 3, where the connections between each neuron are weight values applied to outputs of neurons.

The use of ANN for a computer-aided diagnosis (CAD) is a two-step process involving training on known examples followed by testing on unknown samples. The training procedure itself is composed of two processes. It involves feed-forwarding the input data followed by back-propagation of error by adjusting weights to minimize error on each training epoch.

Feed-Forwarding

All the features in the training set were standardized with zero mean and unit variance. This standardization is necessary to prevent nonuniform learning, in which the weights associated with some features converge faster than others. This can cause unwanted errors in learning. After standardization, a randomly chosen sample was put into the network for feed-forwarding. For each sample, four features were used as inputs; these inputs were processed from the hidden layer to the output layer. After each sample was passed through the network, the output value was calculated as $z = g(x) = f(\sum_{j=1}^{10} w_{1j} f(\sum_{i=1}^{4} w_{ji} x_i))$, where $f(\cdot)$ is an n unipolar sigmoid function with a ramp slope of 0.3. The nonlinear nature of the sigmoid function confines the output values between 0 and 1. In this equation, z is the output value, x_i are input features, and w are weighting values used in the hidden and output layers. After calculating the output value z of the network, the output value was compared with the target value (assigned 1 for malignant and 0 for benign) and the difference was used to determine the training error. This process was repeated for every sample in each training epoch.

Learning—error Back-propagation

After feed forwarding the entire training set, weights of the network were adjusted to meet the minima of the mean square error criterion function, defined as

Table 1
Summary of Logistic Regression Analysis

Features	Coefficient	Predictive Value (%) at the Threshold of P = .5			At 95% Sensitivity	
		Y(0|0)	Y(1|1)	Y(0|0) & Y(1|1)	Threshold	Specificity
Constant	3.34 ± 1.99	79.41	70	75.93	0.01	64.71%
Age	0.27 ± 0.02					
Margin sharpness	−0.03 ± 0.02					
Margin echogenicity	−0.27 ± 0.03					
Angular variance	−38.41 ± 1.91					

Note.—Coefficients were represented as mean ± standard deviation, according to each round-robin iteration.

$J(w) = \frac{1}{2}\|t - z\|^2$, where t and z are vectors of target and output values, respectively. To minimize this criterion function, the weight adjustment value was calculated by $\Delta w = -\eta \frac{\partial J}{\partial w}$, where η is the user-input learning rate, which was set to 0.1 based on the result of general training (see the following section).

As a first step, the general behavior of the network was evaluated using all data to confirm that the network can be trained. This also yielded the network parameters, such as learning rate and the slope of the sigmoid function of the optimal training curve. The optimal curve is defined as a curve that has smooth monotonic decreasing slope and saturates to the minimum training error, such as the curve in Figure 4. The parameters obtained from general training were applied to individual cycles of round-robin training and testing.

Receiver Operating Characteristic Analysis

The LR model test outputs and the ANN test outputs were evaluated for diagnostic performance by ROC analysis. Receiver operating characteristic (ROC) curves for each LR model test outputs and the ANN test outputs were generated empirically, based on the sensitivity and specificity value measured according to various decision threshold from 0 to 1. The shapes of curves were observed and the area under the curve, A_z, was used as a measure of the diagnostic performance. Also for each method, specificity was measured at fixed 95% sensitivity. Because we could not rely on Gaussian distribution for our sample distribution, the standard error for ROC curves was determined according to the method proposed by Hanley and McNeil (27):

$$SE = \sqrt{\frac{A_Z\ (1 - A_z) + (n_m - 1)(Q_1 - A_z^2) + (n_b - 1)(Q_2 - A_z^2)}{n_m n_b}}$$

where A_z is the area under the ROC curve; n_m and n_b are the number of malignant masses and benign masses, respectively; and Q_1 and Q_2 are estimated by:

$$Q_1 = \frac{A_z}{(2 - A_z)}, \quad Q_2 = \frac{2A_z^2}{(1 + A_z)}$$

RESULTS

Logistic Regression and ROC Analysis

Table 1 summarizes the results of LR using four features: age, M-sharp, M-echo, and AVM. The probability of malignancy (P_m) was determined by back-transforming the estimated parameters by the formula

$$P_m(D = 1|X_1,\ X_2,\ .....\ ,\ X_k) = \frac{1}{1 + e^{-(\alpha + \Sigma\beta_i X_i)}}.$$

The ROC curve for this four-feature model is shown in Fig. 5. The area under the ROC curve, A_z (±SD), was 0.853 ± 0.059 with 95% confidence limit (0.76–0.95), providing 64.7% specificity at 95% sensitivity, which is often desirable. In the earlier study (25), the area under the ROC curve for LR analysis was 0.874 ± 0.054, which is slightly higher than observed in this study. The difference between the two values is the result of preprocessing the features used in the previous study and of the difference in the parameter estimation method used: least square iteration versus maxi-

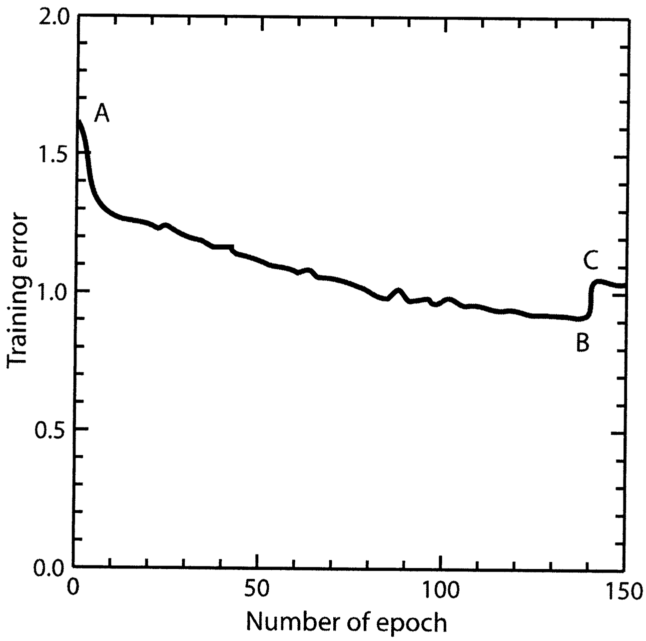

Figure 5. Graph showing the process of training of the artificial neural network analysis. Curve *ABC* is a training curve showing smooth monotonic decrease in training error, which saturates at point B. Beyond B, the network becomes overtrained.

mum-likelihood estimation. The goal of this study was to compare two approaches; therefore, optimization by preprocessing was omitted, because this would have biased results in favor of LR.

Artificial Neural Network and ROC Analysis

Table 2 summarizes the results of ANN. In both the general training and the testing, a batch algorithm—adjusting weights after feed-forwarding the entire training set—was used as learning algorithm. The training curve from the general training is shown in Figure 4. A total of 150 training epochs was used. The training error was minimized and saturated at approximately 140 epochs. The overtraining occurs after 140 epochs, which is marked by an increase in training error (Fig. 4).

The ROC curve for the ANN is shown in Fig. 6. Forty training iterations were employed for each test sample. The reason the number of training iterations is different from the general training is so that we could confirm the overtraining problem with more than 40 training iterations. The area under the ROC curve, A_z ($\pm$SD), was 0.856 $\pm$ 0.058 with 95% confidence limit (0.734–0.936), providing 76.5% specificity at 95% sensitivity.

DISCUSSION

Despite several technical and scientific advancements, unequivocal characterization of breast masses as malignant or benign without biopsy continues to be a difficult problem (10,28). Many authors have suggested the use of CAD as a second reviewer of mammograms or sonograms to reduce the diagnostic error rate (2).

Unfortunately, breast tumors occur in many different forms. They can range from being circumscribed, well-defined masses to ill-defined, irregular, and poorly differentiated lesions. The issue is: can a computer-based pattern recognition system be used to aid diagnosis of breast masses?

There are potentially large numbers of mathematical and statistical approaches for pattern recognition. LR is a popular method to analyze epidemiologic data. Recently, we suggested its use as a pattern classifier for breast masses (25). Another possible approach that has received wide recognition in many disciplines and has direct relevance to breast imaging is the use of ANN to mimic a human observer (2,26). Because both LR and ANN are fundamentally different approaches, we compared their abilities to diagnose breast cancer.

In LR, the probability of malignancy for a sample is estimated according to the logistic formula. The reason for the popularity of LR in epidemiologic study is its ability to perform probability estimation. It uses an appealing S-shaped description of the combined effect of several risk factors to evaluate the risk of a disease. This analysis provides a deterministic model for the data and yields weighting factors for each contributing feature. This information gives the observer an intuitive sense of which of many features dominates the probability of malignancy. It also allows the calculation of the odds ratio, which represents the degree of risk associated with each feature. However, because LR fits the data to a formulated function, it is less flexible and less capable of solving multifaceted problems than ANN.

Although LR fits the data to a formulated function, ANN maps data from the sample space to the decision space. Including this mapping property, there are several reasons why ANN is particularly promising compared with the large group of "universal" pattern classifiers in breast mass diagnosis. The breast mass classification problem appears to have both global and local features, and ANN-style classifiers can robustly accommodate such problem classes. Theory suggests that ANN is particularly

Table 2
Summary of Artificial Neural Network Analysis

			Predictive Value (%) at the Threshold of P = .5			At 95% Sensitivity	
Features	Learning Rate	Slope of Sigmoid	Y(0|0)	Y(1|1)	Y(0|0) and Y(1|1)	Threshold	Specificity
	0.1	0.3	79.41	80	79.63	0.02	76.47%
Age							
Margin sharpness							
Margin echogenicity							
Angular variance							

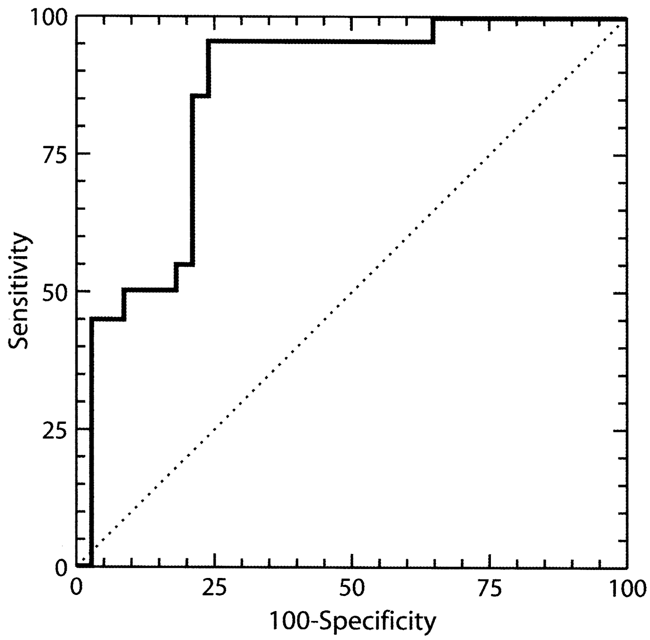

Figure 6. Graph showing receiver operating characteristics (ROC) of the artificial neural network analysis. The area under the ROC curve, A_z, was 0.856 ± 0.058 with 95% confidence limit (0.734–0.936); the curve shows 76.47% specificity at 95% sensitivity.

well suited for "low-frequency" problems, and an intrinsically nonlinear approach may prove profitable. Several training algorithms are possible using ANN. This is particularly attractive for breast cancer diagnosis, in which the problem is not perfectly defined because of the multifaceted nature of tumor biology. Finally, ANN training may be readily adapted to sparse and noisy data and the theories of optimal stopping (of training) and statistical model selection can be applied (29–32).

Although powerful in concept, ANN also has some drawbacks: it has a relatively slow rate of training, though it poses no insuperable obstacle in this context, because training is off-line. The behavior of the network has properties of a black box, not giving exact information on weighting factors of individual components to the user. However, compared with LR, the flexibility and the capability of solving problems are certainly advantages.

According to the results of this study, both the LR and the ANN have nearly identical diagnostic performance as measured by the area under the ROC curve. However, shapes of the ROC curves are different for the two cases; the top-left part of ANN ROC curve is extended more to the left side compared with the LR ROC curve. This means that ANN has better specificity than LR for a fixed sensitivity. For instance, the ANN has 11.8% higher specificity at 95% sensitivity. It is difficult to assess if the advantage of ANN over LR in the local region of a ROC curve reaches statistical significance. However, given the ability of ANN to separate complex and nonlinear feature space, such a benefit can be expected. The choice of 95% sensitivity was made primarily because clinical observers prefer to operate at this or higher levels of sensitivity.

Because the diagnostic performance of both LR and ANN compare well with the published results (2,4–10,33), it is reasonable to assume that either can be used for CAD. Considering that only a basic form ANN (multilayer perceptron model) was used in this study, it is reasonable to anticipate that fine tuning, or front-end preprocessing, could further improve CAD performance. On a relative basis, ANN has higher specificity at the sensitivity of 95% and thus is a preferable choice as a pattern classifier compared with LR in this region of the ROC curve.

The data used in this study represent cases that were recommended for biopsies by radiologists. Because 34 cases proved to be benign, biopsies could have been avoided in 63% (34/54) cases. If LR-CAD was used as

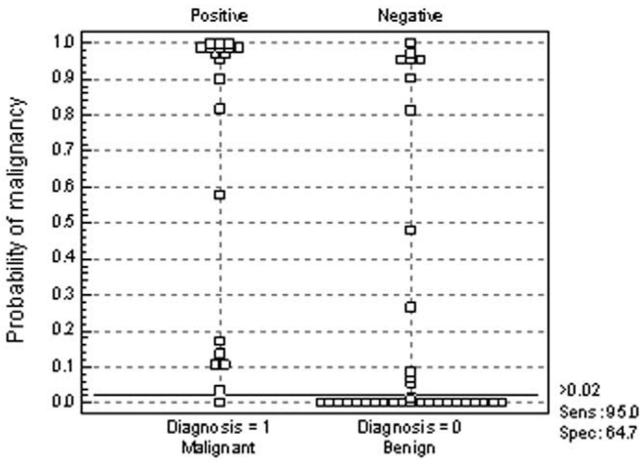

Figure 7. Graph illustrating the biopsy decision for the logistic regression according to thresholds. The horizontal solid line represents threshold of 0.02, where sensitivity is 95% and specificity is 64.71%

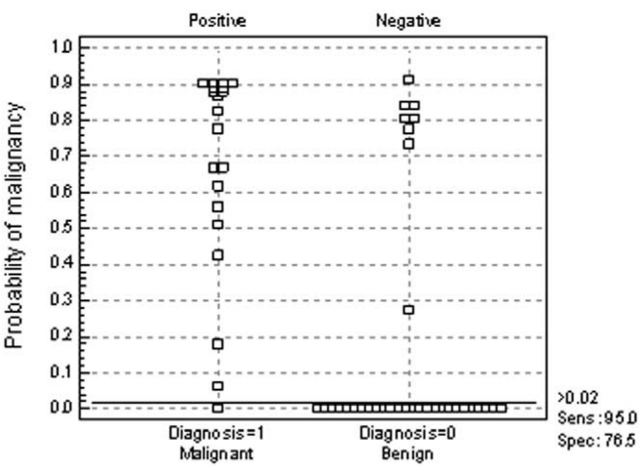

Figure 8. Graph illustrating the biopsy decision for the artificial neural network according to thresholds. The horizontal solid line represents threshold of 0.02, where sensitivity is 95% and specificity is 76.47%.

a second opinion, then 22 of 34 biopsies (64.7%) could have been avoided (Fig. 7). Similarly, if ANN was used as second opinion, 26 of 34 cases (76.5%) would not have been recommended for biopsy (Fig. 8). These results show the possibility of reducing biopsies with CAD. Although both approaches (LR and ANN) reduced the number of unnecessary biopsies, ANN had better performance, because it eliminated four additional biopsies. It is important to note that this benefit of reducing biopsies was at the expense of missing 1 of 20 (5%) malignant masses. Both ANN and LR misdiagnosed the same case.

Finally, the conclusions of this study are based on a relatively small number of samples. Future analysis with more data acquired from various sources and scanners is necessary to demonstrate the generalizability of ANN and LR. However, because both LR and ANN showed acceptable diagnostic performance as CAD methods, the results of this study are promising. Because the results suggest that the two classifiers exhibit regional difference in sensitivity, LR being superior in a high-sensitivity region whereas ANN has better performance in the high-specificity region, combining the two approaches so that each method complements the other has the possibility of improving the overall performance.

ACKNOWLEDGMENT

The authors thank Theodore W. Cary for help in IDL programming and in the preparation of the manuscript.

REFERENCES

1. Wingo PA, Tong T, Bolden S. Cancer statistics, 1995. CA Cancer J Clin 1995; 41:8–30.
2. Chen CM, Chou YH, Han KC, et al. Breast lesions on sonograms: computer-aided diagnosis with nearly setting independent features and artificial neural networks. Radiology 2003; 226:504–514.
3. Richter K, Heywang-Kobrunner SH, Winzer KJ, et al. Detection of malignant and benign breast lesions with an automated US system: results in 120 cases. Radiology 1997; 205:823–830.
4. Sehgal CM, Arger PH, Rowling SE, et al. Quantitative vascularity of breast masses by Doppler imaging: regional variations and diagnostic implications. J Ultrasound Med 2000; 19:427–442.
5. Goldberg V, Manduca A, Ewert DL, et al. Improvement in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence. Med Phys 1992; 19:1475–1481.
6. Garra BS, Krasner BH, Horii SC, et al. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. Ultrasonic Imaging 1993; 15:267–285.
7. Dumane VA, Shankar PM, Piccoli CW, et al. Computer aided classification of masses in ultrasonic mammography. Medi Phys 2002; 29:1968–1973.
8. Shankar PM, Dumane VA, Piccoli CW, et al. Classification of breast masses in ultrasonic B-mode images using a compounding technique in the Nakagami distribution domain. Ultrasound Med Biol 2002; 28: 1295–1300.
9. Drukker K, Giger ML, Horsch K, et al. Computerized lesion detection on breast ultrasound. Med Phy 2002; 29:1438–1446.
10. Chen DR, Chang RF, Kuo WJ, et al. Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. Ultrasound Med Biol 2002; 28:1301–1310.
11. West D, West V. Model selection for a medical diagnostic decision support system: a breast cancer detection case. Artif Intell Med 2000; 20:183–204.
12. Sargent DJ. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. Cancer 2001; 91(8 Suppl): 1636–1642.
13. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002; 35:352–359.
14. Vandewalle J. Advanced statistical techniques: how to use your data more effectively? [comment]. Ultrasound Obstet Gynecol 1999; 13:8–10.

15. DiRusso SM, Chahine AA, Sullivan T, et al. Development of a model for prediction of survival in pediatric trauma patients: comparison of artificial neural networks and logistic regression. J Pediatr Surg 2002; 37: 1098–1104.
16. Bostwick DG, Burke HB. Prediction of individual patient outcome in cancer: comparison of artificial neural networks and Kaplan-Meier methods. Cancer 2001; 91(8 Suppl):1643–1646,.
17. Clermont G, Angus DC, DiRusso SM, et al. Predicting hospital mortality for patients in the intensive care unit: a comparison of artificial neural networks with logistic regression models [see comment]. Crit Care Med 2001; 29:291–296.
18. Gerber TC, Foley DA, Zheng Y, et al. Differentiation of intracardiac tumors and thrombi by echocardiographic tissue characterization: comparison of an artificial neural network and human observers [erratum appears in Echocardiography 2000; 17:iv]. Echocardiography 2000; 17: 115–216.
19. Freeman RV, Eagle KA, Bates ER, et al. Comparison of artificial neural networks with logistic regression in prediction of in-hospital death after percutaneous transluminal coronary angioplasty. Am Heart J 2000; 140: 511–520.
20. Nestorov IS, Hadjitodorov ST, Petrov I, et al. Empirical versus mechanistic modelling: comparison of an artificial neural network to a mechanistically based model for quantitative structure pharmacokinetic relationships of a homologous series of barbiturates. AAPS Pharmsci 1999; 1:E17.
21. Silver DL, Hurwitz GA. The predictive and explanatory power of inductive decision trees: a comparison with artificial neural network learning as applied to the noninvasive diagnosis of coronary artery disease. J Investig Med 1997; 45:99–108.
22. Eng J. Predicting the presence of acute pulmonary embolism: a comparative analysis of the artificial neural network, logistic regression, and threshold models. AJR Am J Roentgenol 2002; 179:869–874.
23. Nguyen T, Malley R, Inkelis S, et al. Comparison of prediction models for adverse outcome in pediatric meningococcal disease using artificial neural network and logistic regression analyses. J Clin Epidemiol 2002; 55:687–695.
24. Ongphiphadhanakul B, Rajatanavin R, Chailurkit L, et al. Prediction of low bone mineral density in postmenopausal women by artificial neural network model compared to logistic regression model. J Med Assoc Thailand 1997; 80:508–515.
25. Sehgal CM, Cary TW, Kangas SA, et al. Computer-based margin analysis of breast ultrasound for differentiating malignant and benign masses. Philadelphia, PA: Department of Radiology, University of Pennsylvania, 2004.
26. Drukker K, Giger ML, Vyborny CJ, et al. Computerized detection and classification of cancer on breast ultrasound. Acad Radiol 2004; 11: 526–535.
27. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143:29–36.
28. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology 1992; 184:613–617.
29. Wang C, Judd SJ, Venkatesh SS. When to stop: on optimal stopping and effective machine size in learning. Conference on Neural Information Processing Systems. Denver, Co: November 1993.
30. Wang C, Venkatesh SS. Machine size selection for optimal generalization. workshop on applications of descriptional complexity to inductive, statistical, and visual inference. New Brunswick, NJ: July 1994.
31. Wang C, Venkatesh SS. Temporal dynamics of generalization in neural networks. In: Touretzky DS, Tesauro G, Leen TK, eds. Advances in neural information processing systems 7. Cambridge, Ma: MIT Press, 1995.
32. Wang C, Venkatesh SS. Criteria for approximation error and complexity trade-off in learning. Proceedings of the Eighth Annual Workshop on Computational Learning Theory. Baltimore, Md: ACM Press, 1995.
33. Chen CM, Chou YH, Han KC, et al. Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks. Radiology 2003; 226:504–514.