Prediction of Breast Cancer Malignancy Using an Artificial Neural Network

Carey E. Floyd, Jr., Ph.D.,† Joseph Y. Lo, Ph.D.,*† A. Joon Yun, M.D.,* Daniel C. Sullivan, M.D.,* and Phyllis J. Kornguth, M.D., Ph.D.**

Background. **An artificial neural network (ANN) was developed to predict breast cancer from mammographic findings. This network was evaluated in a retrospective study.**

Methods. **For a set of patients who were scheduled for biopsy, radiologists interpreted the mammograms and provided data on eight mammographic findings as part of the standard mammographic workup. These findings were encoded as features for an ANN. Results of biopsies were taken as truth in the diagnosis of malignancy. The ANN was trained and evaluated using a jackknife sampling on a set of 260 patient records. Performance of the network was evaluated in terms of sensitivity and specificity over a range of decision thresholds and was expressed as a receiver operating characteristic curve.**

Results. **The ANN performed more accurately than the radiologists ($P < 0.08$) with a relative sensitivity of 1.0 and specificity of 0.59.**

Conclusions. **An ANN can be trained to predict malignancy from mammographic findings with a high degree of accuracy. *Cancer* 1994;74:2944–8.**

Key words: breast neoplasms, diagnosis; computers, neural network; computers, diagnostic aid; mammography.

Introduction

Breast cancer is a serious health problem. In 1993, an estimated 182,000 new cases were diagnosed in the US and an estimated 46,000 deaths resulted from this disease.[1] While mammography remains the most sensitive technique for early detection of breast cancer, a significant fraction of patients referred for biopsy as the result of mammography findings do not have a malignancy. Although specific, biopsy is an invasive, costly, and emotionally stressful procedure. In an effort to reduce the number of benign cases sent to biopsy, an artificial intelligence technique was investigated to predict the outcome of biopsy from radiographic findings. An artificial neural network (ANN) was developed that takes radiologic findings as inputs and predicts the outcome of biopsy as an output.

ANNs are computer algorithms whose structure and function are based on models of the structure and learning behavior of biological neural networks. These algorithms are typically employed to classify a set of patterns into one of several classes. The classification rules are not written into the algorithm, but are learned by the network from examples.

Artificial neural networks have been developed for a wide variety of computational problems in cognition, pattern recognition, and decision making.[2–5] The term "neural network" is applied to a parallel structure of (simple) computation units arranged in layers mimicking the physiologic structure of the brain. There are multiple connections between units within and between layers. These connections have strengths or "weights" that are "learned" by the network. The training paradigm is either "supervised," where sample input-output pairs are presented[3,6,7] or "unsupervised," where the network organizes itself.[8,9] Information in the network is stored in these interconnection weights.

A variety of medical tasks has been successfully performed using such networks, including analysis of EKG patterns,[10] decision-making in pathology,[11] texture analysis in ultrasound,[12] lesion detection in SPECT images,[13] image boundary detection,[14] differential diagnosis from chest radiographs,[15,16] prediction of pulmonary embolism from ventilation/perfusion scans,[17] breast cancer analysis,[18] and decision-making in mammography.[19,20] Here an ANN is described that has been

From the Department of *Radiology, Duke University Medical Center, Durham, NC, and the Department of †Biomedical Engineering, Duke University, Durham, NC.

†Current address: Department of Radiology, University of Pennsylvania Medical Center, Philadelphia PA.

Supported in part by grant CA14236 awarded by the National Institutes of Health.

Address for reprints: Carey E. Floyd, Jr., Ph.D., Box 3949, Duke University Medical Center, Durham, North Carolina 27710.

Received June 24, 1994; revision received August 4, 1994; accepted August 4, 1994.

trained to predict breast malignancy from mammographic elements.

Materials and Methods

Artificial Neural Networks

The ANN for malignancy prediction was implemented as a three-layer backpropagation architecture with one hidden layer. Input feature values are the mammographic findings assigned by the radiologists. These findings were assigned numeric values as described below. The network was trained on the biopsy results for each patient, where 0.2 was used for patients with benign lesions and 0.8 was used for patients with malignancy. This "trained knowledge" was contained in the internal numeric "weights" of the network. The training algorithm iteratively modified the numeric values of these weights to decrease the training error of the network. Once trained, the weights were fixed. When a set of input feature values are presented to the trained network, an output is generated that represents the classification (malignant or benign) of the prediction. This output is based on the trained knowledge that the network learned in the training step.

The network was trained using a backpropagation supervised training algorithm. Supervised training is a technique in which a set of representative input/output pairs is presented to the network. Through an iterative algorithm, the internal network weights are adjusted to decrease the difference between the network prediction and the true result for the training cases. As different examples (from the set of training input/output pairs) are presented, the network weights are adjusted to minimize the mean squared error. Because this error is accumulated over the entire training set, the network weights will adjust to find an average response. It is important that the training set include examples that represent the full range of possible inputs so that the network will correctly classify any input set. Note that the new input sets do not have to be identical to any of the training patterns for the network to correctly classify the new input pattern.

Case Selection

The examples used were mammographic features with corresponding biopsies. The group consisted of the results of the examinations of 260 patients (92 with malignancies, 168 with benign outcomes) randomly selected from examinations at Duke University Medical Center that were verified by surgical biopsy between January 1991 and May 1992. Craniocaudal, mediolateral oblique, and (optional) magnification views had been obtained by screen-film technique. Radiologists interpreted the films and entered the mammographic findings on forms as part of the standard procedure before dictation. There were no patients with missing data. Malignant or benign outcome from biopsy was recorded.

Table 1. Coding of Input Node Features

Node feature	Finding	Value
mass size	no mass	0.0
	mass	size (mm)
mass margin	no mass	0.0
	well circumscribed	0.2
	microlobulated	0.4
	obscured	0.6
	indistinct	0.8
	spiculated	1.0
asymmetric density	none	0.0
	asymmetry	size (mm)
architectural distortion	none	0.0
	distortion	1.0
calcification number	none	0.0
	< 5	0.33
	> 5 and < 10	0.66
	> 10	1.0
calcification morphology	none	0.0
	not suspicious	0.33
	moderately suspicious	0.66
	highly suspicious	1.0
calcification density	none	0.0
	dense	0.33
	mixed	0.66
	faint	1.0
calcification distribution	none	0.0
	scattered	0.33
	intermediate	0.66
	clustered	1.0

Defining the Features List

At the time of this study, a lexicon was used that differed from the ACR Breast Imaging Reporting and Data system. The checklist findings entry form consisted of a list of eight radiographic features (Table 1). For the application to neural networks, numeric values were assigned to these features in ascending order of indication of malignancy. The eventual performance of the network does not rely on the relative scaling of the features in Table 1, only on the relative ranking. That is, a finding with an assigned value of 0.4 is not necessarily twice as likely to indicate malignancy as a finding with

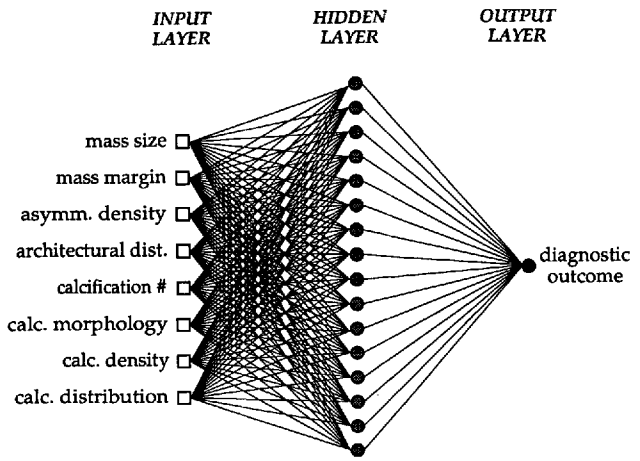

Figure 1. Architecture of neural network for predicting biopsy results of radiographic findings.

a value of 0.2, but does indicate malignancy with a strength that lies between that of findings with values of 0.2 and 0.6. In addition, the radiologists assigned an overall impression of malignancy on a scale from one to five (1 = benign; 2 = probably benign; 3 = indeterminate; 4 = probably malignant; 5 = malignant). This impression was used to evaluate the performance of the radiologists; it was not used as an input to the ANN. Three radiologists read the cases. Each report was read by only one radiologist. The radiologists' diagnostic findings and impressions were gathered from the paper forms; the reports were not read specifically for this study. Thus the three radiologists were treated as a single observer.

Constructing the Neural Network

A three-layered backpropagation neural network (Fig. 1) was created for the classification task. The input layer consisted of eight nodes that represented the eight radiographic features from the data entry form (excluding the overall radiologic impression). The hidden layer consisted of 16 nodes (the optimal number of hidden nodes for a network is difficult to predict). The number of hidden nodes was varied between 2 and 20. There most accurate network training performance was obtained with 16 hidden nodes. The output layer consisted of a single node representing diagnostic outcome; 0.0 for benign and 1.0 for malignant. After a series of trials, the training parameters were optimized: number of nodes in the hidden layer = 16, learning coefficient for connections from input to hidden layer = 0.5, learning coefficient for connections from hidden layer to output layer = 0.3, momentum coefficient = 1.0, and training interval = 200 iterations. All patterns were presented to the network at each iteration. The network was implemented on a computer workstation (SPARCstation 10 512, Sun Microsystems Inc., Mountain View, CA) using locally written software. Training time was typically 10 minutes for 200 iterations on all 260 cases. Once trained, the evaluation of all 260 cases required much less than 1 second.

Training and Testing the Network

The network was trained and evaluated on all cases using the jackknife technique.[21] The jackknife technique is useful when the number of cases is limited. It allows utilization of all available cases for training but also provides a meaningful evaluation of the generalizing ability of the trained network. With this technique, one case was selected for evaluation and the network was trained on the remaining 259 cases. The trained network was then evaluated on the single case that was left out. The selected case was put back in and another case was removed. This process was repeated until all cases had been used for evaluation. Then the network was trained on all cases. The training was "supervised"; for each case the network was provided with both the input of mammographic elements and the corresponding biopsy diagnosis. For training, the value of the output node was 0.2 if the biopsy pathology was benign and 0.8 if the result was malignant. When presented with this testing set, the network generated an output between 0.0 and 1.0 reflecting its predictive value for malignancy. The predictions of the radiologists and the network were compared using receiver operating characteristic (ROC) analysis (program CLABROC by Charles Metz.)[22] Differences in performance were evaluated by comparing the areas under the ROC curves

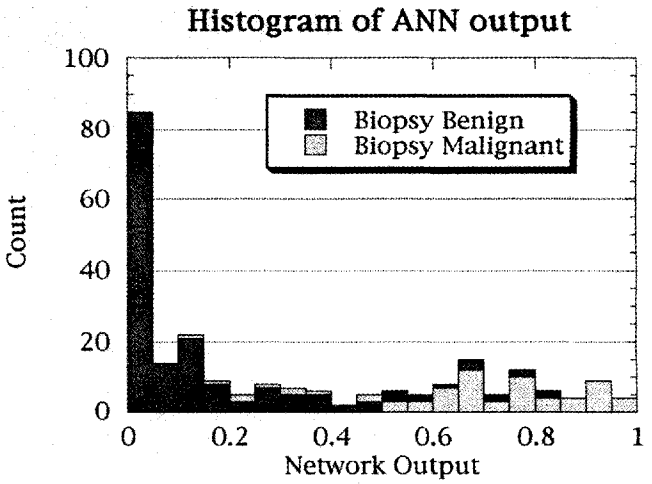

Figure 2. Histogram for the output values of the trained network.

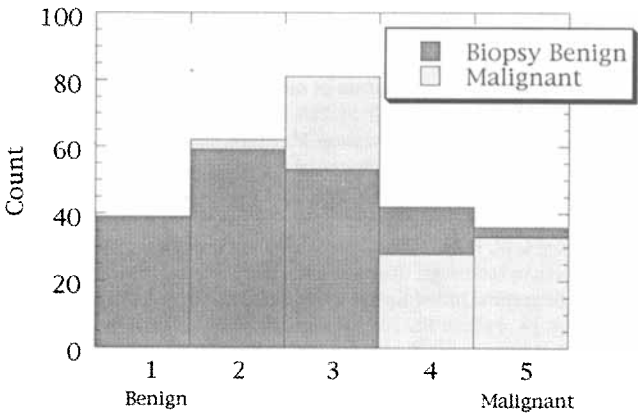

Figure 3. Histogram for the radiologists' rating.

(A_Z). This comparison was also performed using CLABROC.

Results

A histogram for the output values of the trained network is shown in Figure 2. A histogram for the radiologists' rating of the same mammographic elements is shown in Figure 3. These histograms can be interpreted by examining the number of true and false positive diagnoses that result from setting a decision threshold. As can be seen from Figure 2, if a threshold for the ANN is set at 0.1 (that is, if all cases with network output less than 0.1 were not sent to biopsy), then 99 of the 168 benign cases and all of the malignant cases would be correctly identified (relative sensitivity, 1.0; specificity, 0.59). This can be compared to Figure 3 for the radiologists' impression, where if a threshold was set at category 1 (that is, if all cases assigned category 1 were not sent to biopsy), then 38 of the benign cases and all of the malignant cases would be correctly identified (relative sensitivity, 1.0, specificity, 0.23). The term "relative sensitivity" is used because it is unlikely that all malignant cases were sent to biopsy (mammography is not 100% sensitive). Here we are examining the use of ANNs to improve specificity for those cases that were sent to biopsy.

ROC curves comparing the performance of the trained network with the radiologists is shown in Figure 4. The area indices (A_Z) are 0.94 (±0.01) for the ANN and 0.91 (±0.02) for the radiologists. The difference was statistically significant with a two-tailed P value of less than 0.08.

Discussion

An ANN technique has been evaluated to predict malignancy from mammographic findings. The ANN was evaluated for the subset of mammographic cases that were sent to biopsy. This is a distinct subpopulation of the women who receive mammograms and is not representative of the entire population.

For the cases sent to biopsy, the ANN displays significantly better diagnostic performance than the radiologists when the network output is compared to the radiologists' categorical assessment. This is consistent with the results of a recent study by Wu et al.,[20] which also used an ANN to predict malignancy from mammographic elements.

For cases that a radiologist would consider sending to biopsy, a computer aid could be implemented as an ANN that would form a prediction for the outcome of biopsy based on the radiologist's findings. For those cases where any of the relevant findings are present, the output of the network would be presented to the radiologist for consideration. While not demonstrated here, it is feasible that this network prediction could be useful for the radiologist to consider in addition to the mammographic elements and the patient's medical history. The present study suggests that numeric decision-making techniques such as artificial neural networks may have a useful role in improving the accuracy and consistency of medical diagnosis.

Future studies should test the value of adding medical history information to the input feature set for the network. In addition, a prospective study is planned to test whether a diagnostic suggestion supplied by the ANN can improve the diagnostic accuracy of radiologists.

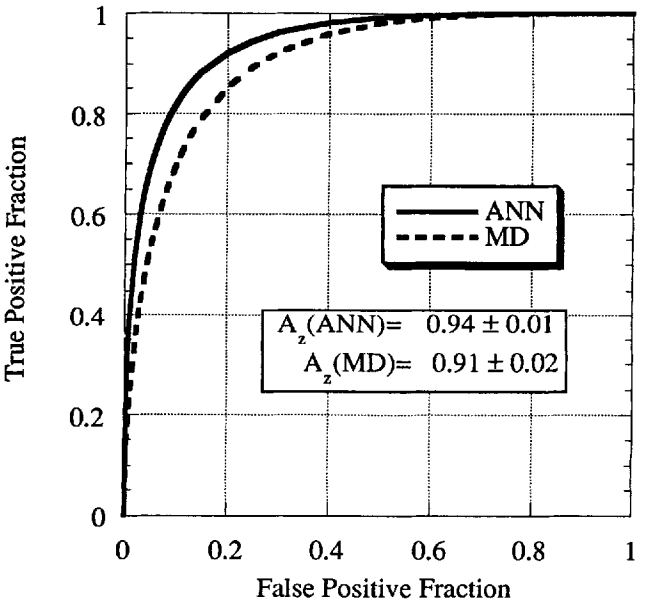

Figure 4. ROC comparison of network and radiologists.

References

1. Boring CC, Squires TS, Tong T. Cancer statistics. *CA Cancer J Clin* 1993;43:7–26.
2. Anderson JA, Rosenfeld E, editors. Neurocomputing: foundations of research. Cambridge: The MIT Press, 1988.
3. Rumelhart DE, McClelland JL, editors. Parallel distributed processing: explorations in the microstructures of cognition. Cambridge: The MIT Press, 1986.
4. Wasserman PD. Neural computing: theory and practice. New York: Van Nostrand Reinhold, 1990.
5. Boone JM, Gross GW, Greco-Hunt V. Neural networks in radiologic diagnosis: I. Introduction and illustration. *Invest Radiol* 1990;25:1012–6.
6. Rosenblatt F. Principles of neurodynamics. New York: Spartan Books, 1959.
7. Widrow B, Hoff ME. Adaptive switching circuits. In: 1960 IRE WESCON Convention Record. New York: IRE, 1960.
8. Hebb DO. The Organization of behavior. New York: John Wiley & Sons, 1949.
9. Kohonen T, editor. Self-organization and associative memory. In: Series in information sciences, vol. 8. Berlin: Springer Verlag, 1984.
10. Cios KJ, Chen K, Langenderfer RA. Use of neural networks in detecting cardiac diseases from echocardiographic images. *IEEE Eng Med Biol* 1990;9:58–60.
11. Dytch HE, Wied GL. Artificial neural networks and their use in quantitative pathology. *Anal Quant Cytol Histol* 1990;12:379–93.
12. DaPonte JS, Sherman P. Classification of ultrasonic image texture by statistical discriminant analysis and neural networks. *Comput Med Imaging Graph* 1991;15:3–9.
13. Floyd CE, Jr, Tourassi GD. An artificial neural network for lesion detection in SPECT images. Invest Radiol 1992;27:667–72.
14. Amartur SC, Piraino D, Takefuji Y. Optimization neural networks for the segmentation of magnetic resonance images. *IEEE Trans Med Imaging* 1992;11:215–20.
15. Asada N, Doi K, MacMahon H, Montner SM, Giger ML, Abe C, et al. Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study. *Radiology* 1990;177:857–60.
16. Gross GW, Boone JM, Greco-Hunt V, Greenberg B. Neural networks in radiologic diagnosis: II. interpretation of neonatal chest radiographs. *Invest Radiol* 1990;25:1017–23.
17. Scott JA, Palmer EL. Neural network analysis of ventilation-perfusion lung scans. *Radiology* 1993;186:661–4.
18. Dawson AE, Austin RE, Weinberg DS. Nuclear grading of breast carcinoma by image analysis: classification by multivariate and neural network analysis. *Am J Clin Pathol* 1991;95(1Suppl):S29–37.
19. Wu Y, Doi K, Giger ML, Vyborny CJ, Schmidt RA, Nishikawa RM, et al. Application of artificial neural networks in mammography for the diagnosis of breast cancer. *Radiology* 1991;181(P):143.
20. Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology* 1993;187:81–7.
21. Tourassi GD, Floyd CE, Sostman HD, Coleman RE. An artificial neural network approach for the diagnosis of acute pulmonary embolism. *Radiology* 1993;189:555–8.
22. Metz CE, Shen J-H, Herman BA. New methods for estimating a binormal ROC curve from continuously-distributed test results. Presented at the 1990 Joint Meetings of the American Statistical Society and the Biometric Society. Anaheim CA, 1990.