

## Original Investigations

# Prescreening Entire Mammograms for Masses with Artificial Neural Networks: Preliminary Results

Barry L. Kalman, PhD, William R. Reinus, MD, Stan C. Kwasny, PhD  
Andrew Laine, PhD, Lawrence Kotner, MD

**Rationale and Objectives.** The authors evaluated the feasibility of combining wavelet transform and artificial neural network (ANN) technologies to prescreen mammograms for masses.

**Methods and Materials.** Fifty-five mammograms (29 with masses and 26 without) were digitized to 100-mm resolution and processed by using wavelet transformation. These wavelets were subjected to a linear output sequential recursive auto-associative memory ANN and cluster analysis with feature vector formation. These vectors were used in two separate experiments—one with 13 cases and another with seven cases held out in a test set—to train feed-forward ANNs to detect the mammograms with a mass. The experiments were repeated with rerandomization of the data, four and six times, respectively.

**Results.** There was a statistically significant correlation ( $P < .01$ ) between the network's prediction of a mass and the presence of a mass. With majority voting, the feed-forward ANNs detected masses with 79% sensitivity and 50% specificity.

**Conclusion.** Although preliminary, the combination of wavelet transform and ANN is promising and may provide a viable method to prescreen mammograms for masses with high sensitivity and reasonable specificity.

**Key Words.** Computers, neural network; breast neoplasms, diagnosis; breast radiography

**B**ecause breast cancer is now estimated to strike one in eight adult American women [1, 2], many national institutions are promoting large-scale breast cancer screening programs [3–5]. Application of computer-aided diagnosis techniques to mammography screening programs may offer substantial benefit in terms of cost reduction and increased effectiveness of the screening process. The use of computers to directly prescreen mammograms may eventually permit a substantial reduction in the number of studies that must be viewed by a radiologist. Of course, this implies that computers must be able to directly interpret digitized images and that this process must be fully automated. To date, this capability has been

From the Mallinckrodt Institute of Radiology,  
Barnes-Jewish Hospital, St Louis, MO.

Address reprint requests to W. R. Reinus, MD,  
Mallinckrodt Institute of Radiology, Barnes-Jewish  
Hospital, North Campus, 216 S Kingshighway  
Blvd, St Louis, MO 63110.

Received December 16, 1996, and accepted for  
publication after revision February 21, 1997.

**Acad Radiol**; 4:405–414  
©1997, Association of University Radiologists

limited by available technologies. Several rule-based schemes that use thresholding, subtraction, or both have been developed for computer analysis of digitized mammograms [6-9]. Although these techniques have produced promising results, they have been hampered by high false-positive detection rates [10, 11].

Mammograms show only an estimated 3% of their actual information content [12]. Improvements in the visibility of mammographic information content will probably improve detection of small tumors. It is unlikely, however, that state-of-the-art screen-film radiography alone can be improved to display more information. Wavelet transformation, an image enhancement technique, has been used successfully to enhance the visibility of image information content in mammograms, including both masses and microcalcifications [13-24]. Another advantage of wavelet image representations is that they can provide high magnitudes of data compression without loss of important image features [25].

Artificial neural networks (ANNs) have been suggested as an alternative to traditional rule-based methods for computer-aided detection of mammographic lesions [26-28]. ANNs learn the importance of image features on the basis of example training images. In general, they have proved very adept at pattern recognition problems [29-35]. Because of technologic limitations, however, most of the investigators attempting direct digitized data analysis of mammograms with ANNs have used small regions of interest selected from an entire image [27,28]. Others have extracted features, either qualitative or quantitative, for network training [36] or have incorporated ANNs into other computer-aided diagnosis schemes to improve lesion detection [37, 38].

In this pilot study, we investigated the potential of a set of customized ANNs to directly screen entire digitized mammograms for radiographically dominant masses. We used ANNs to detect masses from wavelet-transformed data from pairs of *entire* digitized mammographic images. We took advantage of wavelet transformation techniques for both feature enhancement and data compression.

## MATERIALS AND METHODS

A data set consisting of 55 mammograms, 29 with masses and 26 without, was digitized to 100- $\mu$ m resolution. The optical density of the darkest area of any image measured 4.1. Each of the patients had undergone

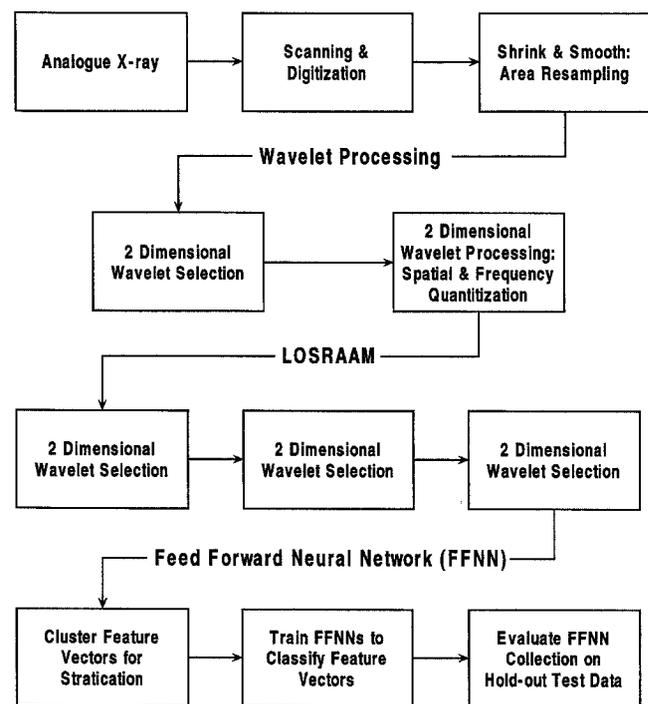


FIGURE 1. Schematic diagram of image processing with wavelet and neural network processing. LOSRAAM = linear-output sequential recursive auto-associative memory.

breast biopsy because of a radiographically detected mass or microcalcifications or because of a palpable abnormality not detected on a mammogram. All mammograms were pathologically correlated with the results of biopsy. Because this is a pilot study, the mammograms were selected from a 380-case database to be representative of various mass appearances, including a variety of sizes, locations within the breast, and edge appearances.

The average age of patients in the chosen data set was 57.0 years  $\pm$  12.9. The masses ranged in size from 5 to 35 mm, with a mean of 14.1 mm  $\pm$  6.4. On the basis of pathologic findings, 11 masses corresponded to benign lesions and 18 corresponded to cancer. The edges of six, 12, and 11 masses were classified as predominantly smooth, ill defined, and spiculated, respectively. Similar to a method described by Chan et al [37], mammographic abnormalities, including both masses and microcalcifications, were scored according to a subjective five-point scale of visibility ranging from 1 (easiest to see) to 5 (most difficult to see). In each case the same radiologist (L.K.) scored the difficulty of the finding on entry into the database. These scores were reviewed by another radiologist (W.R.R.) for accuracy,

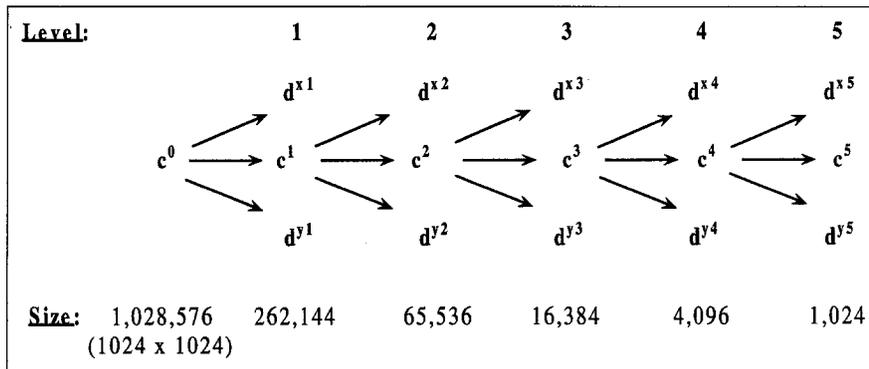


FIGURE 2. Diagram of wavelet hierarchy shows number of coefficients resulting at each level.

and disagreements were resolved by the radiologists together in review of the images. The masses in our data set had an average visibility of  $2.3 \pm 1.2$ . Smooth, spiculated, and ill-defined masses had visibility scores of  $2.0 \pm 1.4$ ,  $2.1 \pm 1.2$ , and  $2.7 \pm 1.1$ , respectively.

### Data Preparation

The overall method of data acquisition and processing is shown in Figure 1. The mammograms, consisting of two images each, were digitized by using a film digitizer system (FBI; Eastman Kodak, Rochester, NY). This system incorporates a 100- $\mu$ m focal spot film digitizer (model 200; Lumisys, Sunnyvale, CA). Each image measured  $9.375 \times 6.825$  inches ( $23.43 \times 17.06$  cm); after digitization, the images were on the order of  $1,920 \times 1,536$  pixels. All image processing and neural network training was done on a SPARC Center 2000 (Sun Microsystems, Sunnyvale, CA) with 20 available parallel processors. To accommodate our wavelet transformation software, each image was resized by using an area-sampling algorithm based on that used by Newman and Sproull [39], from a rectangular matrix to a square one ( $1,024 \times 1,024$  pixels, each 1 byte in depth). This algorithm determines the value of each pixel in the new image in proportion to the area from which it is derived in the original image. After the image was resized, the entire data set was contained in 115,343,360 bytes.

### Wavelet Transformation

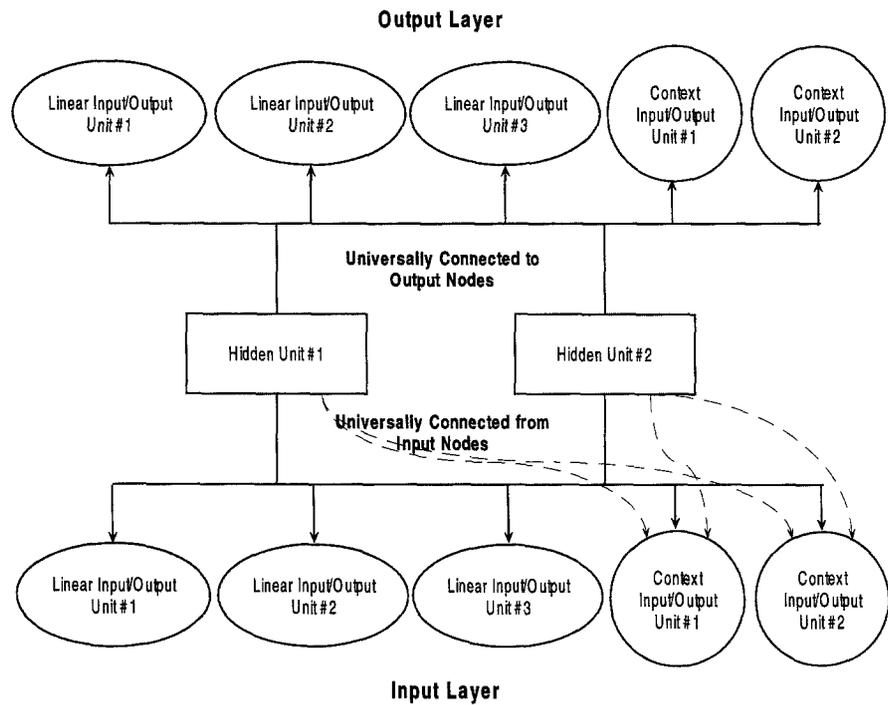
Multiresolution (five-level) and multidirection (two-dimensional) wavelet analysis with quadratic spline wavelets was used to transform each square image [40]. These wavelets are equivalent to the first-order derivative of a smoothing function, and so they enhance the edges of image objects. This algorithm transformed the data set to 699,392 4-byte floating point coefficients within the interval  $(-1, 1)$ . The wavelet coef-

ficients can be viewed hierarchically descending along the two dimensions (x followed by y) at each of the five levels of resolution, terminating with the remaining (nondecomposed) coefficients at level five (Fig 2). Laine and Song [16] showed that high-quality mammogram reconstruction can be performed with a truncated wavelet hierarchy wherein coefficients below a certain threshold are discarded. We tested thresholds between 0.25 and 0.75 and found that a threshold of 0.4 was optimal for analysis in that this threshold provided the best results with our later methods (see below). After truncation of the wavelet transform, with use of 0.4 as the coefficient threshold, the data set was represented by 174,841 triplets (wavelet coefficient, hierarchy level, position within level) of 4-byte floating point integers with an average of 1,589 triplets per image. Thus, the entire data set was contained in 2,098,092 bytes after wavelet transformation and truncation, a compression ratio of 55:1 from the original 110 images in  $1,024 \times 1,024$  by 1-byte format.

### LOSRAAM Analysis and Clustering

Because two-dimensional wavelet transforms quantize an image in terms of space and spatial frequency and can be ordered linearly, images can be processed recursively to determine prominent features. On the basis of our previous experience parsing natural language (a problem with similar characteristics to the current one), we used a neural network approach derived from sequential recursive auto-associative memory to parse the wavelet coefficients and hierarchy data. This method is fully described elsewhere [41]. In this case, because the wavelet coefficients are continuous, linear output instead of sigmoidal output was used. This variation is therefore called linear output sequential recursive auto-associative memory, or LOSRAAM.

Because the objective of training the LOSRAAM net-



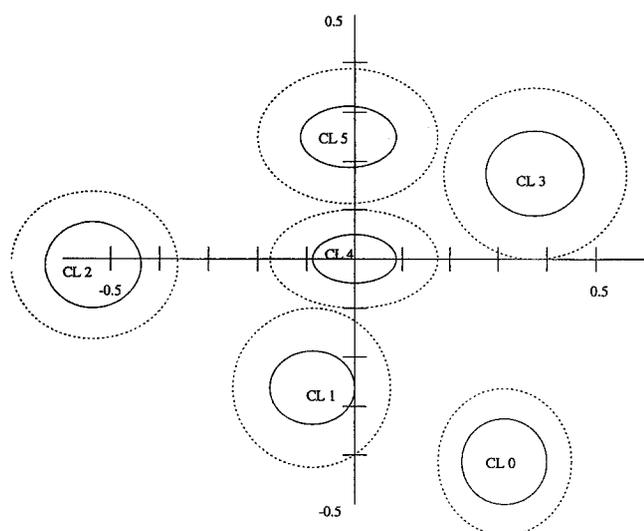
**FIGURE 3.** LOSRAAM neural network. All units in the input layer are connected to both of the units in the hidden layer (solid arrows). Each of the two hidden units is connected to each of the units in the output layer (solid arrows). In addition, each of the hidden units is connected in feed-back to the two context input-output units in the input layer (dashed arrows). The objective of the network is to set each of the corresponding units in the output layer equal to the values in the units in the input layer. (See text for more explanation.)

work is to have the output exactly match the input, the network is constructed by using an input layer and a mirror-image output layer. Each consisted of three input-output units, which corresponded to the values in wavelet triplets, and two context units (Fig 3). Between these two layers is a layer with two hidden units. The number of hidden units chosen for the LOSRAAM architecture was determined empirically by testing architectures with one to three units. Use of two units proved superior to either the one- or three-unit experiments. The values of the context units in the input layer are set by using the values developed in the hidden units from input of the immediately preceding wavelet coefficient triplet. The values of the initial context units input with the first wavelet coefficient triplet in the series are set randomly. Each of the three layers is fully connected with adjacent layers in a feed-forward direction (Fig 3). Training is carried out on the entire data set until a threshold is reached and no better prediction can be obtained. Because this network is used as a classification system for the wavelet coefficients, it is not tested on data not used in training the network.

Once this network is fully trained, the coefficients and their context units can be given to the LOSRAAM network and it will predict the preceding wavelet coefficient within a predetermined error tolerance. In this

sense, the context units associated with each coefficient triplet become operators that predict the prior adjacent wavelet coefficient triplet and its context units. Thus, given any but the first wavelet coefficient information, the network can approximately predict the prior wavelet coefficient, its hierarchical level, its position within the level, and the two context units that will predict the next preceding wavelet coefficient triplet. Thus, the entire set of wavelet coefficients can be predicted from the last pair of hidden (context) units.

The context unit pairs arising from serial evaluation of the wavelet coefficient triplets were collected as two-dimensional vectors. These were subjected to cluster analysis according to a method modified from Fu [42]. Fu's method was modified to handle data that clustered at any point, including at the origin. This analysis, with use of a sensitivity factor of 0.5, yielded six identifiable and discrete states (Fig 4, Table 1). From these, a six-element feature vector was created for each image [43]. Each element in the feature vector represented the number of times the corresponding state from the above cluster analysis was found in each image. Representing each image as a six-element feature vector, with each element contained in 4 bytes, gives a final compression ratio for the data set of 43,691:1.



**FIGURE 4.** Graph of LOSRAAM clusters. Solid lines represent 1 standard deviation and dotted lines represent 2 standard deviations of point distribution about the cluster centroid. Each of the six clusters developed during LOSRAAM analysis is distinct from the others. CL = cluster.

### ANN Architecture

Finally, in two separate experiments, two sets of feed-forward neural networks (FFNNs) were trained to classify the image pairs corresponding to each case for the presence of a mass. The FFNN architectures included six input units, which corresponded to the scalars from each dimension of the image feature vectors, and five output units. A middle layer with hidden units was also used.

The hidden unit layer contained two units in the first experiment and four units in the second experiment. Although derived empirically, these numbers of hidden units ensure that overfitting of the data will not occur. This claim is based on work by MacKay [44], in which an Occam factor (OF) is computed as follows:  $(\text{patterns} \times \text{EO}) / \text{EP}$ , where  $\text{EO} = \log_2(\text{number of output categories})$  is the number of effective outputs and EP the effective parameters. Effective parameters are a function of the network architecture and depend on the number of effective outputs, the number of hidden units ( $H$ ), and the number of effective inputs (EI). The effective input is the number of inputs remaining after application of singular value decomposition to the data to determine linear independence. The effective parameter is related to the number of adjustable weights in the network and is calculated as follows:  $\text{EI} \times \text{EO} + H \times (\text{EI} + \text{EO}) + H + \text{EO}$ . An Occam factor greater than 3 is believed to be required to avoid overfitting of the train-

**TABLE 1: Results of LOSRAAM Clustering of Wavelet Coefficients**

Cluster No.	Population*	Vector State		Standard Deviation‡
		Centroid†		
0	42,384	0.309, -0.418	0.100, 0.081	
1	12,661	-0.074, -0.270	0.090, 0.081	
2	32,401	-0.531, -0.045	0.093, 0.064	
3	55,648	0.372, 0.185	0.117, 0.085	
4	6,409	-0.006, 0.005	0.087, 0.051	
5	25,338	-0.022, 0.248	0.096, 0.065	

\*Number of wavelet coefficients falling within the cluster.

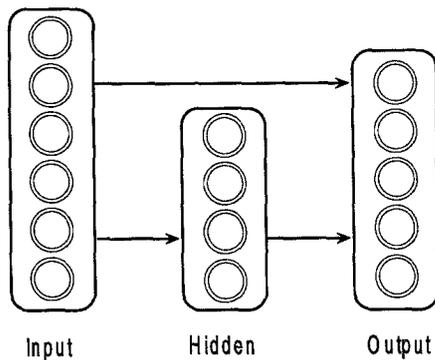
† The centroid of the cluster is expressed as a two-dimensional vector.

‡ Standard deviations of the centroids along each of the unit axes.

ing data. Our networks have an Occam factor approximately equal to 8.0 and 5.5 for the first and second experiments, respectively.

The five output units were divided into two groups, the first containing two categorization units (mass and no mass) and the second containing three hint units [45, 46] that indicated the location of a mass, if any. During training, the mass and no-mass outputs were initially targeted at either on (+1) or off (-1), depending on the presence of a mass. The output from these two nodes form a two-dimensional vector that we configured to allow vector best-match criteria for the network's decision. The direction cosine was calculated for each output vector. The network's prediction was then classified according to the ideal classification vector with which the output vector had the largest direction cosine. This method, which uses two nodes and hence a two-dimensional vector, avoids the problem of a possible zero vector for which interpretation is difficult.

The location data in the hint units of the output layer were used solely as targets during network training. These targets provided additional constraints so that training was guided properly in classifying images. Because this location information was not provided as input data, the three hint units and the weights connecting them to the other layers may be discarded after training without any effect on the performance of the network. In addition, when unknowns from the test set are presented to the network, these three units obtain discrete values for each test case presented and so potentially may be used to localize lesions. Although le-



**FIGURE 5.** Map of current ANN architecture shows three layers: input, hidden, and output. These layers contain six, four, and five units, respectively. Three of the output units are hint units and two are classifiers (ie, mass and no mass). The input layer is connected in a forward direction to the hidden and output layers, whereas the hidden layer is connected only to the output layer.

sion localization is therefore possible with this design, this experiment was not designed to evaluate this feature of the networks and so the hint units were discarded in the network testing phase.

The first hint unit distinguished between the upper and lower half of the breast, the second unit distinguished between the outer and inner half of the breast, and the third unit specified a subareolar versus a central location. The only valid combinations were upper outer, upper inner, lower outer, lower inner, subareolar, and central. For the purpose of training, hint units not relevant to the location designation were initially targeted at zero, whereas relevant hint units were targeted at either on (+1) or off (-1). For no-mass cases, all three units were targeted to zero and did not contribute to weight updates during training. In our experience, hints have proved to be an invaluable aid to training because they serve to constrain the number of possible solutions the ANN can achieve. The units of the layers of the FFNNs were totally connected with weights to the units of any layer forward in the network architecture (Fig 5). This means that the input layer was connected to both the output layer and the hidden layer, whereas the hidden layer was connected only to the output layer. ANNs with only one layer of weights are called perceptrons [47] and are capable of solutions only in the form of hyperplanes that divide the input space into linearly separable regions [48, 49]. Even highly nonlinear problems, such as the interpretation of mammograms, have some linearly separable ele-

**TABLE 2: Distribution of Feature Vectors among Clusters according to Mass and No-Mass Images**

Mammogram	Cluster No.				Total
	0	1	2	3	
Mass	26	14	18	0	58
No-mass	25	16	10	1	52
Total	51	30	28	1	110

**TABLE 3: FFNN Prediction of Mass on Test Set Cases from Experiment 1**

Run No.	Threshold No. of Votes Required to Diagnose a Mass*							
	5	6	7	8	9	10	11	12
Presence of Mass Correctly Predicted Out of 8 Possible Cases with Masses								
1	6	6	6	6	6	5	5	5
2	5	5	5	5	5	5	5	5
3	7	6	6	6	6	5	5	5
4	6	6	6	6	6	6	6	5
Average sensitivity (%)	75	72	72	72	72	66	66	63
Absence of Mass Correctly Predicted Out of 5 Possible Cases with Masses								
1	2	2	2	2	3	3	3	3
2	2	2	2	2	2	2	2	2
3	2	2	2	2	3	4	5	5
4	2	2	2	2	2	2	2	4
Average specificity (%)	40	40	40	40	40	55	60	70

\*Total number of votes was 18 (nine networks with two images per case).

ments. This architecture, by connecting inputs directly to outputs, provides these linear components with direct representation in the overall solution.

**FFNN Training**

Our approach to training neural networks for a variety of applications is described elsewhere [50]. FFNN training was performed by using the conjugate gradient method modified for use with a self-scaling error function, optimized to reduce the number of derivative computations, and monitored for restart conditions with a technique described by Powell [51]. A more thorough discussion of our training methods is provided elsewhere [46].

**TABLE 4: FFNN Prediction of Mass on Test Set Cases from Experiment 2**

Run No.	Threshold No. of Votes Required to Diagnose a Mass*				
	8	9	10	11	12
Presence of Mass Correctly Predicted Out of 4 Possible Cases with Masses					
1	4	6	6	6	6
2	3	5	5	5	5
3	4	6	6	6	6
4	2	6	6	6	6
5	3	2	1	1	0
6	4	4	4	4	3
Average sensitivity (%)	83	79	71	71	63
Absence of Mass Correctly Predicted Out of 3 Possible Cases without Mass					
1	2	2	2	2	3
2	2	2	2	2	2
3	2	2	2	2	3
4	2	2	2	2	2
5	0	0	1	1	1
6	2	2	2	2	2
Average specificity (%)	39	50	56	67	67

\*Total number of votes was 18 (nine networks with two images per case).

Two separate experiments were performed. In the first, a test set containing 26 images (13 cases, eight with mass and five without) was created. In each case, the remaining images were used for training. Before training, we used cluster analysis with a sensitivity factor of 0.5 to analyze the feature vectors representing the images [42]. This analysis showed that the images were contained within four distinct and separate classes (109 images were within three classes and one image was in the fourth) (Table 2). We constructed the test sets from among the 55 cases by using the following technique. Image pairs that made up an individual case were segregated according to their classification pairing. This yielded six classification pairs. Cases were selected at random from classification pair groups. This technique ensured equal representation of the classifications among the test set groups. The test sets for the two experiments did not enter into the training process in any way and were used solely for unbiased testing of the networks generated during training.

Those cases not selected for the test set in each experiment were used for training. In each experiment network training was repeated nine times, each time with initial randomization of the network input values. This process yielded nine distinct ANNs. Training was terminated when the performance of the network was optimized for both the mass and no-mass images in the training set (ie, no better prediction of cases could be obtained). Although this method of determining completion of network training may have created networks that were not optimal in terms of their ability to generalize, it eliminated the introduction of training bias.

The resulting collection of nine networks was then given the test set images for analysis. Each network produced one vote per image as to the presence or absence of a mass. Thus, each case received 18 votes. The accuracy of the networks at various voting thresholds was examined. The first experiment was repeated four times and the second experiment was repeated six times, and a new random test set was selected each time. These experiments were designed to measure the stability of the network predictions by using different-sized test sets.

## RESULTS

Analysis of mass and no-mass images failed to reveal discrete separation of the two according to clusters at either the LOSRAAM or the feature vector level of data processing. Both mass and no-mass images were approximately equally distributed among three of the four feature vector states (Table 2). The fourth cluster contained a feature vector from only a single image. No statistically significant bias was found in this distribution to suggest that any one cluster correlated with images with masses or images without masses ( $P = .3734$ ,  $\chi^2$  test). This finding is consistent with the nonlinearity of the problem of mass detection and suggests that image data cannot be separated into simpler elements.

Tables 3 and 4 show the predictions of the nine networks from each experiment on the test set cases at different voting thresholds for each separate run of the experiments. The two images corresponding to the two views from the same breast for each test set case are shown together. The tables are divided horizontally, showing the number of correct predictions for mass cases in the top half and the number of correct predictions for the no-mass cases in the bottom half.

We evaluated the predictions from the first experi-

ment (Table 3) for the likelihood that the FFNNs are making their predictions on the basis of relevant information content on the two-view mammograms. We calculated the  $\chi^2$  for the network's predictions of a mass when a mass was actually present and when a mass was not present (Tables 5, 6). Results of the  $\chi^2$  test show that when a mass is present on the mammogram, the networks predict its presence statistically significantly more often than would be expected by random chance ( $P < .01$ ). Conversely, if no mass is present, the prediction of the presence of a mass is random ( $P = .6869$ ). These findings indicate that the networks are evaluating image content that is directly related to the presence of a mass.

## DISCUSSION

Our preliminary results show that, even with small data sets, a moderate degree of sensitivity (79%) can be obtained over repeated experiments while maintaining a reasonable level of specificity (50%) (Tables 3, 4). Repeated experiments with two separate-sized test sets show that results are consistent and therefore reproducible. Although the difference was not statistically significant, the networks trained on the larger training set appeared to perform slightly better overall than the networks trained on the smaller training set. There appears to be more variance among the predictions of the former than the latter, however. The exact reason for this is not apparent from these data.

The ultimate goal of computer-aided mammogram diagnosis is to develop a prescreening tool that (a) reduces the number of normal mammograms that a radiologist must read and (b) directs the radiologist's attention to specific areas of suspicion. The first part of these goals requires that sensitivity be maximized. Although the sensitivity of combined wavelet-ANN approach is not currently adequate for this purpose, the use of a postprocessing voting scheme with multiple networks allows the sensitivity and specificity of the system to be adjusted to the clinical indication. Thus, in the case of tumor screening, voting thresholds can be chosen at the highest vote threshold that allows near-perfect sensitivity.

Creation of a clinically useful automated prescreening device will require a system (not necessarily based on ANNs) that can detect all mammographic manifestations of malignancy (eg, masses and clustered microcalcifications). Nishikawa et al [38] have suggested that

**TABLE 5: FFNN Prediction of the Presence of a Mass from Paired Orthogonal Mammographic Views When a Mass Is Present**

Paired Prediction*	Observed	Expected	$\chi^{2\dagger}$
PP	133	116	2.49
PN	100	134	8.63
NN	55	38	7.61

\*N = no mass, P = mass.

† Total = 18.73 (2df:  $P < .01$ ).

strategies for computer-aided diagnosis of mammographic lesions combine separate algorithms to evaluate each of these two harbingers of malignancy. With use of wavelet-ANN technology, two separate networks could be used to prescreen mammograms—one to screen for masses and the other for clustered microcalcifications.

Although specificity is somewhat less important to such a device, its optimization cannot be considered a secondary issue. The minimum acceptable level of specificity will be determined on the basis of the clinical situation and the number of linked networks. Assuming that each network screens the target mammograms entirely independently of the others in the system, their ability to reduce work flow is exponentially related to their individual specificities. In the case of two networks, one screening for masses and the other for clustered microcalcifications, individual network specificities must be substantially better than 50%. Otherwise, in a worst case scenario, there could very well be no real reduction in the number of mammograms viewed by the radiologist. Hence, our current specificities are not sufficient to support clinical use.

High false-positive rates with computer-aided diagnosis techniques have been a problem for the detection of both masses and microcalcifications on digitized mammograms, in the range of two to three per image [6-11, 37, 52]. With use of a majority voting scheme on only the test set images, the false-positive rate (by case) was 50% in each experiment. This performance is marginal in terms of being able to reduce the number of mammograms that must be viewed by a radiologist assuming similar performance for a system detecting microcalcifications. However, we emphasize that these data are preliminary. It is likely that with larger data sets, improved image resolution, and more sophisticated wavelet transforms, the specificity of this technique can be improved substantially.

**TABLE 6: FFNN Prediction of the Presence of a Mass from Paired Orthogonal Mammographic Views When No Mass Is Present**

Paired Prediction*	Observed	Expected	$\chi^{2†}$
PP	38	34	0.47
PN	80	88	0.73
NN	62	58	0.28

\*N = no mass, P = mass.

† Total = 1.48 (2df:  $P = .4771$ ).

Because of the nature of network training and assuming no overt bias in our case selection, the results with a small pilot data set such as this one should represent a lower bound of performance using our technique. A larger data set will allow more representative training sets thereby permitting better ANN training and presumably better generalization. The larger the data set, the greater will be its variety. The greater the data set variety, the less will be the expected bias in the networks from the point of view of both training and testing. Additionally, testing different wavelet transform algorithms, feature vector designs, network architectures, and voting schema may yet lead to further improvement in the performance of this architecture. Thus, this technology must be tested and validated on a large data set for further validation of our current promising results.

The complexity of data processing in our current methods is dictated by several factors. First, to make image data practical for ANN analysis, its magnitude must be greatly decreased without destroying the essential information contained within the image. Thus, wavelet transformation and stratification of the data become important. Second, the data must be handled in such a way as to make analysis clinically practical and eventually cost-effective by using available hardware. Third, the system must allow for the possibility of both linear and nonlinear information as part of the solution. Non-linearity in mass identification is apparent from the ability of the FFNN to detect masses despite the lack of correlation among the four feature vector states with the presence of a mass on the images (Table 2). Fourth, data handling must be designed to allow for adjustable sensitivity and specificity depending on the problem posed. Finally, the system should show minimal variance over a given data set. The introduction of voting schemes that use multiple ANNs helps address these last two issues.

A major drawback of our method is its inability to

pinpoint the suspicious region within the breast and indicate this area on the image (although it is theoretically able to generally localize an abnormality to within a quadrant of the breast). Since many mammographic abnormalities are subtle and can be missed, even under rigorous conditions of evaluation, the ability of a screening device to direct attention to suspicious areas of a mammogram may increase early or subtle lesion detection and, hence, improve patient survival. Because of the high degrees of compression employed in our current techniques, mass localization on the image is not possible. Modification of this technique to process smaller areas of the images may provide an ability to localize lesions more specifically. Indeed, this partial image processing may need to be done only when the initial screen of the entire image has located a suspicious area. This would ultimately save on processing time and allow a greater throughput of mammograms.

Although preliminary, our results suggest that the combination of wavelet and ANN technology has promise as a method to eventually screen mammograms for masses. There are a number of modifications of our approach that may further improve both the neural network's sensitivity and specificity for mass detection. Efforts for improvement may be directed toward improving the wavelet transform used for initial image analysis; changing the wavelet truncation threshold, altering the LOSRAAM architecture; changing the sensitivity factor for feature vector analysis; and improving the design of the FFNNs. Many of these experiments will require a larger data set for testing. In addition, the overall strategy of analysis may be modified (eg, by training the network on specific image abnormalities within the data and then testing on entire images). Finally, testing can be improved with larger data sets by excluding test cases from all phases of training. We believe that further investigation of our combined wavelet-ANN approach is warranted.

#### ACKNOWLEDGMENT

The authors express their appreciation to Andrew Laine, PhD, for providing the edge-enhancing wavelet transform used in the preliminary analysis of our data.

#### REFERENCES

1. Miller BA, Ries LAG, Hankey BF, et al. SEER cancer statistics review: 1973–1991. National Institutes of Health publication no. 94-2789. Bethesda, MD: National Cancer Institute, 1994.
2. Feuer EJ, Wun LM, Boring CC, et al. The lifetime risk of developing breast cancer. *J Natl Cancer Inst* 1993;85:892–897.

3. Mettlin C, Smart CR. Breast cancer detection guidelines for women aged 40 to 49 years: rationale for the American Cancer Society reaffirmation of recommendations. *CA Cancer J Clin* **1994**;44:248–255.
4. Dodd GD. American Cancer Society guidelines from the past to the present. *Cancer* **1993**;72(suppl 4):1429–1432.
5. Volkens N. NCI replaces guidelines with statement of evidence. *J Natl Cancer Inst* **1994**;86:14–15.
6. Lau TK, Bishof WF. Automated detection of breast tumors using the asymmetry approach. *Comput Biomed Res* **1991**;24:273–295.
7. Chan HP, Doi K, Vybomy CJ, Lam KL, Schmidt RA. Computer-aided detection of microcalcifications in mammograms: methodology and preliminary clinical study. *Invest Radiol* **1988**;23:664–671.
8. Yin FF, Giger ML, Vybomy CJ, Doi K, Schmidt RA. Comparison of bilateral-subtraction and single-image processing techniques in the computerized detection of mammographic masses. *Invest Radiol* **1993**;28:473–481.
9. Kegelmeyer WP Jr, Pruneda JM, Bourland PD, Hillis A, Riggs MW, Nipper ML. Computer-aided mammographic screening for spiculated lesions. *Radiology* **1994**;191:331–337.
10. Yin FF, Giger ML, Doi K, Metz CT, Vybomy CJ, Schmidt RA. Computerized detection of masses in digital mammograms: analysis of bilateral subtraction images. *Am Assoc Phys Med* **1991**;18:955–963.
11. Nishikawa RM, Giger ML, Doi K, et al. Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms. *SPIE* **1993**;1905:422–432.
12. Brodie I, Gutcheck RA. Radiographic information theory and application to mammography. *Med Phys* **1982**;9:79.
13. Laine A, Schuler S, Fan J. Mammographic feature enhancement by multiscale analysis. *IEEE Trans Med Imaging* **1994**;13:1–28.
14. Laine A. Multiscale wavelet representations for mammographic feature analysis. Presented at the Proceedings of National Cancer Institute Breast Imaging Workshop: State-of-the-Art and New Technologies, Image Enhancement Techniques: Computer Science, Bethesda, MD, September 4–6, **1991**.
15. Laine A, Song S. Multiscale wavelet representations for mammographic feature analysis. Presented at the Conference on Mathematical Methods in Medical Imaging, San Diego, CA, July 23–25, **1992**.
16. Laine A, Song S. Wavelet processing techniques for digital mammography. Presented at the Conference on Visualization in Biomedical Computing, Chapel Hill, NC, October 13–16, **1992**.
17. Laine A, Schuler S. Hexagonal wavelet processing of digital mammography. *Proc SPIE* **1993**;1989:559–573.
18. Laine A, Schuler S, Fan J. Mammographic feature enhancement by multiscale analysis. *IEEE Trans Med Imaging* **1994**;13:1–28.
19. Wickerhauser MV. Applied wavelet analysis from theory to software. Wellesley, MA: Peters, **1994**.
20. Wei D, Chan HP, Helvie MA, et al. Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis. *Med Phys* **1995**;22:1503–1513.
21. Qian W, Kallergi M, Clarke LP, et al. Tree structured wavelet transform segmentation of microcalcifications in digital mammography. *Med Phys* **1995**;22:1247–1254.
22. Qian W, Clarke LP, Li HD, Clark R, Silbiger ML. Digital mammography: M-channel quadrature mirror filters (QMFs) for microcalcification extraction. *Comput Med Imaging Graph* **1994**;18:301–314.
23. Lucier BJ, Kallergi M, Qian W, et al. Wavelet compression and segmentation of digital mammograms. *J Digit Imaging* **1994**;7:27–38.
24. Clarke LP, Kallergi M, Qian W, Li HD, Clark RA, Silbiger ML. Tree-structured nonlinear filter and wavelet transform for microcalcification segmentation in digital mammography. *Cancer Lett* **1994**;77:173–181.
25. Golberg MA, Pivovarov M, Mayo-Smith WW, et al. Application of wavelet compression to digitized radiographs. *AJR* **1994**;163:463–468.
26. Wu Y, Giger ML, Vybomy CJ, Schmidt RA, Metz CE. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology* **1993**;187:81–87.
27. Goldberg V, Manduca A, Ewert DL, Gisoid JJ, Greenleaf JF. Improvement in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence. *Med Phys* **1992**;19:1475–1481.
28. Wu Y, Doi K, Giger ML, Nishikawa RM. Computerized detection of clustered microcalcifications in digital mammograms: applications of artificial neural networks. *Med Phys* **1992**;19:555–560.
29. Boone JM, Gross GW, Greco-Hunt V, Greenberg B. Neural networks in radiologic diagnosis. II. Introduction and illustration. *Invest Radiol* **1990**;25:1012–1016.
30. Gross GW, Boone JM, Greco-Hunt V, Greenberg B. Neural networks in radiologic diagnosis. II. Interpretation of neonatal chest radiographs. *Invest Radiol* **1990**;25:1017–1023.
31. Raff U, Newman FD. Lesion detection in radiologic images using an autoassociative paradigm: preliminary results. *Med Phys* **1990**;17:234–240.
32. Boone JM, Sigillito VG, Shaber GS. Neural networks in radiology: an introduction and evaluation in a signal detection task. *Med Phys* **1990**;17:234–241.
33. Reinus WR, Wilson A, Kwasny S, Kalman B. Parallel distributed processing systems as a means of diagnosing solitary lesions of bone. *Invest Radiol* **1994**;29:606–611.
34. Miller AS, Blott BH, Hames TK. Review of neural network applications in medical imaging and signal processing. *Med Biol Eng Comp* **1992**;30:449–464.
35. Parks RW, Long DL, Levine DS, et al. Parallel distributed processing and neural networks: origins, methodology, and cognitive functions. *Int J Neurosci* **1991**;60:195–214.
36. Suckling J, Dance DR, Moskokvic E, Lewis DJ, Blacker SG. Segmentation of mammograms using multiple linked self-organizing neural networks. *Med Phys* **1995**;22:145–152.
37. Chan H, Lo SB, Sahiner B, Lam KL, Helvie MA. Computer-aided detection of mammographic microcalcifications: pattern recognition with an artificial neural network. *Med Phys* **1995**;22:1555–1567.
38. Nishikawa RM, Giger ML, Doi K, et al. Computer-aided detection and diagnosis of masses and clustered microcalcification from digital mammograms. *SPIE* **1993**;205:422–432.
39. Newman WF, Sproull R. Principles of interactive computer graphics. New York, NY: McGraw-Hill, **1979**;402–404.
40. Mallat S, Zhang S. Characterization of signals from multiscale adages. *IEEE Trans Pattern Anal Machine Intell* **1992**;14:710–732.
41. Kwasny SC, Kalman BL. Tail-recursive distributed representations and simple recurrent networks. *Connect Sci* **1995**;7:61–80.
42. Fu L. Neural networks in computer intelligence. St Louis, MO: McGraw-Hill, **1994**.
43. Dai X. Wavelet applications in process sensor data analysis. Doctoral dissertation. St Louis, MO: Washington University, **1996**.
44. MacKay DJC. Bayesian interpolation. *Neural Computation* **1992**;4:415–447.
45. Abu-Mostafa YS. Machines that learn from hints. *Sci Am* **1995**;272:64–69.
46. Yu Y, Simmons RF. Extra output biased learning. Technical report AI90-128. Austin, TX: University of Texas, March **1990**.
47. Minsky ML, Paper SA. Perceptrons. Cambridge, MA: MIT Press, **1988**.
48. Hertz J, Krogh A, Palmer R. Introduction to the theory of neural computation. Redwood City, CA: Addison-Wesley, **1991**.
49. Weiss SM, Kulikowski CA. Computer systems that learn. San Francisco, CA: Morgan Kaufmann, **1990**.
50. Kalman B, Kwasny SC. High performance training of feed-forward and simple recurrent networks. *Neurocomputing* **1997**;14:63–84.
51. Powell MJD. Restart procedures for the conjugate gradient method. *Math Programming* **1977**;12:241–254.
52. Lu P, Giger ML, Schmidt RA, et al. Computer-aided detection of masses on digital mammograms: processing and feature extraction with multiple-gray-level resolution (abstr). *Radiology* **1993**;189(P):186.