Texture Analysis of Breast Tumors on Sonograms

Dar-Ren Chen, Ruey-Feng Chang, Yu-Len Huang, Yi-Hong Chou, Chui-Mei Tiu, and Po-Pang Tsai

We performed a feasibility study to determine if the texture features extracted from sonograms can be used to predict malignant or benign breast pathology by the proposed artificial neural network and to compare the diagnostic results with the radiologists’ results. A total of 1,020 images (4 different rectangular regions from the 2 orthogonal imaging planes of each tumor) from 255 patients were used as samples. When a sonogram was performed, 1 physician identified the region of interest in the sonogram; then, a neural network model, using 24 autocorrelation texture features, classified the tumor as benign or malignant. Three radiologists who were unfamiliar with the samples also classified these images. The receiver operating characteristic (ROC) area index for the proposed neural network system is 0.9840 ± 0.0072. The neural network identified 35 of 36 malignancies and 211 of 219 benign tumors using all 4 regions of interest. The radiologists, on average, identified 19 of 36 malignancies, with 12 tumors called indeterminate and 4 tumors called benign. We conclude that benign and malignant breast tumors can be distinguished using interpixel correlation in digital ultrasonic images.

Copyright © 2000 by W.B. Saunders Company

THE CURRENT PROCEDURES for detection and diagnosis of breast cancer show the difficulty of both maximizing sensitivity and specificity. Stavros et al.1 in 1995 and Skaane and Engedal2 in 1998, found the sensitivity of breast ultrasound for malignancy was 98.4% and 99.55%, the specificity was 67.8% and 29%, the positive predictive value (PPV) was 38% and 66%, and the negative predictive value (NPV) was 99.5% and 98%, respectively. In a majority of articles, the positive biopsy rate for cancer is low, between 10% and 31%.3-5; most mammographers would probably agree that too many biopsies (both needle and excisional) are performed for benign breast disease.6 Therefore, the challenge facing breast imagers is how to decrease the number of benign biopsies without missing the detection of cancer. How to improve feature analysis and refine criteria of recommendation for biopsy is important. Ideally, imaging follow-up (ie, follow-up at 6 months) could be performed rather than biopsy.

Recent progress of computer-aided diagnosis (CAD) in mammography and ultrasonography has suggested that neural networks can assist physicians in diagnosis.7-9 Neural networks are computer programs that learn to make diagnostic predic-

From the Departments of General Surgery and Radiology, China Medical College & Hospital, Taichung, Taiwan; the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan; and the Department of Radiology, Veterans General Hospital, Taipei, Taiwan.

Address reprint requests to Dar-Ren Chen, MD, Department of General Surgery, China Medical College & Hospital, 2 Ver-Der Rd, Taichung, Taiwan. E-mail: dchen88@mails13.hinet.net

Copyright © 2000 by W.B. Saunders Company
0887-2171/00/2104-0004$10.00/0

Seminars in Ultrasound, CT, and MRI, Vol 21, No 4 (August), 2000: pp 308-316
lesion using neural networks. Because the shape of the tumor is not spherical and the sonographic texture is heterogeneous in multiple scan planes, we did not know what type of region would be most useful for image analysis, or how many views of a tumor should be used. We decided that regardless of which method is used, the breast needs to be examined in at least 2 orthogonal imaging planes when a mass is identified.

A total of 255 patients with tumors were scanned in 2 planes and were reviewed in a blind study by 3 radiologists experienced in breast ultrasound. They were asked to classify these tumors on a scale of 1 to 3 as follows: 1 = benign, 2 = indeterminate, 3 = malignant. The sensitivity, specificity, and predictive values of negative and positive results and overall accuracy were calculated for the classification.

MATERIALS AND METHODS

Data Acquisition

The inclusion criterion of these 255 consecutive patients was the identification of a tumor on sonographic examination. The images were collected from March 1, 1998 to April 30, 1999, by 1 breast surgeon; the patients’ ages ranged from 18 to 68 years (mean, 36 years); tumors were from 0.8 to 3.6 cm in diameter (mean, 1.8 cm). A total of 1,020 sonograms of ROIs from 255 patients including 36 cancers, 57 fibrocystic nodules, 118 fibroadenomas, 1 intraductal papilloma, 3 lipomas, 38 cysts, and 2 galactoceles were used as case samples. (Note that each case contained 4 ROI images in the analysis.) The ultrasonic appearances were then correlated either with the fine-needle aspiration, core biopsy, or surgical excision. The breast was scanned transversely and longitudinally. The ultrasonic images were captured in 2 orthogonal imaging planes of each tumor (Fig 1). Sonography was performed using an Aloka SSD 1200 (Tokyo, Japan) scanner and a 7.5-MHz linear transducer with freeze-frame capability. No acoustic standoff pad was used in any of the samples. The sonographic gain setting remained unchanged throughout the entire period of study, except for changes made to obtain the best view. When a sonogram is performed, an analog video signal is transmitted from the VCR output of the scanner to a portable notebook computer; the data was then digitized by a frame grabber Video CATcher (Top Solution Technology Co., Taipei, Taiwan) that is connected to the printer port of the computer. The capturing resolutions of the portable computer and the external frame grabber are 736 × 566 pixels for an National Television Standards Committee (NTSC) video screen picture. The study used the Prolab’s ProImage software package, which is bundled with the frame grabber, to capture the real-time digital image. The monochromatic ultrasound image is digitized into 8 bits (ie, 256 gray levels) with a resolution of 58 × 58 pixels in a 1-cm × 1-cm rectangle (Fig 1). In this study, 4 different rectangular regions from the 2 orthogonal imaging planes of each tumor were used for analysis. These were: (1) a region that extended beyond the lesion margins by 1 to 2 mm in all directions (shown as TA and LA.

Fig 1. Breast scanned transversely and longitudinally. The 4 ROI images for a tumor: (A) full longitudinal ROI (LA), (B) inside longitudinal ROI (LB), (C) full transverse ROI (TA), and (D) inside transverse ROI (TB).
in Fig 1), and (2) the largest rectangular region that would fit inside the lesion (shown as TB and LB in Fig 1). This artificial neural network relies on the physician locating all sonographic abnormalities. The features of the 4 ROI images are extracted to present the texture information and the tumor is classified as probably benign or probably malignant by a hierarchical neural network (HNN) diagnostic system using the autocorrelation features.

**Image Analysis**

This study used the correlation between neighboring pixels within the images as classifying features of the tumor. The normalized autocorrelation coefficients\(^{18}\) can be used to reflect the interpixel correlation within an image. The two-dimensional (2-D) normalized autocorrelation coefficient between pixel \((i, j)\) and pixel \((i + \Delta m, j + \Delta n)\) in an image with size \(M \times N\) can be defined as

\[
\gamma(\Delta m, \Delta n) = \frac{A(\Delta m, \Delta n)}{A(0, 0)},
\]

where

\[
A(\Delta m, \Delta n) = \frac{1}{(M - \Delta m)(N - \Delta n)} \sum_{x=0}^{M-1-\Delta m} \sum_{y=0}^{N-1-\Delta n} f(x, y)f(x + \Delta m, y + \Delta n).
\]

Moreover, the 2-D autocorrelation coefficients are further modified into a mean-removed version to generate the similar autocorrelation features for images with different brightness but with a similar texture. This modified version is expressed as

\[
A'(\Delta m, \Delta n) = 1 - \frac{1}{(M - \Delta m)(N - \Delta n)} \sum_{x=0}^{M-1-\Delta m} \sum_{y=0}^{N-1-\Delta n} |f(x, y) - \bar{f}| (f(x + \Delta m, y + \Delta n) - \bar{f}),
\]

where \(\bar{f}\) is the mean value of \(f(x, y)\). The absolute value is adopted in the earlier equation because when the gray level of a pixel subtracts the mean, a negative value may be produced. This work uses the modified version of the 2-D normalized autocorrelation matrix for the input of the neural network.

The dimension of the matrix can be fixed for any size of image. In this article, because both \(\Delta m\) and \(\Delta n\) are 5, processing an ultrasonic image produces a \(5 \times 5\) autocorrelation matrix (ie, 25 autocorrelation coefficients). The value of \(\gamma(0, 0)\) is always 1 for a normalized autocorrelation matrix. Thus, except for the element \(\gamma(0, 0)\), other autocorrelation coefficients are formed as a 24-D image feature vector.

**Multiview HNN Diagnostic System**

An HNN model\(^{19}\) consists of functionally similar or different neural network models called *component networks* concatenating one to another, as shown in Fig 2. The component networks learn separately but coherently. That is, the output of a component network is taken as the input of the component network on top of it. The HNN model will provide the powerful capability of information abstraction.

Accordingly, this study uses a 2-phase HNN model to combine the information in the 4 ultrasonic ROI images and then to differentiate between benign and malignant tumors. Four MLP networks are used as the component networks in the first phase. Each component MLP network recognizes the breast tumors using the corresponding ROI image. The output values of these 4 MLP networks are taken as the input of a component MLP network in the final phase. The output value of the second phase MLP network is used to decide whether a tumor is benign or malignant. When the output value of a multiview ultrasonic breast image is near enough to 1, the system will classify it as malignant. On the contrary, when the output value is close to 0, the tumor will be diagnosed as benign. The output of networks ranged continuously from 0 to 1. Figure 3 shows the structure of the proposed HNN tissue classification model. Each MLP neural network in the first phase contains 25 input nodes, 10 hidden nodes, and a single output node. The final phase MLP network comprises 5 input nodes, 2 hidden nodes, and a single output node.

**Training and Testing**

The k-fold cross validation method\(^{20}\) was used to evaluate the performance of the HNN diagnostic system. The 255 samples in the database are randomly divided into k groups. In this study, k equals 10. The test samples of the first group were removed from the 255 training samples. The remaining samples were trained, and the output of the test
group samples was determined. The test samples were rejoined with the other training samples, and the second group was removed and used as the test group. This process was repeated until all groups were used once as the test group. In this method, all groups were eventually used to both train and test the network. The results of the HNN diagnostic system and each MLP diagnostic subsystem were plotted as receiver operating characteristic (ROC) curves (software package LABROC1 by Professor C. E. Metz, University of Chicago). The diagnostic accuracy for the neural network was compared with the results for the radiologists’ impression of the likelihood of malignancy.

RESULTS

The LA MLP diagnostic subsystem has an $A_z$ value of 0.9495 ± 0.0165 (SD), as shown in Fig 4A. The LB MLP diagnostic subsystem obtains an $A_z$ value of 0.8605 ± 0.0442, as shown in Fig 4B. The TA MLP diagnostic subsystem has an $A_z$ value of 0.9237 ± 0.0296, as shown in Fig 4C. The TB MLP diagnostic subsystem has an $A_z$ value of 0.7348 ± 0.0553, as shown in Fig 4D. Finally, the HNN diagnostic system achieved an area under the ROC curve of 0.9840 ± 0.0071, as shown in Fig 5. Figure 6 shows the false-negative rate and false-positive rate. Figure 7 shows the first screen view of the proposed HNN diagnostic system. In this screen, the 4 ROI images would be loaded. Then, the autocorrelation matrices could be produced, as shown in Fig 8. Finally, the output value of the HNN model is used to decide whether a tumor is benign or malignant. The diagnostic screen is given in Fig 9. In this program, the output value of the HNN diagnostic system is multiplied by 100.
**Performance of Proposed HNN Compared With 3 Experienced Radiologists**

With a threshold of 0.35, the network correctly identifies 35 of 36 malignant tumors and 211 of 219 benign tumors. Table 1 lists the number of misdiagnosed cases of the neural network at a threshold of 0.35 for each test set. The sonographic-histological correlation by the proposed HNN system and the radiologists is shown in Table 2. The category of not benign is the number including malignant and indeterminate cases of the sonographic classification according to Stavros et al. This represents the total number of lesions requiring biopsy according to their sonographic classification. Thus, the diagnostic performance of the neural network and the radiologists for malignancy is shown in Table 3.

**DISCUSSION**

Safe, conservative management of solid, well-defined nodules is a goal of breast imagers. The diagnostic criteria, excellent techniques, equipment capabilities, comfort level of the examiners, and medicolegal demands all play a role. Sickles
experience supports the safety in follow-up for masses with probably benign mammographic features with the use of careful selection criteria. Stavros et al.\(^1\) and Skaane and Engedal\(^2\) report the NPV approaching 100% in the classification of solid breast masses with a high degree of confidence. In this study, 3 radiologists, evaluating sonograms, also have a high NPV compared with the proposed HNN system. The NPV of breast biopsy, that is, the proportion of cases deemed benign that were actually benign by histology, were as high as 96% and higher by the HNN and radiologists. In practice, this suggests that experienced radiologists using strict published criteria and HNN should be able to diagnose all of the cancers with fewer benign biopsies. Radiologists 2 and 3 have nearly identical diagnostic performances. They work in the same hospital and their training backgrounds are alike. For radiologist 1, although the sensitivity rate (97.22%) and NPV (99.32%) are high, the specificity (67.58%) and PPV (33.01%), are relatively low. This is because of the number of benign lesions categorized as indeterminate. There were 71 indeterminate lesions of 211 benign cases. I believe that most radiologists would prefer to perform a benign biopsy rather than miss a cancer. However, that resulted in too many biopsies being performed. The number of biopsies of benign lesions proposed by the HNN system was only 8 compared with 71, 31, and 33 by the radiologists. Therefore, many unnecessary benign biopsies with the resultant patient discomfort, ex-
pense, potential complications, cosmetic change, and anxiety can be avoided. It is desirable to maintain a high PPV for suspect findings on breast images. The PPV of breast biopsy, that is, the proportion of cases deemed malignant at diagnosis that were actually malignant by histology, would have been much improved from 33.01%, 49.18%, and 47.19%, respectively, for radiologists alone to 81.4% for CAD with a network. Therefore if 100 biopsies were recommended, 81 cases would be expected to yield malignancy. Tersegno\textsuperscript{22} and Bassett et al\textsuperscript{23} reported that the PPV of mammography is 10% to 35% in the United States. Previous investigators have attempted to increase the PPVs of breast biopsies to a proposed goal of 41% to 51% set by Hall et al\textsuperscript{24} and Ciatto et al.\textsuperscript{25}

In an attempt to predict more accurately if a lesion is benign or malignant, breast ultrasound has been found to be helpful. In 1997, Raza and Baum\textsuperscript{26} used Doppler ultrasound to evaluate 86 solid breast lesions. When gray-scale and power Doppler ultrasound findings were analyzed, they concluded that 100% of the sensitivity and NPV could be achieved while maintaining the 50% rate of PPV. CAD
techniques—HNN in particular—provides one possible solution to improving the PPV of breast biopsy. However, a matter of concern is the rate of false negatives for the radiologists and the proposed CAD system, in this situation, patients with breast cancer inaccurately diagnosed as benign who, therefore, might not be biopsied and consequently would escape early diagnosis. The false-positive rate (8 of 211 cases) for the proposed HNN system is less important clinically, once cancer is classified, it can be subsequently excluded by negative histology. The improvement in specificity for benign lesions potentially has even greater clinical importance.

In this study, the use of CAD significantly increased the diagnostic performance compared with that of the radiologists. The proposed CAD can easily be implemented on an existing commercial diagnostic ultrasound machine. For most currently available diagnostic ultrasound machines, all that would be required for the CAD system is a personal computer with the CAD software. The advantage of using texture analysis is that a physician (1) does not take time to set complicated parameters as in other methods using neural networks and (2) they can get a quick second opinion. The performance of the proposed CAD system has a relatively high diagnostic accuracy while maintaining both high sensitivity and specificity. It can be used in office-based breast clinics as a secondary reading for physicians.

In this study, a 4-image analysis of each tumor obtains better results than if they are used separately. Whatever group of images used for analysis, TA, LA, TB, or LB (Fig 1), each can get an accurate result. These results confirm results of our previous

<table>
<thead>
<tr>
<th>Table 1. Number of Missed Cases of the HNN Diagnostic System at a Threshold of 0.35 for Each Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Set</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Classification of Breast Nodules by Proposed HNN System and 3 Radiologists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sonographic Classification</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>Benign</td>
</tr>
<tr>
<td>Indeterminate</td>
</tr>
<tr>
<td>Malignant</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

NOTE: radiologist 1: 36 carcinomas were diagnosed as not benign in 35 cases (TP = 35), 1 carcinoma was misdiagnosed as benign (FN = 1), 219 benign tumors were diagnosed correctly in 148 cases (TN = 148), 71 benign tumors were categorized as not benign (malignant plus indeterminate) (FP = 71).

Radiologist 2: 36 carcinomas were diagnosed as not benign in 30 cases (TP = 30), 6 carcinomas were misdiagnosed as benign (FN = 6), 219 benign tumors were diagnosed correctly in 188 cases (TN = 188), 31 benign tumors were categorized as not benign (FP = 31).

Radiologist 3: 36 carcinomas were diagnosed as not benign in 30 cases (TP = 30), 6 carcinomas were misdiagnosed as benign (FN = 6), 219 benign tumors were diagnosed correctly in 188 cases (TN = 186), 33 benign tumors were categorized as not benign (FP = 33).

HNN: 36 carcinomas were diagnosed as probably malignant in 35 cases (TP = 35), 1 carcinoma was misdiagnosed as benign (FN = 1), 219 benign tumors were diagnosed correctly in 211 cases (TN = 211), 8 benign tumors were categorized as probably malignant (FP = 8).

*Histological finding.
Abbreviations: TP, true positive; FN, false negative; TN, true negative; FP, false positive.

<table>
<thead>
<tr>
<th>Table 3. Summary of Accuracy Among Radiologists and HNN System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiologist</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>Accuracy (%)</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>Specificity (%)</td>
</tr>
<tr>
<td>PPV (%)</td>
</tr>
<tr>
<td>NPV (%)</td>
</tr>
</tbody>
</table>

NOTE. Accuracy, (TP + TN)/(TP + TN + FP + FN); sensitivity, TP/(TP + FN); specificity, TN/(TN + FP); PPV, TP/(TP + FP); NPV, TN/(TN + FN).
Abbreviations: TP, true positive; TN, true negative; FP, false positive; FN, false negative.
study that benign and malignant tumors can be distinguished using interpixel correlation in digital ultrasonic images. However, to incorporate two orthogonal imaging planes and 4-image analysis, the result appears more promising.

Breast sonographic characteristics between benign and malignant nodules show a substantial overlap. It is impossible to distinguish all benign from all malignant nodules. Finding an objective method to identify a subgroup in which the certainty of malignancy is enough and avert benign biopsy when possible is the goal of this study. CAD techniques may provide a significant measure of confidence in the evaluation of a subgroup of lesions that are likely benign so that the number of benign biopsies can be reduced.

REFERENCES