Chung-Ming Chen, PhD
Yi-Hong Chou, MD
Ko-Chung Han, MS
Guo-Shian Hung, MS
Chui-Mei Tiu, MD
Hong-Jen Chiou, MD
See-Ying Chiou, MD

Breast Lesions on Sonograms: Computer-aided Diagnosis with Nearly Setting-Independent Features and Artificial Neural Networks[1]

PURPOSE: To develop a computer-aided diagnosis (CAD) algorithm with setting-independent features and artificial neural networks to differentiate benign from malignant breast lesions.

MATERIALS AND METHODS: Two sets of breast sonograms were evaluated. The first set contained 160 lesions and was stored directly on the magnetic optic disks from the ultrasonographic (US) system. Four different boundaries were delineated by four persons for each lesion in the first set. The second set comprised 111 lesions that were extracted from the hard-copy images. Seven morphologic features were used, five of which were newly developed. A multilayer feed-forward neural network was used as the classifier. Reliability, extendability, and robustness of the proposed CAD algorithm were evaluated. Results with the proposed algorithm were compared with those with two previous CAD algorithms. All performance comparisons were based on paired-samples *t* tests.

RESULTS: The area under the receiver operating characteristic curve (A_z) was 0.952 ± 0.014 for the first set, 0.982 ± 0.004 for the first set as the training set and the second set as the prediction set, 0.954 ± 0.016 for the second set as the training set and the first set as the prediction set, and 0.950 ± 0.005 for all 271 lesions. At the 5% significance level, the performance of the proposed CAD algorithm was shown to be extendible from one set of US images to the other set and robust for both small and large sample sizes. Moreover, the proposed CAD algorithm was shown to outperform the two previous CAD algorithms in terms of the A_z value.

CONCLUSION: The proposed CAD algorithm could effectively and reliably differentiate benign and malignant lesions. The proposed morphologic features were nearly setting independent and could tolerate reasonable variation in boundary delineation.

© RSNA, 2003

Breast cancer is one of the leading causes of death for women in many countries (1). For early detection of breast cancer, mammography is currently the most widely used screening modality, but it has a low negative predictive value. Many investigators have found that more than 60% of masses referred for breast biopsy on the basis of mammographic findings are actually benign (2,3).

Breast sonography was shown to be an effective adjunct to mammography in reducing the number of negative biopsy results (4–8). For example, with deliberately devised sonographic features, Stavros et al (7) were able to attain the overall sensitivity, specificity, accuracy, positive predictive value, and negative predictive value of 98.4%, 67.8%, 72.9%, 38%, and 99.5%, respectively. Potentially effective as it is, breast sonography remains controversial for screening because interpretation of the ultrasonographic (US) images is greatly influenced by the scanning techniques and the sonographic features of the suspected abnormality. Breast sonologists with different experiences might have different

interpretations of the sonograms. To minimize the effect of the operator-dependent nature inherent in US, many computerized approaches have been proposed to assist differentiation between benign and malignant breast lesions (9–15).

The general idea of computer-aided diagnosis (CAD) for breast sonography is to convert the visually extractable sonographic features into mathematic models and to characterize the lesions with the mathematic features based on the classification schemes. The mathematic features may be categorized into two classes, namely, the regional features and the morphologic features. The regional features characterize the image properties evolved from the intensity distribution (eg, echogenicity, echotexture), whereas the morphologic features describe the shape and contour of the lesion. As an example, with use of the mathematic features that quantify lesion margin, shape, homogeneity, and posterior acoustic attenuation pattern, Giger et al (12) achieved values for the area under the receiver operating characteristic (ROC) curve (A_z) of 0.94 and 0.87 for the entire database and the equivocal database on the basis of linear discriminant analysis (LDA).

Although promising performances have been reported, CAD for breast sonography is still impractical for routine use because previous mathematic features depend on either the setting of the US systems or the contour extraction process. It is easy to show that most regional features vary nonlinearly with the system setting. For instance, the co-occurrence matrix used by Garra et al (10) may fluctuate with such system parameters as the time-gain compensation, total gain, and focal depth. To avoid this problem, many previous CAD algorithms necessitate that all breast images be obtained with the same system parameter setting (13–15). This constraint is clinically undesirable. On the other hand, since the morphologic features (eg, the contour gradients [11]) are derived from the contour, they are more susceptible to the contour extraction process than are the regional features. Ideally, this problem may be solved by using automatic contour extraction schemes. However, automatic contour extraction on a US image is a difficult task in general, and no satisfactory approaches exist so far, to our knowledge.

The purpose of this study was to develop a CAD algorithm with setting-independent features and artificial neural networks to differentiate benign from malignant breast lesions. More specifically, this study was aimed to design a set

of morphologic features that were nearly independent of not only the system setting but also the contour extraction process.

MATERIALS AND METHODS

The proposed CAD algorithm was composed of three essential components, namely, feature extraction, feature selection, and classification. To relax the constraints on the system settings, the morphologic features rather than the regional features were adopted. The potential dependence of the morphologic features on the contour extraction process was minimized by capturing important topologic properties of the lesions, which may not vary drastically with the delineated contour. Feature selection was necessary to alleviate dimensionality (16). A set of essential morphologic features that yielded the curse of the best performance was selected on the basis of stepwise logistic regression (17). Classification was accomplished with a multilayer feed-forward neural network (MFNN) (18) on the basis of the essential morphologic features. The advantage of the MFNN is that arbitrarily complex convex separation surfaces can be approximated.

Study Subjects and Image Acquisition

Two sets of breast US images were used in this study that were selected randomly from the database of a medical center in Taiwan. The image data were collected during a period of 4 years. The institutional review boards agreed that the patient images and clinical information could be used for study without written consent if anonymity was maintained. This regulation was carefully followed in the present study.

The first set of sonograms was obtained from September 9, 1996, to June 6, 2000, in 160 female patients (age range, 16–85 years; mean age, 46 years). The sonograms depicted 160 breast lesions, including 42 cysts, 49 fibroadenomas, and 69 carcinomas, that were pathologically proven. They were stored directly (by using the system built-in function) on a US system (HDI 3000; Advanced Technological Laboratory, Bothell, Wash) equipped with a broadband 5–10-MHz linear electronically focused transducer and cine loop capability.

The second set of sonograms was obtained from January 1, 1997, to December 31, 1998, in 111 women (age range, 18–82 years; mean age, 42 years). The

sonograms depicted 111 breast lesions that were pathologically proven, including 40 fibroadenomas and 71 carcinomas. They were obtained with the same US system that was used for the first set. Unlike the first set of US images, the second set comprised hard-copy images. The lesions were extracted from these sonograms by first digitizing these images with film scanners (HP6300C; Hewlett-Packard, Palo Alto, Calif).

No constraint was imposed on the system settings during acquisition of these images. The sonologists were free to adjust the system settings to obtain the best views. In both sets, the lesion boundaries were delineated manually. The first set of lesions served as the primary basis for performance evaluation and comparison of the complete morphologic and regional information preserved in the directly stored US images. To take into account the potential variation of delineation among different persons, the first set of lesions was delineated by four graduate students (K.C.H. and others), and each student was supervised by one of four attending physicians (Y.H.C., C.M.T., H.J.C., S.Y.C.) with 22, 7, 7, and 3 years of experiences in breast sonography, respectively. For each lesion, size variation was defined as the ratio of the SD to the mean of the sizes of the four delineated lesion boundaries. The mean ± SD of variations of lesion size for all lesions in the first set was 9.1% ± 7.7.

The second set of lesions served as a larger number of samples with which to evaluate the extendability and robustness of the proposed CAD algorithm. The second set of lesions was delineated by only one graduate student (G.S.H.), under the supervision of an attending physician (Y.H.C.) with 22 years of experience in breast sonography. Two of the five graduate students (K.C.H., G.S.H.) were involved in development of the CAD algorithm.

Feature Extraction

Seven morphologic features were extracted from each lesion to account for such sonographic features as shape, contour, and size. Five of these morphologic features were newly developed, including the number of substantial protuberances and depressions (NSPD), lobulation index (LI), elliptic-normalized circumference (ENC), elliptic-normalized skeleton (ENS), and long axis to short axis (L:S) ratio. The other two features were clinically useful indicators (19): depth-to-width (D:W) ratio and size of the lesion.

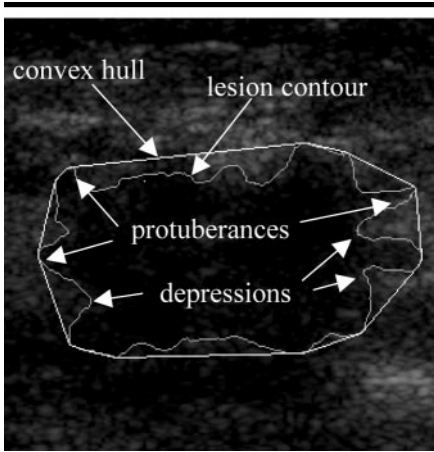*NSPD.*—The spiculation (7) and irregu-

Figure 1. Sonogram shows the inner gray contour as the lesion border and the outer white polygon as the convex hull of the lesion. Protuberances and depressions in the malignant breast lesion are indicated.
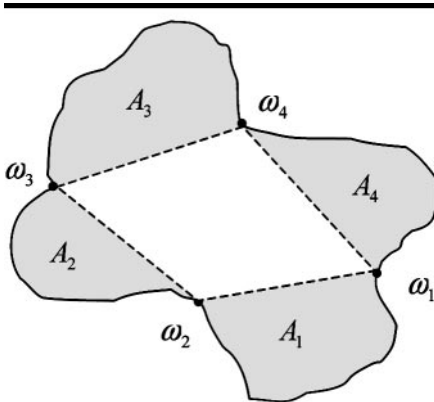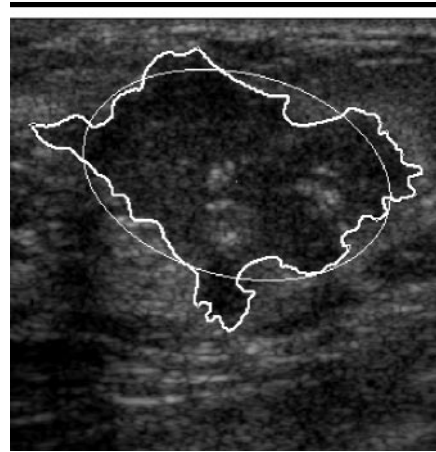

Figure 3. Sonogram shows the equivalent ellipse (thin line) of a malignant breast lesion, the boundary of which is marked by the thick line.

Figure 2. Schematic shows lobes in a lesion, where $\{\omega_1, \omega_2, \omega_3, \omega_4\}$ are four representative concave points in Ω and $\{A_1, A_2, A_3, A_4\}$ are the sizes of the lobes.

lar shape and contour (8) of a lesion are two important sonographic features that characterize a malignant breast lesion. The NSPD is an effective descriptor in a lesion to quantify these two sonographic features. With a geographic analogy, a protuberance and a depression are like a peninsula and a bay, respectively. As an example, Figure 1 shows typical protuberances and depressions in a malignant breast lesion. Since protuberances and depressions may easily result from a wobbly delineation process, as described in the Appendix, only the substantial protuberances and depressions defined by the representative convex and concave points, respectively, were used to characterize a breast lesion.

As derived in the Appendix, given a threshold θ_p, let $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda p\}$ and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_d\}$ be the set of repre-

sentative convex and concave points of a lesion boundary, where p and d are the numbers of points in each set. The NSPD, denoted by $NSPD(\theta_p)$, is defined as $p + d$, where $\theta_p \in \{20°, 30°, 40°, 50°, 60°\}$. Ideally, a malignant breast lesion has a larger NSPD.

LI.—LI was devised to characterize the size distribution of the lobes in a lesion. As illustrated in Figure 2, a lobe is defined as the gray region enclosed by the lesion contour and the dashed line connected by two adjacent representative concave points. The size of the lobe is the area of the gray region. Suppose a breast lesion has N_1 lobes and the size of the ith lobe is A_i, $i = 1$, to N_1. Let A_{max} and A_{min} denote the sizes of the largest and the smallest lobes. The LI is then defined as

$$LI = \frac{A_{max} - A_{min}}{\dfrac{1}{N_1} \displaystyle\sum_{i=1}^{N_1} A_i}.$$

The LI can correctly characterize a benign lesion with multiple large lobes of similar sizes. This type of benign lesion may be misclassified as a malignant lesion with the NSPD.

ENC.—Anfractuosity is a common morphologic characteristic of malignant lesion boundaries that provides at least two visually appreciable geometric features. One feature is the multiple protuberances and depressions that may be well described with the NSPD. The other feature is the lengthened circumference due to the circuitous boundaries that define the protuberances and depressions. Since the boundary of a smaller lesion would appear to be more winding than

that of a larger lesion with the same circumference, the circumference itself is not a good descriptor with which to characterize the anfractuosity of the lesion boundary. Alternatively, a more reasonable approach is quantification of the anfractuosity with the percentage of circumference increment relative to a lesion-dependent baseline. An ideal baseline would be a smooth curve such that the lesion boundary would look like twining around the curve.

To quantify the anfractuosity of a lesion contour, the circumference ratio of the lesion and its equivalent ellipse is proposed, which is termed the ENC. The equivalent ellipse of a lesion (20,21) is an ellipse with the same area and center of mass as those of the lesion when the interiors of the lesion and its equivalent ellipse are both set to the same constant gray level. For instance, Figure 3 shows the equivalent ellipse and boundary of a malignant lesion. Perceptually, one can see that the equivalent ellipse roughly captures the shape of the lesion, and the lesion boundary meanders around the equivalent ellipse.

ENS.—A skeleton is an effective representation of a region (22) that is used frequently in such areas as computer vision and pattern recognition. Let R denote a region and B_R the set of boundary points of the region R. The skeleton of a region, R, is a set of points X that satisfy for each skeleton point $x \in X$, where x is within R and there exist at least two boundary points, p_i and p_j, in B_R such that $d(x, p_i) = d(x, p_j) = \min \{d(x, p_k)|p_k \in B_R\}$, where $d(\cdot)$ is any preferred distance metric (eg, Euclidean, city block). With the malignant breast lesion shown in Figure 3 as an example, the skeleton of the lesion is given in Figure 4.

The skeleton is sensitive to the anfractuous property of the lesion boundary. The more protuberances and depressions contained in the lesion boundary, the more complex the skeleton is. Therefore, it seems to be reasonable to quantify the shape complexity by the number of points in the skeleton. Nevertheless, the number of skeleton points is also a function of the lesion size. Just as for the ENC, to eliminate the size effect, it is suggested that the number of skeleton points be normalized by the circumference of the equivalent ellipse of the lesion, which gives the ENS.

In addition to these four descriptors— NSPD, LI, ENC, and ENS—which capture the contour and shape characteristics, three more mathematic features are con-

sidered to incorporate two clinically useful indicators. The first feature is the D:W ratio of the lesion (11,12,19). The depth and the width of a lesion are the horizontal and vertical edge lengths, respectively, of the minimal circumscribed rectangle of the lesion. The larger the D:W ratio, the more likely the lesion is malignant. Since the D:W ratio may vary with the scanning angle and the compressing pressure, however, we suggest use of another quantity to describe the shape of the lesion, namely, the L:S ratio. The L:S ratio is the length ratio of the major (long) axis to the minor (short) axis of the equivalent ellipse of the lesion. Clearly, the L:S ratio is independent of the scanning angle but may be affected by the compressing pressure. The last feature is the size of the lesion (ie, the area within the lesion boundary). Clinically, the larger the breast lesion, the more likely the lesion is malignant.

Feature Selection and Classification

Feature selection is usually used to select a set of features that potentially yield the best performance with the given classifier. These selected features are referred to as the substantial features. The classifier is then trained with the substantial features to determine the mathematic model that describes the relation of these features.

The substantial features were selected for each training data set on the basis of the forward sequential search approach (23) with the logistic discrimination function (17). To minimize the estimation bias (16), the classification accuracy $\eta(Y)$ for a feature set Y was evaluated by means of the leave-one-out cross-validation strategy for each training data set. More specifically, set the counter $n_c = 0$, for every l_i in the training data set Φ with m data, then construct the logistic discrimination function with the training data $\Phi_i = \Phi - \{l_i\}$ with the feature set Y. If l_i is predicted correctly with the derived logistic discrimination function, then increase n_c by 1. After all l_i values have been evaluated, then compute $\eta(Y) = n_C/m$.

Feature selection was performed in two stages for each training data set. In the first stage, the best NSPD value was selected from five candidate NSPD values that corresponded to five θ_p values (ie, $\theta_p \in \{20°, 30°, 40°, 50°, 60°\}$). Then, in the second stage, the selected NSPD value along with the other six features were used to select the essential features that yielded the best classification accuracy for the underlying training data set.
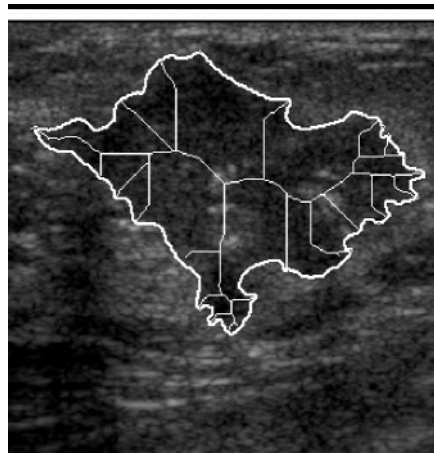

Figure 4. Sonogram shows the skeleton of a malignant breast lesion. The boundary of the lesion is marked by the thick line, and the skeleton is indicated by the thin line segments.

The classifier used in the present study is an MFNN. Once the essential features were selected by means of the logistic discrimination function for a set of training data Φ, the training data were used to train the MFNN to divide the training data into benign and malignant categories. As depicted in Figure 5, the MFNN used in this study was a two-layer feed-forward neural network with one hidden layer. The number of inputs for the MFNN was set to be the same as the number of essential features, and the number of neurons in the output layer was set to 1 for the underlying two-class classification. Some suggestions were made previously (eg, the Kolmogorov theorem [24]) to determine the number of neurons in the hidden layer, but none of them led to satisfactory performance. Instead, the number of neurons in the hidden layer was determined through exhaustive experiments to be two to 10 neurons. As a result, the number of neurons in the hidden layer was set to two because results with two neurons gave the best performance for almost all cases evaluated.

The training algorithm used to train the MFNN was the widely used error back-propagation training algorithm (18). The training data with the essential features were fed into the MFNN in a cyclic manner. For each datum, the estimated output o was computed on the basis of the synaptic weights determined in the previous iterations and the sigmoidal activation function φ. The discrepancy between the desired output and the estimated output was back propagated to modify the synaptic weights until the discrepancy was within the acceptable
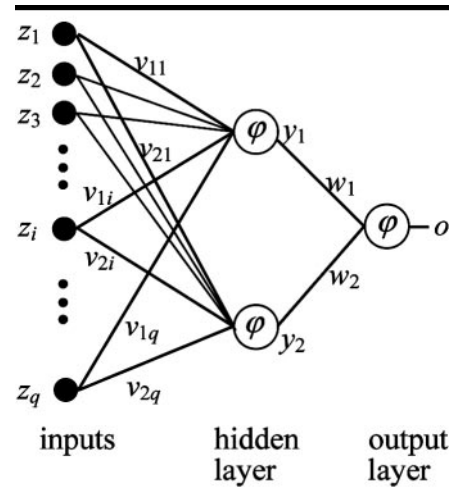

Figure 5. Two-layer feed-forward neural network used for classification. z_i = the ith feature, v_{ij} = the weight of the synapse that connects the jth input to the ith neuron in the hidden layer, y_i = the ith neuron in the hidden layer, w_i = the weight of the synapse that connects the ith neuron in the hidden layer to the output, φ = the sigmoidal activation function, and o = the output.

range. Since the final output o was a number between 0 and 1, a threshold T_{NN}, where NN is neural network, was required to assign the datum to the benign or the malignant category. In the present study, T_{NN} was determined on the basis of the value that resulted in a dichotomization with the best classification accuracy for the training data, or T_{NN} varied from 0 to 1 to generate the ROC curve.

Comparative Performance Analysis

Five experiments were conducted in the present study for performance analysis. Two performance measures were reported for each analysis. One measure was the A_z value, which was calculated by using commercially available statistical software (SPSS for Windows, version 10; SPSS, Chicago, Ill). The other measure was the best classification accuracy (TP + TN)/(TP + TN + FP + FN) along with the associated sensitivity (TP/[TP + FN]), specificity (TN/[TN + FP]), positive predictive value (TP/[TP + FP]), and negative predictive value (TN/[TN + FN]), where TP is the number of true-positive findings (ie, a malignant lesion is considered to be malignant); TN, true-negative; FP, false-positive; and FN, false-negative.

For performance comparison, the A_z values were used because the best classification accuracy is not necessarily the preferred criterion for classification. Some-

Radiology

times, one would rather have a higher sensitivity or specificity than have the best accuracy. Except for the third experiment, all performance measures were derived on the basis of the leave-one-out cross-validation strategy.

To evaluate the reliability of the proposed morphologic features, in the first experiment, denoted as C160, the proposed CAD algorithm was evaluated with the four sets of boundaries drawn independently by four people for each of the 160 breast lesions in the first set of US images. To justify the necessity for feature selection, the second experiment, denoted as C160A, repeated the first experiment but without feature selection (ie, all seven features were used by the MFNN classifier). A paired-samples t test was used to test if incorporation of feature selection would yield a better performance (ie, if C160 was significantly better than C160A), with the significance level set at $\alpha = .05$.

Results of the third experiment validated the extendibility of the proposed CAD algorithm. Since the two sets of US images used in the present study originated from two archiving media, degradation for the boundary definition of the lesions that was caused by the acquisition procedure or the archiving medium was potentially different for each set. Therefore, these two sets of US images might be considered as samples from two different sample spaces. In this experiment, we attempted to investigate how well the classifier derived on the basis of images from one sample space could be extended to those from the other sample space. Two implementations were performed. One implementation was performed with the first set of US images as the training set and the second set as the prediction set, which was denoted as C271f. In reverse, the other implementation was performed with the second set of US images as the

training set and the first set as the prediction set, which was denoted as C271r. The training set was trained on the basis of the leave-one-out cross-validation strategy. Recall that only the first set of US images had four sets of boundaries. Paired-samples t tests were used to test if C271f and C271r had the same or better performance than did C160 and if C271f and C271r had the same performance with the significance level set at $\alpha = .05$.

The fourth experiment, denoted as C271LC, was performed to investigate the robustness of the proposed CAD algorithm. That is, we attempted to evaluate how well the performance achieved in the first experiment with a smaller sample size could be reproduced with a larger number of samples from heterogeneous sample spaces. All 271 breast lesions were involved in the fourth experiment, and the leave-one-out cross-validation strategy was used. To validate the robustness of the proposed CAD algorithm, paired-samples t test was used to determine if C271LC had the same performance as C160 with the significance level set at $\alpha = .05$.

For comparative study, in the fifth experiment, the proposed CAD algorithm was compared with two previous CAD algorithms with the first set of breast lesions. The first algorithm was proposed by Giger et al (12), which was denoted as LDA_{Giger}. The Giger algorithm included four mathematic features, namely, normalized radial gradient, D:W ratio, coarseness, and the mean gray-level difference between the region of interest within the lesion and that posterior to the lesion, denoted by $\mu_1 - \mu_2$. The classification scheme was the LDA. Since the MFNN is usually superior to the LDA, for a fair comparison with our approach, as a modified implementation, which was denoted as $MFNN_{Giger}$, the MFNN was used to replace the LDA. The number of neu-

rons in the hidden layer was also determined by searching in the range of two to 10 neurons. As a result, 10 neurons were used in the hidden layer for the modified Giger algorithm.

The second CAD algorithm to be compared with the proposed CAD algorithm was proposed by Chen et al (13), which was denoted as $MFNN_{Chen}$. The Chen algorithm was based solely on a texture feature (ie, normalized autocorrelation coefficients obtained from a rectangular region of interest that enclosed the lesion). The size of the region of interest was 1–2 mm extended from the lesion margin in all directions. The feature vector contained 5×5 autocorrelation coefficients. The classifier used by Chen et al (13) was an MFNN—with 25 inputs, 10 hidden nodes, and one output node—that was also trained with the error back-propagation training algorithm.

To test if C160 had significantly better performance than that of LDA_{Giger}, $MFNN_{Giger}$, and $MFNN_{Chen}$, paired-samples t tests were applied with the significance level set at $\alpha = .05$. Moreover, a paired-samples t test was used to determine the relative performance among these three implementations.

In addition to these five experiments, the performance of each individual feature, including the proposed seven features and the Giger features, was evaluated by means of logistic discrimination analysis (17) based on the leave-one-out cross-validation strategy. Paired-samples t tests were used to compare the performances of every pair of individual features. It should be emphasized that all algorithms were evaluated for statistical robustness with four collections of lesion boundaries. As a summary, Table 1 lists the notations for the implementations performed in this study.

TABLE 2
Results of First Experiment with First Set of Breast Lesions

Experiment	ROC Curve		Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
	A_z	SE*					
C160-1	0.938	0.023	91.9	95.6	87.0	90.6	93.8
C160-2	0.941	0.021	90.6	97.8	81.2	87.3	96.6
C160-3	0.962	0.014	94.4	97.8	89.9	92.7	96.9
C160-4	0.966	0.015	94.4	95.6	92.8	94.6	94.1
μ_{C160}	0.952	0.018	92.8	96.7	87.7	91.3	95.4
σ_{C160}	0.014	0.004	1.9	1.3	4.9	3.1	1.6

Note.—PPV = positive predictive value, NPV = negative predictive value.
  * Standard error of the ROC curve.

TABLE 3
Means and SDs of Performances Achieved in C160A, C271f, C271r, and C271LC

Mean and SD	ROC curve		Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
	A_z	SE*					
μ_{C160A}	0.940	0.022	91.1	94.5	86.6	90.3	92.4
σ_{C160A}	0.019	0.005	1.4	2.7	3.2	2.0	3.4
μ_{C271f}	0.982	0.010	94.8	90.6	97.2	94.9	94.9
σ_{C271f}	0.004	0.001	0.9	1.3	1.6	2.8	0.6
μ_{C271r}	0.954	0.016	91.4	95.1	86.6	90.4	93.2
σ_{C271r}	0.016	0.004	1.7	2.9	3.6	2.3	3.7
μ_{C271LC}	0.959	0.011	92.8	96.7	87.7	91.3	95.4
σ_{C271LC}	0.005	0.001	1.9	1.3	4.9	3.1	1.6

Note.—PPV = positive predictive value, NPV = negative predictive value.
  * Standard error of ROC curve.

RESULTS

Results of the first experiment for the reliability evaluation of the proposed CAD algorithm are reported in Table 2. For each implementation, the A_z value and standard error of the ROC curve were listed, as well as the performance data at the best classification accuracy, including accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. Table 2 shows that the proposed CAD algorithm has achieved reasonably high performances for all four sets of lesion boundaries in terms of the A_z value and the classification accuracy, the means of which, denoted as μ_{C160}, were as high as 0.952% and 92.8%, respectively. Moreover, differences in performance induced by the potential variations in boundary definitions of the four sets of lesion boundaries were not significant. The SDs of the A_z values and the best classification accuracies, denoted as σ_{C160}, were 0.014% and 1.9%, respectively. On the basis of the experimental results, the proposed CAD algorithm potentially may be capable of tolerating the variations in boundary definition due to manual delineation by different persons.

For the second experiment, the mean and SD of the A_z value, which are labeled as μ_{C160A} and σ_{C160A}, respectively, and the performance data at the best classification accuracy are summarized in Table 3. The test hypothesis and result of the paired-samples t test, labeled as T_1, to compare C160 with C160A are given in Table 4. Since $P = .020 < \alpha = .05$, the null hypothesis should be rejected (ie, C160 was superior to C160A at the 5% significance level), which justified the advantage of adopting feature selection in the proposed CAD algorithm. The means and SDs of the A_z values and the performance data at the best classification accuracy for C271f, C271r, and C271LC are provided in Table 4. These three implementations attained reasonably high performances, which were greater than 0.954% and 91.4% for the mean A_z and the mean best accuracy, respectively.

For pairwise performance comparisons between C160 versus C271f, C160-C271r, C271f-C271r, or C160-C271LC, the test hypotheses and results of four corresponding paired-samples t tests—labeled as T_2, T_3, T_4, and T_5, respectively—are listed in Table 5. With $P = .019 < \alpha = .05$, results of the paired-samples t test T_2 suggested that the null hypothesis be rejected, which implied that the performance of C271f was significantly higher than C160 at the 5% significance level. Results of the paired-samples t tests T_3–T_5 suggest that these three null hypotheses be accepted, since all three P values were greater than $\alpha = .05$. In other words, at the 5% significance level, the performance of C160 was the same as that of C271r and C271LC, and there was no significant difference between the performances of C271f and C271r. From the test results of T_2–T_4, it might be concluded that in comparison with use of only the first set of US images, the proposed classifier derived on the basis of the images from one sample space might be generalized to the images from another sample space without performance degradation at the 5% significance level. Moreover, the test result of T_5 validated that the proposed CAD algorithm was robust in the sense that its performance would be the same for both small and large sample sizes at the 5% significance level.

The significant features selected for classification varied with the collection of lesion boundaries, and, more specifically, they varied with the training set. Table 5 lists the most frequently selected features for each collection of lesion

TABLE 4
Results of Paired-Sample *t* Tests to Compare Means of Different Implementations on the Basis of the Proposed CAD Algorithms

Comparison between Two Algorithms*	Paired Differences			*t*	*P* Value
	Mean	95% CI[†]			
		Lower	Upper		
T_1: $\begin{cases} H_0: \mu_{C160} - \mu_{C160A} \leq 0 \\ H_A: \mu_{C160} - \mu_{C160A} > 0 \end{cases}$	1.18×10^{-2}	3.86×10^{-3}	NA	3.506	0.020[‡]
T_2: $\begin{cases} H_0: \mu_{C160} - \mu_{C271f} \geq 0 \\ H_A: \mu_{C160} - \mu_{C271f} < 0 \end{cases}$	-2.98×10^{-2}	NA	-1.01×10^{-2}	-3.567	0.019[‡]
T_3: $\begin{cases} H_0: \mu_{C160} - \mu_{C271r} = 0 \\ H_A: \mu_{C160} - \mu_{C271r} \neq 0 \end{cases}$	-1.75×10^{-3}	-6.32×10^{-3}	2.82×10^{-3}	-1.219	0.31
T_4: $\begin{cases} H_0: \mu_{271f} - \mu_{C271r} = 0 \\ H_A: \mu_{271f} - \mu_{C271r} \neq 0 \end{cases}$	2.80×10^{-2}	-1.43×10^{-3}	5.74×10^{-2}	3.028	0.056
T_5: $\begin{cases} H_0: \mu_{C160} - \mu_{C271LC} = 0 \\ H_A: \mu_{C160} - \mu_{C271LC} \neq 0 \end{cases}$	-7.25×10^{-3}	-2.15×10^{-3}	7.00×10^{-3}	-1.619	0.204

* H_0 = null hypothesis, H_A = alternative hypothesis.
[†] NA = not applicable.
[‡] Significant at the level of α = .05.

TABLE 5
Most Frequently Selected Features for Each Collection of Lesion Boundaries

US Image Set	B1*	B2*	B3*	B4*
1[†]	NSPD and 40°, ENC	NSPD and (40°), ENS	NSPD and 40°, ENS	NSPD and 50°
1 + 2[‡§]	NSPD and 40°, D:W ratio	NSPD and 40°, ENS	NSPD and 40°	NSPD and 40°

Note.—Given *N* lesion boundaries, the frequency of a feature was the number of times that the feature was selected by the *N* possible training sets in the leave-one-out cross-validation process.
* *Bn*: the nth set of lesion boundaries drawn from the first set of US images, *n* = 1–4.
[†] First set of US images.
[‡] Four combinations of the first and second sets of US images, each of which contained one set of lesion boundaries from the first set and the lesion boundaries from the second set.
[§] Features selected for only the second set of US images, with one set of delineated lesion boundaries, were NSPD and 40°, L:S ratio, and D:W ratio.

boundaries. Recall that feature selection was performed on the basis of the leave-one-out cross-validation strategy. Given *N* lesion boundaries in a collection, the frequency of a feature was the number of times that the feature was selected for the *N* possible training sets in the leave-one-out cross-validation process. From Table 5, it is clear that NSPD was the most important feature, which might be either NSPD and 40° or NSPD and 50°. Furthermore, ENS, ENC, L:S ratio, or D:W ratio might be combined with NSPD to achieve a better performance.

With the first set of US images, the mean and SD of the A_z value and the performance data at the best classification accuracy for LDA_{Giger}, $MFNN_{Giger}$, and $MFNN_{Chen}$ are provided in Table 6. The performances attained with LDA_{Giger} and $MFNN_{Chen}$ were substantially lower than those reported in references 12 and 13. For pairwise performance compari-

sons among C160, LDA_{Giger}, $MFNN_{Giger}$, and $MFNN_{Chen}$, the test hypotheses and results of the paired-samples *t* tests are given in Table 7. With *P* values less than α = 0.05, all null hypotheses should be rejected. It might be concluded that the relative performances of these four algorithms were $MFNN_{Chen} < LDA_{Giger} < MFNN_{Giger} < C160$, where *A* < *B* denotes that *A* is worse than *B* at the 5% significance level.

To evaluate the performance of each individual feature, Figures 6 and 7 show the mean A_z and the mean best accuracy for each proposed feature with the first set of US images and all 271 US images, respectively. Results of paired-samples *t* tests suggested that NSPD, LI, ENS, and ENC are better than lesion size, L:S ratio, and D:W ratio at the 5% significance level in both Figures 6 and 7. For example, consider four pairwise comparisons, including NSPD lesion size, LI lesion size,

ENC lesion size, and ENS lesion size. The *P* values for these four pairs were .01, .043, .025, and .03, respectively, when the first set of US images were used, and they were .005, .02, .017, and .005, respectively, when all 271 US images were used. The null hypothesis for each pair *A* versus *B* was $\mu_A - \mu_B \leq 0$, where μ_A and μ_B stand for the mean A_z values attained with features *A* and *B*, respectively. Since all these *P* values were less than α = 0.05, NSPD, LI, ENC, and ENS were better than lesion size at the 5% significance level. Furthermore, results of paired-samples *t* tests suggest that the performances of NSPD, LI, and ENC remained the same for different sample sizes at the 5% significance level. That is, NSPD, LI, and ENC were robust. The *P* values of these three tests were .204, .405, and .587, respectively. Notably, the mean A_z values and the mean best accuracies of NSPD in both Figures 6 and 7 were greater than 0.94 and 0.91, respectively. In particular, use of NSPD or ENS alone could outperform LDA_{Giger}, $MFNN_{Giger}$, and $MFNN_{Chen}$ at the 5% significance level.

Figure 8 demonstrates the mean A_z and mean best accuracy for each Giger feature with the first set of US images. On the basis of results of paired-samples *t* tests, it could be shown that the proposed NSPD, ENC, and ENS were significantly better than all the Giger features, with *P* values less than .001 when the first set of US images were used. The proposed LI and lesion size had the same performance as normalized radial gradient at the 5% significance level.

TABLE 6
Means and SDs of Performances Achieved in LDA$_{Giger}$, MFNN$_{Giger}$, and MFNN$_{Chen}$

Data	ROC curve		Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
	A_z	SE*					
$\mu_{LDA_{Giger}}$	0.844	0.033	81.6	87.4	73.9	82.2	82.6
$\sigma_{LDA_{Giger}}$	0.023	0.003	2.4	7.1	12.2	5.9	6.0
$\mu_{MFNN_{Giger}}$	0.896	0.027	84.8	83.2	87.0	89.6	80.1
$\sigma_{MFNN_{Giger}}$	0.014	0.003	1.8	5.6	4.9	3.1	4.7
$\mu_{MFNN_{Chen}}$	0.731	0.040	71.7	72.1	71.4	66.1	77.5
$\sigma_{MFNN_{Chen}}$	0.033	0.002	2.4	8.2	7.7	4.6	4.3

Note.—PPV = positive predictive value, NPV = negative predictive value.
  * Standard error of the ROC curve.

TABLE 7
Paired-Sample *t* Tests to Compare Means of C160, LDA$_{Giger}$, MFNN$_{Giger}$, and MFNN$_{Chen}$

Comparison between Two Algorithms*	Paired Differences				*t*	*P* Value‡
	Mean	95% CI				
		Lower	Upper†			
T_6: $\begin{cases} H_0: \mu_{C160} - \mu_{LDA_{Giger}} \leq 0 \\ H_A: \mu_{C160} - \mu_{LDA_{Giger}} > 0 \end{cases}$	1.08×10^{-1}	7.60×10^{-2}	NA		7.944	.002
T_7: $\begin{cases} H_0: \mu_{C160} - \mu_{MFNN_{Giger}} \leq 0 \\ H_A: \mu_{C160} - \mu_{MFNN_{Giger}} > 0 \end{cases}$	5.60×10^{-2}	3.81×10^{-2}	NA		7.343	.003
T_8: $\begin{cases} H_0: \mu_{C160} - \mu_{MFNN_{Chen}} \leq 0 \\ H_A: \mu_{C160} - \mu_{MFNN_{Chen}} > 0 \end{cases}$	2.21×10^{-1}	1.69×10^{-1}	NA		10.147	.001
T_9: $\begin{cases} H_0: \mu_{MFNN_{Giger}} - \mu_{LDA_{Giger}} \leq 0 \\ H_A: \mu_{MFNN_{Giger}} - \mu_{LDA_{Giger}} > 0 \end{cases}$	5.20×10^{-2}	2.86×10^{-2}	NA		5.226	.007
T_{10}: $\begin{cases} H_0: \mu_{MFNN_{Giger}} - \mu_{MFNN_{Chen}} \leq 0 \\ H_A: \mu_{MFNN_{Giger}} - \mu_{MFNN_{Chen}} > 0 \end{cases}$	1.65×10^{-1}	1.26×10^{-1}	NA		10.116	.001
T_{11}: $\begin{cases} H_0: \mu_{LDA_{Giger}} - \mu_{MFNN_{Chen}} \leq 0 \\ H_A: \mu_{LDA_{Giger}} - \mu_{MFNN_{Chen}} > 0 \end{cases}$	1.13×10^{-1}	6.27×10^{-2}	NA		5.319	.006

* H_0 = null hypothesis, H_A = alternative hypothesis.
  † NA = not applicable.
  ‡ Differences in all comparisons significant at the level of $\alpha = .05$.

DISCUSSION

The performance of radiologists in the interpretation of mammograms and breast sonograms is said to be less than ideal (2,7). Although performance can often be improved by having two radiologists review mammograms or two sonologists review US images, this strategy is not easily available. Therefore, varieties of CAD efforts have been attempted in the imaging evaluation of breast diseases. Artificial neural networks have been applied in mammography and proved to be of potential help in the mammographic decision-making task of distinguishing between benign and malignant lesions (25). To our knowledge, however, there is only a limited number of articles that deal with CAD in breast sonography (9–15), and the previously proposed CAD algorithms are not yet mature for clinical utility.

On the basis of the high performance mean and low performance variation achieved with four sets of lesion boundaries in the first, third, and fourth experiments in the present study, the proposed CAD algorithm was shown to be an effective and robust approach to differentiation of benign from malignant breast lesions. Results of the third and fourth experiments further validated the extendibility and robustness of the proposed CAD algorithm. At the 5% significance level, the promising results obtained in these two experiments suggest that with the proposed CAD algorithm, the classifier trained by the images directly captured and stored in the electronic storage media may be applied to the hard-copy images and vice versa. Moreover, the proposed CAD algorithm was robust in the sense that performance remained the same for both small and large sample sizes.

The high performance of the proposed CAD algorithm resulted mainly from the effective and reliable morphologic features and incorporation of feature selection. Results of the evaluation of each individual feature showed that NSPD, ENC, and ENS were better than all the Giger features and even only one of the first two would outperform LDA$_{Giger}$, MFNN$_{Giger}$, and MFNN$_{Chen}$ at the 5% significance level. On the other hand, NSPD, LI, and ENC were shown to give reliable performance for different sample sizes. They were intrinsically reliable because a small local variation in contour delineation would not lead to a dramatic change in feature values. For example, the NSPD was used to count NSPDs in a lesion boundary. Reasonable variation in contour delineation might alter the shape of the lesion boundary but would not cause a big change in the NSPD. For LI and ENC, any reasonable variation in local delineation would be diluted by their own normalization factors, which were the mean area of the lobes for LI and
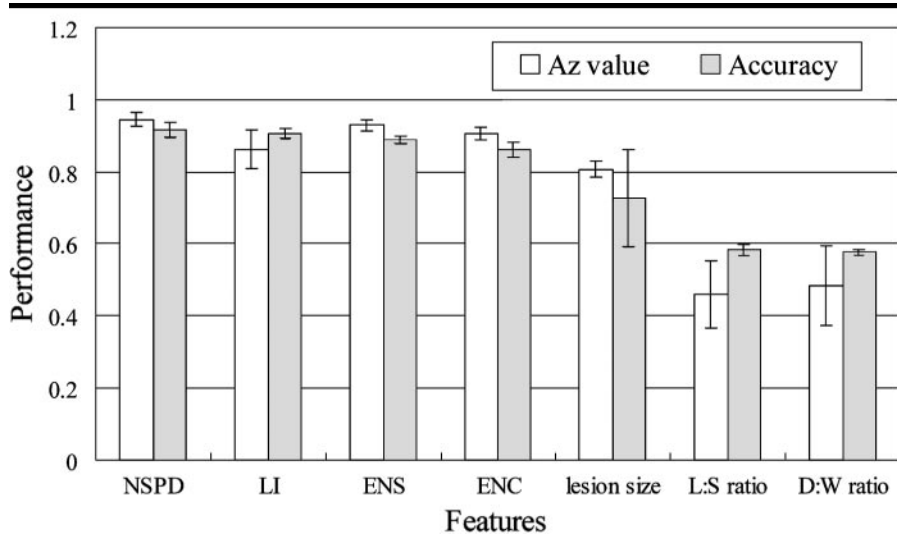
Figure 6. In bar graph, each pair of bars illustrates the mean A_z value and mean best classification accuracy achieved with a proposed morphologic feature with the first set of US images. Error bars indicate 1 SD. The first four features (ie, NSPD, LI, ENS, and ENC) are better than the other three at the 5% significance level.
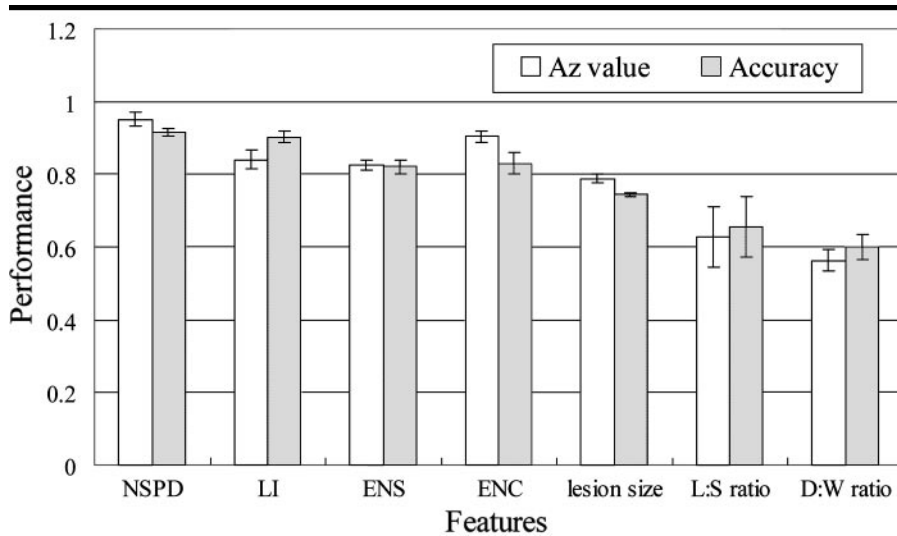

Figure 7. In bar graph, each pair of bars depicts the mean A_z value and mean best classification accuracy achieved with a proposed morphologic feature when applied to four collections of lesion boundaries. Each collection comprised one of the four sets of lesion boundaries in the first set of US images and the lesion boundaries in the second set of US images. Error bars indicate 1 SD. At the 5% significance level, the first four features (ie, NSPD, LI, ENS, and ENC) are better than the other three. Moreover, the performances of NSPD, LI, and ENC are the same with the first set of US images and with all 271 images.

the circumference of the equivalent ellipse for ENC. Since these normalization factors were usually on the order of 100 or 1,000, the potential value changes in LI and ENC would be small relative to the dynamic ranges of these four features.

On the basis of the effective features, feature selection was a necessary and beneficial step to further integrate the differential power of each individual feature, while accounting for the problem of "curse of dimensionality." The curse of dimensionality suggests that the sampling density of the training data is too low to promise a meaningful estimation of a high-dimensional classification function with all seven features with the available finite number of training data (16). As verified in the second experiment, performance with feature selection

(ie, C160) was superior to that without feature selection (C160A). It should be emphasized that feature selection is basically a learning process, and the best features vary as the training data change. This finding means that for a practical CAD system, feature selection should be performed frequently to allow learning from the changing training data sets.

The proposed CAD algorithm was shown to be better than the algorithms of Giger et al (12) and Chen et al (13). Setting dependence is one of the major reasons that the previous CAD algorithms are impractical for clinical use in the differentiation of benign from malignant breast lesions. This problem is particularly serious for those algorithms based on the regional features. For example, the Chen algorithm (13) was able to attain A_z values and classification accuracy as high as 0.956 and 95%, respectively, when the system setting was basically fixed. Results of the fifth experiment, however, showed that the Chen algorithm performed poorly with the first set of US images, which were acquired without any constraint imposed on the system setting. Any nonlinear change in the system setting may cause a nonnegligible variation in the normalized autocorrelation coefficients used with the Chen algorithm, even for the same lesion.

Similarly, the Giger algorithm (12) also had the setting-dependence problem because of the two regional features involved (ie, coarseness and mean gray-level difference between the region of interest within the lesion and that posterior to the lesion [$\mu_1 - \mu_2$]). Worse than the normalized autocorrelation coefficients, these two regional features may give different values, even with a linear change in the system setting. On the other hand, the morphologic feature of normalized radial gradient was sensitive to the local delineation as a result of the gradient type of information. That is, a small zigzag in the contour might result in a drastic change in the gradient. The experimental result showed that the performance of normalized radial gradient was substantially worse than that of NSPD, ENS, and ENC. Figure 8 reveals that none of these four features could provide sufficient differential power by itself.

The setting dependence of the regional features and the high sensitivity to the local delineation might account for the discrepancy between the high performance reported by Giger et al (12) and the low performance achieved in our fifth experiment. Although the performance improved with MFNN as the clas-

sifier, the best performance of the Giger mathematic features was still inferior to that of the proposed CAD algorithm. Moreover, the reliability of the Giger features remained questionable because of their setting and operator dependence.

The bar graphs in Figures 6 and 7 suggest that the mathematic features based on the aspect ratio of the lesion (ie, D:W ratio and L:S ratio) were not effective for differentiating a malignant from a benign lesion, though D:W ratio is considered as a clinically useful indicator (19). In particular, the L:S ratio was devised to eliminate the dependence on the scanning angle that is inherent in the D:W ratio. The low classification accuracy of the L:S ratio seemed to imply that the aspect ratio of a lesion is not a useful indicator for lesion malignancy.

In conclusion, setting independence is clearly a crucial property for a CAD algorithm to be used in practice. To assist differential diagnosis of benign and malignant breast lesions without imposing constraints on system settings, we propose, on the basis of findings in the present study, a new CAD algorithm with nearly setting-independent morphologic features and an artificial neural network as the classifier. The proposed morphologic features were by no means comprehensive, though the experimental results supported that NSPD, LI, and ENC are effective and reliable. We believe that further exploration of the setting-independent regional features that may faithfully characterize echotexture, sound transmission, and angular margin would be required to form a complete set of mathematic features for CAD of breast lesions.

APPENDIX

Suppose the contour of a breast lesion is composed of N consecutive discrete points. Let p_i stand for the ith point in the contour of a breast lesion, assuming that the points in the contour are numbered clockwise. For each point p_i, the depression depth h_i is the shortest distance from p_i to the convex hull of the breast lesion. The convex hull of a breast lesion is the smallest convex set of points that enclose the lesion, as depicted in Figure 1. Denote $ñ_i$ as the normal vector of p_i in the contour of a breast lesion. The k-curve angle of p_i in the contour is defined as $θ_i(k) = \mathrm{sgn}(ñ_{i+k} \times ñ_{i-k})\cos^{-1}(ñ_{i+k} \cdot ñ_{i-k})$, where $ñ_{i+k} \times ñ_{i-k}$ is the outer product of $ñ_{i+k}$ and $ñ_{i-k}$, $ñ_{i+k} \cdot ñ_{i-k}$ is the inner product of $ñ_{i+k}$ and $ñ_{i-k}$, and sgn takes the polarity of $ñ_{i+k} \times ñ_{i-k}$, which would be positive if $ñ_{i+k} \times ñ_{i-k}$ points upward.

A point p_i is defined as a convex point if

$θ_i(k) \geq θ_p$, where $θ_p$ is a prespecified positive threshold value. Similarly, a point p_i is defined as a concave point if $h_i \geq \sqrt{2}$ pixels and h_i is the local maximum among the neighborhood of p_i and $θ_i(k) \leq θ_d$, where $θ_d$ is a prespecified negative threshold value. The threshold $\sqrt{2}$ pixels, which is the distance between two diagonal pixels, is set to eliminate undesirable depression caused by the unsteady delineation process. If two consecutive convex points do not have any concave point in between, the one with the smaller k-curve angle is eliminated. Likewise, if two consecutive concave points do not have any convex point in between, the one with the smaller depression depth is removed. Let $\Lambda = \{\lambda_1, \lambda_2, ..., \lambda_p\}$ and $\Omega = \{\omega_1, \omega_2, ..., \omega_d\}$ be the set of points after redundant points have been removed, where p and d are the numbers of points in each set. Then, each point λ_i in Λ is called a representative convex point that defines a substantial protuberance and each point ω_i in Ω is called a representative concave point that defines a substantial depression.

Empirically, $θ_d$ was determined in consideration of two conflicting observations. On one hand, it is common to find a depression with a slowly varying contour (ie, the k-curve angle is small) so that $θ_d$ should be kept as small as possible. On the other hand, it is easy to generate a depression with a small k-curve angle and a small depression depth simply owing to a wobbly delineation process, which may be considered as a noise. As a compromise, $θ_d$ was set to $-20°$, which tolerated $10°$ of aberration from the ideal contour for each side of a concave point.

To determine k, consider a depression with the smallest possible depression depth

(ie, $h_i = \sqrt{2}$ pixels). One may easily obtain that $k \approx 8$ pixels for $θ_d = -20°$ by approximating the depression as a triangle and determining k with the Pythagorean theorem. When the discrete property of a digital image is taken into account, it is appropriate to use either $k = 7$ or 8 to evaluate a depression under the lower-bound condition (ie, given $h_i = \sqrt{2}$ pixels and $θ_d = -20°$). In this study, k was set to 7 to allow the smaller depressions and protuberances. Since there was no reasonable constraint for $θ_p$, we decided to determine $θ_p$ through learning from data. More specifically, five $θ_p$ values were considered in this study (ie, $θ_p \in \{20°, 30°, 40°, 50°, 60°\}$), the best of which was determined by using feature selection.

References

1. Pisani P, Parkin DM, Ferlay J. Estimates of the worldwide mortality from 18 major cancers in 1985: implications for prevention and projections of future burden. Int J Cancer 1993; 55:891–903.
2. Brown ML, Houn F, Sickles EA, Kessler LG. Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. AJR Am J Roentgenol 1995; 165:1373–1377.
3. Zheng Y, Greenleaf JF, Gisvold JJ. Reduction of breast biopsies with a modified self-organizing map. IEEE Trans Neural Networks 1997; 8:1386–1396.
4. Fornage BD, Sneige N, Faroux MJ, Andry E. Sonographic appearance and ultrasound-guided fine-needle aspiration biopsy of breast carcinomas smaller than 1cm. J Ultrasound Med 1990; 9:559–568.
5. Bassett LW, Ysrael M, Gold RH, Ysrael C. Usefulness of mammography and sonography in women less than 35 years of age. Radiology 1991; 180:831–835.
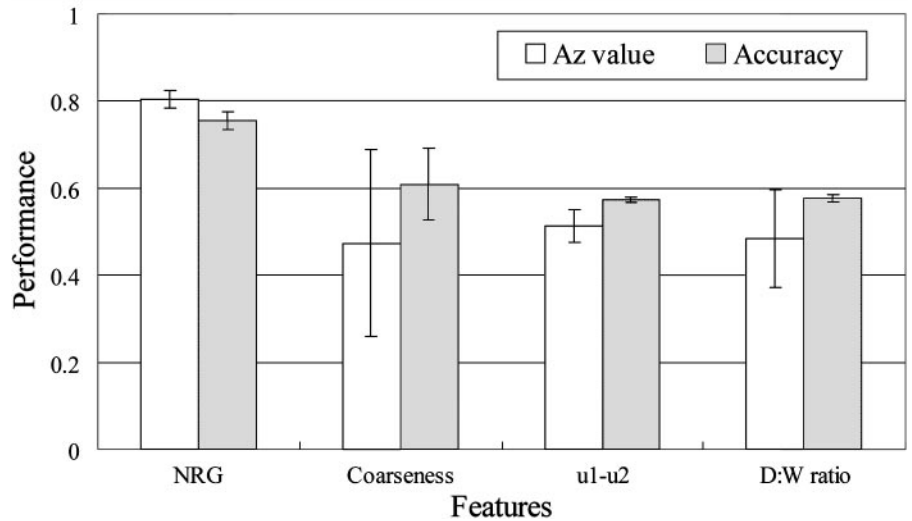6. Jackson VP. The role of US in breast imaging. Radiology 1990; 177:305–311.

Figure 8. In bar graph, each pair of vertical bars indicates the mean A_z value and mean best classification accuracy achieved with a Giger feature with the first set of US images. Error bars indicate 1 SD. All these features (*NRG* = normalized radial gradient) are worse than NSPD, ENS, and ENC at the 5% significance level.

7. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology 1995; 196:123–134.

8. Skaane P, Engedal K. Analysis of sonographic features in the differentiation of fibroadenoma and invasive ductal carcinoma. AJR Am J Roentgenol 1998; 170: 109–114.

9. Goldberg V, Manduca A, Ewert DL, Gisvold JJ, Greenleaf JF. Improvement in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence. Med Phys 1992; 19:1475–1481.

10. Garra BS, Krasner BH, Horii SC, Ascher S, Mun SK, Zeman RK. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. Ultrason Imaging 1993; 15:267–285.

11. Ruggiero C, Bagnoli F, Sacile M, Rescinito CG, Sardanelli F. Automatic recognition of malignant lesions in ultrasound images by artificial neural networks. In: Proceedings of the 20th Annual International Conference of the IEEE/EMBS. New York, NY: Institute of Electrical and Electronic Engineers, 1998; 872–875.

12. Giger M, Al-Hallaq H, Hui Z, et al. Computerized analysis of lesions in US images of the breast. Acad Radiol 1999; 6:665–674.

13. Chen DR, Chang RF, Huang YL. Computer-aided diagnosis applied to US of solid breast nodules by using neural networks. Radiology 1999; 213:407–412.

14. Chen DR, Chang RF, Huang YL. Breast cancer diagnosis using self-organizing map for sonography. Ultrasound Med Biol 2000; 26:405–411.

15. Chang RF, Kuo WJ, Chen DR, Huang YL, Lee JH, Chou YH. Computer-aided diagnosis for surgical office based breast ultrasound. Arch Surg 2000; 135:696–699.

16. Cherkassky V, Mulier F. Learning from data: concepts, theory, and methods. New York, NY: Wiley, 1998.

17. Dillon WR, Goldstein M. Multivariate analysis: method and applications. New York, NY: Wiley, 1984.

18. Zurada JM. Introduction to artificial neural systems. Boston, Mass: PWS, 1992.

19. Tohno E, Cosgrove DO, Sloane JP. Ultrasound diagnosis of breast diseases. Edinburgh, Scotland: Churchill Livingstone, 1994; 50–73.

20. Prokop RJ. The technique of standard moments for global feature object representation. Thesis. Cornell University, Ithaca, NY, 1990.

21. Reeves AP, Prokop RJ, Anfrews SE, Kuhl FP. Three-dimensional shape analysis using moments and Fourier descriptors. IEEE Trans PAMI 1988; 10:937–943.

22. Haralick RM, Shapiro LG. Computer and robot vision. New York, NY: Addison-Wesley, 1993.

23. He DC, Wang L, Guibert J. Texture discrimination based on an optimal utilization of texture features. Pattern Recognition 1989; 21:141–146.

24. Tsoukalas LH, Uhrig RE. Fuzzy and neural approaches in engineering. New York, NY: Wiley, 1997.

25. 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; 187:81–87.

Radiology