Computer-Aided Analysis of Ultrasound Elasticity Images for Classification of Benign and Malignant Breast Masses

**OBJECTIVE.** The purpose of this study was to evaluate computer-aided analysis of ultrasound elasticity images for the classification of benign and malignant breast tumors.

**MATERIALS AND METHODS.** Real-time ultrasound elastography of 140 women (mean age, 46 years; age range, 35–67 years) with nonpalpable breast masses (101 benign and 39 malignant lesions) was performed before needle biopsy. A region of interest (ROI) was drawn around the margin of the mass, and a score for each pixel was assigned; scores ranged from 0 for the greatest strain to 255 for no strain. The diagnostic performances of a neural network based on all six features were compared with visual assessment of elasticity images and BI-RADS assessment using B-mode images.

**RESULTS.** The values for the area under the receiver operating characteristic curve (AUC) of the six elasticity features—mean hue histogram value, skewness, kurtosis, difference histogram variation, edge density, and run length—were 0.84, 0.69, 0.63, 0.75, 0.68, and 0.71, respectively. The sensitivity, specificity, positive predictive value, and negative predictive value of the neural network based on all six features were 92% (36/39), 74% (75/101), 58% (36/62), and 96% (75/78), respectively, with an AUC value of 0.89, which is significantly higher than the AUC of 0.81 for visual assessment by radiologists (p < 0.01) and 0.76 for BI-RADS assessment using B-mode images (p = 0.002).

**CONCLUSION.** Computer-aided analysis of ultrasound elasticity images has the potential to aid in the classification of benign and malignant breast tumors.

Ultrasound elastography was developed to depict the stiffness of tissue by measurement of the strain induced by probe compression [1, 2]. Organs to which pressure can be applied are candidates for elastographic evaluation, and these organs include the breast, thyroid, prostate, liver, vessels, and lymph nodes [1, 3–8]. Recent studies have shown that elastography had almost the same diagnostic performance as conventional ultrasound, with 78–88% sensitivity, 80–98% specificity, and 80–90% accuracy, for the differentiation of benign and malignant solid breast masses [9–13]. Several diagnostic criteria including lesion visualization, relative brightness, and lesion size were proposed after elasticity and B-mode images were compared [4]. For color-coded images, a 5-point visual scoring system based on the degree and uniformity of color in a hypoechoic lesion has been used [10].

The classification of elasticity images is, however, dependent on the examiner, and significant interobserver variability has been found in reader studies [13–15]. Therefore, a quantitative method or computer-aided analysis of elasticity images is needed to evaluate lesion stiffness objectively and to aid in the task of classification of benign and malignant breast tumors. The use of computer-aided analysis of B-mode images or color Doppler images of breast lesions has shown the potential to improve the diagnostic accuracy of radiologists in distinguishing between malignant and benign breast lesions [16, 17]. Recently, computer-aided analysis of elastographic images obtained at endoscopic ultrasound was used for the differentiation of benign pancreatic lesions from malignant pancreatic lesions, and the method showed promising results, with a value for the area under the receiver operating characteristic (ROC) curve (AUC) of 0.96 [18]. To our knowledge, however, computer-aided analysis of ultrasound elastography has not been used for the classification of benign and malignant nonpalpable breast tumors.
The purpose of this study was to evaluate computer-aided analysis of ultrasound elasticity images for the classification of benign and malignant nonpalpable breast tumors.

Materials and Methods

Database

Between May 2006 and November 2006, 171 consecutive women who had been scheduled to undergo an ultrasound-guided percutaneous 14 gauge-core needle biopsy were examined with ultrasound elastography. Of the 171 women, 23 women with a palpable mass and eight women with inadequate images due to improper compression were excluded from the study.

A total of 140 nonpalpable breast masses (101 benign and 39 malignant lesions) in 140 women (mean age, 46 years; age range, 35–67 years) constituted this study group. Lesions manifested as clinically occult mammographic lesions in 40 women, nipple discharge or discomfort in 15 women, and an incidental ultrasound lesion in 85 women. This study was approved by the local ethics committee, and informed consent was obtained from all of the included patients.

The malignant masses included infiltrating ductal carcinoma (n = 35), infiltrating lobular carcinoma (n = 1), and ductal carcinoma in situ (DCIS) (n = 3). The benign lesions were fibroadenomas (n = 50), papillomas (n = 10), and fibrocystic changes (n = 41). The diameters of the lesions based on histology results were 5–30 mm (mean, 13.2 mm) for invasive cancers and 5–25 mm for DCIS lesions. The size of the benign lesions as determined by ultrasound was 4–20 mm (mean, 10.1 mm).

According to the American College of Radiology (ACR) BI-RADS [19], the final assessments of the 140 solid breast masses determined before biopsy based on ultrasound were as follows: category 3 (probably benign lesions) for nine masses, category 4 (suspicion of malignancy) for 116 masses, and category 5 (highly suggestive of malignancy) for 15 masses. Biopsy was performed in the nine probably benign lesions because of patient or referring clinician preference on clinical grounds. Elasticity images were not used to determine the need for biopsy. In two patients with papilloma at core biopsy, discordance between the imaging finding (irregular-shaped mass) and pathology result was found, and surgical excision was performed. No upgrade was found in either case. Surgical excision of 39 masses was performed because of malignant findings after a previously performed percutaneous needle biopsy.

Data Acquisition

Elasticity images were obtained using a scanner (EUB-8500, Hitachi Medical) with a 14–6-MHz linear transducer by one of five radiologists with 1–10 years of experience in the performance of breast ultrasound and with knowledge of the clinical and mammographic findings. The five operators had 4–6 months’ experience using ultrasound elastography before this study. For data acquisition, a region of interest (ROI) box was set to include the region from the subcutaneous fat layer to the superficial portion of the pectoralis muscle layer, and transverse and longitudinal real-time imaging of the target mass was performed. The target lesion was vertically compressed as the operator applied very light pressure to the transducer. Operators avoided using high levels of pressure, which manifest nonlinear properties of tissue elasticity; in such circumstances, the association between pressure and strain is no longer proportional.

The elasticity images were displayed with the use of 256-color mapping for each pixel according to the degree of strain using a scale from red (highest strain, softest), green (average strain, intermediate), to blue (no strain, hardest). The radiologist who performed the real-time imaging selected two representative elasticity and B-mode static images of the masses. The images were sent to a PACS and were then saved as bit-map files on a hard disk. The mean pixel resolution was 95 pixels/cm for both the elasticity and B-mode images.

Computer-Aided Analysis

Because the original color-scale elasticity images were converted to translucent images and were superimposed on the corresponding B-mode images for this ultrasound elastography system, original color-scale elasticity images were obtained by subtracting the B-mode images from the color-coded elasticity images. Then, an ROI was drawn manually around the margin of a mass on a B-mode image (Fig. 1) by the radiologist who performed the real-time elastography and two expert radiologists with 10–15 years’ experience in the practice of breast ultrasound in consensus. This ROI was superimposed on subtracted color-scale elasticity images using a paint program (Image J, version 5.3, National Institutes of Health). The value of each pixel in a subtracted color-scale elasticity image was assigned a value from 0 for the greatest strain to 255 for no strain.

Six elasticity features—mean hue histogram value, skewness, kurtosis, difference histogram variation, edge density, and run length—were computed by one of the investigators to evaluate the findings of benign and malignant tumors. The mean hue histogram value was calculated as the total color value of the pixels within the tumor divided by the total number of pixels inside the tumor boundary. Skewness was defined as a measure of symmetry [20]. A distribution, or data set, is symmetric if it appears the same to the left and right of the center point. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution [21]. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. Difference histogram variation is one of the texture features that can be extracted from a histogram of gray-

Fig. 1—50-year-old woman with infiltrating ductal carcinoma.

A. B-mode image shows 1.2-cm hypoechoic mass with ill-defined margins (arrow). Final assessment was suspicious (BI-RADS category 4).

B. Elasticity image with color coding shows strain at periphery of hypoechoic lesion and sparing of center of lesion. Elasticity score was 3.

C. Subtracted (B) – |A| color-scale image with region of interest around tumor margin (arrow) shows elasticity images without B-mode information. In this case, computer-aided analysis based on six elasticity features classified lesion as malignant.
level differences [20]. Edge density was defined as the mean value of edge pixels in a unit area [20]. Run length was defined as a measurement of runs of consecutive points that all have the same gray level. Run length provides information about the coarseness of image texture in specified directions [22]. A detailed description of the method for calculating difference histogram variation, edge density, and run length is provided in Appendix 1.

A general multilayer perceptron neural network [23] with the back-propagation learning rule was used to classify solid breast tumors on the basis of the values of the six features. The values produced by the output node of the neural network were between 0 and 1. We chose a threshold of 0.5 to classify the benign and malignant tumors after conducting experiments. If the output value was equal to or higher than 0.5, a tumor was classified as malignant. If the output value was less than 0.5, a tumor was regarded as benign. The results obtained using the multilayer perceptron neural network were compared with those obtained using Bayesian classification, a classic linear classification method [24].

Visual Assessment by Radiologists

Two breast radiologists retrospectively reviewed the elasticity and B-mode images of the 140 patients without knowledge of mammographic and histology information by consensus. The elasticity and B-mode images were mixed and separated using the Image J program so that the reviewers could look at the elasticity image and B-mode image individually for blind review. A 21-inch (53-cm) video monitor and PACS software were used in a darkened room.

On the basis of the color pattern in the mass, each image was assigned an elasticity score on a 5-point scale according to the classification proposed by Itoh et al. [10]. A score of 1 indicated even strain for the entire hypoechoic lesion. A score of 2 indicated strain in most of the hypoechoic lesion with some areas of no strain. A score of 3 indicated strain at the periphery of the hypoechoic lesion with sparing of the center of the lesion. A score of 4 indicated no strain in the entire hypoechoic lesion. A score of 5 indicated no strain in the entire hypoechoic lesion or in the surrounding area.

For conventional ultrasound, the radiologists were given instructions that the malignancy risks of each category determined by ultrasound were according to the ACR BI-RADS final assessment [12]. Category 4 was subclassified into 4A, 4B, and 4C. Thus, category 4A included risks from >2% to ≤10%, category 4B included risks from >10% to ≤50%, and category 4C included risks from >50% to <95%.

Data and Statistical Analysis

The mean values and SDs of the six elasticity features were calculated for the benign and malignant tumors. Significant differences between the values of the six features for benign and malignant tumors were evaluated with the independent-samples Student’s t test. An assumption of equal variances of the two populations was determined with the Levene test, and Student’s t test results then were interpreted accordingly. That is, the Student’s t test was applied when equal variances were assumed and the Welch t test was applied when the assumption was not fulfilled. The performance of the values for the features was also evaluated with ROC curve analysis by use of a computer program (LABROCI 1993, Charles E. Metz, University of Chicago).

The diagnostic performance of the neural network and that of the Bayesian classifier based on the values of the six elasticity features used for classification of solid breast tumors were evaluated with sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) and AUC analysis. The elasticity score and neural network probability of nonpalpable breast masses according to the BI-RADS category on conventional ultrasound were analyzed. In addition, ROC curve analysis was performed to compare the diagnostic performances of the neural network based on the values of the six features, visual assessment of the elasticity images by the radiologists, and the combination with conventional ultrasound.

Table 1—Receiver operating characteristic curves of values for six features and for neural network classifier (all features). Area under receiver operating characteristic curve (AUC) value of mean hue histogram value feature was significantly higher than AUC value of other five features (p < 0.001). Value for neural network with all six features produced best performance, with AUC value of 0.89.

Table 2—Comparison of mean values for six features and for neural network classifier (all features). Significant differences between the values of the six features for benign and malignant tumors. Differences between benign and malignant breast tumors were statistically significant for all six features (p < 0.01).

Table 3—Comparison of diagnostic performance of the neural network and that of the Bayesian classifier based on the values of the six elasticity features used for classification of solid breast tumors.

Fig. 2—Receiver operating characteristic curves of values for six features and for neural network classifier (all features). Area under receiver operating characteristic curve (AUC) value of mean hue histogram value feature was significantly higher than AUC value of other five features (p < 0.001). Value for neural network with all six features produced best performance, with AUC value of 0.89.
Visual assessment of the elasticity images by the radiologists showed 54% (21/39) sensitivity, 91% (92/101) specificity, 70% (21/30) PPV, and 84% (92/110) NPV when a cutoff between an elasticity score of 3 and 4 was used. When a cutoff point between an elasticity score of 2 and 3 was used, elastography had 97% (38/39) sensitivity, 40% (40/101) specificity, 38% (38/99) PPV, and 98% (40/41) NPV. One malignant lesion with an elasticity score of 2 that was classified as benign by the neural network analysis was a 0.5-cm infiltrating ductal carcinoma in the background of DCIS. Nine benign lesions with an elasticity score of 4 included four fibroadenomas, two papillomas, and three cases of fibrocystic changes. They were assessed as category 4 in seven cases and category 3 in two cases on conventional ultrasound, and three of the seven lesions were classified as benign by the neural network analysis.

Table 1: Elasticity Score and Neural Network Probability of 140 Nonpalpable Breast Masses According to BI-RADS Category

<table>
<thead>
<tr>
<th>BI-RADS Category and Elasticity Score</th>
<th>No. of Lesions</th>
<th>No. of Lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Benign (n=101)</td>
<td>Malignant (n=39)</td>
</tr>
<tr>
<td>BI-RADS category 3 (n=9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>BI-RADS category 4A (n=92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>BI-RADS category 4B (n=13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>BI-RADS category 4C (n=11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>BI-RADS category 5 (n=15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

*One malignant lesion misclassified as benign on ultrasound elastography was a 0.5-cm infiltrating ductal carcinoma in the background of ductal carcinoma in situ.

Discussion

Our study results showed that computer-aided analysis of breast lesions on ultrasound elastography is feasible and has the potential to be used for the classification of benign and malignant tumors. The A value was 0.89 for the computer-assisted analysis, which is significantly higher than the A value of 0.81 for visual assessment with a 5-point scoring system used by the radiologists (p = 0.01) and 0.76 for BI-RADS category on conventional ultrasound (p = 0.002). When combined with conventional ultrasound, the A value of visual assessment by the radiologists and the neural network analysis increased to 0.88 and 0.93, respectively (p < 0.001).

In our study, the A values were used as indicators of performance because different threshold or cutoff points can result in differences in sensitivity and specificity. The neural network showed no improvement from conventional ultrasound in terms of sensitivity and specificity when we use BI-RADS subcategory 4A or 4B as a cutoff point. Our database included only nonpalpable suspicious cases, BI-RADS category 4 or 5 lesions (except for the nine probably benign lesions), on conventional ultrasound that required an interventional procedure to determine tissue diagnosis.

For interpretation of color-coded elasticity images, a 5-point visual scoring system has been widely used and a cutoff point between an elasticity score of 3 and 4 has been recommended for the classification of benign and malignant tumors [10]. Our study results, however, suggest that a lower cutoff point be-
between an elasticity score of 2 and 3 instead of an elasticity score of 3 and 4 should be used to aid in reducing the number of benign biopsies performed in clinical practice. Only 54% (21/39) sensitivity, 91% (92/101) specificity, and 84% (92/110) NPV were obtained with use of a cutoff between an elasticity score of 3 and 4, whereas 97% (38/39) sensitivity, 40% (40/101) specificity, and 97% (40/41) NPV were obtained with use of a cutoff between an elasticity score of 2 and 3. Only one malignant lesion (a 0.5-cm infiltrating ductal carcinoma in the background of a DCIS) showed an elasticity score of 2. However, differentiating a lesion with an elasticity score of 2 from a lesion with a score of 3 is sometimes difficult, especially when assessing small lesions.

By definition, an elasticity score of 2 indicated strain in most of the hypoechoic lesion with some areas of no strain and an elasticity score of 3 indicated strain at the periphery of the hypoechoic lesion with sparing of the center of the lesion. The categorization based on visual assessment is simple and fast, but it is subjective and inconsistent. To overcome these problems, various quantification methods have been described in the literature with the use of elasticity images [12, 15, 25]. Sohn et al. [12] used the area ratio, which was defined as the ratio of the area without strain to the area of the lesion on elastography and was automatically calculated by the ultrasound machine. Another example is the strain index; the strain index is a measure of the relative stiffness of breast lesions and is calculated by comparison with the subcutaneous fat and the same depth of breast tissue [25]. The diagnostic performance based on the area ratio, or strain index, of the lesion was better than the elasticity score.

In our study, the six elasticity features were chosen to quantify the degree and uniformity of the color within the lesion boundary. The first three elasticity features—mean hue histogram value, skewness, and kurtosis—are related to color intensity and distribution of the hue histograms. The remaining three features—difference histogram variation, edge density, and run length—are related to the texture of tumors. Notably, the Az value for the mean hue histogram value feature was significantly higher than the Az values of the other five features (p < 0.001). This finding concurs with a recent report by Saftoiu et al. [18] describing computer-aided analysis of pancreatic masses with the use of elasticity images. Those investigators used only the mean hue histogram value for their neural network analysis and reported an excellent performance, with an A of 0.96 for the classification of benign and malignant pancreatic masses. In addition to the elasticity features, B-mode features based on mass morphology and echogenicity can be used to differentiate benign from malignant breast lesions. A neural network based on a combination of elasticity and B-mode features to classify benign and malignant breast tumors should be tested.

Some limitations of the current study should be considered. First, only two elasticity images per case selected by the radiologists were used for the analysis; this could be a possible bias. Fluctuation of the strain image patterns depends on the compression magnitude, and misclassification may occur because an improper amount of pressure was applied during scanning or because inadequate images were selected for interpretation. Second, we drew the tumor contours by hand and this could introduce considerable bias. Methods to automate the segmentation of breast lesions have been reported [26]. Third, the threshold for the neural network to classify the benign and malignant tumors was established after data collection, which could be a limitation of the study. Last, the ROC studies were not performed by the radiologists with aid of the computer-aided analysis system in our study.

In conclusion, we have developed a method for computer-aided analysis of ultrasound elasticity images for the classification of benign and malignant breast tumors. The results obtained showed that the neural network based on elasticity features can be successfully used to classify benign and malignant breast tumors. A biopsy may not be required when a BI-RADS category 4A or 4B lesion has an elasticity score of 1 or 2 and a low probability (< 0.5) at the computer-aided analysis.

References
**Ultrasound Elastography of Breast Masses**

1. **Difference histogram variation**
   
   Gray-level differences define the equation:
   
   \[ g(d) = |u(k, l) - u(k + d_1, l + d_2)| \]

   Let \( d = (d_1, d_2) \) be the displacement vector between two image pixels and \( u(k, l) \) be the gray level in point \((k, l)\). If \( P_g(g, d) \) is the histogram of the gray-level differences at the specific distance \(d\), then difference histogram variation \( \delta_d^2 \) is defined as follows:

   \[
   \delta_d^2 = \frac{1}{N} \sum_{k=1}^{N} \sum_{g=1}^{G} (g_k - \mu_d^2)^2 P_g(g, d)
   \]

2. **Edge density**

   A pixel location \((m, n)\) was declared as an edge location if its gradient \(g(m, n)\) exceeded some threshold \(t\). The location of edge points constituted an edge map \(e(m, n)\), which was defined as follows:

   \[
   e(k, l) = \begin{cases} 
   1, & (k, l) \in I_g \\
   0, & \text{otherwise}
   \end{cases}
   \]

   where

   \[ I_g = \{(m, n); g(m, n) > t\} \]

   Given an edge map, the edge density was measured by use of the average number of edge pixels per unit area.

3. **Run length**

   In a directional texture, run lengths that occur along a given line should depend on the direction of the line. The feature used was called the “run percentage” \((RP)\), for which the equation is as follows:

   \[
   \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} Q_{R \cdot L(i, j)}
   \]

   where \( Q_{R \cdot L(i, j)} \) is the number of run lengths \(j\) for gray level \(i\) in some direction \(0\), \(N_g\) is the number of gray levels of the image, \(N_r\) is the number of different run lengths, and \(P\) is the total number of image pixels. Run lengths for \(0\), \(45^\circ\), \(90^\circ\), and \(135^\circ\) and the sample mean and SD were estimated.