

A Comparison of Logistic Regression Analysis and an Artificial Neural Network Using the BI-RADS Lexicon for Ultrasonography in Conjunction with Interobserver Variability

Sun Mi Kim · Heon Han · Jeong Mi Park ·
Yoon Jung Choi · Hoi Soo Yoon · Jung Hee Sohn ·
Moon Hee Baek · Yoon Nam Kim · Young Moon Chae ·
Jeon Jong June · Jiwon Lee · Yong Hwan Jeon

Published online: 24 January 2012
© Society for Imaging Informatics in Medicine 2012

Abstract To determine which Breast Imaging Reporting and Data System (BI-RADS) descriptors for ultrasound are predictors for breast cancer using logistic regression (LR) analysis in conjunction with interobserver variability be-

tween breast radiologists, and to compare the performance of artificial neural network (ANN) and LR models in differentiation of benign and malignant breast masses. Five breast radiologists retrospectively reviewed 140 breast masses and described each lesion using BI-RADS lexicon and categorized final assessments. Interobserver agreements between the observers were measured by kappa statistics. The radiologists' responses for BI-RADS were pooled. The data were divided randomly into train ($n=70$) and test sets ($n=70$). Using train set, optimal independent variables were determined by using LR analysis with forward stepwise selection. The LR and ANN models were constructed with the optimal independent variables and the biopsy results as dependent variable. Performances of the models and radiologists were evaluated on the test set using receiver-operating characteristic (ROC) analysis. Among BI-RADS descriptors, margin and boundary were determined as the predictors according to stepwise LR showing moderate interobserver agreement. Area under the ROC curves (AUC) for both of LR and ANN were 0.87 (95% CI, 0.77–0.94). AUCs for the five radiologists ranged 0.79–0.91. There was no significant difference in AUC values among the LR, ANN, and radiologists ($p>0.05$). Margin and boundary were found as statistically significant predictors with good interobserver agreement. Use of the LR and ANN showed similar performance to that of the radiologists for differentiation of benign and malignant breast masses.

S. M. Kim · H. Han (✉) · H. S. Yoon · J. H. Sohn · M. H. Baek ·
J. Lee · Y. H. Jeon
Department of Radiology,
Kangwon National University College of Medicine,
192-1 Hyoja 2-dong,
Chuncheon, Kangwon-do 200-701, Republic of Korea
e-mail: kimsmlms@paran.com

S. M. Kim
Department of Radiology,
Seoul National University Bundang Hospital,
Seongnam-Si, Republic of Korea

J. M. Park
Division of Breast Imaging and Intervention,
Department of Radiology, Carver College of Medicine,
University of Iowa Hospitals and Clinics,
Iowa City, IA, USA

Y. J. Choi
Department of Radiology, Kangbuk Samsung Hospital,
Sungkyunkwan University School of Medicine,
Seoul, Republic of Korea

Y. N. Kim · Y. M. Chae
Department of Health Informatics,
Graduate School of Public Health, Yonsei University,
Seoul, Republic of Korea

J. J. June
Department of Statistics, Seoul National University,
Seoul, Republic of Korea

Keywords Breast · Ultrasonography · Artificial neural network · Breast neoplasm · Logistic regression

Introduction

With recent improvements in ultrasonographic technology, including the use of all-digital high-frequency transducers up to 13 MHz and the use of color and power Doppler and harmonic imaging, sonography has become a standard breast imaging procedure [1–7]. In addition, these improvements have allowed expansion of the indications for use of breast ultrasound (US), including tumor differentiation, pre-operative staging, follow-up after cancer treatment, and interventional diagnosis [8–11].

The main limitation of breast US is that it is a highly operator-dependent modality. In addition, observer variability in sonographic interpretation has an effect on mass characterization and management. To overcome these limitations, the American College of Radiology has now developed a sonographic breast imaging reporting and data system (BI-RADS) [12]. Until the third edition of BI-RADS, the lexicon was only applicable to mammography. Use of mammographic BI-RADS has resulted in unified understanding and implications of various mammographic terms, as well as increased observer agreement in mammographic readings [13, 14]. Recent reports on observer variability with the use of ultrasonographic BI-RADS have shown good results with kappa statistic agreement [15–17]. Although a high level of agreement has been shown for some descriptors, a final assessment showed a low level of agreement (fair agreement) [14, 16]. Therefore, this low level agreement in final assessment leads to a lack of consistency in patient management.

Logistic regression (LR) analysis and an artificial neural network (ANN) have been applied for tumor characterization [18, 19]. The LR model is superior for use in examination of possible causal relationships between independent and dependent variables and in understanding the effects of predictors on outcome variables [20]. The ANN constitutes a non-algorithmic approach to information processing and prediction of the probability of cancer [21]. Use of the ANN has resulted in improved interpretation of mammography and can aid in improvement of recommendations for performance of a biopsy [22–24]. Only one study of a small group of 24 malignant and 30 benign lesions has been conducted for comparison of LR and an ANN for use in breast cancer imaging using breast US [18]. Significant predictors according to LR using ultrasonographic BI-RADS and for evaluation of each predictor's interobserver variability have not been reported.

The purpose of this study was to determine which BI-RADS descriptors for US are statistically significant predictors, using LR in conjunction with interobserver variability among breast radiologists, and to compare the diagnostic performance of ANN and LR statistical models.

Materials and Methods

Case Selection

A computerized search of the electronic medical records, including surgical, pathologic, and breast ultrasonographic findings was performed for identification of pathologically confirmed ultrasonographic breast masses between January 2001 and February 2003 at our medical center. Because the study was retrospective, patients signed a general consent form to cover all diagnostic studies, and neither Institutional Review Board approval nor patient informed consent was necessary. Sonograms of 140 solid masses of 140 patients, who ranged in age from 22 to 70 years (mean age, 44.3 years), were selected for the study. All of the masses had a known diagnosis based on an ultrasonography-guided hook wire localization and excision biopsy. Seventy lesions (50%) were confirmed as malignant and 70 lesions (50%) were benign. Histologic types of malignancies included invasive ductal carcinomas for 55 lesions, ductal carcinomas in situ for 11 lesions, invasive lobular carcinomas for two lesions, an invasive tubular carcinoma for one lesion, and a secretory carcinoma for one lesion. Benign lesions included fibroadenomas for 54 lesions, papillomas for eight lesions, sclerosing adenosis for three lesions, adenosis for two lesions, ductectasia for two lesions, and a tubular adenoma.

Assessment of Ultrasonographic Findings

Ultrasonography was performed in the transverse (axial) and longitudinal (sagittal) planes using a HDI 3000 or 5000 ultrasound scanner (Philips-Advanced Technology Laboratories, Bothell, WA, USA) equipped with a 5–12-MHz linear array transducer. The most experienced breast radiologist (JM Park) selected one key image from each case on a picture archiving and communication system and converted images into tag image file format (TIFF) files with 300 dots per inch. All TIFF files were arranged in arbitrary order in Microsoft Power Point.

Five subspecialty-trained breast radiologists (SM Kim, HS Yoon, YJ Choi, MH Beak, and JH Son) with 7, 5, 3, 1, and 1 year of experience, respectively, performed a retrospective review of all of the images. All five investigators were familiar with the use of ultrasonographic BI-RADS descriptors during their daily work and no formal training for the descriptions was involved in this study. All of the observers performed an independent review of all 140 images without knowledge of clinical information, mammographic findings, and pathologic results of each case, as well as the incidence ratio of malignant to benign lesions. All observers described each lesion using the BI-RADS lexicon and categorized the final assessment (Table 1). Of the seven categories (from 0 to 6) in the BI-RADS, the categories of 0

Table 1 BI-RADS descriptors for breast ultrasonography

Shape	Oval, round, irregular
Orientation	Parallel, non-parallel
Margin	Circumscribed Indistinct, angular, microlobulated, spiculated
Lesion boundary	Abrupt interface, echogenic halo
Echo pattern	Anechoic, hyperechoic, complex, hypoechoic, isoechoic
Posterior acoustic features	Absent, enhancement, shadowing, combined
Calcifications	Macrocalcifications, microcalcifications in the mass, microcalcifications out of the mass
Associated findings	Ductal change, Cooper’s ligament change, edema, architectural distortion, skin thickening, skin retraction/irregularity
Final assessment	Category 2, category 3, category 4, category 5

(incomplete assessment), 1 (normal), and 6 (biopsy-proven malignancy) were already excluded in this study.

Statistical Analysis for Interobserver Variability

Interobserver variability in selection of the BI-RADS lexicon in each category was assessed with the kappa statistic. Kappa values range from 1.0 (complete agreement) to -1.0 (complete disagreement). A value of $k=0$ corresponds to random agreement. The kappa statistic, as described by Landis and Koch, was noted as slight agreement for $0 < k < 0.20$, fair agreement for $0.21 < k < 0.4$, moderate agreement for $0.41 < k < 0.6$, substantial agreement for $0.61 < k < 0.8$, and perfect agreement for $0.81 < k < 1.0$ [25].

LR Procedure

A histologic diagnosis of malignancy for a breast mass was entered as a dependent variable in the LR model and was coded as 0 for absent (benign) and 1 for present (malignant). The regression equation derived from each of the training subsets is applied to the just removed cross-validation subset for estimation of the individual probability of having breast cancer.

If Y is denoted as an indicator of cancer, the probability of cancer, given x_i , is as follows [19, 26]:

$$P(Y = 1|x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}$$

where $x_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ is the i th observation with the k covariate BI-RADS lexicon descriptors, \exp is the base of the natural logarithm, and β_0 is the intercept and $\beta_j (j = 1, \dots, k)$ are coefficients corresponding to the j th BI-RADS lexicon descriptors. We used stepwise LR for selection of significant predictors for breast cancer.

ANN Procedure

We used a three-layer back-propagation ANN, known as multilayered perceptrons. This ANN includes at least three layers of neurons (an input layer, an output layer, and at least one hidden layer).

$x = (x_1, x_2, \dots, x_k)$ are BI-RADS lexicon descriptors and $z = (z_1, z_2, \dots, z_m)$ is a feature value created by the sigmoid function of the input variable. y is the linear combination z . For each i ,

$$z_{ij} = \frac{\exp(\alpha_{0j} + \alpha_{1j}x_{i1} + \dots + \alpha_{kj}x_{ik})}{1 + \exp(\alpha_{0j} + \alpha_{1j}x_{i1} + \dots + \alpha_{kj}x_{ik})}, \quad j = 1, \dots, m$$

$$y = \beta_1 z_1 + \dots + \beta_m z_m$$

Coefficients are obtained by minimizing $\sum_{i=1}^n \sum_{j=1}^m (y_i - \beta_j z_{ij})^2$. An ANN model is a type of general additive model and has good performance for fitting a complex model. Commercially available software (SAS Enterprise Miner, version 4; SAS Institute, Cary, NC, USA) was used for construction of the ANN. Neurons are tied together with weighted connections. The number of nodes in the input corresponds to the number of input variables. The output layer has one node, including values from zero to one, indicating the level of malignancy. Our ANN software consisted of a neuron in the hidden layer.

Statistical Analysis

All of the BI-RADS lexicon descriptors were used for stepwise LR analysis. Descriptors included shape, orientation, margin, boundary, echogenicity, posterior acoustic features, calcifications, and associated findings.

Responses of the five radiologists for the BI-RADS lexicon descriptors were pooled. For binary data, if three or more radiologists gave a positive response, the pooled response was considered positive; otherwise, the pooled response was considered negative. For ordinal data, the

pooled response was the median value of the five radiologists' responses. The frequencies of the vote were 28.1% in 3–2, 27% in 4–1, and 44.9%, 5–0. Pooled data were used in the subsequent statistical analysis. The 140 sets of data were divided randomly into 70 for the training set and 70 for the test set. The training set consisted of 41 malignancies and 29 benign masses where as test set included 29 malignancies and 41 benign masses.

With the training set, an optimal subset of independent variables was determined by performance of forward stepwise selection using the likelihood-ratio statistic ($p=0.05$ for entry and $p=0.10$ for removal) as a selection criterion. With the determined subset of independent variables, a final model was constructed by fitting the LR and ANN on the entire training set. The constructed LR and ANN models were then validated using the test set.

Receiver operator characteristic (ROC) curve analysis was used for comparison of the diagnostic performance of the ANN, LR, and the radiologists in order to distinguish between benign and malignant breast masses. Area under the ROC curve (AUC value) was calculated for each fitted ROC curve using MedCalc software (MedCalc, Mariakerke, Belgium). MedCalc software provides the empirical ROC curve and non-parametric estimate of the area under the empirical ROC curve with its 95% CI, based on the method developed by Hanley et al. [27]. A comparison between two paired ROC curves is available and the statistical significance of the difference between two AUCs is calculated using the z test [27]. p values <0.05 were considered to indicate a significant difference. At the cutoff value yielding the highest accuracy, the sensitivity, specificity, and accuracy of each model and the radiologists were compared.

Results

Interobserver Agreement

Interobserver agreements for the ultrasonographic descriptors are shown in Table 2. Overall agreement of shape ($k=0.46$), orientation ($k=0.51$), overall agreement of margin ($k=0.53$), lesion boundary ($k=0.45$), and calcifications ($k=0.47$) were moderate, and posterior features ($k=0.40$) and final assessment ($k=0.40$) were fair.

For each descriptor, moderate agreement was obtained for an oval and irregular shape, parallel and non-parallel orientation, circumscribed margin, abrupt interface, and echogenic halo of boundary, absent and combined posterior features, microcalcifications in the mass, and categories 3 and 5. Fair agreement was obtained for an indistinct, angular, microlobulated and spiculated margin, isoechoic echo pattern, enhanced and shadowing posterior features, macrocalcifications, and category 4 for the final assessment. Slight

agreement was obtained for a round shape, anechoic, hyperechoic, complex, and hypoechoic echo pattern, microcalcifications out of the mass, and category 2 of the final assessment.

Optimal Variable Selection and Training of the Multiple Linear Regression Model

According to the results of stepwise selection, the margin and boundary were determined as the optimal subset of independent variables, while the remaining variables were excluded. From the entire training set, regression coefficients (β) for the margin and boundary were estimated as 2.29 (95% CI, 0.79, 3.79) and 1.23 (–0.14, 3.6), respectively. The equation of the LR model constructed on the training set was determined as follows:

$$\text{Logit}(p) = -1.73 + 2.29 \times \text{margin} + 1.23 \times \text{boundary}$$

Performance of the ANN and LR Models

Area under the ROC curve (AUC) values for LR and the ANN were 0.87 (95% CI, 0.77, 0.94) and 0.87 (0.77, 0.94) for the testing set, respectively. Values for performance of the radiologists were 0.88 (95% CI, 0.79, 0.95), 0.79 (95% CI, 0.68, 0.88), 0.85 (95% CI, 0.75, 0.93), 0.91 (95% CI, 0.82, 0.97), and 0.85 (95% CI, 0.75, 0.93), respectively (Table 3, Fig. 1). There was no significant difference in the AUC value among the use of LR, the ANN, and the performance of the radiologists ($p>0.05$). At the cutoff value yielding the highest accuracy, the accuracy of LR, the ANN, and the five radiologists were 0.80, 0.80, 0.83, 0.70, 0.80, 0.86, and 0.80, respectively (Table 3).

Discussion

The primary aim of this study was to determine statistically significant predictors from BI-RADS ultrasonographic descriptors using LR in conjunction with interobserver variability. We used a stepwise LR procedure for isolation of significant predictors. Among all ultrasonographic BI-RADS features, margin and boundary were shown to be statistically significant predictors. Agreements for margin and boundary were relatively higher, compared with other descriptors, such as echo pattern or a posterior feature.

Breast ultrasonography has traditionally been used for detection and diagnosis of cysts. Improvements in image quality from the use of new ultrasound techniques have expanded the role of US, and US is now an essential tool for use in differentiation of benign from malignant breast masses [8]. The most significant shortcoming of the use of

Table 2 Interobserver variability in description and final assessment according to US BI-RADS Lexicon

BI-RADS lexicons	Descriptors and final category	Percentage of agreement (n=140) ^a	k value
Shape	Oval	15 (21)	0.49
	Round	0	0.04
	Irregular	33.6 (47)	0.49
Orientation	Overall	48.6 (68)	0.46
	Parallel	49.3 (69)	0.51
	Non-parallel	8.6 (12)	0.51
Margin	Overall	57.9 (81)	0.51
	Circumscribed	15 (21)	0.53
	Not circumscribed	42.1 (59)	0.53
Boundaries	Indistinct	6.4 (9)	0.27
	Angular	2.9 (4)	0.28
	Microlobulated	1.4 (2)	0.25
	Spiculated	0	0.25
	Overall	57.1 (80)	0.53
	Abrupt interface	35 (49)	0.45
	Echogenic halo	12.1 (17)	0.45
Echo pattern	Overall	47.1 (66)	0.45
	Anechoic	0	0.09
	Hyperechoic	0	0.09
	Complex	1.4 (2)	0.08
	Hypoechoic	16.4 (23)	0.16
	Isoechoic	2.1 (3)	0.21
	Overall	20 (28)	0.15
Posterior features	Absent	18.6 (26)	0.45
	Enhancement	10.7 (15)	0.38
	Shadowing	1.4 (2)	0.28
	Combined	1.4 (2)	0.41
	Overall	32.1 (45)	0.40
Calcifications	Absent	68.6 (96)	0.58
	Macrocalcifications	0	0.34
	Microcalcifications in the mass	1.4 (2)	0.52
	Microcalcifications out of the mass	0	0.06
	Overall	70 (98)	0.47
Final assessment	Category 2	0	0.08
	Category 3	7.9 (11)	0.47
	Category 4	12.9 (18)	0.32
	Category 5	5.7 (8)	0.49
	Overall	26.4 (37)	0.40

Data are combined for all five readers

^aPercentage of agreement and number of observations in agreement (in parentheses)

US is that performance and interpretation of breast ultrasound is subjective. To overcome this limitation, ultrasonographic BI-RADS was developed for characterization of breast masses by qualitative assessment of lesion features within an image [12]. A variety of image features based on shape, orientation, margin, lesion boundary, echogenicity, posterior acoustic features, calcifications, and associated findings have been used. Taken together, these qualitative descriptors have been shown to improve the specificity of ultrasonographic findings [28].

Despite the use of ultrasonographic BI-RADS, the interobserver variability of the use of ultrasonographic BI-RADS terminology has shown agreement ranging from slight-to-substantial [14–17]. Reported agreements of shape, orientation, and boundary have been moderate-to-substantial, and, for margin, have been fair-to-substantial [14–17]. Our findings also showed moderate agreement in the assessment of shape, orientation, margin, boundary, and calcifications. In general, determination of a parallel or non-parallel orientation to the skin of the mass can be easily assessed,

Table 3 Receiver-operating characteristic curves for performance of LR, the ANN, and the radiologists on the test set

ROC analysis	LR	ANN	Radiologist 1	Radiologist 2	Radiologist 3	Radiologist 4	Radiologist 5
AUC value	0.87 (0.77, 0.94)	0.87 (0.77, 0.94)	0.88 (0.79, 0.95)	0.79 (0.68, 0.88)	0.85 (0.75, 0.93)	0.91 (0.82, 0.97)	0.85 (0.75, 0.93)
Cutoff value yielding the highest accuracy	0.19	0.26	3	3	4	3	3
Sensitivity %	93.1 (77.2, 99.0)	93.1 (77.2, 99.0)	96.6 (82.2, 99.4)	93.1 (77.2, 99)	51.7 (32.5, 70.5)	100 (87.9, 100)	89.7 (72.6, 97.7)
Specificity %	70.7 (54.5, 83.9)	70.7 (54.5, 83.9)	73.2 (57.1, 85.8)	51.2 (35.1, 67.1)	100 (91.3, 100)	75.6 (59.7, 87.6)	75.6 (59.7, 87.6)
Accuracy	0.80 (54.5, 83.9)	0.80 (54.5, 83.9)	0.83 (71.9, 90.8)	0.70 (57.8, 80.3)	0.80 (54.5, 83.9)	0.86 (75.2, 82.9)	0.80 (54.5, 83.9)

Numbers in parentheses are 95% CI

ANN artificial neural network, AUC area under the receiver-operating characteristic curve, LR logistic regression, ROC receiver-operating characteristic curve

explaining the relatively good interobserver variability. For assessment of the shape of a mass, there were three descriptors, including oval, round, or irregular. An oval shape can show up to three gentle lobulations; with greater than four lobulations, the mass should be considered as having an irregular shape. However, there is no consistency in assessment of the number of lobulations. For assessment of margin, there are four descriptors for not being circumscribed, including indistinct, spiculated, angular, and microlobulated. Slight-to-fair agreement was noted in determination of a non-circumscribed margin. When the mass margin was simplified as circumscribed or non-circumscribed, interobserver variability was seen as relatively high-to-moderate.

Although agreements of these descriptors were relatively higher, only a fair level of agreement was found for the final assessment in this study. The reported kappa agreements in the final assessment category, as fair and substantial, were also

relatively low [14–17]. For determination of the final assessment, a breast US specialist usually attempts to identify a typical benign feature or suspicious finding. A spiculated margin, irregular shape, and non-parallel orientation are known to have a high predictive value for malignancy and a circumscribed margin, oval shape, and parallel orientation are highly predictive of a benign lesion [28]. If there are any suspicious findings, a biopsy is recommended. If there are typical benign findings suggestive of cysts, or hyperechoic lesions, suggestive of fibroadenomas, follow-up is recommended. Variability in the description of breast masses by the observers resulted in inconsistency in determination of the final assessment. Therefore, management of a breast mass detected on breast US can vary; however, the use of BI-RADS and a relatively high level of agreement were observed for some descriptors.

The secondary aim of this study was to compare the diagnostic performance of LR and the ANN. LR and the ANN are the most frequently used computer models in clinical risk estimation [19]. The advantage in use of the ANN is the capacity to model complex non-linear relationships between independent and predictor variables. Compared with the ANN, LR analysis may be preferred due to improved interpretation of individual predictors [20]. LR and the ANN showed statistically similar performances, compared with the radiologists, even if only two descriptors, margin and boundary, were used in construction of models.

This study differs from an automatically feature extracted CAD system in its use of BI-RADS features selected by experienced radiologists. The standardized BI-RADS sonographic lexicon is known to be helpful in distinguishing benign from malignant solid masses [28]. A large number of structured reporting software programs have recently been developed. Some of them are being filled with BI-RADS descriptors; in this system, ANN or LR models can aid in the final decision regarding a US mass. In this study, according to results of stepwise LR, margin and boundary were found to be statistically significant predictors. This

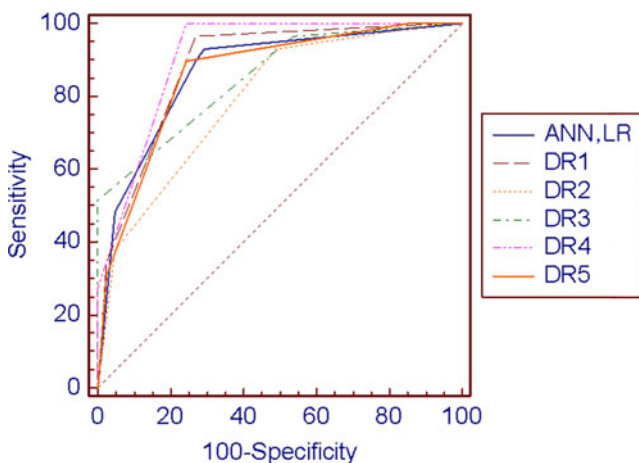


Fig 1 Receiver-operating characteristic curves for performance of logistic regression, artificial neural network, and radiologist. Area under the ROC curve values for logistic regression (LR), the artificial neural network (ANN) were 0.87 (95% CI, 0.77, 0.94) and 0.87 (0.77, 0.94) for the testing set. Performance of the radiologists were 0.88 (95% CI, 0.79, 0.95), 0.79 (95% CI, 0.68, 0.88), 0.85 (95% CI, 0.75, 0.93), 0.91 (95% CI, 0.82, 0.97), and 0.85 (95% CI, 0.75, 0.93), respectively

result regarding morphologic features can be applied to an ultrasound CAD system.

This study has several limitations. Cases were selected retrospectively from surgically excised biopsy-proven masses. The number of category 2 lesions was relatively small. Selection bias may contribute to an over-categorization in some cases. Since this was a retrospective study, real-time ultrasonography was not available at the review point and the masses were interpreted with the use of static images. Observers in this study did not have the opportunity to take advantage of real-time US benefits. Although an experienced operator performed all of the breast ultrasound examinations, some operator dependency while capturing images may have occurred. In their review of images, the radiologists used the ultrasonographic BI-RADS lexicon and had their own standardizations, but no formal training for use of the BI-RADS lexicon. Also, the radiologists worked independently, which may have affected interobserver reproducibility. Use of the BI-RADS lexicon was not strange to the readers; however, a training session may aid in increasing their confidence and in reducing variation. For reducing the fact of substantial variability in the readers, we used pooling data from all five of the radiologist's responses and used median value or dominated value. Such pooling represents an average radiologist's response but does not reflect actual practice. Our reviewers interpreted ultrasonography only, which does not reflect actual practice, which involves categorization, based on both mammographic and ultrasonographic results. This may have lowered the radiologists' performance. Thus, an analysis with more data, including mammographic findings for the LR and ANN models is necessary for comparison with radiologists.

Ultrasonographic BI-RADS descriptors are useful for evaluation of breast masses and offer standard classifications. Among BI-RADS descriptors, margin and boundary were found to be statistically significant predictors according to the results of stepwise LR and showed good interobserver agreement. Although relatively high interobserver agreements were obtained, agreement for the final assessment was only fair. Use of LR and the ANN showed similar performance with that of the radiologists for differentiation of benign and malignant breast masses, despite use of only margin and boundary descriptors. These descriptors can be applied to a CAD approach for breast ultrasound.

References

- Jackson VP, Reynolds HE, Hawes DR: Sonography of the breast. *Semin Ultrasound CT MR* 17:460–475, 1996
- Radiology ACo: ACR Standards 2000–2001. American College of Radiology, Reston, 2000
- Baker JA, Soo MS: Breast US, assessment of technical quality and image interpretation. *Radiology* 223:229–238, 2002
- Baker JA, Soo MS, Rosen EL: Artifacts and pitfalls in sonographic imaging of the breast. *AJR* 176:1261–1266, 2001
- Rizzato GJ: Towards a more sophisticated use of breast ultrasound. *Eur Radiol* 11:2425–2435, 2001
- Schroeder RJ, Bostanjoglo M, Rademaker J, Maeurer J, Felix R: Role of power Doppler techniques and ultrasound contrast enhancement in the differential diagnosis of focal breast lesions. *Eur Radiol* 13:68–79, 2003
- Dempsey PJ: The history of breast ultrasound. *J Ultrasound Med* 23:887–894, 2004
- 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* 196:123–134, 1995
- Berg WA, Gutierrez L, Ness-Aiver MS, Carter WB, Bhargavan M, Lewis RS, et al: Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology* 233:830–849, 2004
- Mendelson EB: Problem-solving ultrasound. *Radiol Clin North Am* 42:909–918, 2004
- Mehta TS: Current uses of ultrasound in the evaluation of the breast. *Radiol Clin North Am* 41:841–856, 2003
- Radiology ACo: Breast Imaging Reporting and Data System, Ultrasound, 4th edition. American College of Radiology, Reston, 2003
- Baker JA, Kornguth PJ, Floyd Jr, CE: Breast imaging reporting and data system standardized mammography lexicon: observer variability in lesion description. *AJR* 166:773–778, 1996
- Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS: BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology* 239:385–391, 2006
- Lee HJ, Kim EK, Kim MJ, Youk JH, Lee JY, Kang DR, et al: Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. *Eur J Radiol* 65:293–298, 2008
- Abdullah N, Mesurolle B, El-Khoury M, Kao E: Breast imaging reporting and data system lexicon for US: interobserver agreement for assessment of breast masses. *Radiology* 252:665–672, 2009
- Park CS, Lee JH, Yim HW, Kang BJ, Kim HS, Jung JI, et al: Observer agreement using the ACR Breast Imaging Reporting and Data System (BI-RADS)-ultrasound, First Edition (2003). *Korean J Radiol* 8:397–402, 2007
- Song JH, Venkatesh SS, Conant EA, Arger PH, Sehgal CM: Comparative analysis of logistic regression and artificial neural network for computer-aided diagnosis of breast masses. *Acad Radiol* 12:487–495, 2005
- Ayer T, Chhatwal J, Alagoz O, Kahn Jr, CE, Woods RW, Burnside ES: Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radiographics* 30:13–22, 2010
- Tu JV: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 49:1225–1231, 1996
- Hecht-Nielsen R: Replicator neural networks for universal optimal source coding. *Science* 269:1860–1863, 1995
- Baker JA, Kornguth PJ, Lo JY, Floyd Jr, CE: Artificial neural network: improving the quality of breast biopsy recommendations. *Radiology* 198:131–135, 1996
- Baker JA, Kornguth PJ, Lo JY, Williford ME, Floyd Jr, CE: Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology* 196:817–822, 1995
- 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* 187:81–87, 1993

25. Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 33:159–174, 1977
26. Chou YH, Tiu CM, Hung GS, Wu SC, Chang TY, Chiang HK: Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis. *Ultrasound Med Biol* 27:1493–1498, 2001
27. Hanley JA, McNeil BJ: A method comparing the areas under receiver operator characteristic curves derived from the same cases. *Radiology* 148:839–843, 1983
28. Hong AS, Rosen EL, Soo MS, Baker JA: BI-RADS for sonography: positive and negative predictive values of sonographic features. *AJR* 184:1260–1265, 2005