Bayesian Probability of Malignancy With BI-RADS Sonographic Features

Ghizlane Bouzghar, MD, Benjamin J. Levenback, BS, Laith R. Sultan, MD, Santosh S. Venkatesh, PhD, Alyssa Cwanger, BS, Emily F. Conant, MD, Chandra M. Sehgal, PhD

Objectives—The purpose of this study was to develop a quantitative approach for combining individual American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) sonographic features of breast masses for assessing the overall probability of malignancy.

Methods—Sonograms of solid breast masses were analyzed by 2 observers blinded to patient age, mammographic features, and lesion pathologic findings. BI-RADS sonographic features were determined by using American College of Radiology criteria. A naïve Bayes model was used to determine the probability of malignancy of all the sonographic features together and with age and BI-RADS mammographic features. The diagnostic performance for various combinations was evaluated by using the area under the receiver operating curve (A_z).

Results—Sonographic features had high positive and negative predictive values. The A_z values for BI-RADS sonographic features for the 2 observers ranged from 0.772 to 0.884, which increased to 0.866 to 0.924 when used with patient age and BI-RADS mammographic features. The benefit of adding age and mammographic information was more marked for the observer with lower initial diagnostic performance. Age-specific analysis showed that diagnostic performance varied with age, with higher performance for patients aged 45 years and younger and patients older than 60 years compared to those aged 46 to 60 years. In 85% of cases, the diagnosis of the observers matched. When the consensus between the observers was used for diagnostic decisions, a high level of diagnostic performance (A_z, 0.954) was achieved.

Conclusions—A naïve Bayes model provides a systematic approach for combining sonographic features and other patient characteristics for assessing the probability of malignancy to differentiate malignant and benign breast masses.

Key Words—Bayes probability of malignancy; breast cancer imaging; Breast Imaging Reporting and Data System (BI-RADS); negative predictive value; positive predictive value

Received May 16, 2013, from the Departments of Radiology (G.B., B.J.L., L.R.S., A.C., E.F.C., C.M.S.) and Electrical Engineering (S.S.V.), University of Pennsylvania, Philadelphia, Pennsylvania USA. Revision requested July 1, 2013. Revised manuscript accepted for publication August 1, 2013.

We thank Theodore W. Cary, Susan M. Schultz, and Karen Apadula for help with data analysis and patient studies. The images used in this study were also used for computer analysis of manually drawn margins of the masses in reference 14. This work was supported in part by National Institutes of Health grant CA130946.

Address correspondence to Chandra M. Sehgal, PhD, Department of Radiology, University of Pennsylvania, 1 Silverstein, 3400 Spruce St, Philadelphia, PA 19104 USA.

E-mail: chandra.sehgal@uphs.upenn.edu

Abbreviations

A_z, area under the receiver operating characteristic curve; BI-RADS, Breast Imaging Reporting and Data System; NPV, negative predictive value; PPV, positive predictive value

doi:10.7863/ultra.33.4.641

Breast cancer is the second leading cause of cancer death in women after lung cancer.[1] Although mammography has proven to be the most effective tool for detecting breast cancer at its earliest and most treatable stage,[2] breast sonography has become an important adjunct to diagnostic mammography. Breast sonography is typically performed on palpable or mammographically identified masses to determine their cystic or solid nature. Since cysts constitute only 25% of all breast lesions, it leaves a large

 | J Ultrasound Med 2014; 33:641–648 | 0278-4297 | www.aium.org

number of lesions in the indeterminate and solid-nodule categories, which require aspiration or biopsy. Although well tolerated, these procedures do have some risk, induce patient discomfort and anxiety, and increase overall health care costs.[3–5] There continues to be a need for improving imaging methods to reduce the number of unnecessary biopsies. With the increased resolution and improvements in near-field imaging, sonography is being used increasingly to characterize solid breast masses.

To standardize lesion characterization and reporting procedures, a grading system that parallels the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) mammographic system was developed for breast sonography in which the risk category is assessed by evaluating the lesion shape, margin, echogenicity, and other features defined by a standardized lexicon.[4,6] Raza et al[7] have demonstrated the use of the BI-RADS sonographic system and the factors that influence decision making. Hong et al[8] observed high positive and negative predictive values for the BI-RADS sonographic features. Although these results are encouraging, it is less clear how individual lesion features are combined to assess the BI-RADS category. The current practice is to weight individual features intuitively toward the final assessment of the BI-RADS category. Although helpful, the intuitive assignment of the BI-RADS category from individual features is subjective and likely to vary between observers[9–12] as well as within the same observer repeated over time. An approach that combines individual features objectively could overcome this limitation.

In this article, we propose a quantitative technique based on a Bayesian model to combine individual BI-RADS sonographic features for assessing the overall probability of malignancy. Since the predictive values of the sonographic features are influenced by the age of the patients,[13] we evaluated the effect of patient age on the diagnostic performance of the proposed method. Finally, we evaluated the difference in the diagnostic performance of 2 observers who independently assessed the sonographic features on the same set of sonograms.

Materials and Methods

Image Acquisition

This study was approved by the Institutional Ethics Committee. A total of 264 masses were obtained from 246 patients with biopsy-proven solid masses and known BI-RADS mammographic data. Each consecutive patient was approached, and written consent for sonography for the study was obtained. Fifty-two masses were BI-RADS category 5 (highly suspicious); 127 were category 4 (suspicious); 9 were category 3 (probably benign); 14 were category 2 (benign); 10 were category 1 (negative); 23 were category 0 (incomplete study); and BI-RADS categories for the remaining 29 were not available. Some patients had more than 1 lesion, and all lesions were included in the study because it was difficult to identify which lesion to leave out objectively. Sonograms were acquired with a broadband 12–5-MHz transducer and an HDI 5000 scanner (Philips Healthcare, Bothell, WA). Five to 7 images were acquired per patient in radial and antiradial planes.

Sonogram Analysis

Images were analyzed for sonographic features by 2 physician observers (G.B. and L.R.S.) according to the American College of Radiology guidelines.[6] Before analyzing the images, the observers, with 2 years of prior training in general radiology, underwent self-study of training cases consisting of breast images with known BI-RADS features and pathologic findings. The observers were blinded to patient age, race, physical examination findings, family history, the mammographic report, and the histologic diagnosis. Sonographic features were assessed for each case in radial and antiradial planes. When the assessed features in both planes were different, images that showed features with greater clarity were selected. BI-RADS defines echogenicity relative to subcutaneous fat.[6] Some authors have studied echogenicity relative to fat and relative to surrounding tissue.[4] In this study, we used both approaches, comparing the echogenicity of the mass to the surrounding tissue as well as to subcutaneous fat. The positive predictive values (PPVs) of the individual features determined by the observers were measured by taking the ratio of true-positive findings to the sum of true- and false-positive findings. Similarly, the negative predictive values (NPVs) of the individual features were measured by taking the ratio of true-negative findings to the sum of true- and false-negative findings. The statistical significance (P value) of the predictive values was evaluated by the Pearson χ^2 test. To evaluate the contribution of each feature individually, the cases with a given feature present were also compared to the cases without that feature by the χ^2 test.

Analysis of Combined BI-RADS Sonographic Features

The probability of malignancy, $P(M|F)$, given multiple features (F) was determined by a Naïve Bayes model[14]:

$$(1)\quad P(M \mid F) = P(M \mid F_1, F_2, \ldots F_N) = \frac{P(M) \times \prod_{i=1}^{N} P(F_i \mid M)}{[P(M) \times \prod_{i=1}^{N} P(F_i \mid M)] + [P(B) \times \prod_{i=1}^{N} P(F_i \mid B)]},$$

where the probability that the feature F_i is present given the mass is malignant, $P(F_i|M)$, was determined by taking the ratio of the number of malignant cases with features F_i over the total number of malignant cases. The probability that the feature F_i is present given the mass was benign, $P(F_i|B)$, was determined by taking the ratio of the number of benign cases with features F_i over the total number of benign cases. The prior probability that the mass was malignant, $P(M)$, or benign, $P(B)$, was determined by the ratio of the number of malignant or benign cases to the total number of cases studied. The probability of malignancy for each case based on all the sonographic features was determined by Equation 1 using leave-one-out cross-validation. The diagnostic performance of the measured probabilities was assessed by measuring the area under the receiver operating characteristic curve (A_z; MedCalc Software, Ostend, Belgium). The age of the patient and BI-RADS mammographic features were used as additional features along with sonographic features to assess whether they could be used to enhance diagnostic performance. The statistical significance of the area under 2 different receiver operating characteristic curves was calculated by the method developed by DeLong et al.[15]

To evaluate the age-specific influence of patient age on diagnostic performance, the patients were divided into 3 groups: patients aged 45 years and younger, patients aged 46 to 60 years, and patients older than 60 years. Age thresholds of 45 and 60 years were used for grouping because these values were close to the mean ages of the patients with benign and malignant masses as described below. For each patient group, diagnostic performance was measured separately.

Finally, we combined the outputs of both observers to determine whether the combined use improved the diagnostic performance. Toward this goal, Bayesian probabilities of the observers were compared, and at different probability thresholds, agreement between the observers was determined. For example, given a threshold T, cases in which both observers assigned a probability value higher than T were classified as malignant, and cases in which both observers assigned a probability value lower than T were classified as benign. Receiver operating characteristic curve analysis was performed on the cases with agreement, and the cases with disagreement (in which one observer classified a case as benign and the other observer classified it as malignant) were considered indeterminate, which needed further evaluation by other diagnostic methods.

Results

General Characteristics

Of the 264 lesions, 85 (32%) were malignant and 179 (68%) were benign. Among the malignant lesions, invasive ductal carcinoma was the most common (65 [76%]). Other diagnoses included invasive lobular carcinoma (6 [7%]), ductal carcinoma in situ (6 [7%], including 1 papillary carcinoma in situ case), adenocarcinoma (3 [3%]), and 1 remaining case, which was diagnosed as mucinous mammary carcinoma (a rare form of invasive ductal carcinoma). Of the benign masses, 44% were found to be fibroadenoma, 33% were identified as miscellaneous fibrocystic changes, 6% were sclerosing adenosis, and the remaining 17% were identified as benign lesions without atypia in the histopathologic reports. The mean age of the entire patient population ± SD was 51.5 ± 14.7 years. The mean age of the patients with malignant masses was 58.8 ± 12.1 years compared to 48.0 ± 14.5 years for those with benign cases. The difference in the mean ages of the malignant and benign groups was statistically significant ($P < .0001$).

BI-RADS Sonographic Descriptor Analysis

Figure 1 shows an example of infiltrating ductal carcinoma with BI-RADS features suggestive of malignancy. The mass appears markedly hypoechoic with an irregular border and strong posterior acoustic shadowing; it is taller than wide with a nonparallel orientation and spiculated margin. Fibroadenoma, on the other hand, has benign characteristics: oval shape, well-defined margin, parallel orientation, isoechogenicity relative to fat, and posterior acoustic enhancement (Figure 2). The PPVs and NPVs for the individual features are summarized in Table 1. Features predictive of malignancy were irregular shape (PPV, 56%; NPV, 90%), nonparallel orientation (PPV, 60%; NPV, 86%), indistinct margin (PPV, 45%; NPV, 70%), spiculation (PPV, 72%; NPV, 75%), echogenic halo (PPV, 64%; NPV, 77%), and shadowing (PPV, 61%; NPV, 79%). Cooper ligament changes, skin retraction, calcifications, diffuse vascularity in surrounding tissue, and an angular margin were also predictive of malignancy, but the numbers of cases observed were too few to provide meaningful evaluation. The PPVs for mass shape, mass orientation, mass margin, lesion boundary, echo pattern, and posterior features were statistically significant ($P \leq .001$). The PPVs of the surrounding tissue features, calcifications, and vascularity did not reach statistical significance ($P \geq .1$).

Table 2 summarizes the predictive values of echogenicity of the mass relative to the surrounding tissue and to fatty tissue. Although the predictive values of echogenicity

Figure 1. Sonogram of a breast mass diagnosed as infiltrating ductal carcinoma showing features highly suggestive of malignancy, including a spiculated margin with an irregular border, a taller-than-wide shape, a nonparallel orientation, marked hypoechogenicity, and a strong posterior acoustic shadow.

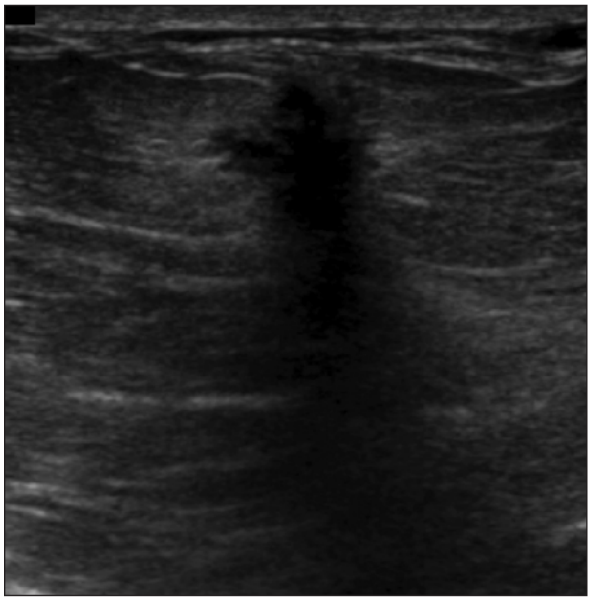


Figure 2. Sonogram of a fibroadenoma showing features usually associated with benign breast masses, including an oval shape, a well-defined margin, a parallel orientation, isoechogenicity (compared to subcutaneous fat), and posterior acoustic enhancement.

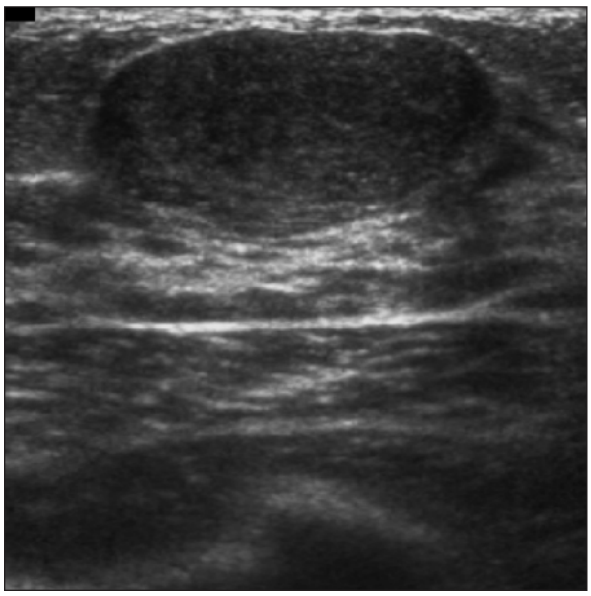


Table 1. Predictive Values of BI-RADS Sonographic Features

Descriptor	n	PPV, %	NPV, %	*P′*	*P″*	PPV, %[8]
Mass shape				<.001	<.001	
Oval	129	10	55			16
Irregular	136	56	90			62
Mass orientation				<.001	<.001	
Parallel	155	13	40			22
Not parallel	107	60	86			69
Mass margin				<.001	<.001	
Circumscribed	9	6	67			10
Indistinct	37	45	70			46
Angular	19	49	69			60
Microlobulated	151	17	47			51
Spiculated	39	72	75			86
Diffuse	9	56	69			
Lesion boundary				<.001	.16	
Abrupt interface	214	25	52			29
Echogenic halo	50	64	77			70
Echo pattern				<.001	.14	
Anechoic	4	12.5	68			50
Hyperechoic	1	0.0	68			0
Complex	58	36	69			10
Hypoechoic	162	38	77			40
Isoechoic	40	4	62			16
Posterior acoustic features				<.001	<.001	
None	69	25	65			21
Enhancement	94	14	58			33
Shadowing	74	61	79			52
Combination pattern	26	31	68			50
Surrounding tissues				.57	<.001	
Duct enlargement	3	20	68			
Cooper ligament changes	3	80	68			
Edema	0	–	68			
Architectural distortion	3	0	67			
Skin thickening	12	29	68			
Skin retraction/irregularity	1	100	68			
Calcifications				.15	<.001	
Macrocalcifications	6	45.5	68			
Microcalcifications outside mass	11	57	69			
Microcalcifications inside mass	8	60	69			
None	242	30	44			
Vascularity				.16	<.001	
Not assessed	10	10.5	67			
Not present	21	17	66.5			
Present in lesion	137	39	75			
Present immediately adjacent lesion	219	35	82			
Diffused vascularity in surrounding issue	3	40	68			
Special cases				.42	<.001	
Clustered microcyst	1	0	68			
Complicated cysts	1	0	68			
Mass in or on skin	1	0	68			
Lymph nodes present	0	–	68			

P′ indicates *P* value for PPV; and P″, *P* value for NPV.

relative to fatty tissue were highly significant, they were not significant relative to the surrounding tissue; ie, comparing the mass echogenicity to fat provided better results (NPV, 95.3; PPV, 4.7) than comparing it to surrounding tissue (NPV, 66.6; PPV, 33.3). On an individual feature basis, only hypoechoic and isoechoic features demonstrated statistical difference (Table 2).

The mean probabilities of malignancy ± SD determined by the naïve Bayes model using all the sonographic features assessed by the 2 observers were .17 ± .29 and .17 ± .32 for the benign group versus .77 ± .33 and .68 ± .40 for the malignant group. The difference between the benign and the malignant groups was statistically significant ($P < .00001$, Student *t* test).

The diagnostic performance of the probabilities determined by using all the sonographic features together and in combination with age and BI-RADS mammographic features for the 2 observers is summarized in Table 3. The A_z values for all the BI-RADS sonographic features for the observers were 0.884 ± 0.025 and 0.772 ± 0.032. The difference was significant ($P < .001$). When patient age was also included with the sonographic features, the performance (A_z) of the Bayes probability estimates improved by a small amount to 0.894 ± 0.024 for observer 1 and more noticeably to 0.818 ± 0.030 for observer 2 (Table 3). The improvement in the performance with age was significant for both observers (observer 1, $P = .014$; observer 2, $P < .001$). Adding BI-RADS mammographic categories further improved the diagnostic performance of the probability estimates to 0.924 ± 0.021 and 0.866 ± 0.027 for observers 1 and 2, respectively (Table 3). The improvement in the performance with sonographic features, age, and BI-RADS mammographic features compared to sonographic features only was significant for both observers (observer 1, $P = .002$; observer 2, $P < .001$).

Age-specific analysis of the diagnostic performance of observer 1 shows that the groups with patients aged 45 years and younger and patients older than 60 years had high performance, with A_z values of 0.946 ± 0.052 and 0.917 ± 0.034, respectively. The performance of each of these individual groups was better than the overall performance (A_z, 0.884 ± 0.025). The group with patients aged 46 to 60 years had an A_z of 0.837 ± 0.043 and thus had lower diagnostic performance compared to the performance of the whole group. The data for observer 2 showed a similar trend but with lower performance; the A_z values for the groups with patients aged 45 years and younger, 46 to 60 years, and older than 60 years were 0.775 ± 0.094, 0.722 ± 0.053, and 0.794 ± 0.054, respectively (Table 4).

Table 2. Predictive Values of Echo Patterns Compared to Fat Versus Compared to Surrounding Tissue

	Comparison to Fat					Comparison to Surrounding Tissue				
Echo Pattern	**n**	**NPV, %**	**PPV, %**	**2 ×2 *P***	***P*(All)**	**n**	**NPV, %**	**PPV, %**	**2 × 2 *P***	***P* (All)**
Anechoic	3	100	0	.32	<.0002	3	100	0	.32	.6867
Hyperechoic	1	100	0	.57		3	33.3	66.7	.29	
Complex	69	63.8	36.2	.39		70	65.7	34.3	.60	
Hypoechoic	129	55.8	44.2	<.001		184	69	31	.60	
Isoechoic	64	95.3	4.7	<.001		6	66.7	33.3	.96	

P(all) compares the significance of all features together; 2 × 2 *P* assesses the significance of each feature individually.

Table 3. Comparison of Bayes Probabilities and Diagnostic Performance of the Sonographic Features When Used Alone and With Age and BI-RADS Mammographic Features

Parameter	**Observer 1**	**Observer 2**	**Consensus**
Bayes probability of malignancy, P(M|F) ± SD			
Benign masses	.17 ± .29	.17 ± .32	.04 ± .06
Malignant masses	.77 ± .33	.68 ± .40	.94 ± .10
Diagnostic performance, A_z ± SE			
Sonographic features alone	0.884 ± 0.025 (0.836–0.934)	0.772 ± 0.032 (0.709–0.835)	0.914 ± 0.022 (0.871–0.957)
Sonographic features with age	0.894 ± 0.024 (0.847–0.941)	0.818 ± 0.030 (0.759–0.877)	0.929 ± 0.020 (0.889–0.968)
Sonographic features with age and BI-RADS	0.924 ± 0.021 (0.883–0.965)	0.866 ± 0.027 (0.813–0.919)	0.954 ± 0.016 (0.922–0.986)

Values in parentheses are 95% confidence intervals.

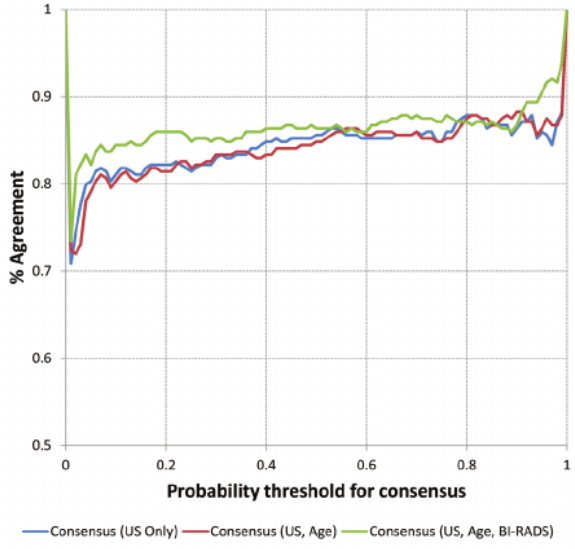
A consensus curve for the combined use of the Bayesian probabilities from the 2 observers is shown in Figure 3. The A_z values were 0.914 ± 0.022 with 84.5% ± 3.5% agreement for sonographic features, 0.929 ± 0.020 with 84.3% ± 3.8% agreement for sonographic features and age, and 0.954 ± 0.016 with 86.6% ± 3.0% agreement for sonographic features, age, and BI-RADS mammographic features. For this curve, the diagnosis had specificity of 0.75 and 0.88 at sensitivity of 0.95 and 0.90, respectively.

Table 4. Age-Specific Diagnostic Performance of the Sonographic Features

	A_z ± SE	
Age, y	Observer 1	Observer 2
≤45	0.946 ± 0.052 (0.844–1)	0.775 ± 0.094 (0.591–0.959)
46–60	0.837 ± 0.043 (0.752–0.920)	0.722 ± 0.053 (0.618–0.826)
>60	0.917 ± 0.035 (0.848–0.986)	0.794 ± 0.054 (0.688–0.900)

Values in parentheses are 95% confidence intervals.

Figure 3. Percent agreement between observers 1 and 2 as a function of the probability threshold for differentiating malignant and benign masses. The Consensus (US Only) curve represents agreement between the observers when only sonographic features were used for diagnostic decisions. The Consensus (US, Age) curve represents agreement between the observers when sonographic features were used in conjunction with the ages of the patients. The Consensus (US, Age, BI-RADS) curve represents agreement between the observers when sonographic features were used in conjunction with the ages of the patients and BI-RADS mammographic features.



Discussion

The BI-RADS lexicon for sonography is based on the established lexicon used successfully in mammography and attempts to provide a common language to avoid ambiguity in interpreting, reporting, and teaching breast sonography. It is a data control system that provides a lexicon for describing lesions, establishes levels of suspicion for breast cancer, and indicates the required subsequent steps for evaluation and treatment of breast cancer. Proper and consistent use of the BI-RADS sonographic lexicon has numerous advantages, including facilitating communication of final assessment categories that clearly indicate management recommendations, data tracking for self-audits, and clinical review of outcome summaries.[7,8]

The BI-RADS sonographic lexicon includes 6 morphologic features of solid breast masses: shape, orientation, margin, lesion boundary, internal echo pattern, and posterior acoustic features.[16] Rahbar et al[17] found that the features most likely to predict a benign diagnosis in solid masses were a round or oval shape, a circumscribed margin, and a width-to-anteroposterior ratio greater than 1.4. Features most predictive of malignancy were an irregular shape, a microlobulated or spiculated margin, and a width-to-anteroposterior ratio of 1.4 or less.

In our study, BI-RADS sonographic features that showed a high predictive value of malignancy ($P < .001$) included an irregular shape, spiculations, a nonparallel orientation, shadowing, and the presence of an echogenic halo. Features that were found to have a high NPV were an oval shape, a parallel orientation, a circumscribed margin, an abrupt interface, an anechoic or hyperechoic pattern, the presence of posterior enhancement, and the absence of vascularity. Hong et al[8] calculated the predictive values of BI-RADS sonographic features in a study conducted on 403 breast masses. Although the individual values were different, as one would expect from the different composition of the database, the overall outcome was the same: ie, the BI-RADS features are predictive, with PPVs and NPVs ranging from 45% to 72% and 15% to 94%, respectively.

We found that comparing the echogenicity of the mass against fat yielded better results than comparing it to surrounding tissue. These results support an earlier study by Stavros et al,[4] which showed that more useful information can be gained by comparing nodule echogenicity to a structure that has echogenicity near the middle of the grayscale spectrum. Breast terminal ductal lobular units, breast periductal elastic tissue, and fat have echogenicity that is near the middle of the grayscale spectrum. Of these, only fat is uniformly present in all patients. Also, the

American College of Radiology recommends that fatty tissue be used as the reference tissue for comparison of echogenicity on breast sonography,[6] setting the standard for "isoechogenicity" within a breast. However, to compare echogenicity of solid nodules to subcutaneous fat, sonographic parameters must be set so that fat is portrayed as gray rather than black to increase the chances of accurately assessing the echogenicity of a solid nodule and identifying other lesions.[4] The results of this study also show that of the 5 echo patterns, only hypoechoic and isoechoic features contributed to the observed difference (Table 2).

Although sonographic features are used to determine BI-RADS sonographic assessments, the process by which such assignments are made is subjective. The weight assigned to individual features for BI-RADS characterization is observer dependent and likely to vary with the observer's clinical training and experience. In this article, we demonstrate the use of a Bayesian model to combine individual features based on their PPVs and NPVs. The results are encouraging and show that when using the sonographic features alone, the Bayesian method of weighting provides a systematic approach for combining features to obtain a high level of diagnostic performance, with an A_z of approximately 0.884; ie, 88.4% of the time, a randomly chosen patient from the malignant group has a test value (probability of malignancy) higher than that of a patient chosen randomly from the benign group. However, the performance can vary with the observer. In this study, observer 2 had lower performance, with an A_z of approximately 0.772. The difference in the performance between the observers is primarily a result of the difference in the assessment of the individual features. These results are consistent with previous studies that have shown that the overall agreement for assessing BI-RADS sonographic features ranges between fair and moderate, as measured by κ statistics.[9–11] A stronger emphasis on training in analyzing sonograms could contribute to a reduction in intraobserver variability and promote a more uniform diagnosis.

Although there is a significant difference in the mean age of patients with malignant lesions compared to those with benign lesions,[18–20] the results of this study show that adding age as a feature only resulted in a small improvement (ΔA_z, 0.010, or 1%) in the diagnostic performance when the original performance was relatively high (A_z, 0.884). However, when the performance was lower (A_z, 0.772), the benefit of adding age to the analysis was more pronounced, and the performance (ΔA_z) increased by 0.046, or 6%. The reason for this difference is not known but could have been related to how much of the diagnostic information present at a given age was accounted for in the sonographic analysis by the observers. When the overlap between the features is small, there is greater improvement in the diagnostic performance. Raza et al[7] also found that the age of the patient was an important clinical factor in determining the likelihood of malignancy. They suggested that for an older patient, the threshold for biopsy should be lowered, and biopsy may be warranted even for probably benign imaging features. The age-specific analysis in this study showed that all age groups did not have equal diagnostic performance. The patient group aged 46 to 60 years had lower performance compared to the patients aged 45 years and younger and the patients older than 60 years. Although further validation with larger studies is needed, the implication of these results is that more attention should be given to certain patient age groups when analyzing sonograms: eg, the age group of 46 to 60 years in this study.

As one may expect for complex tasks of image evaluation, there are differences in the performance between observers. This factor raises the question of whether assessment of the individual observers can be combined to improve diagnostic decisions. In principle, several approaches are possible. In this study, consensus between the probability estimates of the observers was used to guide diagnostic decisions. The cases in which the observers did not agree were considered cases that needed further evaluation by other diagnostic methods and not suitable for decisions. The results show that by combining the probability estimates of 2 observers, the diagnostic performance increased substantially for all configurations: sonographic features alone, sonographic features with age, and sonographic features with age and mammographic features. In each case, the combined performance was better than that of each individual observer. For consensus-based decisions using sonographic features, age and BI-RADS mammographic features had a high level of diagnostic performance, and an A_z of 0.954 was achieved. This finding represents specificity of 0.75 and 0.88 at sensitivity of 0.95 and 0.90, respectively. Other configurations using sonographic features only or with age yielded lower performance but still had A_z values exceeding 0.9. The gain in performance, however, comes at the cost of approximately 15% of cases requiring further evaluation. In effect, the consensus-based method acts as a filter for identifying cases in which a diagnosis can be made with higher confidence.

In conclusion, this study shows that a Bayesian model can be used with BI-RADS sonographic features only or in combination with other patient-specific information to determine the probability of malignancy. These quantitative estimates of probabilities have a high level of diagnostic performance for differentiating malignant from benign

breast masses. This study also demonstrates that combining the probability estimates of 2 observers, analogous to double reading, improves the diagnostic performance markedly. Although the results are encouraging, further studies with larger data sets involving multiple users are needed to demonstrate the clinical applicability of the proposed approach.

References

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2006. 2006; 56:106–130.
2. Kopans DB. Standardized mammographic reporting. 1992; 30:257–261.
3. Mendelson EB, Berg WA, Merritt CRB. Toward a standardized breast ultrasound lexicon, BI-RADS: ultrasound. 2001; 36:217–225.
4. Stavros T, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. 1995; 196:123–134.
5. Barr RG. Breast ultrasound: a bright future. 2001; 45:8–13.
6. American College of Radiology. BI-RADS: ultrasound. In: . 5th ed. Reston, VA: American College of Radiology; 2003:1–149.
7. Raza S, Goldkamp AL, Chikarmane SA, Birdwell RL. Ultrasound of breast masses categorized as BI-RADS 3, 4, and 5: pictorial review of factors influencing clinical management. 2010; 30:1199–1213.
8. Hong AS, Rosen EL, Soo MS, Baker JA. BI-RADS for sonography: positive and negative predictive values of sonographic features. 2005; 184:1260–1265.
9. 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. 2009; 252:665–672.
10. Lee HJ, Kim EK, Kim MJ, et al. Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. 2008; 65:293–298.
11. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: inter-observer variability and positive predictive value. 2006; 239:385–391.
12. Baker JA, Kornguth PJ, Soo MS, Walsh R, Mengoni P. Sonography of solid breast lesions: observer variability of lesion description and assessment. 1999; 172:1621–1625.
13. Fu CY, Hsu HH, Yu JC, et al. Influence of age on PPV of sonographic BI-RADS categories 3, 4, and 5. 2011; 32(suppl 1):S8–S13.
14. Cary TW, Cwanger A, Venkatesh SS, Conant EF, Sehgal CM. Comparison of naïve Bayes and logistic regression for computer-aided diagnosis of breast masses using ultrasound imaging. In: Bosch JG, Doyley MM (eds). Vol 8320. Bellingham, WA: SPIE; 2012:83200M-1–83200M-7.
15. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. 1988; 44:837–845.
16. Jackson VP. Management of solid breast nodules: what is the role of sonography? 1995; 196:14–15.
17. Rahbar G, Sie AC, Hansen GC, et al. Benign versus malignant solid breast masses: US differentiation. 1999; 213:889–894.
18. Sehgal CM, Cary TW, Kangas SA, et al. Computer-based margin analysis of breast sonography for differentiating malignant and benign masses. 2004; 23:1201–1209.
19. Harvey P, Arger PH, Conant EF, Sehgal CM. Differentiation of the solid benign and malignant breast masses by quantitative analysis of the ultrasound images. In: Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2009:530–533.
20. Arger PH, Sehgal CM, Conant EF, Zuckerman J, Rowling SE, Patton JA. Interreader variability and predictive value of ultrasound descriptions of solid breast masses: pilot study. 2001; 8:335–342.