

Computerized Assessment of Breast Lesion Malignancy using DCE-MRI:

Robustness Study on Two Independent Clinical Datasets from Two Manufacturers

Weijie Chen, PhD, Maryellen L. Giger, PhD, Gillian M. Newstead, MD, Ulrich Bick, MD, Sanaz A. Jansen, PhD, Hui Li, PhD, Li Lan, MSc

Rationale and Objectives: To conduct a preclinical evaluation of the robustness of our computerized system for breast lesion characterization on two breast magnetic resonance imaging (MRI) databases that were acquired using scanners from two different manufacturers.

Materials and Methods: Two clinical breast MRI databases were acquired from a Siemens scanner and a GE scanner, which shared similar imaging protocols and retrospectively collected under an institutional review board–approved protocol. In our computerized analysis system, after a breast lesion is identified by the radiologist, the computer performs automatic lesion segmentation and feature extraction and outputs an estimated probability of malignancy. We used a Bayesian neural network with automatic relevance determination for joint feature selection and classification. To evaluate the robustness of our classification system, we first used Database 1 for feature selection and classifier training, and Database 2 to test the trained classifier. Then, we exchanged the two datasets and repeated the process. Area under the receiver operating characteristic curve (AUC) was used as a performance figure of merit in the task of distinguishing between malignant and benign lesions.

Results: We obtained an AUC of 0.85 (approximate 95% confidence interval [CI] 0.79–0.91) for (a) feature selection and classifier training using Database 1 and testing on Database 2; and an AUC of 0.90 (approximate 95% CI 0.84–0.96) for (b) feature selection and classifier training using Database 2 and testing on Database 1. We failed to observe statistical significance for the difference AUC of 0.05 between the two database conditions ($P = .24$; 95% confidence interval $-0.03, 0.1$).

Conclusion: These results demonstrate the robustness of our computerized classification system in the task of distinguishing between malignant and benign breast lesions on dynamic contrast-enhanced (DCE) MRI images from two manufacturers. Our study showed the feasibility of developing a computerized classification system that is robust across different scanners.

Key Words: Computer-aided diagnosis; breast MRI; robustness.

©AUR, 2010

Acad Radiol 2010; 17:822–829

From the Department of Radiology, Committee on Medical Physics, The University of Chicago, 5841 South Maryland Avenue, MC-2026, Chicago, IL 60637 (W.C., M.L.G., G.M.N., U.B., S.A.J., H.L., L.L.); Division of Imaging and Applied Mathematics, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD (W.C., current address); Department of Radiology, Charité Campus Mitte, Berlin, Germany (U.B., current address). Received January 25, 2010; accepted March 9, 2010. Supported in part by NIH R33-113800, P50-CA125183, and DOE grant DE-FG02-08ER6478. M.L.G. is a stockholder in R2 Technology/Hologic and receives royalties from Hologic, GE Medical Systems, MEDIAN Technologies, Riverain Medical, Mitsubishi, and Toshiba. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interest that would reasonably appear to be directly and significantly affected by the research activities. Certain commercial materials and equipment are identified in order to adequately specify experimental procedures. In no case does such identification imply recommendation or endorsement by the Food and Drug Administration, nor does it imply that the items identified are necessarily the best available for the purpose. **Address correspondence to:** M.L.G. e-mail: m-giger@uchicago.edu

©AUR, 2010

doi:10.1016/j.acra.2010.03.007

Magnetic resonance imaging (MRI) is being increasingly used in clinical practice for the detection and characterization of breast cancer (1–6). The American College of Radiology Imaging Network Trial 6667 Investigators Group has recently reported that MRI can detect cancer that is missed by mammography and clinical examination at the time of the initial breast cancer diagnosis (3). In 2007, the American Cancer Society published guidelines for breast screening with MRI as an adjunct to mammography (4) in high-risk women. The expanding clinical applications of breast MRI call for developments of advanced computer tools to assist the radiologists in image interpretation and patient workup. First, computer-aided analysis may improve diagnostic accuracy as well as reduce intra- and inter-observer variability as it does in traditional modalities (7–10); there is evidence that computerized analysis of breast MRIs complements clinical reading (11). Furthermore, a typical breast MRI study acquires a substantial amount of four-dimensional data, and navigation and interpretation of these large datasets is time-consuming and even

challenging for radiologists (12,13). These reasons have motivated us to investigate computerized techniques for breast MRI analysis (14–17).

Historically, researchers and clinicians started performing breast MRI at the two ends of the spectrum of imaging techniques: one used high temporal resolution techniques attempting to distinguish between benign and malignant lesions by enhancement characteristics (called the “dynamic school” in a review by Kuhl et al) (18), whereas another (called the “static school”) (18) used high spatial resolution techniques attempting to distinguish between benign and malignant lesions by characteristic morphological features. This disparity was mainly from technical limitations and tradeoffs between spatial and temporal resolution taken at different institutions. Correspondingly, computerized breast lesion characterization in MRIs has been an active area of research for many years (19–33) and many of these studies have focused on some particular characteristic features of breast lesions on MRIs. Recent multicenter studies (34) and clinical consensus (35,36) have shown that both morphological and kinetic features are important and should be combined in breast MRI lesion characterization. Consensus has also been reached, in principle, that breast MRIs should be obtained with both fairly high spatial resolution and fairly high temporal resolution, which is achievable now because of technical progress in recent years (1). However, the image acquisition protocol for breast MRI is not yet standardized across manufacturers and institutions (1). This fact raises the question about the robustness of a computerized image analyses to different MRI scanners or acquisition parameters—a question that is insufficiently studied in the current literature.

In this research, we conducted a preclinical evaluation of our computerized system, which takes a suspicious breast lesion identified by radiologists as input and performs automatic lesion segmentation and characterization to aid in the discrimination between malignant and benign lesions. Our purpose is to conduct a preclinical evaluation of the robustness of our computerized system for breast lesion characterization on two breast MRI databases that were obtained from two different MRI scanners.

MATERIALS AND METHODS

Breast MRI Databases

Two breast MRI databases were retrospectively collected under a protocol approved by the Institutional Review Board for this Health Insurance Portability and Accountability Act–compliant study. Informed consent was waived. The two databases share similar imaging protocols with slightly different imaging acquisition parameters as summarized in Table 1.

Database 1 consisted of 121 primary lesions (mass or non-mass, focus lesions excluded) from 121 patients imaged on a 1.5T Siemens system (Siemens Vision, Siemens, Erlangen, Germany). Seventy-seven lesions were classified as malignant

and 44 lesions as benign based on retrospective review of final pathology reports. The average lesion size was $4.1 \pm 0.7 \text{ cm}^3$ (range, 0.03–78 cm^3 ; median, 2.0 cm^3). The average age of the patients was 51 years (range, 21–85; median, 50). Database 2 consisted of 181 primary lesions (mass or nonmass, focus lesions excluded) from 161 patients imaged on a 1.5T General Electronic (GE) system (GE Medical Systems, Milwaukee, WI). Of these, 97 lesions were classified as malignant and 84 lesions as benign. The average lesion size was $3.4 \pm 0.4 \text{ cm}^3$ (range, 0.09–27 cm^3 ; median, 1.2 cm^3). The average age of the patients was 52 years (range, 24–87; median, 52). The distribution of lesions by pathologic subtype is shown for both systems in Figure 1. In both databases, fibroadenomas were the most common benign lesion, whereas invasive ductal carcinomas composed the majority of malignant lesions. The difference between the average lesion sizes was not found to be statistically different (two-tailed *P* value of .34 in unpaired *t*-test).

Computerized Analysis of Breast Lesions on MRI images

We have developed a computerized breast MRI analysis system that aims to characterize a suspicious breast lesion as benign or malignant. After a breast lesion is identified by the radiologist on MRIs, the computer performs automatic lesion segmentation and analysis and outputs an estimated probability of malignancy (Fig 2). In our method, the suspicious lesion is first automatically segmented in three dimensions. Then, the characteristic kinetic curve (CKC) is automatically identified from the four-dimensional lesion data (ie, the time course of the three-dimensional [3D] lesion). Computerized algorithms are subsequently used to extract image features (ie, numerical mathematical descriptors) that are useful in the task of distinguishing between malignant and benign lesions. Kinetic features are derived from the computer-identified CKC, morphological features are derived from the 3D lesion segmentation, and texture features are derived from the 3D lesion at the first postcontrast time frame. Finally, a subset of selected features is used to train a neural network that outputs an estimated probability of malignancy. Details of these various stages within the breast MRI computer-aided diagnosis workstation are expanded in the next paragraphs.

The breast lesions are automatically segmented in 3D using a fuzzy-*c* means (FCM)–based approach, the details on which we have reported previously (15). As an initial step, a box-shaped volume of interest containing the suspicious lesion is selected by a human operator. The postcontrast series of the volume of interest are normalized voxel-by-voxel by the precontrast intensities. FCM clustering is then applied to partition the voxels in the volume of interest into two classes, lesion and non-lesion, using the relative-enhancement values of the postcontrast series as a feature vector. Subsequently, a connected-component labeling operation in terms of 3D connectivity is performed to remove the spurious and noisy

TABLE 1. Breast Magnetic Resonance Imaging Acquisition Protocols

	Database 1	Database 2
Scanner	1.5T Siemens vision	1.5T GE Medical system
Sequence	T1-weighted three-dimensional spoiled gradient echo sequence	
Repetition time	8.1 ms	7.7 ms
Echo time	4 ms	4.2 ms
Fat suppression	No	No
Acquisition matrix	256 × 128	256 × 128
Temporal resolution	69 seconds	68 seconds
Planar spatial resolution	1.25 mm × 1.25 mm	1.25 mm × 1.25 mm–1.6 mm × 1.6 mm
Slice thickness	2–3 mm	3–4 mm
View, number of slices	Coronal, 64 slices	Coronal, 60 slices
Gadolinium DiethyleneTriamine Pentaacetic Acid (Gd-DTPA) dose	0.2 mmol/kg	Fixed 20 mL of 0.5 mmol/mL
Flow rate		2 mL/second
Saline flush		20 mL
Number of lesions	77 malignant, 44 benign	97 malignant, 84 benign

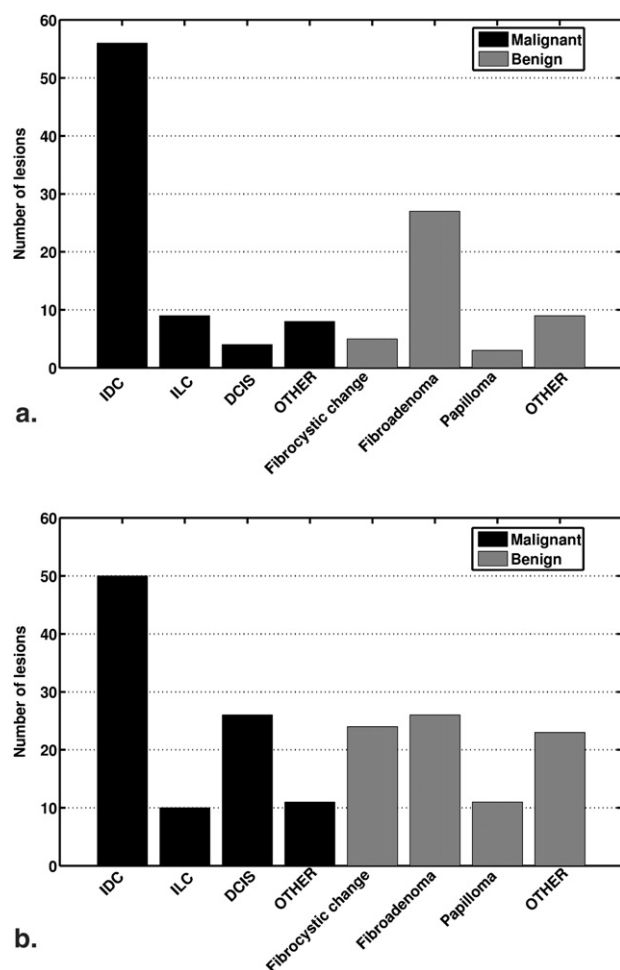


Figure 1. Distribution of patients over their primary lesion pathology for (a) Database 1 and (b) Database 2. IDC: invasive ductal carcinoma; ILC: infiltrating lobular carcinoma; DCIS: ductal carcinoma in situ.

structures and yield a 3D lesion object. A hole-filling operation is performed on the 3D lesion object to include those lesion voxels with necrosis-induced or noise-related low

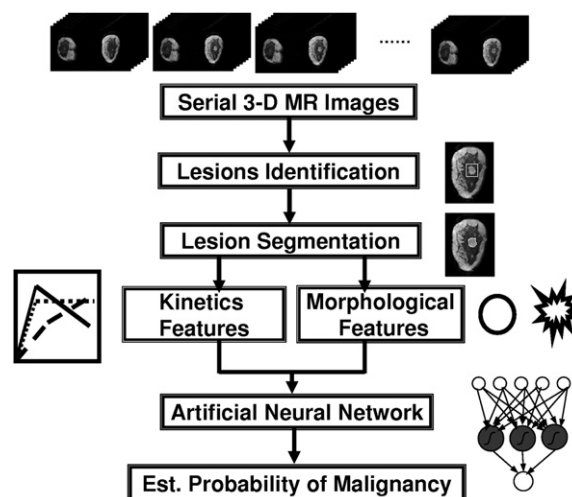


Figure 2. Schematic diagram of our computerized analysis and interpretation scheme. 3-D MR: three-dimensional magnetic resonance.

enhancement. Lesion segmentation by this computerized approach has been found to be highly correlated with the manual delineation by an experienced breast MRI radiologist (15).

The next key step in our approach is to identify the representative kinetic curve of the segmented lesion. A kinetic curve that is generated by simple averaging over the voxels within the entire lesion is suboptimal to characterize the lesion due to inhomogeneity of the contrast agent uptake within the lesion. In the clinical practice, the curve is usually generated from a small region of interest drawn manually by clinicians. Visually searching and manually outlining the most-enhancing region of interest from four-dimensional data is a time-consuming procedure and is affected by inter- and intra-observer variability. Thus, we had developed an efficient and automatic approach for this curve identification task (16).

In our approach, the FCM clustering technique is employed to partition the voxels within the lesion into a number of subsets with each subset having kinetic curves similar to a prototype kinetic curve. The prototype curves are compared and the curve that is most-enhancing at the first postcontrast timeframe is selected as the CKC of the lesion. Kinetic features including the uptake rate, washout rate, time to peak, and maximum uptake are extracted from the CKCs. In the task of distinguishing between benign and malignant lesions, the performance of the kinetic features based on the FCM-identified CKCs was found to be better than that of the features based on the curves obtained by averaging over the entire lesion, and similar to the performance of the features based on kinetic curves generated from regions drawn within the lesion by an experienced radiologist (16).

We also investigated computerized methods for quantifying the morphological and textural characteristics of breast lesions on MRIs. Investigation of shape and margin features is motivated by the clinical observation that benign lesions tend to have smooth or lobulated borders, whereas spiculated and irregular margins are indicators of malignancy. A prior study from our group has demonstrated that 3D analyses are more advantageous than two-dimensional analyses for shape and margin features (20).

Investigation of texture features is motivated by the clinical observation that the spatial pattern of contrast enhancement of a malignant lesion tends to be more inhomogeneous than that of a benign lesion. To further quantify the spatial properties of enhancement in breast lesions on MRIs, we extended the traditional two-dimensional gray-level co-occurrence matrix method to a volumetric texture analysis approach, which allows extraction of texture features from 3D segmented lesions (17). Compared to the traditional texture feature extraction method that calculates the features from a rectangular region of interest containing the lesion, our approach has two advantages: it improves the counting statistics by using the 3D information and it renders the texture characterization more accurate by excluding the influence of the normal structure that would be included by a regular-shaped region of interest. Our results showed that our volumetric approach significantly improved the performance of texture features, as compared to the traditional two-dimensional approach, in the task of distinguishing between malignant and benign lesions (17).

Additionally, spatial-temporal features were extracted to describe the kinetics of spatial variance of the enhancement within the 3D lesion (14).

The qualitative description of all these features (mathematical descriptors) is summarized in Table 2.

Computerized Classification of Breast Lesions on MRIs

Computerized classification is performed using an automatic classifier to combine multiple features into a numerical index (ie, an estimated probability of malignancy). A feature selection step is needed to reduce the dimensionality of the feature space and hence reduce the complexity of the classifier. In this

task, we used a Bayesian neural network (BNN) with automatic relevance determination (ARD) priors for joint feature selection and classification. BNN with ARD priors has been investigated and described in the literature (37–40) and its use in medical imaging has been demonstrated (41). In this approach, a hierarchical model is designed to use a group of hyperparameters to control the magnitude of the neural network weight parameters. In particular, each feature has a hyperparameter that indicates the relevance of this feature to the classifier output—the larger the hyperparameter, the less relevance of the feature.

A subset of the most relevant features can then be used to train the neural network to estimate the probability of malignancy, which evaluated by receiver operating characteristic analysis. The area under the receiver operating characteristic curve, or AUC, is used as a performance figure of merit in the task of distinguishing between malignant and benign breast lesions. ROCKIT software (42) was used in our analyses.

To evaluate the robustness of our classification system, we first used Database 1 (DB1) for feature selection and classifier training, and used Database 2 (DB2) as a test dataset to assess the performance of the trained classifier. Then, we exchanged the training/test role of the two datasets and repeated the process. The repetition is an independent and identical procedure except switching the role of the two datasets—that is, the results in the first training/test procedure were not employed in any way (eg, adjusting classifier parameters) in the repeated procedure.

RESULTS

We used the BNN with ARD priors to determine the relevance of the 28 features under investigation. For clarity, Figure 3a shows the relative importance of the top 14 features as determined by training the BNN with ARD priors using DB1. Figure 3b shows the corresponding results for DB2. Between the two databases (DB1 and DB2), 8 features were commonly ranked as among the top 14 features. Although the kinetic feature (F_{k2}) is ranked as the most relevant feature by both datasets, other features are not ranked in the exact same order.

We then examined the classification performance of BNN using the top n ($n = 2, 3, \dots, 28$) features as input. The following results are shown in Figure 4: (a) feature selection and classifier training using DB1 and testing the classifier on DB2; (b) feature selection and classifier training using DB2 and testing the classifier on DB1; (c) feature selection and classifier training using DB1 and testing the classifier on DB1; and (d) feature selection and classifier training using DB2 and testing the classifier on DB2. The results in (a) and (b) represent the assessment on independent datasets whereas the results in (c) and (d) represent resubstitution results that are optimistically biased. It is observed from Figure 4 that when the number of features is higher than

TABLE 2. Summary of Computerized Features in the Classification of Breast Lesions on Dynamic Contrast-Enhanced Magnetic Resonance Imaging

Feature		Description	
F _{k1}	Max. enh.	Maximum contrast enhancement	–
F _{k2}	Time to peak	Time frame at which the maximum enhancement occurs	Small
F _{k3}	Uptake rate	Uptake speed of the contrast enhancement	Large
F _{k4}	Washout rate	Washout speed of the contrast enhancement	Large
F _{k5}	Curve shape index	Difference between late and early enhancement	Small
F _{k6}	Max. var. of enh.	Maximum variance of contrast enhancement	–
F _{k7}	Time to peak	Time frame at which the maximum variance occurs	Small
F _{k8}	Var. incr. rate	Increasing speed of the enhancement-variance	Large
F _{k9}	Var. dec. rate	Decreasing speed of the enhancement-variance	Large
F _{m1}	Energy	Local image homogeneity	Small
F _{m2}	Contrast	Local image variations	Large
F _{m3}	Correlation	Image linearity	Large
F _{m4}	Variance homogeneity	How spread out the gray-level distribution is	Large
F _{m5}	Inverse difference moment	Local image homogeneity	Small
F _{m6}	Sum average	The overall brightness (average gray level)	Large
F _{m7}	Sum variance	How spread out the distribution of the sum of the gray levels of voxel-pairs is	Large
F _{m8}	Sum entropy	The randomness of the sum of gray levels of neighboring voxels	Large
F _{m9}	Entropy	The randomness of the gray levels	Large
F _{m10}	Difference variance	Variations of difference of gray levels between voxel-pairs	Large
F _{m11}	Difference entropy	The randomness of the difference of neighboring voxels' gray levels	Large
F _{m12}	IMC1	Nonlinear gray-level dependence	Small
F _{m13}	IMC2	Nonlinear gray-level dependence	Large
F _{m14}	Max corr. coeff	Nonlinear gray-level dependence	Large
F _{m15}	Circularity	Similarity of lesion shape to an effective sphere	Small
F _{m16}	Irregularity	Deviation of three-dimensional lesion surface from a sphere surface	Large
F _{m17}	Margin sharpness	Mean of the image gradient at the lesion margin	Small
F _{m18}	Var. of Margin sharpness	Variance of the image gradient at the lesion margin	Small
F _{m19}	Variance of RGH	Indicates how well the enhancement structures in a VOI extend in a radial pattern originating from the center of the VOI	Small

“small” (“large”): malignant lesions tend to have smaller (larger) feature values than do benign ones; “–”: failing to find significant difference of feature values between malignant and benign lesions; IMC, information measure of correlation; RGH, radial gradient histogram; VOI, volume of interest.

10, the AUC achieves a plateau for all four conditions with average AUC values of 0.84, 0.90, 0.95, and 0.92, for (a), (b), (c), and (d), respectively.

We then used ROCKIT to perform a statistical comparison of the two “independent assessment” conditions (a) and (b) when the top 14 features were used and the results are plotted in Figure 5. We obtained an AUC of 0.85 for condition (a) and an AUC of 0.90 for condition (b). We failed to observe statistical significance for the difference AUC of 0.05 between the two conditions ($P = .24$; 95% confidence interval $[-0.03, 0.1]$).

DISCUSSION

Our results demonstrate the robustness of our computerized classification system in the task of distinguishing between malignant and benign breast lesions on DCE-MRI images. Both independent assessments—that is, training on one distinct database and evaluating performance on another—yielded fairly high classification performances, which shows no statistical significance between them. The two datasets

were obtained in two institutions on scanners from two different manufacturers. The two scanners used a similar MRI protocol with slightly different image acquisition parameters and different contrast agent dosage. The significance of the study showed the feasibility of developing a computerized classification system that is robust across different scanners.

It should be noted that the failure to observe statistically significant difference does not mean that the difference may not actually exist (43). The retrospective analysis of the two existing datasets may have limited the ability to demonstrate the robustness of our technique because the different mix of lesion subtypes in the two datasets (Fig 1) may have contributed to the observed performance difference. Despite this limitation, the results still show encouraging degree of robustness. Although the conclusion may not be directly generalizable to all MRI protocols in current clinical practice, the computer characterization architecture of the system may be applicable to other imaging protocols. For example, the classifier can be retrained for the system to be used on a different imaging protocol.

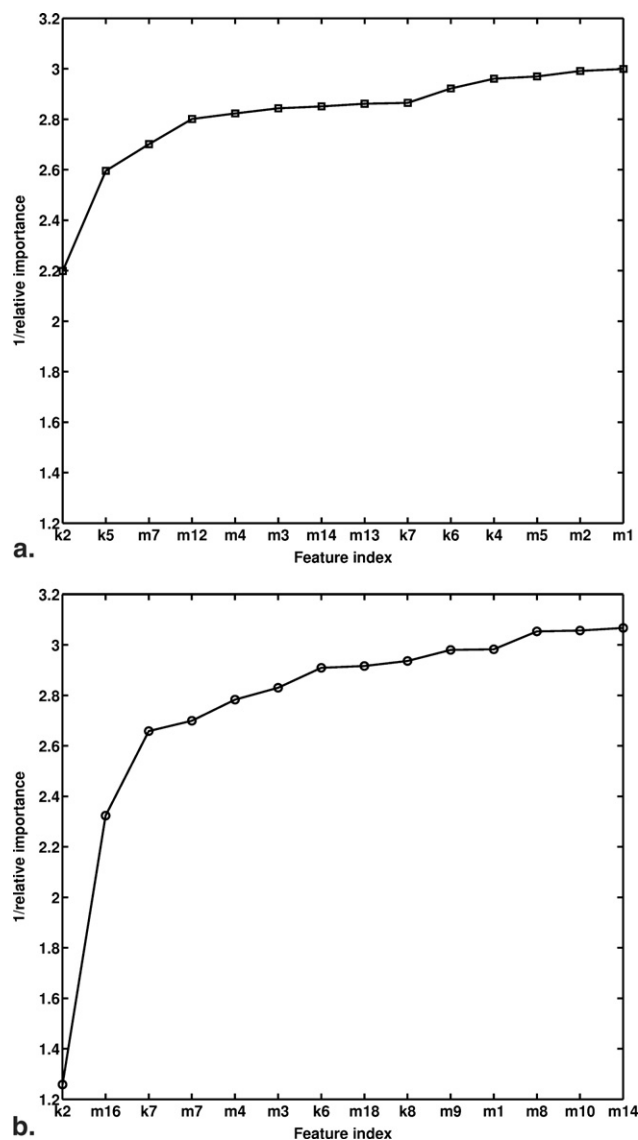


Figure 3. The relative importance of computerized image features to the classification task as assessed by the Bayesian neural network model with automatic relevance determination priors using (a) Database 1 and (b) Database 2. The results for the top-ranked 14 features are plotted. See Table 2 for the definition of the features.

Indeed, image acquisition protocols in breast MRI are not standardized and optimal imaging parameters are still under investigation (1,44). This is probably the fundamental obstacle at present for the development of computerized systems because a computerized system is typically designed on a dataset available from a specific institution or manufacturer, and robustness concerns can always be raised when a newly designed protocol is introduced to clinical practice. Research on the standardization and optimization of breast MRI protocols using methods such as the task-based approach outlined elsewhere (45) are expected to facilitate the development of computerized analysis systems.

The imaging technique used in our study is similar to that launched by Kaiser et al (46), called “the archetype of dynamic

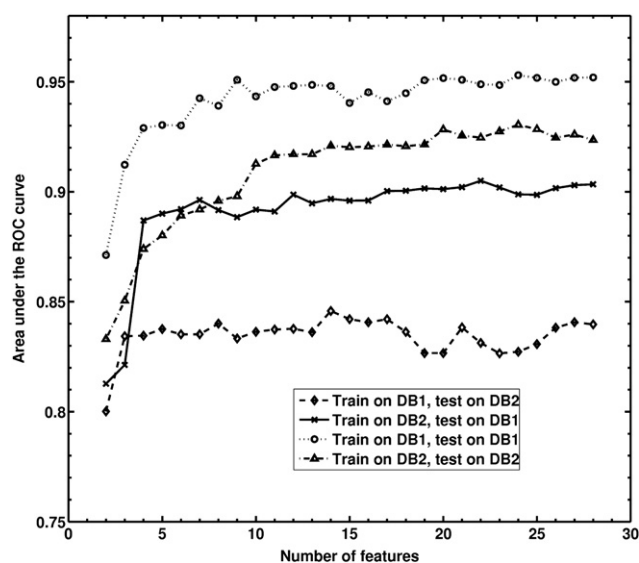


Figure 4. The classification performance of the computerized system as a function of the number of top-ranked features. Database 1: DB1; Database 2: DB2.

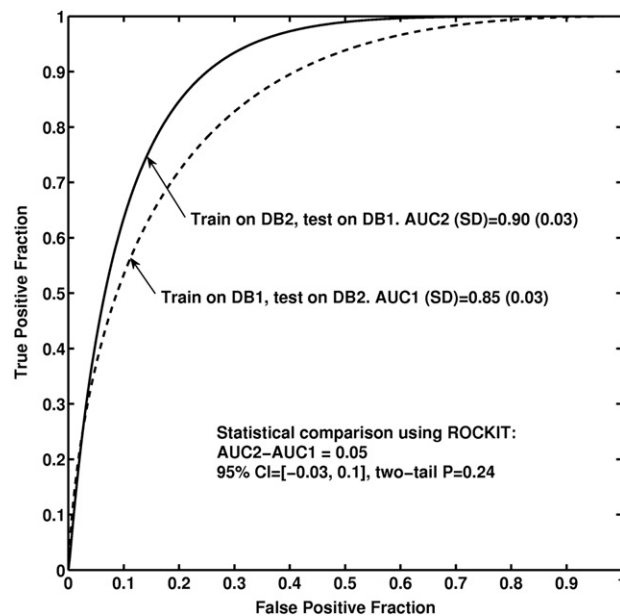


Figure 5. Receiver operating characteristic curves and statistical comparison of the two “independent assessment” conditions: (a) feature selection and classifier training using Database 1 (DB1) and testing the classifier on Database 2 (DB2) (dash); (b) feature selection and classifier training using DB2 and testing the classifier on DB1 (solid). AUC: area under the curve.

breast MRI” by Kuhl et al (18), which involves acquisition of one precontrast and a series of postcontrast images of both breasts at a temporal resolution of roughly 60 seconds. This type of breast MRI acquisition sequence has the advantage of being able to provide both morphological and kinetic information from one MRI examination, and was representative of early dynamic MRI protocols. However, because of

technical limitations, both the spatial resolution and the temporal resolution are moderate in this protocol. A recent study by Schnall et al (12) reported on a multicenter breast MRI study in which a “double exam” protocol was used. In their study, subjects were scanned with a contrast-enhanced 3D T1-weighted gradient echo sequence with high spatial resolution (minimum spatial resolution of $0.7 \times 1.4 \times 3$ mm). Subsequently, patients with suspicious findings were asked to return for a two-dimensional dynamic contrast-enhanced MRI with a temporal resolution of 15 seconds. The computerized morphological features and kinetic features that we have derived may have improved diagnostic performance when applied to MRIs with improved spatial resolution and improved temporal resolution.

The BNN model with ARD priors provides the flexibility for joint feature relevance determination and classification. However, an optimal subset of features may not be determined uniquely for a number of reasons. First, it is possible that different sets of features may yield similar classification performances on the population (ie, with infinite data to perform the feature selection). Second, because of the use of a finite size data sample, even if there exists a unique optimal subset of features for a population, a feature selector would select the optimal subset only with some certain probability less than 1 because of the uncertainty in the individual performance of each feature (47). Third, in practice, different datasets may not represent an actual random sample from the same population, thus increasing the variability in the feature selection results. The observed discrepancy between Figure 3a and Figure 3b may be due to all these reasons. A large dataset would help improve the consistency of feature selection.

In addition, the BNN model with ARD priors that we used only provided the relevance (or relative importance) of the features, with the number of features being chosen by the user. Without defining further objective criterion for this judgment, one cannot use the assessment results on clinical datasets such as those in Figure 4 to determine the exact number of features for practical applications. In this application, one may arbitrarily choose between 10 and 14, because, within this range, we observed no substantial decrease of performance due to increasing the complexity or reducing the feature set. Increasing the patient sample size would definitely reduce this ambiguity.

The output of the BNN is an estimate of the a posteriori probability of malignancy (48). It should be noted that the use of AUC is an assessment of the diagnostic accuracy or the *discrimination* ability of the BNN, but not an assessment of the absolute value or the *calibration* of the BNN output (49). In addition, for clinical use of such probability estimates, some adjustment of the BNN output may be needed to reflect the actual clinical prevalence of malignancy (50). The assessment of calibration and clinical adjustment are beyond the scope of this article.

Our datasets were collected from a consecutive series of patients with known breast lesions planned for open (surgical)

biopsy, which explains the relative high number of malignant cases compared to benign lesions. Although the focus of this article is on diagnostic workup applications, the techniques have potential to be extended to other clinical applications, such as classification of lesions beyond the two primary classes to multiple subtypes of cancers, high-risk screening, and preoperative staging.

Patient motion during image acquisition may introduce inaccuracies in kinetic features. Cases with abrupt and large patient movements between dynamic series had been clinically treated as acquisition failure and were excluded from our datasets. In our datasets, only patient respiratory motion was observed. The motion mostly resulted in additional blurring rather than actual displacement of image structure. However, it is important to note that image alignment of breast volumes at different time frames may improve the accuracy of our analyses.

In conclusion, in this preclinical evaluation, we have demonstrated the robustness of our computerized classification system across two different scanners in the task of distinguishing between malignant and benign breast lesions on dynamic contrast-enhanced MRIs. The system is potentially useful to aid radiologists in the characterization and interpretation of breast MRIs. For actual clinical applications, a clinical study for both stand-alone and observer performance evaluations are needed with a prospective design on a target population (screening or diagnosis) and a sufficiently large random sample of patients from that population.

REFERENCES

1. Kuhl CK. Current status of breast MR imaging. Part 1. Technical issues. *Radiology* 2007; 244:356–378.
2. Kuhl CK. Current status of breast MR imaging. Part 2. Clinical applications. *Radiology* 2007; 244:672–691.
3. Lehman CD, Gatsonis C, Kuhl CK, et al. ACRIN Trial 6667 Investigators Group. MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer. *N Engl J Med* 2007; 356:1295–1303.
4. Saslow D, Boetes C, Burke W, et al. American Cancer Society Breast Cancer Advisory Group. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007; 57:75–89.
5. DeMartini W, Lehman C, Partridge S. Breast MRI for cancer detection and characterization: a review of evidence-based clinical applications. *Acad Radiol* 2008; 15:408–416.
6. Eyal E, Degani H. Model-based and model-free parametric analysis of breast dynamic-contrast-enhanced MRI. *NMR Biomed* 2009; 22:40.
7. Huo Z, Giger ML, Vyborny CJ, et al. Breast cancer: effectiveness of computer-aided diagnosis observer study with independent database of mammograms. *Radiology* 2002; 224:560–568.
8. Chan HP, Sahiner B, Helvie MA, et al. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study. *Radiology* 1999; 212:817–827.
9. Jiang Y, Nishikawa RM, Schmidt RA, et al. Improving breast cancer diagnosis with computer-aided diagnosis. *Acad Radiol* 1999; 6:22–33.
10. Jiang Y, Nishikawa RM, Schmidt RA, et al. Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications. *Radiology* 2001; 220:787–794.
11. Deurloo EE, Muller SH, Peterse JL, et al. Clinically and mammographically occult breast lesions on MR images: potential effect of computerized assessment on clinical reading. *Radiology* 2005; 234:693–701.

12. Schnall MD. Clinical applications of image analysis and visualization for three dimensional breast imaging. In: Armato SG III, Brown MS, eds. Multi-dimensional image processing, analysis, and display: RSNA categorical course in diagnostic radiology physics. Oak Brook, IL: RSNA, 2005; 195-199.
13. Gilhuijs KG. Computer-supported analysis of breast MR imaging in four dimensions: why, when, and how. In: Armato SG III, Brown MS, eds. Multi-dimensional image processing, analysis, and display: RSNA categorical course in diagnostic radiology physics. Oak Brook, IL: RSNA, 2005; 201-213.
14. Chen W, Giger ML, Lan L, et al. Computerized interpretation of breast MRI: investigation of enhancement-variance dynamics. *Med Phys* 2004; 31: 1076-1082.
15. Chen W, Giger ML, Bick U. A fuzzy c-means (FCM) based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. *Acad Radiol* 2006; 16:63-72.
16. Chen W, Giger ML, Bick U, et al. Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI. *Med Phys* 2006; 33:2878-2887.
17. Chen W, Giger ML, Li H, Bick U, Newstead GM. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. *Magn Reson Med* 2007; 58:562-571.
18. Kuhl CK, Schild HH. Dynamic image interpretation of MRI of the breast. *J Magn Reson Imaging* 2000; 12:965-974.
19. Hayton P, Brady M, Tarassenko L, et al. Analysis of dynamic MR breast images using a model of contrast enhancement. *Med Image Anal* 1997; 1:207-224.
20. Gilhuijs KG, Giger ML, Bick U. Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. *Med Phys* 1998; 25:1647-1654.
21. Penn AI, Bolinger L, Schnall MD, et al. Discrimination of MR images of breast masses with fractal-interpolation function models. *Acad Radiol* 1999; 6:156-163.
22. Lucht RE, Knopp MV, Brix G. Classification of signal-time curves from dynamic MR mammography by neural networks. *Magn Reson Imaging* 2001; 19:51-57.
23. Gibbs P, Turnbull LW. Textural analysis of contrast-enhanced MR images of the breast. *Magn Reson Med* 2003; 50:92-98.
24. Szabó BK, Wiberg MK, Boné B, et al. Application of artificial neural networks to the analysis of dynamic MR imaging features of the breast. *Eur Radiol* 2004; 14:1217-1225.
25. Twellmann T, Lichte O, Nattkemper TW. An adaptive tissue characterization network for model-free visualization of dynamic contrast-enhanced magnetic resonance image data. *IEEE Trans Med Imag* 2005; 24: 1256-1266.
26. Penn AI, Thompson S, Brem R, et al. Morphologic blooming in breast MRI as a characterization of margin for discriminating benign from malignant lesions. *Acad Radiol* 2006; 13:1344-1354.
27. Schmid VJ, Whitcher B, Padhani AR, et al. Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. *IEEE Trans Med Imaging* 2006; 25:1627-1636.
28. Williams TC, DeMartini WB, Partridge SC, et al. Breast MR imaging: computer-aided evaluation program for discriminating benign from malignant lesions. *Radiology* 2007; 244:94-103.
29. Meinel LA, Stolpen AH, Berbaum KS, et al. Breast MRI lesion classification: Improved performance of human readers with a back propagation neural network computer-aided diagnosis (CAD) system. *J Magn Reson Imaging* 2007; 25:89-95.
30. Woods BJ, Clymer BD, Kurc T, et al. Malignant-lesion segmentation using 4D co-occurrence texture analysis applied to dynamic contrast-enhanced magnetic resonance breast image data. *J Magnet Reson Imaging* 2007; 25:495-501.
31. Levman J, Leung T, Causer P, et al. Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines. *IEEE Trans Med Imaging* 2008; 2:688-696.
32. Nie K, Chen JH, Yu HJ, et al. Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. *Acad Radiol* 2008; 15:1513-1525.
33. McLaren CE, Chen WP, Nie K, et al. Prediction of malignant breast lesions from MRI features: a comparison of artificial neural network and logistic regression techniques. *Acad Radiol* 2009; 16:842-851.
34. Schnall MD, Blume J, Bluemke DA, et al. Diagnostic architectural and dynamic features at breast MR imaging: multicenter study. *Radiology* 2006; 238:42-53.
35. Ikeda DM, Hylton NM, Kinkel K, et al. Development, standardization, and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance imaging studies. *J. Magn Reson Imaging* 2001; 13:889-895.
36. American College of Radiology. ACR BI-RADS: MRI. In ACR BIRADS: Breast Imaging Reporting and Data System: breast imaging atlas. Reston, VA, 2003.
37. MacKay DJS. Bayesian methods for adaptive models. PhD thesis, California Institute of Technology, Pasadena, CA, 1992.
38. Neal RM. Bayesian learning for neural networks. Lecture notes in statistics. New York: Springer-Verlag, 1996.
39. Bishop CM. Neural networks for pattern recognition. Oxford, UK: Oxford University Press, 1995.
40. Nabney T. Netlab: algorithms for pattern recognition. London, UK: Springer, 2002.
41. Chen W, Zur RM, Giger ML. Joint feature selection and classification using a Bayesian neural network with "automatic relevance determination" priors: potential use in CAD of medical imaging. *Proc SPIE, Med Imaging* 2007. Comp-aided Diagnosis 6514, 65141G.
42. Metz CE, Herman BA, Shen JH. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Stat Med* 1998; 17:1033-1053.
43. Metz CE. Quantification of failure to demonstrate statistical significance: the usefulness of confidence intervals. *Invest Radiol* 1993; 28:59-63.
44. Jansen SA, Shimauchi A, Zak L, et al. Kinetic curves of malignant lesions are not consistent across MR systems: the need for improved standardization of breast dynamic contrast-enhanced MRI acquisitions. *AJR Am J Roentgenol* 2009; 193:832-839.
45. Barrett HH, Myers KJ. Foundations of image science. Hoboken, NJ: Wiley-Interscience, 2003.
46. Kaiser WA, Zeitler E. MR imaging of the breast: fast imaging sequence with and without Gd-DTPA. *Radiology* 1989; 170:681-686.
47. Kupinski MA, Giger ML. Feature selection with limited datasets. *Med Phys* 1999; 26:2176-2182.
48. Kupinski MA, Edwards DC, Giger ML, et al. Ideal observer approximation using Bayesian classification neural networks. *IEEE Trans Med Imaging* 2001; 20:886-899.
49. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; 115:928-935.
50. Horsch K, Giger ML, Metz CE. Prevalence scaling: applications to an intelligent workstation for the diagnosis of breast cancer. *Acad Radiol* 2008; 15:1446-1457.