Published in final edited form as:

Acad Radiol. 2010 September ; 17(9): 1158–1167. doi:10.1016/j.acra.2010.04.015.

Multimodality Computer-Aided Breast Cancer Diagnosis with FFDM and DCE-MRI

Yading Yuan, BS,

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, IL 60637

Maryellen L. Giger, PhD,

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, IL 60637

Hui Li, PhD, Neha Bhooshan, MS, and

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, IL 60637

Charlene A. Sennett, MD

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, IL 60637

Abstract

Rationale and Objectives—To investigate a multimodality computer-aided diagnosis (CAD) scheme that combines image information from full-field digital mammography (FFDM) and dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) for computerized breast cancer classification.

Materials and Methods—From a retrospective FFDM database with 432 lesions (255 malignant, 177 benign) and a retrospective DCE-MRI database including 476 lesions (347 malignant, 129 benign), we constructed a multimodality dataset of 213 lesions (168 malignant, 45 benign). Each lesion was present on both FFDM and DCE-MRI images and deemed to be a difficult case given the necessity of having both clinical imaging exams. Using a manually indicated lesion location (ie, a seed point on FFDM images or a region of interest on DCE-MRI images, the computer automatically segmented the mass lesions and extracted lesion features). A subset of features was selected using linear stepwise feature selection and merged by a Bayesian artificial neural network to yield an estimate of the probability of malignancy. Receiver operating characteristic (ROC) analysis was used to evaluate the performance of the selected features in distinguishing between malignant and benign lesions.

Results—With leave-one-lesion-out cross-validation on the multimodality dataset, the mammography-only features yielded an area under the ROC curve (AUC) of 0.74 ± 0.04 , and the DCE-MRI-only features yielded an AUC of 0.78 ± 0.04 . The combination of these two modalities,

Address correspondence to: Y.Y. yading@uchicago.edu.

It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interest with would reasonably appear to be directly and significantly affected by the research activities.

which included a spiculation feature from mammography and two kinetic features from DCE-MRI, yielded an AUC of 0.87 ± 0.03 . The improvement of combining multimodality information was statistically significant as compared to the use of single modality information alone.

Conclusions—A CAD scheme that combines features extracted from FFDM and DCE-MRI images may be advantageous to single-modality CAD in the task of differentiating between malignant and benign lesions.

Keywords

Breast cancer; computer-aided diagnosis; multimodality imaging; full-field digital mammography; dynamic contrast enhanced magnetic resonance imaging

Breast cancer is a leading cause of mortality in American women, with an estimated 192,370 new cancer cases and 40,170 deaths in the United States in 2009 (1). Although there are limited methods for curing breast cancer, recent statistics show that there has been a steady decrease in the annual death rate from breast cancer among women, from 32.69 in 1991 to 24.00 in 2005 (per 100,000 population). This decrease accounts for nearly 40% of decreases in cancer death rates for women and largely reflects improvements in early detection and treatment (1).

Medical imaging plays a crucial role in reducing breast cancer mortality, with contributions to early detection through screening, diagnosis, image-guided biopsy, treatment planning, and treatment response monitoring (2,3). As the primary imaging modality for early detection and diagnosis of breast cancer, mammography has achieved significant success and has reduced the mortality from breast cancer by 15%–35% (4,5). However, about 15%–20% of cancers are still missed, and 65%–85% of breast biopsies are performed on benign lesions (2,6–8). Consequently, complementary imaging modalities, such as breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) and breast sonography (ie, breast ultrasound imaging), are being investigated to improve the accuracy of breast cancer diagnosis.

DCE-MRI is regarded as the most sensitive additional imaging modality for supplementing mammography (2). Clinical studies have demonstrated its ability to provide accurate diagnosis, extent of disease and multicentricity, and the ability to detect mammographically occult cancers in the contralateral breast (9,10). Peters et al (11) conducted a meta-analysis to determine the diagnostic performance of DCE-MRI in patients with breast lesions. From 251 studies between January 1985 and March 2005, they analyzed 44 eligible studies and obtained an overall sensitivity of 0.9 and an overall specificity of 0.72. Although the sensitivity of breast MRI is encouraging, its specificity is relatively low and varies widely with both cancer prevalence in the studied population and the interpretative criteria used to differentiate malignant lesions from benign ones.

To aid radiologists in distinguishing between malignant and benign lesions, various investigators are developing computerized image analysis methods for characterization and diagnosis of lesions in breast images (12,13). Computer-aided diagnosis (CAD) was initially

introduced for mammography (14–16), and then extended to breast sonography (17–20) and breast MRI (21–23).

Although the results of CAD systems for single imaging modality are encouraging (16,20,24–30), merging information across different modalities is recently attracting more attention. Because different imaging modalities provide complementary information regarding lesions, combining information from two or more modalities may increase diagnostic accuracy. Several investigations have been conducted to combine information from mammography and sonography to improve the diagnostic accuracy of breast cancer. Some studies assessed the performance of the computerized multimodality (mammography and sonography) schemes alone (31,32), whereas other studies evaluated the influence of multimodality CAD systems on radiologists' diagnostic accuracy (33–35). However, to the best of our knowledge, very few studies have investigated combining information from mammography and DCE-MRI in distinguishing between malignant and benign lesions.

This study aims to investigate the potential of computerized methods that use computer-extracted features from both mammograms and DCE-MRI images in the characterization of breast lesions. The main aspect of the proposed multimodality CAD system includes automatic lesion segmentation, feature extraction, and estimation of probability of malignancy (PM) by a classifier that fuses multimodality information.

MATERIALS AND METHODS

Database

A full-field digital mammography (FFDM) database and a DCE-MRI database were used in this study. All images were acquired at the University of Chicago Medical Center between 2002 to 2006, and retrospectively collected under an institutional review board–approved protocol. Data collection and usage were compliant with Health Insurance Portability and Accountability Act regulations. Lesion pathological states were biopsy or aspiration proven, and an expert breast radiologist (C.A.S.) identified the corresponding physical lesions as seen on both modalities based on visual criterion and biopsy-proven reports.

The FFDM database consisted of 432 biopsy-proven mass lesions in 269 patients, of which 177 benign lesions can be seen on 559 mammographic images and 255 malignant lesions can be seen on 942 mammographic images. The number of images per lesion varied from 1 to 13, including both standard and special views. All the images were obtained from GE Senographe 2000D systems (GE Medical Systems, Milwaukee, WI) at 12-bit quantization with a pixel size of 100 μ m.

The DCE-MRI database consisted of 476 mass lesions in 412 patients, of which 129 lesions were benign and 347 lesions were malignant. Images were obtained using a T1-weighted three-dimensional (3D) spoiled gradient echo sequence (repetition time, 7.7 ms; echo time, 4.2 ms, flip angle = 30°). Fat suppression was not employed. The patients were scanned in a prone position using a standard double breast coil on a 1.5T whole-body GE Signa MRI system (GE Medical Systems). For all patients, the first contrast-enhanced images were acquired 20 seconds after IV injection of 20 mL of 0.5 M of gadodiamide (Omniscan, GE

Healthcare, Little Chalfont, UK) followed by a 20-mL saline flush at the rate of 2.0 mL/second. Three to five post-contrast series were then acquired at a time interval of 68 seconds. Each series contained 60 slices with an in-plane spatial resolution ranging from $1.25 \text{ mm} \pm 1.25 \text{ mm}$ to $2.08 \text{ mm} \pm 2.08 \text{ mm}$. Slice thickness varied from 2.0 mm to 5.0 mm depending on breast size.

Based on these two databases, we constructed a multimodality dataset of 213 lesions in 171 patients (age range, 24–88 years, average, 56), of which 45 lesions were benign and 168 lesions were malignant. Each lesion was present on both FFDM and DCE-MRI images (Fig 1). Benign lesions included fibrocystic change (6/45), fibroadenoma (16/45), papilloma (6/45), and other atypical benign disease (17/45). Malignant lesions included invasive ductal carcinoma (106/168), invasive lobular carcinoma (14/168), ductal carcinoma in situ (27/168), and other atypical malignant disease (21/168). Note that, by virtue of having both FFDM and DCE-MRI performed as part of the clinical exam, these lesions could be assumed to be difficult in terms of interpretation.

Computerized Classification Methods

Computerized classification methods for mammography (16,36–38) and DCE-MRI (21,23,28,39) have been described elsewhere and are only summarized here.

A dual-stage segmentation method (37), which involves an active contour model (40) that followed an initial segmentation using a radial gradient index–based method (41), was employed to automatically extract lesions from the normal breast tissue in FFDM. Four morphological features were extracted from the segmented lesions. Margin sharpness is obtained as the average of the gradient magnitude along the margin of the mass, average gray level is measured by averaging the gray level values of each pixel within the segmented lesion, lesion size is defined as the diameter of a circle yielding the same area as the segmented lesion, and contrast measures the difference of average gray levels between the segmented lesion and the surrounding parenchyma. Nine spiculation features were calculated from a spiculation measure that was based on the radial gradient of the pixels within the lesion and its local environment. In addition, a texture feature, which is the standard deviation of the gradient within a mass lesion, was used to quantify the heterogeneity of the lesion. The detailed descriptions of these features can be found elsewhere (16,36,38). Because the number of mammograms for each physical lesion varied, for each feature, the computer calculated its average over all the mammograms of that particular lesion.

For each 3D DCE-MRI scan, fuzzy c-mean (FCM) clustering–based method was used to automatically categorize the voxels in the selected region around the lesion into two classes (lesion and non-lesion). In the FCM method, the membership map was based on the voxel values at the various time points (39). Once segmented, morphological, spiculation, and kinetic features were extracted from 3D lesions. The morphological features included lesion size, circularity, irregularity, and two margin sharpness features (21). Circularity indicates the conformity of a lesion to a sphere, irregularity measures the roughness of the lesion surface, and the margin sharpness features quantify the voxel-value gradients and their variations along the margin of the lesion, respectively. The spiculation feature measures the

variance of radial gradient histogram obtained from the pixels within a box encompassing the suspect lesion (21). Five kinetic features were used to quantify the time course of signal intensity within the lesion and another four kinetic features quantify the time course of enhancement-variance over the lesion (23,28). The characteristic kinetic curve of a lesion was automatically identified by a second FCM clustering procedure. In this procedure, the signal-time curves obtained from each voxel within the segmented 3D lesion were categorized into a number of prototypic curves (classes), and the membership curve with the highest initial enhancement was selected as the representative characteristic curve of the lesion (28).

Linear stepwise feature selection with Wilks lambda criterion (42) was employed to select a subset of effective features for the task of distinguishing between benign and malignant lesions. The selected features were then merged using a n -5-1 Bayesian artificial neural network (BANN) (43,44) to yield an estimate of PM, where n is the number of selected features.

Performance Evaluation and Statistical Analysis

All the reported classification performance in this article were based on leave-one-lesion-out (LOLO) cross-validation. Ideally, feature selection should be conducted in an independent dataset to avoid bias in performance evaluation. Nevertheless, because of the limited dataset in this study, we incorporated feature selection into the LOLO procedure to reduce the bias to some extent (45). We employed linear stepwise feature selection at each round of LOLO cross-validation, and constructed a histogram of the frequency at which features were automatically selected. Those features that were selected most frequently served as input to the BANN classifier.

Receiver operating characteristic (ROC) analysis (46,47) was used to assess the performance of the outputs of BANN in the task of differentiating between benign and malignant lesions. The area under the maximum likelihood-estimated binormal ROC curve (AUC) was used as an index of performance. ROCKIT software (version 1.1 b, available at http://xray.bsd.uchicago.edu/krl/KRL_ROC/software_index6.htm) (48) was used to determine the P value of the difference between two AUC values, and the Holm t -test (49) for multiple tests of significance was employed to evaluate the statistical significance, with an overall α -level of 0.05.

RESULTS

Classification Performance of Single-Modality CAD

We first assessed the overall classification performance of each single-modality CAD method on the two single-modality databases. The selected feature subset with the entire FFDM database included three morphological and two spiculation features: lesion size, gray level, contrast, region of interest (ROI)-based full width at half maximum (FWHM) of radial gradient histogram, and margin-based FWHM. The LOLO cross-validation using a BANN to merge the selected features yielded an AUC of 0.75 ± 0.02 on the FFDM database. The selected feature subset with the entire DCE-MRI database included one morphological and

three kinetic features: lesion irregularity, peak location of enhancement dynamics, peak location of enhancement–variance dynamics, and maximum enhancement of enhancement–variance dynamics. The LOLO cross-validation using a BANN to merge the selected features yielded an AUC of 0.79 ± 0.02 on the DCE-MRI database.

Classification Performance of Single-modality CAD on the Multimodality Subset

We considered two scenarios when evaluating the classification performance of single modalities on the multimodality dataset. In the first scenario (scenario 1), for each imaging modality, the same features used in the previous section served as inputs to the BANN classifier. Recall that these features were selected based on the entire single-modality databases. During the LOLO cross-validation, the classifier training procedure used all the cases in the entire single-modality databases, except the testing one in the multimodality dataset. In this scenario, an AUC of 0.71 ± 0.04 was obtained with selected FFDM features and an AUC of 0.77 ± 0.04 was obtained with selected DCE-MRI features.

In the second scenario (scenario 2), both feature selection and BANN training were restricted to the multimodality dataset only. The selected FFDM features included one morphological feature (lesion size) and two spiculation features (ROI-based FWHM and lesion-based FWHM). The LOLO cross-validation using a BANN to merge the selected features yielded an AUC of 0.74 ± 0.04 . The selected DCE-MRI features included three features from enhancement dynamics (maximum enhancement, peak location, and shape index of enhancement dynamics), and one feature from enhancement-variance dynamics (maximum enhancement-variance). The LOLO cross-validation using a BANN to merge the selected features yielded an AUC of 0.78 ± 0.04 .

We failed to show statistically significant difference between these two scenarios. The AUC values based on FFDM features had a P value of .30 between Scenario 1 and 2, whereas the AUC values based on DCE-MRI features had a P value of .66. ROC curves resulting from these two scenarios are shown as Figure 2.

CLASSIFICATION PERFORMANCE OF MULTIMODALITY CAD

We evaluated the classification performance of the CAD method when multimodality features were available. Based on 15 FFDM features and 15 DCE-MRI features, the selected feature subset included a spiculation feature from FFDM (ROI-based FWHM) and two enhancement dynamic features from DCE-MRI (curve shape index and peak location). Figure 3 shows the relationship among these three features.

A BANN classifier merged the selected multimodality features and yielded an estimate of probability of malignancy. Figure 4 shows the PM distribution in a leave-one-lesion-out cross-validation. For comparison, we also showed the distributions of PM values resulting from selected single-modality features in scenario 2. The separation between malignant and benign lesions increased with multimodality features, indicating improved classification performance. The PM values from multimodality features yielded an AUC of 0.87 ± 0.03 in the task of distinguishing between malignant and benign lesions. ROC curves resulting from multimodality features and single-modality features are shown as Figure 5.

Table 1 shows the AUC values and the associated standard error (SE) of the BANN outputs using multimodality features and single-modality features only, respectively. Also shown are the 95% of confidence intervals (CI) of the difference of AUCs obtained from multimodality features (AUC_M) and single-modality features (AUC_S) (ie, $AUC = AUC_M - AUC_S$). The improvement of classification performance by using multimodality features was statistically significant compared to the use of single-modality features (overall significance level $\alpha^T = 0.05$) (49).

DISCUSSION

In this study, we investigated the performance of a CAD scheme with computerized features extracted from mammography alone, DCE-MRI alone, and the combination of these two modalities. Merging of the two modalities at this feature data level is preferred because of the different acquisition patient positions and the difficulty in integration of the actual multimodality image data. Our results demonstrate that combining information from mammography and DCE-MRI yielded significantly higher diagnostic accuracy, in terms of AUC values, than the use of single-modality information only, in the task of differentiating between malignant and benign lesions. The promising results suggest that a multimodality computerized analysis is advantageous in characterizing breast lesions as compared to traditional single-modality methods.

DCE-MRI can be used as a complementary imaging modality to mammography for problem solving in the clinical scenarios where mammographic findings are equivocal. Because different imaging modalities characterize different physical properties of breast lesions, the fusion of modalities has the potential to improve the diagnostic accuracy. Mammography represents the two-dimensional projection of the 3D distribution of x-ray attenuation of tissue. One of the strengths of mammography is its capability to delineate the details of mass lesions because of its relatively high spatial resolution (100 μ m in this study). DCE-MRI represents the 3D distribution of free induction decay signals of tissue. Besides the volumetric information in 3D space, DCE-MRI provides functional information that is indicative of increased vascular density and vascular permeability changes associated with angiogenesis (50). In the previous studies on single-modality CAD systems, spiculation and kinetic features have been justified as the best features when distinguishing malignant and benign lesions for mammography and DCE-MRI, respectively. By appropriately selecting and merging multimodality features, our computerized image analysis method successfully combined the strengths of these two imaging modalities, as demonstrated in Figure 6.

Feature selection plays an important role in our computerized characterization of breast lesions with multimodality features. Figure 7 shows the histogram of the frequency at which features were automatically selected at each step of the LOLO procedure.

In an attempt to optimize the classification performance of single-modality features on the multimodality dataset, we considered two scenarios in this study. On one hand, because a large training set usually improves the generalization performance of BANN classifiers, scenario 1 used the entire single-modality database for feature selection and BANN training. On the other hand, the multimodality CAD method was developed and evaluated on the

multimodality dataset. To conduct a fair comparison, we restricted the feature selection and BANN training of single-modality features within the multimodality dataset in scenario #2. The multimodality classification outperformed single-modality classification by using the same computerized approach in any of these two scenarios. It is evident from our results that merging features from different modalities is very promising in distinguishing between malignant and benign breast lesions.

There are two limitations in this study. First, the majority of the lesions (168 of 213) in the multimodality dataset were malignant. This imbalance is expected, because usually only those patients who have Breast Imaging Reporting and Data System scores of 3 or above are suggested to obtain DCE-MRI in clinical practice. We also observe a similar imbalance in the entire DCE-MRI database, in which 347 of 432 lesions were malignant. Although data imbalance is an interesting problem in machine learning because it may deteriorate the classification performance (51,52), we focus on the comparison between multimodality CAD and single-modality CAD in this study, and defer the investigation on how the imbalance data affect the performance of the proposed analysis in future work. The other limitation is that the evaluation of classification performance in this study was based on the computer output alone. However, our ultimate goal is to help radiologists in their diagnostic decision process; thus, in the future, we will conduct an observer study to investigate the effect of the proposed multimodality CADx system on radiologists' performance in discriminating malignant and benign lesions when both FFDM and DCE-MRI CADx are simultaneously available.

CONCLUSION

In this article, we have presented a multimodality computerized analysis method to characterize breast lesions on FFDM and DCE-MRI images, and compared its performance to the use of single-modality information alone. Our investigation indicates that multimodality CAD is a promising way to distinguish between malignant and benign lesions. With leave-one-lesion-out cross-validation, the automatically selected multimodality features, including a ROI-based FWHM from mammography and curve shape index and peak location of enhancement dynamic curve from DCE-MRI, yielded an AUC of 0.87 ± 0.03 . The improvement by using multimodality features was statistically significant as compared to the use of single modality alone.

ACKNOWLEDGMENT

The authors are grateful to Weijie Chen, PhD, for help with the DCE-MRI database. Supported in part by a US Army Breast Cancer Research Program (BCRP) Predoctoral Traineeship Award (W81XWH-06-1-0726), by a Lawrence H. Lanzl Graduate Student Fellowship in Medical Physics (Committee on Medical Physics, the University of Chicago), by Breast SPORE (P50-CA125183), by DOE grant (DE-FG02-08ER6478), and by Cancer Center Support Grant (5-P30CA14599). M.L. Giger is a stockholder in R2 Technology/Hologic and receives royalties from Hologic, GE Medical Systems, MEDIAN Technologies, Riverain Medical, Mitsubishi, and Toshiba.

REFERENCES

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2009. *CA Cancer J Clin.* 2009; 59:225–249. [PubMed: 19474385]

2. Heywang-Kobrunner, SH.; Dershaw, DD.; Schreer, I. Diagnostic breast imaging: mammography, sonography, magnetic resonance imaging, and interventional procedures. New York: Thieme Medical Publisher; 2001.
3. Kuhl CK. The current status of breast MR imaging: part II; clinical applications. *Radiology*. 2007; 244:672–691. [PubMed: 17709824]
4. Humphrey LL, Helfand M, Chan B, et al. Breast cancer screening: a summary of the evidence for the U.S. preventive services task force. *Ann Intern Med*. 2002; 137:347–360. [PubMed: 12204020]
5. Institute of Medicine. Saving women's lives: Integration and innovation: a framework for progress in early detection and diagnosis of breast cancer. Washington D National Academic Press; 2005.
6. Harms SE. Technical report of the international working group on breast MRI. *J Magn Reson Imaging*. 1999; 10:980–980. [PubMed: 10581514]
7. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. *Radiology*. 1992; 184:613–617. [PubMed: 1509041]
8. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammography by US radiologists: findings from a national sample. *Arch Intern Med*. 1996; 156:209–213. [PubMed: 8546556]
9. Lehman CD, Gatsonis C, Kuhl CK, et al. MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer. *N Engl J Med*. 2007; 356:1295–1303. [PubMed: 17392300]
10. Kuhl CK, Schrading S, Bieling HB, et al. MRI for diagnosis of pure ductal carcinoma in situ: A prospective observational study. *Lancet*. 2007; 370:485–492. [PubMed: 17693177]
11. Peters NH, Rinkes IHB, Zuithoff NP, et al. Meta-analysis of mr imaging in the diagnosis of breast lesions. *Radiology*. 2008; 246:116–124. [PubMed: 18024435]
12. Giger ML, Chan HP, Boone J. Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. *Med Phys*. 2008; 35:5799–5820. [PubMed: 19175137]
13. Elter M, Horsch A. CADx of mammographic masses and clutered micro-calcifications: a review. *Med Phys*. 2009; 36:2052–2068. [PubMed: 19610294]
14. Gale AG, Roebuck EJ, Riley P, et al. Computer aids to mammographic diagnosis. *Br J Radiol*. 1987; 60:887–891. [PubMed: 3311272]
15. Swett HA, Miller PA. ICON: a computer-based approach to differential diagnosis in radiology. *Radiology*. 1987; 163:555–558. [PubMed: 3550885]
16. Huo Z, Giger ML, Vyborny CJ, et al. Automated computerized classification of malignant and benign masses on digitized mammograms. *Acad Radiol*. 1998; 5:155–168. [PubMed: 9522881]
17. Garra BS, Krasner BH, Horii SC, et al. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. *Ultrason Imaging*. 1993; 15:267–285. [PubMed: 8171752]
18. Huber S, Delorme S, Knopp MV, et al. Breast tumors: computer-assisted quantitative assessment with color Doppler US. *Radiology*. 1994; 192:797–801. [PubMed: 8058950]
19. Giger ML, Al-Hallaq H, Huo Z, et al. Computerized analysis of lesions in ultrasound images of the breast. *Acad Radiol*. 1999; 6:665–674. [PubMed: 10894069]
20. Horsch K, Giger ML, Venta LA, et al. Computerized diagnosis of breast lesions on ultrasound. *Med Phys*. 2002; 29:157–164. [PubMed: 11865987]
21. Gilhuijs KGA, Giger ML, Bick U. Automated analysis of breast lesions in three dimensions using dynamic magnetic resonance imaging. *Med Phys*. 1998; 25:1647–1654. [PubMed: 9775369]
22. 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. [PubMed: 10898034]
23. Chen W, Giger ML, Lan L, et al. Computerized interpretation of breast MRI: investigation of enhancement-variance dynamics. *Med Phys*. 2004; 31:1076–1082. [PubMed: 15191295]
24. Sahiner B, Chan H, Petrick N, et al. Improvement of mammographic mass characterization using spiculation measures and morphological features. *Med Phys*. 2001; 28:1455–1465. [PubMed: 11488579]
25. Varela C, Timp S, Karssemeijer N. Use of border information in the 25. Classification of mammographic masses. *Phys Med Biol*. 2006; 51:425–441. [PubMed: 16394348]

26. Huang S, Chang R, Chen D, et al. Characterization of spiculation on ultrasound lesions. *IEEE Trans Med Imaging*. 2004; 23:111–121. [PubMed: 14719692]
27. Sahiner B, Chan H, Roubidoux MA, et al. Computerized characterization of breast masses on three-dimensional ultrasound volumes. *Med Phys*. 2004; 31:744–754. [PubMed: 15124991]
28. 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. [PubMed: 16964864]
29. Zheng, Y.; Baloch, B.; Englander, S., et al. Medical image computing and computer-assisted intervention. Vol. 4792. *MICCAI*; 2007. Segmentation and classification of breast tumor using dynamic contrast-enhanced MR images; p. 393–401.
30. 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. [PubMed: 17507720]
31. Drukker K, Horsch K, Giger ML. Multimodality computerized diagnosis of breast lesions using mammography and sonography. *Acad Radiol*. 2005; 12:970–979. [PubMed: 16087091]
32. Jesneck JL, Lo JY, Baker JA. Breast mass lesions: computer-aided diagnosis methods with mammographic and sonographic descriptors. *Radiology*. 2007; 244:390–398. [PubMed: 17562812]
33. Horsch K, Giger ML, Vyborny CJ, et al. Classification of breast lesions with multimodality computer-aided diagnosis: observer study results on an independent clinical data set. *Radiology*. 2006; 240:357–368. [PubMed: 16864666]
34. Horsch K, Giger ML, Metz CE. Potential effect of different radiologist reporting methods on studies showing benefit of CAD. *Acad Radiol*. 2008; 15:139–152. [PubMed: 18206613]
35. Sahiner B, Chan H-P, Hadjiiski LM, et al. Multimodality CADx: ROC study of the effect on radiologists' accuracy in characterizing breast masses on mammograms and 3D ultrasound images. *Acad Radiol*. 2009; 16:810–818. [PubMed: 19375953]
36. Li H, Giger ML, Yuan Y, et al. Evaluation of computer-aided diagnosis on a large clinical full-field digital mammographic dataset. *Acad Radiol*. 2008; 15:1437–1445. [PubMed: 18995194]
37. Yuan Y, Giger ML, Li H, et al. A dual-stage method for lesion segmentation on digital mammograms. *Med Phys*. 2007; 34:4180–4193. [PubMed: 18072482]
38. Huo Z, Giger ML, Vyborny CJ, et al. Analysis of spiculation in the computerized classification of mammographic masses. *Med Phys*. 1995; 22:1569–1579. [PubMed: 8551981]
39. Chen W, Giger ML, Bick U. A fuzzy c-mean (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. *Acad Radiol*. 2006; 13:63–72. [PubMed: 16399033]
40. Chan TF, Vese LA. Active contours without edges. *IEEE Trans Image Processing*. 2001; 10:266–277.
41. Kupinski MA, Giger ML. Automated seeded lesion segmentation on digital mammograms. *IEEE Trans Med Imaging*. 1998; 17:510–517. [PubMed: 9845307]
42. Lachenbruch, PA. Discriminant analysis. London, UK: Hafner; 1975.
43. Bishop, CM. Neural networks for pattern recognition. Oxford, UK: Oxford University; 1995.
44. Kupinski MA, Edwards DC, Giger ML, et al. Ideal observer approximation using Bayesian classification neural networks. *IEEE Trans Med Imaging*. 2001; 20:886–899. [PubMed: 11585206]
45. Li Q, Doi K. Reduction of bias and variance for evaluation of computer-aided diagnostic schemes. *Med Phys*. 2006; 33:868–875. [PubMed: 16696462]
46. Metz CE. ROC methodology in radiologic imaging. *Invest Radiol*. 1986; 21:720–733. [PubMed: 3095258]
47. Metz CE, Herman BA, Shen J. Maximum likelihood estimation of receiver operating characteristic ROC curves from continuously-distributed data. *Stat Med*. 1998; 17:1033–1053. [PubMed: 9612889]
48. Metz CE, Herman BA, Roe CA. Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets. *Med Decision Making*. 1998; 18:110–121.

49. Glantz, SA. Primer of biostatistics. New York: McGraw-Hill; 2002.
50. Kuhl CK, Schild HH. Dynamic imaging interpretation of MRI of the breast. J Magn Reson Imaging. 2000; 12:965–974. [PubMed: 11105038]
51. Japkowicz N, Stephen S. The class imbalance problem: a systematic study. Intell Data Anal. 2002; 6:429–449.
52. Weiss GM, Provost F. Learning when training data are costly: The effect of class distribution on tree induction. J Artificial Intell Res. 2003; 19:315–354.

Figure 1.

Example of a malignant lesion imaged by both mammography and dynamic contrast-enhance magnetic resonance imaging (DCE-MRI). (a) A mammographic region of interest (ROI) in CC view. *Left*: original image; *right*: image with computer-delineated contour superimposed. (b) The corresponding ROI in MLO view. *Left*: original image; *right*: image with computer-delineated contour superimposed. (c) The corresponding MRI. *Left*: a MRI slice containing the same mass lesion; *middle*: lesion with computer-delineated contour

superimposed; *right*: the computer-identified characteristic kinetic curve of the lesion.
MLO, medio-lateral oblique; CC, cranial-caudal.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2.

Receiver operating characteristic curves of each modality evaluated in two scenarios on (a) mammography alone and (b) dynamic contrast-enhance magnetic resonance imaging (DCE-MRI) alone. In scenario 1, both feature selection and classifier training were based on the entire single modality database, whereas in scenario 2, feature selection and classifier training were restricted within the multimodality dataset only. AUC: area under the curve.

Figure 3.

Scatter plots of multimodality features resulting from linear stepwise feature selection method: (a) curve shape index (MRI) vs. ROI-based spiculation (FFDM), (b) peak location (MRI) vs. ROI-based spiculation (FFDM), and (c) peak location (MRI) vs. curve shape index (MRI). MRI, magnetic resonance imaging; FFDM, full-field digital mammography; ROI, region of interest.

Figure 4.

The distributions of probability of malignancy (PM) values in leave-one-lesion-out (LOLO) cross-validation, when the input features to Bayesian artificial neural network were: (a) full-field digital mammography (FFDM) features alone, (b) dynamic contrast-enhance magnetic resonance imaging (DCE-MRI) features alone, and (c) FFDM and DCE-MRI features combined. Note that here only the PM values from the multimodality subset was used in the LOLO validation; thus, cancer prevalence is the same in all three figures.

Figure 5.

Receiver operating characteristic curves of computer-aided diagnosis method performed on full-field digital mammography (FFDM) features only (*dash line*), dynamic contrast-enhance magnetic resonance imaging (DCE-MRI) features only (*dot dash line*), and the combination of FFDM and DCE-MRI features (*solid line*).

Figure 6.

Classification results for a malignant lesion (*top*) and a benign lesion (*bottom*). The solid lines in the left two mammographic images are computer-delineated contours. The probability of malignancy (PM) values on the left was estimated by a Bayesian artificial neural network (BANN) based on mammographic features alone. The right column shows the corresponding magnetic resonance images and the computer-identified characteristic kinetic curves of these two lesions. PM values on the right were estimated by a BANN based on dynamic contrast-enhance magnetic resonance imaging (DCE-MRI) features alone. By

combining information from FFDM and DCE-MRI, the proposed multimodality computer-aided diagnosis increased the PM value of the malignant lesion to 0.97, and reduced the PM value of the benign lesion to 0.19.

Figure 7.

Histogram of the frequency at which features were automatically selected at each step of leave-one-lesion-out procedure ($n = 213$). FFDM, full-field digital mammography; DCE-MRI, dynamic contrast-enhance magnetic resonance imaging; ROI, region of interest.

Classification Performance of the Multimodality CAD Method in a Leave-one-lesion-out Cross-validation for the Task of Differentiating Malignant and Benign Lesions, and the Comparison with the Performance of Single-modality CAD

TABLE 1

Modality	$AUC_M \pm SE$	$AUC_S \pm SE$	P Value	Significance Level	95% CI of AUC
FFDM and DCE-MRI	0.87 ± 0.03	—	—	—	—
FFDM (scenario 1)	—	0.71 ± 0.04	$<10^{-4}$	0.0125	[0.09, 0.23]
FFDM (scenario 2)	—	0.74 ± 0.04	.002	0.0250	[0.05, 0.20]
DCE-MRI (scenario 1)	—	0.77 ± 0.04	.009	0.0500	[0.03, 0.15]
DCE-MRI (scenario 2)	—	0.78 ± 0.04	.004	0.0375	[0.02, 0.13]

CAD, computer-aided diagnosis; FFDM, full-field digital mammography; DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging.
The value after “ $\pm$ ” is the standard error (SE) associated with each AUC . The two-tailed P value and 95% CI of AUC were calculated by ROCKIT. The significance level column represents the significance level of individual tests adjusted with Holm t -test (overall significance level $\alpha^* = 0.05$).