

Characterization of mammographic masses based on level set segmentation with new image features and patient information

Jiazheng Shi,^{a)} Berkman Sahiner, Heang-Ping Chan, Jun Ge, Lubomir Hadjiiski, Mark A. Helvie, Alexis Nees, Yi-Ta Wu, Jun Wei, Chuan Zhou, Yiheng Zhang, and Jing Cui
Department of Radiology, The University of Michigan, Ann Arbor, Michigan 48109-0904

(Received 6 July 2007; revised 9 November 2007; accepted for publication 12 November 2007; published 21 December 2007)

Computer-aided diagnosis (CAD) for characterization of mammographic masses as malignant or benign has the potential to assist radiologists in reducing the biopsy rate without increasing false negatives. The purpose of this study was to develop an automated method for mammographic mass segmentation and explore new image based features in combination with patient information in order to improve the performance of mass characterization. The authors' previous CAD system, which used the active contour segmentation, and morphological, textural, and spiculation features, has achieved promising results in mass characterization. The new CAD system is based on the level set method and includes two new types of image features related to the presence of microcalcifications with the mass and abruptness of the mass margin, and patient age. A linear discriminant analysis (LDA) classifier with stepwise feature selection was used to merge the extracted features into a classification score. The classification accuracy was evaluated using the area under the receiver operating characteristic curve. The authors' primary data set consisted of 427 biopsy-proven masses (200 malignant and 227 benign) in 909 regions of interest (ROIs) (451 malignant and 458 benign) from multiple mammographic views. Leave-one-case-out resampling was used for training and testing. The new CAD system based on the level set segmentation and the new mammographic feature space achieved a view-based A_z value of 0.83 ± 0.01 . The improvement compared to the previous CAD system was statistically significant ($p=0.02$). When patient age was included in the new CAD system, view-based and case-based A_z values were 0.85 ± 0.01 and 0.87 ± 0.02 , respectively. The study also demonstrated the consistency of the newly developed CAD system by evaluating the statistics of the weights of the LDA classifiers in leave-one-case-out classification. Finally, an independent test on the publicly available digital database for screening mammography with 132 benign and 197 malignant ROIs containing masses achieved a view-based A_z value of 0.84 ± 0.02 . © 2008 American Association of Physicists in Medicine.
[DOI: 10.1118/1.2820630]

Key words: computer-aided diagnosis, mammography, breast masses, level set, segmentation, classification

I. INTRODUCTION

Breast cancer is the second leading cause of cancer death of American women between 40 and 55 years of age.¹ Early diagnosis and treatment can improve survival rate of breast cancer patients.²⁻⁴ Screening mammography is currently the most cost-effective method to detect early breast cancer.⁵⁻⁷ Breast masses are one of the important mammographic indicators of malignancy. Clinical studies show that only a minority of biopsied masses are malignant.⁸⁻¹⁰ It is therefore important to improve methods for diagnosis of malignancy and the positive predictive value of mammography.

Computer-aided diagnosis (CAD) systems can be used to assist radiologists in mammographic mass detection and classification.¹¹ Automated mass detection has been applied to digitized screen-film mammograms (SFM) and full field digital mammograms (FFDM) and achieved encouraging performance.¹² Several investigations have shown that CAD has the potential to improve radiologist's accuracy in assessing malignant and benign masses.^{11,13-18} This study focuses

on automated classification of malignant and benign masses. Our purpose is to develop a new segmentation algorithm and to design new mammographic features in combination with patient information for improving computerized characterization of mammographic masses.

Segmentation is a crucial step for computerized classification of malignant and benign breast masses. Image segmentation is a common step in the field of computer vision. Popular segmentation methods include local and global thresholding, region growing, deformable models, random fields, and edge detection.^{11,19} Segmentation of mammographic masses is more challenging than segmenting other natural or artificial images because breast masses on mammograms typically have fuzzy and irregular edges and low contrast. The American College of Radiology Breast Imaging Reporting and Data System²⁰ (BI-RADS) classifies breast mass margins into five major categories, of which only circumscribed mass margin might have an abrupt transition between the lesion and the surrounding tissue. To segment this type of medical images, researchers have developed or re-

fined several breast mass segmentation methods. Guliato *et al.* refined the region growing method using fuzzy set theory.²¹ The method considered the uncertainty of the boundary around the mass. Huo *et al.* also used the region growing method for mass segmentation. Region growing was terminated by using the multiple transition point technique.^{16,22} Timp and Karssemeijer proposed an automated segmentation technique based on dynamic programming.²³ Our previous investigations also developed a fully automated, two-step segmentation method.¹⁵ It used *K*-means clustering for initial segmentation, which was then refined by the active contour (AC) method.

Features can be extracted from original regions of interest (ROIs), segmented masses, and patient information. Most CAD systems in the literature explore mammographic features based on segmented masses. Useful features include: (1) statistics of mass intensity, e.g., average and variance; (2) morphological features, e.g., perimeter and area of masses, normalized radial length,²⁴ and the Fourier descriptors;²⁵ (3) textural features, e.g., run-length statistics (RLS) and spatial gray level dependence (SGLD);²⁶ and (4) features that can describe mass margins, such as spiculation measure.^{14,27}

Our previous CAD system, which used AC segmentation, and morphological, textural, and spiculation features, has shown encouraging results in classifying malignant and benign masses. In this study, we introduced the increasingly popular method, level set (LS) segmentation, and added two new types of image features related to the presence of microcalcifications within the mass and abruptness of the mass margin, in combination with patient age, to improve the CAD system. Breast masses with microcalcifications are more likely to be malignant. However, since the probability that a breast mass contains microcalcification is low, the microcalcification feature has not been extensively used for the purpose of mass classification. We used our previously developed automated algorithm²⁸ to detect microcalcifications within the ROI containing the mass and extracted a feature related to the likelihood of microcalcification presence in the ROI from our detection results. Circumscribed mass margin is a predominantly benign feature.²⁰ Although some of the features in our previous CAD system may be correlated with this characteristic, we designed new features specifically for characterizing the abruptness of the transition between the mass and the surrounding tissue in this study. Finally, patient age is an objective feature which can be collected conveniently.²⁴ In this study, we studied the usefulness of combining patient age with our image features for breast mass classification.

II. METHODS

Figure 1 presents the flowchart of our method for mammographic mass classification. It involves three major stages: mass segmentation, feature extraction, and classification. In this study, we combined stepwise feature selection and linear discriminant analysis (LDA) classification as a wrapper algorithm²⁹ during classifier training. Feature selection³⁰ is an important process during the development of the CAD

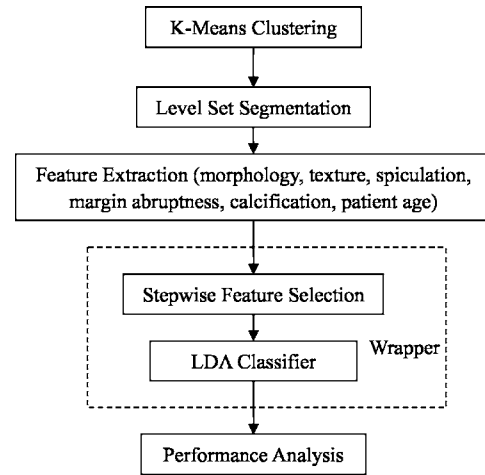


FIG. 1. The flowchart of the newly developed CAD system for mammographic mass classification.

system. However, once the system is developed, the features will be fixed when the system is applied to unknown test cases.

II.A. Data sets

Our primary data set was collected with Institutional Review Board (IRB) approval from patient files in the Radiology Department at the University of Michigan who had undergone breast biopsy. The mammograms were acquired with MinR2000 screen-film systems (Eastman Kodak, Rochester, NY), and were digitized with a LUMISCAN 85 laser film scanner (Lumisys, Los Altos, CA). The biopsied mass location was identified by a Mammography Quality Standard Act (MQSA) radiologist (M.A.H.) with 20 years of experience in mammography using available clinical information and a bounding box containing the mass was defined. The data set consisted of 427 biopsy-proven masses (200 malignant and 227 benign) in 909 ROIs (451 malignant and 458 benign) from multiple views including craniocaudal (CC), mediolateral oblique (MLO), and/or lateral.

The MQSA radiologist provided a likelihood of malignancy (LM) rating for each mass on each mammographic view on a scale of 1 (most benign appearance) to 10 (most malignant appearance). For most masses, the LM rating was provided in the same reading session as the identification of the biopsied mass location and bounding box. Because the identification procedure utilized available clinical information including pathology report, the radiologist was aware of the outcome of the biopsy. However, he was asked to make an assessment of malignancy based on the visual appearance of the mass on the mammogram. Figure 2(a) shows the distribution of the LM ratings of the masses. As expected from the fact that all cases underwent biopsy, the two distributions partially overlap. Figure 2(b) shows the distributions of the size of the malignant and benign masses. The size was measured as the longest dimension of the lesion by the radiologist. It ranged from 4 to 100 mm with a mean size of 17 mm. The distributions of the size of malignant and benign masses

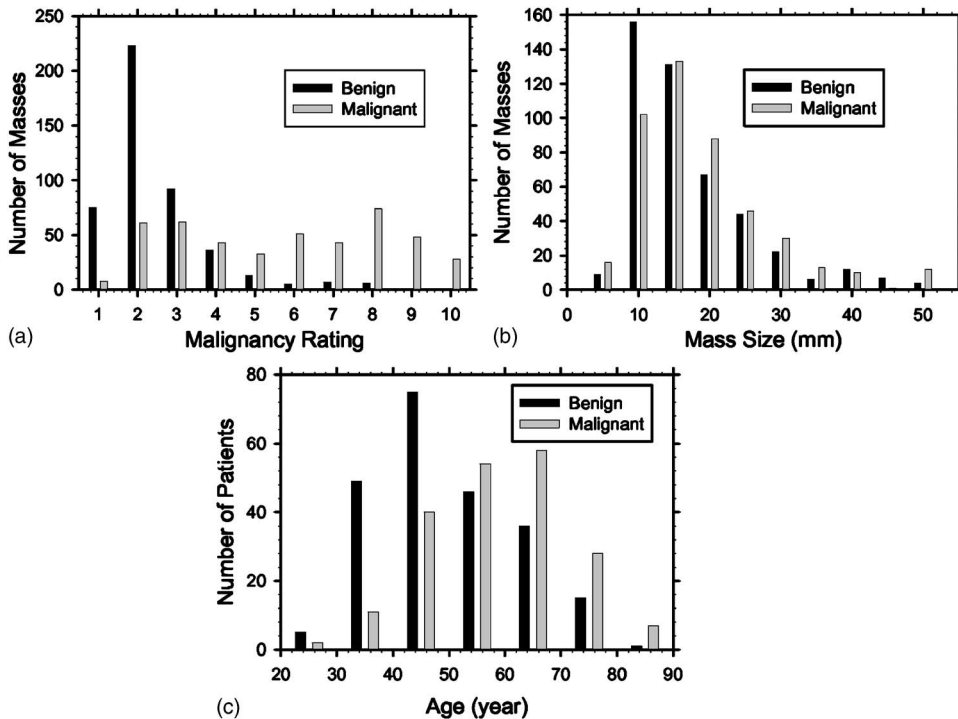


FIG. 2. Characteristics of the data set included in this study. (a) The distribution of the LM rating of the masses by the radiologist. (b) The distributions of mass size. To show the size for the majority of the masses in detail, 16 masses with sizes over 50 mm were grouped to the bar representing mass size=50 mm. (c) The distributions of ages for patients with malignant or benign masses.

are similar to each other. Figure 2(c) shows the distributions of patient age for the malignant and benign cases. Patient age ranged from 23 to 86 years (mean: 54 years). The age distributions for benign and malignant cases overlap to a great extent.

In order to evaluate the performance of the microcalcification likelihood feature, discussed below, a MQSA radiologist (A.N.) identified locations of microcalcification clusters on the digitized ROIs using a graphical user interface that allowed windowing, leveling, and magnification. Assessment of microcalcification presence was performed in an independent session. The radiologist was not informed of the biopsy outcome of the masses when the microcalcifications were marked. The radiologist was asked to judge whether clustered microcalcifications was present regardless of the appearance of the masses or the microcalcifications. Thirty-seven benign and 115 malignant ROIs contained microcalcification clusters, while the remaining 757 ROI were determined to be free of microcalcification clusters.

We used a second data set to independently test the performance of the newly developed CAD system. We previously evaluated³¹ the accuracy of a mass detection system on a subset of the digital database for screening mammography³² (DDSM) publicly available from the University of South Florida. The data set for that evaluation consisted of 142 two-view mammograms (CC and MLO) acquired between 1992 and 1998 and digitized with a Lumisys 200 laser film scanner. In this study, we used the same set of 142 mammogram pairs for our mass classification system. All images in this data set had a pixel size of 50 μm and 12-bit gray-level resolution. The data set consisted of 167 biopsy-proven masses (66 benign and 101 malignant) in 329

ROIs (132 benign and 197 malignant) from 136 patients. Patient age ranged from 31 to 86 years (mean: 56 years).

II.B. Initial mammographic mass segmentation

Mass segmentation was performed in a ROI containing the radiologist-identified mass. The ROI was selected to include the bounding box described above, plus a 5-mm-wide band surrounding the bounding box that contains breast tissue around the mass. The ROI was preprocessed by background gray-level correction, which was designed to reduce the nonuniformity caused by the overlapping breast structures and the location of the lesion on mammograms. We then used the K -means clustering algorithm followed by a morphological opening operation, for initial mass segmentation.³³

II.C. Level set segmentation

The LS method, also called implicit AC in the literature, provides a flexible mechanism allowing topological change of target objects, for example, object splitting and merging. Evolving curves in two-dimensional (2D) images or surfaces in three-dimensional (3D) volumes based on the LS numerical technique have been used in various applications of image processing, computer graphics, and computational geometry.^{34–36} For example, the LS method is used for restoring images degraded by noise and blurring or estimating object surface from given objects sample points. We have previously utilized the 3D LS method for segmenting head and neck lesions on CT scans.³⁷ In this study, the breast masses were collected from mammograms, so we limit our

discussion to 2D curve evolution. The method can be extended in a straightforward manner to 3D mass segmentation in emerging breast imaging modalities.

While the explicit AC model is a Lagrangian integral formation tracking the curve, the LS model is an Eulerian formulation capturing the curve.³⁴ It can work on a Cartesian grid without having to parameterize target object boundaries. Mathematically, let $\phi(x, y) = c$ be an isocontour in a 2D space where each location has a unique level c . The isocontour $\phi(x, y) = 0$, which encloses a region Ω , is specifically named as the zero LS in the literature. For the purpose of segmenting a mammographic mass ROI, we first need to construct the initial zero LS, which can be performed by using a general object model (e.g., a circle) or using the boundary of an initially segmented mass (e.g., the K -means clustering result). For iterative evolution, we incorporate the time parameter and track the zero LS. Mathematically, the evolution can be described as

$$\begin{aligned} \phi(x(t), y(t), t) &= 0, \\ \forall t &\geq 0, \end{aligned} \quad (1)$$

with the initial condition

$$\phi(x(0), y(0), 0) = \phi(x(0), y(0)).$$

Differentiating both sides of Eq. (1) and using the chain rule yields the partial differential equation (PDE)

$$\frac{\partial \phi}{\partial t} + \frac{\partial \phi}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial \phi}{\partial y} \frac{\partial y}{\partial t} = \frac{\partial \phi}{\partial t} + \vec{V} \cdot \nabla \phi = 0, \quad (2)$$

where $\nabla \phi = (\partial \phi / \partial x, \partial \phi / \partial y)$ is the gradient of the distance function and $\vec{V} = (\partial x / \partial t, \partial y / \partial t)$ is the velocity field that guides the curve evolution. The success of the LS segmentation is subject to appropriate design of \vec{V} . In analogy to the explicit AC model, we may incorporate internal forces to guarantee the smoothness of the curve and external forces to attract the curve to desired image edges. The mean curvature $\kappa = -\nabla \cdot \nabla \phi / |\nabla \phi|$ is a popular term³⁴ to constrain the curve evolution and the edge information $\vec{G} = \nabla I(x, y)$ of the original image $I(x, y)$ is used to construct external force. Based on the methods reviewed by Suri *et al.*,³⁵ this study formulates the LS PDE for mammographic mass segmentation as follows:

$$\frac{\partial \phi}{\partial t} + (|\vec{V}_0| - b\kappa)|\nabla \phi| + \vec{G} \cdot \nabla \phi = 0, \quad (3)$$

where $|\vec{V}_0| = 1 / (1 + |\vec{G}|)$ and $b = 2|\vec{V}_0|$ balance the effect of internal and external forces. The intuitive idea behind Eq. (3) is that the curve should move fast on a flat surface, while it should evolve slowly as it approaches object edges. Equation (3) is a Hamilton–Jacobi equation. It can be solved numerically.³⁴ One disadvantage of using the LS segmentation is its high computational cost, but researchers have proposed several solutions to increase the efficiency, e.g., the fast marching method which evolves curves only in one direction and the narrow band method which updates the

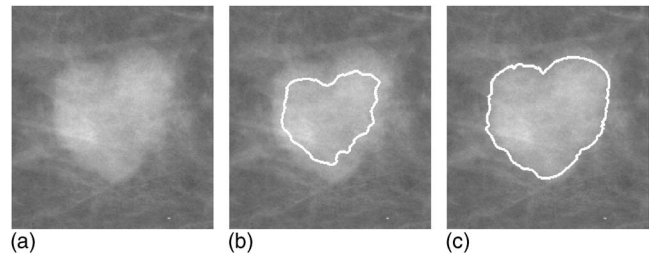


FIG. 3. An example of mammographic mass segmentation using the LS method. (a) Original ROI. (b) Initial boundary obtained by K -means clustering and morphological opening. (c) LS boundary.

signed distance function only in the region near the zero LS.³⁸

Because mammograms typically have low contrast and boundaries of breast masses are vague or obscured by soft tissue or even are not visible on the mammograms at all, segmentation leak is a challenge for mammographic mass segmentation. We borrow the idea of the region growing method to reduce leak. During the curve evolution, each update of the signed distance function maintains the high homogeneity of mass gray level. We compute the average gray level $\mu_{\Omega}(k)$ of the region that is enclosed by the zero LS at the k th iteration. For each pixel in $I(x, y)$ on the zero LS at the $(k+1)$ th iteration, i.e., $\phi(x, y, k+1) = 0$, the local mean $\mu_l(x, y, k+1)$ is computed by a 3×3 average. The purpose of local averaging is to reduce the effect of noise. We update the zero LS as

$$\phi(x, y, k+1) = \begin{cases} 0 & \text{if } \mu_l(x, y, k+1) > \rho \mu_{\Omega}(k) \\ \phi(x, y, k) & \text{otherwise} \end{cases}. \quad (4)$$

ρ was chosen experimentally to be 0.7 in our study. Increasing the ratio ρ may reduce segmentation leak, but it also may cause undersegmentation.

Figure 3(a) shows an example ROI which contains a microlobulated mass that was biopsy-proven to be an invasive ductal carcinoma. The mass edge is blurred and partially overlapped by other soft tissues, so the traditional segmentation methods are prone to segmentation leak. Figure 3(b) shows the boundary resulting from the K -means clustering and morphological opening. The boundary is then refined by the LS segmentation as shown in Fig. 3(c). The LS segmentation covers most mass edges visually, though it still slightly undersegmented on the upper right corner of the mass.

We compared the segmentation results of the newly developed LS method and our previous AC method on our primary data set of 909 masses. Three measures the average minimum distance, the Hausdorff distance, and the area overlap measure, that we previously chose for comparing AC segmentation to radiologists' manual segmentation¹⁵ were used for the current comparison. The first two measures use the minimum Euclidean distance (MED) between a point on one contour and the other contour in the Cartesian plane. The Hausdorff distance computes the maximum of MED while the average

minimum distance computes the mean. The area overlap measure is defined as the ratio of the intersection to the union of the two segmented areas. Their mathematical definitions can be found in the literature.¹⁵

II.D. Feature extraction

We designed two new types of image features, the presence of microcalcifications and margin abruptness, for our new CAD system. These new image features are combined with our previously developed 20 RLS features, three spiculation features, and twelve morphological features including the normalized radial length (NRL) features^{14,15} to form a new image feature space. We compared these features extracted from contours obtained by the AC segmentation and the LS segmentation to evaluate the effectiveness of the segmentation methods. The new mammographic features are described in detail below.

Patient age is a feature which can be collected conveniently and has been included in CAD systems by several groups.^{17,39,40} As described below, we investigated classifiers that relied on purely mammographic features, as well as classifiers that merged the mammographic features with the patient age information.

II.D.1. Microcalcification likelihood feature

Although it is clinically known that masses with microcalcification are more likely to be malignant, this feature has received little attention in CAD. Samulski *et al.* attempted to incorporate the number of calcifications found in the segmented mass region to characterize mass malignancy,¹⁸ but its classification capability was not reported. We have developed an automated system for microcalcification detection on digitized screen-film⁴¹ and full field digital mammograms.²⁸ This study tailored the automated microcalcification detection system to mass ROIs and summarized the output as a feature that describes the likelihood that the ROI contained a microcalcification cluster. Briefly, we used the same image enhancement and individual microcalcification segmentation methods as in our previous publications,⁴¹ with the exception that the number of prescreened signals in this study was lower ($N=40$), because the detection was performed in an ROI as opposed to the entire mammogram. After individual microcalcification candidate locations were determined, we used a trained convolution neural network (CNN) to estimate the likelihood that the signal corresponds to a true microcalcification. We thus obtained a CNN score for each microcalcification candidate. We defined the average of the top three scores in a ROI as the microcalcification likelihood feature.

II.D.2. Margin abruptness features

Breast mass shape and margin are important features to differentiate malignant and benign masses. We have explored spiculation features in our previous studies to improve mammographic mass characterization. In this study, we propose new features to estimate the abruptness of the transition be-

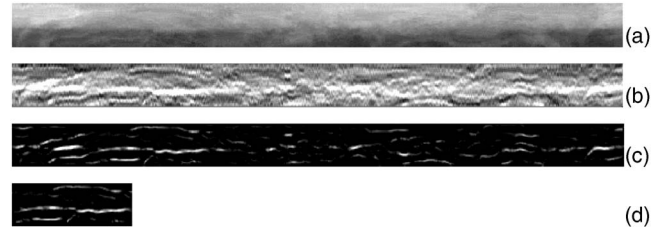


FIG. 4. Extraction of the margin abruptness feature. (a) RBST image from the margin of the mass in Fig. 3(c). (b) The gradient image of (a) in the vertical direction. (c) The HLE-RBST image. (d) The segment that contains the longest line in terms of the DFS algorithm.

tween the mass and the surrounding tissue and the length of the longest segment of this well-circumscribed mass boundary. We first segmented the mass images using either the LS or AC methods. The rubber-band straightening transform²⁶ (RBST) was used to map a band of pixels surrounding the mass onto a rectangular image. Figure 4(a) shows an example of the resulting image after applying the RBST transform to the mass in Fig. 3(c). The mass edge was stretched into several vague lines, which were enhanced by finding the gradient magnitude using the Sobel operator as shown in Fig. 4(b).

After the RBST gradient magnitude was obtained, Hessian analysis⁴² was applied to obtain a Hessian line enhanced RBST (HLE-RBST) image. The Hessian response, which essentially is based on the tensor analysis of a data volume in terms of the second-order derivatives, has many applications in computer-aided diagnosis, for example, microcalcification analysis in mammograms⁴³ and nodule detection and registration in thoracic CT scans.^{44,45} The eigenvalues of the Hessian matrix contain important information about the shape of the object around a pixel.⁴⁶ For better utilizing the Hessian eigenvalues and reducing noise in the second-order derivatives, a common technique is to compute the Hessian matrices at multiple scales. At each scale, the image volume is convolved with an isotropic Gaussian function with a different standard deviation. In this study, we used four different standard deviations with $\sigma_i = 4 \times 3^{i/3}$, $i \in \{0, 1, 2, 3\}$ pixels. We then defined the line response function as described in the literature⁴⁵ to enhance line structures and selected the maximum response under the four different scales. Figure 4(c) shows the resulting HLE-RBST image, in which line structures were enhanced. An ideally circumscribed mass margin corresponds to a long connected line on the HLE-RBST image. We used the depth first search⁴⁷ (DFS) algorithm to search the longest line, starting from candidate locations on the HLE-RBST image. A candidate location is defined as a pixel whose Hessian response is larger than the average response in the entire HLE-RBST image. Figure 4(d) shows the segment of Fig. 4(c) that contains the longest line by using the DFS algorithm.

Two features were extracted for measuring the margin abruptness: normalized line length and normalized line Hessian response. The normalized line length measures the ratio of the horizontal span of the longest line [i.e., the width of Fig. 4(d)] to the length of the whole RBST image [i.e., the

width of Fig. 4(a)]. The normalized line Hessian response measures the ratio of the average Hessian response along the longest line to the average Hessian response on the segment of HLE-RBST image that contains the line in Fig. 4(d).

II.E. Classification

The primary data set was partitioned into training and test sets using leave-one-case-out resampling. For a given round, this resampling method used one case for testing and the remaining cases for training. The case chosen for testing was changed at each round, until all cases had been used once for testing. If a case had multiple views, all views were grouped in either the test set or the training set in the same round, but not both. In each resampling round, the wrapper algorithm²⁹ with stepwise feature selection^{48,49} was used to select the most effective features and then trained a LDA classifier. The designed classifier was applied to all views of the left-out case. The stepwise procedure selected effective features using the Wilks lambda criterion and F statistics.⁴⁹ Since our primary data set had 427 cases, 427 different sets of features were selected and 427 classifiers were trained.

The discriminant scores from leave-one-case-out testing were analyzed with the receiver operating characteristic (ROC) methodology.⁵⁰ The classification performance was measured in terms of the area A_z under the ROC curve. Two types of ROC curves are presented, view-based and case-based. In view-based ROC analysis, the mass on each mammogram was considered an independent sample. The numbers of malignant and benign mass ROIs in the primary data set were therefore 451 and 458, respectively. For case-based ROC analysis, the LDA scores of the same mass seen on different views were averaged. The numbers of malignant and benign mass ROIs were 200 and 227, respectively.

To compare the effect of new features, we designed classifiers in three feature spaces: (1) the previous mammographic feature space that contained the morphological, RLS, and spiculation features, (2) the new mammographic feature space that contained the microcalcification likelihood feature and the margin abruptness features in addition to (1), and (3) the new combined feature space that contained the patient age feature in addition to (2). The same set of features was extracted from the masses segmented by the LS and the AC methods. The view-based systems with the LS and AC segmentation methods are referred to as LS-V and AC-V, respectively. Similarly, the corresponding case-based systems are referred to as LS-C and AC-C, respectively.

II.F. Radiologist classification

As shown in Fig. 2(a), a LM rating was provided by an expert radiologist for all masses based on their appearance on individual mammographic views. Using these ratings, we are able to estimate a view-based ROC curve for the radiologist. To compare with the case-based computer classifiers, a case-based ROC curve was simulated for the radiologist by averaging the LM ratings of the same mass from different mammographic views. To compare with the computer classifiers that use patient age, we designed LDA classifiers that

combined the radiologists' LM rating with patient's age. Similar to the computer classifiers, the weights of the classifiers were trained and tested using a leave-one-case-out scheme.

II.G. Statistical comparisons

We performed ROC analysis on pairs of conditions to investigate a primary interest and to explore a few other issues. Our primary interest was to compare the performance of the CAD system based on LS segmentation and the new mammographic feature space compared to the CAD system based on AC segmentation and the previous mammographic feature space. The other issues of interest include (1) the effect of LS and AC segmentation methods on the mass classification performance of the CAD system, (2) the effect of patient age on mass classification, and (3) performance of the newly developed CAD system compared to the radiologist's assessment.

III. RESULTS

III.A. Segmentation

Automated breast mass segmentation is an important step for the CAD system. As discussed in Sec. II, we quantitatively compared the difference between the LS and AC segmentation methods using the average minimum distance, the Hausdorff distance, and the area overlap measure. For the 909 ROIs in the primary data set, the average minimum distance, the Hausdorff distance, and the area overlap measure were 1.12 ± 0.78 mm, 3.47 ± 1.91 mm, and 0.73 ± 0.10 (mean \pm standard deviation), respectively.

III.B. Classification performance of individual features

Table I shows the A_z values of the features extracted from LS segmentation and used individually to characterize mass malignancy. The average spiculation measure had the best A_z of 0.78 ± 0.02 . As demonstrated in our previous investigation, the spiculation features were very effective for mass characterization. RLS features had A_z values ranging between 0.52 and 0.64 and NRL features had A_z values ranging between 0.55 and 0.64. With an A_z of 0.68 ± 0.02 , patient age also had a limited classification capability. The two margin abruptness features, the normalized line length and normalized line Hessian response, had A_z values of 0.64 ± 0.02 and 0.60 ± 0.02 , respectively.

We used the microcalcification detection results of the MQSA radiologist as the reference standard to evaluate the microcalcification likelihood feature as an indicator for the presence of microcalcifications. The data set for this evaluation therefore contained 152 positive and 757 negative ROIs. ROC analysis indicated that the microcalcification likelihood feature achieved an A_z value of 0.90 ± 0.01 in distinguishing between ROIs that were positive and negative for microcalcification presence. In terms of distinguishing between the malignant and benign masses, however, the microcalcification likelihood feature had an A_z of 0.64 ± 0.02 .

TABLE I. A_z values of individual features in characterizing mass malignancy for features extracted from the LS-segmented masses.

Feature name	A_z	Feature name	A_z
Patient age	0.68	Perimeter	0.71
Microcalcification likelihood	0.64	Fourier descriptor	0.68
Normalized line length	0.64	Square ratio	0.66
Normalized line Hessian response	0.60	NRL zero crossing count	0.64
Average spiculation measure	0.78	NRL mean	0.62
Average spiculation measure of above $\pi/4$	0.77	Area	0.61
Percentage of spiculation measure above $\pi/4$	0.75	Circle ratio	0.60
Gray-level nonuniformity (GLN)	0.59–0.64	NRL entropy	0.57
Run length nonuniformity (RLN)	0.56–0.63	NRL standard deviation	0.57
Run percentage (RP)	0.52–0.62	NRL area ratio	0.55
Short runs emphasis (SRE)	0.57–0.62	Circularity	0.52
Long runs emphasis (LRE)	0.52–0.55	Contrast	0.50

III.C. Comparisons of classification accuracy

We first evaluated how well the new segmentation and new mammographic features can work together to improve the performance of our previous CAD. Figure 5(a) shows the ROC curves for the LS-V system in the new mammographic feature space ($A_z=0.83\pm 0.01$) and the AC-V system in the previous mammographic feature space ($A_z=0.80\pm 0.01$). Figure 5(b) is the counterpart of Fig. 5(a) for case-based classification. ROC analysis⁵⁰ using the CLABROC program indicated that the differences between the view-based A_z values of the two systems [Fig. 5(a)] and between the case-based A_z values [Fig. 5(b)] were statistically significant ($p=0.02$ and $p=0.03$, respectively).

To investigate the effect of the segmentation method on the classification accuracy, we compared the A_z values obtained with the LS and AC methods in the new mammographic feature space. Replacing the LS segmentation with the AC segmentation in the new mammographic feature space reduced the A_z value from 0.83 ± 0.01 to 0.82 ± 0.01 for view-based classification ($p=0.23$) and from 0.85 ± 0.02 to 0.83 ± 0.02 for case-based classification ($p=0.18$).

We also investigated how computer-extracted mammographic features can be combined with patient information. Patient age, which is known to have a positive correlation

with breast cancer risk, was used as a nonimage feature. In the new combined feature space, the use of the patient age improved the A_z value of the LS-V system from 0.83 ± 0.01 to 0.85 ± 0.01 ($p=0.0005$) and that of the AC-V system from 0.82 ± 0.01 to 0.84 ± 0.01 ($p=0.006$). For case-based systems, the use of the patient age improved the A_z value of the LS-C system from 0.85 ± 0.02 to 0.87 ± 0.02 ($p=0.09$) and that of the AC-C system from 0.83 ± 0.02 to 0.85 ± 0.02 ($p=0.16$).

Figure 6(a) includes four ROC curves obtained for the comparison of the LS system to radiologist's LM ratings: (1) radiologist's view-based LM ratings alone ($A_z=0.86\pm 0.01$), (2) radiologist's view-based LM ratings combined with patient age ($A_z=0.87\pm 0.01$), (3) the LS-V system designed in the new mammographic feature space ($A_z=0.83\pm 0.01$), and (4) the LS-V system in the new combined space ($A_z=0.85\pm 0.01$). The radiologist's ratings had a higher A_z value than the computer classifier with ($p=0.05$) or without ($p=0.34$) the age feature. Figure 6(b) shows the corresponding case-based ROC curves. As in view-based classification, the radiologist's ratings had a higher A_z value than the computer classifier with ($p=0.46$) or without ($p=0.21$) the age feature.

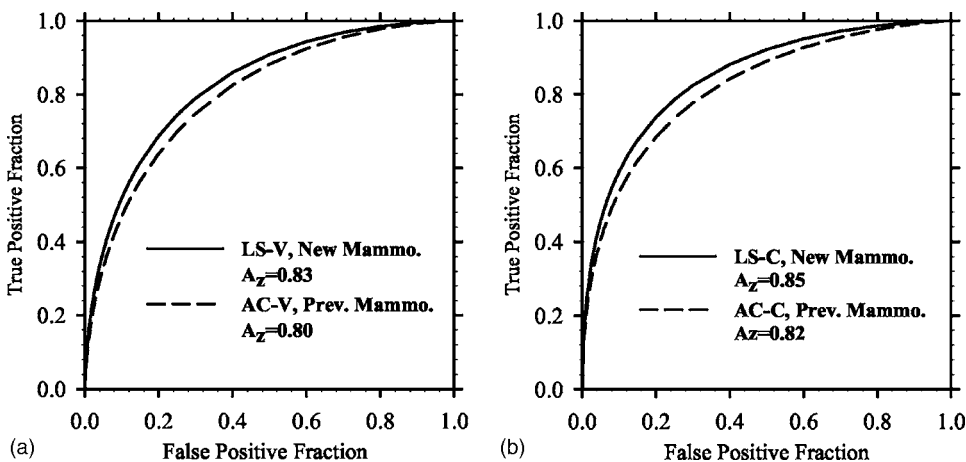


FIG. 5. Classification performances of the LS system in the new mammographic feature space and the AC system in the previous mammographic feature space. (a) View-based analysis. (b) Case-based analysis.

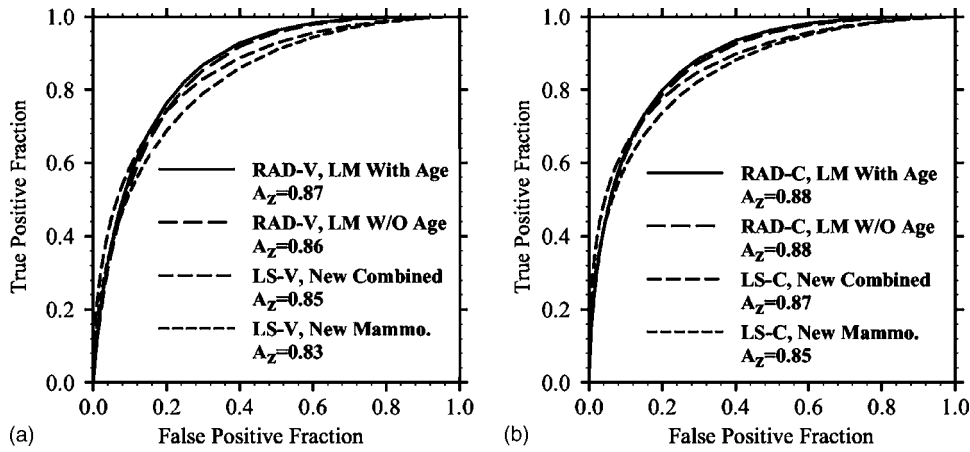


FIG. 6. Classification performances of the LS system and the radiologist (RAD). (a) View-based analysis. (b) Case-based analysis.

Tables II and III, respectively, summarize the A_z values of various computer classifiers and the radiologist's rating with and without including patient age.

III.D. Statistics of selected features for classification

We used the leave-one-case-out resampling for performance evaluation in our primary data set. It included 427 cases, so 427 LDA classifiers were trained. For the LS system in the new combined feature space, the average number of selected features was 7.9. A total of 11 features were selected at least once for at least 1 of the 427 classifiers. In order to study the consistency of these classifiers, we investigated how many times each feature was selected and how similar the weights were for the selected features. The first and second columns in Table IV, respectively, list the selected features and how many times they were selected among the 427 classifiers. Only 9 of the 11 features are shown; two RLS features that were selected less than three times were not included. Three out of the five features that were selected in all resampling rounds were the newly developed features, although their individual A_z values were low (refer to Table I). Among the morphological features, only the mass area was selected. The RLS texture features were selected less frequently. The coefficient of variation, which expresses the standard deviation as a percentage of the mean for the weights of each feature in the LDA classifiers, is also shown in Table IV. A small coefficient of variation indicates a small variation in the LDA weights of a feature as the training set changes in the leave-one-case-out resampling

procedure. This variation was less than 4% for any selected feature, which demonstrates the consistency of the newly developed CAD system for this data set.

III.E. Independent test

In order to verify the classification accuracy of the newly developed CAD system, a subset of the DDSM database, described in Sec. II, was used as an independent test set. Five features that were consistently selected in the 427 LDA classifiers for the primary data set with the LS system (Table IV), i.e., patient age, mass area, average spiculation measure, microcalcification likelihood, and normalized line length, were used for training a LDA classifier for the LS system using the primary data set. Similarly, five features, i.e., patient age, mass perimeter, average spiculation measure, microcalcification likelihood, and normalized line length, consistently selected for the primary data set with the AC system were used for training another LDA classifier for the AC system using the primary data set. The two trained LDA classifiers were then tested in the independent test data set. The view-based A_z values for the LS-V and the AC-V systems were 0.84 ± 0.02 and 0.81 ± 0.02 , respectively. The case-based A_z values for the LS-C and the AC-C systems were 0.85 ± 0.03 and 0.83 ± 0.04 , respectively. The difference of the LS and AC systems was not significantly different, using either view-based ($p=0.17$) or case-based ($p=0.25$) analysis.

TABLE II. A_z values of the classifiers using the AC and LS segmentation methods, under view-based and case-based analyses, in different feature spaces.

Segmentation	Feature space					
	Previous mammographic		New mammographic		New combined (with age)	
	View-based	Case-based	View-based	Case-based	View-based	Case-based
Active contour	0.80 ± 0.01	0.82 ± 0.02	0.82 ± 0.01	0.83 ± 0.02	0.84 ± 0.01	0.85 ± 0.02
Level set	0.81 ± 0.01	0.84 ± 0.02	0.83 ± 0.01	0.85 ± 0.02	0.85 ± 0.01	0.87 ± 0.02

TABLE III. A_z values obtained by using the radiologist's LM rating in combination with patient age under view-based and case-based analyses. For case-based analysis, the LM rating of the radiologist was simulated by averaging the view-based LM ratings from different views. A LDA classifier was designed for combining the radiologist's LM rating with patient age.

Feature space	Evaluation mode	
	View-based	Case-based
Radiologist's LM rating	0.86±0.01	0.88±0.02
Radiologist's LM rating with patient age	0.87±0.01	0.88±0.02

IV. DISCUSSION

In this study, we investigated how LS segmentation and the new features may improve the CAD performance. When compared with the AC segmentation in the same feature space, LS segmentation demonstrated slightly higher performance in terms of the ROC curves, but the improvement was not statistically significant. One reason is that the segmentation of mammographic masses is inherently a very challenging task so that both the AC and the LS methods may fail to achieve accurate segmentation for some masses. Our LS segmentation was designed to reduce segmentation leak by maintaining the homogeneity of the mass. The side effect was that the homogeneity constraint may cause undersegmentation of some masses. Further improvement of mammographic mass segmentation is important for breast cancer CAD systems.

Our current and previous studies have demonstrated that the mass spiculation features were powerful in characterizing mass malignancy. Spiculated mass margin often indicates high likelihood of malignancy, while circumscribed mass margin often indicates benignity. In this study, we developed features that described the abruptness of the mass margin. Our results showed that one of the new margin abruptness features was consistently selected by stepwise feature selection and, in combination with other mammographic features, improved the CAD performance. However, the accuracy of the individual margin abruptness features in classifying the masses as malignant or benign was low. Another feature de-

veloped in this study was related to the presence of microcalcification clusters in the mass ROIs. Similar to the abruptness of the mass margin, microcalcification by itself had limited classification capability, but was selected consistently in stepwise selection. Patient age has been incorporated in several CAD systems by other groups.^{17,24,39,40} Our study shows that when combined with view-based features, patient age can improve the accuracy of mass characterization. Because patient age is an objective feature that can be collected conveniently, it may be a cost-effective adjunct to the mammographic features.

For most masses, the LM rating was provided by the MQSA radiologist in the same reading session as the identification of the bounding box using available clinical information. There is a possibility that the radiologist's ROC curve may have been optimistically biased by this information. However, we believe that a comparison of the computer classifiers with the radiologist's classification is useful for providing a reference about the performance of the CAD system. When patient age was not used, radiologists' LM rating achieved a higher A_z value than that using the computer classifier scores, for both view-based ($p=0.05$) and case-based ($p=0.21$) analysis. In order to compare the performance of the CAD system in the new combined feature space (which included patient age) to that of the radiologist, we designed LDA classifiers that combined the radiologist's LM ratings with patient age. Our results indicated that the difference in the A_z values between the CAD system and the radiologist, both with patient age information, did not achieve statistical significance, ($p=0.34$ and 0.46 for view-based and case-based analysis, respectively) although the radiologist's performance was still better. Since radiologists may use patient age in different ways (e.g., nonlinear combination) for patient management, future work is needed to establish the optimal use of patient age in a computer classifier.

This investigation presented several p values for various statistical comparisons. While it is true that searching in different subgroups or with different hypotheses until some p value happens to fall below the alpha level of 0.05 is unwarranted, there is debate in the biostatistical literature concerning when adjustment for multiple testing is needed or how adjustment should be made.^{51,52} In this study, we use an alpha level of 0.05 for statistical significance for our primary interest, which is the comparison of the performance of our new and old CAD systems. For statistical tests on other exploratory issues, if the p value is less than 0.05, we do not make a comment about whether the test result is statistically significant because the significance may depend on the adjustment of the alpha level for multiple hypothesis testing. If the p value is larger than 0.05, we state that the test does not achieve significance because this would be the case either without or with an adjustment of the alpha level.

V. CONCLUSION

We investigated the LS segmentation method and designed new types of mammographic features including abruptness of the mass margin and the presence of microcal-

TABLE IV. The number of times each feature was selected and the coefficients of variation of the weights of each feature in the LDA classifiers for the LS system. Two RLS features that were selected less than three times were not included.

Feature name	Number of times selected	Coefficient of variation (%)
Patient age	427	1.0
Mass area	427	3.6
Average spiculation measure	427	3.0
Microcalcification likelihood	427	2.5
Normalized line length	427	1.7
Horizontal GLN 90	398	1.8
Horizontal RLN 90	400	1.5
Vertical GLN0	399	1.0
Vertical SRE0	27	2.4

cifications in the mass. We also investigated the effect of combining patient age with mammographic features for improving computerized characterization of masses. The individual performance of the new features was poorer than some of our previous features, especially the spiculation features. However, the new features were consistently selected using the stepwise feature selection method for LDA classification. Statistical tests showed that our new CAD system based on LS segmentation and the new mammographic feature space performed significantly better than our previous CAD system that was based on AC segmentation and the morphological, RLS, and spiculation features. Patient age by itself had limited capability in classifying breast mass malignancy. However, when combined with mammographic features, patient age improved the performance of the CAD system. Our experimental results also indicate that the new CAD system may have the potential to approach the performance of an expert radiologist in classification of masses.

ACKNOWLEDGMENTS

This work is supported by USPHS Grant Nos. CA95153 and CA118305 and USAMRMC Grant No. DAMD17-02-1-0214. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program.

NOMENCLATURE

AC	=	active contour
AC-C	=	AC segmentation for case-based systems
AC-V	=	AC segmentation for view-based systems
A_z	=	area under the ROC curve
BI-RADS	=	breast imaging reporting and data system
CAD	=	computer-aided diagnosis
CC	=	craniocaudal
CNN	=	convolution neural network
DDSM	=	digital database for screening mammography
DFS	=	depth first search
FFDM	=	full field digital mammograms
HLE-RBST	=	Hessian line enhanced RBST
IRB	=	Institutional Review Board
LDA	=	linear discriminant analysis
LM	=	likelihood of malignancy
LS	=	level set
LS-C	=	LS segmentation for case-based systems
LS-V	=	LS segmentation for view-based systems
MED	=	minimum Euclidean distance
MLO	=	mediolateral oblique
MQSA	=	Mammography Quality Standard Act
NRL	=	normalized radial length
PDE	=	partial differential equation
RAD	=	radiologist
RBST	=	Rubber-band straightening transform
RLS	=	run-length statistics
ROC	=	receiver operating characteristic
ROI	=	original region of interest
SFM	=	digitized screen film mammograms
SGLD	=	spatial gray level dependence

^{a)}Author to whom correspondence should be addressed. Present address: Department of Radiology, University of Michigan, 1500 E. Medical Center Drive, CGC B2103, Ann Arbor, MI 48109-0904. Telephone: (734)647-8553; Fax: (734)615-5513. Electronic mail: sjiazhen@med.umich.edu

- ¹American Cancer Society, "Statistics for 2005," www.cancer.org, 2005.
- ²C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," *Cancer* **75**, 1619–1626 (1995).
- ³S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," *AJR Am. J. Roentgenol.* **171**, 29–33 (1998).
- ⁴B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," *Cancer* **91**, 1699–1703 (2001).
- ⁵L. Tabar *et al.*, "Reduction in mortality from breast cancer after mass screening with mammography," *Lancet* **1** (8433), 829–832 (1985).
- ⁶H. C. Zuckerman, in *Breast Cancer, Diagnosis and Treatment*, edited by I. M. Ariel and J. B. Cleary (McGraw-Hill, New York, 1987).
- ⁷R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**, 613–617 (1992).
- ⁸F. M. Hall, J. M. Storella, D. Z. Silverstone, and G. Wyshak, "Nonpalpable breast lesions: Recommendations for biopsy based on suspicion of carcinoma at mammography," *Radiology* **167**, 353–358 (1988).
- ⁹G. Hermann, C. Janus, I. S. Schwartz, B. Krivisky, S. Bier, and J. G. Rabinowitz, "Nonpalpable breast lesions: Accuracy of prebiopsy mammographic diagnosis," *Radiology* **165**, 323–326 (1987).
- ¹⁰J. E. Meyer, T. J. Eberlein, P. C. Stomper, and M. R. Sonnenfeld, "Biopsy of occult breast lesions: Analysis of 1261 abnormalities," *JAMA, J. Am. Med. Assoc.* **263**, 2341–2343 (1990).
- ¹¹H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai, and H. N. Du, "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognit.* **39**, 646–668 (2006).
- ¹²J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge, and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Med. Phys.* **32**, 2827–2838 (2005).
- ¹³A. Retico, P. Delogu, M. E. Fantacci, and P. Kasae, "An automatic system to discriminate malignant from benign massive lesions on mammograms," *Nucl. Instrum. Methods Phys. Res. A* **569**, 596–600 (2006).
- ¹⁴B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and L. M. Hadjiiski, "Improvement of mammographic mass characterization using spiculation measures and morphological features," *Med. Phys.* **28**, 1455–1465 (2001).
- ¹⁵B. Sahiner, N. Petrick, H. P. Chan, L. M. Hadjiiski, C. Paramagul, M. A. Helvie, and M. N. Gurcan, "Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization," *IEEE Trans. Med. Imaging* **20**, 1275–1284 (2001).
- ¹⁶Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, and C. E. Metz, "Computerized classification of benign and malignant masses on digitized mammograms: A study of robustness," *Acad. Radiol.* **7**, 1077–1084 (2000).
- ¹⁷J. L. Jesneck, L. W. Nolte, J. A. Baker, C. E. Floyd, and J. Y. Lo, "Optimized approach to decision fusion of heterogeneous data for breast cancer diagnosis," *Med. Phys.* **33**, 2945–2954 (2006).
- ¹⁸M. Samulski, N. Karssemeijer, P. Lucas, and P. Groot, "Classification of mammographic masses using support vector machines and Bayesian networks," *Proc. SPIE* **6514**, 65141J (2007).
- ¹⁹D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng.* **2**, 315–337 (2000).
- ²⁰American College of Radiology *Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas)*, 4th ed. (American College of Radiology, Reston, VA, 2003).
- ²¹D. Guliato, R. M. Randayyan, W. A. Carnielli, J. A. Zuffo, and J. E. L. Desautels, "Segmentation of breast tumors in mammograms by fuzzy region growing," *J. Electron. Imaging* **12**, 369–378 (2003).
- ²²M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Trans. Med. Imaging* **17**, 510–517 (1998).
- ²³S. Timp and N. Karssemeijer, "A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography," *Med. Phys.* **31**, 958–971 (2004).
- ²⁴J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic le-

- sions using computer-aided image analysis," *IEEE Trans. Med. Imaging* **12**, 664–669 (1993).
- ²⁵L. Shen, R. M. Rangayyan, and J. E. L. Desautels, "Application of shape analysis to mammographic calcifications," *IEEE Trans. Med. Imaging* **13**, 263–274 (1994).
- ²⁶B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," *Med. Phys.* **25**, 516–526 (1998).
- ²⁷Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick, P. Lu, D. E. Wolverton, and R. A. Schmidt, "Analysis of spiculation in the computerized classification of mammographic masses," *Med. Phys.* **22**, 1569–1579 (1995).
- ²⁸J. Ge, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, J. Wei, M. A. Helvie, and C. Zhou, "Computer aided detection of clusters of microcalcifications on full field digital mammograms," *Med. Phys.* **33**, 2975–2988 (2006).
- ²⁹H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.* **17**, 491–502 (2005).
- ³⁰A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample size performance," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 153–158 (1997).
- ³¹N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evaluation of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients," *Radiology* **224**, 217–224 (2002).
- ³²M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer, R. Moore, K. Chang, and S. Munishkumar, in *Digital Mammography*, edited by N. Karssemeijer, M. Thijssen, J. Hendriks, and L. van Erning (Kluwer Academic, Dordrecht, 1998).
- ³³B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue on mammograms," *Med. Phys.* **23**, 1671–1684 (1996).
- ³⁴S. J. Osher and R. P. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces* (Springer, New York, 2003).
- ³⁵J. S. Suri, K. C. Liu, S. Singh, S. N. Laxminarayan, X. L. Zeng, and L. Reden, "Shape recovery algorithms using level sets in 2-D/3-D medical imagery: A state-of-the-art review," *IEEE Trans. Inf. Technol. Biomed.* **6**, 8–28 (2002).
- ³⁶J. A. Sethian, *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision and Materials Sciences* (Cambridge University Press, Cambridge, 1996).
- ³⁷E. J. Street, L. M. Hadjiiski, B. Sahiner, S. Gujar, M. Lbrahim, and H. P. Chan, "Volume analysis of head and neck lesions on CT scans using 3D level set segmentation," *Radiological Society of North America* **2006**, 510–510 (2006).
- ³⁸J. A. Sethian, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science* (Cambridge University Press, Cambridge, 1999).
- ³⁹M. Kallergi, "Computer-aided diagnosis of mammographic microcalcification clusters," *Med. Phys.* **31**, 314–326 (2004).
- ⁴⁰S. Gupta, P. F. Chyn, and M. K. Markey, "Breast cancer CAD(x) based on BI-RADS (TM) descriptors from two mammographic views," *Med. Phys.* **33**, 1810–1817 (2006).
- ⁴¹H. P. Chan, S. C. B. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network," *Med. Phys.* **22**, 1555–1567 (1995).
- ⁴²A. F. Frangi, W. Niessen, R. Hoogeveen, T. v. Walsum, and M. Viergever, "Model-based quantitation of 3D magnetic resonance angiographic images," *IEEE Trans. Med. Imaging* **18**, 946–956 (1999).
- ⁴³R. Nakayama, Y. Uchiyama, K. Yamamoto, R. Watanabe, and K. Namba, "Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms," *IEEE Trans. Biomed. Eng.* **53**, 273–283 (2006).
- ⁴⁴J. Shi, B. Sahiner, H.-P. Chan, L. M. Hadjiiski, C. Zhou, P. N. Cascade, E. A. Kazerooni, Y.-T. Wu, and J. Wei, "Pulmonary nodule registration in serial CT scans based on rib anatomy and nodule template matching," *Med. Phys.* **34**, 1336–1347 (2007).
- ⁴⁵Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans," *Med. Phys.* **30**, 2040–2051 (2003).
- ⁴⁶Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis, "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Med. Image Anal.* **2**, 143–169 (1998).
- ⁴⁷T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms* (MIT Press, Cambridge, MA, 2001).
- ⁴⁸B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner, and L. M. Hadjiiski, "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size," *Med. Phys.* **27**, 1509–1522 (2000).
- ⁴⁹N. R. Draper, *Applied Regression Analysis* (Wiley, New York, 1998).
- ⁵⁰C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Stat. Med.* **17**, 1033–1053 (1998).
- ⁵¹M. Aickin and H. Gensler, "Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods," *Am. J. Public Health* **86**, 726–728 (1996).
- ⁵²T. V. Perneger, "What's wrong with Bonferroni adjustments," *Br. Med. J.* **316**, 1236–1238 (1998).