Characterization of masses in digital breast tomosynthesis: Comparison of machine learning in projection views and reconstructed slices

Heang-Ping Chan,[a] Yi-Ta Wu,[b] Berkman Sahiner, Jun Wei, Mark A. Helvie, and Yiheng Zhang[c]
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109

Richard H. Moore and Daniel B. Kopans
Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts 02114

Lubomir Hadjiiski and Ted Way[d]
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109

(Received 7 January 2010; revised 31 March 2010; accepted for publication 3 May 2010; published 15 June 2010)

Purpose: In digital breast tomosynthesis (DBT), quasi-three-dimensional (3D) structural information is reconstructed from a small number of 2D projection view (PV) mammograms acquired over a limited angular range. The authors developed preliminary computer-aided diagnosis (CADx) methods for classification of malignant and benign masses and compared the effectiveness of analyzing lesion characteristics in the reconstructed DBT slices and in the PVs.

Methods: A data set of MLO view DBT of 99 patients containing 107 masses (56 malignant and 51 benign) was collected at the Massachusetts General Hospital with IRB approval. The DBTs were obtained with a GE prototype system which acquired 11 PVs over a 50° arc. The authors reconstructed the DBTs at 1 mm slice interval using a simultaneous algebraic reconstruction technique. The region of interest (ROI) containing the mass was marked by a radiologist in the DBT volume and the corresponding ROIs on the PVs were derived based on the imaging geometry. The subsequent processes were fully automated. For classification of masses using the DBT-slice approach, the mass on each slice was segmented by an active contour model initialized with adaptive *k*-means clustering. A spiculation likelihood map was generated by analysis of the gradient directions around the mass margin and spiculation features were extracted from the map. The rubber band straightening transform (RBST) was applied to a band of pixels around the segmented mass boundary. The RBST image was enhanced by Sobel filtering in the horizontal and vertical directions, from which run-length statistics texture features were extracted. Morphological features including those from the normalized radial length were designed to describe the mass shape. A feature space composed of the spiculation features, texture features, and morphological features extracted from the central slice alone and seven feature spaces obtained by averaging the corresponding features from three to 19 slices centered at the central slice were compared. For classification of masses using the PV approach, a feature extraction process similar to that described above for the DBT approach was performed on the ROIs from the individual PVs. Six feature spaces obtained from the central PV alone and by averaging the corresponding features from three to 11 PVs were formed. In each feature space for either the DBT-slice or the PV approach, a linear discriminant analysis classifier with stepwise feature selection was trained and tested using a two-loop leave-one-case-out resampling procedure. Simplex optimization was used to guide feature selection automatically within the training set in each leave-one-case-out cycle. The performance of the classifiers was evaluated by the area (A_z) under the receiver operating characteristic curve.

Results: The test A_z values from the DBT-slice approach ranged from 0.87 ± 0.03 to 0.93 ± 0.02, while those from the PV approach ranged from 0.78 ± 0.04 to 0.84 ± 0.04. The highest test A_z of 0.93 ± 0.02 from the nine-DBT-slice feature space was significantly ($p = 0.006$) better than the highest test A_z of 0.84 ± 0.04 from the nine-PV feature space.

Conclusion: The features of breast lesions extracted from the DBT slices consistently provided higher classification accuracy than those extracted from the PV images. [DOI: 10.1118/1.3432570]

Key words: digital breast tomosynthesis, computer-aided diagnosis, mass, SART

I. INTRODUCTION

In mammography, the presence of overlapping dense fibroglandular tissue not only reduces the sensitivity of cancer detection, but also the conspicuity of the features of a lesion. Digital breast tomosynthesis mammography (DBT) is a new imaging modality that may alleviate this problem. DBT is a limited-angle tomographic technique in which a series of

projection view (PV) images is acquired as the x-ray source is rotated over a limited angular range or moved over a limited distance about the breast. Tomographic slices of the imaged volume from the series of PV images are then generated by a tomosynthesis reconstruction technique.[1–5] Although DBT can only provide quasi-three-dimensional (3D) structural information due to the limited-angle information, it can reduce the camouflaging effects of fibroglandular tissues. The advantages of DBT are that the in-plane spatial resolution is similar to that of digital mammograms and, with a detector of high detective quantum efficiency, the total dose required for a DBT scan can be kept at nearly the same or only slightly higher than that of a regular mammogram. DBT is one of the promising methods that may improve the detection and characterization for breast lesions.[6]

DBT has not been approved for routine clinical use. Development of computer-aided detection (CADe) and computer-aided diagnosis (CADx) systems for DBT is still at an early stage and no studies have been reported to date on its effects on radiologists' interpretation of DBT. CADe for mammography has been shown to improve breast cancer detection in laboratory observer studies and prospective clinical trials. Although lesions may be more easily visualized in DBT, the number of images in a DBT exam is much larger than that in a mammographic exam. With the help of a well-designed DBT display workstation, the radiologist's efficiency in reading each slice can be much higher than that in reading a regular mammogram, due to the correlation between adjacent DBT slices and the less complex background. However, an initial study showed that the time required for interpretation of a DBT case was still substantially longer than that for mammograms.[7] With the increase in radiologists' workload, the chance for oversight of subtle lesions may not be negligible. It is expected that CADe will play an important role in DBT, especially for detection of microcalcifications. CADx for mammography has not been implemented in clinical practice. However, laboratory observer studies have demonstrated that CADx can provide significant improvement in radiologists' characterization of masses and microcalcifications.[8–12] Computerized image analysis can extract diagnostic information that may not be perceived easily or consistently by human readers.

Efforts are underway to develop computerized detection systems for DBT, but most of the studies are still preliminary.[13–18] Very few studies have been conducted for computerized lesion characterization in DBT to date. Chan *et al.*[19,20] investigated the feasibility of classification of malignant and benign masses using the reconstructed DBT slices as input. Palma *et al.*[21] performed mass segmentation on PV images and merged the extracted information from all PVs for classification using a fuzzy decision tree. A pilot data set of 23 lesions was classified into spiculated and circumscribed masses using a leave-one-out method. Chan *et al.*[22] conducted a pilot study to compare the accuracy of classifying malignant and benign masses when feature analysis was performed on one and five DBT slices or PV images.

In DBT, computerized lesion detection or characterization can be performed in several ways. One common approach is to perform image analysis in the individual 2D PV images and then combine the information from the multiple PVs. The advantage of the PV approach is that it is independent of the reconstruction method. The subsequent fusion of information from the multiple PVs takes advantage of the correlation among the images to reduce noise and false positives. The second approach is to combine the information in the PVs first by tomosynthesis reconstruction and then perform image analysis in the individual DBT slices. This approach takes advantage of the image reconstruction technique in combining the spatial information accurately before image analysis. The signal-to-noise ratio in the reconstructed images may be improved due to the reduction in both the random noise and the structured background. Other approaches include performing image analyses on the PVs, the DBT slices, or the 3D volume at the various processing steps in combination. In this early study with a limited data set, we focused on the development of machine learning techniques using the first two approaches and the comparison of the dependence of their classification performance on the utilization of multiple images.

II. MATERIALS AND METHODS

II.A. Data set

The DBT mammograms used in this study were collected by the Breast Imaging Research laboratory at the Massachusetts General Hospital (MGH) with the approval of the Institutional Review Board. Patients with suspicious breast lesions were recruited with written informed consent. The DBT system was a first-generation GE prototype. The system had a flat panel CsI/a:Si detector with a pixel pitch of 0.1×0.1 mm^2. It acquired 11 PVs in 5° increments over an arc of 50° degrees. The breasts were imaged in the mediolateral oblique (MLO) view. The total dose for the 11 PVs of a DBT scan was designed to be about 1.5 times that of a single standard film mammogram. We reconstructed the DBTs using a simultaneous algebraic reconstruction technique.[5] The reconstructed slices had a pixel size of 0.1×0.1 mm^2 and a slice interval of 1 mm.

In each DBT volume, the volume of interest (VOI) containing the mass was identified by a Mammography Quality Standards Act-approved radiologist. A data set of 107 masses from 99 patients and 102 breasts was used. Fifty-six malignant masses were obtained from 55 patients, one of whom had bilateral malignant masses. Fifty-one benign masses were obtained from 44 patients; five breasts had two benign masses and two patients had bilateral benign masses. All malignant masses were biopsy-proven. The benign masses were biopsy-proven or by follow-up. The VOI was marked as a rectangular region of interest (ROI) on the "central" slice which was defined as the slice where the mass was best visualized, and bounded from the top and bottom by the first and last slice where the mass became almost invisible. The mass ROI was extracted from each DBT slice with a margin of about 4 mm around the bounding box. For a given mass, the corresponding ROIs on the PVs were located by forward projection of the central slice of the mass to the PVs using

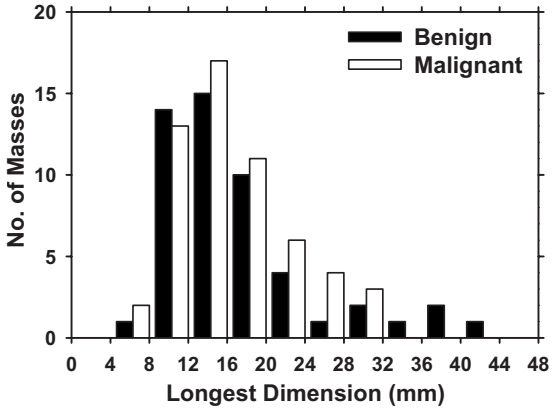


FIG. 1. Histogram of the longest diameter of the masses in the data set. The longest diameter was measured at the central slice where the mass was best visualized as determined by radiologist.

the known geometry of the DBT system. This ensured that the corresponding ROIs on the PVs would be analyzed for each mass. The longest diameter of the mass on the central DBT slice was estimated by the radiologist using an electronic caliper. Figure 1 shows the histogram of the longest diameter of the masses in the data set.

II.B. Image analysis

The automated mass classification scheme used for analysis of both DBT slices and PV images consists of several steps: Segmentation of the mass within the ROI, feature extraction, and feature classification. For the mass segmentation and feature extraction steps, we applied our algorithms previously developed for digitized film mammograms directly to the DBT slices and PVs without retraining the parameters. The details of the methods can be found in the literature.[23–25] The training cases of film mammograms used in the previous studies were obtained from patient files at the University of Michigan and were completely independent of the DBT cases recruited at the MGH. For the feature classification step, a linear discriminant analysis (LDA) classifier with stepwise feature selection was designed using a two-loop leave-one-case-out resampling method, as described below.

II.B.1. Mass segmentation

In this study, we used a two-dimensional approach to image analysis in DBT so that the DBT slices or the PV images were treated similarly. The mass segmentation method[23,24] is summarized below. The mass ROI from an individual DBT slice or PV was used as input. Background correction[25] was applied to the ROI by first estimating a low frequency background image. The gray level of a given pixel in the background image was estimated as the inverse-distance weighted average of the mean pixel values from four small regions on the four sides of the mass ROI. The estimated background image was then subtracted from the ROI to reduce the low frequency intensity variation in the ROI. An adaptive *K*-means clustering algorithm[26] was applied to the background-corrected ROI to segment the pixels into two classes: The mass and the surrounding tissue. A feature vector composed of the gray level values of the original image and its median-filtered image was used for characterizing each pixel in the ROI. The clustering algorithm estimated whether a pixel would belong to the mass class or the background class based on the ratio of the Euclidean distances from the pixel to the mass and the background cluster centers in the feature space. If the clustering extracted more than one object from the ROI, the largest connected object near the center of the ROI was chosen as the mass. Rows 1 and 2 of Figs. 2 and 3 show examples of the original image and segmentation with clustering for a spiculated mass on PVs and DBT slices, respectively. The selected object was filled and its border was eroded by using a morphological erosion operation with a circular mask of 15 pixels in diameter. The boundary of the eroded object was then used as the initial contour for an active contour (AC) model. The initial contour was iteratively deformed to push the contour toward the mass boundary by minimization of a cost function with five energy terms, namely, curvature, continuity, image gradient, object homogeneity, and a balloon force, which represented the internal and external forces for segmentation of breast masses on mammograms. The parameters of the segmentation algorithms were chosen in our previous studies[23,24] such that the segmented object included mainly the body of the mass without the potential spiculations. Row 3 of Figs. 2 and 3 show examples of the AC segmentation results.

After AC segmentation, we used a spiculation extraction technique developed previously to enhance structures that might represent spiculations, and segment the spiculations if they were determined to be present.[23,24] Briefly, image analysis was performed in a 30-pixel wide band surrounding the segmented mass. At each pixel p in the band, the mean of the gradient directions for the pixels within a neighborhood was calculated. The neighborhood was defined as a sector centered at the pixel p, and bounded by an arc of 4 mm radius and two radii at angles of $\pm 45°$ about the normal from the pixel to the mass border. Since the spiculations were extending from the mass approximately in the radial direction and normal to the mass border, if the pixel p lay on the path of a spiculation, then the gradients in the sector would have a dominant direction close to 90° relative to the normal, and hence the mean of the spiculation measure would be high. After spiculation analysis was performed over the entire band, the mean spiculation measures computed at the pixels of the band formed a spiculation likelihood map, as shown in row 4 of Figs. 2 and 3. The approximately radial line structures with high pixel values in the spiculation likelihood map indicated the possible locations of spiculations along the mass border. A threshold was then applied to the spiculation likelihood map to segment possible spiculations. Based on the number and size of the segmented possible spiculations, the mass in a given image was classified either as spiculated or nonspiculated. If the mass in the image was classified as

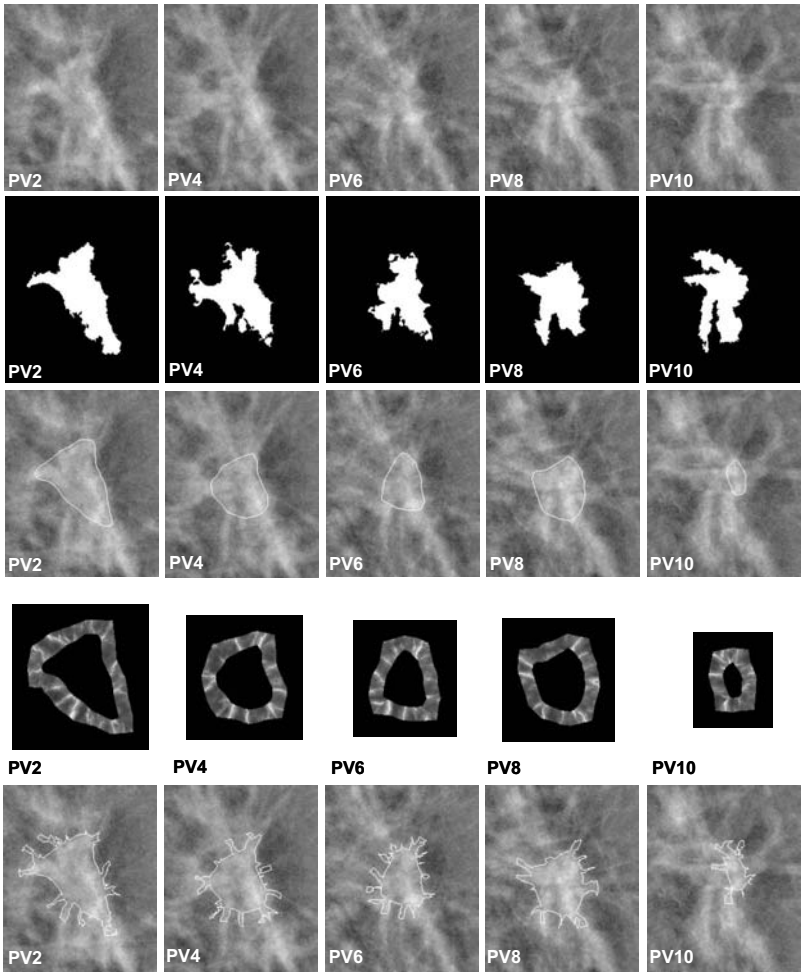


FIG. 2. Projection view images of a spiculated mass. The DBT scan had 11 PVs but only five are shown. PV6 is the central PV at a projection angle of 0°. Row 1: PVs. Row 2: Binary images obtained by adaptive *K*-means clustering. Row 3: Mass boundaries segmented by the AC method. Row 4: Spiculation likelihood map. Row 5: Segmented mass boundary and spiculations.

spiculated, then the spiculations segmented with the threshold described above were joined to the body of the mass segmented with the AC method. Examples of this final segmentation are shown in the last row of Fig. 2, in which all different projection views of the mass were classified as spiculated. The last row of Fig. 3 shows the final segmentation results of the same mass in the reconstructed DBT volume.

II.B.2. Feature extraction

Three types of image features, namely, spiculation, texture, and morphological features, were extracted from the segmented mass to describe the size, shape, margin characteristics, and texture in the breast tissue surrounding the mass. A total of 40 features were extracted from the DBT slices or from the PVs of each mass. The features were summarized in Table I and described below.

II.B.2.a. Texture features. Important texture information exists in the tissue surrounding the mass margin. Malignant masses are more likely to have spiculated and ill-defined borders and the texture in the surrounding tissue may be different from that of benign masses. To facilitate the extraction of texture features that radiate from the borders of the mass, we previously developed a transformation, referred to as the rubber band straightening transform (RBST), that maps the pixel values in a band of pixels surrounding the mass onto the Cartesian plane.[23] The RBST produces a rectangular image of the margin region, in which moving along a row of the image approximately corresponds to traversing a closed path at a constant distance from the segmented mass border in the original image, and moving along a column of the image approximately corresponds to traveling in a direction normal to the segmented mass border in the original image. The possible spiculations in the radial directions, therefore, align approximately in the vertical direction in the RBST image. The mass border is obtained by applying a morphological opening operation to the AC segmentation results described in Sec. II B 1. Circular structuring elements with diameters of 9 and 3 pixels were used for the erosion and dilation operations, respectively. The band surrounding the mass margin for RBST was chosen to be 40 pixels wide, as determined in previous studies. Examples of the RBST images for a mass on a DBT slice and on a PV image are shown in Figs. 4(a) and 4(b), respectively.

Sobel filtering was applied to the RBST image in both the horizontal and vertical directions to enhance the gradients, resulting in horizontal and vertical gradient magnitude images. The vertical and horizontal gradient magnitude images corresponding to the examples of RBST images were also

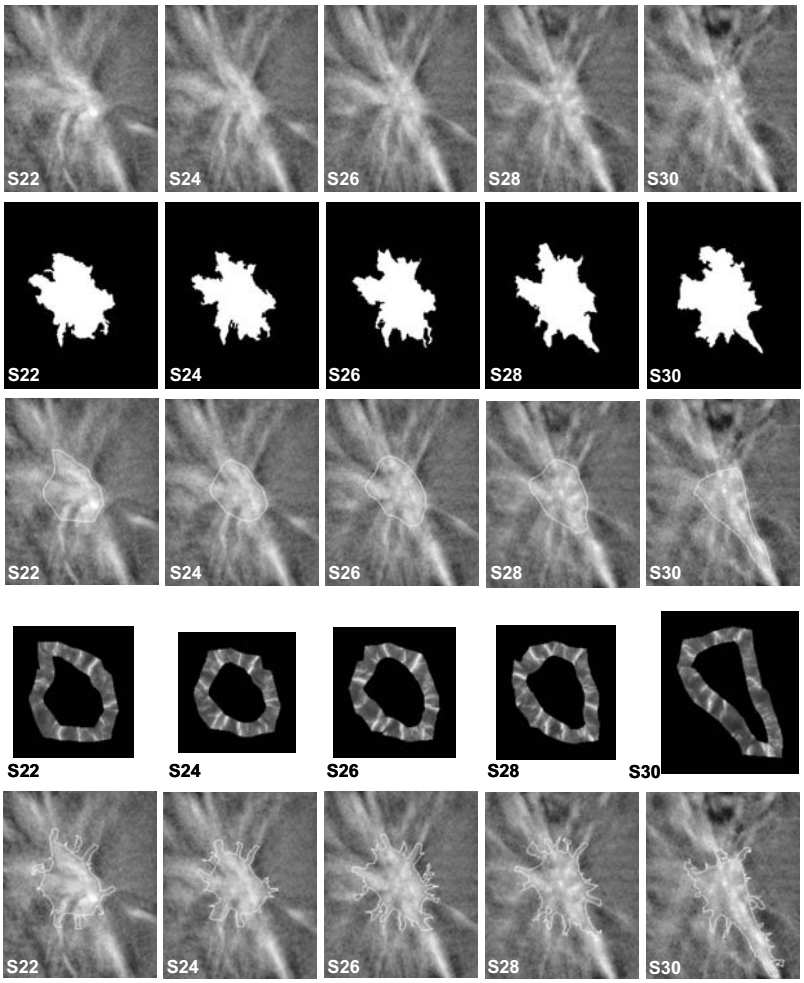


FIG. 3. Five DBT slices of a spiculated mass. S26 is the central slice identified by an experienced radiologist. Row 1: DBT slices. Row 2: Binary images obtained by adaptive *K*-means clustering. Row 3: Mass boundaries segmented by the AC method. Row 4: Spiculation likelihood map. Row 5: Segmented mass boundary and spiculations.

shown in Fig. 4. The run-length statistics (RLS) matrix[27] that described the statistics of runs of consecutive, collinear image pixels having the same gray level value was extracted from each of the gradient magnitude images in two directions (0° and 90°). Five texture features including the short-runs emphasis, long runs emphasis, gray level nonuniformity, run-length nonuniformity, and run percentage were calculated from each of the matrices. A total of 20 RLS features were thus obtained for each image.

II.B.2.b. Spiculation features. We previously developed image analysis methods for extracting spiculation information surrounding the mass margin on mammograms.[24,28] Based on these methods, we extracted a total of eight spiculation features from each image. The first three features were related to the spiculation measure at the pixels along the border of the mass body, as segmented by the AC method. These were the average spiculation measure (AVG), percentage of border pixels with a spiculation measure larger than 45° (PERC_ABV), and the average of the spiculation measure for those pixels with a spiculation measure larger than 45° (AVE_ABV). The remaining five features were related to the information extracted from the thresholded spiculation likelihood map. The first feature was the number of connected objects determined to be spiculations in the spiculation likelihood map (NObj). The second feature was the number of spiculations normalized by the number of border pixels of the mass body (N2E). The third feature was the total area of the suspected spiculations in the spiculation likelihood map normalized by the area of the mass body (SRatio). The fourth feature was the product of NObj and the total area of the suspected spiculations in the spiculation likelihood map normalized by the area of 30-pixel wide band (NSpic2). The last feature was the product of NObj and the third feature (NSpic). The analysis was applied to the DBT slices and PVs.

II.B.2.c. Morphological features. In addition to the texture and spiculation features, 12 morphological features were extracted to characterize the shape and size of the mass. These included the Fourier descriptor, rectangularity, circularity, perimeter-to-area ratio, contrast, perimeter, area, and five features derived from the normalized radial length (NRL) measure (mean, standard deviation, area ratio, zero crossing count, and entropy). The detailed description of these features can be found elsewhere.[24]

II.B.2.d. Feature spaces. In this study, image features were extracted from either the individual PVs or reconstructed DBT slices. Each mass was therefore characterized by multiple sets of features, each obtained from an individual image in one of the approaches. There can be numerous ways to merge the information from the multiple images

TABLE I. Three types of features extracted from the DBT slices or the projection view images after segmentation of the mass.

Run-length statistics texture features	Spiculation features	Morphological features
Short runs emphasis (H, V, 0°, 90°)[a]	Average spiculation measure	Fourier descriptor
Long runs emphasis (H, V, 0°, 90°)	Percentage of border pixels with a spiculation measure > 45°	Rectangularity
Gray level nonuniformity (H, V, 0°, 90°)	Average of the spiculation measure for pixels with a spiculation measure > 45°	Circularity
Run length nonuniformity (H, V, 0°, 90°)	No. of connected objects determined to be spiculations in the SLM[b] (NObj)	Perimeter-to-area ratio
Run percentage (H, V, 0°, 90°)	No. of spiculations normalized by the no. of border pixels of the mass body (N2E)	Contrast
	Total area of the suspected spiculations in the SLM normalized by the area of the mass body (SRatio)	Perimeter
	Product of NObj and the total area of the suspected spiculations in the SLM normalized by the area of 30-pixel wide band (NSpic2)	Area
	Product of NObj and SRatio (NSpic)	NRL[c] Mean
		NRL Standard deviation
		NRL Area ratio
		NRL Zero crossing count
		NRL Entropy
Total no. of features **20**	**8**	**12**

[a]The run-length statistics texture features were extracted from the horizontal (H) and vertical (V) gradient magnitude images, and in two directions (0° and 90°) from each image.
[b]SLM = spiculation likelihood map.
[c]NRL = normalized radial length measures.

within each approach or merge information from the two approaches. However, since the data set was relatively small, we focused on combining the feature information extracted within each approach. The corresponding features extracted from the multiple images were simply combined by averaging to avoid the need to train additional parameters.

In DBT, the spatial resolution in the direction perpendicular to the plane of the DBT slices, or the depth direction, is very low. There is no clear boundary of the mass in the depth direction. As a result, it was not known how many slices should be used for analysis of the mass features. We investigated the dependence of the classification accuracy on the number of slices over which the corresponding features were averaged. The range was varied from one to 19 slices centered about the central slice of the mass, resulting in eight different feature spaces (i.e., one, three, five, seven, nine, 11, 15, and 19 slices). Each feature space contained the same 40 feature descriptors but a feature in a given feature space was obtained by averaging over the specific number of slices for that feature space. For a given mass, if the radiologist-marked VOI enclosing the mass contained fewer than the number of slices on either side of the central slice required for a given feature space, the features would be averaged up to the maximum number of slices on either side in the VOI. Therefore, the number of slices included in the averaging could be asymmetric about the central slice, which was defined in Sec. II A.

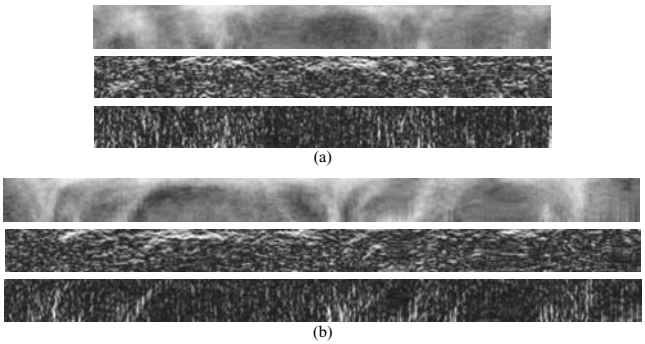


FIG. 4. Transformation of the mass margin to an image in rectangular coordinates using the RBST method. (a) RBST image of mass in PV6 of Fig. 2 and its vertical and horizontal gradient magnitude images. (b) RBST image of mass in S26 of Fig. 3 and its vertical and horizontal gradient magnitude images. In the RBST image, moving along a row approximately corresponds to traversing a closed path at a constant distance from the segmented mass border and moving along a column approximately corresponds to traveling in a direction normal to the segmented mass border in the original image. The radially oriented spiculations are therefore approximately aligned in the vertical direction of the RBST image.

For the PV approach, there were 11 PVs in each scan for this data set. Each PV was acquired with a fraction of the x-ray exposure of a mammogram. In this DBT system, the incident exposure was equally divided among the PVs. However, the primary beam exposure to the detector decreased as the projection angle increased due to the increased path length in the breast tissue. The image quality of the PV images would decrease as the projection angle increased. The dependence of the classification performance on the number

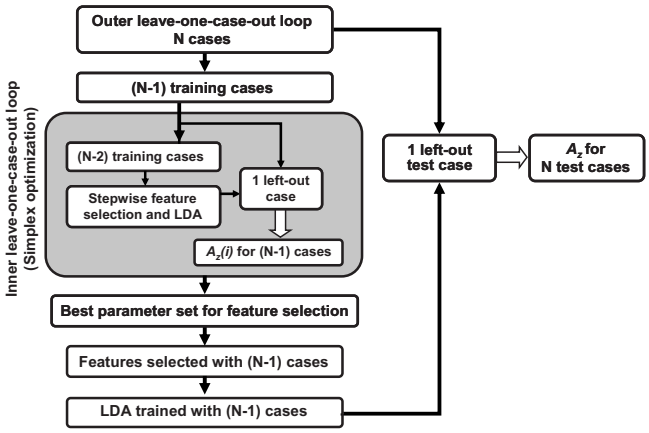


FIG. 5. A two-loop leave-one-case-out scheme for design of classifier.

of PVs over which the corresponding features extracted from the individual PVs were averaged was studied. The range was varied from one to 11 PVs centered about the central PV at 0° incident angle, resulting in six different feature spaces (i.e., one, three, five, seven, nine, and 11 PVs).

II.C. Feature selection and classifier design

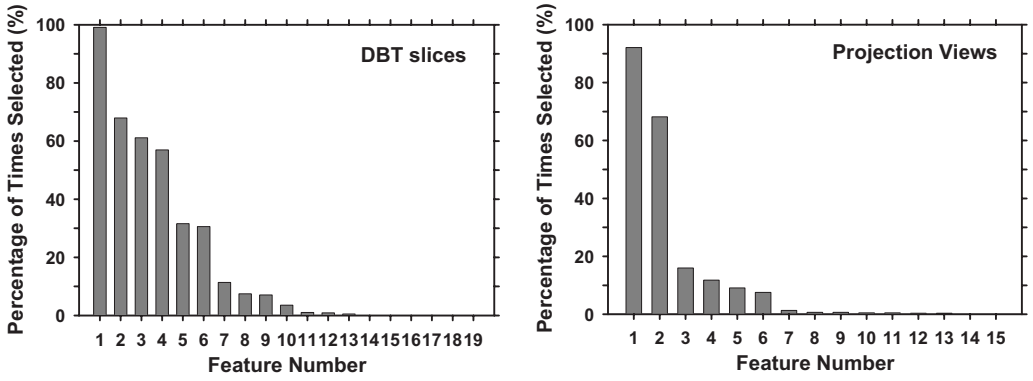
In each feature space, an LDA classifier was trained with stepwise feature selection to find the most effective features from the available feature pool as input predictor variables. The LDA classifier was chosen because we have found in a previous study[29] that the LDA classifier is better generalized to unknown cases than other more complex classifiers when the training sample size is limited. A two-loop leave-one-case-out resampling procedure was designed to train the LDA classifier and test its performance using N available cases,[30] where $N=99$ in the current study. The procedure contained two leave-one-case-out loops as shown in Fig. 5. In each cycle of the "outer" loop, one case including all mass ROIs from the same patient was left out as the independent test case. The remaining $(N-1)$ cases served as the training set to determine the best parameters for stepwise feature selection in an "inner" leave-one-case-out loop and to determine the LDA classifier weights. In each cycle of the inner loop, one case was left out for validation and $(N-2)$ cases were available for training. Simplex optimization was used to automatically search for the best set of parameters (F_{in} and F_{out} thresholds for the F-statistics to determine if a feature should be included or removed from the selected feature pool, and the tolerance threshold on the correlation of the selected features) for the stepwise feature selection procedure. In the simplex search process, for a given set of parameters, stepwise feature selection and LDA were performed with the $(N-2)$ cases and applied to the left-out case for testing; after the inner loop was cycled through the $(N-1)$ cases, the area under the receiver operating characteristic (ROC) curve $A_z(i)$ was estimated. The $A_z(i)$ was used as a figure-of-merit for guiding the simplex search. The parameters that could select a set of features to reach the highest $A_z(i)$ within the search space was taken as the best set of parameters, which was then applied to the entire training set of $(N-1)$ cases to select the features and formulate an LDA classifier. The LDA classifier was tested with the left-out case in the outer loop to obtain a test discriminant score for each mass in the case. The left-out case in the outer loop was not involved in feature selection or LDA formulation and was thus independent of training. The same process was repeated for all N cases in the outer loop in a round-robin manner. After the two-loop resampling procedure was completed, the test scores for all masses in the N cases were collected and evaluated by ROC analysis using the software developed by Metz *et al.*[31] The classification accuracy of the classifier was estimated as the test A_z.

The two-loop leave-one-case-out feature selection and classification procedure was performed in each of the feature spaces for either the PV or the DBT-slice approach. The test ROC curves and the A_z values were then compared.

III. RESULTS

In each feature space, a set of features was selected for each of the outer leave-one-case-out cycle, resulting in 99 sets of features. For the DBT approach, the average number of features selected in the eight feature spaces ranged from 3.46 to 4.00, and the overall average was 3.80 features. The four most consistently selected features included two spiculation features (AVG, NSpic2), one RLS texture feature (short-runs emphasis in the 90° direction extracted from the vertical gradient magnitude image), and one morphological feature (rectangularity). The percentage of times that a given feature was selected relative to the 99 sets of selected features in each feature space and in the eight DBT feature spaces was plotted in Fig. 6(a). The four most frequently selected features corresponded to the four named above. The dependence of A_z on the number of DBT slices used in the feature spaces is shown in Fig. 7. The classification accuracy was relatively constant over the range of three to nine slices ($A_z=0.92$ to 0.93) and fell off when more slices were included in the feature averaging. The central slice was quite efficient in capturing the image information. Although its A_z was about 0.02 lower than the highest value, the difference did not achieve statistical significance. The test ROC curves for the LDA classifiers in the feature spaces using the information from one DBT slice and nine slices are compared in Fig. 8.

The average number of features selected in the PV feature spaces ranged from 1.77 to 2.71, and the overall average was 2.1 features. The two most consistently selected features were both spiculation features (NSpic2, AVG), which were the same as the top two features in the DBT feature spaces although the order was reversed. The percentage of times that a given feature was selected relative to the 99 sets of selected features in each feature space and in the six PV feature spaces was plotted in Fig. 6(b). The two most frequently selected features corresponded to the two named above. The dependence of the classification performance on the number of PVs used in the feature spaces is also shown in Fig. 7. There was an increasing trend of A_z from



(a) (b)

FIG. 6. Histogram of the percentage of times that a given feature was selected, relative to the 99 sets of selected features in each feature space and in all feature spaces for a given approach. The features in the histogram were ordered from the most to the least frequently selected. (a) DBT feature spaces: The maximum number of times a feature would be selected was 99×8 so that 100% corresponded to 792 times. The mean number of features in a selected feature set was 3.8. (b) PV feature spaces: The maximum number of times a feature would be selected was 99×6 so that 100% corresponded to 594 times. The mean number of features in a selected feature set was 2.1.

0.79 ± 0.04 at one PV to 0.84 ± 0.04 at nine PVs and fell off slightly to 0.83 ± 0.04 when the features were averaged over 11 PVs. The difference in A_z between one PV and nine PVs did not achieve statistical significance ($p = 0.28$), however. The test ROC curves for the LDA classifiers in the central PV and nine-PV feature spaces are also compared in Fig. 8. The classification accuracy in the nine-DBT-slice feature space ($A_z = 0.93 \pm 0.02$) was significantly better ($p = 0.006$) than that in the nine-PV feature space ($A_z = 0.84 \pm 0.04$).

IV. DISCUSSION

Studies have shown that the characteristics of mass margins can be visualized much more easily in DBT than in mammography[6,32] The current study showed that the image features extracted from the reconstructed DBT slices could provide significantly higher accuracy than those from the PVs for classification of malignant and benign masses, and the most useful features were those from the mass margin. This result indicates that the reduction in tissue overlap also facilitates machine learning of mass characteristics.

The information content in the reconstructed DBT images is, in principle, the same as that in the set of PVs because postprocessing does not create new information but only helps bring out the information already recorded in the images. The result in this study that the features extracted from the DBT slices are more effective than those from the PVs may be attributed to the accurate fusion of the information in the individual PVs by the tomosynthesis reconstruction process using the known geometry of the DBT system. The reconstruction effectively utilizes the dose from all PVs to reduce the random noise and enhances the correlated structures in the DBT slices, which facilitate image analysis and efficient utilization of the information content. In the PV approach, the image features are extracted from very noisy

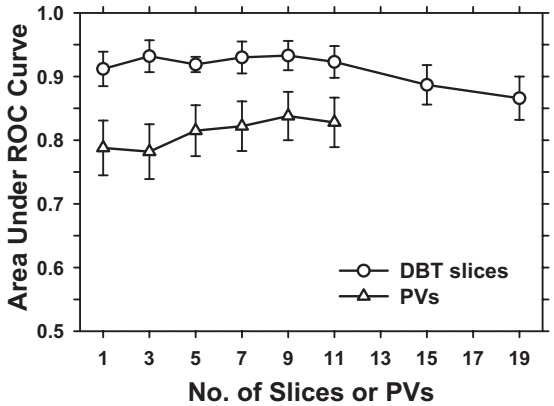


FIG. 7. Dependence of classification accuracy (A_z) of the LDA classifiers on the number of DBT slices or PVs over which the corresponding features were averaged. The classification accuracy using the DBT-slice approach was significantly higher than that using the PV approach.

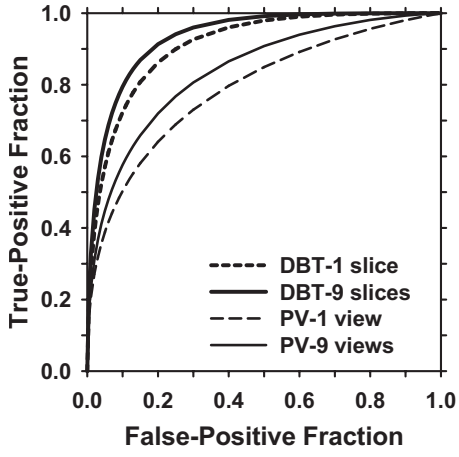


FIG. 8. ROC curves of the LDA classifiers in four of the feature spaces: (1) Features extracted from a single DBT slice intersecting approximately the center of the mass; (2) features extracted from nine DBT slices centered at the central slice and averaged; (3) features extracted from a single PV image at the center of the tomosynthesis scan (0° projection angle); and (4) features extracted from nine PVs centered at the 0° PV and averaged.

images, each of which only utilizes about 9% of the total dose. The subsequent fusion of information by averaging the features extracted from the individual PVs is relatively global and cannot recover the information that may have been masked by the noise in the feature extraction process.

In either the PV or the DBT-slice approach, the averaging of features from multiple PVs or DBT slices improved the classification accuracy over that from a single image although the improvement did not achieve statistical significance. The averaging over multiple images reduces the variance of the features and increases class separation, which is especially important for the PV approach because of the high noise in the individual PVs. This effect can be observed from the stronger dependence of the classification accuracy on averaging over multiple images in the PV approach (Fig. 7). However, for both approaches, the performance eventually decreased when the number of images increased beyond certain values. As demonstrated in Fig. 4, the DBT slices far from the central slice will contain mainly residual density from the mass without detailed structures due to the poor depth resolution in limited-angle reconstruction. Averaging the features extracted from these slices with those from the slices intersecting the main body of the mass will introduce incorrect information and thus degrade the classification performance of the features. This effect is somewhat masked in Fig. 7 because small masses for which the VOI encompassed a small number of slices would reach the maximum A_z performance at a smaller number of slices, whereas large masses would reach the maximum at a larger number of slices. With a data set that contained masses of a range of sizes as in this study, the overall curve would be an average from many curves, each of which had a maximum occurring at a different number of slices so that observed trend might not be as strong as what would be observed if all masses were the same size. It is not practical in this study to train classifiers for subgroups of masses of similar sizes because of the limited number of samples. Nevertheless, the trend in Fig. 7 still indicates that the number of DBT slices of a mass over which feature information is analyzed should be properly adapted to the mass size.

For image analysis performed directly on the PVs, the signal-to-noise ratio in a PV will depend on the incident angle of the x-ray beam to the breast; the larger the angle, the longer the path length in the breast tissue. The longer path length will result in greater beam hardening, more scattered radiation, and less transmitted photons incident on the detector, and thus lower signal-to-noise ratio. The longer x-ray path traversing the breast will also intercept more breast structures, resulting in more complex structured noise. The information from the PVs at large projection angles is thus not as reliable as those from the PVs at small angles. There is a tradeoff, therefore, between reducing the variance and reducing the quality of the feature information when more large-angle PVs are included in the averaging. It is possible that a weighted averaging or more sophisticated fusion method can better utilize information from all PVs. However, these methods will introduce additional parameters to be trained and thus require a larger set of training samples. This and other improvements in the CADx methods will be pursued in future studies.

Because of the relatively limited DBT data set available in this study, we applied mass segmentation and feature extraction algorithms previously developed for digitized film mammograms directly to the DBT slices and PVs without retraining the methods or parameters. The mass segmentation and spiculation extraction algorithms appeared to adapt well to the digital images having different signal and noise characteristics, as demonstrated in the examples shown in Figs. 2 and 3. The overall classification performance in terms of A_z was also reasonably high. The results indicate that CAD algorithms developed with film mammograms can be adapted to digital mammograms and DBT slices without excessive effort if the methods and parameters are designed to be adaptive. We will further optimize the algorithms for each type of image when large data sets become available and compare the classification performance with and without retraining. In addition, we used radiologist-marked VOI as input to the system in this study. The impact of the variations in the VOI and central slice location, either determined by radiologists or by an automated detection algorithm, on the classification performance will have to be investigated in the future.

In this study, we focused on comparison of the PV approach and the DBT-slice approach, and the effectiveness of fusing multiple-image information in each approach. There can be numerous variations to these approaches such as performing the PV and DBT approaches in parallel and then combine the information from the two approaches, or performing image analyses on the PVs, DBT slices, or the 3D volume at different processing steps in combination. Although the information content in the original PV images is the same, the PV and DBT processing analyze the information in different ways; the extracted information may represent different characteristics of the lesions and provide complementary information to improve the overall accuracy, as demonstrated in our previous mass detection study.[18]

Although development of CAD for DBT is similar to that for mammography to a certain extent, the additional 3D information and the flexibility of using different approaches pose greater challenges as well as opportunities for designing computerized lesion detection and characterization algorithms. Many different computer-vision techniques can be designed at each step in each approach. In particular, the design and application of 3D segmentation and feature extraction techniques to the DBT image volume might be a promising approach. In addition, the image quality of DBT depends on the image acquisition parameters such as the tomographic angle and the angular increment in the scan, and the reconstruction technique used for tomosynthesis. The best combination of techniques and parameters will depend not only on the approach taken, but also on the DBT image quality. Studies to date have only explored very limited areas of image analysis in DBT. Much more efforts will be needed to design effective computer-vision methods that can fully exploit the image information in DBT.

ACKNOWLEDGMENTS

This work is supported in part by USPHS Grant Nos. R33 CA120234 and RO1 CA91713 (PI: Paul Carson). The development of the prototype digital breast tomosynthesis system and the collection of the DBT cases were supported by a USAMRMC Grant No. DAMD17-98-1-8309 awarded to the MGH. The content of this paper does not necessarily reflect the position of the funding agencies and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the ROCKIT program provided on the University of Chicago website http://xray.bsd.uchicago.edu/krl/index.htm.

[a]Author to whom correspondence should be addressed. Email: chanhp@umich.edu; Telephone: 734-936-4357.
[b]Present address: Advanced Technology Center, Information and Communications Research Laboratories, Industrial Technology Research Institute (ITRI), Taiwan.
[c]Present address: Hologic, Inc. Massachusetts.
[d]Present address: Microsoft Corp., Redmond, Washington.

[1]L. T. Niklason *et al.*, "Digital tomosynthesis in breast imaging," Radiology **205**, 399–406 (1997).
[2]T. Wu, A. Stewart, M. Stanton, T. McCauley, W. Phillips, D. B. Kopans, R. H. Moore, J. W. Eberhard, B. Opsahl-Ong, L. Niklason, and M. B. Williams, "Tomographic mammography using a limited number of low-dose cone-beam projection images," Med. Phys. **30**, 365–380 (2003).
[3]S. Suryanarayanan, A. Karellas, S. Vedantham, S. P. Baker, S. J. Glick, C. J. D'Orsi, and R. L. Webber, "Evaluation of linear and nonlinear tomosynthetic reconstruction methods in digital mammography," Acad. Radiol. **8**, 219–224 (2001).
[4]J. T. Dobbins and D. J. Godfrey, "Digital x-ray tomosynthesis: Current state of the art and clinical potential," Phys. Med. Biol. **48**, R65–R106 (2003).
[5]Y. Zhang, H.-P. Chan, B. Sahiner, J. Wei, M. M. Goodsitt, L. M. Hadjiiski, J. Ge, and C. Zhou, "A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis," Med. Phys. **33**, 3781–3795 (2006).
[6]M. A. Helvie, M. A. Roubidoux, L. M. Hadjiiski, Y. Zhang, P. L. Carson, and H.-P. Chan, "Tomosynthesis mammography vs conventional mammography: Comparison of breast masses detection and characterization," *RSNA Program Book* (The Radiological Society of North America, Oak Brook, 2007), p. 381.
[7]D. Gur, G. S. Abrams, D. M. Chough, M. A. Ganott, C. M. Hakim, R. L. Perrin, G. Y. Rathfon, J. H. Sumkin, M. L. Zuley, and A. I. Bandos, "Digital breast tomosynthesis: Observer performance study," AJR, Am. J. Roentgenol. **193**, 586–591 (2009).
[8]H. P. Chan, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. S. Gopal, "Improvement of radiologists' characterization of mammographic masses by computer-aided diagnosis: An ROC study," Radiology **212**, 817–827 (1999).
[9]Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, "Improving breast cancer diagnosis with computer-aided diagnosis," Acad. Radiol. **6**, 22–33 (1999).
[10]Z. M. Huo, M. L. Giger, C. J. Vyborny, and C. E. Metz, "Breast cancer: Effectiveness of computer-aided diagnosis—Observer study with independent database of mammograms," Radiology **224**, 560–568 (2002).
[11]L. M. Hadjiiski, H. P. Chan, B. Sahiner, M. A. Helvie, M. Roubidoux, C. Blane, C. Paramagul, N. Petrick, J. Bailey, K. Klein, M. Foster, S. Patterson, D. Adler, A. Nees, and J. Shen, "Improvement of radiologists' characterization of malignant and benign breast masses in serial mammograms by computer-aided diagnosis: An ROC Study," Radiology **233**, 255–265 (2004).
[12]L. M. Hadjiiski, H. P. Chan, B. Sahiner, M. A. Helvie, M. Roubidoux, C. Blane, C. Paramagul, N. Petrick, J. Bailey, K. Klein, M. Foster, S. Patterson, D. Adler, A. Nees, and J. Shen, "Breast masses: Computer-aided diagnosis with serial mammograms," Radiology **240**, 343–356 (2006).
[13]H. P. Chan, J. Wei, B. Sahiner, E. A. Rafferty, T. Wu, M. A. Roubidoux, R. H. Moore, D. B. Kopans, L. M. Hadjiiski, and M. A. Helvie, "Computerized detection of masses on digital tomosynthesis mammograms—A preliminary study," in Proceedings of the Seventh International Workshop on Digital Mammography (IWDM), 2004, pp. 199–202.
[14]I. Reiser, R. M. Nishikawa, M. L. Giger, T. Wu, E. Rafferty, R. H. Moore, and D. B. Kopans, "Computerized detection of mass lesions in digital breast tomosynthesis images using two- and three dimensional radial gradient index segmentation," Technol. Cancer Res. Treat. **3**, 437–441 (2004).
[15]H.-P. Chan, J. Wei, B. Sahiner, E. A. Rafferty, T. Wu, M. A. Roubidoux, R. H. Moore, D. B. Kopans, L. M. Hadjiiski, and M. A. Helvie, "Computer-aided detection system for breast masses on digital tomosynthesis mammograms—Preliminary experience," Radiology **237**, 1075–1080 (2005).
[16]H.-P. Chan, J. Wei, T. Wu, B. Sahiner, E. A. Rafferty, L. M. Hadjiiski, M. A. Helvie, M. A. Roubidoux, R. H. Moore, and D. B. Kopans, "Computer-aided detection on digital breast tomosynthesis (DBT) mammograms: Dependence on image quality of reconstruction," *RSNA Program Book* (The Radiological Society of North America, Oak Brook, 2005), p. 269.
[17]I. Reiser, R. M. Nishikawa, M. L. Giger, T. Wu, E. A. Rafferty, R. H. Moore, and D. B. Kopans, "Computerized mass detection for digital breast tomosynthesis directly from the projection images," Med. Phys. **33**, 482–491 (2006).
[18]H. P. Chan, J. Wei, Y. H. Zhang, M. A. Helvie, R. H. Moore, B. Sahiner, L. Hadjiiski, and D. B. Kopans, "Computer-aided detection of masses in digital tomosynthesis mammography: Comparison of three approaches," Med. Phys. **35**, 4087–4095 (2008).
[19]H. P. Chan, Y. T. Wu, B. Sahiner, Y. Zhang, R. H. Moore, D. B. Kopans, L. M. Hadjiiski, and M. A. Helvie, "Digital breast tomosynthesis mammography: Computerized classification of malignant and benign masses," Med. Phys. **34**, 2645 (2007).
[20]H. P. Chan, Y. T. Wu, B. Sahiner, Y. Zhang, R. H. Moore, D. B. Kopans, M. A. Helvie, L. M. Hadjiiski, and T. Way, "Analysis of mass characteristics on digital breast tomosynthesis (DBT) mammograms: Application to computer-aided diagnosis," *RSNA Program Book* (The Radiological Society of North America, Oak Brook, 2007), p. 315.
[21]G. J. Palma, G. Peters, S. Muller, and I. Bloch, "Masses classification using fuzzy active contours and fuzzy decision trees," Proc. SPIE **6915**, 691509-1–691509-11 (2008).
[22]H. P. Chan, Y. T. Wu, B. Sahiner, Y. H. Zhang, J. Wei, R. H. Moore, D. B. Kopans, M. A. Helvie, L. Hadjiiski, and T. Way, "Digital tomosynthesis mammography: Comparison of mass classification using 3D slices and 2D projection views," Proc. SPIE **6915**, 691506-1–691506-6 (2008).
[23]B. 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).
[24]B. 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).
[25]H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space," Phys. Med. Biol. **40**, 857–876 (1995).
[26]B. 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).
[27]M. M. Galloway, "Texture classification using gray level run lengths," Comput. Graph. Image Process. **4**, 172–179 (1975).
[28]L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. N. Gurcan, "Analysis of temporal change of mammographic features: Computer-aided classification of malignant and benign breast masses," Med. Phys. **28**, 2309–2317 (2001).
[29]H. P. Chan, B. Sahiner, R. F. Wagner, and N. Petrick, "Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers," Med. Phys. **26**, 2654–2668 (1999).
[30]T. W. Way, B. Sahiner, H.-P. Chan, L. Hadjiiski, P. N. Cascade, A. Chughtai, N. Bogot, and E. Kazerooni, "Computer aided diagnosis of pulmonary nodules on CT scans: Improvement of classification perfor-

mance with nodule surface features," Med. Phys. **36**, 3086–3098 (2009).
[31]C. 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).
[32]D. B. Kopans, "Digital tomosynthesis and other applications," in *Categorical Course in Diagnostic Radiology: Breast Imaging-Digital Mammography: RSNA Program Book* (The Radiological Society of North America, Oak Brook, 2005), p. 130.