Computerized detection of breast cancer on automated breast ultrasound imaging of women with dense breasts

Karen Drukker,^{a)} Charlene A. Sennett, and Maryellen L. Giger

Department of Radiology, MC2026, The University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637

(Received 17 June 2013; revised 31 October 2013; accepted for publication 15 November 2013; published 10 December 2013)

Purpose: Develop a computer-aided detection method and investigate its feasibility for detection of breast cancer in automated 3D ultrasound images of women with dense breasts.

Methods: The HIPAA compliant study involved a dataset of volumetric ultrasound image data, “views,” acquired with an automated U-Systems Somo●V[®] ABUS system for 185 asymptomatic women with dense breasts (BI-RADS Composition/Density 3 or 4). For each patient, three whole-breast views (3D image volumes) per breast were acquired. A total of 52 patients had breast cancer (61 cancers), diagnosed through any follow-up at most 365 days after the original screening mammogram. Thirty-one of these patients (32 cancers) had a screening-mammogram with a clinically assigned BI-RADS Assessment Category 1 or 2, i.e., were mammographically negative. All software used for analysis was developed in-house and involved 3 steps: (1) detection of initial tumor candidates, (2) characterization of candidates, and (3) elimination of false-positive candidates. Performance was assessed by calculating the cancer detection sensitivity as a function of the number of “marks” (detections) per view.

Results: At a single mark per view, i.e., six marks per patient, the median detection sensitivity by cancer was 50.0% (16/32) $\pm$ 6% for patients with a screening mammogram-assigned BI-RADS category 1 or 2—similar to radiologists’ performance sensitivity (49.9%) for this dataset from a prior reader study—and 45.9% (28/61) $\pm$ 4% for all patients.

Conclusions: Promising detection sensitivity was obtained for the computer on a 3D ultrasound dataset of women with dense breasts at a rate of false-positive detections that may be acceptable for clinical implementation. © 2014 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4837196>]

Key words: computer-aided detection, 3D automated breast ultrasound, breast cancer, breast density

1. INTRODUCTION

Screening mammography has been shown to reduce breast cancer mortality.^{1,2} A considerable number of studies aiming to quantify breast cancer risk based on mammographic images has been published.^{3–6} While women with “dense breasts” are at higher risk of developing breast cancer, conventional x-ray mammography performance is considerably lower for these women. Although the overall sensitivity of mammography is 70%–90%, the sensitivity can range from 30% to 98% depending on whether the breast density is extremely dense or mostly fatty replaced.⁷ The poor sensitivity of mammography in women with BI-RADS composition/density 3 or 4 has resulted in several states passing legislation requiring women be informed of their breast density and the possible need for additional screening with modalities other than mammography (see <http://www.areyoudenseadvocacy.org/dense/>).

Multiple studies have shown that supplemental whole breast screening ultrasound can increase cancer detection rate from 2.3 to 4.6 cancers per 1000 women screened, especially in women with heterogeneous or extremely dense parenchyma.^{8–11} Screening whole breast ultrasound has advantages and disadvantages when compared to screening

mammography. Sonography is widely available, has no ionizing radiation, is similar in cost to mammography, and is effective in distinguishing cysts from solid masses as well as characterizing benign and malignant masses. Traditional hand held screening breast ultrasound, however, is limited by operator dependency which decreases reproducibility, has a lower specificity and positive predictive value than mammography, and requires a long acquisition time.¹²

Due to the limitations of hand held screening ultrasound, automated whole breast ultrasound equipment has been developed to acquire reproducible, operator independent large volume datasets that can be processed, reviewed, and manipulated on independent workstations. The amount of stored 3D data can be a challenge for radiologists to review and interpret while maintaining an acceptable recall rate, sensitivity, specificity, and positive predictive value. For this reason, computer-aided detection methods may be helpful for automated whole-breast 3D breast ultrasound (and other 3D imaging modalities) in spite of the debated usefulness of such methods for 2D screening x-ray mammography.^{13,14}

Automated 3D breast ultrasound is a relatively new imaging modality (after early attempts decades ago failed because of poor image quality) with the first such system (U-Systems, a GE Healthcare Company) only recently approved by the

FDA as an adjunct screening modality for women with dense breasts. In research settings, computerized lesion detection on 3D whole-breast ultrasound has gained interest in recent years.^{15–17} Past studies concerned themselves with detection of both benign and malignant tumors, whereas here we focus on malignant tumors only. Moreover, these previous studies provided no information regarding breast density and the visibility of tumors on mammography, making a direct comparison to our work difficult. Ikeda *et al.*¹⁵ analyzed a dataset of 109 breasts with a total of 36 masses (including 16 cancers) acquired with an ASU-1004 transducer (Aloka Co, Ltd., Japan), and achieved a sensitivity of 80.6% for all masses at 3.8 false-positives per breast. Using data obtained with this same transducer, Chang *et al.*¹⁶ achieved a sensitivity of 92.3% at 1.76 false-positives per breast, for a dataset of 25 patients with 26 lesions of unspecified pathology and size. The most recent study¹⁷ evaluated detection of both benign and malignant tumors (136 tumors total) in images acquired with the system used in our work, and obtained 70% sensitivity at 2.7 false-positives per view (with the views acquired per breast ranging from 1 to 3).

The purpose of the current study was to develop and investigate a computer-detection method for the detection of breast cancer in women with heterogeneous or extremely dense breast parenchyma imaged with 3D automated breast ultrasound (ABUS). We believe this is currently of great interest, especially given the recent FDA approval of ABUS as an adjunct screening method for women with dense breasts. Moreover, the current study included women for whom breast cancer was deemed occult on their screening x-ray mammograms (XRM) in clinical interpretation [BI-RADS assessment category 1 (negative) or 2 (normal with benign findings)].

2. MATERIALS AND METHODS

2.A. Data set

Image data were obtained on the Somo•V[®] Automated Breast Ultrasound System (ABUS) (U-Systems, a GE Healthcare Company) for 185 asymptomatic women with a mammogram-assigned BI-RADS Composition/Density Category 3 or 4 (heterogeneously dense or extremely dense). The dataset was collected in a HIPAA compliant manner at 13 clinical sites across the United States for a reader study.¹⁸ The reader study was overseen by our Institutional Review Board (IRB). The data collection was performed under a separate multicenter protocol and institution-specific IRBs governed patient enrollment which involved informed consent. The study discussed here had access to deidentified data only.

Of the 185 women included in this study, 52 had breast cancer and 133 did not (Table I). Women with existing benign breast findings were excluded from the noncancer cases. The mean age was 54.0 ± 10.8 and none of the women were BRCA1/2 gene carriers. The cancer cases were biopsy-proven cancers that were diagnosed as a result of findings through any work-up of the patient at most 365 days after the original digital screening mammogram (XRM). More detailed case selection criteria and demographics can be found

TABLE I. Overview of the data set (obtained from a clinical repository study).

	Cancer XRM-assigned BI-RADS 1 or 2 (Group A)	Cancer XRM-assigned BI-RADS 0 (Group B)	Noncancer (Group N)	Total
Patients	31	21	133	185
Cancers	32 ^a	29	...	61
“Views”	186 ^b	126 ^b	798 ^c	1110
“Views” with cancer visible	52	38	...	90

^aOf which 16 cancers were XRM-occult also in retrospective review.

^bIncluded in main analysis (leave-one-case-out).

^cIndependent test set for assessment of false-positive marks.

elsewhere (see <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cftopic/pma/pma.cfm?num=p110006>).

For 31 cancer patients (group A), the clinical XRM-assigned BI-RADS Assessment Category was 1 (negative) or 2 (normal with benign findings), and for 21 cancer patients (group B) the XRM clinically assigned BI-RADS was 0 (needs additional imaging). One patient had multifocal disease in group A and seven patients presented with multifocal cancer in group B, respectively.

The dataset consisted of volumetric image data or “views.” Note that throughout, a “view” is defined as a volumetric (3D) image set containing up to 350 2D images. The ABUS system automatically optimizes image parameters and scans the breast in B-mode ultrasound with a 14–6 MHz transducer that acquires a series of up to 350 2D axial images to a depth up to 5 cm with a 15.4×17 cm wide field of view. Typical resolution was $0.15 \text{ mm} \times 0.3 \text{ mm}$ in the axial image plane of acquisition and 0.6 mm in between axial planes. Although each view contains volumetric image data, multiple views per breast differing in angle of compression and transducer position, are generally obtained to ensure coverage of all breast tissue. In this study, three views per breast were acquired (which was sufficient for all patients) resulting in six views per patient. The angles of acquisition were anterior-posterior, medial, and lateral. Note that a tumor was usually not visible in all three views. For the 31 patients with BI-RADS 1 or 2 (group A), the 32 cancerous tumors were visible in 52 views, and for the 21 patients with BI-RADS 0 (group B), the 29 tumors were visible in 38 views. Each 3D view can be visualized in three orthogonal planes (relative to the transducer orientation): axial, sagittal, and coronal (with the sagittal and coronal images being constructed from the axial images) (Fig. 1).

2.B. Reference standard

Two board-certified “gold-standard” radiologists, each reading >2000 mammograms per year in their practice and with experience in ABUS, independently reviewed all cancer cases (all views on XRM and ABUS). These radiologists independently confirmed the location of malignancies to serve as the gold-standard truth regarding location in each

FIG. 1. Example case of group A: a 61-year-old woman with invasive breast cancer visible in retrospect as an architectural distortion on screening XRM and a 4.5 mm mass on ABUS. In-house developed viewing software depicting the anterior-posterior view of the right breast with the tumor “truth” location marked in axial, sagittal, and coronal planes. The bottom row images are zoomed-in versions of the top row images. Note that truth locations were determined by the “gold-standard” radiologists using a commercially available Somo VIEWer workstation (U-Systems, a GE Healthcare company) and that visualizations provided in this paper are for research purposes only.

applicable view. For 16 of the 31 patients in group A, the cancer was XRM occult also in retrospect. Tumor size was defined as the maximum of the measurements from pathology (per the clinical pathology report), XRM, and ABUS (per the “gold-standard” radiologists). The median tumor size was 12 mm (Fig. 2).

Another board-certified expert radiologist with over 22 years of experience manually outlined each tumor in three orthogonal image planes for each truth location. The latter served as the reference standard for our computerized detection method and all candidate detections within an ellipsoid determined by the bounding box of the radiologist outlines were defined as potential true-positive detections. All other candidates were defined as false-positive detections.

FIG. 2. Tumor size for patients in group A (XRM-assigned BI-RADS of 1 or 2, 31 patients with 32 cancers) and for patients in group B (XRM-assigned BI-RADS of 0, 21 patients with 29 cancers). We failed to find a statistically significant difference in sizes between tumors in groups A and B (t -test $p \gg 0.05$).

2.C. Computerized 3D sonographic image analysis

The computerized image analysis consisted of three steps (Fig. 3). It is important to note that each step was completely automated without human intervention.

2.C.1. Step 1: Detection of initial tumor candidates

The first step was applied on a slice-by-slice basis in the axial plane. Analysis commenced by automatically identifying image regions corresponding to the lungs and regions outside the breast (air) to eliminate them from further analysis. Adaptive gray level thresholding was used to identify the breast border and designate the region inside the border as imaged breast anatomy. The value for the gray level threshold varied by axial slice and was empirically determined, based on a subset of ten views from different patients. The threshold value was subsequently defined as 85% of the maximum gray level for a given axial slice.

Because tumors can have substantially different appearances on breast ultrasound—e.g., anechoic, hypoechoic, and hyperechoic—the use of a fixed gray level threshold to identify candidate tumors is not a viable approach. Hence, we used an unsupervised k-means clustering technique¹⁹ with three clusters to identify image regions as “bright” (parenchyma), “dark” (such as the nipple), and “candidate regions.” The latter refers to the regions where cancers, if any, tend to occur. Subsequently, the locations of tumor candidates, i.e., seed-points, were determined as the geometric centers and morphological skeletal branch-points of the candidate regions (matlab[®] image processing toolbox, The Mathworks, Inc.). For the determination of branch-points *only*, each slice was

Material and methods

Method details

185 women w. dense breasts
 133 normal
 52 breast cancer
 3 views per breast
 1110 views total

Unsupervised technique
 Slice-by-slice (2D)

3D

Supervised technique

View
 (3D whole-breast US image)

Step 1

Detection of initial tumor
 candidates

Step 2

Characterization of candidates

Step 3

Elimination of false-positive
 candidates

Pre-selected number of 'marks' (detections) per view

Prepare: eliminate lungs/air
 Identify: candidate regions
 Identify: candidate 'seed-points'

Segment: gray-value based 3D segmentation
 Characterize: 3D and 2D features

Classify: Bayesian Neural Net
 Assess: Leave-one-case-out cross-validation (*)
 Assess: Detection sensitivity for number of marks per view

FIG. 3. Overview of the fully automated scheme for detection of breast cancer on 3D ABUS images. *Step 3: Leave-one-case-out cross-validation for image data of the 52 cancer patients only [with a case being a patient (six 3D views)]. Views of normal patients were used as an independent test set.

subsampled to a resolution of $2 \text{ mm} \times 2 \text{ mm}$. The locations of branch-points were subsequently projected back onto the original coordinate space. The center points of candidate regions were determined using the original image data to ensure preservation of small candidate regions.

2.C.2. Step 2: Characterization of candidates

Although the initial identification of candidate regions was performed on 2D slices, characterization of the tumor candidates was conducted in three dimensions. The seed-points representing the initial tumor candidates were used to initiate 3D candidate segmentation. Adaptive gray-level thresholding was used to segment the tumor candidates from the surrounding tissue. Being adaptive, the threshold value varied for each candidate seed-point, and for a given seed-point it was defined as the minimum gray value of a small 2-mm-cubed region around that seed-point. Subsequently, the corresponding tumor candidate was defined as the continuous region containing the seed-point resulting from thresholding within the entire 3D image. Other regions, if any, not containing the given seed point were discarded. Each segmented tumor candidate was characterized by 18 mathematical descriptors, i.e., features. These features (Table II) described candidate morphology and echogenicity, and were designed based on both the expected diagnostic merit and the need for efficient calculations given a very large number of seed-points.

2.C.3. Step 3: Elimination of false-positive candidates

In the third step, the tumor candidate features served as input to a Markov-Chain Monte Carlo Bayesian Neural Net (MCMC-BNN) classifier²⁰ with 16 hidden units using all features conjunctively in the elimination of false-positive candidates. The classifier was trained, tested, and evaluated in the task of distinguishing between true-positive detections (seed-points representing actual cancer) and false-positive detections (seed-points representing “everything else”). In this step, classifier *training* was performed using data from the cancer patients only. For cancer patients, classifier training and testing was performed using leave-one-case-out cross-validation. (Here, “case” refers to a patient, i.e., candidate features from six views.) For normal patients, classifier training was performed using data from the cancer patients with the data from normal patients forming an independent test set. Note that normal patients were included in the analysis to gain insight into the nature of false-positive detections in normal views.

The fraction of candidate seed-points representing actual cancers was expected to be very small, hampering efficient classifier training because of the predicted overwhelming number of false-positive detections. Hence, in classifier *training*, features from 500 false-positive candidate seed-points per view were randomly selected as being representative. During classifier testing, however, features from *all* candidate seed-points were classified using the trained classifier. The random selection of features representing false-positive

TABLE II. Features (mathematical descriptors) used in the elimination of false-positive candidates.

Type	D ^a	Plane	Feature category	
			Feature	Definition
Morphology	3D	N/A	Volume	Size
			Sphericity	Conformity of volume to spherical shape
			Irregularity	Deviation of surface from spherical surface
			RGI ^b	Margin sharpness and shape
Echogenicity	3D	N/A	Mean	Average appearance
			Variance	Uniformity of appearance
			Skewness	Asymmetry of gray-level distribution
Morphology	2D	Axial	Depth/width	Orientation with respect to skin
			Depth/width	Orientation with respect to skin
		Sagittal	Area	Size in coronal plane
			Relative area	Size with respect to bounding rectangle
		Coronal	Centroid shift	Deviation from circular shape
			Radius	Effective radius of bounding circle
			Circularity	Conformity to circular shape
			Mean	Average appearance in coronal plane
Echogenicity	2D	Coronal	Variance	Uniformity of appearance
			Skewness	Asymmetry of gray level distribution
			Entropy	Randomness of gray-values

^aD = Dimensionality.^bRGI = Radial gradient index.

seed-points for classifier training was repeated 40 times (resulting in 40 leave-one-case-out cross-validation analyses for the cancer patients and 40 independent tests for the normal patients).

In classifier testing, for each view all seed-points were ordered by decreasing likelihood of being a true-positive detection based on the classifier output. The computerized detection method was designed to operate at a preselected number of detections, “marks,” per *view* with marks being placed in decreasing likelihood of being a true-positive, i.e., the seed-point with the highest classifier output for a given view receiving the first mark. Subsequently, cancer detection sensitivity was calculated as a function of the number of marks per view (range 1–25 marks per view). We will highlight performance at two operating points: a single mark per view and three marks per view served as example operating points since the resulting number of false-positive detections per view (and patient) need to be low enough to be clinically feasible. A true-positive was defined as a physical cancer represented by one (or more) marks. Note that the computer detection sensitivity was assessed by cancer, differing from a by-patient evaluation only for the few patients with multifocal disease. The number of marks per patient equaled six times the number of marks per view based on the availability of six views per patient, and, by design, normal “healthy” patients received the same number of marks as cancer patients.

3. RESULTS

3.A. Step 1: Detection of initial tumor candidates

For all 1110 3D views (groups A+B+N), the average number of candidate seed-points identified per view

was 6102 [range (2534–10 186), 95% confidence interval (2872;10 022)] [Fig. 4(b)]. On a by-view basis, there were two instances in which an imaged cancer was missed. In both instances, however, the cancer was detected in another view, thus resulting in a by-patient cancer detection sensitivity of 100% at the completion of Step 1.

3.B. Step 2: Characterization of candidates

In general, true-positives were more irregular in shape (difference in relative area $\Delta = -0.32$), smaller (difference in mean area $\Delta = -11.3 \text{ mm}^2$), less circular ($\Delta = -0.05$), with a smaller equivalent radius ($\Delta = -0.23 \text{ mm}$) in the coronal plane than false-positives.

3.C. Step 3: Elimination of false-positive candidates

For each view, marks (detections) were placed on the seed-points in order of decreasing classifier output, with the seed-point corresponding to the highest classifier output receiving the first mark [Fig. 4(c)]. When applied to normal views all marks were, by definition, false-positive detections and were largely caused by small areas of shadowing due to tissue interfaces, such as the edge of a rib [Figs. 5(b) and 6(a)] or the interface between echogenic parenchyma and isoechoic fat [Figs. 5(a) and 5(c)]. Other lesser causes of false-positive detections were normal structures in the breast such as fat lobules, as well as imaging artifacts such as shadowing caused by suboptimal transducer contact with the skin [Figs. 5(a) and 5(b)]. The nature of false-positive detections in normal views and those with pathology were found to be similar in nature with many occurring along the chest wall immediately adjacent to ribs [Figs. 5(b) and 6(a)]. Ribs themselves were not a

FIG. 4. An axial image-plane (348×568 pixels) in the anterior-posterior view ($348 \times 568 \times 278$ voxels) of the right breast of the example case (Fig. 1): (a) image-plane, (b) Step 1: image with candidate seed-points superimposed (white), and (c) Step 3: image with the candidate seed-point with the highest classifier output of all 7901 seed-points for this view and hence receiving the first mark (true-positive detection, white circle, classifier output $P = 0.91$) and candidates that would be eliminated at any reasonable value for the desired number of marks per view (black circles, with the next-highest classifier output for this image being $P = 0.06$).

major cause of false-positives due to their distinctive shape. Nipples generally did not cause many false-positive marks, since these tend to appear darker than the candidate regions as identified by the k -means clustering method (with the clusters being “dark,” “bright,” and “candidate,” see Sec. 2).

For a given number of marks per view (range 1–25 marks per view), the computerized cancer detection was found to

FIG. 5. Three axial image planes of a normal anterior-posterior view for a 64-year-old patient and the false-positive marks placed by the automated detection method ranking 1, 2, and 3 (from top to bottom) for this 3D view with classifier outputs $P = 0.82$, 0.68 , and 0.65 , respectively. Some “landmarks” are indicated in (b).

be able to identify small cancers [Fig. 6(b)] and to perform similarly for patients with XRM-assigned BI-RADS Assessment Category 1 or 2 (group A, Fig. 7) and XRM-assigned BI-RADS 0 (group B). At a single mark per view, i.e., six marks per patient, the median detection sensitivity by cancer was $50.0\% (16/32) \pm 6\%$ (standard deviation) for patients with XRM-assigned BI-RADS category 1 or 2 (group A), and $45.9\% (28/61) \pm 4\%$ overall (groups A+B). At 3 marks per view, these numbers were $68.8\% (22/32) \pm 5\%$ (group A) and $63.9\% (39/61) \pm 3\%$ (groups A+B), respectively. Note again

FIG. 6. Example images of the anterior-posterior view of the left breast of a 67-year-old patient with a 5.4 mm XRM-occult invasive node-negative cancer, with (a) axial and coronal images of the location receiving the first mark for this view (false-positive detection, black circle, classifier output $P = 0.67$) and (b) axial and coronal images of location receiving the second mark (true-positive detection, white circle, classifier output $P = 0.58$).

FIG. 7. Box-plot of the detection sensitivity as a function of the number of final “marks” per view for patients in group A (31 patients, 186 views total) obtained from 40 leave-one-case-out analyses with randomly selected false-positives in classifier training. On each box, the central mark is the median, and the edges denote the 25%–75% percentile with “whiskers” extending to the extremes not considered outliers and outliers marked individually.

that, by design, all views, including “normal” views, received the same number of marks in this method. Hence, the number of marks for a breast cancer patient in group A or B was the same as for a patient in group N (“normals”), with for the latter all marks being false-positives.

4. DISCUSSION

Novel aspects of the sonographic computerized detection method presented here are the combined use of 2D and 3D tumor descriptors and the use of randomly selected false-positive candidates in classifier training. In addition, we applied our method to a clinically relevant and challenging data set of ABUS images.

While the number of tumor candidates (seed-points) obtained in the first step of our method may seem large (6102 per view, on average), it is substantially smaller than what would have been obtained through the placement of a $2\text{ mm} \times 2\text{ mm} \times 2\text{ mm}$ uniform grid inside the candidate regions. For the latter, an average of 170 thousand seed-points per view would have been obtained.

It is often stated²¹ that the coronal plane correlates better with breast anatomy and it is interesting to note that, in the second step, the four individual features with the best performance in the task of distinguishing between true-positive and false-positive detections were all obtained from the 2D coronal plane, possibly because this tends to be the plane in which spiculations are best visible. The clinical significance of our findings is unclear, however, and warrant further investigation.

We aimed to prevent the introduction of database bias given the relatively modest number of cancer patients in the dataset, and we therefore purposely did not exhaustively investigate different approaches (such as different classifiers), classifier architecture, parameter settings (such as the number of randomly selected false-positive seed-points

in classifier training), or feature selection. The choices we made here in this respect were “educated guesses” based on our extensive experience in breast image analysis using different classifiers^{22,23} and dimension reduction techniques.²² After obtaining our initial results presented here, we did investigate whether a slightly more “sophisticated” candidate segmentation method, using a Gaussian constraining function in the candidate segmentation step²⁴ would result in a better discrimination between false-positive and true-positive candidates. This, however, did not have the intended effect and instead substantially deteriorated performance. In the future, it may be possible to obtain better performance through further optimization using a larger dataset.

All cases used in the current study were retrospectively selected from a prospective clinical repository study and those from group A and group N were used in an associated retrospective reader study^{18,25} (Table I). The sequential-design reader study involved 17 radiologist observers (readers) who first interpreted each case using XRM on a Hologic Securview DX (Hologic, Inc) workstation, followed by the interpretation of given case on ABUS using a SomoVIEWer workstation (U-Systems, a GE Healthcare Company). Cases were interpreted *without* computer-aided detection for XRM (or for ABUS) but “oversight-error” was expected to be minimal in the cancer-enriched observer study setting. The impact of ABUS interpretation on radiologist performance was assessed including only cancer cases with XRM-assigned BI-RADS assessment category of 1 or 2 and “normal” cases, and a statistically significant improvement in readers’ performance was demonstrated in detecting mammographically occult cancers with a combination of XRM and ABUS as compared to XRM alone. The location-specific sensitivity at a BI-RADS cut-off of 4a for all readers after ABUS interpretation was 49.9% (range 32.3%–67.7% for individual readers) with a specificity of 84.0% (range 67.7%–92.5%).

In comparison, our computerized detection method yielded, at a single mark per view, a median sensitivity of 50.0% (16/32) (Fig. 7) which was similar to the location-specific sensitivity obtained on average by the 17 radiologists in the reader study (49.9%) for this dataset (group A+N). Five cancers were missed by all readers, of which four were visible in a single view only (e.g., Fig. 1) and one was visible in 2 views. At 3 marks per view, our computerized detection method identified 3 out of the 5 (60.0%) cancers missed by all 17 readers. Of the 11 cancers missed by more than 80% of the readers (14/17)—including the five missed by all—six (54.5%) were detected by the computer at 3 marks per view. Of the two example cases of patients with small node-negative invasive cancers, one was missed by all readers and detected at a single mark per view with our method (Fig. 4), and the other one was correctly identified on ABUS by 10 of the 17 readers (without any computer aid) and detected at 2 marks per view by our method (Fig. 6). Note that the reader study was completely independent from the work presented here, and that the potential impact of our computerized detection method on reader performance in the reader study cannot be determined.

There were several limitations to this pilot study. The first limitation was that we assessed the stand-alone performance of the proposed computerized detection method and its potential impact as a computer-aided detection system on radiologist performance remains to be investigated. The second limitation was that, by design, a fixed number of marks per view was generated, also for “normal” views without pathology. Because of the heterogeneous appearance of breast cancer and interpatient variability in conspicuity (as illustrated in the example figures), we designed our method to be relatively unaffected by differences in lesion appearance (including echogenicity) and differences in image acquisition settings. Had the intention been to detect “larger” hypoechoic lesions, rather than—preferably small and not necessarily anechoic—cancers, the method could have been not only more straightforward but also much less useful since one generally does not need the help of a computer to detect an obvious hypoechoic lesion. Instead, k-means clustering was used to identify “indeterminate” (neither very “dark” nor “bright”) regions for extraction of candidate seed-points. Since k-means clustering is an unsupervised method, the appearance of “landmarks” such as the nipple and Coopers ligaments determined what was “dark” and “bright.” The steps after candidate segmentation (characterization and elimination of false-positive candidates), however, involved supervised methods, which were inherently dataset-based (such as classifier training and testing) rather than view- or image-based. This resulted in substantial differences in classifier output values (scores) obtained for different 3D views. For example, the highest score for the 3D view in Fig. 4 was 0.91 (a true-positive detection) and that for the 3D view in Fig. 6 was 0.67 (a false-positive detection). A seed-point within the small invasive cancer in the latter view received the second-highest score for that view of 0.58. A data set-wide threshold value for the classifier output of 0.58 would have resulted in unacceptably large numbers of false-positives in many other views such as the one in Fig. 4. Hence, we opted for a view-based approach at the cost of placing a predetermined number of marks in each and every view. How many marks per view (and hence per patient), trading off sensitivity and number of false-positive marks, would be feasible in clinical practice is open to debate. The fact that none of the normal patients received no false-positive mark could be seen as a serious drawback by some, but we would like to point out that the human detection task is much more complicated here than, e.g., for 2D x-ray mammography, in part because of the vast amount of 3D data to be analyzed. Whether a fixed number of marks per views would in practice be a serious drawback also depends on how easily the false-positive marks generated by our method can be rejected by radiologists (Fig. 5). One should also note that within the current FDA approval of ABUS as an adjunct screening modality to mammography, the fraction of healthy patients is limited to women with dense breasts (although the assessment of breast density is somewhat subjective in spite of the BI-RADS lexicon).

A third limitation was that we used only image data obtained with equipment from a single vendor. In previous work, however, we demonstrated robustness of image analysis for

hand-held ultrasound with respect to acquisition equipment,²⁶ image selection,²⁷ population,²⁸ classifiers,²⁹ and dimension reduction techniques.²² Although, as noted above, our method may be further optimized by making it more robust with respect to lesion appearance and image acquisition, overall it is expected that variations in image quality will have a lesser impact for automated 3D breast ultrasound than for hand-held ultrasound since the imaging technique is much more reproducible and image acquisition protocols are more standardized.

The computerized method obtained promising performance for cancer detection in 3D automated breast ultrasound for women with dense breasts. It operated at a level of marks per view (and hence per patient) that may be practical for future clinical applications while maintaining sensitivity at a level comparable to radiologist observers for this dataset.

ACKNOWLEDGMENTS

M.L.G. is a stockholder in R2 Technology/Hologic, co-founder of and stockholder in Quantitative Insights, Inc., and receives royalties from Hologic, GE Medical Systems, MEDIAN Technologies, Riverain Medical, Mitsubishi, and Toshiba. K.D. received royalties from Hologic. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interest that would reasonably appear to be directly and significantly affected by the research activities.

^a)Electronic mail: kdrukker@uchicago.edu

¹K. Bock, B. Borisch, J. Cawson, B. Damtjerna, C. de Wolf, P. Dean, A. den Heeten, G. Doyle, R. Fox, A. Frigerio, F. Gilbert, G. Hecht, W. Heindel, S. H. Heywang-Kobrunner, R. Holland, F. Jones, A. Lernevall, S. Madaai, A. Mairs, J. Muller, P. Nisbet, A. O'Doherty, J. Patnick, N. Perry, L. Regitz-Jedermann, M. Rickard, V. Rodrigues, M. R. Del Turco, A. Scharpantgen, W. Schwartz, B. Seradour, P. Skaane, L. Tabar, S. Tornberg, G. Ursin, E. Van Limbergen, A. Vandenbroucke, L. J. Warren, L. Warwick, M. Yaffe, and M. Zappa, “Effect of population-based screening on breast cancer mortality,” *Lancet* **378**, 1775–1776 (2011).

²L. Tabár *et al.*, “Reduction in mortality from breast cancer after mass screening with mammography: Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare,” *Lancet* **325**, 829–832 (1985).

³J. J. Heine, C. G. Scott, T. A. Sellers, K. R. Brandt, D. J. Serie, F. F. Wu, M. J. Morton, B. A. Schueler, F. J. Couch, J. E. Olson, V. S. Pankratz, and C. M. Vachon, “A novel automated mammographic density measure and breast cancer risk,” *J. Natl. Cancer Inst.* **104**, 1028–1037 (2012).

⁴H. Li, M. L. Giger, L. Lan, J. Bancroft Brown, A. MacMahon, M. Mussman, O. I. Olopade, and C. Sennett, “Computerized analysis of mammographic parenchymal patterns on a large clinical dataset of full-field digital mammograms: Robustness study with two high-risk datasets,” *J. Digit. Imaging* **25**, 591–598 (2012).

⁵H. Li, M. L. Giger, O. I. Olopade, and L. Lan, “Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment,” *Acad. Radiol.* **14**, 513–521 (2007).

⁶H. Li, M. L. Giger, Z. Huo, O. I. Olopade, L. Lan, B. L. Weber, and I. Bonta, “Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: Effect of ROI size and location,” *Med. Phys.* **31**, 549–555 (2004).

⁷M. T. Mandelson, N. Oestreicher, P. L. Porter, D. White, C. A. Finder, S. H. Taplin, and E. White, “Breast density as a predictor of mammographic detection: Comparison of interval- and screen-detected cancers,” *J. Natl. Cancer Inst.* **92**, 1081–1087 (2000).

- ⁸W. Buchberger, A. Niehoff, P. Obrist, P. DeKoekkoek-Doll, and M. Dunser, "Clinically and mammographically occult breast lesions: Detection and classification with high-resolution sonography," *Semin Ultrasound CT MR* **21**, 325–336 (2000).
- ⁹W. A. Berg, Z. Zhang, D. Lehrer, R. A. Jong, E. D. Pisano, R. G. Barr, M. Bohm-Velez, M. C. Mahoney, W. P. Evans, 3rd, L. H. Larsen, M. J. Morton, E. B. Mendelson, D. M. Farria, J. B. Cormack, H. S. Marques, A. Adams, N. M. Yeh, and G. Gabrielli, "Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk," *J. Am. Med. Assoc.* **307**, 1394–1404 (2012).
- ¹⁰P. Crystal, S. D. Strano, S. Shcharynski, and M. J. Koretz, "Using sonography to screen women with mammographically dense breasts," *Am. J. Roentgenol.* **181**, 177–182 (2003).
- ¹¹R. J. Hooley, K. L. Greenberg, R. M. Stackhouse, J. L. Geisel, R. S. Butler, and L. E. Philpotts, "Screening US in patients with mammographically dense breasts: Initial experience with Connecticut Public Act 09-41," *Radiology* **265**, 59–69 (2012).
- ¹²K. M. Kelly and G. A. Richwald, "Automated whole-breast ultrasound: Advancing the performance of breast cancer screening," *Semin Ultrasound CT MR* **32**, 273–280 (2011).
- ¹³J. J. Fenton, G. Xing, J. G. Elmore, H. Bang, S. L. Chen, K. K. Lindfors, and L. M. Baldwin, "Short-term outcomes of screening mammography using computer-aided detection: A population-based study of medicare enrollees," *Ann. Intern. Med.* **158**, 580–587 (2013).
- ¹⁴R. M. Nishikawa, J. J. Fenton, and C. G. Orton, "Point/counterpoint: Computer-aided detection should be used routinely to assist screening mammogram interpretation," *Med. Phys.* **39**, 5305–5307 (2012).
- ¹⁵Y. Ikeda, D. Fukuoka, T. Hara, H. Fujita, E. Takada, T. Endo, and T. Morita, "Development of a fully automatic scheme for detection of masses in whole breast ultrasound images," *Med. Phys.* **34**, 4378–4388 (2007).
- ¹⁶R. F. Chang, K. C. Chang-Chien, E. Takada, C. S. Huang, Y. H. Chou, C. M. Kuo, and J. H. Chen, "Rapid image stitching and computer-aided detection for multipass automated breast ultrasound," *Med. Phys.* **37**, 2063–2073 (2010).
- ¹⁷W. Moon, Y. Shen, M. Bae, C. Huang, J. Chen, and R. Chang, "Computer-aided tumor detection based on multi-scale blob detection algorithm in automated breast ultrasound images," *IEEE Trans. Med. Imaging* **32**, 1191–1200 (2012).
- ¹⁸M. L. Giger, D. P. Miller, J. Bancroft Brown, M. F. Inciardi, C. E. Metz, Y. Jiang, R. Brem, R. M. Nishikawa, A. Edwards, and J. Papaioannou, "Clinical reader study examining the performance of mammography and automated breast ultrasound in breast cancer screening," *98th Assembly and Annual Meeting of Radiological Society of North America* (Chicago, IL, 2012).
- ¹⁹C. A. F. Seber, *Multivariate Observations* (John Wiley and Sons, Inc., Hoboken, NJ, 1984).
- ²⁰C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Mach. Learn.* **50**, 5–43 (2003).
- ²¹T. Tan, B. Platel, H. Huisman, C. I. Sanchez, R. Mus, and N. Karssemeijer, "Computer-aided lesion diagnosis in automated 3D breast ultrasound using coronal spiculation," *IEEE Trans. Med. Imaging* **31**, 1034–1042 (2012).
- ²²A. R. Jamieson, M. L. Giger, K. Drukker, H. Li, Y. Yuan, and N. Bhooshan, "Exploring nonlinear feature space dimension reduction and data representation in breast Cdx with Laplacian eigenmaps and t-SNE," *Med. Phys.* **37**, 339–351 (2010).
- ²³A. R. Jamieson, M. L. Giger, K. Drukker, and L. L. Pesce, "Enhancement of breast CADx with unlabeled data," *Med. Phys.* **37**, 4155–4172 (2010).
- ²⁴M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Trans. Med. Imaging* **17**, 510–517 (1998).
- ²⁵K. Drukker, K. J. Horsch, L. L. Pesce, and M. L. Giger, "Interreader scoring variability in an observer study using dual-modality imaging for breast cancer detection in women with dense breasts," *Acad. Radiol.* **20**, 847–853 (2013).
- ²⁶K. Drukker, M. L. Giger, and C. E. Metz, "Robustness of computerized lesion detection and classification scheme across different breast US platforms," *Radiology* **237**, 834–840 (2005).
- ²⁷N. P. Grusauskas, K. Drukker, M. L. Giger, C. A. Sennett, and L. L. Pesce, "Performance of breast ultrasound computer-aided diagnosis: Dependence on image selection," *Acad. Radiol.* **15**, 1234–1245 (2008).
- ²⁸N. P. Grusauskas, K. Drukker, M. L. Giger, R. F. Chang, C. A. Sennett, W. K. Moon, and L. L. Pesce, "Breast US computer-aided diagnosis system: Robustness across urban populations in South Korea and the United States," *Radiology* **253**, 661–671 (2009).
- ²⁹K. Drukker, L. Pesce, and M. Giger, "Repeatability in computer-aided diagnosis: Application to breast cancer diagnosis on sonography," *Med. Phys.* **37**, 2659–2669 (2010).