Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation

Brad M. Keller[a] and Diane L. Nathan
Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Yan Wang
Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Yuanjie Zheng, James C. Gee, Emily F. Conant, and Despina Kontos
Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 24 November 2011; revised 18 May 2012; accepted for publication 14 June 2012; published 25 July 2012)

Purpose: The amount of fibroglandular tissue content in the breast as estimated mammographically, commonly referred to as breast percent density (PD%), is one of the most significant risk factors for developing breast cancer. Approaches to quantify breast density commonly focus on either semiautomated methods or visual assessment, both of which are highly subjective. Furthermore, most studies published to date investigating computer-aided assessment of breast PD% have been performed using digitized screen-film mammograms, while digital mammography is increasingly replacing screen-film mammography in breast cancer screening protocols. Digital mammography imaging generates two types of images for analysis, raw (i.e., "FOR PROCESSING") and vendor postprocessed (i.e., "FOR PRESENTATION"), of which postprocessed images are commonly used in clinical practice. Development of an algorithm which effectively estimates breast PD% in both raw and postprocessed digital mammography images would be beneficial in terms of direct clinical application and retrospective analysis.

Methods: This work proposes a new algorithm for fully automated quantification of breast PD% based on adaptive multiclass fuzzy c-means (FCM) clustering and support vector machine (SVM) classification, optimized for the imaging characteristics of both raw and processed digital mammography images as well as for individual patient and image characteristics. Our algorithm first delineates the breast region within the mammogram via an automated thresholding scheme to identify background air followed by a straight line Hough transform to extract the pectoral muscle region. The algorithm then applies adaptive FCM clustering based on an optimal number of clusters derived from image properties of the specific mammogram to subdivide the breast into regions of similar gray-level intensity. Finally, a SVM classifier is trained to identify which clusters within the breast tissue are likely fibroglandular, which are then aggregated into a final dense tissue segmentation that is used to compute breast PD%. Our method is validated on a group of 81 women for whom bilateral, mediolateral oblique, raw and processed screening digital mammograms were available, and agreement is assessed with both continuous and categorical density estimates made by a trained breast-imaging radiologist.

Results: Strong association between algorithm-estimated and radiologist-provided breast PD% was detected for both raw ($r = 0.82$, $p < 0.001$) and processed ($r = 0.85$, $p < 0.001$) digital mammograms on a per-breast basis. Stronger agreement was found when overall breast density was assessed on a per-woman basis for both raw ($r = 0.85$, $p < 0.001$) and processed (0.89, $p < 0.001$) mammograms. Strong agreement between categorical density estimates was also seen (weighted Cohen's $\kappa \geq 0.79$). Repeated measures analysis of variance demonstrated no statistically significant differences between the PD% estimates ($p > 0.1$) due to either presentation of the image (raw vs processed) or method of PD% assessment (radiologist vs algorithm).

Conclusions: The proposed fully automated algorithm was successful in estimating breast percent density from both raw and processed digital mammographic images. Accurate assessment of a woman's breast density is critical in order for the estimate to be incorporated into risk assessment models. These results show promise for the clinical application of the algorithm in quantifying breast density in a repeatable manner, both at time of imaging as well as in retrospective studies. © *2012 American Association of Physicists in Medicine*. [http://dx.doi.org/10.1118/1.4736530]

Key words: digital mammography, breast density, breast cancer risk estimation, quantitative imaging

I. INTRODUCTION

Breast cancer, which is estimated to affect one in eight women over the course of their lives, is the most commonly diagnosed cancer in women and is the second leading cause of cancer death in women in the United States after lung cancer.[1] Work by Gail *et al.* has shown that several factors are associated with an increased risk for developing breast cancer, such as current age, age at menarche, age at first live birth, and number of first-degree relatives with breast cancer.[2] Knowledge of individual risk is critical for the creation of tailored screening recommendations[3–5] and to determine if preventative strategies might be initiated.[6]

Beginning with the pioneering work of Wolfe,[7,8] multiple studies have established that the relative amount of fibroglandular tissue in the breast as estimated mammographically, often referred to as breast percent density (PD%), is an image-derived metric that has been shown to be an independent risk factor for breast cancer, in fact the most significant after age.[9] Currently, the most commonly used methods to assess breast density rely either on subjective, visual assessment by radiologists in distinct categories[10] or through interactive, semiautomated image thresholding.[11] Categorical assessment, such as with the American College of Radiology (ACR) 4-class breast-imaging reporting and data systems (BIRADS) system[10] (Fig. 1), has been reported to have good agreement between readers,[12] with reported Cohen's weighted kappa[13] coefficients ranging between $\kappa = 0.61$ and $\kappa = 0.84$.[11,14] Categorical methods for the assessment of density are known to have relatively lower levels of agreement[15] when assessing moderately dense breasts versus completely fatty or dense breasts. Interactive image-thresholding methods, which require user interaction, are known to be both time consuming and susceptible to reader-variability,[16] limiting translation of such tools into routine clinical workflow.

The majority of automated algorithms presented in the literature to date for the assessment of breast density have been developed for digitized screen-film mammograms (SFM).[17–22] However, digital mammography (DM) is increasingly replacing film mammography in breast cancer screening.[23] Digital mammography images could have a number of different image characteristics compared to SFM images that may need to be considered when adapting techniques developed on SFM to DM. For example, digital mammography is known to have a higher dynamic range than screen-film mammography.[24] This in turn results in higher contrast and a richer gray-level intensity profile of the breast tissue that does not necessarily fulfill the unimodal or bimodal profile assumptions often made for digitized film mammography.[11,21] Another consideration is that the inherent granularity of the film used in screen-film mammography is absent in digital mammography, leading to different noise properties between the two image types.[25] Developing a method tailored to the image characteristics of digital mammograms is thus critical for improved assessment of breast density in clinical practice.

Very limited work has investigated automated assessment of breast density in *full-field digital* mammography.[26–31] Most of these works have shown that the physics of image acquisition in digital mammography, different than that screen-film mammography, is useful in the estimation of breast density. For example, Lu *et al.* show that breast density can be estimated by a regression model of image-acquisition parameters.[27] Oliver *et al.* describe a statistical, principle component analysis-based segmentation scheme based on local texture properties.[28] Heine *et al.* have recently proposed a calibration method coupled with first-order histogram statistics, namely, mean and standard deviation, as an automated surrogate for breast density assessment from digital mammograms for the purposes of estimating cancer risk directly.[29] However, most of these works only provide limited validation and some, such as the work by Heine *et al.* described above, explicitly requires the use of raw mammograms.

In addition to the two-dimensional area measures described above, it has been recently proposed that volumetric breast density estimation may allow for improved assessment of the fibroglandular tissue content of the breast,[32] which may in turn to lead to improved breast cancer risk assessment.[33] Recent work has investigated automated methods to assess dense tissue volume.[30,31,34–36] Both area-based and volume-based density estimates have been shown to be indicative of the risk for breast cancer. However, studies to date have yet to conclusively show that volumetric density measures can surpass area-based measures in breast cancer risk assessment.[37,38] Therefore, it remains beneficial to continue investigations of area-based density measures, given that area breast percent density is currently the most validated imaging metric in breast cancer risk assessment.

A limitation common to digital mammography breast density estimation algorithms presented to date is the focus of density analysis on a single type of digital mammographic image, while, in general, *two* types of images are generated during the digital mammography imaging process. Digital mammography image acquisition initially generates an image which is proportional to the x-ray attenuation through the breast, known as the raw image (i.e., FOR PROCESSING; often with a 14-bit gray-level depth). Then, vendor specific postprocessing algorithms are applied to increase lesion conspicuity before radiological presentation, creating what is known as the processed image (i.e., FOR PRESENTATION; often with a 12-bit gray-level depth). While it is reasonable to assume that breast density estimation should be obtained directly from the raw images since they retain the original relationship with the physical properties of the breast

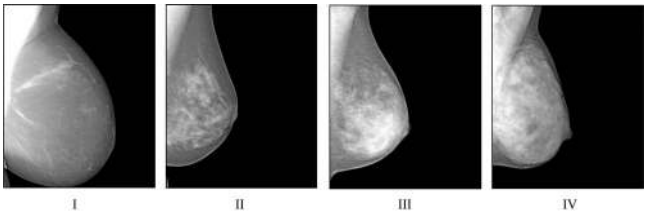

FIG. 1. Sample digital mammograms of BIRADS categories I–IV digital mammograms in order of increasing percent density. (I) <25%; (II) 26%–50%; (III) 51%–75%; (IV) >75%.

tissue,[39] most medical centers use only the postprocessed images for clinical purposes, which are thus, in turn, the only ones archived out of consideration of cost and storage constraints. Therefore, application of methods requiring raw mammograms, such as the method by Heine *et al.*[40] described above, could be difficult in cases where large retrospective epidemiologic or multicenter studies would be desired. Therefore, development of an algorithm that can effectively estimate breast PD% in both raw and postprocessed digital mammogram would be beneficial both in terms of direct clinical application and retrospective research-related studies.

Previous work has shown that certain physics parameters of DM image acquisition, such as x-ray tube current and exposure, correlate to PD% measured from raw mammograms.[27] Image gray-level characteristics such as first-order histogram statistics and texture descriptors have also been shown to be useful in identifying dense tissue in digitized mammograms.[21] In addition, patient-specific characteristics such as age and breast thickness are also known to be associated with measured PD% from digitized film mammograms.[18] Therefore, in this work, we introduce an innovative methodology optimized for the quantification of breast percent density that is tailored to the characteristics of the specific image and specific patient, and that is equally applicable to both raw and postprocessed mammographic images, thus increasing clinical utility. Therefore, we propose a new algorithm for fully automated quantification of breast PD% based on adaptive multiclass fuzzy c-means (FCM) clustering and support vector machine (SVM) classification that is optimized for the imaging characteristics of both raw and processed digital mammography as well as on individual patient and image characteristics. Our algorithm involves a series of steps; (i) breast region and pectoral muscle segmentation; (ii) z-score normalization of the gray-level intensity values within the segment (iii) adaptive histogram-based determination of the optimal number clusters for FCM segmentation; and (iv) dense tissue cluster merging based on features predictive of PD% through a SVM aggregation classifier. A flowchart outlining these steps is provided in Fig. 2. We validate our algorithm by comparing to radiologist-provided estimates of PD% (considered here as our gold standard) and ACR-BIRADS categorical density on a set of 81 cases with bilateral mediolateral oblique (MLO) normal digital screening mammograms, in both raw and vendor postprocessed format (a total of 324 images), which cover the full spectrum of breast densities seen in clinical practice.

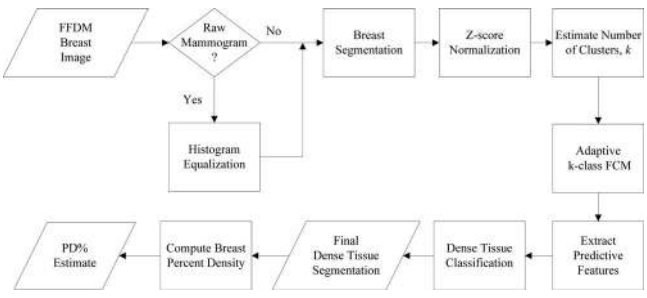

FIG. 2. Flowchart of the proposed algorithm.

II. METHODS AND MATERIALS

II.A. Study population and DM image acquisition

This study was in compliance with the Health Insurance Portability and Accountability Act (HIPAA) and received institutional review board (IRB) approval. Anonymized full-field DM images acquired as part of a separate IRB-approved multimodality breast cancer screening trial previously completed in our department (July 2007–March 2008) were retrospectively analyzed. Women who participated in the trial were asymptomatic volunteers who presented for annual screening mammography and had given written informed consent before their participation. Bilateral, MLO view DM imaging was performed using a full-field digital mammography unit (Senographe DS; GE Healthcare, Chalfont St Giles, UK). The raw digital mammograms were acquired at a 100 μm isotropic resolution using a 14-bit gray-level depth. The raw mammograms were processed using *PremiumView*™ (GE Healthcare), a vendor-specific, embedded adaptive histogram equalization algorithm[41,42] which produces 12-bit gray-level postprocessed images. Of the 83 women originally enrolled in the trial, two were excluded from this analysis: one due to a diagnosis of breast cancer, the other due to insufficient image quality. The remaining 81 women were, therefore, available for retrospective breast density analysis.

II.B. Radiologist estimation of percent breast density

Area-based breast PD% was estimated by a trained breast-imaging radiologist for these 81 women on a per-breast basis using a validated, interactive, image-thresholding tool for breast PD% estimation (Cumulus, Ver. 4.0, Univ. Toronto) (Ref. 43) in both the raw and processed images, for a total of 324 digital mammograms. As a result of the different visualization of the breast tissue between raw and postprocessed images, tailored approaches were adopted for the estimation breast PD%.[39] Briefly, for the postprocessed DM images, the digital mammograms were first windowed by the radiologist for optimal display. Following this, the background air region was excluded via a manually determined intensity threshold, therefore, allowing the breast boundary to be designated. The pectoral muscle region was subsequently excluded via manual delineation of the pectoral muscle edge. The remaining portion of the image was designated as the breast tissue, and the total area of this region is computed by the software. Following identification of the breast, a second, user-defined gray-level intensity threshold is selected in order to define the gray-level cut-off between fibroglandular and adipose tissue, and those pixels within the delineated breast region above this gray-level threshold are designated as fibroglandular tissue. PD% is then computed as the percentage of the breast area occupied by fibroglandular tissue. A similar process is used to estimate PD% from raw images, except that since the raw images are not optimized for clinical visualization and interpretation, the digital mammographic image were rewindowed by the radiologist before each of the segmentation and thresholding steps described above. Two readings per image were performed by the radiologist, each six months apart, and the

TABLE I. Distribution of assigned BIRADs density categories for raw and processed DM images.

	Assigned BIRADs category			
DM image type	I	II	III	IV
Raw	75	57	26	4
Processed	69	73	16	4

average of the two reading was considered as our gold-standard in order to minimize the effects of intrareader variability. The intrareader correlation and 95% confidence interval (CI) between the two radiologist PD% readings was found to be $r = 0.95$ ($p < 0.001$; CI: 0.93–0.97) for the raw DM images and $r = 0.93$ ($p < 0.001$; CI: 0.90–0.95) for the processed DM images. Table I provides the distribution of BIRADs density categories assigned to the 324 DM images in this dataset.

II.C. Automated breast density estimation

Prior to automated breast density estimation, the DM images are standardized in a three-step process. First, histogram equalization is applied as a standard preprocessing step to improve the relative tissue contrast in the raw digital mammograms.[41,44] The histogram equalization method used in this work involves log-normalization of the original raw DM intensity values which are then converted to calibrated optical density values based on a validated characteristic curve (Yaffe *et al.*, University of Toronto).[35] Figure 3 shows an illustration comparing the effect of this histogram equalization algorithm on the intensity distribution of a raw digital mammogram as compared to vendor-processing. Second, in order to both smooth image noise as well as reduce computational time, all digital mammograms are down-sampled by a factor of 4 from an original image pixel-resolution 2294 $\times$ 1914 to a 574 $\times$ 479 resolution using bicubic interpolation. Finally, to take advantage of the symmetric anatomy between left and right breasts, all images considered in this work are aligned so that the chest wall is on the left image edge; effectively all right MLO view mammograms are reflected about the Y axis of image.

II.C.1. Breast tissue area segmentation

MLO-view mammograms can be geometrically subdivided into three distinct regions: the chest wall, including the pectoral muscle; the breast tissue, comprised primarily of adipose and fibroglandular tissues; and the background air region. In order to quantify breast density, accurate segmentation of the breast from the other regions within the image is necessary. Thus, breast area segmentation is performed to first identify the tissue-air interface and then to identify the boundary between breast tissue and the pectoral muscle.

Although previous studies discuss applying a direct threshold to identify the air region in a mammogram,[20,22] few explicitly discuss how the value of that threshold is determined. Background air will often have a nearly uniform gray-level intensity of approximately zero in digital mammography, indicative of the fact that the air region is the most radio-lucent portion of the image. However, appropriate determination of this threshold becomes important for those few images where this assumption does not necessarily hold, as is illustrated in Fig. 4, which could occur for various reasons such as miscalibration of the mammography unit or due to vendor image postprocessing algorithms.

Therefore, in this work, the body-air interface boundary is determined by a threshold based on the gray-level intensity histogram, independent of any prior assumptions. The gray-level value of the edge, E, of the first major component, or region, of the gray-level histogram, H, can be computed as

$$E = \min(t) : t = \{H(i) \geq \max(H(I)) \cdot P_E : i \in I\}, \quad (1)$$

where i is the set of gray-level intensity values found in image I, and P_E is an adjustable proportionality used to define the edge of the significant components of the histogram. This in turn allows the computation of an appropriate air-threshold, th, by calculating

$$th = \min(g) : g = \{H(i) \leq \max(H(I)) \cdot P_{TH} : i > E\}, \quad (2)$$

where P_{TH} is an adjustable proportionality used to define which gray-level regions of the histogram comprise a nominal fraction of the image. Both P_E and P_{TH} can be either learned or explicitly defined; in this work, we found that $P_E = 20\%$ and $P_{TH} = 1\%$ performed reasonably well to allow for accurate identification of the breast-air boundary, an illustration of which is provided in Fig. 4.

Once the air-tissue interface is identified, we determine the boundary between the pectoral muscle and breast tissue areas using a previously validated algorithm based on a straight line Hough transform.[20] Briefly, a canny edge operator[45] is applied to create an edge map, which is then collected in a two-dimensional, (ρ,θ)-line-parameterization histogram of Hough space. The pectoral-boundary line can then be approximated by a straight line between $\theta = 40^\circ$ and $\theta = 80^\circ$ whose angle and location in image-space is defined by the (ρ,θ) location of the global maxima of the Hough-space histogram.

II.C.2. Adaptive fuzzy c-means clustering

Once the breast tissue region within a given digital mammogram is identified, we perform an unsupervised, k-class FCM clustering of the breast region gray-level intensities, where the number of clusters, k, is adaptively determined on a per-image basis, in order to partition the breast tissue into sub-regions of relatively homogenous gray-level intensity clusters.[26] The output of FCM clustering is known to be dependent on the *a priori* definition of the number of clusters. Therefore, given that the overall distribution of the pixel intensity values in a digital mammogram may be thought of as an admixture of several different Gaussian distributions, we adaptively compute k for each image based on the properties of the histogram of the pixel intensity values within the breast region. To accomplish this, the gray-level intensity histogram

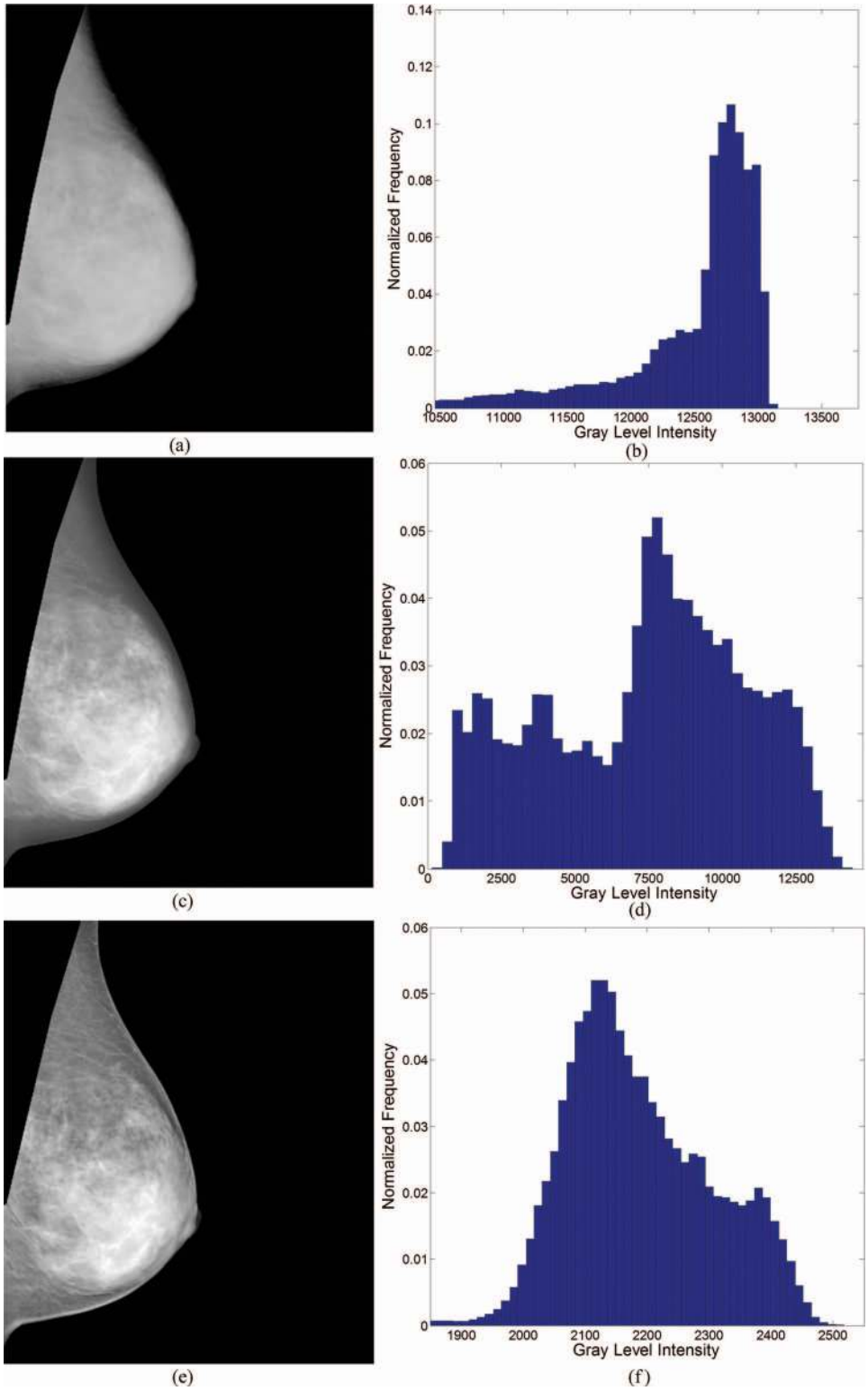

FIG. 3. Comparison of gray-level intensity distributions of the breast region in "For Processing" (i.e., "Raw," images a and b), histogram equalized raw (i.e., images c and d) and "For Presentation" (i.e., "Processed"; images e and f) digital mammograms of a BIRADS III category woman.

of the breast tissue is first normalized to have a zero-mean and unit standard deviation (i.e., z-score normalization). Z-score normalization in this manner allows for rescaling of the histogram percentiles and FCM intensity-centroids computed from the different DM images to a common range without altering first-order histogram statistics (i.e., skewness, kurtosis) as they are useful in distinguishing very fatty and very dense breasts.

The normalized histogram is then smoothed using a discrete Gaussian kernel using the formalization by Harris[46] such

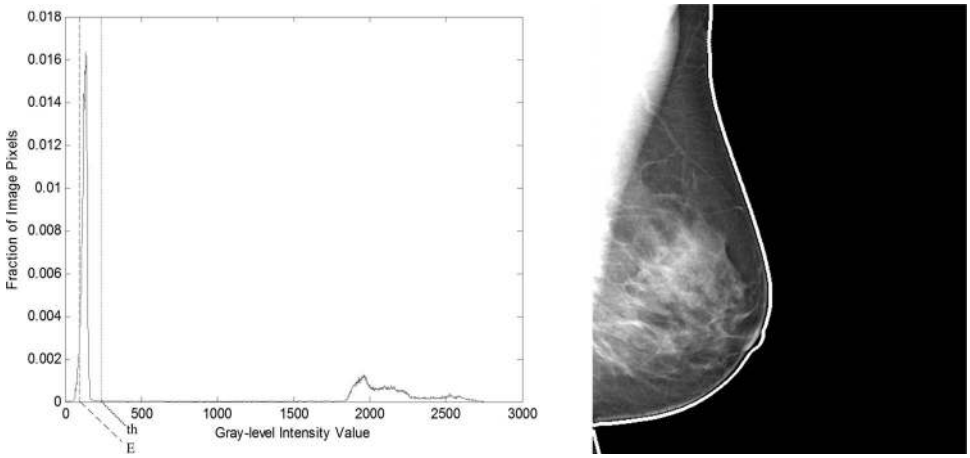

FIG. 4. Illustration of adaptive air threshold detection on a digital mammogram with nonzero air pixels. (Left) Histogram showing the location of the first major rise in gray-level values, E, (long-dashed line) and the computed air threshold, th, (short-dashed). (Right) Identified breast-air interface contour (white line).

that the kernel shape is invariant with respect to the kernel width, in order to minimize variation in histogram intensity values (Fig. 5). The parameters (kernel width and alpha) can be defined explicitly or learned via an optimization step, and the effect of these settings on algorithm performance is explored in this work. After smoothing, the number of local peaks present in the histogram is assigned as k and can be defined as the number of zero-crossings in the first derivative with negative second derivatives such that

$$k = |\{H'(g_n) = 0 : g_n \in M_B; \quad H''(g_n) < 0\}|, \tag{3}$$

where $H(g_n)$ is the smoothed histogram of the normalized gray-level intensity values, g_n, within the segmented breast region, M_B. It is worth noting that for the purposes of this analysis, k is restricted to be at least two in order to ensure the presence of at least one dense and one adipose cluster. Furthermore, while the upper limit of k can theoretically be unbounded, in this work an upper bound of $k = 13$ is imposed in order to ensure computational efficiency as well as minimize the chance of over-fitting.

Once the appropriate number of clusters is computed for a given image, k-class FCM clustering[47] is performed on the z-score normalized gray-level values in the breast region. The FCM algorithm applied in this work uses a standardized initialization of intensity cluster centroids distributed evenly from a normalized intensity of -4σ to 4σ, where σ is the standard deviation on the intensity value distribution in the breast. These cluster centers are then iteratively adjusted to minimize a weighted sum of squares error function,[48] ultimately yielding cluster centroids and a cluster-membership matrix for every intensity value in the breast region. After cluster centroid optimization, every pixel is assigned to the cluster for which that pixel's intensity value has the highest membership score. In this manner, clusters ultimately represent regions of the image of similar gray-level intensity and thus regions of similar x-ray attenuation properties. An example histogram for a $k = 6$ case and the resultant FCM clustering is shown in Figs. 6(b) and 6(c), respectively.

II.C.3. Cluster classification and percent density calculation

Breast PD% is defined as the ratio of dense tissue area to the total breast area as seen in mammography. Standard

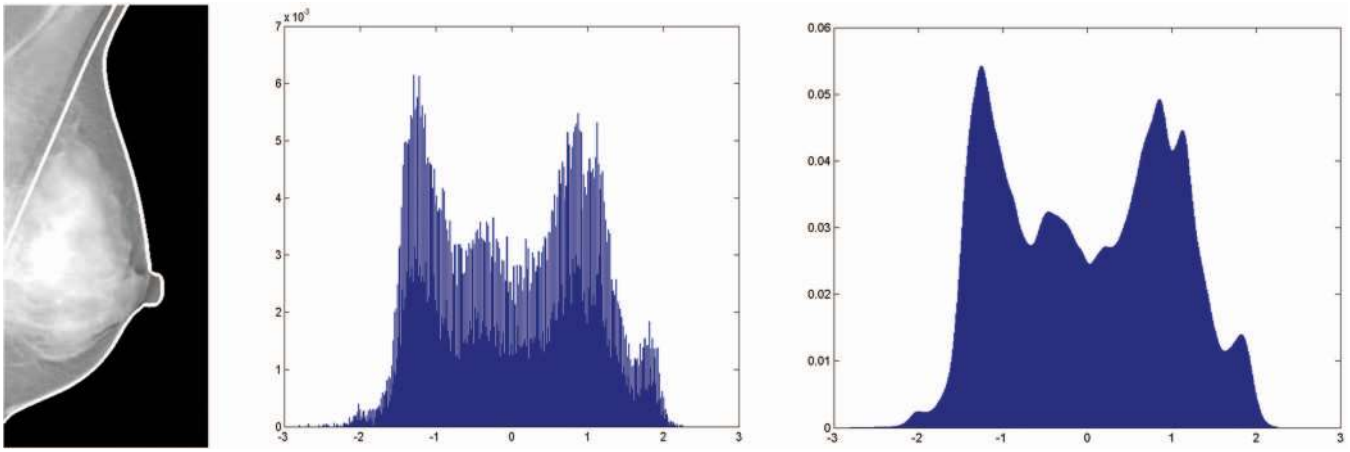

FIG. 5. Effect of gray-level histogram smoothing. (Left) Original mammogram with breast area outlined in white, (Center) Z-score normalized gray-level intensity histogram constructed at a 0.01 bin-width, (Right) Histogram postsmoothing with a Gaussian kernel of width = 50, alpha = 5.

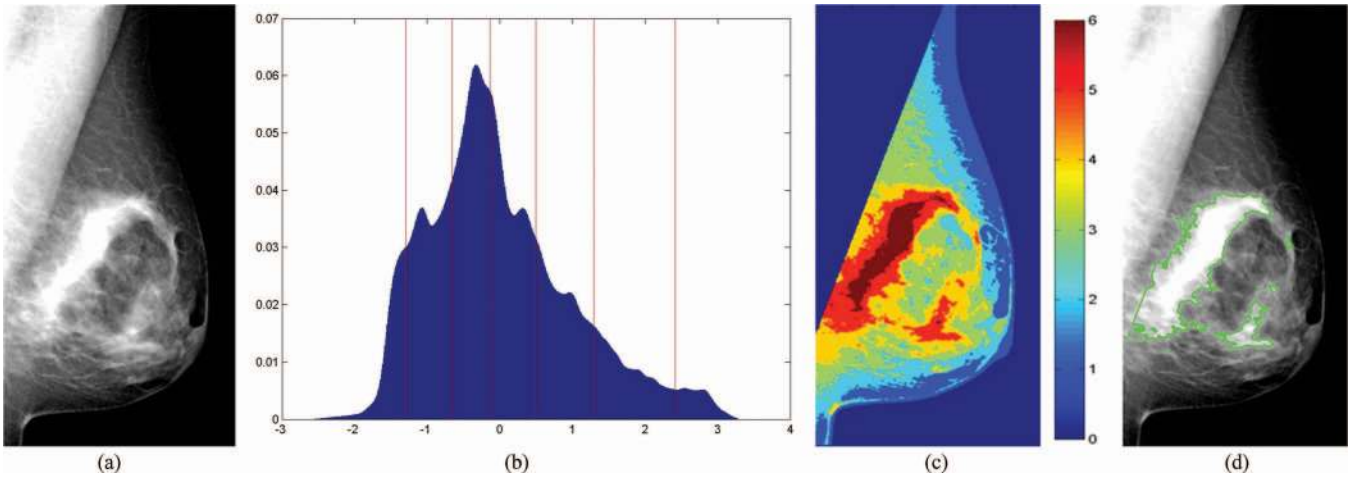

FIG. 6. Segmentation algorithm for a $k = 6$ mammogram. (a) Original mammogram; (b) normalized, smoothed breast-pixel intensity histogram with FCM cluster centroids (vertical lines); (c) pixel cluster-membership represented by shading; (d) final dense tissue segmentation combining clusters 5–6.

practice is to define a gray-level threshold in order to divide the breast area into two distinct regions: one that is predominantly composed dense tissue and one that is composed of fat. To transform the multicluster output of the FCM algorithm into this standard two-class paradigm for the assessment of breat PD%, a SVM classifier is implemented to determine which of the FCM clusters are predominantly dense for a given mammographic image, which in turn are aggregated into a single dense-tissue segmentation output. In this work, a linear SVM classifier is implemented in order to minimize parameter space and reduce the likelihood of overfitting.

Given known associations found between various patient and image characteristics with breast PD%,[18,21,27,49] the SVM considers established image features to guide cluster aggregation for the specific mammogram being analyzed. More specifically, the SVM incorporates image acquisition parameters and patient characteristics, such as x-ray dose, patient age, and breast thickness, shown to correlate to PD%,[27] which are global features for all detected clusters. In addition, the SVM considers both whole breast and per-cluster gray-level histogram statistics (e.g., skewness and kurtosis), gray-level texture (e.g., energy and entropy), and morphological shape descriptors (e.g., compactness and perimeter length) previously validated in the literature,[50–53] as these descriptors have been used to classify images into distinct BIRADs categories.[21] Lastly, a set of features is computed to compare changes to the values of the individual features described above when a new cluster is sequentially added to the merged density segmentation. In this way, a total of 86 features are initially considered.

Cluster-based training labels for the SVM classifier are established by defining by a dichotomous, dense vs nondense, labeling of all the clusters detected by the FCM clustering algorithm such that for each image, the N-highest-intensity clusters are assigned a dense label, where N is selected in order to minimize the error between the resulting PD% for a given cluster-merging and the radiologist-provided estimate for each image. Effectively, this converts the output clustered image of the unsupervised FCM algorithm into a pixelwise ground truth segmentation of breast density for use in training the SVM-classifier based on the radiologist-provided PD% estimate for a given digital mammogram. To accomplish this, for each mammogram in the dataset, we begin by defining C as an image containing the cluster index of pixel i generated by the prior k-cluster FCM step, where the cluster indices are assigned to the M pixels within the breast area such that their rank is in increasing order of their centroid intensities. In this way, the corresponding breast density estimate, PD, for a particular grouping of the upper η cluster indices is

$$\mathrm{PD}(\eta) = \frac{1}{M}\sum_{i=1}^{M}\delta(C_i \geq \eta), \tag{4}$$

where δ is an indicator function on the condition X such that

$$\delta(X) = \begin{cases} 1; & X \text{ is true} \\ 0; & X \text{ is false} \end{cases}. \tag{5}$$

Using Eqs. (4) and (5), we can then determine an optimal η^* that minimizes the error between $\mathrm{D}(\eta)$ and the radiologist-provided estimate of breast density for a given image, PD_R, by computing

$$\eta^* = \mathrm{argmin}\left|\mathrm{PD}_R - \frac{1}{M}\sum_{i=1}^{M}\delta(C_i \geq \eta)\right| : \eta \in \{1,\ldots,k\}. \tag{6}$$

A final, pixelwise labeled breast density image, L, can then be created by computing

$$L_i = \delta(C_i \geq \eta^*), \tag{7}$$

where L_i is the label (i.e., dense vs nondense) for the ith pixel in the breast region, which results in all pixels in a given cluster having an identical label. In order to ensure that there is always at least one cluster assigned as a dense tissue region and one as a nondense region, an additional restriction is imposed requiring η^* to be at least 2.

The SVM classifier is then trained on these density-labeled images in a leave-one-woman-out cross-validation fashion. In each cross-validation fold, the classifier is trained on the

density-labeled FCM-clustered mammograms (n = 160 images) of the 80 women selected for the training set, and then applied to the two (i.e., left and right) out-of-sample FCM-clustered mammograms of the woman selected as the testing sample. In order to reduce the dimensionality of the input feature variables, a stepwise feature-selection stage is nested within each leave-one-woman-out loop of the cross-validation. For this study, feature selection is performed by systematically adding and removing[54] features from the SVM-model classifier based on the statistical significance of the correlations of the features to total breast PD% in the training set. In this manner, it can be expected that the computed features will drive cluster classification in such a way that individual clusters are more likely to be classified as dense if the feature vector considered by the SVM classifier for a particular image is indicative of a relatively dense breast. The SVM is then trained using only those features selected in the given loop and applied to the test set.

After classification of the test set FCM-clustered images, the clusters assigned a dense label are sequentially merged in order of decreasing mean intensity into single segmentation of the dense tissue region until a cluster classified as nondense is found. The final merged segmentation is then selected as the dense tissue segmentation, with the remainder of the breast area designated as fatty tissue. An example of the cluster aggregation result can be seen in Fig. 6(d). From this final density segmentation, we calculate the standard mammographic percent density metric, PD%, by computing

$$\text{PD\%} = \frac{|M_D|}{|M_B|} \cdot 100\%, \tag{8}$$

where MD and MB are the dense tissue segmentation and breast tissue segmentation, respectively.

II.D. Statistical analysis of algorithm performance

To determine the degree of association between algorithm-derived and radiologist-provided estimates of mammographic breast PD%, linear regression analysis was performed and the Pearson product-moment correlation coefficient,[55] r, and 95% CI of r, was computed between the two PD% estimates on a per-breast basis. The association between per-woman estimates of PD%, generated by averaging each woman's estimates of left and right PD%, was also assessed. The feature selection rate of the cross-validation stage was also computed and the features consistently selected across over 90% folds of the cross-validation are reported in order to assess which of the considered features in this work are most robust in assessing breast percent density.

To evaluate the consistency of the algorithm in assessing breast PD% from either raw or processed mammograms, a repeated measures two-way analysis of variance (ANOVA) on estimated PD% using reader (algorithm vs radiologist) and presentation (raw vs processed) as factors was performed, as ANOVA would allow for the identification of any systemic biases between algorithm and radiologist estimates in either the raw or processed image presentation. A comparison evaluating the distribution of absolute differences between algorithm and radiologist estimates of PD% between raw and processed images was also performed using an F-test for equal variances in order to determine if there is a statistically significant difference in the estimation of PD% between the two image presentations.

Performance of the algorithm in terms of correct *categorization* of a particular woman's density was also assessed. Categorical agreement between the radiologist and algorithmic estimates of per-breast and per-woman density category was determined by computing Cohen's κ statistic[56] and 95% CI using a quadratic weighting schema[13,14,57] on the conversion of the radiologist-provided and algorithm-estimated PD% to corresponding ACR BIRADS categories, using the standard thresholds of 25%, 50%, and 75% to separate the four classes.[10,14]

Finally, to assess the robustness of the algorithm performance based on parameter set used in the determination of k, we also vary the three parameters used to construct the histogram over a range of values: the bin-width of the histogram, b, the width of the discrete Gaussian smoothing kernel, w, and the alpha of the discrete Gaussian smoothing kernel, α. This is done to assess both the impact of the specific parameter and the impact of the selection of k on the FCM and SVM density assessment scheme. Pearson's r is used to assess the agreement between the algorithm and radiologist estimates of PD% across the range of evaluated parameters.

III. RESULTS

III.A. Algorithm performance

Algorithm-estimated PD% was strongly associated with radiologic per-breast PD% estimates in both raw ($r = 0.82$, CI: 0.76–0.86; $p < 0.001$) and processed ($r = 0.85$, CI: 0.80–0.89; $p < 0.001$) mammograms when analyzed using an empirically chosen baseline set of parameter settings (raw: $b = 0.01$, $w = 30$, and $\alpha = 5$; processed: $b = 0.01$, $w = 50$, and $\alpha = 5$), as seen in Fig. 7. Nested feature selection found that a subset of 18 features, listed in Table II, were consistently selected across at least 90% of cross-validation folds for both the raw and processed datasets. When combining left and right breast PD% estimates into a per-woman average composite PD% score, agreement increased, with Pearson correlations of $r = 0.85$ (CI: 0.78–0.90) and $r = 0.89$ (CI: 0.84–0.93) for raw and processed mammograms, respectively ($p < 0.001$).

Repeated measures two-way ANOVA demonstrated that there were no statistically significant differences between the estimates ($p > 0.1$) due to either presentation of the image (raw vs processed) or method of PD% assessment (radiologist vs algorithm), even when interactions are considered (Table III), both on a per-breast and a per-woman basis. Figure 8 provides both the per-breast and per-woman distributions of the PD% estimates for the different combinations of presentation and assessment method. Similarly, no statistically significant systematic difference was found in the variance of absolute errors of the algorithm-estimated breast PD% between raw and processed images (F-test: $p > 0.1$).

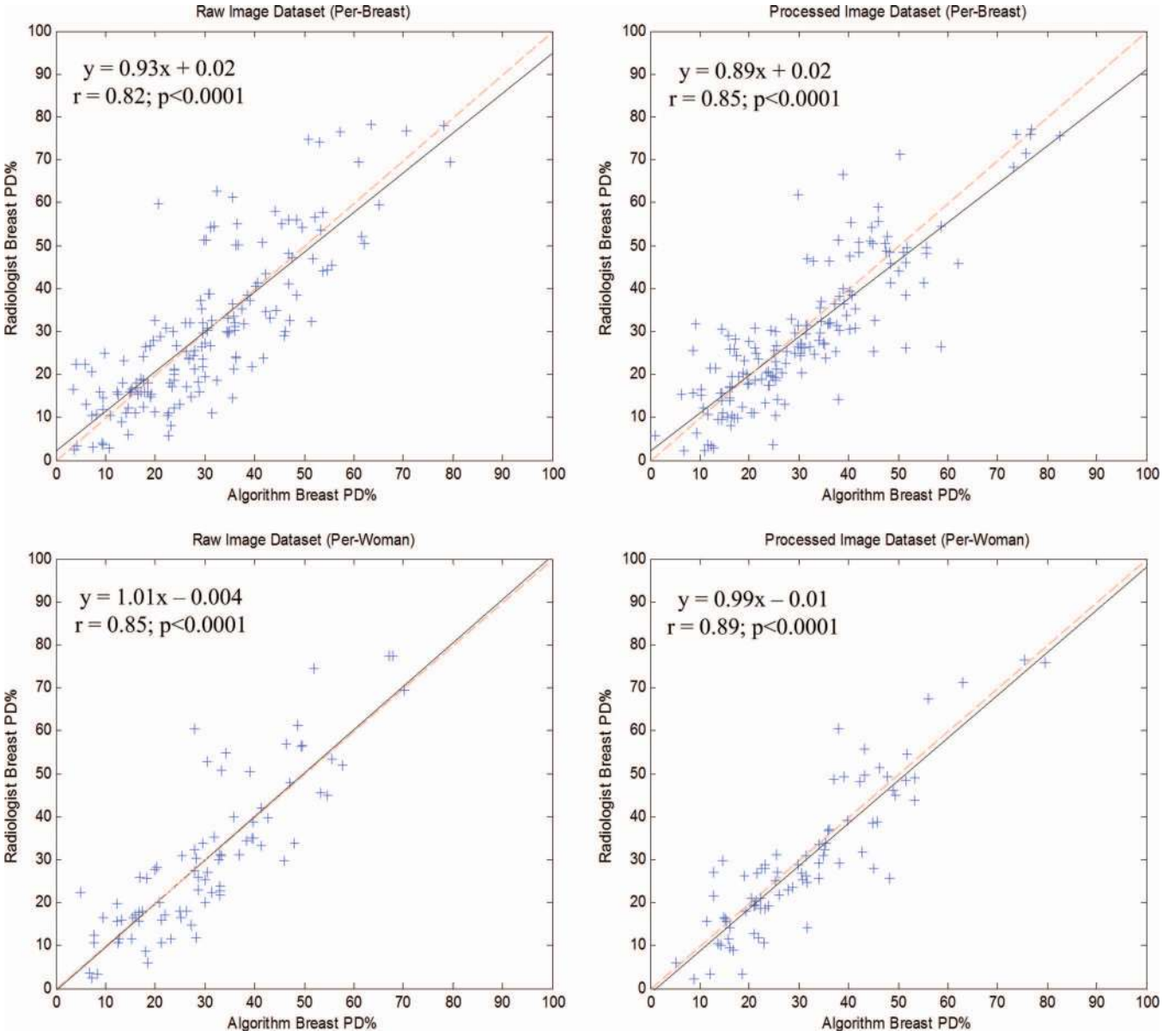

FIG. 7. Scatter plots of per-breast (top row) and per-woman (bottom row) algorithm-estimated (x axis) and radiologist-provided (y axis) PD% for raw (left column) and processed (right column) DM image sets. Regression (solid) and unity (dashed) lines are provided for reference.

Strong categorical agreement is also observed between algorithm-assessed and radiologist-provided density after conversion of the continuous PD% score to the equivalent ACR BIRADS breast density categories (weighted Cohen's $\kappa \geq 0.79$; $p < 0.0001$) in raw and processed images for both per-breast and per-woman analysis (Table IV). Figure 9 illustrates this finding by providing boxplots of algorithm-assessed PD% scores as a function of the radiologist-provided ACR BIRADS estimates.

III.B. Robustness analysis

When assessing the robustness of the algorithm performance as a function of the histogram-construction parameters (i.e., bin width, b, Gaussian kernel width, w, and kernel variance, α) used to determine the number of clusters, k, for the adaptive FCM clustering, it was found that good performance ($r > 0.7$) was obtained for the majority of parameter settings considered in this analysis (Fig. 10). Of the three parameters varied, performance of the algorithm appeared to be most dependent on the bin-width, b. Specifically, when a small histogram bin-width was used, $b = 0.005$, the performance of the algorithm on both raw and processed mammograms was very good ($r > 0.8$) across almost all bin-width settings. Furthermore, performance was effectively constant for the processed mammograms when using a small bin-width, independent of the smoothing kernel parameter. A large bin-width setting, $b = 0.02$, appeared to be associated with rapidly decreasing performance as heavier Gaussian smoothing was applied in both raw and processed images. Finally, it appeared that variance of the Gaussian kernel, α, had minimal effect on overall algorithm performance, relative to the other two parameters.

TABLE II. Features with a 90%+ selection rate in the (a) raw and (b) processed mammogram datasets.

Global features	Cluster-merging features	Intercluster difference features
	(a) Raw	
Patient age	Z-score means	Z-score means
Breast area	Number of unconnected areas	Cluster area normalized by convex hull areas
Breast thickness	Euler number	Compactness
Breast perimeter length	Equivalent circular diameter	Perimeter length
Gray-level variance	Cluster perimeter length	
X-ray tube voltage	Cluster area normalized by convex hull area	
X-ray exposure	Homogeneity	
	(b) Processed	
Breast thickness	Z-score variance	Z-score means
X-ray exposure	Number of unconnected areas	Number of unconnected areas
X-ray tube voltage	Y-axis extent	Y-axis extent
Mean intensity of original breast image	Homogeneity	Major axis length of ellipsoid of inertia
Fifth percentile of the Z-score histogram	Cluster perimeter length	Equivalent circular diameter
Z-score range	Cluster area normalized by convex hull area	
Breast area normalized by convex hull area		

IV. DISCUSSION

In this study, we introduce a new algorithm for fully automated quantification of breast PD% based on adaptive FCM clustering and SVM classification, optimized for characteristics of digital mammography and of the individual woman. The proposed fully automated algorithm is successful in accurately estimating breast PD% from digital mammographic images. Strong association with radiologist-provided estimates of PD% is obtained for both quantitative and categorical PD% estimates, independent of whether raw or processed digital mammograms were used for analysis. The strength of association between found between algorithm and radiologist estimates of breast density in this work ($r = 0.82$–0.89) also outperform those reported in prior studies of automated density assessment in digital mammography ($r = 0.82$–0.83).[26,27] Overall agreement between algorithm and radiologist estimates of breast density also increased when averaged per-woman density estimates were considered, most likely due to the reduction of interbreast variations in density assessment for a given woman.

TABLE III. Repeated Measures ANOVA Tables for per-breast (Top) and per-woman (Bottom) estimation of breast PD%. No systematic difference due to presentation type (raw vs processed) or method of estimation (radiologist vs algorithm) was found, even if interaction was considered ($p > 0.1$).

Per-breast analysis: Source	SSE	D.F.	MSE	F	*p*
Presentation	0.1	1	0.11	0.0	0.984
Methodology	73.8	1	73.8	0.27	0.603
Presentation*methodology	60.5	1	60.5	0.22	0.638
Per-woman analysis: Source	SSE	D.F.	MSE	F	p
Presentation	0.1	1	0.06	0.0	0.988
Methodology	36.9	1	36.9	0.14	0.708
Presentation*methodology	30.3	1	30.3	0.12	0.734

The ability of the algorithm presented in this work to have good performance on both raw and processed digital mammograms is beneficial for several reasons. By being able to assess breast density from either raw or processed mammograms, the same algorithm could be incorporated into the clinical workflow at one of several optional points, such as at the point of image acquisition, thus operating on raw data, or at the clinician's workstation analyzing processed images. Furthermore, given the fact that most clinical centers typically use and archive only the postprocessed images due to cost and storage limitations, the same algorithm could be applied retrospectively to any number of datasets.

A major challenge in the validation of breast density estimation algorithms is establishing "ground truth" from projection images such as mammograms.[32] Given that the underlying, fibroglandular tissue properties of the breast which give rise to breast density are difficult to assess without individualized analysis of histologic breast tissue samples from mastectomy specimens[58] or core biopsies,[59] it is difficult to establish gold-standard PD% estimates. It has been shown that PD% estimates made by expert readers with interactive semiautomated software are indicative of cancer risk and thus may serve as surrogate measures of the true fibroglandular tissue content. Therefore, despite the subjective nature of user-assisted methods, expert readings provided by experienced breast-imaging radiologists remain a useful for generating breast percent density estimates for use in validating fully automated algorithms.

In order to understand the clinical utility of the proposed algorithm, it becomes important to place the reported performance of the algorithm in the context of interreader variability, with the assumption that a well performing algorithm would likely demonstrate a level of variability comparable to interobserver variability, when compared to radiologist gold-standard estimates of breast density. Several groups have reported on interreader agreement rates in studies evaluating

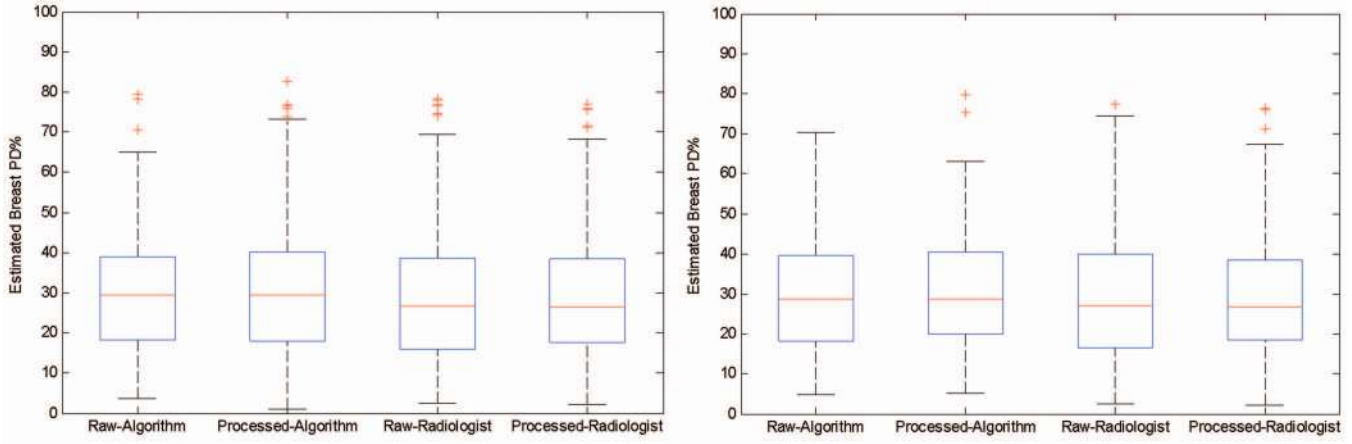

FIG. 8. Distributions of per-breast (left) and per-woman (right) assessed PD% as a function of image presentation and assessment-method. Two-way ANOVA indicated no significant groupwise differences ($p > 0.1$).

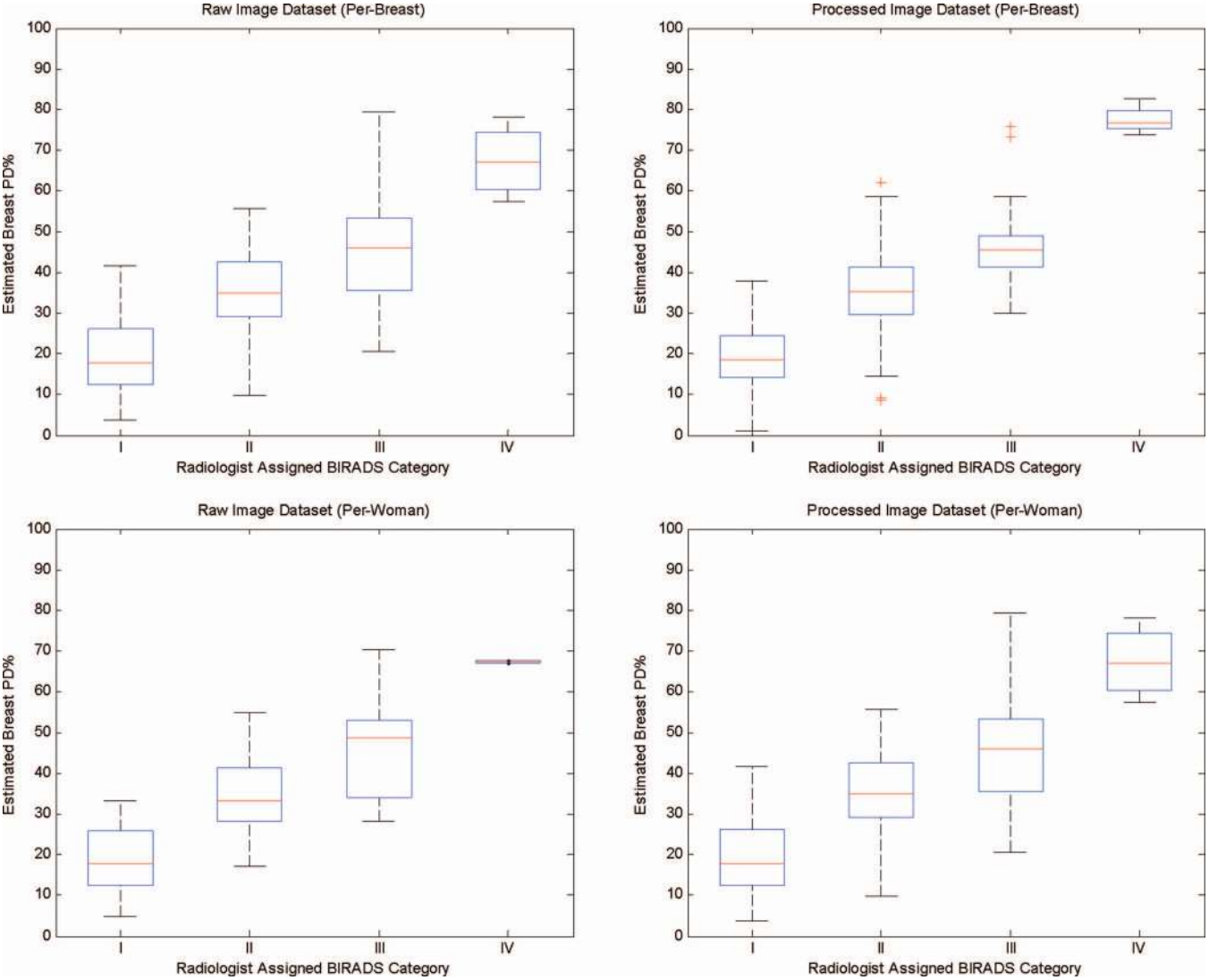

FIG. 9. Per-breast (top) and per-woman (bottom) box-plots of algorithm-estimated PD% in raw (left) and processed (right) DM images vs radiologist-provided categorical ACR BIRADS density scores. BIRADS categories were assigned using the standard thresholds on continuous PD%: (I) < 25%; (II) 25%–50%; (III) 51%–75%; (IV) >75%.

TABLE IV. Categorical agreement assessed using quadratic-weighted Cohen's κ between radiologist and algorithm computed BIRAD density. Strong agreement ($\kappa \geq 0.79$; $p < 0.001$) is seen between the two categorical estimates when assessed either per woman or per breast for both raw and processed images. The κ values and 95% confidence intervals are reported.

Analysis	Raw	Processed
Per-breast	0.80 (CI: 0.76–0.83)	0.80 (CI: 0.77–0.84)
Per-woman	0.79 (CI: 0.74–0.83)	0.84 (CI: 0.77–0.89)

the repeatability of categorical ACR-BIRADS and continuous PD%. For example, when looking to assess interobserver variability of four radiologists in categorizing breast density from screen-film mammograms, Ooms *et al.* found that interreader agreement based on Cohen's weighted-kappa varied between 0.65 and 0.84, with an average overall kappa of 0.77.[14] The weighted kappa values found in this work for agreement between algorithmic and radiologist-provided categorical estimates of breast density in both raw and processed digital mammograms ($\kappa = 0.79$–0.84, Table IV) indicate very high levels of agreement[12] and are within the upper portion of this range, suggesting good relative performance of the algorithm. In terms of quantitative PD% estimation, Conant *et al.* reported interclinician correlations of continuous PD% estimates ranging from $r = 0.90$ to $r = 0.95$ in processed digital mammograms.[42] Agreement between the algorithm and radiologic estimates of PD% ($r = 0.82$–0.89) on the dataset used in this study approaches this reported interclinical range, thus also suggesting relatively good performance of the algorithm

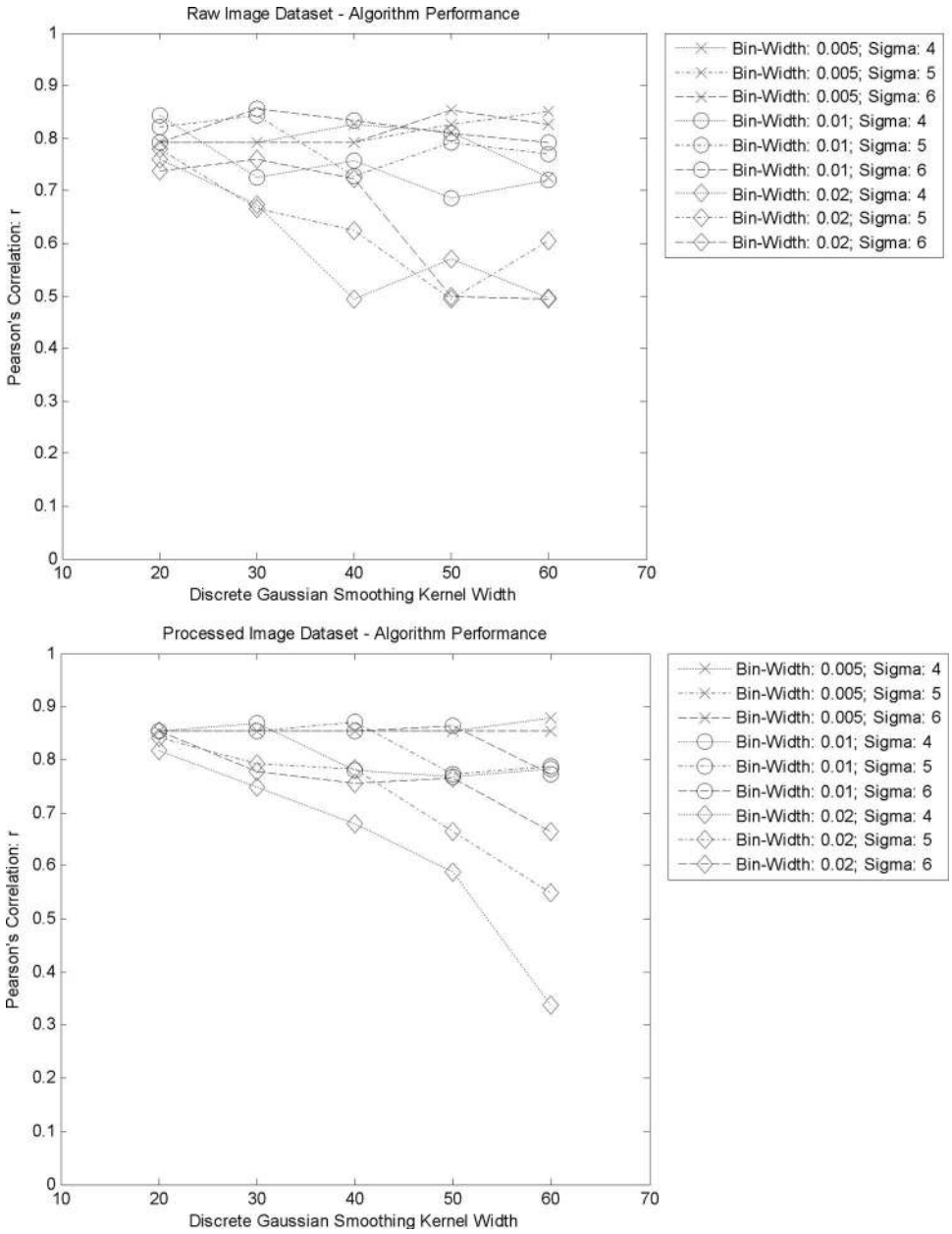

FIG. 10. Cross-validation performance as a function of histogram-construction parameters (i.e., bin width, b, Gaussian kernel width, w, and kernel variance, α) for raw (top) and processed (bottom) digital mammograms.

in assessing breast PD%. It should be noted, however, that as different authors and studies have access to different datasets, comparisons to the literature can only be considered as a reference for putting the different results in context, but not as a direct comparison between different methods.

One interesting finding was the consistently improved performance of the algorithm in assessing PD% on vendor processed images versus raw digital mammograms, both on per-breast and per-woman analysis. While further investigation is warranted to fully elaborate on this possible difference, on hypothesis is that this difference may be due to the presence of a relatively sharp, well-defined air-tissue boundary as well as increased contrast in the fibroglandular regions of the processed image, compared to the raw mammographic images, which may in turn be beneficial to the performance of breast density estimation algorithms.

Given that image intensity histogram construction and smoothing could be one of the largest sources of variability in the performance of the proposed algorithm, we investigated robustness of algorithm performance as function of bin-width and smoothing kernel parameters (i.e., width and alpha). Performance was found to be relatively consistent across a range of parameters, only degrading when larger bin-widths or more aggressive smoothing (i.e., larger Gaussian kernels) was used. The raw mammograms did show increased sensitivity to the parameter selection over the processed images, most likely, as discussed above, due to the effects of vendor processing enhancing the contrast between fibroglandular and adipose tissue regions in the postprocessed images. It was also found that when a small histogram bin-width was used, specifically $b = 0.005$, the performance of the algorithm was effectively constant, independent of the parameter values used to define the smoothing kernel, particularly when analyzing the processed dataset. This may suggest that an optimal parameterization would primarily be focused on a small bin-width, with kernel smoothing used to remove noise in the histogram only, which will be fully investigated in future work.

There are some limitations to the presented study. First, a second dataset with radiologist-provided estimates of breast PD% was not available for independent validation, so repeatability of performance could not be assessed outside of the leave-one-out cross-validation scheme. Second, only digital mammograms from a single manufacturer were analyzed in this study, limiting the assessment of the generalizability of the algorithm across different vendors. It may be expected that methods developed for raw digital mammograms may be most suitable for assessing breast density, as these images are less influenced by vendor-specific postprocessing algorithms.[16] Future studies will focus on evaluating the generalizability of our algorithm as well as determining what vendor-specific retraining may be required. Furthermore, only a single rater's assessment of breast percent density was available, limiting comparisons of interreader agreement to those previously reported in the literature. Future work should also seek to incorporate peripheral enhancement into the histogram equalization preprocessing method applied to the raw digital mammography images. Finally, the methodology presented in this work focused entirely on image intensity information; future work could also account for additional properties, such as the spatial clustering of dense tissue.

While evaluating the performance of our algorithm in terms of reader agreement and interreader variability is useful for understanding the potential efficacy of the algorithm, this provides limited information about the ability of the algorithm to assess breast cancer risk in a consistent manner. To account for this, further studies are underway at our institution to evaluate the predictive role of the estimated breast PD% measures, as provided by our algorithm, in assessing breast cancer risk. The long-term goal of this research is to translate our fully automated PD% estimation algorithm into a robust breast cancer risk assessment tool for use in clinical practice.

V. CONCLUSION

We have proposed and demonstrated the efficacy of a fully automated algorithm for breast percent density estimation in digital mammography. We are able to obtain strong correlation between the output of the computerized algorithm and radiologist-provided estimates of breast density. The ability of the algorithm to analyze both raw and processed digital mammograms with an equal level of accuracy increases its versatility, both in terms of direct clinical application and retrospective research-related studies. Overall, these findings indicate that the proposed algorithm can provide clinically relevant information from digital mammography for the assessment of breast density. The fully automated method could thus be used to accelerate the clinical translation of density-based cancer risk stratification in clinical practice, paving the way for new personalized screening and prevention strategies.

ACKNOWLEDGMENTS

This work was supported in part by American Cancer Society Grant No. RSGHP-CPHPS-119586, United States Department of Defense Concept Award No. BC086591, and National Institutes of Health PROSPR Grant No. 1U54CA163313-01. The authors would also like to thank Dr. Ahmed Ashraf (University of Pennsylvania) and Dr. Martin Yaffe (University of Toronto) for useful discussions pertaining to this work, as well as to Dr. Yaffe for providing *Cumulus* and the histogram equalization tools.

[a]Author to whom correspondence should be addressed. Electronic mail: brad.keller@uphs.upenn.edu; Telephone: 215-615-0826.

[1]A. Jemal, R. Siegel, J. Xu, and E. Ward, "Cancer statistics, 2010," Ca–Cancer J. Clin. **60**, 277–300 (2010).

[2]M. H. Gail, L. A. Brinton, D. P. Byar, D. K. Corle, S. B. Green, C. Schairer, and J. J. Mulvihill, "Projecting individualized probabilities of developing breast cancer for white females who are being examined annually," J. Natl. Cancer Inst. **81**, 1879–1886 (1989).

[3]C. D. Lehman, "Role of MRI in screening women at high risk for breast cancer," J. Magn. Reson Imaging **24**, 964–970 (2006).

[4]C. D. Lehman, J. D. Blume, P. Weatherall, D. Thickman, N. Hylton, E. Warner, E. Pisano, S. J. Schnitt, C. Gatsonis, M. Schnall, G. A. DeAngelis, P. Stomper, E. L. Rosen, M. O'Loughlin, S. Harms, and D. A. Bluemke, "Screening women at high risk for breast cancer with mammography and magnetic resonance imaging," Cancer **103**, 1898–1905 (2005).

[5]C. D. Lehman, C. Isaacs, M. D. Schnall, E. D. Pisano, S. M. Ascher, P. T. Weatherall, D. A. Bluemke, D. J. Bowen, P. K. Marcom, D. K. Armstrong, S. M. Domchek, G. Tomlinson, S. J. Skates, and C. Gatsonis, "Cancer yield of mammography, MR, and US in high-risk women: prospective multi-institution breast cancer screening study," Radiology **244**, 381–388 (2007).
[6]K. L. Smith and C. Isaacs, "Management of women at increased risk for hereditary breast cancer," Breast Dis. **27**, 51–67 (2006).
[7]J. N. Wolfe, "Breast patterns as an index of risk for developing breast cancer," AJR, Am. J. Roentgenol. **126**, 1130–1137 (1976).
[8]J. N. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern," Cancer **37**, 2486–2492 (1976).
[9]N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. A. Jong, G. Hislop, A. Chiarelli, S. Minkin, and M. J. Yaffe, "Mammographic density and the risk and detection of breast cancer," N. Engl. J. Med. **356**, 227–236 (2007).
[10]C. J. D'Orsi, L. W. Bassett, W. A. Berg, S. A. Feig, V. P. Jackson, and D. B. Kopans, "Breast imaging reporting and data system: ACR BI-RADS," in *Mammography*, 4th ed. (American College of Radiology, Reston, 2003).
[11]K. E. Martin, M. A. Helvie, C. Zhou, M. A. Roubidoux, J. E. Bailey, C. Paramagul, C. E. Blane, K. A. Klein, S. S. Sonnad, and H. P. Chan, "Mammographic density measured with quantitative computer-aided method: Comparison with radiologists' estimates and BI-RADS categories," Radiology **240**, 656–665 (2006).
[12]J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," Biometrics **33**, 159–174 (1977).
[13]J. Cohen, "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit," Psychol. Bull. **70**, 213–220 (1968).
[14]E. A. Ooms, H. M. Zonderland, M. J. Eijkemans, M. Kriege, B. Mahdavian Delavary, C. W. Burger, and A. C. Ansink, "Mammography: Interobserver variability in breast density assessment," Breast **16**, 568–576 (2007).
[15]B. T. Nicholson, A. P. LoRusso, M. Smolkin, V. E. Bovbjerg, G. R. Petroni, and J. A. Harvey, "Accuracy of assigned BI-RADS breast density category definitions," Acad. Radiol. **13**, 1143–1149 (2006).
[16]M. J. Yaffe, "Mammographic density: Measurement of mammographic density," Breast Cancer Res. **10**, 209–218 (2008).
[17]M. G. Kallenberg, M. Lokate, C. H. van Gils, and N. Karssemeijer, "Automatic breast density segmentation: An integration of different approaches," Phys. Med. Biol. **56**, 2715–2729 (2011).
[18]A. Tagliafico, G. Tagliafico, S. Tosto, F. Chiesa, C. Martinoli, L. E. Derchi, and M. Calabrese, "Mammographic density estimation: Comparison among BI-RADS categories, a semi-automated software and a fully automated one," Breast **18**, 35–40 (2009).
[19]C. K. Glide-Hurst, N. Duric, and P. Littrup, "A new method for quantitative analysis of mammographic density," Med. Phys. **34**, 4491–4498 (2007).
[20]N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms," Phys. Med. Biol. **43**, 365–378 (1998).
[21]A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Perez, E. R. Denton, and R. Zwiggelaar, "A novel breast tissue density classification methodology," IEEE Trans. Inf. Technol. Biomed. **12**, 55–65 (2008).
[22]T. Subashini, V. Ramalingam, and S. Palanivel, "Automated assessment of breast tissue density in digital mammograms," Comput. Vis. Image Underst. **114**, 33–43 (2010).
[23]J. S. Mandelblatt, N. Stout, and A. Trentham-Dietz, "To screen or not to screen women in their 40s for breast cancer: Is personalized risk-based screening the answer?," Ann. Intern. Med. **155**, 58–60 (2011).
[24]S. Suryanarayanan, A. Karellas, S. Vedantham, H. Ved, S. P. Baker, and C. J. D'Orsi, "Flat-panel digital mammography system: Contrast-detail comparison between screen-film radiographs and hard-copy images," Radiology **225**, 801–807 (2002).
[25]M. J. Yaffe, "Basic physics of digital mammography," *Digital Mammography* (Springer, Heidelberg, Germany, 2010), pp. 1–11.
[26]B. Keller, D. Nathan, Y. Wang, Y. Zheng, J. Gee, E. Conant, and D. Kontos, "Adaptive multi-cluster fuzzy C-means segmentation of breast parenchymal tissue in digital mammography," Med. Image. Comput. Comput. Assist. Interv. **14**, 562–569 (2011).
[27]L. J. Lu, T. K. Nishino, T. Khamapirad, J. J. Grady, M. H. Leonard, Jr., and D. G. Brunder, "Computing mammographic density from a multiple regression model constructed with image-acquisition parameters from a full-field digital mammographic unit," Phys. Med. Biol. **52**, 4905–4921 (2007).
[28]A. Oliver, X. Llado, E. Perez, J. Pont, E. R. Denton, J. Freixenet, and J. Marti, "A statistical approach for breast density segmentation," J. Digit Imaging **23**, 527–537 (2010).
[29]J. J. Heine, K. Cao, D. E. Rollison, G. Tiffenberg, and J. A. Thomas, "A quantitative description of the percentage of breast density measurement using full-field digital mammography," Acad. Radiol. **18**, 556–564 (2011).
[30]K. Hartman, R. Highnam, R. Warren, and V. Jackson, "Volumetric assessment of breast tissue composition from FFDM images," *Digital Mammography* (Springer, Heidelberg, Germany, 2008), pp. 33–39 .
[31]R. Highnam, S. Brady, M. Yaffe, N. Karssemeijer, and J. Harvey, "Robust breast composition measurement-Volpara™," *Digital Mammography* (Springer, Heidelberg, Germany, 2010), pp. 342–349.
[32]D. B. Kopans, "Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk," Radiology **246**, 348–353 (2008).
[33]C. M. Vachon, C. H. van Gils, T. A. Sellers, K. Ghosh, S. Pruthi, K. R. Brandt, and V. S. Pankratz, "Mammographic density, breast cancer risk and risk prediction," Breast Cancer Res. **9**, 217–225 (2007).
[34]O. Alonzo-Proulx, N. Packard, J. M. Boone, A. Al-Mayah, K. K. Brock, S. Z. Shen, and M. J. Yaffe, "Validation of a method for measuring the volumetric breast density from digital mammograms," Phys. Med. Biol. **55**, 3027–3044 (2010).
[35]O. Pawluczyk, B. J. Augustine, M. J. Yaffe, D. Rico, J. Yang, G. E. Mawdsley, and N. F. Boyd, "A volumetric method for estimation of breast density on digitized screen-film mammograms," Med. Phys. **30**, 352–364 (2003).
[36]S. van Engeland, P. R. Snoeren, H. Huisman, C. Boetes, and N. Karssemeijer, "Volumetric breast density estimation from full-field digital mammograms," IEEE Trans. Med. Imaging **25**, 273–282 (2006).
[37]N. Boyd, L. Martin, A. Gunasekara, O. Melnichouk, G. Maudsley, C. Peressotti, M. Yaffe, and S. Minkin, "Mammographic density and breast cancer risk: Evaluation of a novel method of measuring breast tissue volumes," Cancer Epidemiol. Biomarkers Prev. **18**, 1754–1762 (2009).
[38]J. A. Shepherd, K. Kerlikowske, L. Ma, F. Duewer, B. Fan, J. Wang, S. Malkov, E. Vittinghoff, and S. R. Cummings, "Volume of mammographic density and risk of breast cancer," Cancer Epidemiol. Biomarkers Prev. **20**, 1473–1482 (2011).
[39]D. Li, "Comparison of breast percent density estimation from raw versus processed digital mammograms," Proc. SPIE **7963**, 79631X (2011).
[40]J. J. Heine, E. E. Fowler, and C. I. Flowers, "Full field digital mammography and breast density: Comparison of calibrated and noncalibrated measurements," Acad. Radiol. **18**, 1430–1436 (2011).
[41]B. Chen, W. Wang, J. Huang, M. Zhao, G. Cui, J. Xu, W. Guo, P. Du, P. Li, and J. Yu, "Comparison of tissue equalization, and premium view post-processing methods in full field digital mammography," Eur. J. Radiol. **76**, 73–80 (2010).
[42]E. Conant, D. Li, S. Gavenonis, P. Bakic, A. K. Carton, C. Zhang, A. Maidment, and D. Kontos, "A comparative study of the inter-reader variability of breast percent density estimation in digital mammography: Potential effect of reader's training and clinical experience," *Digital Mammography* (Springer, Heidelberg, Germany, 2010), pp. 114–120.
[43]P. R. Bakic, A. K. Carton, D. Kontos, C. Zhang, A. B. Troxel, and A. D. Maidment, "Breast percent density: Estimation on digital mammograms and central tomosynthesis projections," Radiology **252**, 40–49 (2009).
[44]E. D. Pisano, E. B. Cole, B. M. Hemminger, M. J. Yaffe, S. R. Aylward, A. D. Maidment, R. E. Johnston, M. B. Williams, L. T. Niklason, E. F. Conant, L. L. Fajardo, D. B. Kopans, M. E. Brown, and S. M. Pizer, "Image processing algorithms for digital mammography: A pictorial essay," Radiographics **20**, 1479–1491 (2000).
[45]J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell. **8**, 679–698 (1986).
[46]F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," Proc. IEEE **66**, 51–83 (1978).
[47]J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Kluwer Academic Publishers Norwell, MA, 1981).
[48]X. Li, X. Lu, J. Tian, P. Gao, H. Kong, and G. Xu, "Application of fuzzy c-means clustering in data analysis of metabolomics," Anal. Chem. **81**, 4468–4475 (2009).

[49]S. Petroudi and M. Brady, "Breast density segmentation using texture," *Digital Mammography* (Springer, Heidelberg, Germany, 2006), pp. 609–615.
[50]M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision* (PWS, Pacific Grove, CA, 1999).
[51]A. Materka and M. Strzelecki, "Texture analysis methods: A review," Report No. COST B11 (Institute of Electronics, Technical University of Lodz, Brussels, 1998).
[52]H. Li, M. L. Giger, O. I. Olopade, A. Margolis, L. Lan, and M. R. Chinander, "Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms," Acad. Radiol. **12**, 863–873 (2005).
[53]A. Manduca, M. J. Carston, J. J. Heine, C. G. Scott, V. S. Pankratz, K. R. Brandt, T. A. Sellers, C. M. Vachon, and J. R. Cerhan, "Texture features from mammographic images and risk of breast cancer," Cancer Epidemiol. Biomarkers Prev. **18**, 837–845 (2009).
[54]N. Draper and H. Smith, *Applied Regression Analysis*, Wiley Series in Probability and Statistics (Wiley-Interscience, Hoboken, New Jersey 1998).
[55]K. Pearson, "Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia," Philos. Trans. R. Soc. London, Ser. A **187**, 253–318 (1896).
[56]J. Cohen, "A coefficient of agreement for nominal scales," Educ. Psychol. Meas.**20**, 37–46 (1960).
[57]J. Gao, R. Warren, H. Warren-Forward, and J. F. Forbes, "Reproducibility of visual assessment on mammographic density," Breast Cancer Res. Treat. **108**, 121–127 (2008).
[58]J. A. Harvey, R. J. Santen, G. R. Petroni, V. E. Bovbjerg, M. E. Smolkin, F. S. Sheriff, and J. Russo, "Histologic changes in the breast with menopausal hormone therapy use: Correlation with breast density, estrogen receptor, progesterone receptor, and proliferation indices," Menopause **15**, 67–73 (2008).
[59]K. Ghosh, K. R. Brandt, C. Reynolds, C. G. Scott, V. S. Pankratz, D. L. Riehle, W. L. Lingle, T. Odogwu, D. C. Radisky, D. W. Visscher, J. N. Ingle, L. C. Hartmann, and C. M. Vachon, "Tissue composition of mammographically dense and non-dense breast tissue," Breast Cancer Res. Treat **131**, 267–275 (2012).