Using multiscale texture and density features for near-term breast cancer risk analysis

Wenqing Sun and Tzu-Liang (Bill) Tseng

College of Engineering, University of Texas at El Paso, El Paso, Texas 79968

Wei Qian^{a)}

College of Engineering, University of Texas at El Paso, El Paso, Texas 79968 and Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang 110819, China

Jianying Zhang

Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang 110819, China and College of Biological Sciences, University of Texas at El Paso, El Paso, Texas 79968

Edward C. Saltzstein

University Breast Care Center at the Texas Tech University Health Sciences, El Paso, Texas 79905

Bin Zheng

Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang 110819, China and College of Engineering, University of Oklahoma, Norman, Oklahoma 73019

Fleming Lure

College of Engineering, University of Texas at El Paso, El Paso, Texas 79968 and Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang 110819, China

Hui Yu and Shi Zhou

Department of Radiology, Affiliated Hospital of Guiyang Medical University, Guiyang 550004, China

(Received 25 November 2014; revised 16 April 2015; accepted for publication 24 April 2015; published 18 May 2015)

Purpose: To help improve efficacy of screening mammography by eventually establishing a new optimal personalized screening paradigm, the authors investigated the potential of using the quantitative multiscale texture and density feature analysis of digital mammograms to predict near-term breast cancer risk.

Methods: The authors' dataset includes digital mammograms acquired from 340 women. Among them, 141 were positive and 199 were negative/benign cases. The negative digital mammograms acquired from the "prior" screening examinations were used in the study. Based on the intensity value distributions, five subregions at different scales were extracted from each mammogram. Five groups of features, including density and texture features, were developed and calculated on every one of the subregions. Sequential forward floating selection was used to search for the effective combinations. Using the selected features, a support vector machine (SVM) was optimized using a tenfold validation method to predict the risk of each woman having image-detectable cancer in the next sequential mammography screening. The area under the receiver operating characteristic curve (AUC) was used as the performance assessment index.

Results: From a total number of 765 features computed from multiscale subregions, an optimal feature set of 12 features was selected. Applying this feature set, a SVM classifier yielded performance of $AUC = 0.729 \pm 0.021$. The positive predictive value was 0.657 (92 of 140) and the negative predictive value was 0.755 (151 of 200).

Conclusions: The study results demonstrated a moderately high positive association between risk prediction scores generated by the quantitative multiscale mammographic image feature analysis and the actual risk of a woman having an image-detectable breast cancer in the next subsequent examinations. © 2015 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4919772>]

Key words: breast cancer risk, multiscale features, digital mammography, texture features

1. INTRODUCTION

Breast cancer is the most common cancer and one of the leading causes of cancer death in women over the age of 40.¹ Studies have shown that after 29 yr of follow-up, on average every 1334 screening mammograms saved one life, or for every 1000 women from ages of 40 to 69

with biennial screening, between 8 and 11 breast cancer deaths were prevented.^{2,3} As a result, in the last four decades, the promotion of breast cancer screening along with the advancement of cancer treatment methods significantly reduced mortality rates of breast cancer patients.⁴ However, studies have also revealed the limitations of screening mammography including the quite low specificity.^{3,5} One

study reported that during a 10-yr period, more than half of screened women received at least one false-positive recall and 7%–9% had at least one benign biopsy.⁶ The high false positive may result in psychological harms,⁷ unnecessary imaging tests with more radiation exposure,⁸ and negative biopsies of many cancer-free women who routinely participate in mammography screening, as well as the heavy burden for limited health care resources and associated high costs.⁹ In addition, there is also the concern that some women are treated for cancers that would not impact length of life.² Based on these facts, in 2009, the US Preventive Services Task Force has modified its recommendations that screening begins at an older age and at longer intervals than had been previously recommended.² However, the utility (risk-benefit and cost-benefit) of population-based screening mammography still remains controversial.¹⁰

To improve the utilities of breast cancer screening and prevention programs, more scientists and researchers have realized the significance of developing personalized mammography screening paradigms in which the breast cancer screening intervals and screening methods vary from different individuals.¹¹ Although a number of epidemiology-based breast cancer risk prediction models^{12,13} have been developed and tested, these models do not have clinically acceptable discriminatory power [or positive predictive value (PPV)] of individual women for breast cancer screening.¹⁴ Hence, it is important to develop new breast cancer risk assessment models that have high discriminatory power to stratify women into different risk groups of developing breast cancer within a specified period (or near-term) following one or multiple negative screening examinations. Based on this information, optimally personalized screening regimes with adaptively adjusted screening intervals can be recommended; hence, a small group of high-risk women should be more aggressively screened and monitored for the early breast cancer signs, while the majority of low-risk women can be screened in a longer interval to reduce false-positive recalls and associated potential harms.

Numerous studies have identified that breast density was the most significant risk factor besides woman's age.¹² Studies have reported that women with extremely dense breasts on mammography had a fivefold increased risk for breast cancer compared to age-matched women with breast density of less than 10%, and heterogeneously dense breast tissue was associated with a 1.5- to 2.0-fold increased risk.^{15,16} To measure breast density, Wolfe patterns were introduced in 1976 for the breast cancer risk assessment,¹⁷ and the categorical density scoring of the Breast Imaging-Reporting and Data System (BIRADS) was proposed by the American College of Radiology.¹⁸ However, the large interobserver variability makes these subjectively rated methodologies unreliable.¹⁹ In order to achieve more reliable and accurate results, several research groups have already investigated and tested the possibility of developing computerized schemes to automatically detect and quantify mammographic density.^{20–22} Some schemes have also been developed and tested to predict breast cancer risk using computed mammographic image features.^{23–25} Most of these

studies used the computational features extracted from whole breast or dense region of the breast. All the reported PPVs for risk analysis systems are relatively low, most of the PPVs are below 0.6, and the accuracy is still clinically unacceptable. Because of the heterogeneity of breast tissues, using mammographic tissue density features still faces many challenges. In addition, certain benign breast diseases detected on mammograms can also be a possible indicator for cancer risk.²⁶ These affect the performance of current risk prediction schemes. To overcome these limitations, recent studies have indicated texture features extracted from mammograms can help improve the prediction accuracy, even when breast density is considered.^{27–29} Furthermore, we noticed that in most reported studies, computerized features were measured and extracted from the single region of each mammogram, usually the whole breast area, but only a few considered the internal tissue/structural information.³⁰

The purpose of this study is to develop a new framework for near-term breast cancer risk analysis for multiscale texture measurements on mammograms as well as more accurate and sensitive measurements of breast density. In our previous research, we have designed and implemented a prototype of breast cancer risk assessment approach using mammographic density information.²³ In this study, we proposed and tested two new approaches. First, instead of extracting the features from the whole breast area, we calculated the features at different scales of subregions on each mammogram. Second, by expanding our initial image feature pool size, we developed and investigated a new fusion model to optimally combine the selected feature set that includes different types of multiscale density and texture features. We then investigated the degree to which the computed new features and risk levels were related, as well as the model performance in predicting the likelihood of women having high-risk breast lesions or cancers being detected in the near-term.

2. MATERIALS AND METHODS

2.A. Database

We have a preassembled de-identified database that contains full-field digital mammography (FFDM) cases, and each screening examination has four mammograms including two craniocaudal (CC) views and two MLO views from both right and left breasts. Half of the cases were diagnosed as cancer or abnormal, and the other half cases were diagnosed as normal or benign. The proportions of different subtle cancer types and BIRADS ratings follow the distribution of population source. Among the normal or benign cases, we randomly selected 340 cases which at least have two sequential FFDM examinations during a 3-yr period. The participants aged between 42 and 74 yr old, with a mean age of 55.2 yr and median age of 53.7 yr. All these “prior” FFDM screening examinations were interpreted as “negative” or “definitely benign” (without recall) by radiologists in the original screening practice. However, for some cases after 6–36 months, malignant tissues were confirmed by the radiologists on “current” FFDM examinations or through

biopsy (either core biopsy or surgical biopsy), while the other cases remained normal or benign. The malignant case was defined as breast with ductal carcinoma *in situ* (DCIS), invasive ductal carcinoma, or invasive lobular carcinoma, and benign case refers to adenoma, cyst, or hamartoma. Based on the detections and diagnostic status and the verified pathology reports, we split the cases into four subgroups. Of these women, 68 were subsequently diagnosed as having cancer (interval cancer) before the next screening and they were placed in the first subgroup. The second subgroup included 73 women with positive diagnosis on current FFDM images and later confirmed as having cancer (screening-detected cancer) in biopsies or other pathology examinations. The 67 cases in the third group were those detected abnormalities or suspicious findings on current FFDM examinations but later proved to be benign in the following workups (imaging examinations and/or biopsies). The last subgroup included 132 women who remaining screened negative during the sequential FFDM examinations and not recalled by the radiologists. The current FFDM screening examinations were occurred no later than 3 yr after previous screening mammography (i.e., prior FFDM screening examinations). Questionnaires were provided to all the participants to obtain information of common risk factors including age, body mass index (BMI), family medical and cancer history, hormonal status, smoking history, and age at first term pregnancy.

For the mammographic density among our dataset, 15 cases, 114 cases, 179 cases, and 32 cases were respectively rated by radiologists as almost entire fatty (BIRADS I), scattered fibroglandular (BIRADS II), heterogeneously dense (BIRADS III), and extremely dense (BIRADS IV), and Fig. 1 shows the distribution of the cases in each of four subgroups. All these ratings were from the original clinical report. In this study, the FFDM images acquired from the prior negative examinations were selected and analyzed, and all the mammograms used here were CC view images of right and left breasts. The goal of this study is to develop a new computerized scheme to classify risk assessment between 141 high-risk women (subgroup 1 and 2) who developed cancer and/or abnormalities that were detectable by radiologists before or during the next sequential screening examinations

and 199 low-risk women (subgroup 3 and 4) who remained cancer-free in the next sequential screenings.

2.B. Mammographic subregions segmentation

Prior to automated breast subregion segmentation, the computerized scheme was applied to preprocessing every mammogram in three steps. First, the scheme enhances images by using an adaptive tree-structured nonlinear filtering (TSF) method to reduce image noises and artifacts. Second, a directional wavelet transform (DWT) was used for multiorientation signal decomposition. In the last step, a tree-structured wavelet transform (TSWT) was applied as an efficient multiresolution enhancement method. The details of these algorithms can be found in our previous publications.^{23,31,32}

Once the breast region was identified, we performed a k -class fuzzy c means (FCM) clustering to partite it into several subregions and each one has relatively homogeneous density values. Instead of stratifying breast areas into dense and fatty regions only,^{30,33–35} more subregions were extracted and analyzed from the original breast depicted on mammogram. In this study, the number of clusters k was set to 5, so every breast region was separated into five areas. Based on the level of density values, we sorted and marked these five areas from the densest as area 1 to the fattiest as area 5. Then, the five areas were combined into five different subregions for following feature extraction with a combining rule as listed in Table I. After the combination, subregion 1 is the densest area, subregion 2 contains subregion 1 and the second densest area, and subregion 5 is the whole breast area.

2.C. Computational density and texture feature analysis

In order to characterize spatial and texture information of mammograms, for each extracted subregion (as listed in Table I), five groups of computational features were computed. The first group of features was extracted from the directional texture images which were generated from multiorientation signal decomposition. Researches showed that except for malignant lesions, some benign lesions like fibrous tissue, lipid-filled spaces surrounded by histiocytes, or sclerotic stroma can also cause spiculations or fibroglandular tissue distortions,³⁶ and these distortions can be used as indicators of cancer risk. In this study, ray-tracing algorithm

FIG. 1. Distribution of the four subgroups (interval cancer, screening-detected cancer, benign, negative) of cases in four categories of mammographic density (BIRADS) ratings.

TABLE I. Description of five subregions (from subregion 1 the densest region to subregion 5 the whole breast).

Subregion	Area number
1	1
2	1 + 2
3	1 + 2 + 3
4	1 + 2 + 3 + 4
5	1 + 2 + 3 + 4 + 5

Densest region
↓
Whole breast

FIG. 2. Two examples of directional images. (a), (c) Original regions of interest (ROIs); (b), (d) corresponding directional texture ROI images.

was used to determine the directional features. First, the background detail signals, including weak random directional lines, were eliminated and then strong directional signals were isolated, which were possibly directional features. Figure 2 shows two examples of the directional texture images; the major directional tissue like spiculations or fibroglandular tissues were enhanced and can be easily observed. From this image, five features were computed, including the number of directional tissues, their principle orientations determined by gradient operators, average normalized directional tissue length, the number of directional tissues that were computed based on directional texture image with a length exceeding a given threshold, and the directional angular distribution that was computed as entropy.

The second group of features was computed based on Gabor filters. The features provide spatial response profiles that are similar to the receptive field of human vision since similar textured regions show similar frequency spectra.³⁷ The possibility wave is transformed into a probability matrix, and the resulting coefficients represent the probability of the intensity of the wave.³⁸ The following features were computed in this group: contrast, angular second moment, inverse difference moment, entropy, variance, and correlation.

The third and fourth group included statistical texture features that were, respectively, extracted from the sum and difference histograms (SDHs)³⁹ and the gray level co-occurrence matrices (GLCMs).⁴⁰ The features in both of these groups have the high correlation with mammographic density. From

SDHs, three different relative displacements ($d = 1, 3, 5$ pixels) and four different angles ($\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$) were considered for computational feature extraction. For each pair of parameters, the mean, variance, entropy, energy, correlation, the third and fourth moment of mean, homogeneity, cluster shade, and cluster prominence were calculated. GLCM was created by calculating how often a pixel with gray level value i occurs horizontally adjacent to a pixel with value j . Mean, variance, contrast, correlation, energy, entropy, and homogeneity features were then extracted from these matrices.

The fifth group included pixel intensity features. The features are (1) energy, (2) entropy, (3) inertia, (4) mean, (5) standard deviation, (6) skewness, (7) kurtosis, (8) smoothness, (9) mean gradient, (10) the number of pixels with the maximum gray value in the histogram and the total number of pixels of the whole calculated region, (11) the number of pixels with a gray value larger than the average value of the histogram and the total number of pixels of the breast, (12) the average difference of two adjacent values in the histogram, (13) the average value of the histogram, (14) the standard deviation of the histogram, and (15) the uniformity of intensity in the histogram.

For the features listed above, we individually calculated each of them on every one of the five subregions. Every computational feature mentioned above will result in five multiscale features (i.e., the features computed on five different subregions), and a total number of 765 features were computed. To minimize the risk of overfitting, deleting

redundant features and building an optimal feature set are necessary. Sequential forward floating selection (SFFS) was used as feature selection method in this study, and it keeps the parameters flexibly changing so as to approximate the optimal solution as much as possible.⁴¹ For the traditional sequential forward and sequential backward procedures, once selected, the features cannot be discarded later and the discarded features cannot be selected again, so the result is usually only suboptimal. In SFFS, the nesting of feature subsets was prevented by employing either the (l,r) or generalized (l,r) algorithm which involves successive augmentation and depletion processes. In each step, SFFS “plus- l -minus- r ” features with parameters l and r being “float,” until the optimal feature combination was found.

2.D. Classification and data analysis

From the selected features, a machine learning classifier was performed to make predictions for the likelihood of women developing breast cancer or high-risk abnormality in near future. Among different types of machine learning techniques, support vector machine (SVM) was used in this study. SVM is a powerful data classification method using statistical theories to classify data points into two classes and has been widely used in different applications.⁴² Given a dataset $\{(\vec{x}_i, y_i)\}_{i=1}^n$ with n cases, where $\vec{x}_i$ represent each individual sample and $y_i = \pm 1$ shows its corresponding label. SVM identifies a hyperplane that maximizes the margin of data with different labels. Thus, the purpose of SVM is to find a pair $(\vec{w}, b)$ so that $y_i(\vec{w} \bullet \vec{x}_i + b) > 0$, $i \in (1, n)$. Using a well optimized svm training software package,⁴³ the global minimum or optimal result could be determined relatively easy and reliably and can substantially reduce the risk or the possibility of being trapped inside the local minimum. The radial basis kernel function (RBF) was used in this study, and it can be expressed as $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$.

For the performance evaluation, a tenfold cross-validation method was applied to train and test the SVM classifier. The 340 cases were randomly divided into ten exclusive folds (subgroups), and in each of iterations, nine folds were used as training data and the remaining one fold as the validation data. The process was iteratively repeated ten times until each fold was used once as the validation data. This tenfold cross-validation method made our data set more efficient as every mammogram can be used for both training and validation. For each testing case, the SVM classifier produces a risk classification score ranging from 0 to 1.

After obtaining the SVM classification results of all testing cases, we applied receiver operating characteristic (ROC) data analysis methods to assess the performance of our system in predicting the risk or probabilities of each woman developing high-risk breast lesions in the next sequential examinations. The area under the receiver operating characteristic curve (AUC) computed by ROCKIT program (ROCKIT, <http://www-radiology.uchicago.edu/krl/>) was used as the performance evaluation index. The positive predictive value and negative predictive value (NPV) were also assessed and reported. All the testing results were tabulated for comparison.

To evaluate the degree of association between each group of features and the risk of having breast cancer, we also compared normal cases versus (1) interval cancer cases, (2) screening-detected cancer cases, and (3) all mixed interval and screening-detected cancer cases. SFFS algorithm was applied to each group of features individually, and the AUCs along with the corresponding 95% confidence interval (CI) were calculated for comparison. Kendall's tau coefficient was calculated for the correlation of different feature groups. Some well-known risk factors, including age, BMI, family history of breast cancer, hormonal status, and age at first term pregnancy, were incorporated into the classification for adjusted AUC. To analyze the influence of multiscale features on prediction performance, comparisons were made using features computed from different subregions. To minimize the risk of overfitting, all the experiments were conducted by using tenfold cross-validation method. All the tests were two-sided, and a p value of <0.05 was considered as statistically significant.

3. RESULTS

In this section, we first represent an example of subregion extraction result, and Fig. 3 shows the original image, enhanced image, and five segmented subregions. Subregion 1 is the densest part of the mammogram, and subregion 5 is the whole breast area.

Table II shows the feature selection procedure described in Sec. 2.C. In the first step, the features were selected by SFFS algorithm within each group, and from the total 765 features, 34 were selected. After mixing these 34 selected features from different groups, SFFS was used again to choose the finalized feature set. The result showed that there were 12 finalized features, which included (1) principle orientation of the directional tissue determined by using gradient operators (subregion 4), (2) contrast of Gabor filter image (subregion 4), (3) angular second moment of Gabor filter image (subregion 4), (4) entropy of Gabor filter image (subregion 3), (5) the third moment of mean in SDHs ($\theta = 45^\circ$, $d = 3$) (subregion 1), (6) fourth moment of mean in SDHs ($\theta = 135^\circ$, $d = 3$) (subregion 1), (7) homogeneity in SDHs ($\theta = 45^\circ$, $d = 5$) (subregion 2), (8) variance in GLCMs (subregion 1), (9) energy in GLCMs (subregion 1), (10) mean gradient of gray value (subregion 5), (11) mean of gray value (subregion 4), and (12) standard deviation of the histogram (subregion 4).

Table III shows the risk discrimination results (the AUC values) among three different combinations of our dataset: screening negative (not recalled) cases versus (1) interval cancer cases, (2) screening-detected cancer cases, and (3) all mixed interval and screening-detected cancer cases using the selected image-based features in each of the five feature groups as well as the combined 12 finalized features. All the AUCs showed higher values in interval cancer cases than screening-detected cancer cases. Feature groups 2, 3, 4, and 5 can significantly separate the cancer cases ($p < 0.05$), but group 1 failed to separate cancer cases from control cases significantly ($p = 0.07$). When we incorporated all the

FIG. 3. An example of an original mammogram and its corresponding subregions. (a) Original image, (b) enhanced image, each gray value represents one segmented area, (c) subregion 1, (d) subregion 2, (e) subregion 3, (f) subregion 4, (g) subregion 5.

features together and used them to train the classifier, the p value dropped to as low as 10^{-5} . For both training data and testing data, interval cancer cases generated higher AUC values than screening-detected cancer cases.

The results also showed that the risk scores generated by the SVM using the image-based features in group 1 (directional features) have a low correlation with the scores generated using the features computed from other four feature groups. The correlation of risk scores generated by feature groups 3 and 4 (SDHs and GLCMs features) is relatively higher, since they are all statistical texture features. In addition, group 2 (Gabor features) has a relatively low correlation with the other feature groups, and group 5 (density features) has larger correlation with groups 3 and 4 than the other groups (Table IV).

Table V shows the evaluation of the proposed scheme using every group of features extracted from different subregions. We computed the features from different subregions and

calculated the AUC to stratify a total number of 340 cases in our dataset (including 141 cancer and 199 noncancer cases), respectively. And then within each group, the features extracted from different subregions were combined together. To reduce the redundancy of these features, SFFS method was applied for every feature group, and the AUC was computed using the selected features in each group. The results show that fusing the features computed from multiple (or all) regions typically yielded better performance.

The ROC curves using the features extracted from multiple subregions are shown in Fig. 4, which were generated by the `rockit` program with all 340 cases. The top ROC curve was generated by using both image and nonimage-based features (e.g., woman's risk factor information) and the computed AUC was 0.767 with a standard deviation of 0.019% and 95% CI of 0.730–0.803. Another ROC curve was generated using only image-based features in which AUC was 0.729 with a standard deviation of 0.021% and 95% CI of 0.688–0.771. The significance level of these two ROCs was less than 0.01. Figure 5 shows two similar ROC curves, which were generated by using the features extracted from the whole breast region (only subregion 5 was used). The adjusted AUC (including nonimage-based features) was 0.735 with a standard deviation of 0.020% and 95% CI of 0.702–0.765, while the unadjusted AUC (using image-based features only) was 0.698 with a standard deviation of 0.023% and 95% CI of 0.657–0.739, and the p value of these two ROCs was less than 0.01. Using our proposed multiscale features, 71.5% (243 of 340) of cases were correctly predicted when setting the classifier threshold to 0.54, which maximizes the NPV, whereas 28.5% (97 of 340) of cases were misclassified. Among the cancer cases, 67.6% (46 of 68) of interval cancer cases were correctly predicted

TABLE II. The process of feature selection; the number indicates the amount of total features, selected features, and finalized features in each feature group.

Feature group	Total features	Selected features	Finalized features
1	25	4	1
2	30	5	3
3	600	12	3
4	35	5	2
5	75	8	3
Total	765	34	12

TABLE III. Summary of AUC values in three different stratification subgroups using the selected features from each feature group and finalized features.

Feature group	AUC interval ($n = 200$)		AUC screen ($n = 205$)		AUC all cancers ($n = 273$)	
	Unadjusted	Adjusted for age	Unadjusted	Adjusted for age	Unadjusted	Adjusted for age
Group 1 ($n = 4$)	0.581 ± 0.033	0.623 ± 0.031	0.546 ± 0.034	0.579 ± 0.033	0.560 ± 0.030	0.612 ± 0.031
Group 2 ($n = 5$)	0.662 ± 0.025	0.684 ± 0.025	0.598 ± 0.027	0.622 ± 0.024	0.627 ± 0.025	0.663 ± 0.022
Group 3 ($n = 12$)	0.750 ± 0.026	0.761 ± 0.025	0.679 ± 0.021	0.698 ± 0.020	0.713 ± 0.025	0.736 ± 0.024
Group 4 ($n = 5$)	0.765 ± 0.023	0.777 ± 0.023	0.693 ± 0.022	0.714 ± 0.020	0.727 ± 0.021	0.742 ± 0.023
Group 5 ($n = 8$)	0.734 ± 0.019	0.752 ± 0.021	0.660 ± 0.019	0.688 ± 0.017	0.703 ± 0.018	0.735 ± 0.017
All ($n = 12$)	0.754 ± 0.023	0.772 ± 0.021	0.702 ± 0.021	0.716 ± 0.021	0.729 ± 0.021	0.767 ± 0.019

and 63.0% (46 of 73) of screening-detected cancer cases were correctly predicted. In the benign/normal case subgroup, the overall prediction accuracy was 75.9% (151 of 199). By using the single-scale features (features extracted from the whole breast area), the overall accuracy dropped to 61.5% (209 of 340). The results showed that 58.8% of the interval cancer cases (40 of 68), 53.4% of the screening-detected cancer cases (39 of 73), and 65.3% of the noncancer cases (130 of 199) were correctly predicted in the next sequential screening mammography examinations, respectively (Table VI). The experimental results indicated that our prediction model using multiscale features yielded higher accuracy than using single-scale features. In summary, the PPV using multiscale features was 0.657 (92 of 140) and the NPV was 0.755 (151 of 200).

4. DISCUSSION

Establishing a new optimal personalized breast cancer screening paradigm to replace current population-based uniform screening paradigm is an important approach to improve efficacy of mammography based breast cancer screening. Compared to the existing epidemiology study based risk factors and many genetic testing results (i.e., BRCA1/2 gene mutation positive), mammographic image features are not fixed or constant features. Thus, quantitative analysis of these features can not only detect breast tissue density at one time point but also measure the feature variations that may correlate well with the increased risk trend of having image-detectable breast cancers. Based on this assumption, we developed and presented a new SVM based scheme or risk model using a combination of different density and texture features extracted from different scales of subregions in the breast area. Our new model is different from the conventional

epidemiology study based risk models that focused on predicting long-term or lifetime cancer development risk.¹² Our new risk model aims to identify a fraction of women with high risk of having or developing the mammography-detectable breast cancer in the near-term (i.e., the next sequential mammography screening examination). The study results demonstrated a statistically significant (positive) associations (with $AUC = 0.767 \pm 0.019$) between the model-generated risk prediction scores and the actual near-term risk of a woman having an image-detectable breast abnormality that may lead to the development of breast cancer in the next subsequent examination. This study is also different from the previously reported near-term risk prediction studies^{23,25} that used the image features directly computed from the original image of the whole breast area. In this study, we tested and implemented a new multiscaled feature analysis method and systematically investigated the correlations of a large number of texture features on the multiscaled regions with the risk of cancer development.

In order to develop a high performance and robust mammogram based near-term breast cancer risk model, identifying the effective image features is one of the most important steps. In this study, we investigated a large number of features that have not been previously applied for the purpose of the near-term breast cancer risk prediction. From the study, we found that the features from group 1–4 are texture features associated with the breast disorganizations. These features may reflect the distortion of fibroglandular tissue, distribution of spiculations, breast disorganization, and heterogeneity. The correlation between texture features and density features is not high, but at the same time they are not totally independent (Table IV). The biological explanation for this phenomenon may be that the increase of cancer

TABLE IV. The correlation coefficients between different feature groups.

	Interval ($n = 68$)	Screen ($n = 73$)	Benign ($n = 67$)	Normal ($n = 132$)	All ($n = 340$)
Group 2 vs 3	0.06	0.09	0.11	0.17	0.11
Group 2 vs 4	0.12	0.10	0.15	0.21	0.17
Group 2 vs 5	0.19	0.16	0.25	0.29	0.23
Group 3 vs 4	0.50	0.53	0.58	0.69	0.64
Group 3 vs 5	0.22	0.27	0.34	0.41	0.32
Group 4 vs 5	0.31	0.30	0.39	0.48	0.37

TABLE V. AUCs computed using different features computed and selected from different subregions.

Feature group	AUC					
	Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5	All regions
Group 1	0.503 ± 0.057	0.519 ± 0.052	0.532 ± 0.032	0.560 ± 0.030	0.558 ± 0.026	0.560 ± 0.030
Group 2	0.608 ± 0.027	0.604 ± 0.023	0.609 ± 0.021	0.622 ± 0.019	0.613 ± 0.021	0.627 ± 0.025
Group 3	0.706 ± 0.024	0.695 ± 0.022	0.684 ± 0.026	0.68 ± 0.027	0.676 ± 0.020	0.713 ± 0.025
Group 4	0.716 ± 0.022	0.708 ± 0.026	0.699 ± 0.026	0.683 ± 0.025	0.681 ± 0.022	0.727 ± 0.021
Group 5	0.682 ± 0.019	0.674 ± 0.016	0.671 ± 0.023	0.665 ± 0.020	0.661 ± 0.021	0.703 ± 0.018

risk can be captured as an increase in mammographic density; structural and texture changes of breast tissue can be indicated by the texture features. These texture features do not change with the size of dense tissue regions so they can be used as complementary information for cancer risk prediction, which has not been quantitatively used in previous cancer risk studies.

Since breast or mammographic density is regarded as a strong risk factor for breast cancer,^{12,15,44} we also found that the density features we proposed in this study (feature group 5) showed strong association with breast cancer risk. In this study, we used multiscale features instead of single-scale features, as regions with different densities and texture complexities have different information to contribute differentiating women into high and low cancer risk in the near-term. When extracting these features from different subregions on mammogram according to different densities, a significant AUC value increase (0.042) was observed (Table V). Similar to density features, multiscale texture features also result in higher AUC (Table V). When using all density and texture features (feature group 1–5) from all these subregions, the AUC values increased 0.032 (Figs. 4 and 5). Thus, extracting and fusing features computed from different scale levels of subregions on mammogram may prove to be

an effective way to enhance the risk prediction performance. Some previous studies showed that extracting density features from the dense region³⁰ or retroareolar region⁴⁵ led to better prediction accuracy. In our study, the breast area was divided into five subregions according to intensity of each pixel, and all the density and texture features were computed within different subregions. The results of this study demonstrated that the classifier using multiscale features had superior performance measured in terms of accuracy and AUC than the classifier using features computed from the whole breast area (subregion 5 in Table V).

The results of this study also indirectly support an important hypothesis that quantitative analysis of mammographic image features and their variations can help predict the near-term breast cancer risk. Our dataset includes two types of positive cases: interval cancer and screening-detected cancer. The higher prediction performance in the interval cancer subgroup than the screening-detected cancer subgroup (Table III) indicates a positive association or increasing trend between the mammographic image feature variations and the cancer risk increase (or the decrease of time lag between the negative and positive detections), which also agrees with the previous findings that the interval cancers exhibit higher density.^{28,46} In addition, this study also showed the potential of

FIG. 4. The ROC curves using multiscale features.

FIG. 5. The ROC curves using single-scale features.

TABLE VI. Confusion matrix using our developed scheme to stratify testing cases into positive (cancer) and negative (noncancer) groups with a threshold to maximize the NPV on the classifier.

Predictions		Benign/normal	Cancer		Total
			Interval	Screen	
Multiple region features	Negative	151	22	27	200
	Positive	48	46	46	140
Single region features	Negative	130	28	34	192
	Positive	69	40	39	148
Total		199	68	73	340

further improving risk prediction performance by fusing the image and nonimage-based features (i.e., woman's age and other demographic or genetic features) in the future studies.

Despite the promising results discussed above, this study also has a number of limitations. First, our dataset was relatively small, which is not sufficient to represent the general screening population seen in clinical practice. Thus, the robustness and generalizability of the reported results need to be further tested in future studies. Second, in this study, we separate the breast area into five scales of subregions, but the most suitable number of subregions has not been investigated in this study. Third, this study is a retrospective study, and 41% cases in our dataset are high risk, which is significantly higher than the general population. Thus, this may increase the PPV and hard to make a direct comparison with other models built on the dataset following general population risk distribution. Fourth, the investigated texture features may not be the optimal features. Many of other features can still be explored and tested, especially the texture features based on specific regions on mammogram like Gaussian multiscale features.²⁸ In addition, more nonimage-based risk factors will be incorporated to our system in the future.

In summary, in this study, we tested a new approach to extract multiscale density and texture features of mammograms. Based on these features, we developed a new scheme to predict the likelihood of a woman developing breast cancer in near-term after a negative screening episode. Hence, our study demonstrated that the quantitative mammographic image feature analysis has the potential to significantly increase the discriminatory power (the clinically acceptable positive and negative predictive values) to predict near-term risk of women developing breast cancer, which is an important step to help improve efficacy of mammography screening by eventually establishing a new optimal personalized breast cancer screening paradigm.

ACKNOWLEDGMENTS

This work was partially supported by National Institutes of Health (SC1CA166016), National Science Foundation (DUE-TUES-1246050), and Department of Education (Award No. P031S120131). The authors also thank the Border Biological Research Center (BBRC) Core Facilities at The University of Texas at El Paso (UTEP) for their support, which were funded by RCM1-NIMHD-NIH Grant No. 8G12MD007592.

^{a)}Author to whom correspondence should be addressed. Electronic mail: wqian@utep.edu; Telephone: (915) 747-8090; Fax: (915) 747-7871.

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

²US Preventive Services Task Force, "Screening for breast cancer: US preventive services task force recommendation statement," *Ann. Intern. Med.* **151**(10), 716–726 (2009).

³D. A. Berry, K. A. Cronin, S. K. Plevritis, D. G. Fryback, L. Clarke, M. Zelen, J. S. Mandelblatt, A. Y. Yakovlev, J. D. Habbema, and E. J. Feuer, "Effect of screening and adjuvant therapy on mortality from breast cancer," *N. Engl. J. Med.* **353**, 12–20 (2005).

⁴R. A. Smith, S. W. Duffy, and L. Tabár, "Breast cancer screening: The evolving evidence," *Oncology* **26**(5), 471–486 (2012).

⁵H. G. Kaplan and J. A. Malmgren, "The breast cancer overdiagnosis conundrum: An oncologist's viewpoint," *Ann. Intern. Med.* **158**(1), 60–61 (2013).

⁶R. A. Hubbard, K. Kerlikowske, C. I. Flowers, B. C. Yankaskas, W. Zhu, and D. L. Miglioretti, "Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: A cohort study," *Ann. Intern. Med.* **155**(8), 481–492 (2011).

⁷J. Brodersen and V. D. Siersma, "Long-term psychosocial consequences of false-positive screening mammography," *Ann. Fam. Med.* **11**(2), 106–115 (2013).

⁸M. J. Yaffe and J. G. Mainprize, "Risk of radiation-induced breast cancer from mammographic screening," *Radiology* **258**(1), 98–105 (2011).

⁹D. S. Buist et al., "Influence of annual interpretive volume on screening mammography performance in the United States," *Radiology* **259**(1), 72–84 (2011).

¹⁰L. Berlin and F. M. Hall, "More mammography muddle: Emotions, politics, science, costs, and polarization," *Radiology* **255**(2), 311–316 (2010).

¹¹J. T. Schousboe, K. Kerlikowske, A. Loh, and S. R. Cummings, "Personalizing mammography by breast density and other risk factors for breast cancer: Analysis of health benefits and cost-effectiveness," *Ann. Intern. Med.* **155**, 10–21 (2011).

¹²E. Amir, O. C. Freedman, B. Seruga, and D. G. Evans, "Assessing women at high risk of breast cancer: A review of risk assessment models," *JNCI, J. Natl. Cancer Inst.* **102**(10), 680–691 (2010).

¹³L. Irwig, N. Houssami, and C. van Vliet, "New technologies in screening for breast cancer: A systematic review of their accuracy," *Br. J. Cancer* **90**, 2118–2122 (2004).

¹⁴M. H. Gail and P. L. Mai, "Comparing breast cancer risk assessment models," *JNCI, J. Natl. Cancer Inst.* **102**(10), 665–668 (2010).

¹⁵K. Kerlikowske, A. J. Cook, D. S. Buist, S. R. Cummings, C. Vachon, P. Vacek, and D. L. Miglioretti, "Breast cancer risk by breast density, menopause, and postmenopausal hormone therapy use," *J. Clin. Oncol.* **28**(24), 3830–3837 (2010).

¹⁶C. M. Vachon, T. A. Sellers, E. E. Carlson, J. M. Cunningham, C. A. Hilker, R. L. Smalley, and V. S. Pankratz, "Strong evidence of a genetic determinant for mammographic density, a major risk factor for breast cancer," *Cancer Res.* **67**(17), 8412–8418 (2007).

¹⁷J. N. Wolfe, "Breast patterns as an index of risk for developing breast cancer," *Am. J. Roentgenol.* **126**(6), 1130–1137 (1976).

¹⁸American College of Radiology (ACR), Breast Imaging Reporting and Data System Atlas, BI-RADS® Atlas (American College of Radiology, Reston, VA, 2003).

¹⁹W. A. Berg et al., "Breast imaging reporting and data system: Inter- and intraobserver variability in feature analysis and final assessment," *Am. J. Roentgenol.* **174**(6), 1769–1777 (2000).

- ²⁰R. Highnam, X. Pan, R. Warren, M. Jeffreys, G. D. Smith, and M. Brady, "Breast composition measurements using retrospective standard mammogram form (SMF)," *Phys. Med. Biol.* **51**(11), 2695–2713 (2006).
- ²¹Y. H. Chang, X. H. Wang, L. A. Hardesty, T. S. Chang, W. R. Poller, W. F. Good, and D. Gur, "Computerized assessment of tissue composition on digitized mammograms," *Acad. Radiol.* **9**(8), 899–905 (2002).
- ²²S. Timp, C. Varela, and N. Karssemeijer, "Temporal change analysis for characterization of mass lesions in mammography," *IEEE Trans. Med. Imaging* **26**(7), 945–953 (2007).
- ²³W. Sun, B. Zheng, F. Lure, T. Wu, J. Zhang, B. Y. Wang, and W. Qian, "Prediction of near-term risk of developing breast cancer using computerized features from bilateral mammograms," *Comput. Med. Imaging Graphics* **38**(5), 348–357 (2014).
- ²⁴J. Wei *et al.*, "Association of computerized mammographic parenchymal pattern measure with breast cancer risk: A pilot case-control study," *Radiology* **260**(1), 42–49 (2011).
- ²⁵M. Tan, B. Zheng, P. Ramalingam, and D. Gur, "Prediction of near-term breast cancer risk based on bilateral mammographic feature asymmetry," *Acad. Radiol.* **20**, 1542–1550 (2013).
- ²⁶J. A. Tice, E. S. O'Meara, D. L. Weaver, C. Vachon, R. Ballard-Barbash, and K. Kerlikowske, "Benign breast disease, mammographic breast density, and the risk of breast cancer," *JNCI, J. Natl. Cancer Inst.* **105**(14), 1043–1049 (2013).
- ²⁷A. Manduca, M. J. Carston, J. J. Heine, C. G. Scott, V. S. Pankratz, K. R. Brandt, and J. R. Cerhan, "Texture features from mammographic images and risk of breast cancer," *Cancer Epidemiol., Biomarkers Prev.* **18**(3), 837–845 (2009).
- ²⁸M. Nielsen, G. Karemore, M. Loog, J. Raundahl, N. Karssemeijer, J. D. M. Otten, and C. Christiansen, "A novel and automatic mammographic texture resemblance marker is an independent risk factor for breast cancer," *Cancer Epidemiol.* **35**(4), 381–387 (2011).
- ²⁹L. Haberle *et al.*, "Characterizing mammographic images by using generic texture features," *Breast Cancer Res.* **14**, R59 (2012).
- ³⁰Y. N. Law, M. K. Lieng, J. Li, and D. A. A. Khoo, "Automated breast tissue density assessment using high order regional texture descriptors in mammography," *Proc. SPIE* **9035**, 90351Q (2014).
- ³¹W. Qian, X. Sun, D. Song, and R. A. Clark, "Digital mammography: Wavelet transform and Kalman-filtering neural network in mass segmentation and detection," *Acad. Radiol.* **8**, 1074–1082 (2001).
- ³²W. Qian, D. Song, M. Lei, R. Sankar, and E. Eikman, "Computer-aided mass detection based on ipsilateral multiview mammograms," *Acad. Radiol.* **14**(5), 530–538 (2007).
- ³³C. K. Glide-Hurst, N. Duric, and P. Littrup, "A new method for quantitative analysis of mammographic density," *Med. Phys.* **34**, 4491–4498 (2007).
- ³⁴A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Perez, E. R. E. Denton, and R. Zwiggelaar, "A novel breast tissue density classification methodology," *IEEE Trans. Inf. Technol. Biomed* **12**, 55–65 (2008).
- ³⁵B. M. Keller, D. L. Nathan, Y. Wang, Y. Zheng, J. C. Gee, E. F. Conant, and D. Kontos, "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," *Med. Phys.* **39**(8), 4903–4917 (2012).
- ³⁶T. Franquet, C. De Miguel, R. Cozcolluela, and L. Donoso, "Spiculated lesions of the breast: Mammographic-pathologic correlation," *RadioGraphics* **13**(4), 841–852 (1993).
- ³⁷H. Handels, *Medizinische Bildverarbeitung* (B.G. Teubner Verlag, Leipzig, 2000).
- ³⁸C. H. Wei, Y. Li, and C. T. Li, "Effective extraction of Gabor features for adaptive mammogram retrieval," in *2007 IEEE International Conference on Multimedia and Expo* (IEEE, New York, NY, 2007), pp. 1503–1506.
- ³⁹M. Unser, "Sum and difference histograms for texture classification," *IEEE Trans. Pattern Anal. Mach. Intell. PAMI* **8**(1), 118–125 (1986).
- ⁴⁰R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.* **3**(6), 610–621 (1973).
- ⁴¹P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.* **15**(11), 1119–1125 (1994).
- ⁴²N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* (Cambridge University Press, Cambridge, England, 2000).
- ⁴³C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**(27), 1–27 (2011), software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm.10.1145/1961189.1961199>.
- ⁴⁴H. D. Nelson, B. Zakher, A. Cantor, R. Fu, J. Griffin, E. S. O'Meara, and D. L. Miglioretti, "Risk factors for breast cancer for women aged 40 to 49 years: A systematic review and meta-analysis," *Ann. Intern. Med.* **156**(9), 635–648 (2012).
- ⁴⁵Z. Huo, M. L. Giger, O. I. Olopade, D. E. Wolverton, B. L. Weber, C. E. Metz, and S. A. Cummings, "Computerized analysis of digitized mammograms of BRCA1 and BRCA2 gene mutation carriers 1," *Radiology* **225**(2), 519–526 (2002).
- ⁴⁶N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, and M. J. Yaffe, "Mammographic density and the risk and detection of breast cancer," *N. Engl. J. Med.* **356**(3), 227–236 (2007).