Prediction of Near-term Risk of Developing Breast Cancer Using Computerized Features from Bilateral Mammograms

Wenqing Sun[1], PhD, Bin Zheng[2], PhD, Fleming Lure[1], PhD, Teresa Wu[3,4], PhD, Jianying Zhang[5], PhD, M.D., Benjamin Y. Wang[6], M.D., Edward C. Saltzstein[7], M.D., Wei Qian[1,4], PhD

1. Department of Electrical and Computer Engineering, University of Texas, El Paso, TX
2. School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK
3. School of Computing, Informatics, Decision Systems Engineering, Arizona State University, AZ
4. Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang, China
5. Department of Biology, University of Texas, El Paso, TX
6. Radiology Department, Sierra Providence Health Network, El Paso, TX
7. University Breast Care Center at the Texas Tech University Health Sciences, El Paso, TX

Corresponding author:
Wei Qian, Ph.D.
Professor
Department of Electrical and Computer Engineering
Director: Medical Imaging Informatics Laboratory
College of Engineering, University of Texas, El Paso
500 West University Avenue
El Paso, Texas 79968
PH: (915) 747-8090
FX: (915) 747-7871
http://ee.utep.edu/facultyqian.htm
E-mail addresses: wqian@utep.edu

Prediction of Near-term Risk of Developing Breast Cancer Using Computerized Features from Bilateral Mammograms

Wenqing Sun[1], PhD, Bin Zheng[2], PhD, Fleming Lure[1], PhD, Teresa Wu[3,4], PhD, Jianying Zhang[5], PhD, M.D., Benjamin Y. Wang[6], M.D., Edward C. Saltzstein[7], M.D., Wei Qian[1,4], PhD

ABSTRACT

Asymmetry of bilateral mammographic tissue density and patterns is a potentially strong indicator of having or developing breast abnormalities or early cancers. The purpose of this study is to design and test the Global asymmetry features from bilateral mammograms to predict the near-term risk of women developing detectable high risk breast lesions or cancer in the next sequential screening mammography examination. The image dataset includes mammograms acquired from 90 women who underwent routine screening examinations, all interpreted as negative and not recalled by the radiologists during the original screening procedures. A computerized breast cancer risk analysis scheme using four image processing modules, including image preprocessing, suspicious region segmentation, image feature extraction, and classification was designed to detect and compute image feature asymmetry between the left and right breasts imaged on the mammograms. The highest computed area under curve (AUC) is 0.754±0.024 when applying the new computerized aided diagnosis (CAD) scheme to our testing dataset. The positive predictive value and the negative predictive value were 0.58 and 0.80, respectively.

Key Words: Breast cancer; mammogram; computerized breast cancer risk analysis; bilateral mammographic asymmetry feature; near-term breast cancer risk assessment.

I. Introduction

Breast Cancer is the most common cancer and second leading cause of cancer deaths of women [1]. Scientific evidence has shown that early cancer detection is important to enhance the survival rates of the patients through more effective patient management and treatment [2, 3]. Since the majority of breast cancers are detected in women with no known risk factors defined in the existing epidemiology models [4-5], a uniform mammography screening program in the general population is currently applied and considered important [3]. However, due to the large variability in the depiction of breast abnormalities, the overlapping dense fibroglandular tissue on the projection images and the low cancer prevalence in the screening environment, both detection sensitivity and specificity of screening mammography are relatively low [6-8]. To help radiologists improve detection and diagnosis performances in reading and interpreting screening mammograms, a great amount of research has been conducted to develop CAD systems or schemes including our work on developing a variety of two-dimensional computerized image analysis algorithms optimized to enhance the performance of the traditional CAD systems during the last two decades (e.g., [9-13]). Currently, a number of commercialized CAD schemes, including the one originally developed in our group and then being licensed to Carestream Health, Inc [14, 15], are widely used in the clinical practice to assist radiologists in reading and interpreting mammograms to date.

However, a number of recently reported studies made the debate related to the efficacy (risk-benefit and cost-benefit) of population-based screening mammography more controversial [16, 17]. Although reducing mammography screening interval (e.g., from annual screening to screening every two or more years) has the risk of missing early cancers, the currently frequent X-ray mammography screening and the associated high false-positive recall rates contribute to

higher cost as well as the unnecessary harms to many cancer-free women who routinely participate in the recommended mammography screening. As a result, since more scientists or researchers have realized the significance of establishing a more effective risk evaluation system to improve the accuracy of early breast cancer risk prediction, shifting the current population-based breast cancer screening paradigm to a new and optimal personalized screening paradigm in which women should be screened differently (including both the screening interval and screening imaging methods) has been attracting wide research interest [18-22]. To realize this ultimate goal, it is required to develop and establish breast cancer risk prediction or stratification models that have significantly higher (clinically acceptable) discriminatory power or positive predictive value (PPV) to stratify women into high and low risk groups of having or developing breast cancer in the near-term (e.g., ≤ 5 years). Hence, a small group of high risk women should be more aggressively screened and monitored for the early detection of breast cancer, while the majority of low risk women can be screened in a longer interval to reduce the potential harm (i.e., false-positive [23] and added cancer or other health risk [24]) until their risk statuses are changed. Unfortunately, all existing epidemiology based breast cancer risk models [5, 25] do not have clinically acceptable accuracy for this purpose [26]. As a result, much innovative research work is needed to develop the new near-term risk models that are potentially clinically useful or acceptable for breast cancer screening purpose. Among the different efforts, previous studies have shown bilateral mammographic image feature asymmetry, the difference of imaged breast size [27] and average density between the left and right breast [28]. This might be a strong risk indicator of developing breast cancer in the near-term (e.g., the next sequential screening examinations) because such image phenotype variation may associate with genotype abnormality to break human natural bilateral symmetry in the paired morphological traits, including breasts,

which may lead to cancer development. Instead of using registration or alignment techniques, [28] designed global features from the whole breast area as the measurement of bilateral mammogram asymmetry. Thus the potential misalignment and poor registration between bilateral mammograms results from the differences in breast compression or positioning and changes in breast itself [29] are avoided. However, to the best of our knowledge, all the global asymmetry features are constraint in spatial feature domain, no global morphological features or textural features of asymmetry measurement are discussed in existing literatures.

In this study, the global morphological and textural asymmetry features are designed and we also investigate and test the possibility of converting and/or applying the traditional CAD schemes to breast cancer risk analysis schemes using the current negative mammograms (e.g., the negative mammograms acquired in the "prior" examinations) to predict the likelihood of women having high-risk breast lesions or cancers being detected in the near-term based on the computerized image feature analysis from the bilateral (left and right breast) mammograms. Our hypothesis is that the bilateral image feature difference computed from the left and right breasts should also provide useful information indicating the asymmetry of breast tissue structures that might be directly related to the development of high risk breast lesions.

II. Materials and Method

2.1. Database

From an established in-house full-field digital mammography (FFDM) image database, we randomly selected 90 pairs of bilateral mammograms, including 60 pairs of cranio-caudual (CC) view and 30 pairs of Mediolateral-oblique (MLO) view cases from 90 women who

underwent routine screening mammography examinations. The women's ages range from 32 to 64, with a mean age of 43.7 years and median age of 43. Among the data, 4 cases, 28 cases, 49 cases and 9 cases were respectively rated by radiologists as almost entire fatty (BIRADS I), scattered fibro-glandular (BIRADS II), heterogeneously dense (BIRADS III) and extremely dense (BIRADS IV). All of these mammograms (treated as "prior" mammograms in this study) were interpreted as screening normal (without recall) by the radiologists; however, in the next sequential (annual) FFDM screening examinations, 34 of these women were recalled by radiologists due to the highly suspicious findings depicted on the FFDM images. Through additional imaging examinations and/or biopsies, 20 women were diagnosed and confirmed having cancer. The rest of the 56 women remaining screened negative (not recalled) during the sequential FFDM examinations (as shown in Table 1). In this study, 90 pairs of left and right FFDM images acquired from the "prior" negative examinations were used and analyzed. The goal of this study is to develop a new computerized scheme to classify risk assessment between 34 high risk women who developed abnormalities that were detectable by radiologists during the next sequential screening examinations and 56 low-risk women who remained cancer-free and were not recalled by radiologists in the next sequential screening.

Table 1: Distribution of our testing dataset

2.2. Breast Cancer Risk Analysis Scheme

The proposed breast cancer risk analysis scheme was developed to detect the cases with high possibilities of having highly suspicious breast abnormalities detected in the next sequential screening examinations. Specifically, the scheme was divided into four primary image processing and data analysis modules, including (1) image preprocessing, (2) region

segmentation, (3) feature extraction and selection and (4) machine learning classification. Figure 1 shows the flowchart of the proposed scheme.

Figure 1: Flowchart of breast cancer risk analysis scheme using global asymmetry features from bilateral mammograms

A. Preprocessing module:

In order to enhance image features and reduce image noise, we implemented the following three image preprocessing functions that could have impact on the following procedures as well as the final analysis results. These are (1) concurrent image enhancement; (2) concurrent multiorientation transform; and (3) concurrent multiresolution transform.

First, the image enhancement eliminates noise and artifacts by using adaptive tree-structured nonlinear filtering (TSF). In order to maximize the effect of TSF, the standardization method was used to convert the mammogram to match our algorithm [30]. Then we apply central weighted median filter (CWMF) method and eight different variable filter windows to match different edge signals related to the breast tissues [31]. Compared to the preprocessing module in the single-view system, using either the right side mammogram or the left side mammogram, we need to process each pair of bilateral images at the same time and make sure each pair of corresponding parameters is the same. The CMWF and variable filter windows are able to decrease or suppress noise and increase details of breast tissues. Then, the TSF architecture integrate the designed symmetric filters and variable shape windows into a three-stage system [32]. This is a symmetric multistage filter which combines the advantages of central weighted median filters, linear and curved windows, and multistage operations that sequentially compare filtered and raw image data with the objective of obtaining more robust characteristics for noise

suppression and detail preservation. These steps resulted in spatial feature domain images of both left and right breasts for later use in bilateral spatial feature extraction.

Second, directional wavelet transform (DWT) is used for concurrent multiorientation signal decomposition. The input bilateral images are decomposed by the directional wavelet transform and each image yields two output images. Every case (bilateral mammograms) resulted in another four images: two are directional texture images (textural domain) which are used for directional feature extraction, and the other two are smoothed versions of the original images (morphological images), with directional information removed [33]. The importance of this step has already been shown in our previous breast cancer detection research, and even more necessary for this new application because (1) the directional feature is a good indicator of early cancer risk analysis in which the subtle abnormalities in early stage might be difficult to detect in early stages, and (2) the processed resulting image makes it possible to quantitatively analyze spiculations for potential abnormalities.

Third, the tree structured wavelet transform (TSWT) function is designed as an efficient concurrent multi-resolution enhancement method. A two-channel tree-structured wavelet transform is used here as a highly efficient multi-resolution representation method, as opposed to single-scale region-based image enhancement approach for suspicious areas. It combines the advantages of lower resolution characteristics, which are useful for localization of suspicious areas, with the advantages of higher resolution which is essential for details [33].

After applying this image preprocessing module, the proposed scheme generates 3 images for each mammogram including spatial image, texture image and morphology image [34].

Figure 2: An example of original image and other three images obtained from preprocessing module. a) original mammogram, b) spatial image, c) texture image and d) morphology image.

B. Segmentation module:

In this module, morphology images and texture images were processed with the purpose being to detect and segment the suspicious areas depicted on each image. We designed a spatial fuzzy C-means (FCM) clustering algorithm that works by grouping similar pixels into clusters in the feature domain. Having applied the filters, ROIs (region of interest) containing suspicious tissue structures or high-risk breast lesions were extracted from the preprocessed images, and the number of clusters were chosen by the filter results.

Compared to conventional fuzzy clustering method, we integrated the spatial information in our approach. As the pixels on the images are highly correlated, the spatial information is of great importance to be considered in the process of clustering. We sum the intensity membership value of each pixel in a 5 by 5 sized window as a spatial function, with the spatial function is defined as

$$h_{ij} = \sum_{n \times n} u_{ij} / n^2 \ ,$$

where u_{ij} is the membership value of each pixel. The value will be relatively larger if the majority of its neighborhood belongs to the same cluster [35]. By incorporate the spatial function into the membership function we get:

$$u'_{ij} = \frac{u_{ij}^p h_{ij}^q}{\sum_{k=1}^{c} u_{kj}^p h_{kj}^q} \ ,$$

where p and q are the weights of the membership function and spatial function.

Our clustering loop contains two steps namely cluster assignment step and move centroid step. In order to avoid local optimum, the fuzzy C-means algorithm runs iteratively to calculate the cost function each time so that the best clustering can gradually emerge and be selected. Then, an AND rule in fuzzy logic is used to combine intensity membership and spatial membership together. The reduced resolution image was used in the FCM method, since lower resolution images usually contain more homogenous regions. After the clusters in lower resolution images are labeled, the clusters with the same labels were then merge into high resolution images. As a result, the smoothed regions and detailed edges can be segmented and acquired at the same time. The final clustering is achieved iteratively by calculating cost function that is dependent on the distance of the pixels to the cluster centers in the feature domain [35]. The flowchart is shown in Figure 3, and Figure 4 shows the segmentation results marked on original images based on our spatial fuzzy C-means method.

Figure 3: Flowchart of a spatial fuzzy C-means method

Figure 4: The segmentation result using spatial fuzzy C-means method. For better illustration, the segmentation example is implemented on the mammogram with malignant mass tissue (a) original image (b) enhanced image (c) segmentation result

C. Feature computation and analysis

The third module of the proposed breast cancer risk analysis scheme computes and analyzes the related image features. We divided the features into two groups, one with unilateral

features (based on single breast mammogram), and the other with concurrent features (based on both breast mammograms). For convenience, all the features and their groups are presented in Appendix. Unilateral features depict the information of every single breast mammogram, and concurrent features describe the relationship of the bilateral mammograms.

a) Unilateral Features

For unilateral feature analysis, the proposed scheme computed the following groups of features including (1) morphological features: circularity, normalized deviation of radial length, the area of the extracted region in pixels, boundary irregularity factor; (2) textural features: the number of spiculations, the average spiculation length. The description of these features can be found in our previous papers [11-13].

In addition, we also investigated and used three texture features [36, 37]. These features were computed from the images generated by our modified multi-orientation transform using the following approaches.

1) Gabor Energy Feature. This feature combines the outputs of symmetric and anti-symmetric kernel filter together and defined as:

$$e_{\lambda,\Theta}(x,y) = \sqrt{r^2_{\lambda,\Theta,0}(x,y) + r^2_{\lambda,\Theta,-\left(\frac{1}{2}\right)\pi}(x,y)}$$

where $r(x,y) = \iint I(\xi,\eta)g(x-\xi,y-\eta)d\xi d\eta$. $r_{\lambda,\Theta,0}(x,y)$ and $r_{\lambda,\Theta,-\left(\frac{1}{2}\right)\pi}(x,y)$ are the responses of the linear symmetric and anti-symmetric Gabor filters, respectively. And these are standard formulas of Gabor filters: $g_{\lambda,\Theta,\varphi}(x,y) = e^{-((x'^2+\gamma^2 y'^2)/2\sigma^2)}\cos(2\pi\frac{x'}{\lambda}+\varphi)$,

$x' = x\cos\Theta + y\sin\Theta$, $y' = -x\sin\Theta + y\cos\Theta$, $\sigma = 0.56\lambda$ and $\gamma = 0.5$.

2) Complex Moments Feature. This feature provides the information about the presence or absence of dominant texture orientations by using real and imaginary parts of the complex moments of the local power spectrum.

$$C_{mn}(x, y) = \iint (u + iv)^m (u - iv)^n \, \tilde{p}_{u,v}(x, y) \, du \, dv \qquad , \qquad m, n \in N$$

where $u = \dfrac{1}{\lambda} \cos \Theta$, $v = \dfrac{1}{\lambda} \sin \Theta$, $\tilde{p}_{u,v}(x, y) = p_{\lambda,\Theta}(x, y)$. It is proven that a complex moment of even order m+n is able to discriminate textures with (m+n)/2 dominant orientations [36].

3) Grating Cell Operator Feature. As discussed in [36, 37], grating cell operator contains two steps. The first step is to respond at any position to a set of three parallel bars with appropriate periodicity, orientation and position. The second step integrates the responses of a given preferred orientation and periodicity within a certain area to compute the response of a grating cell.

Alone with the morphological and textural features, a set of 9 spatial image features were also computed as unilateral features: (1) energy, (2) entropy, (3) inertial, (4) mean value, (5) standard deviation, (6) skewness, (7) kurtosis, (8) smoothness and (9) mean gradient.

b) Concurrent Features

The concurrent features represent the degree of bilateral mammogram asymmetry. In this step no alignment or registration between two bilateral breast images is required, as all the features we designed and used here are global features. Hence, for concurrent feature analysis, the features were acquired by calculating the absolute different ratios of certain measurements using features from each corresponding bilateral mammogram. Then using the measurements

listed below, we computed the following bilateral asymmetric features from each paired mammograms. In brief, based on the gray level histogram of the entirely segmented breast area six features $(1 - 6)$ are computed: (1) the ratio between the number of pixels with the maximum gray value in the histogram and the total number of pixels of the whole breast; (2) the ratio between 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; (3) the average difference of two adjacent values in the histogram; (4) the average value of the histogram; (5) the standard deviation of the histogram; and (6) the uniformity of intensity in the histogram. The rest of the seven features were computed to mimic the density BIRADS rated by radiologists based on threshold computation. Features (7-9) are the ratios between the average value of pixels under thresholds of 25%, 50%, and 75% of the maximum pixel value and the average pixel value of the whole breast area. Features (10-12) are the ratios between the number of pixels under each threshold and the total number of pixels of the entire segmented breast area. Feature 13 is the ratio between subtraction of the average value of pixels under 75% threshold minus average value of whole breast area and subtraction of average value of whole breast area minus average value of pixels under 25% threshold. Finally, feature 14 is the ratio of difference between number of pixels under 75% threshold and total number of pixels of whole breast area and subtraction of whole pixel number minus number of pixels under 25% threshold.

Besides computing image features from spatial domain, we also designed and computed bilateral asymmetry features in the textural domain and morphological domain. To our knowledge, no publication has discussed global bilateral textural and morphological features. However, from our research we found that some benign cases in "prior" examination also have certain suspicious dense areas, like lobular hyperplasia or benign masses, and the asymmetry of

these areas might also provide useful information for risk cancer development. If no suspicious areas can be found in the mammogram, the dense areas will be extracted. All these areas mentioned above were called suspicious areas in this study. Unlike the suspicious region defined by radiologists (on average much less than one per typical mammogram), the suspicious areas defined here indicate the regions identified in initial processing steps, which is as many as 839 from all the 90 pairs of cases. Instead of struggling to avoid misalignment of bilateral corresponding areas, we treat the entire segmented suspicious areas depicted on one image as a whole, and the average suspicious region measurement of each bilateral mammogram was used in our study. Listed below are the global features we designed in this study:

1. Ratio of total number of suspicious areas in each mammogram,

2. Ratio of total areas of extracted ROIs in pixels in each mammogram,

3. Ratio of average boundary irregularity factor of ROIs in each mammogram,

4. Ratio of average circularity of ROIs in each mammogram,

5. Ratio of average rectangularity of ROIs in each mammogram,

6. Ratio of number of spiculations in each mammogram,

7. Ratio of the average spiculation length in each mammogram,

8. Ratio of Gabor energy feature in each mammogram,

9. Ratio of complex moments feature in each mammogram,

10. Ratio of cell operator feature in each mammogram,

11. Ratio of entropy of suspicious areas in each mammogram,

12. Ratio of the standard deviations of spiculations in each mammogram.

D. Feature selection and classification:

After computing these features in different image feature domains, we analyzed the correlations among these features, deleting redundant features and building an optimal feature set. Based on Darwin's theory of evolution, all living things are based on the rule of "survival of fittest", by translating this into algorithms we search for the solutions in a more natural way, using the Genetic Algorithms (GA). GA search from a broad spectrum of possible solutions to find the optimal feature set [38]. It is a search technique that evaluates each individual's fitness through a fitness function. In this study, the fitness function for every chromosome is $fitness(m) = Az - N(m)/2000$, where Az is the performance from previous selected feature pool and N(m) is the number of selected features in chromosome m. We first calculated the fitness value with initial population, and then selected the individuals with relatively high values for mutation to produce the next generation. The crossover probability and mutation probability is 0.9 and 0.001 respectively. The process ended when the final condition satisfied. In order to solve the problem that the elimination of one feature will affect the elimination of the next feature, a global search technique was applied in our study [39]. The best set of features to be used in classification was selected by testing every possible combination of candidate features. To achieve this goal we started with a random population that converged when all members in the population had approximately the same fitness.

Having selected the features, the Support Vector Machine (SVM) method was used to classify and/or predict the likelihood of the test cases developing high risk breast lesions detectable in the next sequential FFDM screening examinations in our database. Compared with many other machine learning classifiers, SVM uses unique characteristics. Alone with faster training time, it uses large margin classifiers allowing the SVM to maximize the distance of the

training examples to the decision boundaries. Using a well optimized SVM training software package [40], the global minimum or optimal result could be determined relatively easy and reliably while substantially reducing the risk or the possibility of being trapped inside the local minimum. Limited by our dataset size, a 10-folder cross-validation method was used to train and test our SVM. We used nine folds as training data and the remaining one fold as the validation data. This process was iteratively repeated 10 times until each fold was used once as the validation data. This 10-fold cross-validation method made our data set more efficient as every mammogram can be used for both training and validation.

After obtaining the SVM classification results for all testing cases, we applied receiver operating characteristic (ROC) data analysis methods to assess CAD performance in predicting the probabilities of each case developing high risk breast lesions in the next sequential examinations. The area under the ROC curve computed using the ROCKIT program (ROCKIT, http://www-radiology.uchicago.edu/krl/) was used as the performance evaluation index. In addition, by maximizing the margin length from SVM classification, the dataset was divided into two subgroups of positive and negative cases. The positive and negative predictive value were also assessed and reported. All testing results were tabulated for comparison.

III. Results

To evaluate the efficiency of our proposed features, we applied each of the 12 global bilateral symmetric features to classify the cases in our dataset into two groups of positive and negative for development of breast abnormalities or early cancer. The average performance level measured by AUC is 0.5933 ranging from 0.5316 ±0.0303 to 0.6904 ±0.0231 (Figure 5). All twelve of our proposed global asymmetry features yielded an AUC more than the chance (AUC=0.5), including 5 of them achieving AUC more than 0.6. This means all these features

have certain ability to assess the bilateral breast asymmetry and hence identify a fraction of positive cases.

Figure 5: The distribution of the performance levels (the areas under receiver operating characteristic (ROC) curves and the standard deviations) when applying each of the proposed 12 features to classify 90 cases into the two groups of positive and negative for developing breast abnormalities or cancer.

All the features mentioned in Section C (including all the existing features and newly designed 12 global features) were used for GA feature selection, the following features were selected to build the classifier used in our CAD scheme: (1) the standard deviation of single pixel values of breast area; (2) the skewness of pixel values of single breast area; (3) the ratio between the number of pixels with the maximum gray value in the histogram and the total number of pixels of the whole breast; (4) the standard deviation of the histogram; (5) the ratio of average boundary irregularity factor; (6) the ratio of average circularity; (7) the ratio of number of spiculations. The formulations of the selected features are presented in Table 2. Due to the relatively low correlations among these seven features (Table 3), by combining these features the SVM yielded a significantly improved classification performance with AUC = 0.754 ± 0.024 (P<.001). The statistical analyses were performed using SAS software (SAS Institute, Cary, NC).

Table 2: The formulations of selected features

Table 3: The correlation coefficients of the selected seven features

To compare the performance of CC view mammograms and MLO view mammograms, we applied our algorithm to CC view data and MLO view data respectively and with AUC values (SVM selected as the classifier) being 0.762 ± 0.027 (CC view) and 0.7380 ± 0.026 (MLO view).

The ROC curves are shown in Figure 6. The truth files were acquired from the diagnostic results of the next sequential FFDM examinations and/or the biopsy (for the recalled and positive cases). Table 4 shows that among 34 positive high risk cases (from the truth file), 25 (73.5%) were classified or predicted as having high risk of developing detectable cancer breast lesions in the next sequential screening examinations among 56 benign cases, 38 (67.9%) were classified as low risk cases of developing detectable cancers in the next sequential screening examinations, which means that more than two thirds of the cases, 70.0% (63 out of 90 cases), were predicted correctly. In this dataset, the positive predictive value (PPV) and the negative predictive value (NPV) were 0.58 and 0.80, respectively.

Table 4: Summary of using our CAD scheme to stratify our testing cases into the positive (high risk) and negative (low risk) groups based on analysis of all features except our newly designed features (traditional features) and all combined features (traditional features and new features)

Figure 6: Comparison of three ROC curves generated by only MLO view cases, only CC view cases and both view cases

By removing all global asymmetry morphological and textural features (i.e. the 12 features we proposed in this paper) from the feature pool, we reselected the features and rebuilt the SVM in the same way using the spatial features only. The performance measured by the computed AUC goes down to 0.701 with standard deviation of 0.029 and 95% confidence interval of [0.672, 0.730]. The classification results are summarized in Table 5. The PPV and NPV when using spatial image feature asymmetry resulted in 0.50 and 0.74, respectively. The comparison in Table 4 shows that using our CAD scheme by adding morphological and textural

feature asymmetry computed from two bilateral mammograms could improve the performance of the scheme by increasing both PPV and NPV rate.

IV. Discussion

In this study we investigated and presented a new application value of the CAD schemes of mammograms, which aims to identify a fraction of women with high risk of having or developing detectable breast cancer. Computerized asymmetrical image feature analysis of two bilateral (left and right) mammograms is made available in the near-term after the negative screening examination of interest. This study has a number of unique characteristics. First, subtracting the registered bilateral mammograms to identify suspicious abnormalities based on image feature difference. This has been investigated and tested in early CAD development without much success because these CAD schemes typically yield substantially high false-positive detection rates and often unreliable results [42-45]. Hence, the biggest obstacle to compute and analyze the image feature differences on the two bilateral mammograms using CAD scheme is that the accurate image alignments or registration is required, which has been proven to be very difficult and not robust leading to the high false positive detection rates. Due to the complexity of breast cancers, misalignment is almost unavoidable. In this study, we adopted a different approach that did not require image registration while extracting asymmetry features from bilateral mammograms. Our scheme directly compared the differences of image features that were separately computed from two bilateral images.

Second, although using computerized global spatial or statistical image feature difference from two bilateral images to predict near-term breast cancer risk has been previously tested and

reported [28, 46], the disadvantage of these studies is that only the spatial features were used to analyze the cases while ignoring morphological features and texture features. In this study, we tested a new method that utilized features extracted from three domains to try to enhance performance. This work also aimed to add a new function to an existing CAD scheme to allow for not only detect the suspicious lesions in the positive cases but also predict the near-term risk of the negative cases. Our research showed that adding the morphological and texture features currently used to classify the suspicious lesions or regions to the process of risk prediction could increase prediction performance (as comparison shown in Tables 4). An increase of 0.053 in Az value, 0.08 in PPV rate, and 0.06 in NPV rate were observed after adding our proposed features to the scheme. Hence, in this approach the well-developed image feature analysis methods used in current CAD schemes can be applied to the dual tasks of detecting suspicious breast lesions and predicting future cancer risk. In this study, as we deliberately selected some features in morphological and texture domains and mixed them with other features computed from the spatial feature domain, our scheme was able to correctly predict 73.5% of cases that would have detectable cancer in the next sequential screening examinations. To the best of our knowledge, no similar studies have been reported to date.

Third, breast cancers occur in women with variety of image characteristics. Although current CAD performance depends on the case difficulty (e.g., breast density) [45], our approach is largely independent from these case difficult factors or varying image characteristics based on the assumption that two breasts of negative cases are relatively symmetrical. Similar to the most existing CAD schemes, our scheme is also likely to detect more suspicious regions on both left and right dense breasts and few suspicious regions in two fatty tissue dominated breasts. It is the tissue asymmetry that determines the risk of a woman having or developing breast cancer. Hence,

this approach should be more robust than the current CAD schemes in detecting suspicious breast abnormalities based on single images.

Despite these encouraging results, this was a preliminary study with a number of limitations. In particular, our testing dataset was very small, which was 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.

In summary, developing new near-term breast cancer risk assessment models that enables the stratification of women into two groups of high and low risk. The high discriminatory power (the clinically acceptable positive and negative predictive values) is extremely important to help develop and establish an optimal personalized breast cancer screening paradigm. In this study we tested a new approach that analyzes the bilateral mammographic image feature differences in three feature domains without image registration or alignment. This allows us to predict the likelihood of a woman developing breast cancer in near-term after a negative screening examination of interest. The results showed that it is feasible to apply our modified CAD scheme to predict the near-term breast cancer risk in a relatively large fraction of high risk cases. If successful, the prediction results from this or other similar new CAD-based breast cancer risk models may be used to raise a warning flag. Proper use of these methods assist radiologists in reducing the possibility of missing or overlooking these cases in the next sequential examinations or advise the women being more aggressively screened (e.g., using a short screening interval and other more sensitive imaging modalities) to increase the likelihood of the cancer being detected in an early stage.

Appendix: The feature groups and descriptions are presented below:

Groups		Feature descriptions
Unilateral Features	Morphological features	circularity, normalized deviation of radial length, the area of the extracted region in pixels, boundary irregularity factor
	Textural features	number of spiculations, the average spiculation length, Gabor energy feature, complex moments feature, grating cell operator feature
	Spatial features	energy, entropy, inertial, mean value, standard deviation, skewness, kurtosis, smoothness, mean gradient.
Concurrent Features	Spatial features	ratio between the number of pixels with the maximum gray value in the histogram and the total number of pixels of the whole breast
		ratio between 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
		average difference of two adjacent values in histogram;
		average value of histogram
		standard deviation of the histogram
		uniformity of intensity in the histogram
		ratio between the average value of pixels under threshold of 25% of the maximum pixel value and the average pixel value of the whole breast area
		ratio between the average value of pixels under threshold of 50% of the maximum pixel value and the average pixel value of the whole breast area
		ratio between the average value of pixels under threshold of 75% of the maximum pixel value and the average pixel value of the whole breast area
		ratio between the number of pixels under 25% threshold and the total number of pixels of the entire segmented breast area
		ratio between the number of pixels under 50% threshold and the total number of pixels of the entire segmented breast area
		ratios between the number of pixels under 75% threshold and the total number of pixels of the entire segmented breast area
		ratio between subtraction of average value of pixels under 75% threshold minus average value of whole breast area and subtraction of average value of whole breast area minus average value of pixels under 25% threshold
		ratio of difference between number of pixels under 75% threshold and total number of pixels of whole breast area and subtraction of whole pixel number minus number of pixels under 25% threshold.
	Textural and morphological features	ratio of total number of suspicious areas in each mammogram
		ratio of total areas of extracted ROIs in pixels in each mammogram;
		ratio of average boundary irregularity factor of ROIs in each mammogram
		ratio of average circularity of ROIs in each mammogram
		ratio of average rectangularity of ROIs in each mammogram
		ratio of number of spiculations in each mammogram
		ratio of the average spiculation length in each mammogram
		ratio of Gabor energy feature in each mammogram
		ratio of complex moments feature in each mammogram
		ratio of cell operator feature in each mammogram
		ratio of entropy of suspicious areas in each mammogram
		ratio of the standard deviations of spiculations in each mammogram

Reference:

1. Siegel, R., Naishadham, D., & Jemal, A. (2012). Cancer statistics, 2012. CA: a cancer journal for clinicians, 62(1), 10-29. Tabar L, Vitak B, Chen HH, et al, Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality, Cancer 2001; 91:1724-1731.

2. Smith RA, Cokkindes V, Brooks D, et al, Cancer screening in the United States, 2011, CA Cancer J Clin 2011; 61:8-30.

3. Madigan MP, Ziegler RG, Benichou J, et al, Proportion of breast cancer cases in the United States explained by well-established risk factors, J Natl Cancer Inst 1995; 87:1681-1685.

4. Amir E, Freedman OC, Seruga B, Evans DG, Assessing women at high risk of breast cancer: a review of risk assessment models, J Natl Cancer Inst 2010; 102:680-691.

5. Vachon CM, Sellers TA, Carlson EE, et al. Strong evidence of a genetic determinant for mammographic density, a major risk factor for breast cancer. Cancer Research 2007;67:8412–8.

6. Tang JS, Rangayyan RM, Xu J, Naqa EI., Yang YY, Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances, Information Technology in Biomedicine, IEEE Transactions on 2009; 13 (2): 236-251

7. Mandelson MT, Oestreicher N, Porter PL, et al, Breast density as a predictor of mammographic detection: comparison of interval-and screen-detected cancers, J Natl Cancer Inst 2000; 92:1081–1087.

8. Berg WA, Gutierrez L, NessAiver MS, et al, Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer, Radiology 2004; 233:830-849.

9. Zheng BY, Qian W, Clarke LP, Digital mammograghy: MF-based NN for automatic detection of microcalcifications, IEEE Trans. on Medical Imaging 1996; 589-598.

10. Glide-Hurst CK, Duric N, Littrup P. A new method for quantitative analysis of mammographic density. Medical Physics 2007;34: 4491–8.

11. Qian W, Li LH, Clarke LP, Image feature extraction for mass detection using digital mammography: Influence of wavelet analysis. Medical Physics 1999; 26:402-408.

12. Qian W, Sun X, Song D, Clark RA. A Novel Hybrid Filter Architecture for Image Enhancement and Detection in Digital Mammography. Computerized Medical Imaging and Graphics 2002; 233-242.

13. Qian W, Lei M, Song D, et al, Computer Aided Mass Detection based on Ipsilateral Multi-view Mammograms, Academic Radiology 2007;14:530-38.

14. Clarke, L. P., Qian, W., & Kallergi, M. (1998). U.S. Patent No. 5,825,936. Washington, DC: U.S. Patent and Trademark Office.

15. Clarke, L. P., Qian, W., & Li, L. (1998). U.S. Patent No. 5,799,100. Washington, DC: U.S. Patent and Trademark Office.

16. Berlin L, Hall FM, More mammography muddle: emotions, politics, science, costs and polarization, Radiology 2010; 255:311-316.

17. Jorgensen KJ, Is the tide turning against breast screening? Breast Cancer Research 2012; 14:107-109.

18. Schousboe JT, Kerlikowske K, et al, Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness, Ann Intern Med 2011; 155:10-20.

19. Highnam R, Pan X, Warren R, et al. Breast composition measurements using retrospective standard mammogram form (SMF). Physics in Medicine & Biology 2006; 51:2696–713.

20. Glide-Hurst CK, Duric N, Littrup P. A new method for quantitative analysis of mammographic density. Medical Physics 2007; 34: 4491–8.

21. Chang YH, Wang XH, Hardesty LA, et al. Computerized assessment of tissue composition on digitized mammograms. Academic Radiology 2002; 9:898–905.

22. S. Timp, C. Varela, and N. Karssemeijer, "Temporal change analysis for characterization of mass lesions in mammography," IEEE Trans. Med. Imag., vol. 26, no. 7, pp. 945–953, Jul. 2007, 10.1109/TMI.2007. 897392.

23. Hubbard RA, Kerlikowske K, Flowers CI, et al, Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: A cohort study, Ann Intern Med 2011; 155:481-492.

24. Yaffe MJ, Mainprize JG, Risk of radiation-induced breast cancer from mammographic screening, Radiology 2011; 258:98-105.

25. Irwig L, Houssami N and Vliet C van, New technologies in screening for breast cancer: a systematic review of their accuracy, British Journal of Cancer (2004) 90, 2118–2122. doi:10.1038/sj.bjc.6601836

26. Gail MH, Mai PL, Comparing breast cancer risk assessment models, J Natl Cancer Inst 2010; 102:665-668.

27. Scutt D, Lancaster GA, Manning JT, Breast asymmetry and predisposition to breast cancer, Breast Cancer Research 2006; 8:R14 (doi:10.1186/bcr1388).

28. Zheng, B., Sumkin, J. H., Zuley, M. L., Wang, X., Klym, A. H., & Gur, D. (2012). Bilateral mammographic density asymmetry and breast cancer risk: A preliminary assessment. European Journal of Radiology, 81(11), 3222-3228.

29. van Engeland, Saskia, et al. "A comparison of methods for mammogram registration." Medical Imaging, IEEE Transactions on 22.11 (2003): 1436-1444.

30. Qian, W., Sankar, R., Song, X., Sun, X., & Clark, R. (2005). Standardization for image characteristics in telemammography using genetic and nonlinear algorithms. Computers in biology and medicine, 35(3), 183-196.

31. Qian W, Clarke LP, Kallergi M, Clark RA, Tree-structured nonlinear filters in digital mammography, IEEE Trans Med Imaging 1994; 13(1):25–36

32. Qian W, Clarke LP, Li H, et al Digital mammography: m-channel quadrature mirror filters for microcalcification extraction, Comput Med Imaging Graphics 1994; 18:301–314.

33. Qian W, Sun X, Song D, Clark RA, Digital mammography: wavelet transform and Kalman-filtering neural network in mass segmentation and detection, Acad Radiol. 2001; 8:1074-1082.

34. Qian, W., Li, L., Clarke, L., Clark, R. A., & Thomas, J. (1999). Digital mammography: comparison of adaptive and nonadaptive CAD methods for mass detection. Academic radiology, 6(8), 471-480.

35. Chuang K, Tzeng H, Chen S, et al, Fuzzy c-means clustering with spatial information for image segmentation, Computerized Medical Imaging and Graphics 2006;30:9–15.

36. Grigorescu SE, Petkov N, Kruizinga P, Comparison of texture features based on Gabor filters, IEEE Trans. Image Processing 2002; 11:1160-67.

37. Kruizinga P, Petkov N, Non-linear operator for oriented texture, IEEE Trans. Image Processing 1999; 8:1395-1407.

38. Ross TJ, Fuzzy Logic with Engineering Applications, third edition, Page 198.

39. M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert Syst. With Applicat., vol. 36, no. 2, pt. 2, pp. 3240–3247, Mar. 2009, 10.1016/j.eswa.2008.01.009, 0957-4174.

40. Chang C, Lin C, LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

41. Nishikawa RM. Current status and future directions of computer-aided diagnosis in mammography. Comput Med Imaging Graph. 2007;31:224-35.

42. Hand W, Semmlow JL, Ackerman LV, Alcorn FS, Computer screening of xeromammograms: A technique for defining suspicious areas of the breast, Comput. Biomed. Res. 1979; 12:445-460.

43. Yin FF, Giger ML, Doi K, et al, Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images, Med Phys 1991; 18:955-963.

44. Zheng B, Chang YH, Gur D, Computerized detection of masses from digitized mammograms: Comparison of single-image segmentation and bilateral-image subtraction, Acad Radiol 1995; 2:1056-1061.

45. Gur D, Stalder J, Hardesty LA, et al, CAD performance on sequentially ascertained mammographic examinations of masses: an assessment. Radiology 2004; 233:418-423.

46. Wang X, Lederman D, Tan J, et al, Computerized prediction of risk for developing breast cancer based on bilateral mammographic breast tissue asymmetry, Med Eng Phys 2011; 33:934-942.