Published in final edited form as:

Eur J Radiol. 2012 November ; 81(11): 3222–3228. doi:10.1016/j.ejrad.2012.04.018.

Bilateral Mammographic Density Asymmetry and Breast Cancer Risk: A Preliminary Assessment

Bin Zheng, PhD,

Department of Radiology, University of Pittsburgh, 3362 Fifth Ave, Pittsburgh, PA 15213, USA,
Tel: 1-412-641-2568, Fax: 1-412-641-2582, zhengb@upmc.edu

Jules H. Sumkin, MD,

Department of Radiology, Magee Womens Hospital, 300 Halket Street, Pittsburgh, PA 15213, USA, Tel: 1-412-641-1976, Fax: 1-412-641-2582, sumkinjh@mail.magee.edu

Margarita L. Zuley, MD,

Department of Radiology, Magee Womens Hospital, 300 Halket Street, Pittsburgh, PA 15213, USA, Tel: 1-412-641-1635, Fax: 1-412-641-2582, zuleyml@upmc.edu

Xingwei Wang, PhD,

Department of Radiology, University of Pittsburgh, 3362 Fifth Ave, Pittsburgh, PA 15213, USA, Tel: 1-412-641-2567, Fax: 1-412-641-2582, wangx6@upmc.edu

Amy H. Klym, M.S., and

Department of Radiology, University of Pittsburgh, 3362 Fifth Ave, Pittsburgh, PA 15213, USA, Tel: 1-412-641-2566, Fax: 1-412-641-2582, klymah@upmc.edu

David Gur, ScD

Department of Radiology, University of Pittsburgh, 3362 Fifth Ave, Pittsburgh, PA 15213, USA, Tel: 1-412-641-2513, Fax: 1-412-641-2582, gurd@upmc.edu

Abstract

To improve efficacy of breast cancer screening and prevention programs, it requires a risk assessment model with high discriminatory power. This study aimed to assess classification performance of using computed bilateral mammographic density asymmetry to predict risk of individual women developing breast cancer in near-term. The database includes 451 cases with multiple screening mammography examinations. The first (baseline) examinations of all case were interpreted negative. In the next sequential examinations, 187 cases developed cancer or surgically excised high-risk lesions, 155 remained negative (not-recalled), and 109 were recalled benign cases. From each of two bilateral cranio-caudal view images acquired from the baseline examination, we computed two features of average pixel value and local pixel value fluctuation. We then computed mean and difference of each feature computed from two images. When applying the computed features and other two risk factors (woman's age and subjectively rated mammographic density) to predict risk of cancer development, areas under receiver operating characteristic curves (AUC) were computed to evaluate the discriminatory/classification performance. The AUCs are 0.633 ± 0.030 , 0.535 ± 0.036 , 0.567 ± 0.031 , and 0.719 ± 0.027 when

© 2012 Elsevier Ireland Ltd. All rights reserved.

Correspondence to: Bin Zheng.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Financial Disclosures: There are no financial disclosures from any authors.

using woman's age, subjectively rated, computed mean and asymmetry of mammographic density, to classify between two groups of cancer-verified and negative cases, respectively. When using an equal-weighted fusion method to combine woman's age and computed density asymmetry, AUC increased to 0.761 ± 0.025 ($p < 0.05$). The study demonstrated that bilateral mammographic density asymmetry could be a significantly stronger risk factor associated to the risk of women developing breast cancer in near-term than woman's age and assessed mean mammographic density.

Keywords

Breast cancer risk assessment; Full-field digital mammography; Mammographic density; Risk model

INTRODUCTION

Breast cancer is one of the leading cancers in women over age of 40 [1] and it has been repeatedly shown that earlier detection and treatment reduced patients' mortality and morbidity rates [2]. However, interpreting mammograms is a difficult and time-consuming task due to large variability of depicted breast abnormalities, overlapping dense fibroglandular tissue, and low cancer prevalence in screening environment. As a result, detection sensitivity and specificity of mammography are not optimal [3, 4] and the efficacy of annual mammography screening remains a quite controversial issue [5–7]. Hence, to improve efficacy of breast cancer screening and prevention programs, developing personalized mammography screening paradigms has attracted research interest in which the breast cancer screening intervals and/or screening methods vary for differently individual women [8]. For this purpose, it is important to develop breast cancer risk assessment models that have high discriminatory power (positive predictive value) to predict risk level of an individual women developing breast cancer within a specified period (or near-term) following one or multiple negative screening examinations and then use this information to eventually design and implement optimally personalized screening regimes with adaptively adjusted screening intervals based on the near-term risk level assessment and the change of risk assessment over time. Studies have showed that with the exception of woman's age and carrying specific gene mutations that apply to a very small fraction of the population, breast density measured from mammograms is the strongest and likely heritable breast cancer risk indicator [9]. Women with high mammographic density has a, between four and six times, higher risk of developing breast cancer than age matched women with low breast density in their lifetime [10, 11]. Mammographic density is also associated with genetic risk factors [12] and family history of breast cancer [13]. Unfortunately, subjectively rating mammographic density using the Breast Imaging Reporting and Data System (BIRADS) by radiologists is often unreliable due to the considerable inter- and intra-observer variability [14]. Thus, several computerized schemes have been developed to automatically detect and segment mammographic tissue density [15, 16]. The high correlation levels were achieved between averaged visual and automated mammographic density assessments [17].

Despite significant research efforts to predict the likelihood or risk of women developing breast cancer in the lifetime, the majority of breast cancers are actually detected in women without any known risk factors (i.e., >60% [18]). In addition, despite strong population-based significant associations, the existing breast cancer risk models [9] have quite low discriminatory power at the individual level [19] including using mammographic tissue density to predict breast cancer incidence in women carrying BRCA1/BRCA2 gene mutations [20]. Another potential limitation of the existing risk assessment models is the use of "global" (or general) measures (e.g. age) and focus solely on predicting either long-term or lifetime risk. In contrast, for planning breast cancer screening regimes, it is the accurate

near-term risk estimation (e.g., from 1 to 5 years after one or multiple negative examinations) that possibly could be practically utilized in determining optimally personalized screening regimes that aim to help clinicians decide who need (or do not need) more frequent screening or follow-up for optimizing detection of breast cancer at an earlier stage and at the same time reducing false-positive rates as well as overall cost and anxiety associated with cancer screening programs.

In an attempt to improve classification performance when assessing the individual likelihood of mammographic detection of breast cancer in the near-term following a negative screening examination, we investigated a new image-based breast cancer risk factor namely the bilateral mammographic density asymmetry computed from the entire breast areas (pixels) depicted two mammograms of the left and right breasts. Selection of this new risk factor is based on several underlying scientific evidences and validated experimental observations including that (1) humans naturally show bilateral symmetry in paired morphological traits including breasts [21], (2) mammographic tissue asymmetry is an important radiographic image phenotype related to the biological processes [22], (3) radiologists routinely examine the bilateral mammographic tissue asymmetry pattern and change over time when making clinical decisions [23], (4) computerized schemes are able to achieve more consistent results in assessing mammographic tissue density by eliminating inter-observer variability [24], and (5) computerized image features are more reliable than other risk factors reported by patients themselves in the clinic [25]. Thus, this study aims to investigate and test the feasibility of using the computed bilateral mammographic density asymmetry levels depicted on the negative examinations to predict the risk of women having or developing breast abnormalities or cancer that would be likely detected and verified during the first subsequent screening examination to a negative one in question. Prediction results were also compared with those yielded using several other known risk factors including woman's age and subjectively or computerized assessed mammographic density based on single images.

MATERIALS AND METHODS

From a pre-assembled image database involving the full-field digital mammography (FFDM) examinations conducted during 2006 to 2008 in our medical center [26], we selected all 451 available cases that had images of two (371) or three (80) sequential FFDM examinations during this time period. These cases constituted the dataset used for testing the hypothesis in this study. Except for the last FFDM examination termed as "current" examination, all "prior" examinations in the dataset had been interpreted as "negative" during the originally clinical interpretation and had not been recalled (screening BIRADS 1 or 2). Based on verified status changes (if any) during the "current" examination that took place between six (for interval cancer cases) and up to 18 months post the baseline negative examination in question, we divided cases into three groups. The first group included 187 women who developed pathology-verified cancer or pre-cancer high-risk lesions that were surgically excised adhering to current clinical practice guidelines. Among these, 136 had cancer detected during a periodic mammography examination, 38 had interval cancers, and 13 depicted surgically excised high-risk lesions. The second group included 109 women who were recalled due to the suspicious findings on the screening mammogram, but diagnostic workup of additional imaging procedures and/or a biopsy resulted in benign findings. The third group included 155 women who remained negative (not-recalled). Two and half years after closing the dataset, the recent follow-up status verification found that one invasive lobular carcinoma (ILC) and one ductal carcinoma in situ (DCIS) were subsequently detected one and two years post the "current" negative examination and all other negative cases remained negative, to date.

Bilateral cranio-caudal images from the selected baseline negative FFDM examination of each case were used in the study. A computer-aided detection scheme of FFDM images [26] was applied to process each selected image. The two image features were then computed

from the entire segmented breast area. These were 1) the mean pixel value, $F_1 = \frac{1}{N} \sum_{i=1}^N I_i$, where N is the number of pixels inside the entire segmented breast area depicted on one image and I_i is the digital value (intensity) of selected pixel (i), and 2) the mean pixel value fluctuation. To compute this feature, a convolution kernel (5×5 pixels) was applied to scan the entire breast area. Maximum pixel value difference between the center pixel (i) and all other pixels (k) inside the kernel was computed. After averaging computed values for all

scanned pixels (N), the feature was computed as: $F_2 = \frac{1}{N} \sum_{i=1}^N |Max(I_i - I_k)|$. These features were easy to compute and have been shown to represent averaged mammographic tissue density and global tissue pattern fluctuations [17, 27].

For each case we computed the mean ($\bar{F}_1$ and $\bar{F}_2$) and absolute difference (ΔF_1 and ΔF_2) of the feature values computed from the cranio-caudal view of two mammograms of left and right breasts. In each group of mean and absolute difference feature values, we defined an additional feature that represented the maximum of two computed values following range normalization $\bar{F}_3 = \text{Max}(\bar{F}_1, \bar{F}_2)$ and $\Delta F_3 = \text{Max}(\Delta F_1, \Delta F_2)$. For this purpose, the computed features' values (i.e., $\bar{F}_1$ and $\bar{F}_2$ or ΔF_1 and ΔF_2) were normalized using a monotonic (order preserving) transformation method to preprocess each feature by sorting all 451 values and dividing them into 10 exclusive bins with 45 cases in the first 9 bins and 46 in the last one. Within each bin feature values (F_i) of all cases ($i = 1, 2, \dots, 45$ or 46) were linearly transformed into: $F'_i = 0.1 \times k + 0.1 \times (F_i - F_{MIN}) / (F_{MAX} - F_{MIN})$, where F_{MAX} and F_{MIN} represent the maximum and minimum feature values within a specific bin (k), respectively. As a result, all feature values were transformed to 0 to 1 range with their rank order preserved. We then compared the computed values between paired features (e.g., ΔF_1 and ΔF_2) computed from the same examination and determined the value of third feature (e.g., ΔF_3).

Using the computed mammographic density related features together with two well-known risk factors, namely woman's age and subjectively rated mammographic density (BIRADS), we assessed the feasibility of classifying cases into "positive" and "negative" for having detected breast abnormalities (or cancer) during the first subsequent examination to a negative one. The classification performance was assessed using the areas under the receiver operating characteristic (ROC) curves (AUC). Specifically, the normalized values of each feature computed from all cases in two compared case groups was processed by a maximum likelihood statistical data analysis based ROC fitting program (ROCKIT, <http://www-radiology.uchicago.edu/krl/>) and the results (i.e., AUC and its 95% confidence interval) were recorded and compared. After comparing classification performance using a single feature independently, we also tested a simple equal-weighted fusion method to combine woman's age after a linear rank order preserving range transformation and the computed bilateral mammographic density asymmetry to improve classification performance as compared with a single feature. The initial assessment was conducted using the 187 "positive" and 155 not-recalled "negative" cases. The results were then tested again by adding 109 recalled benign cases to either "positive" or "negative" groups.

RESULTS

Table I summarizes the distributions of woman's age and subjectively rated breast density (BIRADS) for the women in each of the three case groups. The average ages and the standard deviations are 60.5 ± 12.8 , 54.9 ± 9.9 , and 52.3 ± 9.3 for the cancer, not-recalled negative, and recalled benign case groups in this dataset, respectively. The distributions of density BIRADS among three case groups are quite similar. Figure 1 shows three example cases including two sets of cranio-caudal view images acquired from both the baseline and the next sequential screening examinations of three women with age of 44, 62, and 59 years old (from the top to bottom rows). The baseline images were all originally interpreted as negative and the cancers were detected in the next sequential images, which were associated with a spiculated, an irregular, and a poorly defined mass revealed invasive duct carcinoma (IDC) with nuclear grade 2-3 in these three cases, respectively. Figure 2 shows three box-plots that includes the median density asymmetry (ΔF_3) and inter-quartile range. The three median ΔF_3 score for the negative (not-recalled), recalled (benign) and positive groups were 0.34, 0.45, and 0.50, respectively.

Table II summarizes the overall single feature based classification performance levels (AUCs) when using woman's age, subjectively rated breast density, each of the six computed image features ($\bar{F}_1$, $\bar{F}_2$, $\bar{F}_3$, ΔF_1 , ΔF_2 and ΔF_3), and a fusion feature between age and computed density asymmetry (ΔF_3) in classifying between two groups of 187 positive and 155 negative cases. The results indicate that woman's age was a significantly stronger risk predictor (AUC = 0.633 ± 0.030) than the subjective mammographic density BIRADS rating (AUC = 0.535 ± 0.036 , $p = 0.024$) and the computed mean mammographic density ($\bar{F}_3$) with AUC = 0.567 ± 0.031 ($p = 0.043$). The computed bilateral mammographic density asymmetry (ΔF_3) was the strongest single feature based risk predictor (AUC = 0.719 ± 0.027) with a significantly higher prediction power than woman's age ($p = 0.022$) and the mammographic density assessed by either radiologists (BIRADS) or a computer scheme ($p < 0.001$). An equal-weighted fusion feature (combining woman's age and ΔF_3) also yielded a significantly higher classification performance with the AUC increasing from 0.719 ± 0.027 for using ΔF_3 alone to 0.761 ± 0.025 for the fused predictor ($p = 0.011$). Figure 3 shows and compares four fitted ROC-type performance curves.

Additional data analyses also showed consistency of the results. First, the exclusion of 13 high-risk (pre-cancer) cases from the positive group did not materially change the discrimination results (i.e., the classification performance increased minimally to AUC = 0.720 ± 0.028 when using the bilateral mammographic density asymmetry (ΔF_3). Second, the exclusion of two "negative" cases with cancer detected one and two years post the "current" negative examination had little impact on the classification performance (i.e., AUC = 0.720 ± 0.027 using ΔF_3). Third, the inclusion of the recalled group of 109 benign cases in the classification process resulted in a decrease in the discriminative ability. For example, when using ΔF_3 , AUCs decreased to 0.690 with 95%CI [0.638, 0.739] and 0.662 with 95%CI [0.611, 0.712] for inclusion of these recalled cases into either the positive or negative group, respectively, demonstrating that this group is intermediate in terms of overall density asymmetry, which may have led to a recall. When using woman's age, AUCs changed to 0.561 with 95%CI [0.506, 0.614] and 0.656 with 95%CI [0.603, 0.706] for inclusion of the recalled cases into either the positive or negative group, respectively, indicating that younger women were more likely to be recalled in this dataset. However, using ΔF_3 yielded a significantly higher ($p < 0.01$) classification performance than woman's age when the recalled group was included in the "positive" group.

At a 90% specificity level, applying the computed bilateral mammographic density asymmetry scores (ΔF_3) was able to detect (or predict) 20.2% (22/109) of recalled benign

cases and 34.2% (65/187) cancer or high-risk pre-cancer cases to be “positive” for the cancer being detected in the next sequential examinations. Among the 65 detected positive cases, 2 (28.6%), 17 (29.8%), 43 (36.4%), and 3 (60.0%) were rated with density of BIRADS categories 1 to 4, respectively. At an 80% specificity level, using ΔF_3 detected 28.4% (31/109) recalled benign cases and 43.9% (82/187) positive cases. By further dividing the positive cases into three subgroups of high risk lesions, cancer, and interval cancer cases, the detection sensitivity levels were monotonically increase as the time-lag reduced between the negative and positive examinations that lead to cancer detection (Figure 4). The highest sensitivity was achieved in interval cancer group in which cases typically have 6 to 9 months of time-lag between the negative and positive (cancer-detected) examinations (e.g., 47.4% sensitivity at 90% specificity), while the lowest sensitivity was yielded in recalled benign case group with much longer time-lag of cancer-free (e.g., 20.2% sensitivity at 90% specificity).

Table III summarizes the fraction of screening detected cancer cases for whom diagnosis would have been delayed by one screening period (e.g., one year) when applying the risk stratification results generated using ΔF_3 and the fusion score between woman’s age and ΔF_3 to recommend women with “low-risk” prediction scores to possibly skip one screening examination (e.g., change the screening interval from one year to two years). Although at high risk assessment specificity levels (e.g., >50%) the diagnosis of a large fraction of “positive” cases would have been delayed, if the recommendation was made for all cases in the lower quintile (i.e. lower 20% in terms of ΔF_3 scores) the diagnosis of less than 3% (5/187) of the screening detected cancer cases would have been delayed. For example, in our dataset two of 155 cases with “current” negative examinations were subsequently detected one and two years later as having cancer. In the first case with ILC detected one year later, the density asymmetry scores (ΔF_3) increased by 34% (from 0.41 to 0.55) in the two available sequential negative examinations. In the second case with DCIS detected two years later, ΔF_3 scores fluctuated in three sequential negative examinations and were 0.33, 0.25 and 0.31, respectively. Setting negative prediction threshold scores at 3 $\Delta F_3 > 0.31$ or $\Delta F_3 > 0.55$, would result in delaying cancer detection in one or both cases at 46% or 67% specificity levels, respectively.

DISCUSSION

In this preliminary assessment study we investigated and demonstrated that using a new image feature of the computed bilateral mammographic density asymmetry, one could significantly increase positive predictive value of classifying two groups of high and low risk cases in developing high-risk breast abnormalities and/or cancer in the next sequential screening examination than using two currently well-known risk factors (namely woman’s age and mammographic density subjectively rated or computed based on single images). This study has two unique characteristics that make it different from the previous conventional breast cancer risk assessment studies in cancer epidemiology fields. First, although a number of epidemiology-based risk models have shown highly significantly statistical associations with the risk of cancer development in lifetime of individual women as compared to general population [9], no one has clinically acceptable positive predictive value (discriminatory power) [19]. Hence, without a risk prediction model that is able to reliably predict the risk of an individual woman developing breast cancer, in particular, in the relatively near-term (i.e., 1 to 5 years after a negative screening examination), current universally adopted breast cancer prevention and screening programs cannot be optimized. This study focused on developing a new concept and approach to significantly increase positive predictive value in classifying between two groups of high and low risk cases in developing breast abnormalities or cancer in near-term. Thus, the near-term risk assessment

results yielded in this study is not directly comparable with the lifetime risk assessment results yielded by using the conventional epidemiology based risk models.

Second, although mammographic density has been widely recognized as one of strongest known risk factors and used in many of previous risk assessment studies or epidemiology-based models to date [9], previous studies have generated controversial or inconsistent results. Some studies concluded that mammographic density was an strong and/or important breast cancer risk factor [28–30], while others showed that high mammographic density was only a weak and quite confused risk factor [20, 31, 32] because these authors believed that assessing mammographic breast density could be quite biased. Breast density is not only influenced by many woman's life style factors (i.e., using hormonal agents, smoking and drinking alcohol) [33], but also difficult to be accurately measured from mammograms without knowing the exact X-ray exposure value and breast compression ratio [34]. As a result, mammographic density alone has quite low discriminatory power (positive predictive value) in predicting the risk of individual women for developing breast cancer [19]. In this study, we found that (1) the distributions of the subjectively rated mammographic density (BIRADS) across three groups of cases in our test dataset are quite similar (Table I) and (2) the average mammographic density level computed from two single CC view images of the same women has limited discriminatory power (i.e., the maximum classification performance $AUC < 0.567 \pm 0.031$) that is substantially lower than the classification performance levels yielded using either woman's age or the computed bilateral mammographic density asymmetry. This is probably due to the fact that the cancer or high risk lesions can be detected on both dense and non-dense breasts (in all BIRADS categories) in the screening environment. However, the computed bilateral mammographic density asymmetry is largely independent from the actual mammographic density levels (e.g., the correlation coefficients between the computed mean and asymmetry of mammographic density are -0.13 and -0.07 for the positive and negative case groups in our dataset, respectively). The cases with lower mammographic density level but higher bilateral asymmetry have higher risk assessment scores of developing breast cancer in near-term than the cases with both higher mammographic density and bilateral symmetry. In addition, since many potential biases in assessing mammographic density [33, 34] should equally affect on features computed from two bilateral mammograms, subtraction of two features computed from two matched images help reduce or minimize such biases. Thus, this new risk factor focuses on the comparison of density asymmetry between the left and right breasts of one woman should be more robust. Since the two image features namely the mean and difference of mammographic density computed from two bilateral mammograms are not correlated, the bilateral mammographic density asymmetry could be a new and valuable risk-assessment feature that has never been tested and applied in any of existing breast cancer risk models to date. This study demonstrated the feasibility of using this new feature to yield significantly higher positive predictive value in assessing near-term breast cancer risk of individual women than using the conventional approaches to subjectively assess or compute mammographic density from the single images.

In addition, since woman's age is the strongest cancer risk (e.g., with a relative risk of 30 at extremes [9]), we combined age and the computed bilateral mammographic density asymmetry in our risk assessment to predict near-term risk and we found that this combination significantly increase the discriminatory power in classifying between negative (not-recalled) and positive (cancer) cases. However, adding age as a risk factor did not increase performance in classifying the recalled benign cases as "high risk" cases in our test dataset probably because the younger women with dense breasts are more likely to be recalled. The results of this study indicated the potential of that combining this new risk factor/feature with the other known risk factors might help to build a new risk assessment

predictor/model with significantly higher discriminatory power than the existing risk models.

This is a very preliminary study that has a number of limitations. First, our original FFDM database [26] was established to develop a computer-aided detection scheme of FFDM images and assess its performance and reliability based on the concept of a nested case-control type study. Due to the short tracking period (in less than two and half years from middle 2006 to the end of 2008), only a limited number of 451 cases has series of sequential FFDM examinations in the database including a negative baseline FFDM examination. Hence, the testing dataset used in this study is small and could not adequately cover the diversity of cases in screening environment. Second, a large number of computed image features have been previously tested to assess mammographic tissue density and patterns [27] and studies have also shown that using more sophisticated statistical texture analysis methods applied to the manually selected regions of interesting (ROIs) could more accurately classify between high and low risk groups of cases [17, 24]. We selected two simple image features that are computed from the entire segmented breast area depicted on an image. Although these two features are easy to compute and much robust or highly reproducible as compared to the use of the texture-based features computed from the manually or automatically selected ROIs, we recognized that these two features might not be the optimal features for this task. Third, unlike the majority of existing risk models [9] that estimate primarily a fixed long-term or lifetime risk using many risk factors that rarely change over time (i.e., family history of breast cancer), the bilateral mammographic density asymmetric levels could change over time due to the development of breast abnormalities that may lead to cancer detection. Whether and how change of bilateral mammographic density asymmetry over time (on series of multiple negative examinations) could impact the risk assessment performance was not investigated in this study.

In summary, this study demonstrated that the bilateral mammographic density asymmetry was a significantly stronger risk factor (feature) than mammographic density rated or assessed based on single images by either radiologists (BIRADS) or computerized schemes in predicting the likelihood of individual women developing high-risk breast abnormalities and/or cancer in near-term. As a result, developing an optimal, robust, and easy to implement computerized scheme to quantitatively measure or assess bilateral mammographic density asymmetry may have significantly clinical impact. The risk assessment results may provide clinicians and women who participated in screening mammography a useful tool to establish the optimally and adaptively adjustable personalized breast cancer prevention and screening programs. To realize this ultimate goal, much work is needed in this field including building large and diverse image databases with multiply sequential negative FFDM examinations, exploring and selecting optimal image features to more effectively detect bilateral mammographic density asymmetry, and evaluating the performance and reliability of this or similar computerized risk model in the cohort (or prospective) studies.

Acknowledgments

This work is supported in part by Grants CA CA77850 and CA160205 from the National Cancer Institute, National Institutes of Health, USA.

REFERENCES

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer*. 2010; 60:277–300.
2. Smith RA, Cokkindes V, Brooks D, et al. Cancer screening in the United States, 2011. *CA Cancer J Clin*. 2011; 61:8–30. [PubMed: 21205832]

3. Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology*. 2002; 224:861–869. [PubMed: 12202726]
4. Fenton JJ, Wheeler J, Carney PA, et al. Reality check: perceived versus actual performance of community mammographers. *Am J Roentgenol*. 2006; 187:42–46.
5. US Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2009; 151:716–726. [PubMed: 19920272]
6. Berlin L, Hall FM. More mammography muddle: emotions, politics, science, costs and polarization. *Radiology*. 2010; 255:311–316. [PubMed: 20413746]
7. Buist DS, Anderson ML, Haneuse SJ, et al. Influence of annual interpretive volume on screening mammography performance in the United States. *Radiology*. 2011; 259:72–84. [PubMed: 21343539]
8. Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: Analysis of health benefits and cost-effectiveness. *Ann Intern Med*. 2011; 155:10–21. [PubMed: 21727289]
9. 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. [PubMed: 20427433]
10. Saftlas AF, Hoover RN, Brinton LA, et al. Mammographic densities and risk of breast cancer. *Cancer*. 1991; 67:2833–2838. [PubMed: 2025849]
11. Boyd NF, Martin IJ, Stone J, et al. Mammographic densities as a marker of human breast cancer risk and their use in chemoprevention. *Curr Oncol Rep*. 2001; 3:314–321. [PubMed: 11389815]
12. 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 Res*. 2007; 67:8412–8418. [PubMed: 17804758]
13. Martin LJ, Melnichouk O, Guo H, et al. Family history, mammographic density, and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2010; 19:456–463. [PubMed: 20142244]
14. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast imaging reporting and data system: Inter- and intra-observer variability in feature analysis and final assessment. *Am J Roentgenol*. 2000; 174:1769–1777. [PubMed: 10845521]
15. Highnam R, Pan X, Warren R, et al. Breast composition measurements using retrospective standard mammogram form (SMF). *Phys Med Biol*. 2006; 51:2696–2713.
16. Glide-Hurst CK, Duric N, Littrup P. A new method for quantitative analysis of mammographic density. *Med Phys*. 2007; 34:4491–4498. [PubMed: 18072514]
17. Chang YH, Wang XH, Hardesty LA, et al. Computerized assessment of tissue composition on digitized mammograms. *Acad Radiol*. 2002; 9:898–905.
18. Madigan MP, Ziegler RG, Benichou J, et al. Proportion of breast cancer cases in the United States explained by well-established risk factors. *JNCI*. 1995; 87:1681–1685. [PubMed: 7473816]
19. Gail MH, Mai PL. Comparing breast cancer risk assessment models. *J Natl Cancer Inst*. 2010; 102:665–668. [PubMed: 20427429]
20. Passaperuma K, Warner E, Hill KA, Gunasekara A, Yaffe MJ. Is mammographic breast density a breast cancer risk factor in women with BRCA mutations. *J Clin Oncol*. 2010; 28:3779–3783. [PubMed: 20625126]
21. Scutt D, Lancaster GA, Manning JT. Breast asymmetry and predisposition to breast cancer. *Breast Cancer Research*. 2006; 8:R14. [PubMed: 16563179]
22. Scutt D, Manning JT, Whitehouse GH, et al. The relationship between breast asymmetry, breast size and the occurrence of breast cancer. *Br J Radiol*. 1997; 70:1017–1021. [PubMed: 9404205]
23. Blanks RG, Wallis MG, Given-Wilson RM. Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography. *J Med Screen*. 1999; 6:152–158. [PubMed: 10572847]
24. Wei J, Chang HP, Wu Y, et al. Association of computerized mammographic parenchymal pattern measure with breast cancer risk: a pilot case-control study. *Radiology*. 2011; 260:42–49. [PubMed: 21406634]

25. Domchek SM, Antoniou A. Cancer risk models: translating family history into clinical management. *Ann Intern Med.* 2007; 147:515–517. [PubMed: 17909212]
26. Zheng B, Sumkin JH, Zuley M, Lederman D, Wang X, Gur D. Computer-aided detection of breast masses depicted on full-field digital mammograms: a performance assessment. *Br J Radiol.* 2011 (Article in press published online on 21 February 2011).
27. Wang X, Lederman D, Tan J, Wang XH, Zheng B. Computerized detection of breast tissue asymmetry depicted on bilateral mammograms: a preliminary study of breast cancer stratification. *Acad Radiol.* 2010; 17:1234–1241. [PubMed: 20619697]
28. Giatto S, Zappa M. A prospective study of the value of mammographic patterns as indicators of breast cancer risk in a screening experience. *Eur J Radiol.* 1993; 17:122–125. [PubMed: 8223680]
29. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Eng J Med.* 2007; 356:670–675.
30. Olsen AH, Bihrmann K, Jensen MB, Vejborg I, Lynge E. Breast density and outcome of mammography screening: a cohort study. *Br J Cancer.* 2009; 10:1052–1059.
31. van Gils CH, Otten JD, Verbeek AL, Hendriks JH. Mammographic breast density and risk of breast cancer: masking bias or causality? *Eur J Epidemiol.* 1998; 14:315–320. [PubMed: 9690746]
32. Colin C, Prince V, Alette PJ. Can mammographic assessment lead to consider density as a risk factor for breast cancer? *Eur J Radiol.* 2010
33. Harvey J, Bovbjerg VE. Quantitative assessment of mammographic breast density: relationship with breast cancer risk. *Radiology.* 2004; 230:29–41. [PubMed: 14617762]
34. Kopans DB. Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. *Radiology.* 2008; 246:348–353. [PubMed: 18227535]

Figure 1.

Example of using the bilateral mammographic density asymmetry to predict the likelihood of woman having breast cancer detectable in the next sequential mammography screening examination. The left two columns show two CC view images acquired from the baseline (negative) examinations and the right two columns show images acquired from the next screening examinations with detectable cancer (circled regions). From the top to bottom rows, the breast density of the three cases was originally rated by the radiologists as BIRADS 2, 3, and 4, and the corresponding computed bilateral density asymmetric scores are 0.83, 0.94, and 0.35, respectively.

Figure 2. Box plots of the computed bilateral mammographic density asymmetry scores (ΔF_3) of three groups of cases in our test dataset.

Figure 3.

Comparison of five ROC-type performance curves in which the dash line is a reference line ($AUC = 0.5$) and the other four curves sorted from the smallest to the largest AUC values of 0.567 ± 0.031 , 0.633 ± 0.030 , 0.718 ± 0.027 , and 0.761 ± 0.025 , represent classification performance of using the computed mammographic density from single images, woman's age, computed bilateral mammographic density asymmetry, and fusion of age and computed bilateral mammographic density asymmetry, respectively.

Figure 4. Distributions of the prediction sensitivity levels at two specificity levels (80% and 90%) when applying the computed bilateral mammographic density asymmetry scores to the four types of cases.

Table I

Absolute and percentage distribution data of women age and subjectively rated breast density (BIRADS) in three group of 451 cases selected in this study.

	Positive cases	Recalled benign cases	Negative cases
Woman's Age (Years Old)			
A 40	4 (2.1%)	10 (9.2%)	4 (2.6%)
40 < A 50	45 (24.1%)	37 (33.9%)	60 (38.7%)
50 < A 60	46 (24.6%)	38 (34.9%)	45 (29.1%)
60 < A 70	46 (24.6%)	20 (18.3%)	34 (21.9%)
70 < A 80	31 (16.6%)	4 (3.7%)	10 (6.4%)
A > 80	15 (8.0%)	0 (0.0%)	2 (1.3%)
Density BIRADS rating			
1	7 (3.7%)	6 (5.5%)	8 (5.2%)
2	57 (30.5%)	37 (33.9%)	52 (33.5%)
3	118 (63.1%)	62 (56.9%)	92 (59.4%)
4	5 (2.7%)	4 (3.7%)	3 (1.9%)

Table II

Summary of classification performance level using nine risk-related features to classify between two groups of 187 positive (cancer or high-risk) cases and 155 negative (not-recalled) cases.

	AUC	Standard Error	95% Confidence Interval
Women Age	0.633	0.030	[0.574, 0.689]
Density BIRADS rating	0.535	0.036	[0.464, 0.605]
F_1	0.526	0.031	[0.466, 0.587]
F_2	0.529	0.032	[0.467, 0.590]
$F_3 = \text{Max}(F_1, F_2)$	0.567	0.031	[0.506, 0.626]
ΔF_1	0.680	0.029	[0.622, 0.734]
ΔF_2	0.591	0.031	[0.530, 0.649]
$\Delta F_3 = \text{Max}(\Delta F_1, \Delta F_2)$	0.719	0.027	[0.663, 0.770]
Fusion of Age and ΔF_3	0.761	0.025	[0.709, 0.808]

The fraction of potentially delayed screening detected cancer cases when a given fraction of women were recommended to skip one mammography screening cycle examination based on one of the two computed risk scores yielding at different specificity levels of the risk assessment.

Table III

Specificity	80%	65%	50%	35%	20%
Density asymmetry (ΔF_3)	57.4%	45.6%	27.2%	11.8%	2.7%
Fusion of age and ΔF_3	42.2%	35.3%	24.1%	12.3%	4.3%