Published in final edited form as:

Acad Radiol. 2013 December ; 20(12): . doi:10.1016/j.acra.2013.08.020.

Prediction of Near-term Breast Cancer Risk based on Bilateral Mammographic Feature Asymmetry

Maxine Tan¹, Bin Zheng¹, Pandiyarajan Ramalingam², and David Gur²

¹ School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019

² Department of Radiology, University of Pittsburgh, Pittsburgh, PA 15213

Abstract

Rationale and Objectives—The objective of this study is to investigate the feasibility of predicting near-term risk of breast cancer development in women after a negative mammography screening examination. It is based on a statistical learning model that combines computerized image features related to bilateral mammographic tissue asymmetry and other clinical factors.

Materials and Methods—A database of negative digital mammograms acquired from 994 women was retrospectively collected. In the next sequential screening examination (12 to 36 months later), 283 women were diagnosed positive for cancer, 349 were recalled for additional diagnostic workups and later proven to be benign, and 362 remain negative (not-recalled). From an initial pool of 183 features, we applied a Sequential Forward Floating Selection feature selection method to search for effective features. Using 10 selected features, we developed and trained a support vector machine (SVM) classification model to compute a cancer risk or probability score for each case. The area under the receiver operating characteristic curve (AUC) and odds ratios (ORs) were used as the two performance assessment indices.

Results—The AUC=0.725±0.018 was obtained for positive and negative/benign case classification. The ORs showed an increasing risk trend with increasing model-generated risk scores (from 1.00 to 12.34, between positive and negative/benign case groups). Regression analysis of ORs also indicated a significant increase trend in slope ($p=0.006$).

Conclusions—This study demonstrates that the risk scores computed by a new SVM model involving bilateral mammographic feature asymmetry have potential to assist the prediction of near-term risk of women for developing breast cancer.

Keywords

Bilateral mammographic feature asymmetry; Breast cancer; Cancer risk prediction model; Computer-aided detection (CAD) of mammograms

© 2013 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.

CORRESPONDENCE: Maxine Tan, PhD School of Electrical and Computer Engineering, University of Oklahoma, 101 David L. Boren Blvd Norman, OK 73019. maxinety1@gmail.com:.

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.

I. INTRODUCTION

Breast cancer is the most prevalent cancer in women, and is also a very heterogeneous disease (1). The majority of breast cancers are detected among women with no known cancer risk factors (2, 3). Hence, a uniform, population-based mammography screening protocol is currently recommended for all women who are older than the qualifying age. Although scientific evidence has shown that early detection of breast cancer combined with improved treatment strategies has incrementally and significantly reduced patients' mortality and morbidity rates over the last four decades (4, 5), interpreting screening mammograms is quite a difficult task for radiologists due to the large variability of breast abnormalities, overlapping dense fibro-glandular tissues, and the low cancer detection rate (i.e., three to five cancers in 1000 non-baseline screening examinations) (6). These factors substantially reduce the detection sensitivity and specificity of screening mammography (7, 8). One study reported that during a 10-year period, more than half of screened women will receive at least one false-positive recall, and 7 to 9% will have at least one benign biopsy (9). In addition to anxiety in women that can cause long-term psychosocial consequences (10), and other side effects of the false-positive recalls due to cumulative radiation exposure (11) and unnecessary biopsies, limited healthcare resources and associated high costs are other major factors that have to be considered in this issue (12). Thus, the efficacy of current mammography screening remains controversial (13, 14). In order to overcome these limitations, it is desirable to develop personalized screening recommendations based on individualized risk assessment (15); this concept has recently been attracting significant research interests (16-19). The prerequisite of reaching this goal of establishing a new and more effective personalized cancer screening paradigm is to identify cancer risk factors and/or develop risk prediction models with improved discriminatory power, which aim to stratify women into different risk groups whereby different screening methods and intervals can be recommended.

Studies have shown that with the exception of women's age and specific gene mutations that apply to a very small fraction of the population, breast density is the strongest and most-likely heritable breast cancer risk indicator (3). A woman with higher breast density has a higher risk of developing breast cancer (between four to six times more) than another age-matched woman with lower breast density, in her lifetime (20, 21). As the process of directly detecting and/or measuring actual breast density is difficult, mammographic density measured by the percentage of fibro-glandular tissue segmented from mammograms is currently used to report breast density in the breast cancer screening environment. The American College of Radiology has established a Breast Imaging Reporting and Data System (BIRADS) to rate mammographic density into four categories namely 1) almost entirely fatty with percentage of fibro-glandular tissue (FGT) $\leq 25\%$, 2) scattered fibro-glandular density ($25\% < \text{FGT} \leq 50\%$), 3) heterogeneously dense ($50\% < \text{FGT} \leq 75\%$), and 4) extremely dense ($\text{FGT} > 75\%$) (22). However, subjectively-rated mammographic density using BIRADS has been found unreliable due to considerable inter- and intra-observer variability among radiologists (23). In order to achieve more reliable and consistent mammographic density assessment results, a number of research groups have developed various computerized schemes to detect and quantify mammographic density (24-34). In these previous researches, different image features and machine learning-based classifiers were investigated and compared. Studies have reported high correlation (e.g., >0.87 (28)) between averaged visual and automated assessments of mammographic density categories (i.e., between automated/semi-automated tissue segmentation and BIRADS ratings).

In spite of these research efforts, the discriminatory power of assessed mammographic density alone remains quite low (35). In addition, breast density is not only influenced by many lifestyle factors in women (i.e., the use of hormonal agents, smoking or drinking

alcohol) (36) an accurate measurement of breast density from mammograms is also difficult without knowledge of the exact X-ray exposure values, breast positioning and breast compression ratios (37). A number of studies also show that high mammographic density might be a weak and often confusing risk factor (38-40) due to the masking bias and inconsistencies in its assessment. In our previous studies, we have preliminarily investigated a different breast cancer risk assessment approach of using mammographic density, which is based on bilateral mammographic density asymmetry between the left and right breasts (41-43). In this study, we analyze a large number of image features and search for the effective ones. We then train and develop a support vector machine (SVM) classifier as a risk prediction model that combines nine computed bilateral mammographic feature differences (asymmetry) and another risk factor of woman's age, in order to predict the risk or likelihood of breast cancer development, namely, detecting cancer at the next sequential screening mammography examination (12 to 36 months after the negative examination of interest). The detailed experimental methods and results are reported as follows.

II. MATERIALS AND METHODS

Image Datasets

The full-field digital mammography (FFDM) images selected and analyzed in this study were retrospectively collected from an available clinical database at our medical center under an Institutional Review Board (IRB)-approved image data collection protocol. The case inclusion criteria are (1) all women who have at least two sequential FFDM examinations of both left and right breasts previously performed in our medical center and (2) the first FFDM examination was negative (not recalled). An IRB-certified “honest broker” in our research group blindly selected FFDM cases from the verified clinical databases. To minimize the case selection bias, the “honest broker” sequentially searched for cases starting from the date after the FFDM examinations started in our medical center in late 2006. From this date based retrospective searching method, the “honest broker” checked the status of the second FFDM examination that was performed 12 to 36 months after the first FFDM examination and then assigned the case into one of three subgroups (malignant, recalled benign, and not-recalled negative).

From this ongoing case selection process, we assembled an initial testing dataset that includes screening FFDM examinations (cases) acquired from 994 women for this study. Each case includes FFDM images acquired with at least two routine, sequential screening examinations. Based on the diagnostic result of the latest available examination (defined as the “current” examination) of each case, we divided this dataset into three subgroups. The first subgroup includes 283 positive cases whereby either cancers were detected in the FFDM images and later confirmed in pathology examinations (253 cases), or high-risk pre-cancer (tumors) were surgically removed (30 cases). The second subgroup includes 349 cases in which suspicious abnormalities were detected on the FFDM images and the women were recalled, but were later proved to be benign during the following diagnostic workup (imaging and/or biopsy). The third subgroup includes 362 screening negative cases that were not recalled by the radiologists.

The most recent “prior” examination to the latest “current” examinations of all 994 cases was originally interpreted as “negative” by radiologists. Specifically, all these acquired “prior” FFDM screening examinations were interpreted as “negative” (screening BIRADS 1) or “definitely benign” (screening BIRADS 2), and had not been recalled for additional workup. These “prior” examinations could have occurred 12 to 36 months before the “current” examinations. Figure 1 shows an example of two sets of bilateral cranio-caudal (CC) view FFDM images of the left and right breasts, acquired from the “prior” and “current” mammography screening examinations of a woman. In this case, the “prior”

FFDM images were interpreted as negative, and a breast tumor was detected in one of the “current” images, and was later confirmed as malignant by biopsy and pathology examinations. In this study, we used all negative FFDM images acquired from the most recent “prior” screening examinations to compute image features, and train a SVM classifier model for predicting the risk or likelihood of a woman having an image-detectable malignant tumor by radiologists during the next sequential (“current”) mammography screening examination.

In addition to collecting FFDM images, a number of clinical risk factors (namely, woman's age, family history of breast cancer and subjective breast density rating using BIRADS) were also collected for each case. Figure 2 shows the distribution of the cases into four categories of mammographic density BIRADS ratings, which shows that the distribution differences in three case subgroups are small. The average and standard deviation of the subjective density BIRADS ratings are 2.60 ± 0.64 , 2.72 ± 0.57 , and 2.67 ± 0.61 for these three subgroups of not-recalled negative, recalled benign, and positive for cancer cases, respectively. In this limited dataset, the subgroup of recalled benign cases has highest average BIRADS density rating than other two subgroups, which may indicate that radiologists may be more likely to recall the women with dense breasts.

Computation and Measurement of Image Features to Measure Breast Tissue Asymmetry

For each case, a pair of bilateral CC view FFDM images of left and right breasts was selected and analyzed. We applied a computerized scheme to segment the breast regions depicted on each image (44), and then compute image features from the segmented breast regions. Five groups or types of image features were computed and initially selected (as listed in Table 1). The first group includes pixel value (intensity) based statistical features (i.e., mean, standard deviation, skewness, and kurtosis values), and a fractal dimension feature originally proposed in (28), computed on the whole and segmented dense breast regions (28). The dense breast region was defined in (28) as the region of the breast that encompasses the pixel values above the median value of the whole breast region (i.e. 50% of the whole breast region). We used the same method as reported previously (42) to compute these statistical features of image pixel values.

The second feature group includes the gray level run length-based texture features. These consist of short run emphasis, long run emphasis, low gray level run emphasis, high gray level run emphasis, short run low gray level emphasis, short run high gray level emphasis, long run low gray level emphasis, long run high gray level emphasis, gray level non-uniformity, run length nonuniformity, and run percentage (45). To compute these run length texture features, we first reduced the gray level range of an original FFDM image from 4096 to 256 gray levels, at a sub-sampling ratio of 16. A computer scheme computed two run length matrices along 0 and 90 degrees (in the horizontal and vertical directions). The average and maximum values of the two run length features computed along the two directions were then computed to obtain the final feature values using the previously determined computation equations in (45).

The third group of features includes the x -axis and y -axis histogram (cumulative projection) related features (46) of mean, standard deviation, skewness, kurtosis, and median. These features are replicated from an asymmetry detection method proposed by Tzikopoulos et al. in (46). In brief, in order to find the x -axis cumulative projection in the form of a histogram, we computed the number or sum of ‘on’ pixels in every row of the image. Similarly, we obtained the cumulative projection of the y -axis histogram by calculating the sum of ‘on’ pixels in every column of the image. Then, we computed the first-order statistics for each of these histograms, namely the mean, standard deviation, skewness, kurtosis, and median. The

feature computation procedure was repeated for both the whole and dense breast regions, for each case.

The fourth group of features includes the gray level co-occurrence matrix based features (i.e., contrast, correlation, energy, and homogeneity). A standard computation method was used to compute these features using both the whole and dense breast regions for each case.

Finally, the fifth group of features includes three popular risk factors used in many existing epidemiology-based breast cancer risk models. These features are woman's age, family history of breast cancer and the subjectively-rated mammographic density (using BIRADS) by the radiologists.

For each of the mammographic image features listed in Table 1 (groups one to four), we separately computed two features from the CC view of bilateral FFDM images acquired from the left and right breasts of the same woman in one screening examination. To compute the bilateral mammographic image feature differences, we used the methods proposed in (46). Suppose that the feature vector for the left breast is given by f , and for the corresponding right breast, the feature vector is g . We then used the following three equations to compute the image feature differences of two bilateral mammograms between the left and right breasts to assess mammographic tissue density asymmetry:

$$F_{1-60}^{ASYMM} = \frac{|f_i - g_i|}{\max(f_i, g_i)} \quad (1)$$

$$F_{61-120}^{ASYMM} = |f_i - g_i| \quad (2)$$

$$F_{121-180}^{ASYMM} = |f_i - g_i|^3 \quad (3)$$

Using these three equations, we computed and generated a total of 180 image features that are related to the bilateral mammographic tissue density difference (asymmetry). By adding three clinical risk factors (woman's age, family history, and subjective breast density BIRADS rate), we built an initial feature pool that contains 183 features. Due to the *curse of dimensionality* (47), we applied a feature selection method to eliminate the redundant features and also minimize the risk of 'over-fitting' of the classifier during the training procedure or task. In our experiments, we analyzed and compared several feature selection methods including the best first search (48) and greedy stepwise feature selection (49) methods. Based on our experimental and comparison results, we decided to use the Sequential Forward Floating Selection (SFFS) method proposed by Pudil et al. (50). An evaluation criterion based on achieving a small within-class distance and a large between-class distance (51, 52) was also used to guide the selection of an optimal subset of features using the SFFS method.

Optimization and Evaluation of a Support Vector Machine Classifier

After selecting an optimal feature subset, we conducted experiments to train and optimize a statistical machine learning classifier to combine these features, and then used the output (a score) generated by the classifier to predict the risk or likelihood of women having image-detectable breast abnormality or cancer. Although many types of machine learning classifiers (i.e., logistic regression, naive Bayes, artificial neural network) can be applied for this purpose, in this study we trained and optimized a support vector machine (SVM) based classifier using the LIBSVM algorithm (53). The LIBSVM classifier used a radial basis

function (RBF) kernel, defined as $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, $\gamma > 0$, on the training set of instance-label pairs (x_i, y_i) , $i = 1, \dots, l$ where $x_i \in R^n$ and $y \in \{1, -1\}^l$. We used a recommended five-fold cross-validation method with a parallel “grid-search” (54) to determine the penalty parameter of the error term and γ .

After selecting the optimal feature set and determining the SVM structure, we validated our SVM classifier or model performance using a 10-fold cross-validation method whereby the sum of the 283 positive cases and 362 negative cases were randomly divided into 10 exclusive partitions (or subgroups). In each validation (training and testing) cycle, nine subgroups were used to train the classifier, and the trained classifier was then applied to the remaining subgroup. For each testing case, the classifier generated an output score ranging from 0 to 1. A higher score indicates a higher probability/risk of having an image-detectable breast cancer in the next sequential screening examination. This process was iteratively executed 10 times using the different combinations of subgroups. In this way, each of the 645 positive or negative cases was tested once, with a corresponding classifier-generated risk score. In this experiment, the last trained classifier or model was also applied to all 349 benign cases to generate an additional set of 349 risk scores.

As discrimination and calibration are two major components in evaluating accuracy and potential clinical utility of a risk prediction model (55), we used the area under a receiver operating characteristic (ROC) curve (AUC) to assess discrimination, and an adjusted odds ratio (OR) to assess the calibration component of our new SVM-based risk model. First, AUC along with the corresponding 95% confidence interval was computed using a ROC curve fitting program that applies an expanded bi-normal model and the maximum likelihood estimation method (ROCKIT <http://www.radiology.uchicago.edu/krl/>, University of Chicago, USA, 1998). Secondly, the model-generated risk scores were sorted in ascending order, and five threshold values were selected, in order to segment all cases into five subgroups (bins) with approximately-equal number of cases in each subgroup. Thus, except subgroup 5 that has 198 cases, each of the other four subgroups (1 to 4) has 199 cases, summing up to the total number of 994 cases in our dataset. Then, the adjusted odds ratios (ORs) calculated from a multivariate statistical model in all subgroups and an OR increasing trend were computed and analyzed using a publically-available statistical computing software package (*R* version 2.1.1, <http://www.r-project.org>). The results were then tabulated and compared.

III. RESULTS

The SFFS algorithm selected an optimal feature set that includes 10 features. These are (1) woman's age; (2) mean pixel value difference in the dense breast region, computed according to equation (2); (3) mean value of short run emphasis in the whole breast region, computed according to equation (2); (4) maximum value of short run emphasis in the whole breast region, computed according to equation (2); (5) standard deviation of the y -axis cumulative projection histogram on the whole breast region, computed according to equation (2); (6) standard deviation of the y -axis cumulative projection histogram on the whole breast region, computed according to equation (1); (7) median of the x -axis cumulative projection histogram on the dense breast region, computed according to equation (1); (8) mean pixel value in the dense breast region, computed according to equation (3); (9) mean value of short run low gray level emphasis in the whole breast region, computed according to equation (3); and (10) median of the x -axis cumulative projection histogram on the whole breast region, computed according to equation (1). The SVM classifier evaluated in this study was then trained using these 10 features.

Figure 3 shows a ROC curve that was generated by the ROCKIT program using the SVM classifier-generated risk/probability scores for all 994 cases in our dataset. Using ROCKIT, AUC = 0.725 with a standard deviation of 0.018 and 95% confidence interval of [0.689, 0.759]. Figure 4 shows a similar ROC curve, computed only on the positive and negative cases (excluding the second group of benign cases); AUC = 0.716 with a standard deviation of 0.020 and 95% confidence interval of [0.675, 0.754] was obtained. Using the classifier prediction score of 0.5 as a threshold, Table 3 shows a confusion matrix that indicates that 71.3% (709/994) of cases are correctly predicted by our SVM model, whereas 28.7% (285/994) of cases are misclassified, at this prediction threshold. In the positive (cancer) case subgroup, 49.8% (141/283) are correctly predicted. In the negative/benign case subgroup, the overall prediction accuracy is 80.2% (570/711). By separating the negative and benign subgroups, the results show that 79.3% of the cases (287/362) in the not-recalled subgroup and 81.1% of the cases (283/349) in the recalled but proven-to-be-benign case subgroups are correctly predicted or stratified as having low risk of having image-detectable breast cancers in the next sequential screening mammography examination, respectively. This result indicates that the “prior” images of the majority of recalled benign cases have very comparable or even lower bilateral mammographic feature asymmetrical levels as compared to the not-recalled negative cases. The experimental results indicated that using a multi-feature based SVM model yielded significantly higher risk prediction result than using subjectively rated mammographic density based BIRADS as shown in Figure 2 in which AUC values computed by the ROCKIT program are 0.511 ± 0.023 between cancer and non-cancer subgroups of cases and 0.532 ± 0.026 between cancer and not-recalled negative subgroups of cases, respectively.

Table 2 summarizes the adjusted ORs and the corresponding 95% confidence intervals (CIs) computed for five subgroups (bins) of cases. The SVM classifier-generated risk scores gradually increase from subgroup 1 to 5. Using the cases in subgroup 1 as a baseline (reference), ORs and 95% CIs at subgroups 2 to 5 were computed and compared in Table 2. The increasing trend of ORs is shown as a function of increasing classifier-generated risk scores, comparing between positive (cancer) and negative/benign (non-cancer) case groups. The maximum, adjusted OR in subgroup 5 increases from 1.0 (in subgroup 1) to 12.34 with a 95% confidence interval of [7.12, 21.38]. As shown in Figure 5, the slope of the regression trend line between the scores generated by our SVM model and the adjusted ORs is significantly different from the zero slope ($p = 0.006$). This trend indicates that as the bilateral mammographic feature difference increases, the risk of women developing breast cancer in the near-term (e.g., the next sequential mammography screening, in our study) also increases.

IV. DISCUSSION

In this study, we computed and analyzed a large number of different image features related to the bilateral mammographic tissue density asymmetry, and then trained a SVM classifier-based model to predict breast cancer risk in the near term. We preliminarily investigated the association between the model-generated risk 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 demonstrates that our approach and the SVM classification model is able to achieve significantly higher prediction or classification performance than random guesswork (with $AUC = 0.725 \pm 0.020$ for positive and negative/benign prediction), and also shows an increasing trend in odds ratios of near-term cancer risk with the increase of model-generated risk scores.

Although many epidemiology-based breast cancer risk assessment models have been established and used to date (3), these models typically show significant statistical

associations with regards to lifetime risk of developing breast cancer compared to the general population. However, these models have relatively low discriminatory power or positive predictive value at the level of the individual woman for developing breast cancer in the near-term (i.e., < 5 years), which is required for routine personalized screening and clinical practice (35). This study is different. We focused on assessing near-term risk of developing breast cancer from a set of features computed from the bilateral mammographic density and/or tissue asymmetry-related information. Thus, our study and new SVM model provides novel or supplementary information in breast cancer risk prediction, in particular for near-term risk of developing cancer.

Although breast or mammographic density is considered as a strong risk factor in women for developing breast cancer in their lifetime, its discriminatory power to predict near-term cancer risk is low. In this study, when using mammographic density (BIRADS) subjectively rated by radiologists (as depicted in Figure 2) to predict the likelihood or risk of a woman having cancer(s) detected at the next sequential screening mammography examinations, AUC is only 0.51. However, using the proposed method to compute bilateral mammographic image feature asymmetry, we achieve significantly higher discriminatory power in predicting near-term breast cancer risk.

In order to improve the efficacy of breast cancer screening programs, the procedure of delaying mammographic screening age and reducing mammographic screening frequency has been recommended by U.S. Preventive Services Task Force (13), and also supported by many scientific researchers (14). On the other hand, public awareness of breast density as a high cancer risk factor has motivated many states in the United States of America to pass breast density notification legislation for breast cancer prevention, screening and diagnosis (56), which requires more screening for a large percentage of women using mammograms and/or other more advanced screening tools. As a result, the task of achieving a good balance in risk-benefit and cost-benefit analysis for breast cancer screening remains a difficult and unresolved issue. It is thus important to investigate and develop more effective or higher discriminative risk factors and/or risk prediction models, in particular those that enable more effective prediction of near-term risk after one or a series of negative screening examinations. The new model should more-accurately and reliably stratify women into high and low risk of developing breast cancer in the near-term.

For those identified as having a higher risk, annual screening may be necessary and important for detecting cancer at an early and treatable stage, whereas for those with lower risk, reducing screening frequency (e.g., from annual screening to every two-years (13)) will be beneficial to the individual woman (e.g., reducing the probability of false-positive recall) and society (e.g., eliminating costs for unnecessary imaging examinations and other medical procedures). Unlike many currently-known risk factors that maintain constant values (do not change), near-term risk factors may change over time. As a result, screening frequency in women can also be adaptively changed as prediction risk scores increase.

In the last two decades, a large number of computer-aided detection (CAD) schemes of mammograms have been developed and the commercialized schemes have also been used in current clinical practice of assisting radiologists to read screening mammograms; the scheme developed in this study is not another conventional CAD scheme that targets to detect specific abnormalities (i.e., masses or micro-calcification clusters). Our goal is to develop a new breast cancer risk stratification tool based on a new concept and image features that are considered to be associated with the risk of developing breast cancers. The hypothesis is based on the facts that humans naturally show bilateral symmetry in paired morphological traits including breasts, and as breast abnormalities and cancer start to develop in one breast, the asymmetry level of left and right breasts is also likely to increase (57). Hence, bilateral

mammographic density (or other image features of) asymmetry is a dynamic risk factor that might change over time, and may be a useful risk indicator to help adaptively adjust the screening frequency of a woman at different periods of her life.

Although this study demonstrates a preliminarily promising result, this study also has a number of limitations. Firstly, this is a retrospective study with a relatively-small dataset that may involve potential case selection bias and cannot fully represent the diversity in the general breast cancer screening population. Secondly, since this is a laboratory based study, the ratio of positive and negative cases does not represent the actual cancer prevalence ratio in the general breast cancer screening practice. Thirdly, in evaluating performance of our SVM prediction model, we used a 10-fold cross-validation method to increase computational efficiency, which may have resulted in higher bias than using a leave-one-case-out cross-validation method. Fourthly, the same dataset was used for both feature selection and evaluation of SVM model prediction accuracy, which may also have introduced bias in the SVM optimization process. Therefore, this new breast cancer risk stratification idea and model needs to be further examined and validated using larger and more diverse databases in future prospective and/or cohort studies.

Acknowledgments

This work was supported in part by the National Cancer Institute, National Institutes of Health, under Grant R01CA160205.

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin.* 2013; 63(1):11–30. [PubMed: 23335087]
2. Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN. Proportion of breast cancer cases in the United States explained by well-established risk factors. *J Natl Cancer Inst.* 1995; 87(22):1681–5. [PubMed: 7473816]
3. 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(10):680–91. [PubMed: 20427433]
4. Tabar L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer.* 2001; 91(9):1724–31. [PubMed: 11335897]
5. Smith RA, Cokkinides V, Brooks D, Saslow D, Shah M, Brawley OW. Cancer screening in the United States, 2011: A review of current American Cancer Society guidelines and issues in cancer screening. *CA Cancer J Clin.* 2011; 61(1):8–30. [PubMed: 21205832]
6. Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology.* 2002; 224(3):861–9. [PubMed: 12202726]
7. Kopans DB. The positive predictive value of mammography. *Am J Rotentgenol.* 1991; 158:521–6.
8. Fenton JJ, Wheeler J, Carney PA, et al. Reality check: perceived versus actual performance of community mammographers. *Am J Rotentgenol.* 2006; 187:42–6.
9. 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; 155:481–92. [PubMed: 22007042]
10. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. *Ann Fam Med.* 2013; 11(2):106–15. [PubMed: 23508596]
11. Yaffe MJ, Mainprize JG. Risk of radiation-induced breast cancer from mammographic screening. *Radiology.* 2011; 258(1):98–105. [PubMed: 21081671]
12. Buist DSM, Anderson ML, Haneuse SJPA, et al. Influence of annual interpretive volume on screening mammography performance in the United States. *Radiology.* 2011; 259(1):72–84. [PubMed: 21343539]

13. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2009; 151(10):727–37. [PubMed: 19920273]
14. Berlin L, Hall FM. More mammography muddle: emotions, politics, science, costs, and polarization. *Radiology.* 2010; 255(2):311–6. [PubMed: 20413746]
15. Brawley OW. Risk-based mammography screening: an effort to maximize the benefits and minimize the harms. *Ann Intern Med.* 2012; 156(9):662–3. [PubMed: 22547477]
16. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med.* 2008; 148(5):337–47. [PubMed: 18316752]
17. 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(1):10–20. [PubMed: 21727289]
18. Ravesteyn, NTv; Miglioretti, DL.; Stout, NK., et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: A comparative modeling study of risk. *Ann Intern Med.* 2012; 156:609–17. [PubMed: 22547470]
19. Nelson HD, Zakher B, Cantor A, et al. Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. *Ann Intern Med.* 2012; 156(9):635–48. [PubMed: 22547473]
20. Saftlas AF, Hoover RN, Brinton LA, et al. Mammographic densities and risk of breast cancer. *Cancer.* 1991; 67(11):2833–8. [PubMed: 2025849]
21. Boyd NF, Martin IJ, Stone J, Greenberg C, Minkin S, Yaffe M. Mammographic densities as a marker of human breast cancer risk and their use in chemoprevention. *Curr Oncol Rep.* 2001; 3(5): 314–21. [PubMed: 11389815]
22. American College of Radiology. Bi-RADS Breast Imaging Reporting and Data System Breast Imaging Atlas. Reston, VA: 2003.
23. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR Am J Roentgenol.* 2000; 174(6):1769–77. [PubMed: 10845521]
24. Taboces PG, Correa J, Souto M, Gómez L, Vidal J. Computer-assisted diagnosis: the classification of mammographic breast parenchymal patterns. *Phys Med Biol.* 1995; 40(1):103–17. [PubMed: 7708834]
25. Byng JW, Yaffe MJ, Lockwood GA, Little LE, Tritchler DL, Boyd NF. Automated analysis of mammographic densities and breast carcinoma risk. *Cancer.* 1997; 80(1):66–74. [PubMed: 9210710]
26. Heine JJ, Velthuisen RP. Spectral analysis of full field digital mammography data. *Med Phys.* 2002; 29(5):647–61. [PubMed: 12033559]
27. Zhou C, Chan HP, Petrick N, et al. Computerized image analysis: estimation of breast density on mammograms. *Med Phys.* 2001; 28(6):1056–69. [PubMed: 11439475]
28. Chang Y-H, Wang X-H, Hardesty LA, et al. Computerized assessment of tissue composition on digitized mammograms. *Acad Radiol.* 2002; 9(8):899–905. [PubMed: 12186438]
29. Pawluczyk O, Augustine BJ, Yaffe MJ, et al. A volumetric method for estimation of breast density on digitized screen-film mammograms. *Med Phys.* 2003; 30(3):352–64. [PubMed: 12674236]
30. Wang XH, Good WF, Chapman BE, et al. Automated assessment of the composition of breast tissue revealed on tissue-thickness-corrected mammography. *AJR Am J Roentgenol.* 2003; 180(1): 257–62. [PubMed: 12490516]
31. Wei J, Chan H-P, Helvie MA, et al. Correlation between mammographic density and volumetric fibroglandular tissue estimated on breast MR images. *Med Phys.* 2004; 31(4):933–42. [PubMed: 15125012]
32. Li H, Giger ML, Olopade OI, Margolis A, Lan L, Chinander MR. Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms. *Acad Radiol.* 2005; 12(7):863–73. [PubMed: 16039540]

33. Glide-Hurst CK, Duric N, Littrup P. A new method for quantitative analysis of mammographic density. *Med Phys.* 2007; 34(11):4491–8. [PubMed: 18072514]
34. Wei J, Chan HP, Wu YT, et al. Association of computerized mammographic parenchymal pattern measure with breast cancer risk: a pilot case-control study. *Radiology.* 2011; 260(1):42–9. [PubMed: 21406634]
35. Gail MH, Mai PL. Comparing breast cancer risk assessment models. *J Natl Cancer Inst.* 2010; 102(10):665–8. [PubMed: 20427429]
36. Harvey JA, Bovbjerg VE. Quantitative assessment of mammographic breast density: relationship with breast cancer risk. *Radiology.* 2004; 230(1):29–41. [PubMed: 14617762]
37. Kopans DB. Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. *Radiology.* 2008; 246(2):348–53. [PubMed: 18227535]
38. 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(4):315–20. [PubMed: 9690746]
39. Colin C, Prince V, Alette PJ. Can mammographic assessment lead to consider density as a risk factor for breast cancer? *Eur J Radiol.* 2013; 82(3):404–11. [PubMed: 20133095]
40. 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(23):3779–83. [PubMed: 20625126]
41. 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 risk stratification. *Acad Radiol.* 2010; 17(10):1234–41. [PubMed: 20619697]
42. Wang X, Lederman D, Tan J, Wang XH, Zheng B. Computerized prediction of risk for developing breast cancer based on bilateral mammographic breast tissue asymmetry. *Med Eng Phys.* 2011; 33(8):934–42. [PubMed: 21482168]
43. Zheng B, Sumkin JH, Zuley ML, Wang X, Klym AH, Gur D. Bilateral mammographic density asymmetry and breast cancer risk: a preliminary assessment. *Eur J Radiol.* 2012; 81(11):3222–8. [PubMed: 22579527]
44. Zheng B, Leader JK, Abrams GS, et al. Multiview-based computer-aided detection scheme for breast masses. *Med Phys.* 2006; 33(9):3135–43. [PubMed: 17022205]
45. Tang X. Texture information in run-length matrices. *IEEE Transactions on Image Processing.* 1998; 7(11):1602–9. [PubMed: 18276225]
46. Tzikopoulos S, Mavroforakis ME, Georgiou HV, Dimitropoulos N, Theodoridis S. A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. *Comput Methods Programs Biomed.* 2011; 102(1):47–63. [PubMed: 21306782]
47. Powell, WB. *Approximate Dynamic Programming: Solving the Curses of Dimensionality.* 1st edition. Wiley-Interscience; 2007.
48. Xu, L.; Yan, P.; Chang, T. 9th International Conference on Pattern Recognition. IEEE Computer Society Press; Rockville, MD: 1989. Best first strategy for feature selection.; p. 706-8.
49. Caruana, R.; Freitag, D. Greedy attribute selection.. In: Cohen, WW.; Hirsh, H., editors. 11th International Conference on Machine Learning. Morgan Kaufmann; New Brunswick, NJ: Los Altos, CA: 1994. p. 28-36.
50. Pudil P, Novovicova J, Kittler J. Floating search methods in feature selection. *Pattern Recognition Letters.* 1994; 15(11):1119–25.
51. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition.* 4th edition. Academic Press; 2008.
52. Theodoridis, S.; Koutroumbas, K. *Introduction to Pattern Recognition: A MATLAB Approach.* Academic Press; 2010.
53. Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology.* 2011; 2(3):27.
54. Hsu, C-W.; Chang, C-C.; Lin, C-J. *A practical guide to support vector classification* Technical report. National Taiwan University; Taipei 106, Taiwan: 2009.

55. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010; 21(1):128–38. [PubMed: 20010215]
56. The California report. [July 23, 2013] California's breast density notification law goes into effect. 2013. Available at: <http://blogs.kqed.org/stateofhealth/2013/03/29/californias-breast-density-notification-law-goes-into-effect/>.
57. Scutt D, Manning JT, Whitehouse GH, Leinster SJ, Massey CP. The relationship between breast asymmetry, breast size and the occurrence of breast cancer. *Br J Radiol*. 1997; 70(838):1017–21. [PubMed: 9404205]

Figure 1.

An example of a case in the subgroup of positive cases, which shows two sets of bilateral CC view mammograms acquired from the “prior” (left) and “current” (right) screening examinations of a woman. The “prior” images were interpreted as negative and a cancer (pointed by an arrow) was detected on the “current” image by a radiologist, and was later confirmed in biopsy and pathology examinations.

Figure 2.
Histogram distribution of the three groups of cases (negative, benign and positive) in four categories of mammographic density (BIRADS) ratings.

Figure 3.

ROC curve of applying our SVM model to predict the risk of having image-detectable breast cancer in the next sequential mammography screening examination, examined on all positive, negative and benign cases with AUC = 0.725 (reference AUC = 0.5).

Figure 4.

ROC curve of applying our SVM model to predict the risk of having image-detectable breast cancer in the next sequential mammography screening examination, examined on positive and negative cases only (AUC = 0.716; reference AUC = 0.5).

Figure 5.
Trend of the increase in odds ratios with the increase in the risk scores generated by the trained SVM classifier.

Table 1

Computed features according to grouping or type, and their description/ explanation

Feature group/ type	Feature ID	Notes/ description
Generic gray value (statistical)-based features, and a fractal dimension feature [28] *	1-10	Mean, standard deviation, skewness, and kurtosis of pixel values. Fractal dimension feature originally described in [28]
Mean and maximum values of gray level run length-based texture features, computed along 0 and 90 degrees (horizontal and vertical directions) [45]	11-32	Short run emphasis, long run emphasis, low gray level run emphasis, high gray level run emphasis, short run low gray level emphasis, short run high gray level emphasis, long run low gray level emphasis, long run high gray level emphasis, gray level non-uniformity, run length non-uniformity, and run percentage
X-axis and y-axis histogram (cumulative projection) features [50] *	33-52	Mean, standard deviation, skewness, kurtosis, and median values
Gray level co-occurrence matrix-based features *	53-60	<p>Statistics computed from a gray level co-occurrence matrix (created by calculating how often a pixel with gray level value i occurs horizontally adjacent to a pixel with value j), namely:</p> $\text{Contrast} = \sum_{i,j} i - j ^2 p(i, j)$ $\text{Correlation} = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j}$ $\text{Energy} = \sum_{i,j} p(i, j)^2$ $\text{Homogeneity} = \sum_{i,j} \frac{p(i, j)}{1 + i - j }$
Biological features	181-183	Age, family history and breast density

* Features computed on both the whole and dense breast regions

Table 2

Relative, adjusted odds ratios (ORs) and 95% confidence intervals (CIs) with increasing levels of the trained SVM classifier-generated risk scores

Subgroup	Number of Cases (Positive – Negative/Benign)	Adjusted Odds Ratio (OR)	95% Confidence Interval (CI)
1	19 – 180	1.00	baseline
2	33 – 166	1.88	[1.03, 3.44]
3	59 – 140	3.99	[2.28, 7.00]
4	60 – 139	4.09	[2.33, 7.17]
5	112 – 86	12.34	[7.12, 21.38]

Table 3

Confusion matrix of prediction results using our proposed method, and obtained by applying a threshold of 0.5 on the classifier-generated risk/probability scores

Prediction ->	Negative/Benign Cases	Positive Cases
Negative/Benign Cases	570	141
Positive Cases	144	139