Development and Assessment of a New Global Mammographic Image Feature Analysis Scheme to Predict Likelihood of Malignant Cases

Morteza Heidari, **Seyedehnafiseh Mirniaharikandehei**

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

Wei Liu

School of Computer Science, Xi'an University of Post and Communication, Xi'an, Shaanxi 710121, China

Alan B. Hollingsworth

Department of Surgery, Mercy Hospital, Oklahoma City, OK 73120, USA

Hong Liu, **Bin Zheng**

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

Abstract

This study aims to develop and evaluate a new computer-aided diagnosis (CADx) scheme based on analysis of global mammographic image features to predict likelihood of cases being malignant. An image dataset involving 1,959 cases was retrospectively assembled. Suspicious lesions were detected and biopsied in each case. Among them, 737 cases are malignant and 1,222 are benign. Each case includes four mammograms of craniocaudal and mediolateral oblique view of left and right breasts. CADx scheme is applied to pre-process mammograms, generate two image maps in frequency domain using discrete cosine transform and fast Fourier transform, compute bilateral image feature differences from left and right breasts, and apply a support vector machine (SVM) to predict likelihood of the case being malignant. Three sub-groups of image features were computed from the original mammograms and two transformation maps. Four SVMs using three sub-groups of image features and fusion of all features were trained and tested using a 10-fold cross-validation method. The computed areas under receiver operating characteristic curves (AUCs) range from 0.85 to 0.91 using image features computed from one of three sub-groups, respectively. By fusion of all image features computed in three sub-groups, the fourth SVM yields a significantly higher performance with AUC = 0.96±0.01 (p<0.01). This study demonstrates feasibility of developing a new global image feature analysis based CADx scheme of mammograms with high performance. By avoiding difficulty and possible errors in breast lesion segmentation, this new CADx approach is more efficient in development and potentially more robust in future application.

morteza.heidari@ou.edu.

Index Terms—

Computer-aided Diagnosis Scheme; CAD of Mammograms; Breast Cancer Detection; Global Mammographic Image Feature Analysis

Introduction

Mammography is the only clinically acceptable imaging modality to detect breast cancer in current population-based breast cancer screening [1]. Due to the quite low cancer detection yield (i.e., detecting 3.6 cancers per 1000 (0.36%) mammography screenings [2]) and higher recall rate (i.e., ~10%) in breast cancer screening environment, reading and interpreting screening mammograms is difficult and time-consuming for radiologists [3]. To assist radiologists more accurately and efficiently reading mammograms and reducing inter-reader variability, computer-aided detection (CADe) schemes of mammograms have been developed and used in the clinical practice as "a second reader" for the last two decades [4]. Although previous observer performance studies reported that using CADe might help radiologists detect more cancers that may be previously missed or overlooked by radiologists (i.e., [5]), the clinical data analysis studies showed that using CADe increased false-positive recalls and reduced radiologists' performance measured by areas under the receiver operating characteristics curves (i.e., [6]). Thus, the specificity of current mammographic imaging remains lower in clinical practice. Approximately only one in four lesion biopsies are proved to be malignant [7]. The higher false-positive recall rates add anxiety with potentially long-term psychosocial consequences [8] and physical harms to many cancer-free women who participate in mammography screening due to cumulative x-ray radiation and unnecessary biopsies [9]. The high false-positive recall rates also associate with a high economic burden on the healthcare system [10], Thus, in order to help improve efficacy of mammography screening, developing the computer-aided diagnosis (CADx) schemes that aim to assist radiologists in their decision-making to better assess risk of the detected suspicious breast lesions being malignant and reduce the unnecessary biopsies of benign lesions have been attracted broad research interest for the last two decades [11].

Despite great research effort, CADx schemes of mammograms have not been accepted and used in clinical practice. It still faces multiple technical challenges to improve CADx performance and robustness. For example, previous schemes typically include 3 steps which (1) apply image processing algorithms to segment suspicious lesions depicting on mammograms, (2) compute images features from the segmented regions, and (3) train multi-feature fusion-based machine learning classifiers [12]. However, due to the overlap of dense fibro-glandular tissue on mammograms, accurate lesion segmentation is often difficult and unreliable, which can substantially affect performance and robustness of CADx schemes [13]. To overcome this difficulty, researchers recently investigated and applied deep learning techniques to develop CADx schemes without lesion segmentation and hand-crafted feature computation [14, 15]. Although deep learning approach can avoid difficulty in lesion segmentation and manually defining image features, it requires a large and diverse image dataset to train the scheme to minimize the risk of overfitting and validate its performance, which is another difficult task.

To address these challenges, we recently investigated feasibility of developing new computer-aided quantitative image feature analysis schemes or machine learning models based on the global mammographic image features to predict risk of developing breast cancer in a short-term [16, 17] or risk of depicting suspicious lesions on mammograms [18]. Similar global image feature analysis schemes can also be developed using different imaging modalities to predict other clinical outcomes, such as response of breast cancer patients to neoadjuvant chemotherapies using breast MRI [19] and response of ovarian cancer patients to chemotherapy using CT images [20]. By avoiding difficulty and errors in lesion segmentation, our studies have demonstrated advantages of developing and applying the global image feature analysis schemes in CAD-related quantitative image informatics field, which has potential to be more efficient and robust.

In this study, we hypothesized that the similar global image feature analysis schemes can be developed and applied to predict likelihood of cases being malignant if the suspicious lesions (i.e., soft tissue-based masses) are detected by the radiologists on mammograms. Thus, we proposed to investigate a new CADx scheme with 2 unique approaches. First, the new CADx scheme identifies and selects image features computed from the entire breast area depicting on the mammograms of left and right breasts. This global approach is different from previous local region or lesion-based CADx schemes that either require lesion segmentation or define regions of interest (ROI) with the fixed size to cover the suspicious lesions. Second, the new CADx scheme uses the bilateral asymmetrical image features computed from left and right breasts. As a result, unlike the conventional single-image based CADx scheme, this is a multi-image fusion based CADx scheme. Although this is new approach, advantages of developing multi-image fusion-based CADe schemes over the single-image based CADe schemes have been demonstrated in previous studies (i.e., [21]). If successful, this new approach may enable to provide radiologists a new CADx-generated image marker or risk prediction score to support their decision-making in classifying between malignant and benign lesions to increase diagnostic accuracy (including reduction of false-positive recalls and unnecessary biopsies of benign lesions). Thus, objective of this study is to test our hypothesis using a relatively large and diverse digital mammography image dataset.

MATERIALS AND METHODS

Image Dataset

We retrospectively assembled a full-field digital mammography (FFDM) image dataset, which involves the fully anonymized images acquired from 1,959 patients who underwent routine annual mammography screening with the age ranging from 35 to 80 years old. In these patients, suspicious lesions were detected by radiologists in the original mammogram reading and diagnosis. All detected suspicious lesions were recommended and performed biopsy. Based on the histopathology examinations of the biopsy-extracted lesion specimens, 737 cases were confirmed as positive for cancer, while other 1,222 cases had biopsy-approved benign masses.

Each mammography case involves 4 images of craniocaudal (CC) and mediolateral oblique (MLO) view of the left and right breasts. The original FFDM images have a pixel size of

$70\mu m$. Like the conventional CAD schemes of mammograms, all images were subsampled using a pixel averaging method with a 5×5 pixel frame to make the image size of 818×666 pixels and 12-bit pixel depth. Each pixel size is increased to $0.35mm$ [22]. Table I summarizes and compares case distribution of patient age and mammographic density rated by radiologists using BIRADS guidelines. The patients in benign groups are relatively younger than those in malignant group, but there is no significant difference in BIRADS density rating (p=0.878).

Background of Image Features

After segmenting breast area from the surrounding area region depicting on each mammogram [17], we applied a computerized scheme to extract and compute global image features from the original mammograms in spatial domain and the transformed maps in frequency domain. Specifically, the feature extraction algorithm relies on the basic fact that mammography images are highly structured, which means their pixels exhibit strong dependences. Under presence of cancer the pixel dependency would change not only in the region of lesions, but also the surrounding parenchymal tissues in breast area. In addition, since radiologists are quite sensitive to bilateral image feature differences related to the structural information between left and right breasts in detecting suspicious lesions and distinguish malignancy cases from benign ones, we will extract and compute the global bilateral image feature difference between the left and right CC or MLO view images to build the machine learning model for predicting risk of the cases being malignant.

From the original FFDM images, we computed image features and applied the structural similarity index (SSIM) to measure the similarity between 2 bilateral images of the left and right breasts. SSIM was originally proposed to assess image quality based on structural similarity [23]. It has been widely used in medical imaging field with higher correlation to human visual system adopted to extract structural information of images including our previous studies (i.e., [16]). For the SSIM assessment, if we assume two nonnegative image signals like $\mathbf{x} = \{x_i | i = 1,2,\ldots,M\}$ and $\mathbf{y} = \{y_i | i = 1,2,\ldots,M\}$ as two patches of each image that have been aligned to each other, we can calculate SSIM index using the following equation [24]:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1}$$

where $\sigma_{xy} = \frac{1}{M}\sum_{i=1}^{M}(x_i - \mu_x)(y_i - \mu_y)$, $\mu_x = \frac{1}{M}\sum_{i=1}^{M}x_i$, $\sigma_x = \frac{1}{M}\sum_{i=1}^{M}(x_i - \mu_x)^2$, $\sigma_y = \frac{1}{M}\sum_{i=1}^{M}(y_i - \mu_y)^2$ and C_1, C_2 are two positive constant.

Thus, SSIM index values range between zero and one. The maximum value is achieved when the input images are identical. The more the two input images are bilaterally different to each other, the smaller the corresponding calculated SSIM index values are.

In addition, to take the advantages of computer vision that can be relatively easy to acquire and analyze image features in the frequency domain, we performed two transformations to compute bilateral image feature differences in the frequency domain. Specifically, we

applied discrete cosine transform (DCT) and fast Fourier transform (FFT) as two similar and complementary ways to facilitate detecting and analyzing useful image contents change across the whole image just in a small number of components. In general, under these transformations lower spatial frequency coefficients contain more information than higher frequency components.

For DCT transformation, by assuming $f(x,y)$ as an image with size M by N, we applied the following general equation to the image to calculate 2D DCT of the input image:

$$F(m, n) = \frac{2}{2\sqrt{MN}} C(m)C(n) * \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cos\left((2x+1)m\pi / 2M\right) * \cos\left((2y+1)n\pi / 2N\right) \tag{2}$$

In this equation $C(m) = C(n) = 1/\sqrt{2}$ for $m, n = 1$ and $C(m) = C(n) = 1$ otherwise.

DCT transforms the information contained in pixels of special domain to frequency domain. The element in the top left corner of 2D DCT matrix is the DC term and contains a value that is almost always of a great magnitude, which is summation of all pixel values. On a zigzag scanning from the top left to the bottom right corner, the farther away from the DC term, it will have components with the higher frequency with the smaller corresponding magnitude [25].

FFT transformation computes the discrete Fourier transform of its input sequence. If the input image has a specific pattern, this transform can detect it in the magnitude spectrum components. By assuming $f(x,y)$ as an image with size M by N, we used the following general equation to calculate 2D discrete Fourier transform of the input images:

$$F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp\left[2\pi i \left(\frac{xu}{M} + \frac{yv}{N}\right)\right]; \begin{cases} u = 0, 1, \ldots, M-1 \\ v = 0, 1, \ldots, N-1 \end{cases} \tag{3}$$

where u, v are special frequencies and $|F(u, v)| = \sqrt{F_{Re}^2(u, v) + F_{Im}^2(u, v)}$ represents magnitude spectrum, which is useful to extract specific patterns [26].

Texture of each image shows important properties of distribution pattern of the fatty and fibro glandular tissues of breast. Texture also determines local spectral or frequency content of an image. In frequency domain, it is mostly projected to the low frequency coefficients. On the other hand, the noise unrelated to any specific pattern (like additive noise pattern) because of their randomness nature are mostly projected to the high frequency components. Furthermore, in [27] it has been shown that if scanning DCT coefficients in a zigzag order, the absolute DCT coefficient values are somehow correlated to each other, which means that the absolute DCT coefficient values are horizontally, vertically, and diagonally correlated to each other. Magnitude spectrums of FFT components also have the same characteristics in each local area. We will take advantage of these attributes for feature extraction.

Moreover, because of particular characteristics of benign and malignant tumors like intensity, shape and texture [28], the structural patterns of breasts depicting benign or malignant tumors would be different. The larger or higher grade of the malignant tumor for a case is, the larger disturbance or bilateral structural differences caused by the cancer manifests become more obviously. The disturbance can be enlarged in the difference of absolute value of extracted image features from the frequency coefficients of left and right breast by considering fact that the presence of a tumor can disturb the correlation of frequency coefficients.

Data Preprocessing

CADx scheme applies an image preprocessing phase to the whole FFDM images. This step includes two algorithms namely, a cropping operation and an image enhancement. Cropping is just applied to the MLO view to detect and remove chest wall area, and enhancement is done on both CC and MLO views of images to remove or reduce image noise on the black background area as well as written labels.

MLO view images typically have advantage over CC view because almost all the breast area is visible, which means we can extract more information from this view, especially for the global feature extraction methods. The main disadvantage of this view is that images also include chest wall and part of the pectoral muscle regions. Pectoral muscle area is typically brighter compared to the breast tissue and will have negative affect on the extracted features. Due to the great variation or heterogeneity of mammograms in different cases, although developing automated schemes to segment pectoral muscle has been tried before (i.e., [29]), it remains difficult to achieve robust results when applying to a large and diverse dataset. In this study, we used a hand craft method to remove pectoral muscle area. For each MLO image two points are determined at the margin of chest wall, then by plotting a straight line between these two points, the chest wall is determined and the pectoral muscle region is deleted to select the remaining breast region for further analysis. Other non-breast areas (i.e., labels) are also automatically deleted. An example of this preprocessing phase is given in Fig. 1. After the pre-processing phase all images in the dataset are saved in Portable Network Graphics (PNG) format of a lossless mode for feature extraction phase.

Image Feature Extraction

After image pre-processing, the computer-aided scheme is applied to extract and compute relevant image features from the entire breast area segmented on FFDM images. These features are divided into 3 subgroups from both spatial domain and frequency domain. First, from the original FFDM images (spatial domain), the scheme computes SSIM-related features of left and right images of CC or MLO view in a tree structural shape base, which is inspired by the commonly used hierarchical methods in video data processing area for motion estimation purpose [30]. Since each FFDM image has an original size of 818×666 pixels, each left and right image is first divided to 4 sub-blocks with a size of 409×333 pixels each. SSIM is computed using Equation (1) for all pairs of 4 sub-blocks in the matched position of the left and right breast. The sub-block with the smallest SSIM value, which means the highest bilateral asymmetry among these 4 pairs of the matched sub-blocks, is selected. Next, the scheme continues to divide the selected sub-blocks into 4 sub-

blocks again with a size of 205 × 167 pixels each. Four SSIM indices are computed and the new sub-block with the smallest SSIM index value is selected again. Such a process is repeated 6 times or iterations. In the last iteration, the size of the sub-block is reduced to 13 × 11 pixels. From these 6 iterations, the scheme selects 6 SSIM index values representing the highest bilateral asymmetry of breast tissue patterns with gradually deceased sub-block size. Fig. 2 illustrates a block diagram of this process.

In computing SSIM index value, several parameters need to be determined by the experiments. Based on our experimental results in computing SSIM of bilateral FFDM images for cancer risk assessment [16], the default parameter values are set up as 0.05 for constants C_1 and C_2, and 8 for window size used in Equation (1). Additionally, due to the heterogeneity of clinical cases (i.e., the variation lesion size and surround parenchymal tissues), it is not possible to predetermine an optimal sub-block size to compute SSIM index. Thus, in this study, we selected all 6 smallest SSIM index values computed in above iterations to build a SSIM feature pool or vector.

Second, after DCT and FFT transformation, the scheme computes 2 two-dimensional (2D) DCT and FFT matrixes of the whole input image using Equations (2) and (3), respectively. Hence, each image has a 2D matrix of DCT coefficients and a 2D matrix of FFT coefficients. By filtering out the last 10 percent of high frequency components, the redundant information is mostly filtered out with respect to the information related to the pattern of the breast. In this way, frequency domain coefficients are more suitable for feature extraction rather than pixel domain coefficients.

After preprocessing on the frequency coefficients, the 2D matrixes are changed to row format to reduce computational complexity of feature extraction phase. Hence, a sequence like $X = (x_1, x_2, …, x_k)$ represents these coefficients in row format. Then, the following features are extracted and computed.

From the DCT and FFT frequency domain, our scheme computed following statistical moments related features. Based on [31], by assuming that a sequence like $X = (x_1, x_2, …, x_N)$ is a finite population of size N, the scheme can compute an unknown probability density function (PDF) $p(x)$ for this targeted population. The nth row moment for this population is given by:

$$\widehat{m}_n = \frac{1}{N} \sum_{i=1}^{N} x_i^n \tag{4}$$

where the 1st row moment ($n = 1$) is mean (μ) of this population. By centralizing this equation, the scheme calculates the next centralized momentums for the population with:

$$\widehat{m}_n = \sum_{i=1}^{N} p_i(x_i - \mu)^n \tag{5}$$

That is an unbiased estimate of nth moment:

$$m_n = EX^n = \int_{-\infty}^{\infty} p(x)x^n dx \qquad (6)$$

According to Equation (6), $p(x)$ is weighted by x^n, so that any change in the $p(x)$ is polynomially reinforced in the statistical moments. Thus, by considering DCT and FFT components as finite populations of the input images, any changes in their PDF due to presence of malignant lesions is polynomial reinforced in the statistical moments of the computed coefficients. In this study, we utilized the statistical moments to catch bilateral image feature differences in both DCT and FFT maps of left and right breasts. Using Equation (4), the scheme computes mean of the frequency components, and using Equation (5) for $n - 2,3,4$, the scheme computes variance, skewness, and kurtosis of the frequency components. Additionally, the scheme also computes other popular statistical features including entropy, correlation, energy, root mean square level, uniformity, max, min, median, range, and mean absolute deviation from the DCT and FFT maps. Then, the absolute differences of these matched image features from the left and right view maps are computed to represent global bilateral differences of the left and right breasts in DCT and FFT based frequency domains. Table II also lists the 14 features computed from DCT and FFT maps.

In summary, our scheme computes 34 features from two bilateral images or maps of the left and right breasts (as shown in Table II). Since each case has two sets of bilateral images acquired from CC and MLO view, the totally computed image features are 68.

Fig. 3 shows a schematic diagram of the feature extraction phase to show how the scheme extracts each sub-group of features from each of 4 individual images (LCC, RCC, LML,RML) of a case and combine them to create the final feature vector (F_{fusion}) of 34 features in each of CC or MLO views. Specifically, $lccF_{dct}$ and $lccF_{fft}$ are DCT and FFT features computed from CC view image of left breast, while $rccF_{dct}$, $rccF_{fft}$ are DCT and FFT features computed from CC view image of right breast. Similarly, $lmlF_{dct}$, $lmlF_{fft}$ and $rmlF_{dct}$, $rmlF_{fft}$ are DCT and FFT features computed from MLO view images of left and right breast, respectively. Last, $F_{ssim}cc$ is vector of SSIM features related to two bilateral CC view images (LCC, and RCC), and $F_{ssim}ml$ is vector of SSIM features related to two bilateral MLO view images (LML and RML).

After computing these 34 image features from the bilateral images of one view, we computed and generated 2 correlation matrices for CC and MLO view (Fig. 4). The results indicate that majority of these features are not highly correlated (i.e., $r < |0.25|$ as shown by the light to dark blue color in Fig. 4), which can provide complementary information to predict the likelihood of the case being malignant.

Classification Phase

In this phase, we built multiple feature fusion-based machine learning models to predict the likelihood of the cases being malignant. Although many different types of machine learning

classifiers (i.e., artificial neural network, Bayesian belief network, and logistic regression model) can be used for this purpose, based on our previous experience in developing the variety of CAD schemes of medical images, we chose to train and build support vector machine (SVM) based machine learning models to predict the likelihood of the cases being malignant. To achieve high robustness, a popular RBF kernel was selected to build the SVM model, which has demonstrated good performance and low computational cost in our previous studies [32, 33]. Specifically, in each CC or MLO view images, we built 4 SVM models using image features computed from (1) the original FFDM image (6 SSIM based features), (2) DCT maps (14 features), (3) FFT maps (14 features), and fusion of 34 features computed. After comparing the performance of the SVMs trained using only one view images, we also fuse the image features computed from the two view images to retrain and test 4 new SVM models.

Each SVM-based prediction model is applied to the entire image dataset of 1,959 cases to predict likelihood of the cases being malignant. To train each SVM and assess its performance, we applied a 10-fold cross-validation method. The SVM model produces likelihood or prediction scores ranging from 0 to 1 in the testing phase. The higher score indicates the higher risk or likelihood of the case being malignant. Using the prediction scores computed from all 1,959 cases, a receiver operating characteristic (ROC) curve is generated and the area under the ROC curve (AUC value) is computed as an evaluation index.

Then, to evaluate an absolute classification accuracy for the proposed scheme, we also applied an operating threshold (T = 0.5) on the SVM-generated prediction scores. All cases are divided into two malignant and benign classes to generate a confusion matrix. From the confusion matrix, the overall prediction or classification accuracy, sensitivity, specificity and odds ratio (OR) are calculated as well. Furthermore, we sort the SVM-generated detection scores for all cases in an ascending order and select 5 threshold values to segment all cases into 5 sub-groups. Then, based on the available multivariate statistical model included in a statistical software package (R version 2.1.1, http://www.r-project.org), we calculated the adjusted OR values and detected the possible ORs increasing trend with the increased classification scores.

In addition, to test whether we can further reduce dimensionality of the feature space to identify better features, we applied Principal Component Analysis (PCA) as a feature analysis and regeneration method to reduce feature vector size and train SVM models. The performance levels of the SVM models trained with and without applying PCA method were compared. All computation tasks were conducted using MATLAB R2019a package. Fig. 5 illustrates a complete block-diagram of the proposed scheme and testing method.

RESULTS

Fig. 6 shows CC and MLO images of one malignant and one benign case. Using the global bilateral image feature analysis, the SVM-generated prediction scores are 0.82 and 0.37 in these 2 cases, respectively. Table III shows and compares AUC values and overall classification accuracy after applying the operation threshold (T = 0.5). The results show that

using the image features computed from the bilateral MLO view images yielded significantly higher performance than using image features computed from bilateral CC view images ($p < 0.05$).

Table IV summarizes and compares the computed AUC values of 4 SVM models trained using image features computed from both CC and MLO view images. Fig. 7 shows 4 corresponding ROC curves. The results indicate that using 3 subgroups of features computed from the original FFDM images and 2 transformation maps, AUC values range from 0.85 to 0.91. After fusion of all image features computed from 3 subgroups, AUC value of the 4th SVM model significantly increases to 0.96±0.01 (with p<0.01). In addition, the standard deviation after fusion of 3 subgroups of image features is also substantially decreased as comparing to the use of one subgroup of image features, which indicates the increase of reliability of the 4th SVM model performance (AUC value).

Table V shows and compares 4 confusion matrices generated by the SVM-generated prediction scores after applying an operational threshold (T = 0.5). From these confusion matrices, additional performance indices can be computed as shown in Table VI. It shows that SVM trained using subgroup of DCT features yields the highest overall prediction accuracy as comparing to other two SVMs trained using SSIM and FFT features. However, by fusion of all 68 features, the SVM model yields further increased overall prediction accuracy (92%).

After dividing 1,959 testing cases into 5 subgroups of approximately equal number of cases (~392) based on the SVM-generated prediction scores (Table VII), the adjusted odd ratios (OR) increased from 1.0 in the baseline subgroup with lowest classification scores to 25,220 in the 5th subgroup with the highest prediction scores (the highest chance of being malignant). Regression analysis of the adjusted OR data also shows an increase trend of odds ratios with the increase in SVM-generated prediction scores. The slope of the regression trend line between the adjusted ORs and SVM-generated scores is significantly different from zero slope ($p < 0.01$).

By applying a PCA algorithm to reduce feature space dimensionality, we trained and tested SVM models with the increased number of the PCA-regenerated features. The highest AUC value is 0.94 and the highest overall prediction accuracy after applying the same operation threshold of T = 0.5 is 91%, which involves 65 numeric components produced by the PCA algorithm. The performance is slightly lower than the SVM trained using all 68 features as shown in Tables IV and VI.

Moreover, to test the significance of preprocessing phase (as shown in Fig. 1) of whether removal of pectoral muscle regions can boost performance of the CADx scheme, we recomputed all image features from the bilateral MLO view images without applying removing the pectoral muscle and retrained SVM classification models. Table VIII compares the classification performance of the SVM models trained and tested using the image features computed from MLO view images with and without removing pectoral muscle areas depicted on the images. The result shows significant improvement of classification performance by removing the pectoral muscle areas from the MLO view images (p<0.05).

For example, the AUC value of the SVM model using all image features computed from the original mammograms and two transformation maps increases more than 15% (from 0.79 to 0.94).

DISCUSSION

This study has several unique characteristics and generates several new interesting observations. First, although many CADx schemes of mammograms (i.e., as reviewed in [12]) have been previously developed and tested to classify between malignant and benign lesions, their performance is often limited by the difficulty and errors of lesion segmentation due to the fuzziness of lesion boundary and irregular tissue overlap in 2D mammograms [13]. Since unlike CADe schemes that aim to automatically detect suspicious lesions in which correctly cuing location of the lesion is important, CADx schemes apply to the cases in which suspicious lesions and their locations have already been detected by radiologists. The important issue in CADx schemes is to determine the likelihood of the case or the detected lesion being malignant. However, accurately predicting the likelihood of the detected suspicious lesions being malignant remains a difficult task for radiologists, which result in higher false-positive recall rates and higher rates of benign biopsy in current clinical practice. Thus, developing a more accurate and robust CADx scheme as an assistant tool to support radiologists in their decision-making is important no matter whether the CADx scheme uses local (region) or global image feature analysis. In this study, we explored a new approach to develop a unique case-based CADx scheme based on the detection, computation and analysis of globally asymmetrical image features computed from two bilateral images of left and right breasts and assessed its performance using a relatively large image dataset of 1,959 cases. Thus, this new CADx scheme is a multiply image-based scheme that integrates image feature differences computed from 4 view images, which makes it significantly different from other previously single or region-based CADx schemes.

Second, we explored and tested 3 types or subgroups of global image features computed from the original FFDM images and their transformation maps aiming to more accurately predict likelihood of cases being malignant. From a pair of bilateral mammograms, SSIM is used in a quadratic-tree based format that searches through different sub-blocks of original images (in a spiral way) to select areas with the highest level of bilateral asymmetry between left and right images of each case. In this way, area outside of the breasts are removed automatically because of high SSIM values, while the area with the smallest SSIM value, which represents the highest bilateral difference, is selected. The physical meaning of the proposed SSIM-based algorithm can be well described to mimic the image features used by the radiologists to asses and interpret tumors in the clinical practice. However, there is large variation of lesion size and asymmetrical structure of the surrounding parenchymal tissue patterns in the different clinical cases. In order to automatically compensate such variations, we used an iteration approach to compute a SSIM vector with 6 SSIM index values, which represent 6 pairs of the matched sub-blocks with gradually reduced size (from 409×333 to 13×11 pixels). The correlation coefficients of these 6 SSIM index values are relatively low (as shown in Fig. 4). Thus, fusion of these 6 SSIM features can increase prediction power of using SSIM applying to a large and diverse image dataset.

Third, to further take the potential advantages of computer vision over human vision, we explored image features computed from the frequency domain. For example, FFT and DCT have been widely used as two popular frequency domains for image feature extraction in many CADx schemes to classify between malignant and benign lesions (i.e., [13]), and predict tumor response to chemotherapy (i.e., [34]). In this study, we extracted absolutely asymmetrical feature values computed from two bilateral view images (CC or MLO view of the left and right breasts) and investigated their feasibility to predict likelihood of cases being malignant, which is also a new approach in CADx schemes of mammograms. We observed that the DCT-based features yielded the highest AUC value (as shown in Table IV), which shows the importance of identifying an optimal transformation map in the frequency domain for image feature extraction.

Fourth, the previous studies have reported that quantitatively detect and analysis of image features computed from MLO view images typically yielded higher performance than using image features computed using CC view images, such as applying CAD schemes to predict breast cancer risk [35] and detect suspicious lesions [36]. In this study, we systematically analyzed and compared correlation coefficients of image features computed from CC and MLO view images. The results showed that the image features computed from the MLO view images had lower distribution of the correlation coefficients than the image features computed from CC view images (Fig. 4). This supports the results that SVMs trained using MLO view images yield higher prediction accuracy than the SVMs trained using CC view images (i.e., Table III). In addition, by applying a PCA algorithm to search for and regenerate the optimal feature vectors, the best prediction performance yielded when using 65 numeric components produced by the PCA algorithm remains lower than the SVM trained using all 68 image features computed in 3 subgroups. The results indicate that although 68 features build a relatively large feature vector or space, when considering size of our dataset of 1,959 cases, the ratio between the numbers of the cases per class and image features remains relatively bigger (i.e., >10 per class). Thus, this size of feature vector is acceptable in this study.

Fifth, the study also shows that image features computed from the original mammograms and transformation maps contain the complementary information or discriminatory power. Thus, optimally combining or fusing multiple features computed from different feature domains to build a machine learning model or classifier can further significantly increase CADx prediction performance (i.e., AUC value and overall accuracy after applying an operating threshold). Automatically and optimally integrating image features from the different domains is an advantage of using machine learning based schemes over human observers. Additionally, the study results also show that removing pectoral muscle regions from the MLO view images can help increase prediction power to more accurately distinguish between malignant and benign cases. Thus, it is still important to develop algorithms that can more accurately and robustly detect chest wall and remove pectoral muscle regions in mammograms [29].

Sixth, many CADx studies have been previously reported in the literature to classify between malignant and benign lesions. For example, reference [12] presents a table that summarizes 8 previous CADx studies, which used image datasets ranging from 38 to 1,200

cases and yielded performance of AUC values ranging from 0.70 to 0.86. Another CADx scheme that used a dataset of 560 regions of interest and a deep learning model to classify between malignant and benign breast masses reported an AUC of 0.79 [14]. This study used a larger dataset involving 1,959 cases. Although we cannot directly compare the performance between this new case-based CADx scheme and previous CADx schemes reported in the literature due to the use of different image datasets, the high prediction or classification result (i.e., AUC value) of this study is encouraging. Unlike CADe schemes, which detect specific lesions and information of lesion location is important, determining lesion location is less important in CADx schemes because the suspicious lesions have been visually detected and located by radiologists. Thus, both the conventional CADx schemes based on analysis of the image features computed from the segmented lesions and this new CADx scheme based on analysis of the global image feature difference can play the same role to support radiologists in their decision-making of predicting likelihood of the detected lesions being malignant. By avoiding the difficulty and possible errors of breast lesion segmentation, developing a new CADx scheme based on global mammographic image feature analysis approach can potentially be more efficient and robust.

Last, despite the encouraging results and many new observations, we recognize that this is a laboratory based retrospective data analysis study with several limitations. First, although we assembled a relatively large and diverse image dataset, case selection bias is always an issue of concern. Second, the ratio between malignant and benign classes does not represent the actual cancer prevalence ratio in the general clinical practice. Hence, the performance and robustness of this new CADx scheme need to be further assessed and validated in future studies with new image datasets that better represent clinical practice. Third, based on the experience of our previous studies, we only explored and tested the limited numbers and types of image features, as well as the simple SVM models in this study. This may not be an optimal approach. How to identify and select optimal features and machine learning models need to be further investigated in future studies. Furthermore, this is a primary technology development study. Its clinical utility or impact on radiologists' performance in diagnosis of breast cancer using mammograms has not been tested. In summary, despite these limitations, this study has presented a new and novel approach to develop CADx scheme based on the global image feature analysis to predict the likelihood of case being malignant once the suspicious lesions are detected by the radiologists and demonstrated feasibility of this new approach, which may create a new opportunity for researchers in the CAD-related medical imaging informatics field to develop and optimize new computer-aided decision-making supporting tools for future clinical applications.

Acknowledgments

References
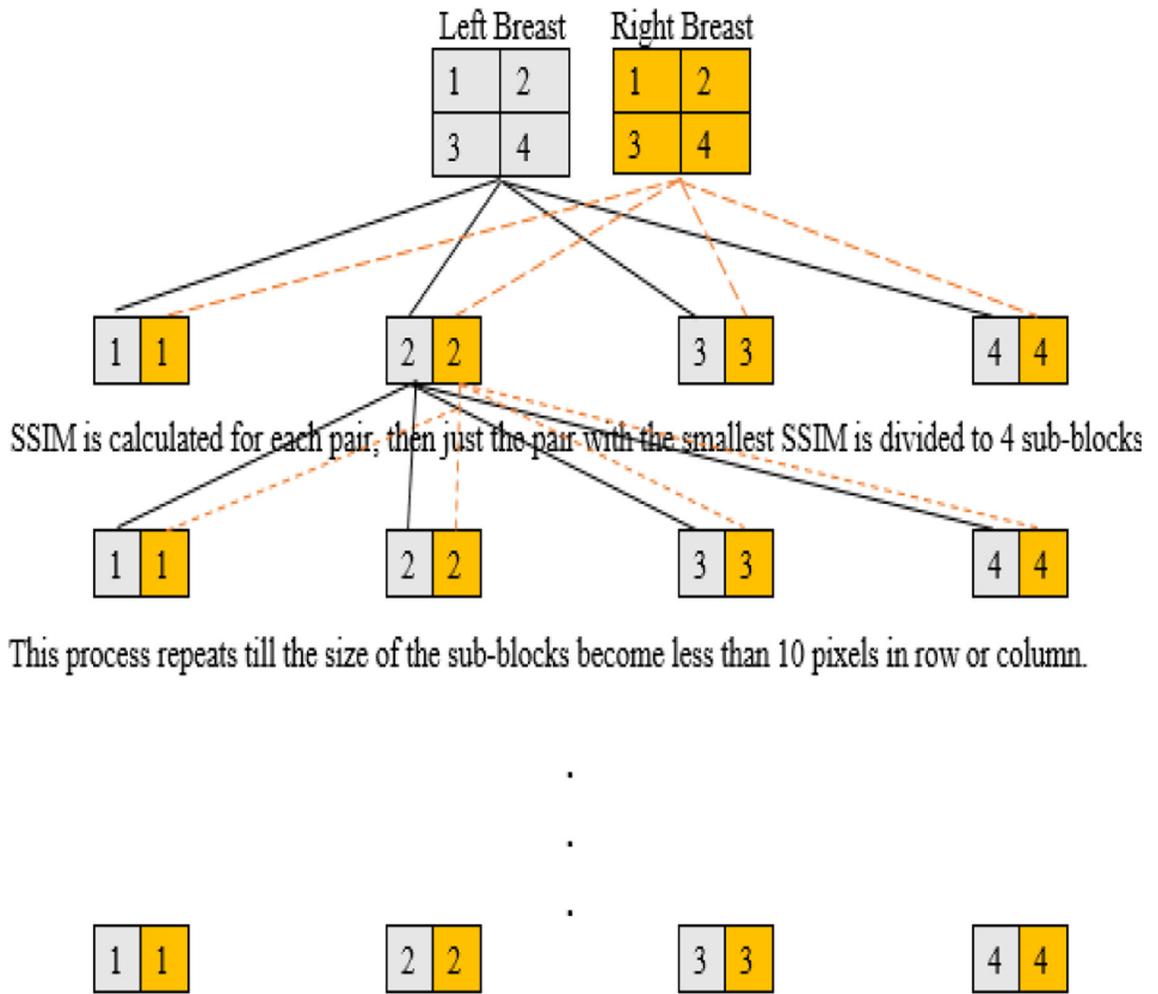
[1]. Baron R, Drucker K, Lagdamen L, et al., "Breast cancer screening: A review of current guidelines," Am J Nurs., vol. 118, no. 7, pp. 34–41, 7 2018.

[2]. Kelly KM, Dean J, Comulada WS, Lee S, "Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts," Eur Radiol., vol. 20, no. 3, pp. 734–742, 3 2010. [PubMed: 19727744]

[3]. Woodard DB, Gelfand AE, Barlow WE, Elmore JG, "Performance assessment for radiologists interpreting screening mammography," Stat. Med vol. 26, no. 7, pp. 1532–1551, 3 2007. [PubMed: 16847870]

[4]. Nishikawa RM, Gur D, "CADe for early detection of breast cancer – current status and why we need to continue to explore new approaches," Acad Radiol., vol. 21, no. 10, pp. 1320–1321, 10 2014. [PubMed: 25086951]

[5]. Huo Z, Giger ML, Vyborny CJ, Metz CE, "Breast cancer: Effectiveness of computer-aided diagnosis – observer study with independent database of mammograms," Radiology, Vol. 224, no. 2, pp. 560–568, 8 2002. [PubMed: 12147857]

[6]. Fenton JJ, Taplin SH, Carney PA, et al., "Influence of computer-aided detection on performance of screening mammography," N Engl J Med, vol. 356, no. 14, pp. 1399–1409, 4 2007. [PubMed: 17409321]

[7]. 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. vol. 155, no. 8, pp. 481–492, 10 2011. [PubMed: 22007042]

[8]. Brodersen J, Siersma VD, "Long-term psychosocial consequences of false-positive screening mammography," Ann Fam Med. vol. 11, no. 2, pp. 106–15, 3 2013. [PubMed: 23508596]

[9]. Yaffe MJ, Mainprize JG, "Risk of radiation-induced breast cancer from mammographic screening," Radiology, vol. 258, no. 1, pp. 98–105, 1 2011. [PubMed: 21081671]

[10]. Buist DS, Anderson ML, Haneuse SJ, et al., "Influence of annual interpretive volume on screening mammography performance in the United States," Radiology, vol. 259, no. 1, pp. 72–84, 4 2011. [PubMed: 21343539]

[11]. Doi K, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," Comput Med Imaging Graph., vol. 31, no. 4, pp. 198–211, 3 2007. [PubMed: 17349778]

[12]. Wang Y, Aghaei F, Zarafshani A, et al., "Computer-aided classification of mammographic masses using visually sensitive image features," J Xray Sci Technol., vol. 25, no. 1, pp. 171–186, 1 2017. [PubMed: 27911353]

[13]. Danala G, Patel B, Aghaei F, et al., "Classification of breast masses using a computer-aided diagnosis scheme of contrast enhanced digital mammograms," Ann Biomed Eng., vol. 46, no. 9, pp. 1419–1431, 9 2018. [PubMed: 29748869]

[14]. Qiu Y, Yan S, Gundreddy R, et al., "A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology," J Xray Sci Technol., vol. 25, no. 5, pp. 751–763, 10 2017. [PubMed: 28436410]

[15]. Gao F, Wu T, Li J, et al., "SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis," Comput Med Imaging Graphs., vol. 70, pp. 53–62, 12 2018.

[16]. Tan M, Zheng B, Leader JK, Gur D, "Association between changes in mammographic image features and risk for near-term breast cancer development," IEEE Trans Med Imaging., vol. 35, no. 7, pp. 1719–1728, 7 2016. [PubMed: 26886970]

[17]. Heidari M, Khuzani A, Hollingsworth AB, et al., "Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm," Phys Med Biol., vol. 63, no. 3, pp. 035020, 1 2018. [PubMed: 29239858]

[18]. Tan M, Qian W, Pu J, et al., "A new approach to develop computer-aided detection schemes of digital mammograms," Phys Med Biol., vol. 60, no. 11, pp. 4413–4427, 6 2015. [PubMed: 25984710]

[19]. Aghaei F, Tan M, Hollingsworth AB, Zheng B, "Applying a new quantitative global breast MRI feature analysis scheme to assess tumor response to chemotherapy," J Magn Reson Imaging., vol. 44, no. 5, pp. 1099–1106, 4 2016. [PubMed: 27080203]

[20]. Wang Y, Thai T, Moore K, et al., "Quantitative measurement of adiposity using CT images to predict the benefit of bevacizumab-based chemotherapy in epithelial ovarian cancer patients," Oncol Lett., vol. 12, no. 1, pp. 680–686, 7 2016. [PubMed: 27347200]
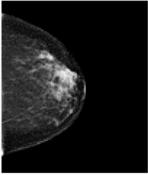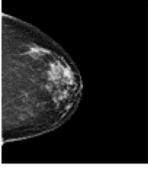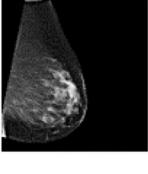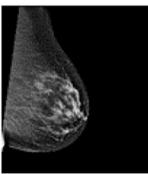
[21]. Zheng B, Leader JK, Abrams GS, et al. "Multiview based computer-aided detection scheme for breast masses," Med Phys., Vol. 33, no. 9, pp. 3135–3143, 9 2006. [PubMed: 17022205]

[22]. Zheng B, Sumkin JH, Zuley M, et al., "Computer-aided detection of breast masses depicted on full-field digital mammograms: a performance assessment," Br J Radiol., vol. 85, no. 1014, pp. e153–161, 6 2012. [PubMed: 21343322]

[23]. Wang Z, Bovik AC, Sheikh HR, et al., "Image quality assessment from error visibility to structural similarity," IEEE Trans Image Process., vol. 13, no. 4, pp. 600–612, 4 2004. [PubMed: 15376593]

[24]. Sampat MP, Zhou Wang S Gupta, et al., "Complex wavelet structural similarity: a new image similarity index," IEEE Trans Image Process., vol. 18, no. 11, pp. 2385–2401, 6 2009. [PubMed: 19556195]

[25]. Rao KR and Yip PC, Discrete Cosine Transform : Algorithms, Advantages, Applications, Academic Press, San Diego, CA, USA, 1990.

[26]. Burrus CS, Fast Fourier Transforms, Google Books Online: http://cnx.org/content/col10550/1.22/.

[27]. Fan X, Lu Y, and Gao W, "A novel coefficient scanning scheme for directional spatial prediction-based image compression," Proc. IEEE Int. Conf. Multimed. Expo, vol. 2, pp. II557–II560, 8 2003.

[28]. Rouhi R and Jafari M, "Classification of benign and malignant breast tumors based on hybrid level set segmentation," Expert Syst. Appl vol. 46, pp. 45–59, 3 2016.

[29]. Yin K, Yan S, Song C, Zheng B, "A robust method for segmenting pectoral muscle in mediolateral oblique (MLO) mammograms," Int J Comput Assist Radiol Surg., vol. 14, no. 2, pp. 237–248, 2 2019. [PubMed: 30288698]

[30]. Revaud J, Weinzaepfel P, Harchaoui Z, et al. "DeepMatching: Hierarchicl deformable dense matching," Int J Comput Vis., vol. 120, no. 3, pp. 300–323, 12 2016.

[31]. Wang Y and Moulin P, "Optimized feature extraction for learning-based image steganalysis," IEEE Trans. Inf. Forensics Secur vol. 2, no. 1, pp. 31–45, 2 2007.

[32]. Lederman D, Zheng B, Wang X, Gur DD, "Improving breast cancer risk stratification using resonance-frequency electrical impedance spectroscopy through fusion of multiple classifiers," Ann Biomed Eng., vol. 39, no. 3, pp. 931–945, 3 2011. [PubMed: 21116847]

[33]. Tan M, Pu J, Zheng B, "Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector machine (SVM) model," Int J Comput Assist Radiol Surg., vol. 9, no. 6, pp. 1005–1020, 11 2014. [PubMed: 24664267]

[34]. Khuzani A, Du Y, Heidari M, et al., "Prediction of chemotherapy response in ovarian cancer patients using a new clustered quantitative image marker," Phys Med Biol., vol. 63, no. 15, pp. 155020, 8 2018. [PubMed: 30010611]

[35]. Tan M, Pu J, Cheng S, et al., "Assessment of a four-view mammographic image feature based fusion model to predict near-term breast cancer risk," Ann Biomed Eng., vol. 43, no. 10, pp. 2416–2428, 10 2015. [PubMed: 25851469]

[36]. Mirniaharikandehei S, Hollingsworth AB, Patel B, et al., "Applying a new computer-aided detection scheme generated imaging marker to predict short-term breast cancer risk," Phys Med Biol., vol. 63, no. 10, pp. 105005, 5 2018. [PubMed: 29667606]
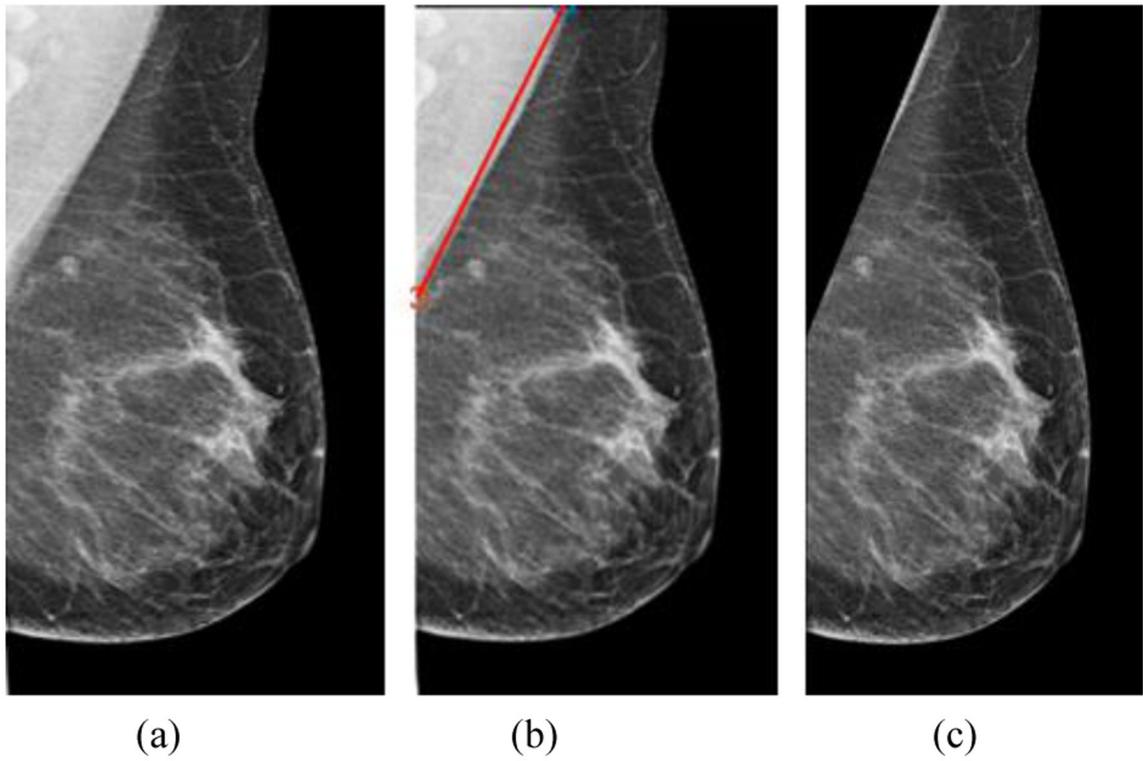
Fig.1.
Preprocessing phase. a) the original image, b) chest wall removal step, c) denoising black area, and written labels removal.

Left Breast   Right Breast

| 1 | 2 | | 1 | 2 |
| 3 | 4 | | 3 | 4 |

| 1 | 1 | | 2 | 2 | | 3 | 3 | | 4 | 4 |

SSIM is calculated for each pair, then just the pair with the smallest SSIM is divided to 4 sub-blocks

| 1 | 1 | | 2 | 2 | | 3 | 3 | | 4 | 4 |

This process repeats till the size of the sub-blocks become less than 10 pixels in row or column.

.

.

.

| 1 | 1 | | 2 | 2 | | 3 | 3 | | 4 | 4 |

Fig. 2.
Block diagram of the proposed method for SSIM feature extraction.

|			
LCC	$lccF_{dct}$ $= [f_{1c}, f_{2c}, \dots, f_{14c}]$ $lccF_{fft}$ $= [f_{1c}, f_{2c}, \dots, f_{14c}]$	$F_{dct}cc$ $= abs(lccF_{dct}$ $- rccF_{dct})$	
RCC	$F_{ssim}cc$ $= [f_{1c}, f_{2c}, \dots, f_{6c}]$ $rccF_{dct}$ $= [f_{1c}', f_{2c}', \dots, f_{14c}']$ $rccF_{fft}$ $= [f_{1c}', f_{2c}', \dots, f_{14c}']$	$F_{fft}cc$ $= abs(lccF_{fft}$ $- rccF_{fft})$ $F_{ssim}cc$ $= [f_{1c}, f_{2c}, \dots, f_{6c}]$	$F_{fusion} = [F_{dct}cc, F_{dct}ml, F_{fft}cc, F_{fft}ml, F_{ssim}cc, F_{ssim}ml]$
LML	$lmlF_{dct}$ $= [f_{1m}, f_{2m}, \dots, f_{14m}]$ $lmlF_{fft}$ $= [f_{1m}, f_{2m}, \dots, f_{14m}]$	$F_{dct}ml$ $= abs(lmlF_{dct}$ $- rmlF_{dct})$	
RML	$F_{ssim}ml$ $= [f_{1m}, f_{2m}, \dots, f_{6m}]$ $rmlF_{dct}$ $= [f_{1m}', f_{2m}', \dots, f_{14m}']$ $rmlF_{fft}$ $= [f_{1m}', f_{2m}', \dots, f_{14m}']$	$F_{fft}ml$ $= abs(lmlF_{fft}$ $- rmlF_{fft})$ $F_{ssim}ml$ $= [f_{1m}, f_{2m}, \dots, f_{6m}]$	

Fig. 3.
Feature extraction phase of the proposed method.
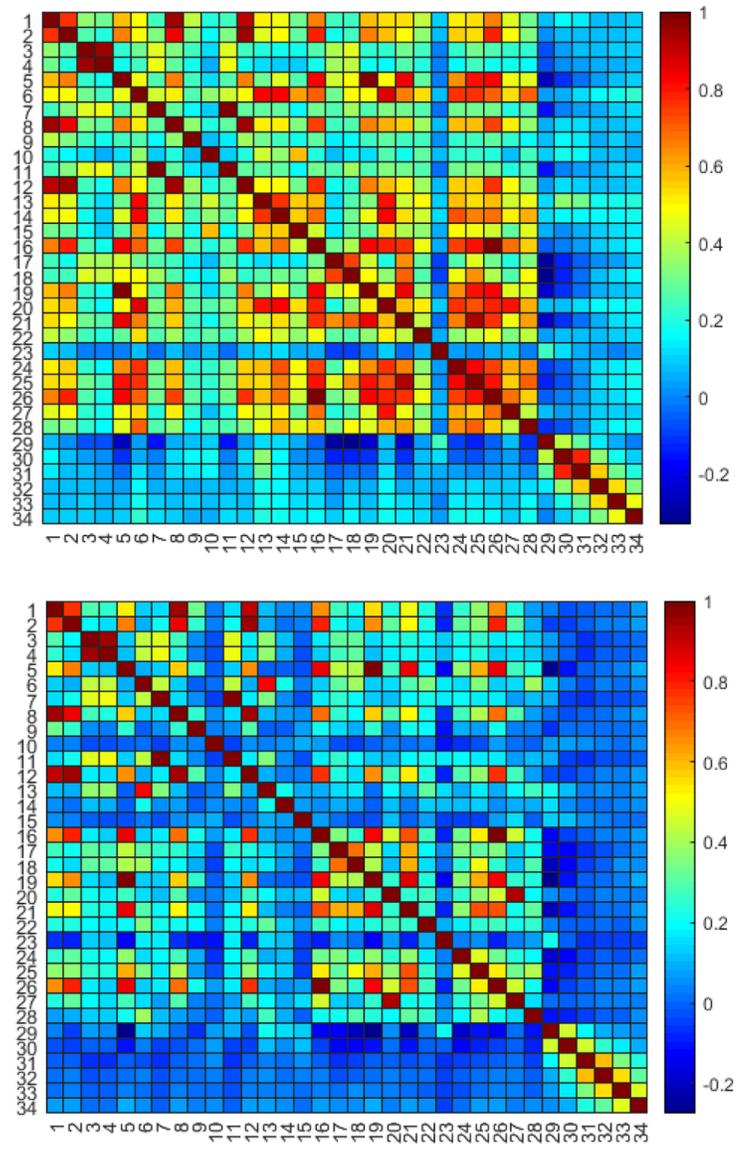
Fig. 4.
Correlation coefficient matrices of 34 image features computed from CC (top) and MLO (bottom) view images.
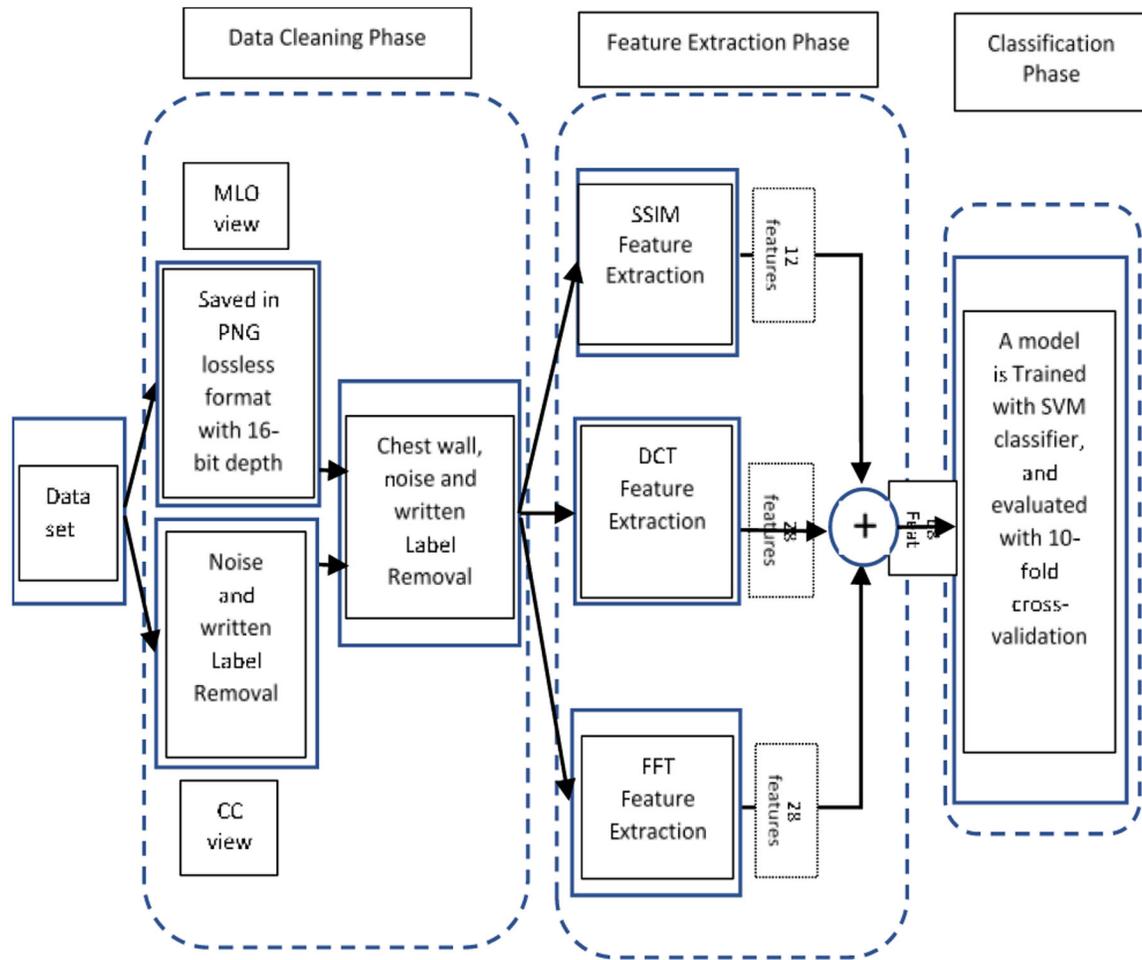
Fig. 5.
A summarized block diagram of the proposed scheme for classification of benign and malignant tissues in mammography imaging.
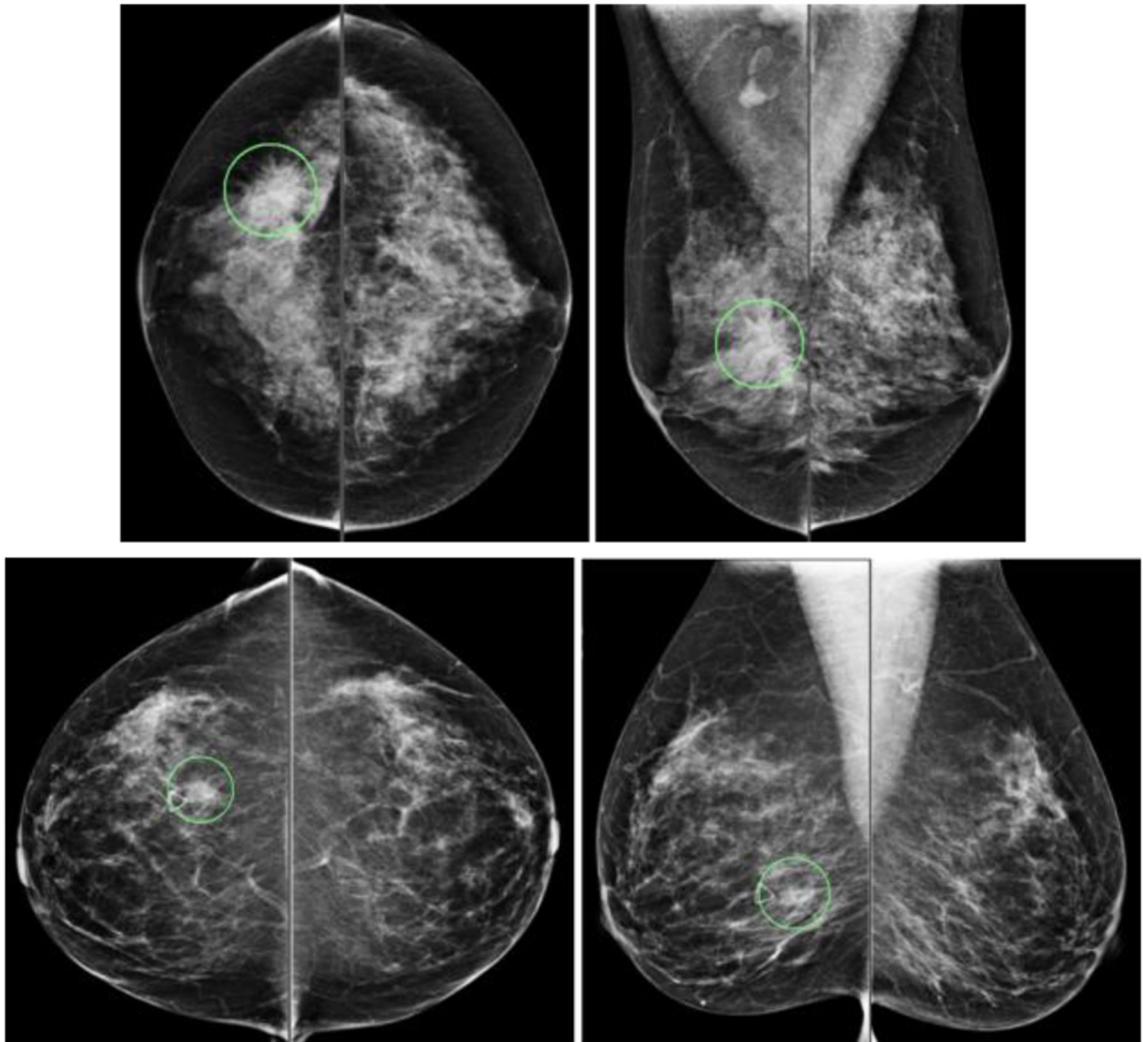
Fig. 6.
Illustration of one malignant case (the first row) and one benign case (the second row). The detected masses are circled (Green Color) in the images.
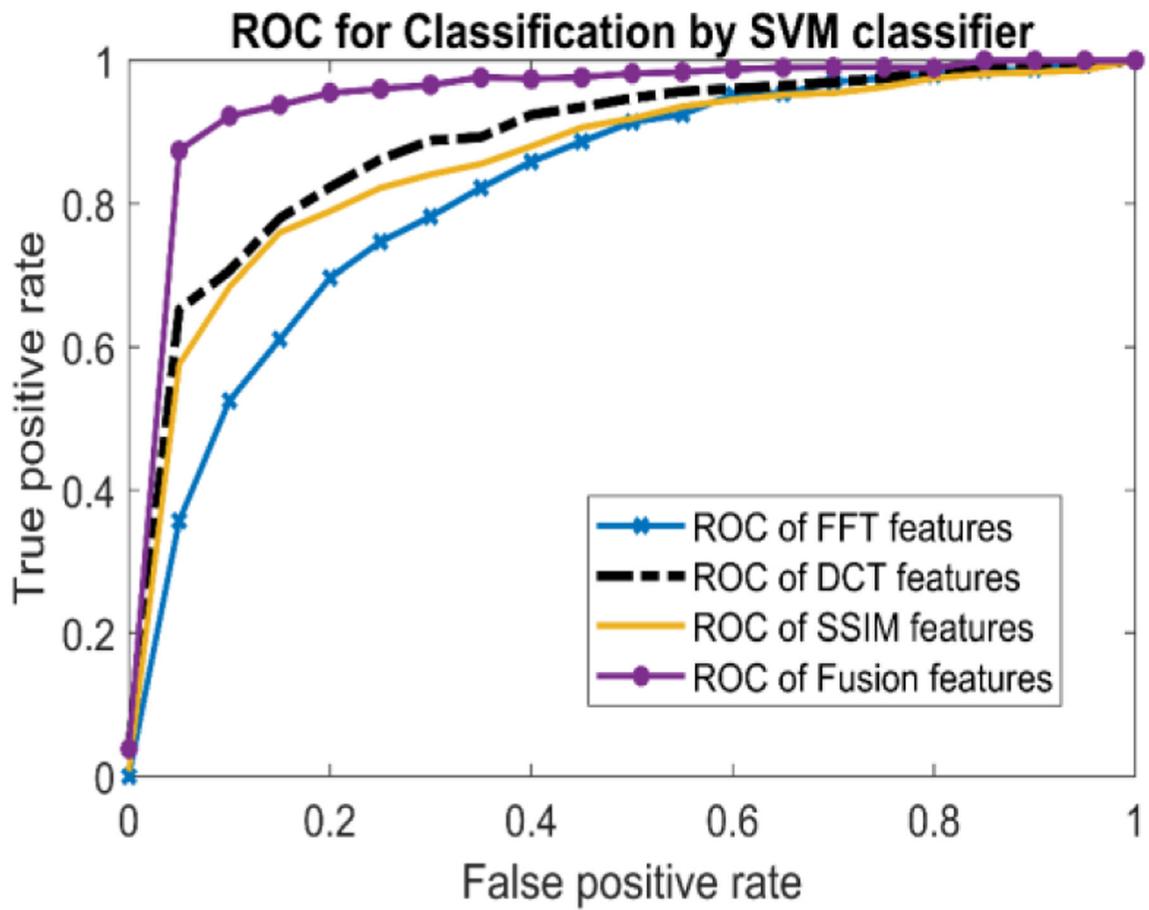
Fig. 7.
Comparison of 4 ROC curves generated by four SVMs trained using image features computed from both CC and MLO view images.

TABLE I

Case number and percentage distribution of patients age and mammographic density rated by radiologists using BIRADS guidelines.

	Subgroup	Malignant Cases	Benign Cases
Density BIRADS	1	39 (5.3%)	58 (4.7%)
	2	286 (38.8%)	412 (33.7%)
	3	401 (54.4%)	702 (57.4%)
p-value = 0.878	4	11 (1.5%)	50 (4.1%)
Age of Patients (years old)	A < 40	25 (3.4%)	50 (4.1%)
	40 ≤ A < 50	141 (19.2%)	561 (45.9%)
	50 ≤ A < 60	189 (25.6%)	335 (27.4%)
	60 ≤ A < 70	180 (24.4%)	187 (15.3%)
	70 ≤ A	202 (27.4%)	89 (7.3%)

Table II

The computed SSIM, DCT and FFT image Features

Feature category	Feature Description
SSIM features computed from original FFDM images	Six SSIM indices computed using Equation (1) from the six pairs of sub-blocks with the gradually reduced size.
Features computed from frequency domain of DCT and FFT transformed maps	1.Mean, 2. variance, 3. skewness, 4. kurtosis, 5. entropy, 6. correlation, 7. energy, 8. root mean square level, 9. uniformity, 10. max, 11. min, 12. median, 13. range, 14. mean absolute deviation

TABLE III

AUC and Accuracy for different sub-group of features on CC view in compare with MLO view.

Feature sub-groups	Number of features	AUC ± STD	Accuracy (%)
FFT, CC view	14	0.63 ± 0.025	66
FFT, MLO view	14	0.84 ± 0.017	77
DCT, CC view	14	0.62 ± 0.026	64
DCT, MLO view	14	0.89 ± 0.015	83
SSIM, CC view	6	0.53 ± 0.026	63
SSIM, MLO view	6	0.78 ± 0.021	71
Fusion, CC view	34	0.65 ± 0.027	67
Fusion, MLO view	34	0.94 ± 0.009	89

TABLE IV

Computed area under ROC curve using individual group of features on both CC and MLO views.

Feature sub-group	Num. of features	AUC	STD	95% CI
FFT features	28	0.85	0.018	[0.80, 0.90]
DCT features	28	0.91	0.013	[0.89, 0.95]
SSIM features	12	0.89	0.016	[0.85, 0.92]
Fusion of all features	68	0.96	0.007	[0.95, 0.97]

TABLE V

Four confusion matrices generated using 4 SVMs trained using features computed from both CC and MLO view.

Feature Group	Predicted	Actual Positive	Actual Negative
SSIM	Positive	528	106
	Negative	209	1116
DCT	Positive	501	54
	Negative	1168	236
FFT	Positive	437	135
	Negative	300	1087
Fusion	Positive	**656**	61
	Negative	81	**1161**

TABLE VI

Accuracy, sensitivity, specificity, and odd ratio of using 4 SVMs trained using different features computed from both CC and MLO views

Feature sub-group	Accuracy (%)	Sensitivity (%)	Specificity (%)	Odds Ratio
FFT features	78	59	89	11.47
DCT features	85	68	96	52
SSIM features	83	71	90	24
Fusion	**92**	**89**	**95**	**154**

TABLE VII

Adjusted odds ratios (ORs) and 95% confidence intervals (CIs) at five subgroups with increasing values of SVM-generated prediction scores.

Subgroup (bin)	Number of Cases (Positive/Negative)	Adjusted OR	95 % CI
1	2–390	1.00	Reference
2	9–383	4.58	0.93–21.34
3	41–351	22.78	5.47–94.85
4	297–95	609.6	149.05–2493.35
5	388–3	25220	4190.90–151766.40

TABLE VIII

Comparison of prediction performance of SVMs trained with and without removal of pectoral muscle areas in MLO view images.

Feature sub-group	Number of features	AUC ± STD	Accuracy (%)
FFT without chest removal	14	0.70 ± 0.021	72
FFT with chest removal	14	0.84 ± 0.017	77
DCT without chest removal	14	0.68 ± 0.022	69
DCT with chest removal	14	0.89 ± 0.015	83
SSIM without chest removal	6	0.62 ± 0.026	61
SSIM with chest removal	6	0.78 ± 0.021	71
Fusion without chest removal	34	0.79 ± 0.017	83
Fusion with chest removal	34	0.94 ± 0.009	89