Computer aided detection of clusters of microcalcifications on full field digital mammograms

Jun Ge,[a] Berkman Sahiner, Lubomir M. Hadjiiski, Heang-Ping Chan, Jun Wei, Mark A. Helvie, and Chuan Zhou
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109-0904

(Received 28 December 2005; revised 27 March 2006; accepted for publication 16 May 2006; published 27 July 2006)

We are developing a computer-aided detection (CAD) system to identify microcalcification clusters (MCCs) automatically on full field digital mammograms (FFDMs). The CAD system includes six stages: preprocessing; image enhancement; segmentation of microcalcification candidates; false positive (FP) reduction for individual microcalcifications; regional clustering; and FP reduction for clustered microcalcifications. At the stage of FP reduction for individual microcalcifications, a truncated sum-of-squares error function was used to improve the efficiency and robustness of the training of an artificial neural network in our CAD system for FFDMs. At the stage of FP reduction for clustered microcalcifications, morphological features and features derived from the artificial neural network outputs were extracted from each cluster. Stepwise linear discriminant analysis (LDA) was used to select the features. An LDA classifier was then used to differentiate clustered microcalcifications from FPs. A data set of 96 cases with 192 images was collected at the University of Michigan. This data set contained 96 MCCs, of which 28 clusters were proven by biopsy to be malignant and 68 were proven to be benign. The data set was separated into two independent data sets for training and testing of the CAD system in a cross-validation scheme. When one data set was used to train and validate the convolution neural network (CNN) in our CAD system, the other data set was used to evaluate the detection performance. With the use of a truncated error metric, the training of CNN could be accelerated and the classification performance was improved. The CNN in combination with an LDA classifier could substantially reduce FPs with a small tradeoff in sensitivity. By using the free-response receiver operating characteristic methodology, it was found that our CAD system can achieve a cluster-based sensitivity of 70, 80, and 90 % at 0.21, 0.61, and 1.49 FPs/image, respectively. For case-based performance evaluation, a sensitivity of 70, 80, and 90 % can be achieved at 0.07, 0.17, and 0.65 FPs/image, respectively. We also used a data set of 216 mammograms negative for clustered microcalcifications to further estimate the FP rate of our CAD system. The corresponding FP rates were 0.15, 0.31, and 0.86 FPs/image for cluster-based detection when negative mammograms were used for estimation of FP rates. [DOI: 10.1118/1.2211710]

Key words: computer-aided detection (CAD), full-field digital mammography (FFDM), artificial neural network

I. INTRODUCTION

Breast cancer is the most frequently diagnosed cancer and ranks second among cancer deaths in women. An estimated 211 240 new cases of invasive breast cancer and an estimated 40 410 breast cancer deaths are expected to occur among women in the United States during 2005.[1] Studies indicate that detection and treatment at an early stage can improve the survival rate of women with breast cancer.[2–4] Mammography is the most effective method to date for the detection of breast cancer. However, it has been reported that a substantial fraction of breast cancers which are visible upon retrospective analyses of the images are missed initially.[5–7] The use of a computer-aided detection (CAD) system as an objective "second reader" is considered to be one of the promising approaches that may help radiologists improve the sensitivity of mammography. The majority of studies to date have shown that CAD can improve radiologists' detection accuracy without substantially increasing the recall rates.[8–13] Since breast imaging specialists detect more cancers and more early-stage cancers, and have lower recall rates than general radiologists,[14] the value of CAD may vary among readers.[15,16]

Microcalcifications account for over 50% of all the nonpalpable lesions detected using mammography.[17] Most microcalcifications represent benign conditions, but approximately 20 to 30 % of microcalcification clusters (MCCs) that are biopsied when no palpable mass is present prove to be malignant.[18] A number of CAD algorithms have been developed for automated detection of microcalcifications on mammograms. These algorithms can be approximately grouped into two categories: (1) sequential stepwise approaches[19–23] and (2) model-based iterative approaches.[24] In sequential stepwise approaches, a prescreening step is usually performed to select microcalcification candidates at a high-sensitivity level. In the subsequent steps, increasingly strict criteria are used to reduce the number of the false positives

(FPs) to an acceptable level. In a model-based iterative approach,[24] one of four labels (background, dot structure, line structure, and microcalcification) is assigned to each pixel. A random field model is used to model the spatial context. The Bayesian method is then used to iteratively update the labeling.

Most of these mammographic CAD algorithms were developed specifically for digitized screen-film mammograms (SFMs). In the last few years, several full-field digital mammography (FFDM) manufacturers have obtained clearance from the Food and Drug Administration (FDA) for clinical use. Recently, a large clinical trial[25] found that the overall accuracy of breast cancer detection by FFDM and by SFM were comparable in the general patient population. However, FFDM provided significantly higher accuracy in patients with dense breasts. For CAD, FFDMs may provide the advantages of having higher signal-to-noise ratio (SNR) and detective quantum efficiency (DQE), wider dynamic range, and higher contrast sensitivity than SFMs. Moreover, the detection results of CAD systems for SFM have been found to be inconsistent for repeated film digitization.[26] The reproducibility of CAD systems can be improved with FFDM because the film digitization step is eliminated. Commercial CAD systems have been modified to be used with FFDMs. One industry-sponsored study[27] reported that a commercial CAD system detected 97% of malignant clustered microcalcifications with 0.55 FP marks/image for FFDMs. This result is similar to the reported FP rate of 0.5 marks/image for SFMs by the same manufacturer. Another industry-sponsored study[28] reported that a CAD system achieved 100% sensitivity at about 2 FP marks/case (four standard views) in detecting nine malignant MCCs on FFDMs. McLoughlin *et al.*[29] reported a cluster detection sensitivity of 90% at about 1 FP marks/image for a data set of 124 FFDMs containing 28 MCCs (either benign or malignant) for a noncommercial system.

We have previously developed a CAD system for the detection of MCCs on digitized SFMs.[19,22,23] We are developing a CAD system for mammograms acquired by an FFDM system. The methodology used for detecting MCCs on digitized mammograms was adapted to FFDMs, and the system parameters were retrained at stages that are sensitive to image noise. The difference-image technique based on a box-rim filter was used at the prescreening stage. At the FP reduction stage, a truncated sum-of-squares error function was used to improve the efficiency and robustness of the training of a convolution neural network (CNN) in our CAD system for FFDMs. The effects of the FP reduction technique on the sensitivity and the FP rate of our system were examined. To evaluate the effects of preprocessing on microcalcification detection, we also compared the performance of our CAD system using GE-processed images as the input and using the raw images as the input.

II. MATERIALS AND METHOD

A. Data sets

Institutional Review Board (IRB) approval was obtained to collect the mammograms in the Department of Radiology at the University of Michigan. The mammograms in this study were acquired with a GE Senographe 2000D FFDM system. The GE system has a CsI phosphor/a:Si active matrix flat panel digital detector with a pixel size of 100 μm $\times$ 100 μm and the raw images were acquired at 14 bits per pixel. The data set contained 96 cases with 192 images. All cases had two mammographic views: the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral (LM or ML) view. The mammogram was assessed by a Mammography Quality Standards Act (MQSA) radiologist and a polygon was drawn to enclose each MCC. The radiologist marked the clusters as c0, c1, c2, based on the degree of concern. There were 96 c0 clusters in the data set, of which 28 were proven by biopsy to be malignant and 68 were proven to be benign. In this study, we concentrated on the detection rather than the classification of the malignant/benign nature of the MCCs so that both malignant and benign microcalcifications were considered to be positive cases. There were 8 c1 and 1 c2 marks that were not biopsied or followed up and they were not counted as true positives (TPs) or FPs in the evaluation. The distribution of the sizes (in mm) for the c0 clusters, estimated as its longest dimension of the bounding polygon, in our data set is shown in Fig. 1(a).

The GE-processed image was displayed on a workstation at full resolution and the coordinates of individual microcalcifications in the image were manually identified. The graphical user interface allowed windowing and zooming of the displayed image to facilitate viewing, and a cursor was available to mark the locations of individual microcalcifications. A total of 2127 microcalcifications were marked for the biopsied clusters. The histogram of the number of manually identified microcalcifications per cluster is shown in Fig. 1(b). The number of microcalcifications per cluster ranged from 3 to 91, with a mean of 8.97 and a standard deviation of 3.15. Since the clusters with a large number of microcalcifications may cause large deviation of the estimated statistics from the true statistics, we excluded the nine clusters with greater than or equal to 30 microcalcifications in the calculation of the mean and standard deviation of microcalcifications per cluster.

The data set of 192 images was separated into two independent, equal-sized subsets with the malignant cases equally distributed. Each subset contained 48 cases with 96 images, of which 14 cases were malignant. Figure 1(b) shows the histograms of the number of manually identified microcalcifications per cluster for the two subsets. Twofold cross validation was chosen for the training and testing of our CAD system. Once the training with one subset was completed, the parameters and all thresholds were fixed for testing with the other subset. The training and test subsets

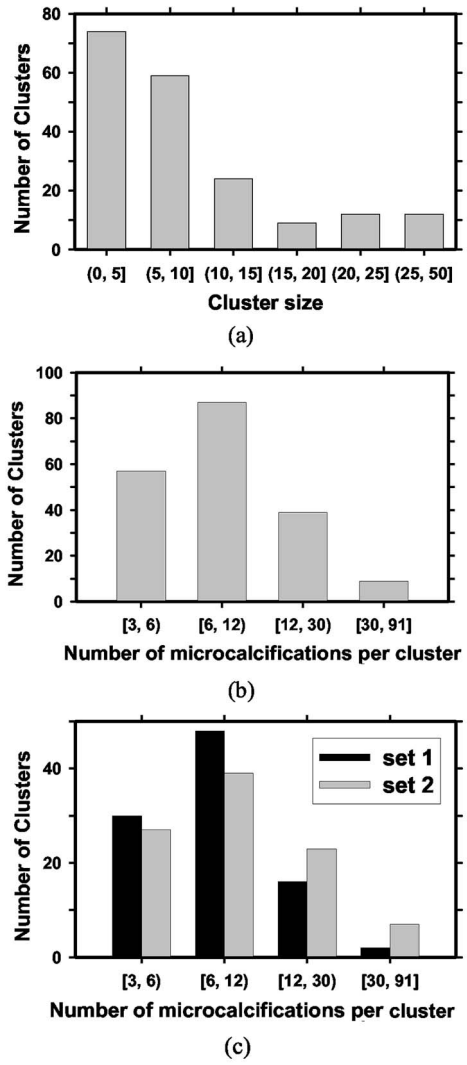

FIG. 1. (a) The distribution of the longest dimension of the clusters for the data set of 192 images. Histogram of the number of microcalcifications per cluster: (b) for the data set of 192 images and (c) for subsets 1 and 2.

were switched and the training process was repeated. The overall detection performance was evaluated by averaging the performances for the two test subsets.

Another data set of 108 cases with 216 FFDM images was collected. These mammograms were negative for microcalcifications such that no clustered microcalcifications were found based on review by experienced breast radiologists although they may contain other mammographic abnormalities such as soft tissue masses. This negative data set was used to evaluate the FP cluster detection rate by our CAD systems.

B. Methods

The computer vision techniques used for detecting MCCs on digitized mammograms in our previous study was adapted to FFDMs. The CAD system includes six stages: (1) preprocessing; (2) image enhancement; (3) segmentation of individual microcalcification candidates; (4) FP reduction for individual microcalcifications using rule-based classifiers and a CNN; (5) regional clustering of microcalcifications; and (6) FP reduction for clustered microcalcifications using stepwise LDA feature selection and classifier. The block diagram of our CAD system is shown in Fig. 2. Details of each stage will be described in this section.

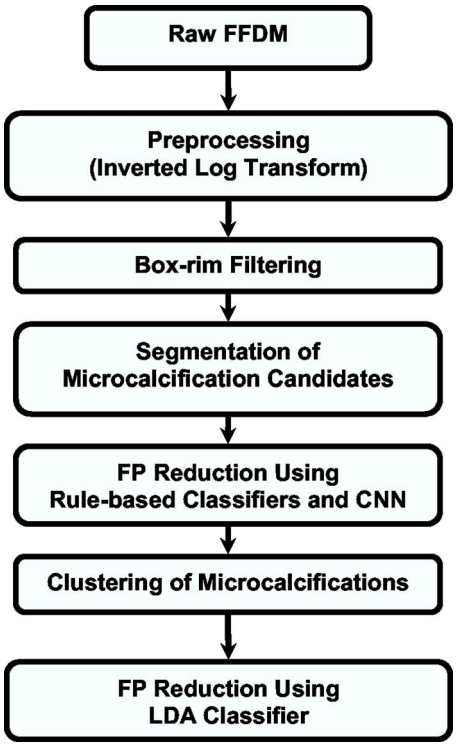

FIG. 2. The block diagram of our CAD system for detection of microcalcification clusters on FFDMs.

1. Preprocessing

FFDMs are generally preprocessed with proprietary methods by the manufacturer of the FFDM system before being displayed to radiologists in clinical practice. The image preprocessing method used depends on the manufacturer of the FFDM system. It has been reported that radiologists prefer images from different FFDM manufacturers for different mammographic reading tasks and for different lesion types on FFDMs.[30] A CAD system is also likely to be sensitive to preprocessing methods to some degree. To develop a CAD system which is less dependent on the FFDM manufacturer's proprietary preprocessing methods, we use the raw FFDM as input to our CAD system. The input raw image is first subjected to boundary segmentation using a two-step algorithm. First, Otsu's method[31] is used to calculate a threshold and binarize the original image. Second, an eight-connectivity labeling method is used to identify the connected regions below the threshold on the binary image. The connected region with the largest area will be considered to be the breast region. Any area external to the breast region is trimmed. Further steps are only applied to the segmented breast area to reduce computation time.

Clinical mammograms are usually viewed in an inverted mode of the raw images, namely, the background of the dis-

played image is dark. In our data set, all clinical mammograms have been preprocessed by GE's proprietary procedure. The pixel depth of clinical mammograms is 12 bits and the background pixel values are zero. In order to process an image in our CAD system with the same format as the clinical mammograms, we applied an inverted logarithmic transformation[32] to the raw pixel values. The transformation function is shown in Fig. 3(a). Figure 3(b) shows a typical raw image and Fig. 3(c) shows its gray-level histogram. The transformed image and its gray-level histogram are shown in Figs. 3(d) and 3(e), respectively. The same transformation was applied to the whole image, but the slope is inversely proportional to the local raw pixel values. The denser (lower raw pixel value) the breast area, the farther the pixel values are stretched by the transformation, leading to a larger contrast enhancement on the transformed image. Thus, an adaptive contrast enhancement function that depends on the local density has been incorporated in this transformation implicitly.

2. Image enhancement

Digital image enhancement techniques have been widely used for enhancement of mammographic images.[19,22,33–36] Specifically, enhancement of the contrast of the mammographic structures of interest is the primary concern. One of the approaches is to apply a nonlinear enhancement function to the spatial frequency components corresponding to the mammographic structures of interest. Then the modified frequency contents are used to reconstruct the image in order to achieve enhancement of the lesions. In the enhancement methods by Stahl *et al.*[36] and Wei *et al.*,[34] a nonlinear multiscale enhancement method based on hierarchically repeated unsharp masking was used to enhance weakly contrasting structures at multiple scales. In their methods, the detail coefficients in the transform domain represent the frequency bands corresponding to the mammographic structures of interest.

Chan *et al.*[19,22] used a difference-image technique to enhance the signal-to-noise ratio (SNR) of the microcalcifications. A signal-enhancement filter is used to enhance the signal and smooth the random noise. A signal-suppression filter is used to remove the signal and again smooth the random noise. The two filtered images are then subtracted to produce a difference image in which the low-frequency structured background is removed and the high-frequency noise is suppressed. The two filters can be combined to a bandpass filter when they are both linear. With appropriate adjustment of the filter kernel, the bandpass filter will enhance the frequency contents of microcalcifications. In this study, we implement the difference-image technique as a 8×8 box-rim filter. The size of the box-rim filter was selected for mammograms of 100 μm pixel size[19,22] such that the spectrum peaks at around 0.2 cycles/pixel. Thus the signals whose diameters are about five pixels will be enhanced.

3. Segmentation of individual microcalcifications

In this step, our goal is to segment the individual microcalcification candidates (signals). We classify the pixels in the breast region into two classes: c_1, pixels associated with microcalcifications, and c_2, pixels which are not associated with microcalcifications. The histogram of the difference image is determined and the gray-level values of the pixels in the box-rim filtered difference image are used as the feature to classify the pixels into c_1 and c_2. An image labeling technique is used to extract the signal. The procedure is performed iteratively until the number of signals is within a predefined range. For our CAD system for SFMs, this range was chosen to be about 3000 to 4000. For FFDMs, it was found by training that the range should be reduced to between 400 and 500. This may be attributed to the fact that the FFDM difference images are less noisy than the digitized SFM difference images, which was observed by comparing the average root-mean-square (rms) noise in the background regions of the two types of difference images. One major noise component in the digitized SFMs may be contributed by the digitization process.

The individual microcalcifications have higher SNR, on average, than the background in their local area. A locally adaptive gray-level thresholding method is used to refine the signal candidates. The local rms noise within a square kernel centered at the signal candidate location is estimated. The central pixels of the kernel that contain the signal candidate are excluded from the rms noise estimation. A pixel is retained only if its value is larger than the mean pixel value by a predefined multiple k of the rms noise, where k is the SNR threshold. From our previous study for SFMs,[19,22] the kernel size must be sufficiently large to give a good estimate of the local background noise fluctuation and was chosen to be 51×51 pixels.

4. False positive reduction for individual microcalcifications

In the FP reduction stage, the microcalcification candidates are classified as either TP or FP using a combination of rule-based feature classification and a trained CNN classifier. Two features, namely, the area and the contrast of the microcalcification candidate, are first used to exclude small-area signals that are likely to be noise and high-contrast signals that are likely to be artifacts or large benign calcifications. The area is calculated as the number of pixels segmented for the candidate. The contrast is calculated as the peak pixel value within the segmented signal region above the average pixel value of the local background. The microcalcification candidate is classified as FP if the area is less than three pixels or the contrast is ten times higher than the background rms noise to exclude large benign calcifications or artifacts.

We used the optimal architecture of CNN selected in our previous study,[23] but retrained the weights connecting the nodes within the CNN for FFDMs. The CNN had one input node group for a 16×16-pixel (1.6×1.6 mm) region of interest (ROI) centered at a signal candidate, two hidden layers, and one output node (a score between 0 and 1).[23] All 14

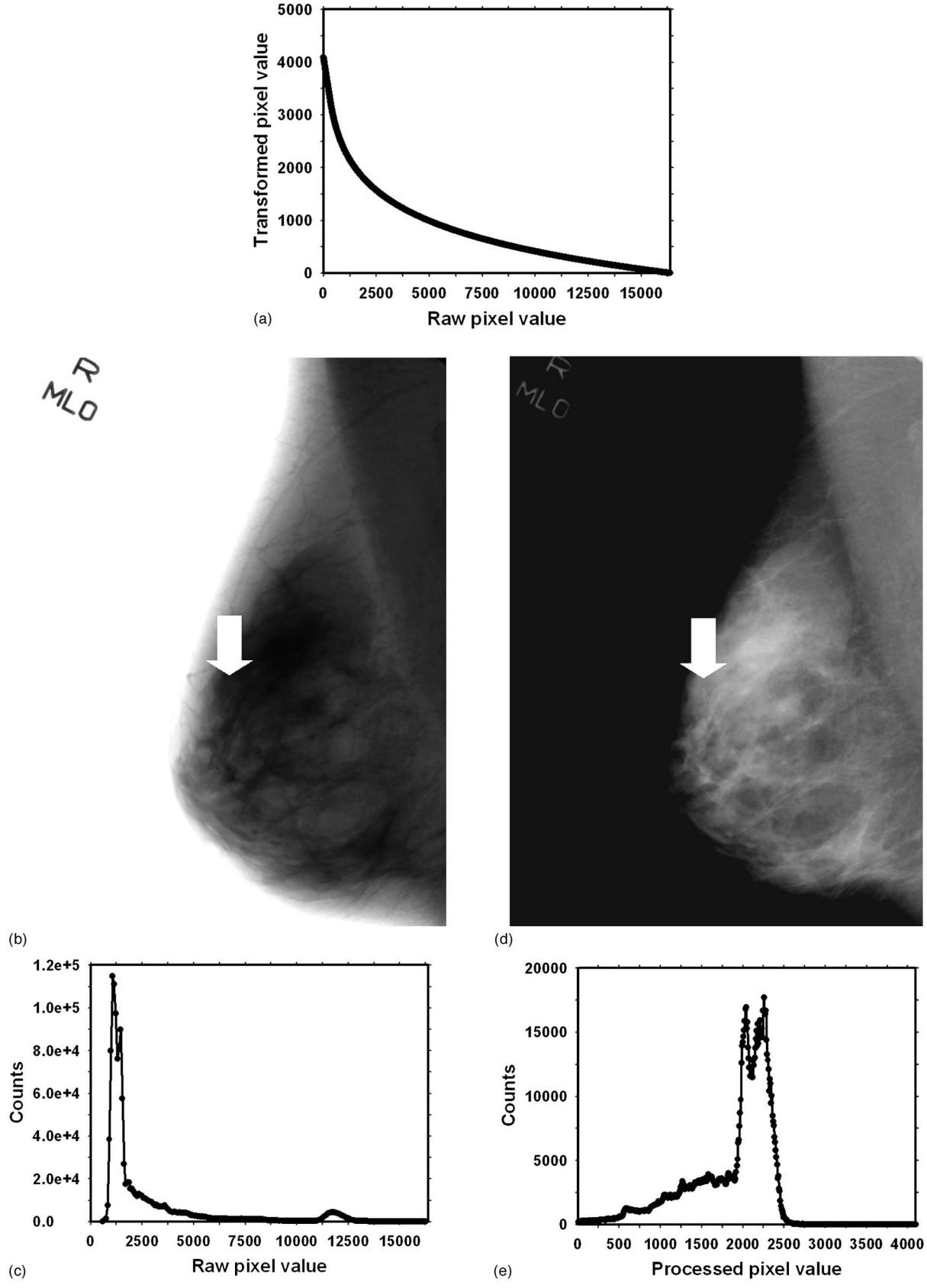

FIG. 3. (a) The inverted logarithmic transformation function from the raw pixel value x to the transformed pixel value y. $y=\alpha \ln(x_0/x)$ for $x>x_T$, $y=4095-\beta x$ for $x\leqslant x_T$, where x_0 is the raw pixel value of the unattenuated beam, $\alpha=833.67$, $\beta=2.541$, and $x_T=320$. (b) A raw image (14 bit), (c) histogram of the raw image, (d) a processed image (12 bit), and (e) histogram of the processed image.

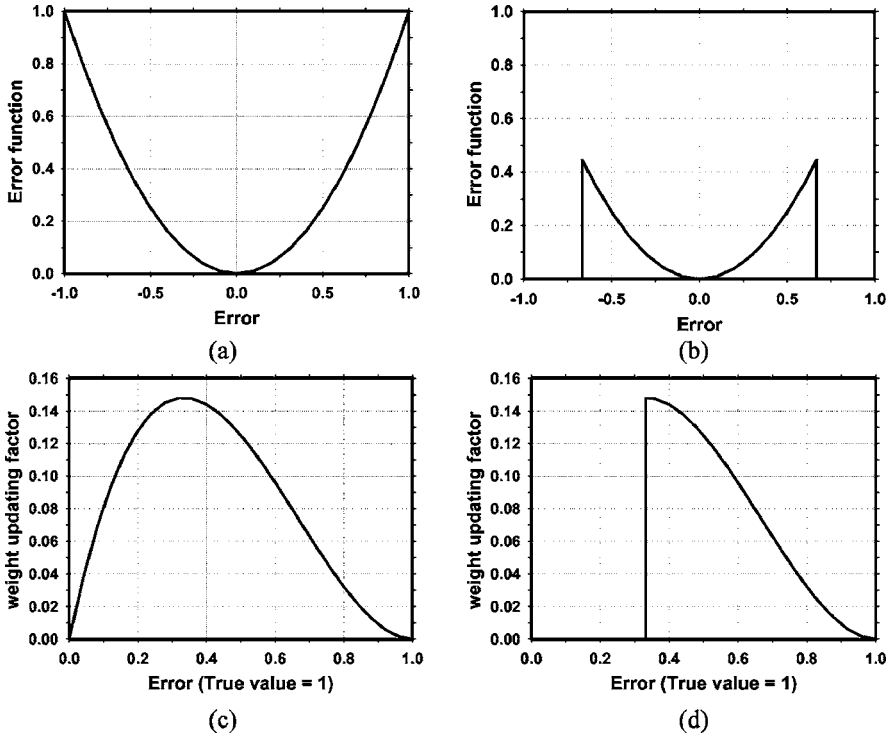

FIG. 4. (a) Sum-of-squares error function, (b) truncated sum-of-squares error function, (c), (d) weight updating factors for (a) and (b), respectively. The weights between the second hidden layer and output layer are updated as $w_{n+1}=w_n-\eta\Sigma\,\gamma x_j$, where η, γ, and x_j are the learning rate, weight updating factor and jth activation value from previous layer, respectively.

node groups in the first hidden layer and ten node groups in the second hidden layer were fully connected. The kernel sizes of the first group of filters between the input and the first hidden layer were 5×5, and those of the second group of filters between the first and second hidden layers were 7×7. The logistic sigmoid function was chosen as the activation function for both the hidden nodes and output nodes. The choice of sum-of-squares error (SSE) function allows a probabilistic interpretation of the CNN output.[37] That is, the CNN output may be interpreted as the probability of correctly classifying the input sample as a true microcalcification ROI. Our CNN was previously trained using back-propagation learning rule with an SSE function. The SSE function has the disadvantage that any mislabeled samples or outliers will send the network a large feedback signal during training to drastically alter the weights. Although we have carefully screened the training ROIs so that we do not expect any mislabeled samples, some true microcalcifications may be outliers because the gray-level variation of the surrounding background may be very large due to, for example, the presence of other true microcalcifications in the ROI. Even with an optimal CNN architecture, the learning curve can oscillate drastically between iterations in the presence of noisy training data. In this study, we used a truncated sum-of-squares error function which prohibits the updating of CNN weights when the absolute difference between the CNN output and the target value is larger than a threshold. The truncated sum-of-squares error function and the associated weight updating factor are illustrated in Fig. 4. At the beginning of CNN training, it is expected that many samples will produce large errors because the CNN weights are far from their optimal values. After some training iterations, however, large errors may indicate outliers. For this reason, the truncated sum-of-squares error function was not applied until after a chosen number of iterations of the training.

When a given subset of the available data set was used for training the CAD system, the cases in the subset were further separated into a training set and a validation set for the training of the CNN classifier as shown in Fig. 5. The input samples to the CNN classifier was 16×16-pixel ROIs, each of which was extracted with its center at either a true or false signal. True positive ROIs were selected from the manually identified individual microcalcifications which were within the true clusters. For some clusters in our data set, the individual microcalcifications were located very close together so that a 16×16-pixel ROI could include more than one microcalcification. The additional microcalcifications within

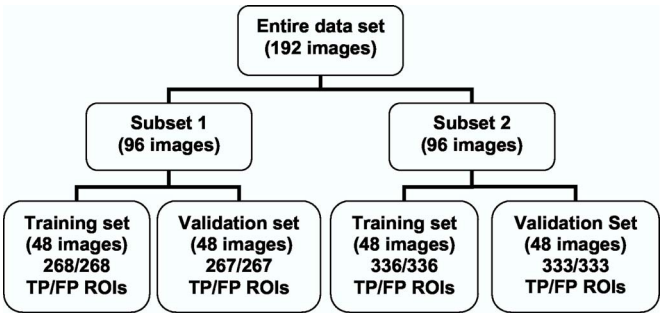

FIG. 5. The data subsets for design of our CAD system. The data set was separated into two independent subsets in a cross-validation training and testing scheme. When a given subset is used for training of the CAD system, the data were further separated into a training set and a validation set for training the CNN classifier and the LDA classifier. The trained CAD was then applied to the other subset for evaluation of its test performance. The numbers of TP and FP ROIs for individual microcalcifications in each of the training and validation sets for CNN training are shown.

the ROI may cause a large variation of the pixel values surrounding the selected microcalcification. The ROIs containing more than six individual microcalcifications were excluded from the training. False positive ROIs were obtained from the output of the adaptive gray-level local thresholding step when the CAD system was applied to the training set. Since there were more FPs than TP microcalcifications, we randomly selected a subset of the FPs such that the numbers of TP and FP ROIs were the same. The numbers of TP and FP ROIs in each of the training and validation sets are shown in Fig. 5. To reduce the effect of the orientation of the individual microcalcification in the training, each ROI was rotated 90°, 180°, and 270° and mirrored to generate eight ROIs as the CNN training inputs. The rotation and mirroring were not intended to generate more training samples but to reduce the fluctuations due to the location and orientation of the microcalcification in the ROI.

The accuracy for classification of true microcalcifications and FPs was evaluated as the area under the fitted ROC curve A_z using the LABROC program. The A_z, in combination with a cost function based on the free-response ROC (FROC) curve for cluster detection, for the validation set within the training subset was used to guide the selection of the CNN weights, as described below. The cost function was defined as

$$C = 100(u - l) - \int_l^u s(f)df,$$

where l and u were the lower and upper limits of the FP range of interest, respectively; f was the average number of FP clusters per image obtained as described in the next section and $s(f)$ was the sensitivity at an FP rate of f.[23] Figure 6 shows the mean-squared error and A_z for both the training set and validation set of subset 1 as a function of the training iteration number. As shown in Fig. 6(b), the validation A_z was improved from about 0.965 to 0.981 when the truncated error metric was applied during training. The training of the CNN was terminated if the mean-squared error was less than 0.005. We observed that after 90 iterations, both the mean-squared curves and A_z curves were almost flat. Therefore, we could select any trained CNN after 90 iterations as the best one in terms of classifying true ROIs and false ROIs. In this study, we selected ten consecutive iterations within the same training run in the flat region of the A_z curve. The ten trained CNNs were applied to the validation set for FP reduction. The trained CNN was selected as the one with the median value of the cost function C estimated from the ten FROC curves.

Once the training with one subset was completed, the CNN and other thresholds were fixed for testing with the other subset. The training and test subsets were then switched and the entire training process was repeated. The overall detection performance was evaluated by averaging the FP rates at the corresponding sensitivity levels along the FROC curves for the two test subsets.

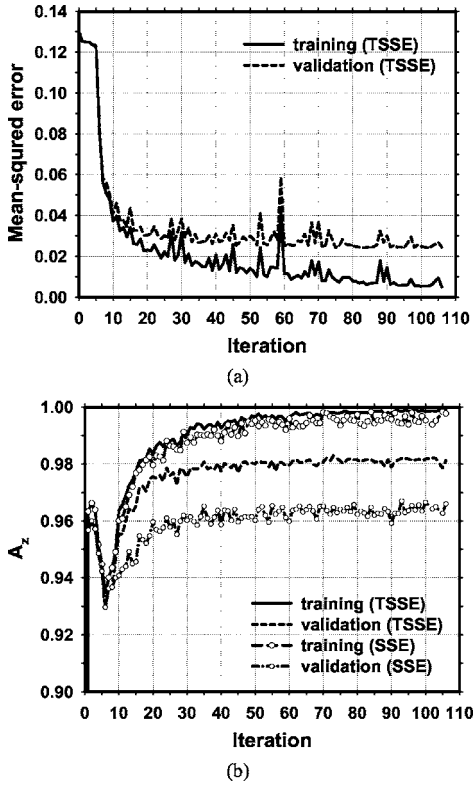

FIG. 6. Dependence of (a) the mean-squared error and (b) the A_z on the number of iterations for both the training and validation sets. In training with the truncated sum-of-squares error function (TSSE), the truncated squared error function was applied after tenth iteration. When SSE was used for training, both the training and validation A_z values were lower.

5. Regional clustering and false positive reduction for clustered microcalcifications

Finally, potential clusters are identified by a regional clustering procedure[19,38] based on the fact that true microcalcifications of clinical interest always appear in clusters on mammograms. In this procedure, a region with a higher concentration of potential signals is given a higher priority as a starting region to grow a cluster. A dynamic clustering algorithm identifies new members in the neighborhood and updates the cluster centroid after each new member is added. A potential signal is included as a member if it is within a preselected distance threshold (5 mm) from the current cluster centroid. A cluster completes growing if no more potential signals in the neighborhood can satisfy the inclusion criterion. A cluster is considered to be positive if the number of its members is greater than three. Although the cluster centroid is dynamically updated when new members are included, the detected cluster regions are typically about 1 cm in the long dimension. The process continues until no more clusters can be grown in the breast region. The remaining signals which are not found to be members of any potential clusters will be considered as isolated noise objects and excluded.

In order to differentiate true clustered microcalcifications from clusters of normal noisy structures, we extracted features from each of the clusters found at the stage of regional

clustering and built an LDA classifier. The size, the mean density, the eccentricity, the moment ratio, and the axis ratio features[39] were first extracted for the individual microcalcifications in the cluster. To quantify the variation of these visibility and shape descriptors in a cluster, the maximum, the average, the standard deviation, and the coefficient of variation (ratio of the standard variation to the average) of each of these five features were then calculated for the cluster. Another feature describing the number of microcalcifications in a cluster is also included, resulting in a set of 21 morphological features. A detailed description of these features can be found in the literature.[39]

We also used features derived from the CNN output value of each individual microcalcification in the cluster. These features included the minimum, the maximum, and the mean of the CNN output values in the cluster, and the average of the first three highest CNN output values. A total of 25 features (21 morphological features, 4 CNN features) were extracted for each of the clusters. In order to obtain the best feature subset and reduce the dimensionality of the feature space to design an effective classifier, feature selection with stepwise linear discriminant analysis (LDA) was applied. Stepwise feature selection involves the selection of three parameters, namely, F_{in}, F_{out}, and tolerance. A discussion of how these parameters are related to the feature selection process and how they affect the performance can be found in the literature.[40] In this study, for a given training subset of 96 images, we first split the subset into a training set and a validation set, each with 48 images, as shown in Fig. 5. An appropriate set of parameters was selected by searching in the parameter space for the combination of F_{in}, F_{out}, and tolerance that could achieve the highest classification accuracy, in terms of A_z, with a relatively small number of features in the validation set. We then used the chosen set of F_{in}, F_{out}, and tolerance parameters to select a final set of features and LDA coefficients using the entire training subset of 96 images which contained 96 TP and over 500 FP clusters. The trained classifier was applied to the test subset as an FP reduction step in the CAD system. Note that only a small subset of the 25 features were selected (see Sec. III) during the classifier design process.

6. Evaluation methods

Computerized detection of MCCs is a complicated task to evaluate, because it involves objects with multiple elements.[41] The scoring method used in this study has been described in detail in our previous study for SFMs.[23] Briefly, there are two sets of inputs to the automatic scoring program. The first consists of the overlay files, in which the extent of each MCC is drawn by an expert radiologist as a polygon. The second consists of outputs of the automated microcalcification detection program, which are the smallest rectangular bounding boxes enclosing the detected MCCs. The scoring program automatically calculates the intersection of the areas enclosed by these rectangles and the polygons. If the ratio of the intersection area to either the rectangle or the polygon area is more than 40%, as determined in the previous study,[23] then the cluster enclosed by the polygon is considered to be detected. If a polygon area intersects with more than one rectangular region, only one TP finding is recorded.

The detection performance of the CAD system was assessed by FROC analysis. FROC curves were presented on a per-cluster and a per-case basis. For cluster-based FROC analysis, the MCC on each mammogram was considered an independent true object; the sensitivity was thus calculated relative to 96 clusters in each of the two test subsets. For case-based FROC analysis, the same MCC imaged on the two-view mammograms was considered to be one true object and detection of either or both clusters on the two views was considered to be a TP detection; the sensitivity was thus calculated relative to 48 clusters in each of the two test subsets.

To evaluate the effect of the preprocessing methods on microcalcification detection, we also trained a CAD system using the GE-processed images as input. This CAD system used the same methods as those described above for the raw images except that the inverted logarithmic transformation was not applied, and that the CNN was retrained specifically for the GE-processed images to obtain the best performance. The training and test subsets contained the same corresponding cases as for the raw image subsets. The training and testing were performed using the cross-validation method as described above. The performance of the CAD system using the GE-processed images as input was quantified by the average test FROC curve and compared with that using the raw images. In addition, the CAD system trained with the raw images as input was applied to the GE-processed images without retraining, except that the inverted logarithmic transformation was turned off, to evaluate the robustness of the system against small differences in the image properties due to preprocessing.

We also used the data set without clustered microcalcifications to evaluate the FP cluster detection rate on negative cases. We applied the two trained CAD systems obtained in the two-fold cross-validation scheme separately to the negative data set for FP detection. For a given CAD system, the FP rate was determined by counting the detected clusters on the negative mammograms while the detection sensitivity was determined by counting the TP clusters on the test subset. A test FROC curve was then derived by combining the sensitivity from the test subset and the FP rate from the negative data set at the corresponding detection thresholds. After the test FROC curve was determined separately for each of the two CAD systems, the two test FROC curves were averaged to obtain an overall FROC curve quantifying the test performance of our approach to clustered microcalcification detection on FFDMs.

The FROC curve captures the inherent tradeoff between sensitivity and FP rate. Therefore, the FPs/image at a certain sensitivity level is typically used as a performance measure. However, the uncertainties stemming from the variations in the data sets and the various processes in the detection system being used are not estimated. Chakraborty and Berbaum proposed the jackknife FROC (JAFROC) analysis[42] for testing the significance of the difference between two FROC

curves. In this analysis, one assigns weights to each lesion in the case, where the weights add up to unity.[42] A figure-of-merit (FOM) which involves the weighted combinations of the ratings of detections is then defined to evaluate the performance. If each case only has one lesion, the weights are all unity and the FOM degenerates to the Wilcoxon statistic. The analysis step of the JAFROC method follows the same basic rationale as jackknife analysis of ROC data.[42] We used this method to test the statistical significance of the difference between the FROC curves in this study.

III. RESULTS

In our CAD system for FFDMs, the global threshold in the segmentation stage and the CNN weights were retrained as described above. We first evaluated the performance of the system by comparing the test FROC curves without the CNN and LDA classifiers. The FROC curves were generated by varying the local SNR threshold in the range of 1.9 to 3.7. Figure 7(a) shows that a cluster-based sensitivity of 98% can be achieved at about 10 FPs/image for both subsets. At the same FP rate, the system detected the cluster on at least one view for all the cases (100% case-based sensitivity) in both test subsets as shown in Fig. 7(b). The FP rate is high because many image structures cannot be differentiated from individual microcalcifications at the global and local thresholding stages. The two rule-based features used in the CAD system, the area and the gray-scale contrast of the microcalcification candidate, are not very effective in reducing FPs.

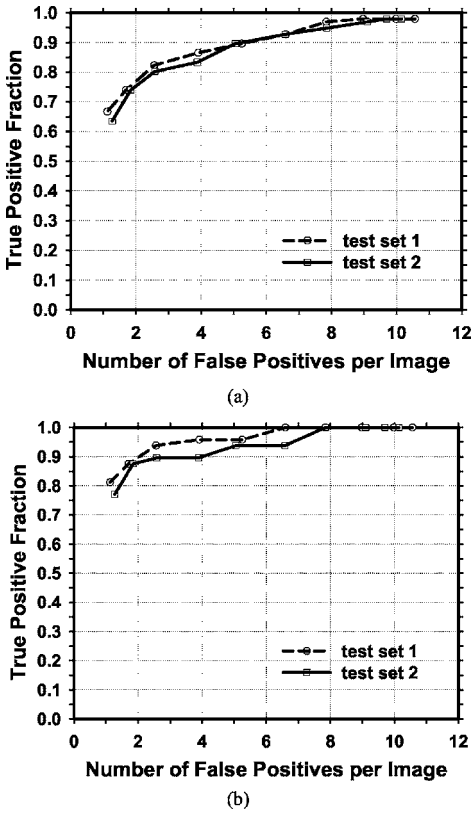

FIG. 7. The test FROC curves from the two independent microcalcification subsets for the CAD system without the FP reduction stages. FROC curves were obtained by varying the local SNR thresholds and the FP rate was estimated from the test subsets with microcalcification clusters. (a) Cluster-based FROC curves and (b) case-based FROC curves.

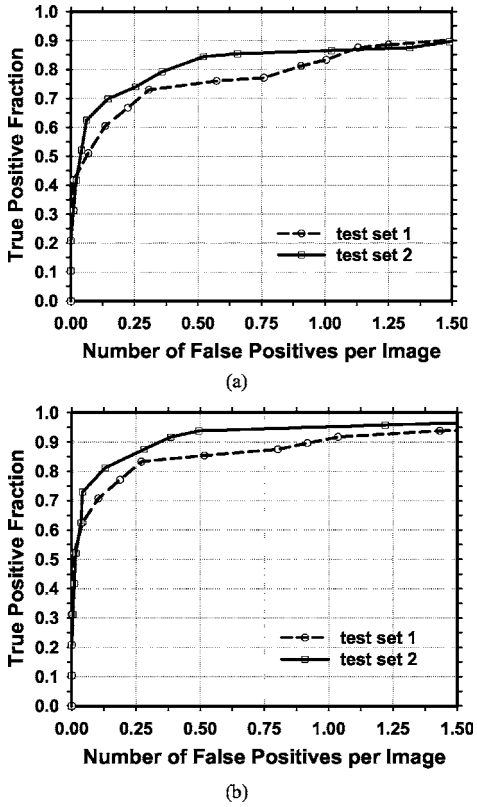

FIG. 8. The test FROC curves from the two independent microcalcification subsets for the CAD system after FP reduction using the CNN and LDA classifiers. FROC curves were obtained by varying LDA classifier threshold and the FP rate was estimated from the test subsets with microcalcification clusters. (a) Cluster-based FROC curves and (b) case-based FROC curves.

Five (two CNN and three morphological) and three (one CNN and two morphological) features were selected from the two independent training sets, respectively, for the LDA classifier. With the CNN classifier and LDA classifier, the performance of the CAD system can be substantially improved. Figure 8 compares the test FROC curves with and without the classifiers. The FP rates at cluster-based detection sensitivities of 70, 80, and 90 % are also summarized in Table I. On average, the CNN and LDA classifiers reduced the FP rate by 86, 74, and 72 %, at cluster-based detection sensitivities of 70, 80, and 90 %, respectively, for the two test subsets. An example of the microcalcification candidates

TABLE I. Comparison of the performance of the CAD system with and without CNN and LDA classifiers at cluster-based detection sensitivities of 70, 80, and 90 %.

	Sensitivity	70%	80%	90%
Test set 1	Without CNN and LDA	1.38	2.32	5.42
	With CNN and LDA	0.27	0.86	1.47
Test set 2	Without CNN and LDA	1.63	2.56	5.25
	With CNN and LDA	0.15	0.38	1.56

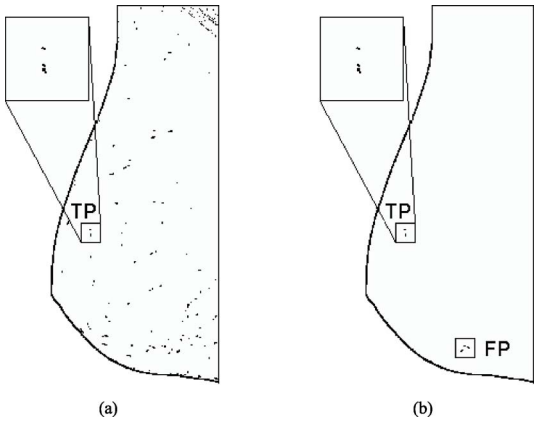

FIG. 9. (a) Microcalcification candidates after the segmentation stage. (b) Detected microcalcification cluster at the output of the CAD system. The gray-level mammogram is shown in Fig. 3(c).

on a test mammogram after the segmentation stage is shown in Fig. 9(a). The detected MCCs at the output of the CAD system are shown in Fig. 9(b). As seen from Fig. 9(b), most of the FP microcalcifications were removed by our CNN and LDA classifiers.

We evaluated the effect of the preprocessing methods on microcalcification detection. Figure 10 shows the average test FROC curves of the CAD systems using the 12-bit GE-processed images as input and the CAD system using the 14-bit raw images as input, and the average FP rates at cluster-based sensitivities of 70, 80, and 90 % are compared in Table II. The FP rates were estimated from the test subsets with microcalcifications. An average FROC curve was derived from the FROC curves for the two test subsets by averaging the FPs/images at the corresponding sensitivities. When the FP rates were estimated from the negative data set without clusters, the two CAD systems demonstrated some differences at different FP ranges as shown in Table II and Fig. 11.

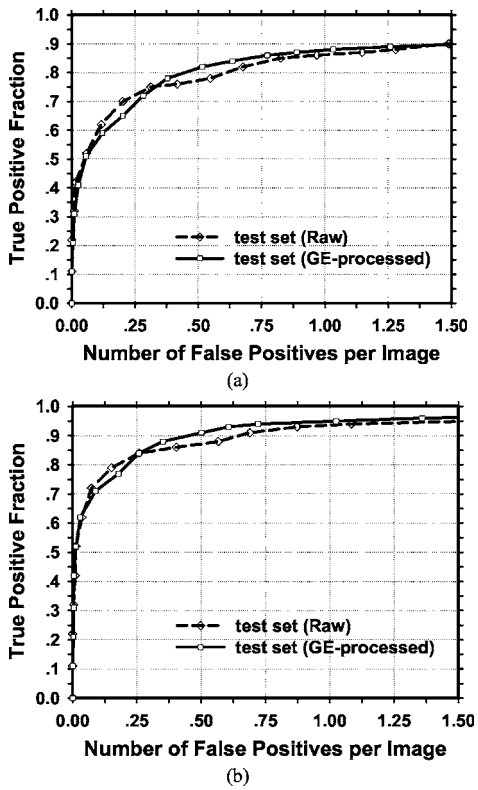

FIG. 10. Comparison of the average test FROC curves obtained from: (1) the CAD system using raw images as input and (2) the CAD system using GE-processed images as input. (a) Cluster-based FROC curves and (b) case-based FROC curves. The FP rates were estimated from the test subsets with microcalcification clusters.

TABLE II. Comparison of the cluster-based performance of the CAD systems using raw images as input and that of the CAD systems using GE-processed images as input. The FP rates were presented as the average from the two test subsets at the corresponding sensitivity levels.

	Sensitivity	70%	80%	90%
Raw	FP based on abnormal set	0.21	0.61	1.49
	FP based on negative set	0.15	0.31	0.86
	FP based on negative set (CAD system using SSE)	0.22	0.46	1.68
GE processed	FP based on abnormal set	0.26	0.45	1.59
	FP based on negative set	0.06	0.19	1.66
	FP based on negative set (CAD system using raw image input)	0.06	0.15	1.87

We applied the JAFROC analysis for testing the significance of the difference between the test FROC curves (using FP rates estimated from the negative data set) generated by the CAD system using raw images as input and the CAD system using GE-processed images as input. The results are summarized in Table III. The FOM from the output of the JAFROC software was 0.83 and 0.87, respectively, on test subsets 1 and 2 for the CAD system using raw images as input, and 0.87 and 0.88, respectively, on the same subsets for the CAD system using GE-processed images as input. The difference between the FOM for our processed images and that for the GE-processed images did not achieve statistical significance ($p > 0.05$) for both test subsets.

We also applied the CAD system trained for the raw images to the data set of the GE-processed images without retraining and the inverted logarithmic transform of the gray levels were turned off. Figure 11 and Table II show that the CAD system achieved cluster-based sensitivities of 70, 80, and 90 % at 0.06, 0.15, and 1.87 FPs/image, respectively, when the FP rates were evaluated on the data set without MCCs. As shown in Table III, the overall test performance was not significantly different ($p > 0.05$) from that of the CAD system retrained for the GE-processed images for both test subsets.

Figure 12 shows the effect of using the truncated sum-of-squares error function in the training of the CNN classifier. The CAD system using the CNN trained with the SSE error function achieved cluster-based sensitivities of 70, 80, and 90 % at 0.22, 0.46, and 1.68 FPs/image, respectively. These FP rates were higher than those obtained with the CNN trained with the truncated sum-of-squares error function. The improvement in the detection performance by using the CNN

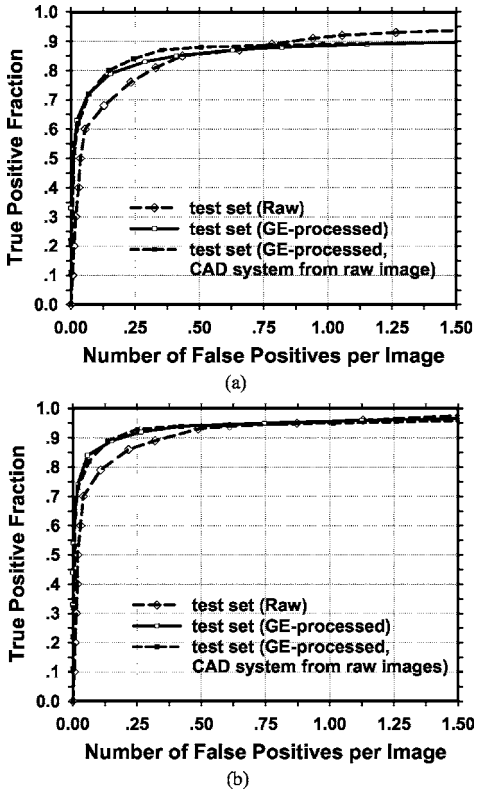

FIG. 11. Comparison of the average test FROC curves obtained from: (1) the CAD system using raw images as input; (2) the CAD system using GE-processed images as input; and (3) the CAD system using raw images as input but applied to the GE-processed images without retraining. (a) Cluster-based FROC curves and (b) case-based FROC curves. The FP rates were estimated from the test subset negative for microcalcification clusters.

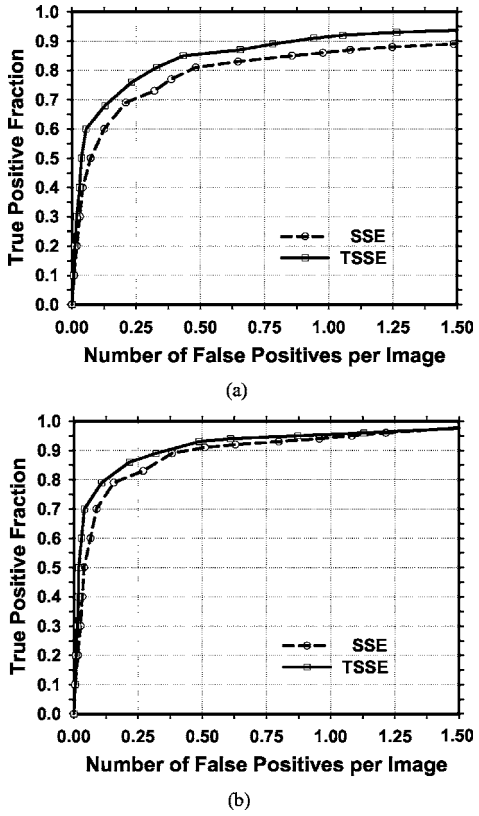

FIG. 12. Comparison of the average test FROC curves for the CAD systems using CNN trained with SSE (sum-of-squares error function) and TSSE (truncated sum-of-squares error function). The CAD system using the raw images as input was used and the FP rates were estimated from the mammograms negative for microcalcification clusters. (a) Cluster-based FROC curves and (b) case-based FROC curves.

trained with the truncated sum-of-squares error function was statistically significant ($p<0.05$) for both test subsets. The CNNs used in all other CAD systems evaluated in this study were therefore trained with the truncated sum-of-squares error function.

The detection performance of a CAD system for malignant clusters is more important than its performance for detecting all clusters. Therefore, we also evaluated the detection accuracy of our CAD system separately for malignant MCCs and for benign MCCs. Figures 13(a) and 13(b) compare the average cluster-based and case-based test FROC curves for detection of malignant and benign clusters. The CAD system achieved cluster-based sensitivities of 70, 80, and 90 % at 0.07, 0.13, and 0.35 FPs/image, respectively, for malignant clusters, as compared to 0.21, 0.39, and 1.22 FPs/image, respectively, for benign clusters. All malignant clusters were detected by the CAD system on at least one view (100% case-based sensitivity) at an average of 0.25 FPs/image.

TABLE III. Estimation of the statistical significance in the difference between the FROC performance of the CAD system using the FFDM raw images as input and that of the CAD system using GE-processed images as input. The FROC curves with the FP rates obtained from the data set without MCCs were compared. (JAFROC software does not provide the standard deviation for FOM.)

	FOM (JAFROC)	
	Test subset 1	Test subset 2
A. Raw image	0.83	0.87
B. GE processed	0.87	0.88
C. GE processed (CAD system from raw images)	0.87	0.89
p value for A and B	0.16	0.79
p value for B and C	0.59	0.27

IV. DISCUSSION

FFDM systems from several manufacturers have obtained FDA approval for clinical use. In these systems, the absorbed x-rays produce electric charges in the detector either directly or indirectly, and the charges arising from many x-rays incident on a detector element are accumulated to produce a signal measurement. The two-dimensional image obtained from these measurements on the detector is generally referred to as the raw image. Since the raw pixel values are a linear function of the absorbed x-ray charges, the signal range between different digital detectors can be normalized linearly with respect to each other. However, each FFDM manufacturer has designed their own proprietary preprocess-

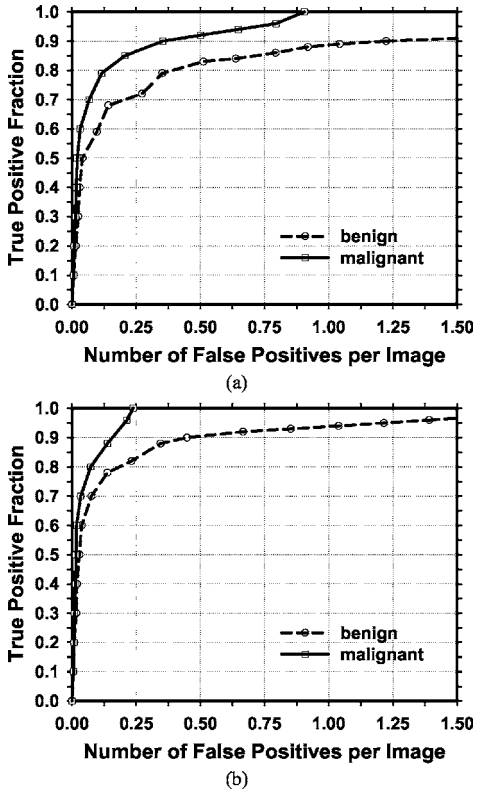

FIG. 13. Comparison of the average test FROC curves for benign and malignant cases. The CAD system using the raw images as input was used and the FP rates were estimated from the mammograms negative for microcalcification clusters. (a) Cluster-based FROC curves and (b) case-based FROC curves.

ing method to enhance the raw image for display. The processed images from different FFDM manufactures can differ in the image properties.[30] Thus it is important to develop a CAD system using raw images as the input to reduce one of the major differences between mammograms produced by different FFDM systems. In this study we applied an inverted logarithmic transformation[32] to the raw pixel values. The histograms of a typical raw image and the inverted logarithmic transformed image are shown in Figs. 3(b) and 3(e), respectively. Since GE did not publish their proprietary FFDM preprocessing method officially, we do not know whether the transformed raw images are different from the GE-processed images only by a thickness correction at the breast periphery. Thus it is prudent to investigate how much difference in CAD performances there is between the inverted logarithmic transformed raw images and the GE-processed images and how sensitive the trained CAD system is to these two processing methods. As shown in Fig. 10, our results showed that there is only a minor difference between the inverted logarithmic transformed raw images and the GE-processed images either with cluster-based or case-based FROC curves. This demonstrates that our CAD system may be easily adapted to raw images from other manufacturers' FFDM systems, using a user-controllable preprocessing method such as the inverted logarithmic transformation with properly chosen parameters. Nevertheless, the additional effects of differences in pixel size and noise properties of different FFDM systems on lesion detection will remain an area of future investigation.

To evaluate the robustness of the CAD system trained with raw images as input, we applied the system to the GE-processed images without retraining, except that the inverted logarithmic transformation was turned off. It was found that the performance of the system was comparable to that retrained using the GE-processed images. The performance was also similar to that obtained using the raw images as input. This result indicates that our CAD system does not need to be retrained to accommodate differences between these two processing methods. Whether it is necessary to retrain our CAD system for processed FFDMs from other manufacturers still needs to be investigated when data sets are available to us in the future.

For the CNN, the computation cost per iteration is high and increases as the number of weights increases. The convergence is slow if the conventional back-propagation training algorithm is used. Many methods[43–45] have been proposed for accelerating the convergence of the back-propagation algorithm. These methods include proper weights initialization,[43] learning rate adaptation,[44] and training with momentum terms.[45] These methods can be applied to CNN to reduce the training cost. There are also many methods[46,47] for improving the generalization ability of artificial neural networks. In this study, we adopted a truncated sum-of-squares error function to reduce the large oscillations of the weights being updated, and thus to improve the training efficiency and the generalization ability of the trained CNNs. We observed that the CNN converged within 100 iterations and the A_z of the validation set increased substantially as demonstrated in Fig. 6. When the sum-of-squares error function was used, the training A_z required a much larger number of iterations (>300) to reach 0.999 whereas the validation A_z could not reach as high a level as that obtained from the truncated error function even after over 300 iterations.

Several clinical trials of breast cancer screening have been conducted to compare FFDM with SFM in screening populations.[25,48–50] Due to important differences in various factors, such as the mammographic equipment, the study design, the sample sizes, and the reader experience, these clinical trials arrived at different conclusions about the advantages or disadvantages of FFDM in comparison to SFM systems. Since the detection of cancers with a computerized program can also be affected by the image properties of the mammograms, it is important to conduct studies to compare the performance of CAD systems between FFDMs and SFMs.

Most of the cases we collected for the project were not clinically normal. We used a set of mammograms that may contain soft-tissue masses but are negative with respect to clustered microcalcifications to evaluate the FP rate of the microcalcification detection system. The "truth" that a mammogram in this negative set did not contain clustered microcalcifications was based on review by experienced breast ra-

diologists. We did not have the follow-up results for many of the mammograms in this negative data set. There is a possibility that some of our negative mammograms may contain clusters that are visible to the radiologists only in retrospect, if they are detected on a follow-up examination. Such a cluster, if detected by our CAD system, may no longer be considered as an FP. Therefore, follow-up for the negative mammograms may decrease the number of FP marks that were counted in this study, which means that the FP rate from these negative mammograms may be slightly overestimated and the FROC curve may be somewhat pessimistically biased.

V. CONCLUSION

In this work, we developed a CAD system for MCCs which uses the raw FFDMs as the input. Our previous CAD system that was developed on digitized screen-film mammograms was adapted to FFDMs. With the use of truncated sum-of-squares error metric, the training of CNN could be accelerated and the classification performance on the test subsets was improved. The CNN in combination with a LDA classifier could substantially reduce FPs with a small tradeoff in sensitivity. The CAD system achieved a cluster-based sensitivity of 70, 80, and 90 % at 0.21, 0.61, and 1.49 FPs/image, respectively. For case-based performance evaluation, a sensitivity of 70, 80, and 90 % were achieved at 0.07, 0.17, and 0.65 FPs/image, respectively. The corresponding FP rates were 0.15, 0.31, and 0.86 FPs/image for cluster-based detection when the FP rates were estimated using negative mammograms without microcalcifications. Further study is underway to improve the CAD system using a larger data set. In addition, we will incorporate joint two-view information[51] for FP reduction in our CAD system for FFDMs.

ACKNOWLEDGMENTS

This work is supported by USPHS Grant No. CA95153 and U. S. Army Medical Research and Materiel Command Grant No. DAMD17-02-1-0214. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program, and to D. Chakraborty, Ph.D., for the JAFROC program.

[a]Electronic mail: gejun@med.umich.edu

[1]American Cancer Society, "Statistics for 2005," www.cancer.org 2005.
[2]C. Byrne, C. R. Smart, C. Cherk, and W. H. Hartmann, "Survival advantage differences by age: evaluation of the extended follow-up of the Breast Cancer Detection Demonstration Project," Cancer **74**, 301–310 (1994).
[3]S. A. Feig and R. E. Hendrick, in *Risk, Benefit, and Controversies in Mammographic Screening. In: Syllabus: A Categorical Course in Physics Technical Aspects of Breast Imaging*, edited by A. G. Haus and M. J. Yaffe (Radiological Society of North America, Inc, Oak Brook, IL, 1993).
[4]C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials," Cancer **75**, 1619–1626 (1995).
[5]J. A. Harvey, L. L. Fajardo, and C. A. Innis, "Previous mammograms in patients with impalpable breast carcinomas: Retrospective vs blinded interpretation," AJR, Am. J. Roentgenol. **161**, 1167–1172 (1993).
[6]R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," Radiology **184**, 613–617 (1992).
[7]V. Beam, D. Sullivan, and P. Layde, "Effect of human variability on independent double reading in screening mammography," Acad. Radiol. **3**, 891–897 (1996).
[8]H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," Invest. Radiol. **25**, 1102–1110 (1990).
[9]L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," Radiology **215**, 554–562 (2000).
[10]T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," Radiology **220**, 781–786 (2001).
[11]R. F. Brem, J. K. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: A multi-institutional trial," AJR, Am. J. Roentgenol. **181**, 687–693 (2003).
[12]S. V. Destounis, P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? initial experience," Radiology **232**, 578–584 (2004).
[13]M. A. Helvie *et al.*, "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection—A pilot clinical trial," Radiology **231**, 208–214 (2004).
[14]E. A. Sickles, D. E. Wolverton, and K. E. Dee, "Performance parameters for screening and diagnostic mammography: specialist and general radiologists," Radiology **224**, 861–869 (2002).
[15]D. Gur, J. H. Sumkin, H. E. Rockette, M. A. Ganott, C. Hakim, L. A. Hardesty, W. R. Poller, R. Shah, and L. Wallace, "Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**, 185–190 (2004).
[16]S. A. Feig, E. A. Sickles, W. P. Evans, and M. N. Linver, "Re. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**, 1260–1261 (2004).
[17]S. Ciatto, L. Cataliotti, and V. Distante, "Nonpalpable lesions detected with mammography: Review of 512 consecutive cases," Radiology **165**, 99 (1987).
[18]D. B. Kopans, *Breast Imaging*, 2nd ed. (Lippincott-Raven Publishers, 227 East Washington Square, Philadelphia, PA, 1997).
[19]H. P. Chan, S. C. B. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network," Med. Phys. **22**, 1555–1567 (1995).
[20]R. M. Nishikawa, Y. Jiang, M. L. Giger, R. A. Schmidt, C. J. Vyborny, W. Zhang, J. Papaioannou, U. Bick, R. Nagel, and K. Doi, "Performance of automated CAD schemes for the detection and classification of clustered microcalcifications," in *Digital Mammography*, edited by A. G. Gale, S. M. Astley, D. R. Dance, and A. Y. Cairns (Elsevier, Amsterdam, 1994).
[21]B. Zheng, Y. S. Chang, M. Staiger, W. Good, and D. Gur, "Computer-aided detection of clustered microcalcifications in digitized mammograms," Acad. Radiol. **2**, 655–662 (1995).
[22]H. P. Chan, K. Doi, S. Galhotra, C. J. Vyborny, H. MacMahon, and P. M. Jokich, "Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography," Med. Phys. **14**, 538–548 (1987).
[23]M. N. Gurcan, H. P. Chan, B. Sahiner, L. Hadjiiski, N. Petrick, and M. A. Helvie, "Optimal neural network architecture selection: Improvement in computerized detection of microcalcifications," Acad. Radiol. **9**, 420–429 (2002).
[24]N. Karssemeijer, "Stochastic model for automated detection of microcalcifications in digital mammograms," Image Vis. Comput. **10**, 369–375 (1992).
[25]E. D. Pisano, C. Gasonis, E. Hendrick, and M. Yaffe, "Diagnostic performance of digital versus film mammography for breast-cancer screening," N. Engl. J. Med. **353**, 1773–1783 (2005).

[26]J. A. Baker, J. Y. Lo, D. M. Delong, and C. E. Floyd, "Computer-aided detection in screening mammogrpahy: Variability in cues," Radiology **233**, 411–417 (2004).
[27]K. F. O'Shaughnessy, R. A. Castellino, S. L. Muller, and K. Benali, "Computer-aided detection (CAD) on 90 biopsy-proven breast cancer cases acquired on a full-field digital mammography (FFDM) system," Radiology **221**(P), 471 (2001).
[28]A. Kshrsagar, K. Young, and S. Stapleton, "Comparison of CAD performance and independent double-reading in screen-film mammography and full-field digital mammography with soft-copy reading: results from the follow-up of the paired Oslo I study," RSNA 2004, Chicago, November 30–December 5 (2004).
[29]K. J. McLoughlin, P. J. Bones, and N. Karssemeijer, "Noise equalization for detection of microcalcification clusters in direct digital mammogram images," IEEE Trans. Med. Imaging **23**, 313–320 (2004).
[30]E. D. Pisano *et al.*, "Radiologists' preferences for digital mammographic display," Radiology **216**, 820–830 (2000).
[31]N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst. Man Cybern. **9**, 62–66 (1979).
[32]A. Burgess, "On the noise variance of a digital mammography system," Med. Phys. **31**, 1987–1995 (2004).
[33]M. A. Gavrielides, J. Y. Lo, R. Vargas-Voracek, and J. C. E. Floyd, "Segmentation of suspicious clustered microcalcifications in mammograms," Med. Phys. **27**, 13–22 (2000).
[34]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, C. Zhou, and Z. Ge, "Computer aided detection of breast masses on full-field digital mammograms: false positive reduction using gradient field analysis," Proc. SPIE Medical Imaging **5370**, 992–998 (2004).
[35]S. Dippel, M. Stahl, R. Wiemker, and T. Blaffert, "Multiscale constrast enhancement for radiographies: laplacian pyramid versus fast wavelet transform," IEEE Trans. Med. Imaging **21**, 343–353 (2002).
[36]M. Stahl, T. Aach, T. Buzug, D. S., and U. Neitzel, "Noise-resistant weak-structure enhancement for digital radiography," Proc. SPIE Medical Imaging **3661**, 1406–1417 (1999).
[37]C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon Press, Oxford, 1995).
[38]H. P. Chan, L. T. Niklason, D. M. Ikeda, K. L. Lam, and D. D. Adler, "Digitization requirements in mammography: Effects on computer-aided detection of microcalcifications," Med. Phys. **21**, 1203–1211 (1994).
[39]H. P. Chan, B. Sahiner, K. L. Lam, N. Petrick, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computerized analysis of mammographic microcalcifications in morphological and texture feature space," Med. Phys. **25**, 2007–2019 (1998).
[40]B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner, and L. M. Hadjiiski, "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size.," Med. Phys. **27**, 1509–1522 (2000).
[41]M. Kallergi, G. M. Carney, and J. Garviria, "Evaluating the performance of detection algorithms in digital mammography," Med. Phys. **26**, 267–275 (1999).
[42]D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: modeling, analysis, and validation," Med. Phys. **31**, 2313–2330 (2004).
[43]G. P. Drago and S. Ridella, "Statistically controlled activation weight initialization (SCAWI)," IEEE Trans. Neural Netw. **3**, 627–631 (1992).
[44]R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," Neural Networks **1**, 295–307 (1988).
[45]X.-H. Yu and G.-A. Chen, "Efficient backpropagation learning using optimal learning rate and momentum," Neural Networks **3**, 517–527 (1997).
[46]S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilema," Neural Comput. **4**, 1–58 (1992).
[47]A. Gupta and S. M. Lam, "Weight decay backpropagation for noisy data," Neural Networks **11**, 1127–1137 (1998).
[48]R. Hendrick *et al.*, "Non-inferiority study of FFDM in an enriched diagnositic cohort: comparison with screen-film mammography in 625 women," in *IWDM 2000: 5th International Workshop on Digital Mammography*, edited by M. J. Yaffe (Medical Physics, Madison, WI, 2001).
[49]E. Cole *et al.*, "Diagnostic accuracy of Fischer SenoScan digital mammography versus screen-film mammography in a diagnostic mammography population," Acad. Radiol. **11**, 876–880 (2004).
[50]P. Skaane, K. Young, and A. Skjennald, "Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading—Oslo I Study," Radiology **229**, 877–884 (2003).
[51]B. Sahiner, M. N. Gurcan, H. P. Chan, L. M. Hadjiiski, N. Petrick, and M. A. Helvie, "The use of joint two-view information for computerized lesion detection on mammograms: Improvement of microcalcification detection accuracy," Proc. SPIE Med. Imaging **4684**, 754–761 (2002).