CrossMark

ORIGINAL ARTICLE

An Automatic Computer-Aided Diagnosis System for Breast Cancer in Digital Mammograms via Deep Belief Network

Mugahed A. Al-antari[1] · Mohammed A. Al-masni[1] · Sung-Un Park[1] ·
JunHyeok Park[1] · Mohamed K. Metwally[1] · Yasser M. Kadah[2] · Seung-Moo Han[1] ·
Tae-Seong Kim[1]

Abstract Computer-aided diagnosis (CAD) offers assistance to radiologists in the interpretation of medical images. A CAD system learns the nature of different tissues and uses this information to diagnose abnormalities. In this paper, we propose a CAD system for breast cancer diagnosis via deep belief network (DBN) that automatically detects breast mass regions and recognizes them as normal, benign, or malignant. In this study, we utilize a standard digital database of mammography to evaluate our proposed DBN-based CAD system for breast cancer diagnosis. We utilize two techniques of ROI extraction: multiple mass regions of interest (ROIs) and whole mass ROIs. In the former technique, we randomly extract four ROIs with a size of 32×32 pixels from a detected mass. In the latter technique, the whole detected breast mass is utilized. A total of 347 statistical features are extracted for both techniques to train and test our proposed CAD system. For classification, we utilized linear discriminant analysis, quadratic discriminant analysis, and neural network classifiers as the conventional techniques. Finally, we employed DBN and compared the results. Our results demonstrate that the proposed DBN outperforms the conventional classifiers. The overall accuracies of a DBN are 92.86% and 90.84% for the two ROI techniques, respectively. The presented work shows the feasibility of a DBN-based CAD system for use as in the field of breast cancer diagnosis.

Keywords Breast cancer classification · Digital mammography · Computer-aided diagnosis (CAD) · Automatic mass detection · Deep learning · Deep belief network (DBN)

Mugahed A. Al-antari and Mohammed A. Al-masni are equally contributed to this work.

✉ Tae-Seong Kim
   tskim@khu.ac.kr

   Mugahed A. Al-antari
   en.mualshz@khu.ac.kr

   Mohammed A. Al-masni
   m.almasani@khu.ac.kr

   Sung-Un Park
   supark@khu.ac.kr

   JunHyeok Park
   psmt2655@khu.ac.kr

   Mohamed K. Metwally
   m.kilany@khu.ac.kr

   Yasser M. Kadah
   ymk@k-space.org

   Seung-Moo Han
   smhan@khu.ac.kr

[1] Department of Biomedical Engineering, College of Electronics and Information, Kyung Hee University, Yongin 17104, Republic of Korea

[2] Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah 21589, Kingdom of Saudi Arabia

1 Introduction

Breast cancer is the second-ranked cause of death for women, after lung cancer. It is generally caused by the growth of uncontrolled abnormal cells that usually arise from the inner milk ducts or lobules [1]. Microcalcifications and masses are two common types of breast cancer and can be benign or malignant. Early detection of breast cancer is critical for patient survival. In general, digital X-ray mammography is most widely used for breast

imaging and screening. The main purpose of mammography is to detect early signs of cancer and to diagnose breast masses [2]. The American Cancer Society (ACS), American College of Radiology (ACR), and American Congress of Obstetricians and Gynecologists (ACOG) suggest that women undergo annual mammograms starting at age 40 [3]. For women between 40 and 50 years of age, the National Cancer Institute (NCI) encourages breast screening one or two times a year [4]. Screening with mammography is performed in two steps. First, the breast is compressed between two small flat plates. Then, a low X-ray dose is applied directly through the breast and acquired by a two-dimensional (2D) panel detector. After that, mammographic images are assessed by radiologists. However, due to a large number of breast images, abnormal lesions can be missed. Computer-aided diagnosis (CAD) could support radiologists by improving the visual screening process and allowing better recognition of breast abnormalities. It has been reported that CAD systems improve the overall accuracy of detection of breast cancer [5–8].

There have been several studies using CAD to detect breast abnormalities in mammograms. Balanica et al. developed a CAD system that utilized the spiculated lesion features (i.e., the edge shapes of lesions) [9]. They adopted a neural network (NN) technique to train their system on 96 cases and distinguished masses as either benign or malignant. Al-Olfe et al. built a CAD system using unsupervised clustering and biclustering classifier techniques [10]. They extracted the first and second statistical and shape features via wavelet transform to classify the masses as either normal or abnormal. Kang et al. utilized two-dimensional (2D) and fractal dimension entropies as texture features [11]. They applied a support vector machine (SVM) classifier to recognize the mass abnormalities. Sharma and Khanna, developed a CAD system based on the manual collection of many regions of interest (ROIs) from two public databases, digital database for screening mammography (DDSM) and image retrieval in medical application (IRMA). They extracted the Zernike features which are rotation and shift invariant, and also relevant to the shape [12]. They used an SVM classifier to evaluate their system and diagnose lesions as malignant or nonmalignant. Recently, a new brand of classifiers has been introduced based on a new principle of deep learning, and it has been extended to address the task of breast tissue classification. Arevalo et al. utilized convolutional neural network (CNN) to explain the content of breast images and recognized masses as benign or malignant [13]. They evaluated their CAD system, showing an area under the ROC curve (AUC) of 82.10%. Jiao et al. designed a CAD system based on CNN to recognize benign or malignant tissues in a total of 600 cases [14]. They extracted multiple hierarchical levels

of features to train CNN. Kooi et al. implemented a deep model using CNN to identify malignant and suspicious normal regions among 398 regions of mammograms [15]. As CNN has been recently adopted for CAD, active investigations are underway to improve its performance to a satisfactory level.

In this paper, we propose a CAD system utilizing deep belief network (DBN), one of the deep learning algorithms [16], to classify abnormalities on mammograms or breast masses into normal, benign, or malignant. We evaluated its performance against those of other conventional classifiers. Our results show that our DBN-based CAD outperforms the conventional CAD systems. The paper is organized as follows. First, we present an overview of our proposed CAD system. Second, we automatically extract ROIs and generate the features for training and testing processes. Third, we train our proposed DBN-based CAD system using the DDSM database. Finally, we compare the performance between the proposed DBN and other classifiers.

2 Materials and Methods

The schematic diagram of our DBN-based CAD system processes is shown in Fig. 1. Our system involves automatic mass detection, ROI extraction techniques, feature extraction, and DBN classifier modules. Our presented system is slightly different in the use of ROIs and involves two techniques, *Technique A* (i.e., *Mass ROIs*), and *Technique B* (i.e., *Whole Mass ROIs*), to study the feasibility of a DBN-based CAD system. These two different ROI selection options are studied to investigate whether they have a significant effect on the performance of the proposed CAD system.

2.1 DDSM Dataset

In this study, we utilize the Digital Database for Screening Mammography (DDSM) to evaluate the proposed system [17]. This dataset is collected by South Florida University and is available online for research purposes [17]. It is collected to represent real breast data with an average size of 3000 × 4800 pixels, resolution of 42 microns, and 16 bits. The DDSM database consists of 2620 cases that are categorized into 43 volumes. Each case involves four breast images, two of them are Mediolateral Oblique (MLO) views and the others are Cranio-Caudal (CC) views of each breast. Benign and malignant masses in all mammograms are recognized and annotated by expert radiologists. In this paper, we utilize 150 mammograms divided equally into normal, benign, and malignant classes.
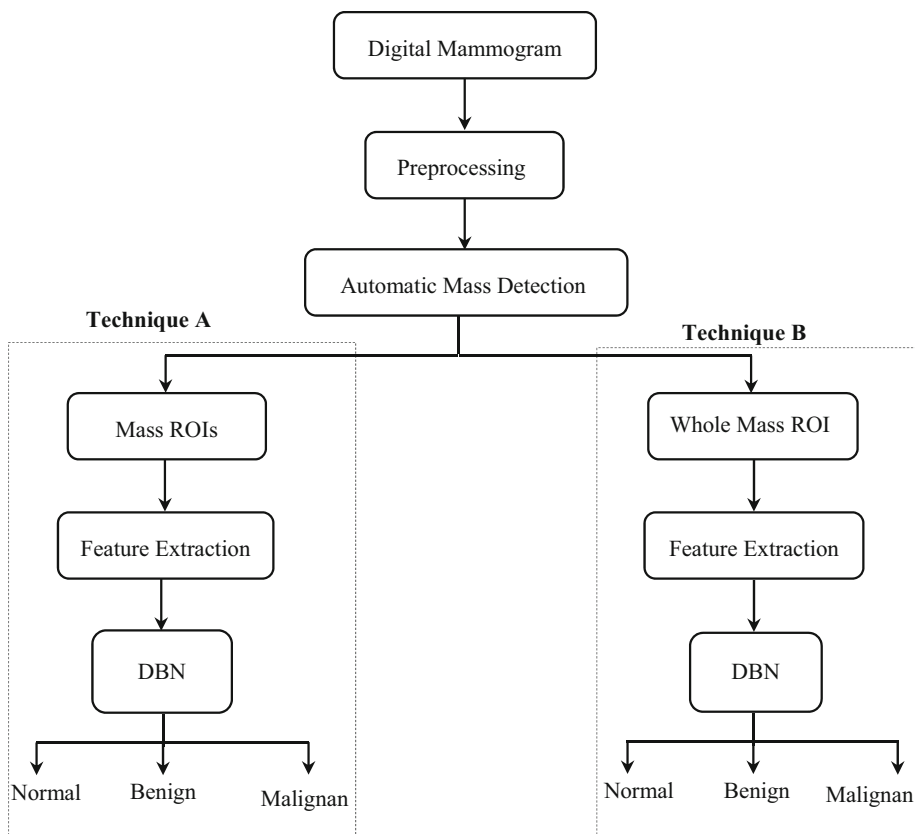
2.2 Preprocessing

Due to breast compression in the scanning process of mammography, breast deformation occurs. The peripheral area of the breast is affected by this compression, which affects the grey level values of breast tissue in these regions [18]. Thus, the intensity values of peripheral areas are always lower than that of the central area. To diagnose correctly, physicians must use certain settings of the window level during inspection of suspicious regions. However, this process can take a long time especially with a huge number of patients, and it is inconvenient. To enhance mammographic images for automatic detection of abnormal masses, we apply a multi-threshold peripheral equalization algorithm [19]. This algorithm enhances and removes irrelevant information from mammograms. The main purpose of this method is to enhance the peripheral area of the mammogram by utilizing multiple thresholds to create multiple images and then average them to produce the smooth transitions between the central and peripheral areas of the mammogram. Thus, physicians can view and inspect the lesions through one window level setting. The algorithm involves the following five sequential steps. First, by utilizing adaptive Otsu thresholding, we segment and separate the breast region from its background to generate a segmented image (I_{seg}) as illustrated in Fig. 2b.

Second, a 2D Gaussian low pass filter (GLPF) is applied to the original mammogram, shown in Fig. 2a, to produce a blurred image (I_{blur}) shown in Fig. 2c. Third, the I_{blur} is multiplied by the segmented image to eliminate the unwanted information that exists outside the breast tissue, as depicted in Fig. 2d. Then, a normalized thickness profile (NTP) of the mammogram is estimated, as shown in Fig. 2e. The NTP is obtained using five threshold values (T_n), as the mean value of the corresponding pixels of the five thresholded images [19]. Each threshold value is computed as follows:

$$T_n = I_{ave} \times F_n; \quad n = 1, 2, \ldots 5, \tag{1}$$

where I_{ave} is the average intensity value of I_{blur}, and F_n is equal to 0.8, 0.9, 1.0, 1.1, and 1.2 and represents the scale parameter used to adjust the threshold value to be around I_{ave}. In this study, we used these five thresholds to increase the intensities of the peripheral regions and eliminate the boundary effect produced if a single threshold is used [19].

To create NTP, we averaged all $\hat{I}_{blur}$ images, which are estimated by rescaling the I_{blur} image according to each threshold value as follows:

Fig. 2 Peripheral density correction using multi-thresholding algorithm. **a** Original mammogram, **b** segmented image (I_{seg}) with adaptive Otsu thresholding, **c** blurred image (I_{blur}), **d** blurred image after multiply by I_{seg}, **e** normalized thickness profile (NTP) of mammogram, and **f** peripheral equalized (I_{peq}) of mammogram

(a) (b) (c)

(d) (e) (f)
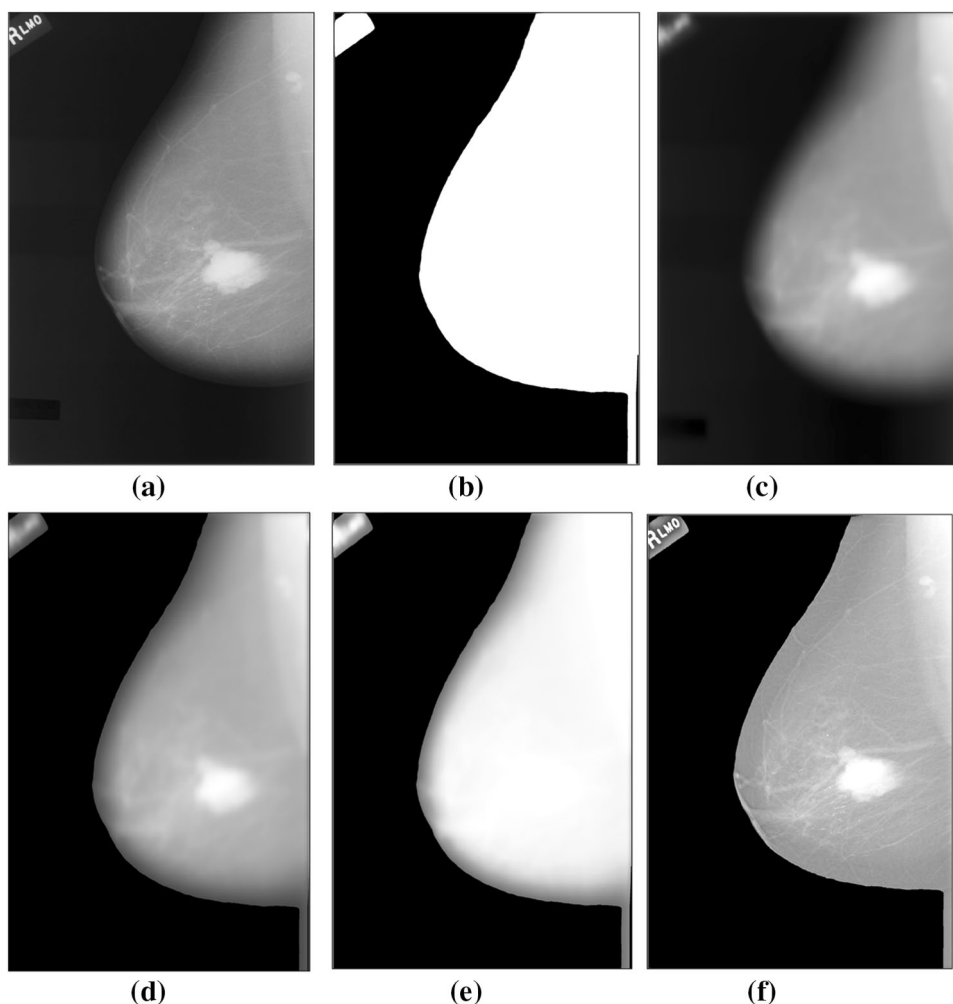
$$\hat{I}_{blur}(i,j) = \begin{cases} \dfrac{I_{blur}(i,j)}{T_n}; & I_{blur}(i,j) \leq T_n \\ 1; & \text{otherwise} \end{cases}. \quad (2)$$

then,

$$NTP = \frac{1}{5} \sum_{n=1}^{5} \hat{I}_{blur}(n), \quad (3)$$

where $i = 1, 2, 3, \ldots, M$, and $j = 1, 2, \ldots, N$, and $M \times N$ is the size of the mammogram. Finally, the peripheral equalized image (I_{peq}) of the mammogram is achieved as follows:

$$I_{peq} = \frac{I_{att}}{(NTP)^r}, \quad (4)$$

where I_{att} is an attenuation image (i.e., the original mammogram, as shown in Fig. 2a). Peripheral equalization of a mammogram is illustrated in Fig. 2f, and r is a constant value in the range of [0.70–1.0] as in [19]. The ratio of signal to noise (SNR) is computed for original and preprocessed mammograms, which are shown in Fig. 2a, f [20], as follows:
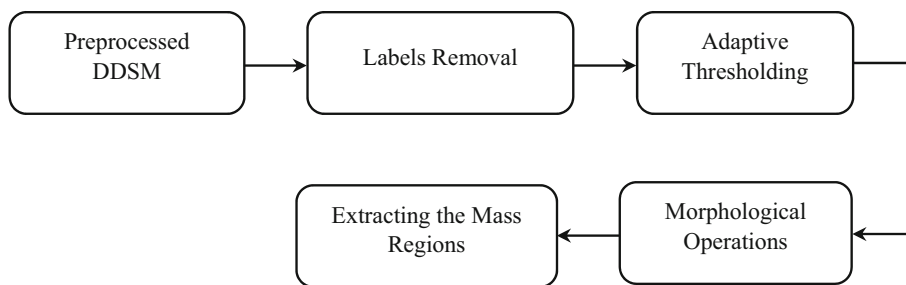
$$SNR \text{ (dB)} = 10 \cdot \log_{10} \left[\frac{\sum_{i=1}^{M} \sum_{j=1}^{N} [I_{att}(i,j)]^2}{\sum_{i=1}^{M} \sum_{j=1}^{N} \left[I_{att}(i,j) - I_{peq}(i,j) \right]^2} \right]. \quad (5)$$

Thus, SNR is estimated to be 17.01 and 17.35 dB when r is equal to 0.7 and 1.0, respectively. So, to achieve our CAD system, we set $r = 1.0$ for all dataset.

2.3 Automatic Mass Detection

One of the important steps in CAD for breast cancer classification is to detect specific masses or suspicious regions on mammograms [21]. In this work, we implemented our automatic mass detection algorithm as shown in Fig. 3. Figure 4a shows a label-removed and extracted breast only image. Figure 4b shows initial suspicious regions identified via the adaptive threshold procedure. We

Fig. 3 Block diagram of our proposed technique for automatic mass detection

Preprocessed DDSM → Labels Removal → Adaptive Thresholding → Morphological Operations → Extracting the Mass Regions

Fig. 4 Proposed algorithm for automatic mass detection. **a** Labels removal, **b** after applying adaptive thresholding, **c** binary image, **d** after applying morphological operations, **e** mass region extraction, **f** mass contour superimposed on the corresponding mammogram

(a)   (b)   (c)

(d)   (e)   (f)
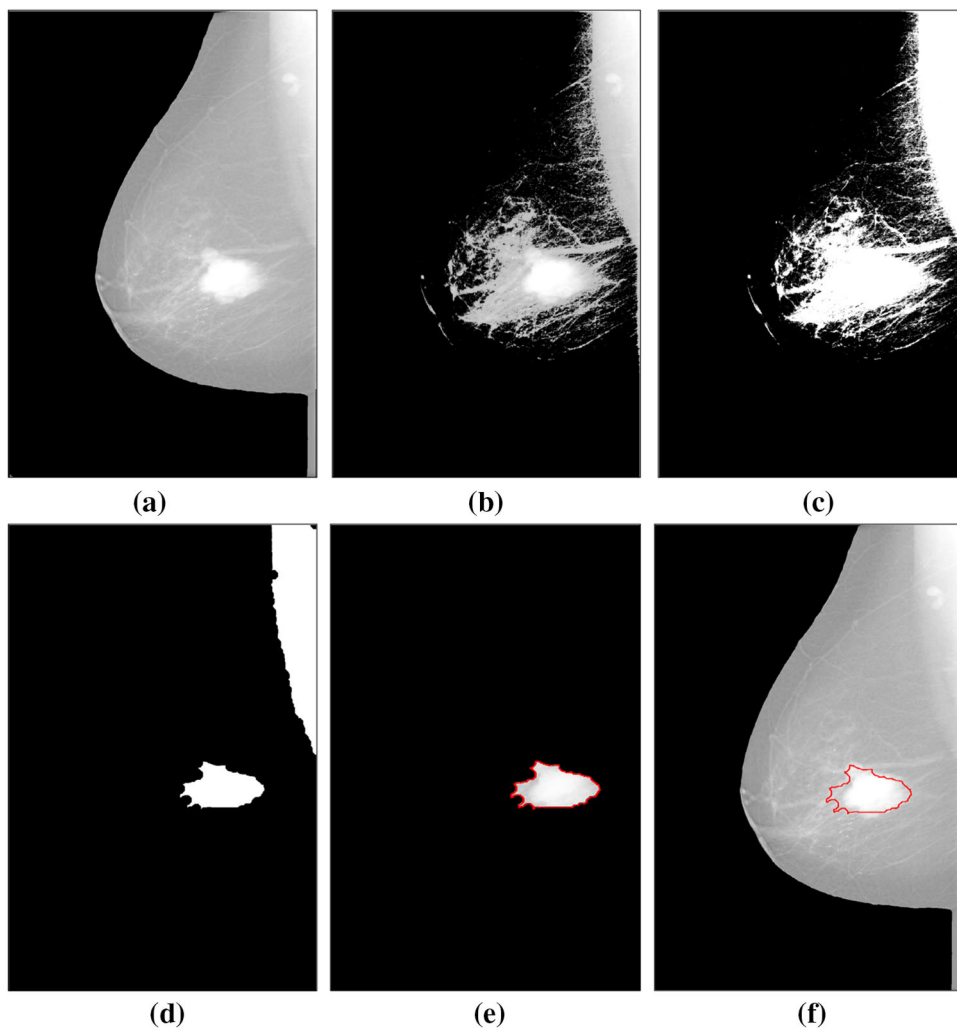
computed the threshold value by aggregating all gray level intensities inside the breast tissues and dividing them by the total number of non-zero pixels (L), which is known as the grayscale intensity of the local background, as follows:

$$T_{thr} = \frac{\sum_{i,j \in I} I(i,j)}{L}; \quad I(i,j) > 0. \tag{6}$$

Then, we applied this threshold (T_{thr}) to the breast image, converting the thresholded image of Fig. 4b to the binary image as in Fig. 4c in order to apply morphological

operations. Consecutive binary morphological operations were applied to determine proper shape and size of the mass. These operations were accomplished using three steps. First, we utilized a fill operation to complete the whole expected suspicious region. Second, erosion with a structuring element of disk type was applied. Finally, we removed the remaining disconnected small areas remaining around the mass after the erosion process. Figure 4d shows the end result of the morphological processing. As it is known that the pectoral muscle exists at the border of the

mammograms in both MLO and CC views. After applying our mass detection algorithm as shown in Fig. 4, the pectoral muscle was disconnected from the mass. Then, we removed the part that is connected to the border which represents the pectoral muscle. Figure 4f shows the whole breast image with the superimposed contour around the extracted mass. In this study, we verified our aforementioned approach on MLO and CC views of 150 mammograms, which were equally divided for each class (i.e., normal, benign, and malignant). In this study, by comparing the detected positions of the mass contour using our method against the ground truth, we achieved an accuracy of 86% as an average of correct mass detection for benign and malignant cases. If the intersection over union ($IOU_{Ground\ truth}^{Extracted}$) between the extracted mass and its ground truth exceeds 50%, we considered the result to be correct. In this study, the next stages of mass classification depended solely on the successfully detected masses. It should be noted that extraction of masses from breast dense regions remains a limitation of our method. Figure 5 shows some sample results compared to the ground truth of the

original mammogram with the contours that drawn manually by expert radiologists during the mass inspection procedure. Some mammograms involved two masses with different sizes, as shown in Fig. 5d. Thus, we collect 56 benign and 56 malignant masses that were extracted correctly. In tumor classification, dense breast tissue remains challenge. So, in order to reduce classification bias, we extracted some normal regions randomly. In total, 56 regions from normal cases were collected.

2.4 ROI Extraction

Once breast masses were detected, we derived two kinds of windows to train and test our proposed CAD system. In *Technique A* (i.e., *Mass ROIs*), we utilized 168 regions (i.e., 56 masses for each benign and malignant region and 56 normal regions). Then, we randomly extracted four non-overlapping ROIs 32×32 pixels in size around the center of each region, extracting 224 ROIs for each class, as shown in Fig. 6a. Thus, a total of 672 ROIs were collected. The size of each ROI was determined in order to obtain
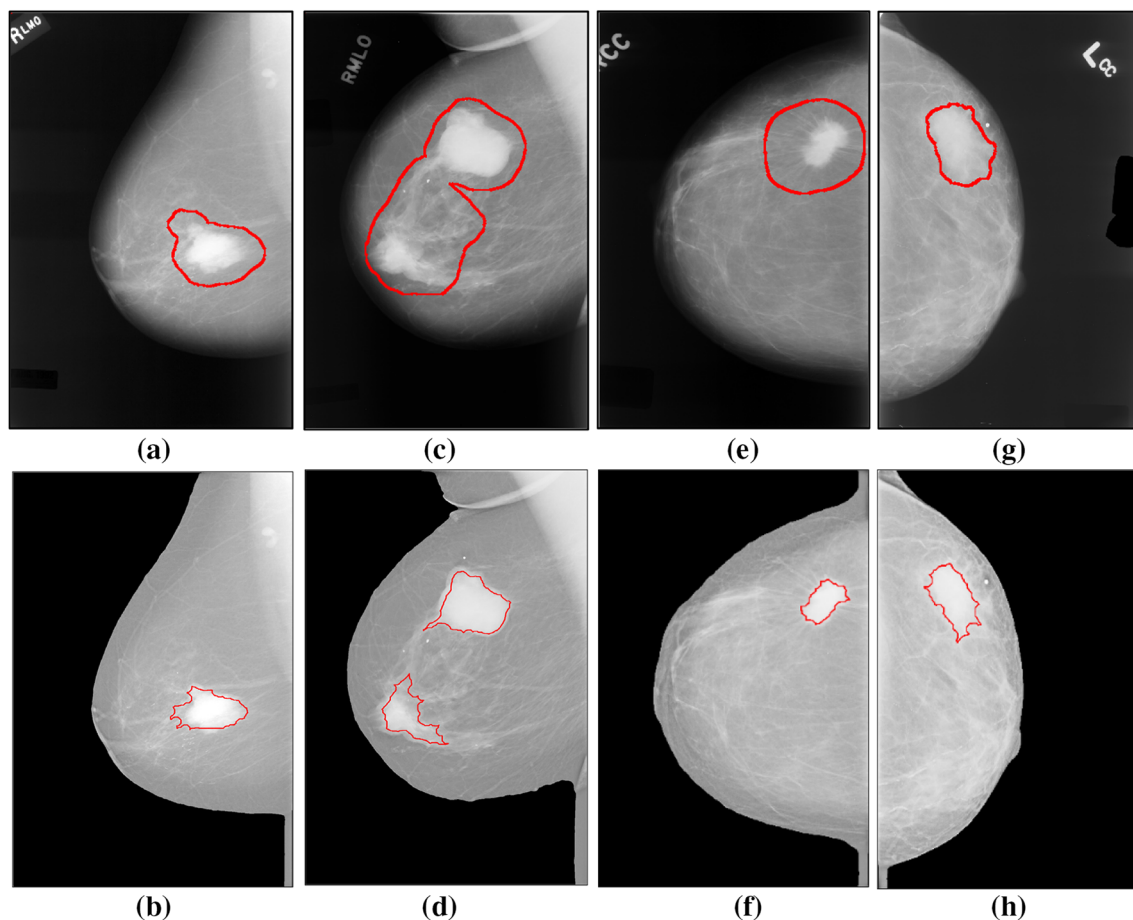
Fig. 5 Examples of proposed automatic mass detection results. The top row shows the ground truth outlined by radiologists while the bottom row illustrates automatically detected masses for the same mammograms

(a)

(b)
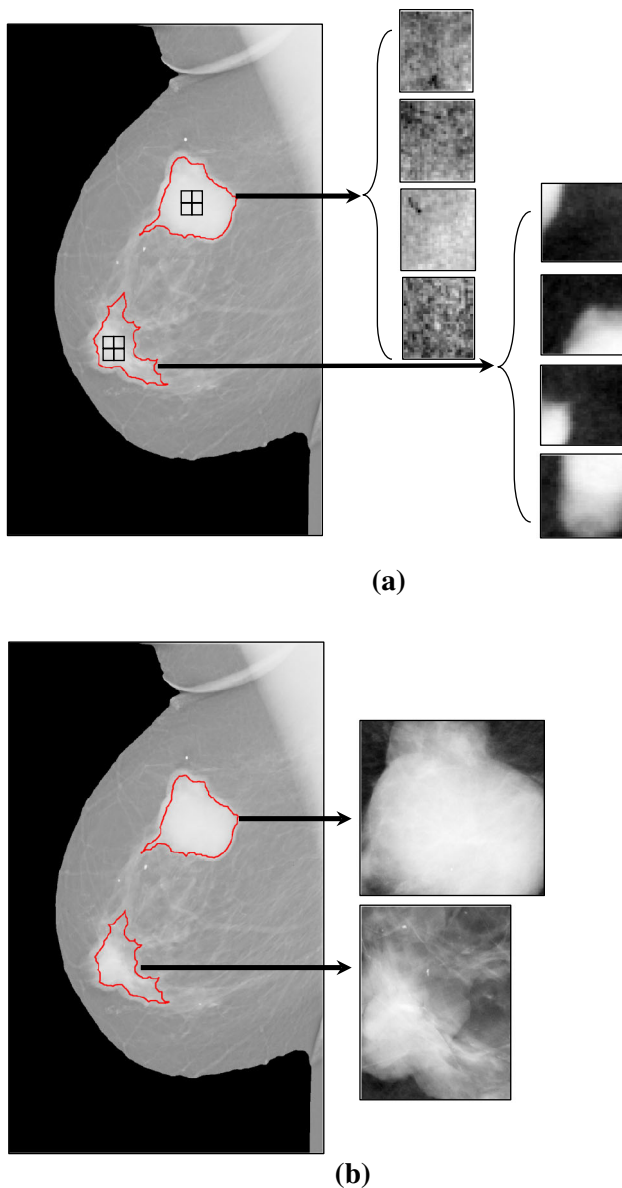
Fig. 6 ROI settings for proposed CAD system. **a** four non-over-lapped ROIs from each mass (i.e., Mass ROIs), and **b** Whole Mass ROIs

reliable statistical features [10, 21, 22] in which the smallest sufficient ROI requires at least 800 pixels, resulting in the ROI size of 32×32 pixels. However, one could try more or less ROIs as long as sufficient number of statistical features can be derived.

In *Technique B* (i.e., *Whole Mass ROIs*), a whole detected mass was utilized as depicted in Fig. 6b. For benign and malignant cases, after extracting the whole ROIs, rectangular boxes were drawn around the irregular shapes of the masses, which numbered 56 benign and 56 malignant. For normal cases, we utilized 56 regions that were manually extracted from normal mammograms. Thus, a total of 168 ROIs were collected.

2.5 Feature Extraction and Selection

The next step was to derive features representing the breast tissue types. These attributes are quantitative measures of breast tissues that are used to describe the salient characteristics of the tissues [23, 24]. In this study, we used the first and higher order statistical features to describe the characteristics of the regions. These statistical features involve both intensity and texture feature types. These various features were utilized in previous studies and were shown to be strong enough to distinguish different lesions [13, 25–30]. A total of 347 statistical features were extracted from the ROIs. After that, we applied some feature selections techniques for the conventional classifiers, but applied the entire features to DBN considering the feature selection capability of deep learning.

2.5.1 First Order Statistical Feature

From the pixels in the ROI, the first order attributes were computed by applying statistical analyses on both grayscale intensities and the histogram of each ROI. Nine features are extracted from the histograms of the ROIs, namely, entropy, modified entropy, standard deviation (SD), modified standard deviation (MSD), energy, modified energy, asymmetry, modified skewness, and range value of the histogram. Other features included mean, SD, smoothness, third moment, entropy, skewness, kurtosis, variance, mode, interquartile range, and percentiles or quintiles at levels 0.1–0.9 [25, 29, 30]. A total of 28 first order features were derived.

2.5.2 Higher Order Statistical Features

To take into consideration the spatial inter-relationships of the pixels as well as their grayscale, the second order attributes were computed on the grayscale co-occurrence matrix (GLCM) as proposed by [31]. The 2D histogram of grayscale intensity for a pair of pixels is called the GLCM. We utilized quantization grayscale (i.e., $L = 32$), angle of orientation (i.e., $\theta = 0°$, $45°$, $90°$, and $135°$), and displacement vector (i.e., $d = 1, 3, 5$, and 9) to create GLCMs [26–28, 32]. The value of d is acceptable in the range of 1–10 [30]. From each value of d, we estimated four GLCMs according to θ, resulting in 16 different GLCMs. Then, we extracted 19 different statistical features from each one, for a total of 304 statistical features. These features included energy, contrast, correlation, homogeneity, entropy, maximum probability, inverse different moment (IDM), variance, sum average, sum entropy, sum variance, difference entropy, difference variance, autocorrelation, dissimilarity, cluster shade, cluster prominence, correlation information 1, and correlation information 2 [21, 24]. On

the other hand, some previous studies have demonstrated that d = 1 and d = 2 for GLCM provided good results of overall accuracy [31, 32]. Thus, we estimated four GLCMs at d = 2 with respect to θ and then averaged them to calculate the average GLCM. Then, we extracted 15 features from this average GLCM. These features are seven invariant moments, entropy, maximum probability, homogeneity, IDM, variance, uniformity, correlation information 1, and correlation information 2 [31]. Finally, a total of 319 higher order features were derived. All extracted features were normalized in the range of [0, 1].

2.5.3 Feature Selection Techniques

To reduce the redundancy in the features and to select the most prominent features, we utilized four feature selection algorithms: sequential backward (SBS), sequential forward (SFS), sequential floating forward (SFFS), and branch and bound (BBS) as described in [33].

2.6 Classifier Designs

The selected features were used to train and test the conventional classifiers of linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and neural network (NN). LDA is used to find the linear combination of features that best separates two or more classes of objects [34, 35]. QDA is a general discriminant function with a quadratic decision boundary that can be used to classify datasets with two or more classes [34, 35]. In this study, we present a 3-class classification problem with multi-dimensional features. Therefore, the task of LDA and QDA is to classify 3 groups with an input of multi-dimensional features. It is known that neural network (NN) is usually used to indicate the feedforward neural networks, while deep neural network (DNN) represents the feedforward NN with many layers [16]. In comparison to the conventional artificial neural network (ANN), deep belief network (DBN) has a slightly different structure. DBN utilizes restricted Boltzmann machines (RBM) in which the layers can be trained using the unsupervised learning algorithm contrastive divergence (CD). This work was devised by Hinton et al. [16] and explained that classification could be achieved in deeper architectures when each layer (RBM) was pre-trained with an unsupervised learning algorithm (i.e., CD algorithm). Then the Network can be trained in a supervised way using backpropagation in order to "fine-tune" the weights. Thus, a DBN-based CAD system can deal with all extracted features without the need for selected algorithms. Therefore, no feature selection method was used, and all features were used for the training and testing of DBN.

2.7 Deep Belief Network (DBN)

Deep belief network (DBN) based on RBM was recently established by [16]. RBM is considered a model of a generative stochastic neural network with connections only between visible nodes (v) and hidden nodes (h). In RBM, there are no connections between nodes or units in the same layer; hence, Gibbs sampling is performed [36]. DBN has a deep structure that generates a prominent model with multiple layers of RBM, as shown in Fig. 7. To develop our DBN for both techniques, we utilized R and Q hidden nodes with first and second hidden layers, respectively. The training of DBN was achieved through two consecutive processes. First, a pre-training process for RBM was performed via unsupervised learning. Then, we utilized a supervised learning method to apply a back propagation algorithm with known labels of breast cancer features to adjust the weights and fine-tune the networks. In DBN, with training data, pre-training assists the neural networks to overcome any problems of over-fitting.

One of the benefits of using DBN is the ability to extract and select the more prominent features from the input data. Figure 7b shows three layers of RBM for both of our techniques. Each layer of RBM is updated depending on the previous one. Once the first layer is prepared by computing the weight matrix, it is considered as an input for the next layer. This process trains RBMs one after another and utilizes their extracted features for learning in the next one. So, the input data during this process is reduced layer by layer. Thus, the selected features at the hidden nodes of the last layer can be considered as a vector of features. In both structures of our DBN, the algorithm of contrastive divergence (CD) with block Gibbs sampling is utilized to update the matrix of weights $\mathbf{w}$ layer by layer [35–37]. In fact, there are five consecutive steps to train RBM. First, all parameters of the network are initialized and set to zero. These parameters are weight matrix $\mathbf{w}$ and two real values of bias vectors (i.e., $\mathbf{A}$ and $\mathbf{B}$) for hidden and visible layers, respectively. Second, the logical state of the first hidden layer is computed as follows:

$$h_1 = \begin{cases} 1; & f(\mathbf{B} + v_1\mathbf{w}^T) > \varphi \\ 0; & \text{otherwise} \end{cases} \tag{7}$$

where $f(z) = 1/(1 + e^{-z})$ is a sigmoid activation function, and φ is an activation threshold. Third, after h_1 is obtained, the state of visible layer v_{recon} is reconstructed corresponding to the following formula:

$$v_{recon} = \begin{cases} 1; & f(\mathbf{A} + h_1\mathbf{w}) > \varphi \\ 0; & \text{otherwise} \end{cases}. \tag{8}$$

The fourth step is to compute the state of hidden layer h_{recon} as follows using v_{recon},

Fig. 7 **a** Structure of our proposed DBN based CAD system for both techniques, **b** pre-training process, and **c** back projection algorithms

(a)

(b)   (c)
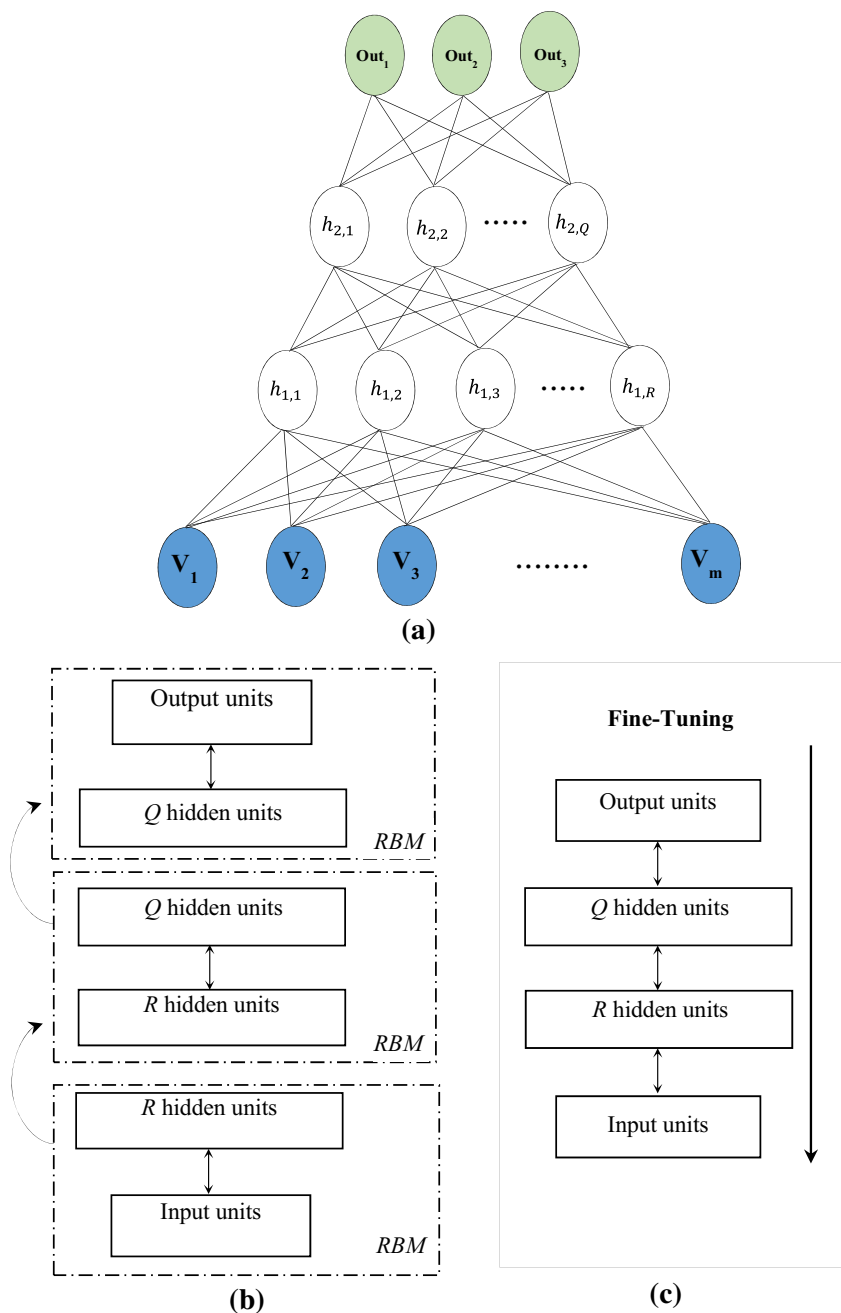
$$h_{recon} = f\left(\mathbf{B} + v_{recon}\mathbf{w}^{T}\right). \qquad (9)$$

Finally, the difference of weight $\Delta\mathbf{w}$ is estimated to compute the current one as follows:

$$\mathbf{w_{k+1}} = \Delta\mathbf{w} + \mathbf{w_k},$$

$$\Delta\mathbf{w} = \left(\frac{h_1 v_1}{\eta}\right) - \left(\frac{h_{recon} v_{recon}}{\eta}\right), \qquad (10)$$

where η is the batch size.

As shown in Fig. 8b–d, v_1 represents the visible layer for each RBM during the pre-training process. This means that, after the first RBM is tuned, the hidden layer of this RBM is considered as the visible layer for the next one and so on. All of these steps are iterated until the number of batches is converged. Figure 8 presents the pre-training process of DBN utilizing the CD algorithm, where both red boxes and arrows demonstrate the current network layers, while blue ones indicate the derived weight and states. Figure 8e shows the learning process of DBN via back propagation, which includes pre-training and fine-tuning of all parameters. For both techniques, all extracted features (i.e., 347 features) were directly utilized as input for DBN through m visible units (i.e., $V_m = V_1, V_2, \ldots V_{347}$), as illustrated in Fig. 7a. In Technique A (i.e., Mass ROIs),

Fig. 8 **a** Proposed DBN layers. **b–d** Schematic diagram of Contrastive Divergence (CD) technique for greedy layer-wise training, and **e** learning process of DBN via back propagation algorithm
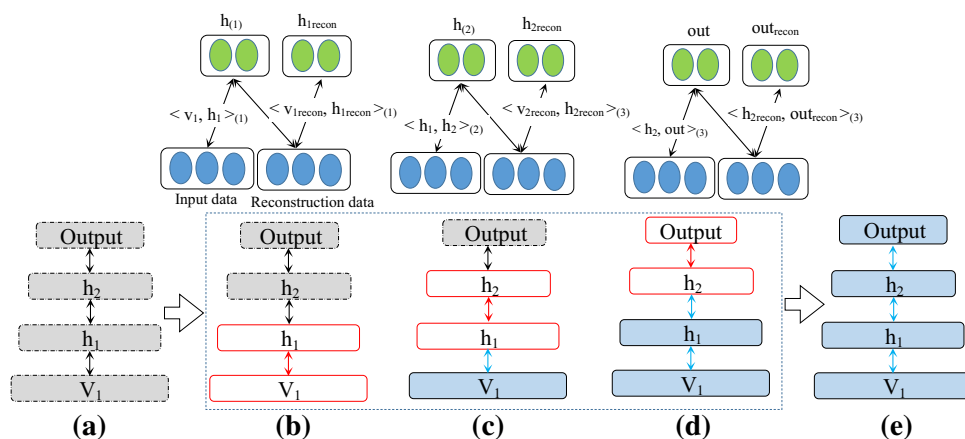
two hidden layers were used with R = 15 and Q = 8 hidden nodes for RBM and the batch size η of 10. In Technique B (i.e., Whole Mass ROIs), we also utilized two hidden layers with R = 20 and Q = 8 hidden nodes and η = 7 for RBM. For both techniques, batch size η of DBN was empirically obtained by trying different values (i.e., trail-and-error based) in order to obtain the best accuracy, as in [36]. In this study, the overall accuracy was derived from 2-fold cross-validation.

2.8 Evaluation of CAD system performance

To evaluate our proposed CAD-based DBN system, we utilized overall accuracy and receiver operator characteristic (ROC) curve with its area under the curve (AUC) [12–14, 23]. The ROC curve presents a trade-off between sensitivity and specificity, where the high rate of classification is investigated when the AUC satisfies a set threshold [13]. The ROC curve is defined based on sensitivity and specificity as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \tag{11}$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \tag{12}$$

where sensitivity represents the ability to measure disease appearance as abnormal, while specificity represents the ability to measure absence of disease as normal. Overall accuracy represents the ability of the system to distinguish between the different classes (i.e., normal, benign, or malignant) and is defined as follows:

$$\text{Overall accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{13}$$

where true positive (TP) and true negative (TN) represent the numbers of studies that are classified correctly as positives and negatives, respectively. False positive (FP) indicates the negative studies that are incorrectly distinguished as positives. False negative (FN) indicates the positive studies that are incorrectly classified as negatives. In this study, we utilized the pairwise comparison method between each two classes to compute all of these metrics, as presented in [38].

3 Results and Discussion

For data preparation in this work, we randomly split the total cases into training and testing datasets. That is, a total of 168 whole cases (or masses) were randomly split into a training dataset (84 cases) and a test dataset (84 cases).

In Technique A (i.e., Mass ROIs), we extracted four ROIs from each mass in the training dataset and collected a

Table 1 Accuracy of the proposed DBN based CAD system against other conventional classifiers

Feature selection method	Technique A: mass ROIs				Technique B: whole mass ROIs			
	LDA	QDA	NN	DBN	LDA	QDA	NN	DBN
SBS	82.14	80.33	82.57		75.00	65.48	80.95	
SFS	77.08	72.02	83.12		78.57	75.00	84.52	
SFFS	78.57	72.46	83.85	**92.86**	78.57	76.19	84.52	**90.48**
BBS	72.25	77.38	77.68		75.00	73.81	78.57	
All features	74.11	69.33	79.17		63.10	62.5	79.52	

Table 2 Confusion matrices of DBN based CAD system for both techniques

|	Technique A: mass ROIs			Technique B: whole mass ROIs		
	Normal	Benign	Malignant	Normal	Benign	Malignant
Normal	**112**	0	0	**28**	0	0
	100%	0.0%	0.0%	**100%**	0.0%	0.0%
Benign	4	**99**	9	0	**25**	3
	3.57%	**88.39%**	8.04%	0.0%	**89.29%**	10.71%
Malignant	0	11	**101**	0	5	**23**
	0.0%	9.82%	**90.18%**	0.0%	17.86%	**82.14%**

Table 3 Performance analysis of DBN based CAD system

Metrics (%)	Technique A: mass ROIs	Technique B: whole mass ROIs
AUC_{NB}	98.23	99.60
AUC_{NM}	100	100
AUC_{BM}	93.54	86.56
Overall accuracy	**92.86**	**90.48**

total of 336 ROIs for training. For testing, we also extracted four ROIs from each mass in the test dataset and classified each mass. In Technique B (i.e., Whole Mass ROIs), we trained the classifiers with the features from the whole mass in the training dataset. After training, we classified each mass from the test dataset. In both techniques, training and testing sets contained equal numbers of cases for each class. In this study, all results were derived as the average of 2-fold cross-validation with all classifiers.
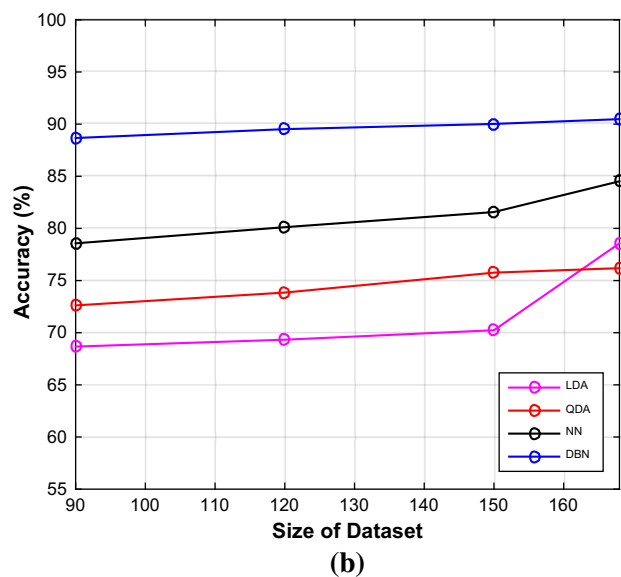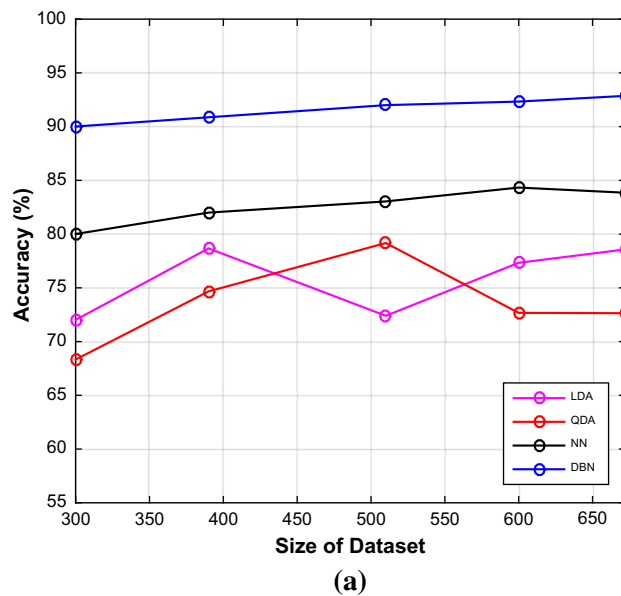
Our two techniques using a CAD system based on DBN were evaluated in terms of overall accuracy and area under the ROC curves. All classifiers are used to classify three kinds of breast tissues of normal, benign, and malignant. The conventional classifiers (i.e., LDA and QDA) depend on the selection feature algorithms to select the most prominent features. The main purpose of using feature selection techniques is to reduce the dimension of features to make the classifiers strong enough for distinguishing between the different lesions of breast tissues. On the other hand, the NN classifier has a good structure to deal with all features, so its performance does not change much with selection algorithms. On the contrary, we directly utilized all extracted features without any dimension reduction as an input of DBN. Overall accuracies of our DBN-based CAD system compared with other classifiers in both techniques are reported in Table 1. The performance of each conventional classifier varied corresponding to the utilized feature selection algorithm. In Technique A, the performance of NN with SFFS was better than those of LDA and QDA, with an accuracy of 83.85%. On the other hand, the performances of QDA and LDA with SBS were better than those of other selection algorithms. On the contrary, DBN achieved a much higher classification rate with an accuracy of 92.86%. In Technique B, DBN again

showed better performance than all other classifiers, with an accuracy of 90.48%. In this technique, all conventional classifiers provided high accuracy with SFFS and SFS algorithms. The performance of QDA with SFFS was slightly increased by 1.19% compared with SFS. From the results in Table 1, we conclude that the performance of DBN in both techniques provides higher accuracy than the others methods. The performance of DBN in Technique A was slightly higher by 2.38% than in Technique B. For conventional classifiers, the NN classifier showed better performance than LDA and QDA for both techniques. This indicates that our DBN-based CAD is preferable to the other classifiers to obtain a highly accurate rate of breast cancer diagnosis. The confusion matrices of our three classes for DBN are shown in Table 2. In Technique A, the error (i.e., misclassification) of mixing benign with normal cases was 3.57%, while the error of mixing malignant with benign was 17.86%. In Technique B, there was no mixing error for benign with normal, but the mixing rate error for confusing benign with malignant was 28.57%. Meanwhile, the ROC curve is considered a detection tool to verify the performance of the CAD system. Due to the misclassification rate between normal and benign for both techniques, the AUC_{NB} (i.e., AUC for ROC curve between 1-specificity for normal and sensitivity for benign) in Technique A was slightly lower than that in Technique B, as reported in Table 3. Due to the false rates of classification between benign and malignant, as shown in Table 2, the values of AUC_{BM} (i.e., AUC for ROC curve between 1-specificity for benign and sensitivity for malignant) were 93.54% and 86.56% for Techniques A and B, respectively. Therefore, the average 97.26% of all AUCs in Technique A was higher than the 95.39% for Technique B. Figure 9 shows an example of ROC curves for DBN against other
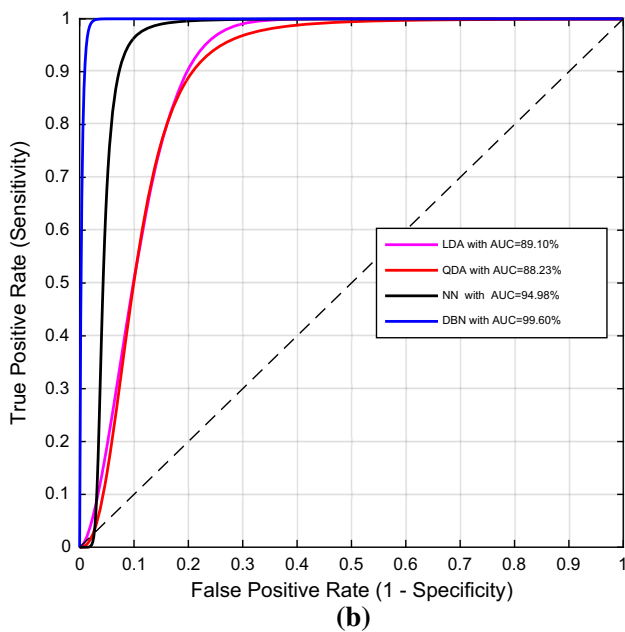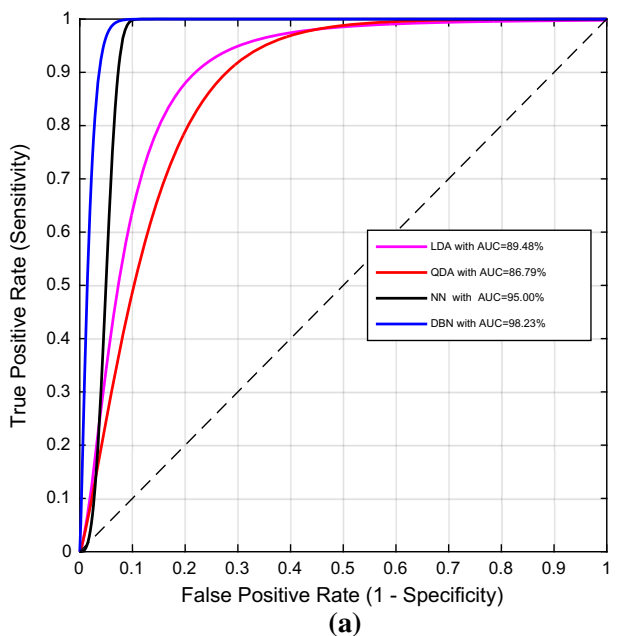
Fig. 9 ROC curves of CAD system for normal against benign classes in **a** Technique A and **b** Technique B

Fig. 10 Comparison of the accuracies between DBN against the conventional classifiers with different sizes of dataset in **a** Technique A and **b** Technique B

classifiers corresponding to 1-specificity for normal and sensitivity for benign and the AUCs (i.e., AUC_{NB} for all classifiers) for both techniques. ROC curves of the conventional classifiers are depicted with the most robust feature selection algorithm of SFFS. Both Techniques A and B with DBN provided much higher AUCs compared with other classifiers. This indicates that a DBN-based CAD system can distinguish different breast abnormalities with higher probability. Figure 10 shows the capability of DBN along with other conventional classifiers with respect

to different numbers of datasets in terms of accuracy and stability, as in [14]. These results show that the performance of DBN improves slightly as the size of the dataset increases. We confirmed the stability of DBN and some improvements in the performance with an increase in dataset size against the conventional classifiers. On the other hand, a much larger dataset on a more powerful machine would likely provide better characterization of the clusters and hence better performance. The presented results of our proposed CAD system utilizing DBN show the feasibility of distinguishing different breast tissues. The usability of deep learning techniques such as DBN or CNN

has been validated for various medical applications such as prostate cancer detection [39], breast cancer detection [13, 14], and discriminative cue integration for medical image annotation [21]. In this work, we developed a DBN-based CAD system to distinguish among three different breast tissue abnormalities, an improvement over previous studies [40] and [41] differentiating only two tissue types (i.e., benign against malignant). Currently, we are investigating the application of a convolutional neural network (CNN) for even better performance with a CAD system [13–15, 39–43].

4 Conclusions

In this paper, we present a DBN-based CAD system to classify among three classes of breast tissues, which are normal, benign, and malignant. We present an automatic mass detection algorithm to identify suspicious masses on mammograms. Then, statistical features are derived from the detected masses. After that, we utilize DBN to investigate its potential for a CAD system with breast tissues. The results of the DBN-based CAD system demonstrate significantly improved performance compared to previous conventional CAD systems. The results indicate that CAD systems with deep learning capability offer great potential for computer-aided detection of breast cancers. For the two proposed techniques, there was a slight difference in performance. However, Technique A could be preferable for microcalcifications' problems, especially when the expected target is relatively small in size. Otherwise, Technique B could be more preferable when the entire mass is considered.

Compliance with Ethical Standards

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Cheng, H., & Cui, M. (2004). Mass lesion detection with a fuzzy neural network. *Pattern Recognition, 37*(6), 1189–1200.
2. Jalalian, A., Mashohor, S., Mahmud, H., Saripan, M., Rahman, A., Ramli, B., et al. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: A review. *Clinical Imaging, 37*(3), 420–426.
3. Verma, B. (2008). Novel network architecture and learning algorithm for the classification of mass abnormalities in digitized mammograms. *Artificial Intelligence in Medicine, 42*(1), 67–79.
4. Grau, A., Vives, J., Font, A., & Ochoa, C. (2016). Positive psychological functioning in breast cancer: An integrative review. *The Breast, 27,* 136–168.
5. Peng, Y., Yao, B., & Jiang, J. (2006). Knowledge-discovery incorporated evolutionary search for microcalcification detection in breast cancer diagnosis. *Artificial Intelligence in Medicine, 37*(1), 43–53.
6. Junior, B., Rocha, D., Gattass, M., Silva, C., & Paiva, A. D. (2013). A mass classification using spatial diversity approaches in mammography images for false positive reduction. *Expert Systems with Applications, 40*(18), 7534–7543.
7. Oliveira, F., Filho, A., Silva, A., Paiva, A., & Gattass, M. (2015). Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM. *Computers in Biology and Medicine, 57,* 42–53.
8. Al-antari, M., AL-masni, M., & Kadah, Y. (2017). Hybrid model of computer-aided breast cancer diagnosis from digital mammograms. *Journal of Scientific and Engineering Research, 4*(2), 114–126.
9. Balanica, V., Dumitrache, I., & Preziosi, L. (2013). Breast cancer diagnosis based on spiculation feature and neural network techniques. *International Journal of Computers Communications & Control, 8*(3), 354–365.
10. Al-Olfe, M., Al-Akwaa, F., Mohamed, W., & Kadah, Y. (2010). Computer-aided diagnosis of digital mammography images using unsupervised clustering and biclustering techniques. *Proceedings SPIE conference on Medical Imaging, 7624,* 1–6.
11. Kang, Y., Ke, L., & Mu, N. (2010). Mass computer-aided diagnosis method in mammogram based on texture features. In *Proceedings of IEEE BMEI Conference on biomedical engineering and informatics* (BMEI) (vol. 1, pp. 354–357).
12. Sharma, S., & Khanna, P. (2015). Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM. *Journal of Digital Imaging, 28*(1), 77–90.
13. Arevalo, J., González, F., Ramos-Pollan, R., Oliveira, J., & Lopez, M. (2016). Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer Methods and Programs in Biomedicine, 127,* 248–257.
14. Jiao, Z., Gao, X., Wang, Y., & Li, J. (2016). A deep feature based framework for breast masses classification. *Neurocomputing, 167,* 221–231.
15. Kooi, T., Gubern A., Mordang J., Mann R., Pijnappel R., Schuur K, Heeten, A., & Karssemeijer N. (2016). A comparison between a deep convolutional neural network and radiologists for classifying regions of interest in mammography. In A. Tingberg et al. (Eds.), *IWDM 2016, LNCS 9699* (pp. 51–56). Switzerland: Springer.
16. Hinton, G., Osindero, S., & The, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*(7), 1527–1554.
17. Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, W. (2001). The digital database for screening mammography. In *Proceedings of the 5th international workshop on digital mammography* (pp. 212–218).
18. Tortajada, M., Oliver, A., Marti, R., Ganau, S., Tortajada, L., Sentis, M., et al. (2014). Breast peripheral area correction in digital mammograms. *Computers in Biology and Medicine, 50,* 32–40.
19. Wu, T., Moore, R., & Kopans, D. (2010). Multi-threshold peripheral equalization method and apparatus for digital mammography and breast Tomosynthesis. *Google patents*, US Patent 7, 764, 823.

20. Gonzalez, R., & Woods, R. (2008). *Digital image processing* (3rd ed.). USA: Prentice Hall. ISBN 013168728.

21. Tommasi, T., Orabona, F., & Caputo, B. (2008). Discriminative cue integration for medical image annotation. *Pattern Recognition Letters, 29,* 1996–2002.

22. Kadah, Y., Farag, A., Zurada, J., Badawi, A., & Yousef, A. (1996). Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images. *IEEE Transactions on Medical Imaging, 15*(4), 466–478.

23. Verma, B., McLeod, P., & Klevansky, A. (2010). Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer. *Expert Systems with Applications, 37*(4), 3344–3541.

24. Nanni, L., Brahnam, S., & Lumini, A. (2012). A very high performance system to discriminate tissues in mammograms as benign and malignant. *Expert Systems with Applications, 39*(2), 1968–1971.

25. Bueno, G., Vallez, N., Deniz, O., Esteve, P., Rienda, M., Arias, M., et al. (2011). Automatic breast parenchymal density classification integrated into a CADe system. *International Journal of Computer Assisted Radiology and Surgery, 6*(3), 309–318.

26. Ramos, R., Nascimento, M., & Pereira, D. (2012). Texture extraction: An evaluation of ridgelet, wavelet and co-occurrence based methods applied to mammograms. *Expert Systems with Applications, 39,* 11036–11047.

27. Vallez, N., Bueno, G., Deniz, O., Dorado, J., Seoane, J., Pazos, A., et al. (2014). Breast density classification to reduce false positives in CADe systems. *Computer Methods and Programs in Biomedicine, 113,* 569–584.

28. Kumar, I., Virmani, J., & Bhadauria, H. (2015). A Review of Breast Density Classification Methods. In *Proceedings of IEEE international conference on computing for sustainable global development* (*INDIACom*) (pp. 1960–1967).

29. Zhang, P., Verma, B., & Kumar, K. (2005). Neural vs. statistical classifier in conjunction with genetic algorithm based feature selection. *Pattern Recognition Letters, 26*(7), 909–919.

30. Dheeba, J., Singh, N., & Selvi, S. (2014). Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of Biomedical Informatics, 49,* 45–52.

31. Haralick, R., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, 3*(6), 610–662.

32. Ou, X., Pan, W., & Xiao, P. (2014). In vivo skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM). *International Journal of Pharmaceutics, 460,* 28–32.

33. Duin, R., Juszczak, P., Paclik, P., Pekalska, E., Ridder, D., Tax, D., & Verzakov, S. (2007). PRTools4.1: A Matlab toolbox for pattern recognition. Delft: Delft University of Technology.

34. Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, 10*(2), 159–203.

35. Thomas, M. C. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers, 14*(3), 326–334.

36. Fischer, A., & Igel, C. (2014). Training restricted Boltzmann machines: An introduction. *Pattern Recognition, 47*(1), 25–39.

37. Nam, S., Park, S., Park, J., & Kim, T-S. (2015). A single depth sensor based human activity recognition via deep belief network. In *Proceedings of* WCSET *of the 4th world conference on applied sciences, engineering & technology* (pp. 15–19).

38. Sampat, M., Patel, A., Wang, Y., Gupta, S., Kan, C., Bovik, A., et al. (2009). Indexes for three-class classification performance assessment—An empirical comparison. *IEEE Transactions on Information Technology in Biomedicine, 13*(3), 300–312.

39. Litjens, G., Sanchez, C., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., et al. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports, 6,* 1–11.

40. Mavroforakis, M., Georgiou, H., Dimitropoulos, N., Cavouras, D., & Theodoridis, S. (2006). Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers. *Artificial Intelligence in Medicine, 37*(2), 145–162.

41. Rojas-Domínguez, A., & Nandi, A. (2009). Development of tolerant features for characterization of masses in mammograms. *Computers in Biology and Medicine, 39*(8), 678–688.

42. Tome, D., Monti, F., Baroffio, L., Bondi, L., Tagliasacchi, M., & Tubaro, S. (2016). Deep convolutional neural networks for pedestrian detection. *Signal Processing: Image Communication.* doi:10.1016/j.image.2016.05.007.

43. Wang, Y., Zhou, S., Gong, Y., & Zheng, N. (2016). Incorporating image priors with deep convolutional neural networks for image super-resolution. *Neurocomputing, 194,* 340–347.