

SCIENTIFIC REPORTS



OPEN

Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning

Jinhua Wang^{1,*}, Xi Yang^{2,*}, Hongmin Cai², Wanchang Tan¹, Cangzheng Jin¹ & Li Li³

Received: 11 February 2016

Accepted: 13 May 2016

Published: 07 June 2016

Microcalcification is an effective indicator of early breast cancer. To improve the diagnostic accuracy of microcalcifications, this study evaluates the performance of deep learning-based models on large datasets for its discrimination. A semi-automated segmentation method was used to characterize all microcalcifications. A discrimination classifier model was constructed to assess the accuracies of microcalcifications and breast masses, either in isolation or combination, for classifying breast lesions. Performances were compared to benchmark models. Our deep learning model achieved a discriminative accuracy of 87.3% if microcalcifications were characterized alone, compared to 85.8% with a support vector machine. The accuracies were 61.3% for both methods with masses alone and improved to 89.7% and 85.8% after the combined analysis with microcalcifications. Image segmentation with our deep learning model yielded 15, 26 and 41 features for the three scenarios, respectively. Overall, deep learning based on large datasets was superior to standard methods for the discrimination of microcalcifications. Accuracy was increased by adopting a combinatorial approach to detect microcalcifications and masses simultaneously. This may have clinical value for early detection and treatment of breast cancer.

Breast cancer is the most common cancer in women worldwide and the second leading cause of female cancer deaths. Mammography is the most efficient method for screening breast cancer and can reduce breast cancer mortality^{1–3}. One of the main early symptoms on mammograms is the appearance of microcalcifications, whose diameter range from 0.1 to 1 mm^{2,4–6}. Early detection and accurate identification of malignant microcalcifications can facilitate early detection, diagnosis and timely treatment of breast cancer^{2,5,7}. However, due to the small size and low contrast compared to the background of images, it is difficult and time for radiologists to make objective and accurate evaluation of microcalcifications^{8–11}. The problem is especially challenging for inexperienced radiologists when facing enormous numbers of mammograms generated in widespread screening^{9–11}. Consequently, there is a need to develop helpful automated tools to overcome these problems and improve diagnostic performance of breast cancer. Advances in computer technologies allow comprehensive and objective analysis of diagnostic features in microcalcifications and masses^{12–14}. Meanwhile, by efficiently analyzing large numbers of images, computer-based methods can minimize intra- and inter-observer performance variability^{12–15}. Through the automatic identification and classification of microcalcifications, computer-based methods can be proposed to aid early detection and diagnosis^{13,15}.

A wide variety of machine learning classifiers have been developed for early diagnosis of breast cancer^{16–18}. The widely used techniques are based on support vector machines (SVM)^{18–20}, k-nearest neighbor (KNN) method^{21,22} and linear discriminant analysis (LDA)^{23,24}. However, the discriminative power of these methods is limited due to the computational costs of identifying definitive features for subset characterization and optimization. Deep learning is a relatively new method in the field of artificial intelligence and machine learning technologies^{25–31}. This approach has achieved considerable successes in multiple applications, including medical research. Deep convolutional neural networks were employed to medical image classification³¹; deep belief nets and active learning were presented for multi-level gene and miRNA feature selection²⁵; convolutional neural networks were used

¹Department of Radiology, Affiliated Nanhai Hospital of Southern Medical University, Foshan 528200, Guangdong, China. ²School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, Guangdong, China. ³Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, Guangdong, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to H.C. (email: hmcai@scut.edu.cn) or L.L. (email: li2@mail.sysu.edu.cn)

	Training group		Testing group	
	Number	Percentage	Number	Percentage
Malignant lesions	323	32.3	107	52.5
Invasive ductal carcinoma	222	22.2	86	42.2
Intraductal carcinoma	8	0.8	2	0.9
Ductal carcinoma <i>in situ</i>	85	8.5	18	8.8
Mucinous carcinoma	4	0.4	0	0
Others	4	0.4	1	0.5
Benign lesions	677	67.7	97	47.6
Fibroadenoma	71	7.1	11	5.3
Fibrocystic changes	491	49.1	58	28.4
Intraductal papilloma	11	1.1	2	0.9
Hyperplasia	6	0.6	2	0.9
Phyllodes tumor	8	0.8	0	0
Inflammation	2	0.2	1	0.5
Follow-up	88	8.8	23	11.3
Microcalcifications only	623	62.3	110	53.9
Masses only	221	22.1	35	17.2
Microcalcifications and masses	156	15.6	59	28.9

Table 1. Distributions of histopathological characteristics of breast lesions for both groups. Note—The follow-up period was at least two years. Only both systematic clinical examination and mammogram showed no malignant findings to the suspicious benign-appearing lesion in this period, can the patient be admitted into the benign group.

to demonstrate an explicit gradient for feature complexity in the ventral pathway of the human brain²⁶; deep learning was applied to determine the sequence specificities of DNA and RNA-binding proteins for identifying causal disease variants²⁷; superpixel and deep learning were used for automatic vaginal bacteria segmentation and classification²⁸; some deep learning-based latent feature representations are proposed for diagnosis of Alzheimer's disease and its prodromal stage, mild cognitive impairment (MCI), such as stacked auto-encoder and deep boltzmann machine^{32,33}. However, only few works have explored deep learning methods to address the automatic classification of identified lesions on mammography. A nice learning framework for breast cancer diagnosis in mammography by convolutional neural networks was reported³⁴. The tested data were preprocessed images. A convolutional sparse autoencoder was proposed for mammographic texture scoring³⁵.

Deep learning comprises a neural network with multiple hidden layers that enhances the recognition accuracy of images, audio and other data types; thereby increasing its versatility for capturing representative features. Deep learning outperforms other state-of-the-art methods in many areas and has solved complicated pattern recognition problems, especially in big data situations^{36–39}. Stacked denoising autoencoder model is one of the most successful deep learning strategies. The deep architecture can be used to discover latent or hidden representation efficiently inherent in the low-level features from modalities, and ultimately to enhance classification accuracy. In this study, with a stacked denoising auto-encoder, an innovative deep learning-based model was employed to retrospectively analyze a large sample of microcalcifications with or without masses on mammography. Its performance and accuracy in classifying and discriminating breast lesions were compared with benchmark models.

Results

The training group consisted of 1000 images, including 677 benign and 323 malignant lesions. The test group consisted of 204 images, including 97 benign and 107 malignant lesions. Table 1 shows the histopathological distributions of the lesions in both groups. Data about microcalcifications and suspicious breast masses were extracted through image segmentation. Both statistical and textural features were used to classify image features and obtain comprehensive characterization of the microcalcifications and masses. A total of 41 quantitative measurements were recorded for each patient. Detailed information is provided in the Appendix File S1. Fifteen microcalcifications features and twenty-six breast masses features were feed into the comparative classifiers, including SVM, LDA, and KNN. These features were selected since they have been shown to improve the performance of standard machine learning classifiers in earlier researches on breast lesions^{18,19,23,24,34,40,41}.

Figure 1 illustrates an automatic detection and segmentation pipeline to identify suspicious microcalcifications and masses in the left breast of a 60-year-old patient with invasive ductal carcinoma. The microcalcifications were extracted from the raw data to delineate the image characteristics (Fig. 1(b)). Figure 2 shows that this method could accurately detect and extract suspicious microcalcifications from the background of a low-density image showing the left breast of a 56-year-old patient with ductal carcinoma *in situ*. This demonstrated the high accuracy and robustness of the image segmentation pipeline. Figure 3 shows the image of the right breast of a 49-year-old patient with fibrocystic changes in which the focal microcalcifications appear low contrast compared with the high-density background. Extraction of suspicious microcalcifications is a challenging task, however, these results demonstrated that our segmentation model was able to accurately identify and extract microcalcifications from the images to facilitate characterization.

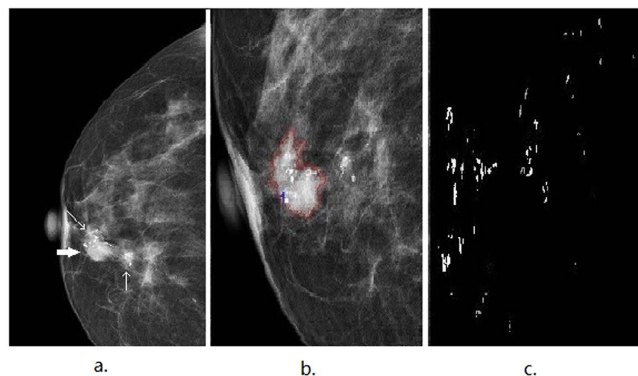


Figure 1. An illustrative example showing segmentation of microcalcifications and breast mass in a mammogram of the left breast of a 60-year-old patient with invasive ductal carcinoma. (a) The craniocaudal (CC) view shows focal clustered microcalcifications (indicated by thin arrows) and an irregular circumscribed mass (indicated by a thick arrow). (b) The suspicious mass is automatically delineated within the red curve. (c) the segmented microcalcifications detected in (b) are used to characterize the features.

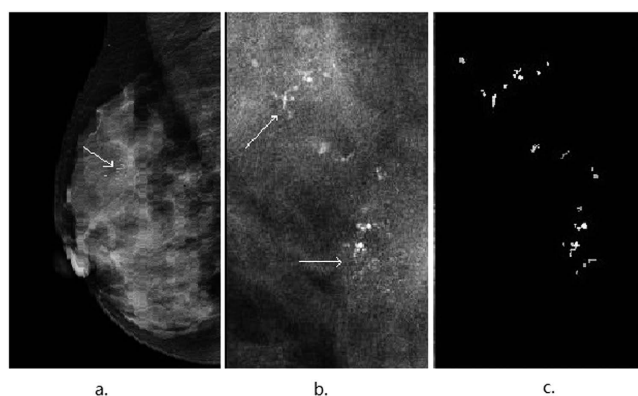


Figure 2. An illustrative example showing segmentation of microcalcifications in a mammogram of the left breast of a 56-year-old patient with ductal carcinoma *in situ*. (a) The mediolateral oblique (MLO) view shows clustered coarse and low density microcalcifications (indicated by thin arrows). (b) The image shows the region of suspicious microcalcifications (indicated by thin arrows). (c) The segmented microcalcifications from (b) are used to characterize the features.

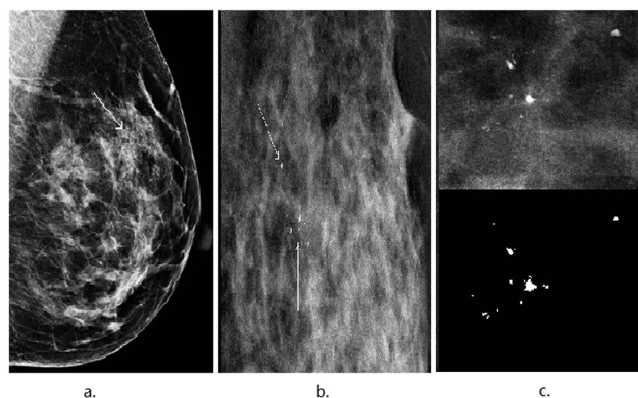


Figure 3. An illustrative example showing segmentation of microcalcifications in a mammogram of the right breast of a 49-year-old patient with fibrocystic changes. (a) The focal microcalcifications (indicated by thin arrows) appear low contrast compared with the dense background in the mediolateral oblique (MLO) view. (b) The region of suspicious microcalcifications is indicated by thin arrows. (c) A zoomed-in view of (b) highlights the segmented microcalcifications.

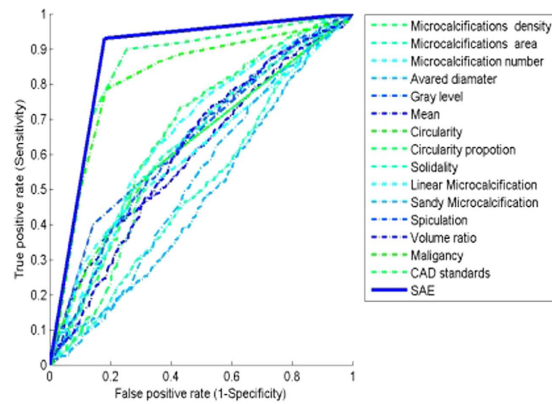


Figure 4. ROC curves for selected microcalcification features. The ROC curves compare the discriminative performances of individual features versus combinations of features.

In order to evaluate the performance and discriminative power of the deep learning model (DL), quantitative measurements for overall classification accuracy (acc), sensitivity, specificity and the area under the receiver operating characteristic (ROC) curve (AUC) were calculated as follows:

$$acc = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%$$

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{FP + TN}$$

where TP , FN , TN and FP represent the true positives, false negatives, true negatives and false positives, respectively.

Previous reports have suggested that the discriminative performances of classifiers can be increased through comprehensive characterization of microcalcifications as opposed to characterization of individual features. In agreement with these reports, our deep learning-based model achieved similar outcomes, as demonstrated by the ROC curves in Fig. 4. Therefore, this approach was used in the following experiments.

Three scenarios for discriminating between malignant and benign lesions were examined: microcalcifications alone; breast masses alone; and microcalcifications and breast masses in combination. The primary aims of the three scenarios were to investigate the discrimination power of microcalcifications, masses or their combination in differentiation of the lesions types. The results were compared to those of SVM, KNN and LDA benchmark classifiers.

The structure of a SAE network is decided by the size of the input layer, the number of hidden layers, and the number of hidden units in each hidden layer. Through the experiments, the data of microcalcifications alone was used as the input in the first scenario; the data of breast masses alone was served as the input in the second scenario; and the data of microcalcifications and breast masses in combination was served as the input in the third scenario. We used the SAE model to classify malignant and benign lesions in three scenarios. The optimal hyper parameters for the three scenarios were estimated by 10 fold cross-validation on training group. For the first scenario, the trained architecture consisted of two hidden layers, and the number of hidden units in each hidden layer was [200, 200], respectively. For the second scenario, the trained architecture consisted of two hidden layers, and the number of hidden units in each hidden layer was [200, 200], respectively. For the third scenario, the trained architecture consisted of two hidden layers, and the number of hidden units in each hidden layer was [400, 400].

In the first scenario, image segmentation yielded 15 features. The overall accuracies were 85.8%, 83.8%, 58.8% and 87.3% for the SVM, KNN, LDA and DL models, respectively. The DL model also achieved the highest specificity and AUC values (0.82 and 0.87, respectively). The results are summarized in Table 2; the ROC curves in Fig. 5(a) provided visual comparisons between the models.

In the second scenario, based on breast masses alone, image segmentation yielded 26 features. The results are summarized in Table 3, and the ROC curves are shown in Fig. 5(b). The overall accuracies were markedly lower in all of the models, at 61.3%, 58.8%, 53.4% and 61.3% for SVM, KNN, LDA and DL respectively. Furthermore, the performance of the DL model was only marginally higher than that of the SVM model. Despite this finding, the sensitivity of the model was approximately 100%, indicating that patients who tested positive all had breast masses. As such, this method may facilitate diagnosis in benign cases; however, it may not serve as a valid diagnostic tool in clinical practice.

In the third scenario, based on a combinatorial approach by analyzing microcalcifications and breast masses simultaneously, image segmentation yielded 41 features. The overall accuracies were 85.8%, 84.3%, 74.0% and 89.7% for the SVM, KNN, LDA and DL models, respectively. Furthermore, the DL model achieved the highest

	Test Dataset				Training Dataset
	accuracy	sensitivity	specificity	AUC	mean \pm std (Accuracy)
SVM	85.8%	0.93	0.79	0.85	0.79 \pm 0.07
KNN (N = 8)	83.8%	0.95	0.74	0.84	0.77 \pm 0.07
LDA	58.8%	0.63	0.55	0.59	0.61 \pm 0.05
SAE	87.3%	0.93	0.82	0.87	0.82 \pm 0.05

Table 2. Diagnostic performances of different classification models through microcalcification features (15 features). The proposed SAE achieved superior performance in terms of the four measurements. The best measurements were highlighted in bold.

specificity and AUC values (0.90 and 0.90, respectively). In comparison, the SVM model achieved 78% specificity and an AUC value of 0.85. The results are summarized in Table 4 and the ROC curves are shown in Fig. 5(b).

These findings confirmed that by accessing a large dataset, the deep learning model produced a higher number of representative segmentation features and exhibited greater overall accuracy for discriminating between malignant and benign breast lesions through mammography compared to standard models. Furthermore, the discriminative power of the deep learning model was greatest if a combinatorial approach was applied to characterize microcalcifications and breast masses simultaneously.

Discussion

Mammography is considered the primary imaging modality for early detection and treatment of breast cancer; however, achieving accurate diagnoses through mammography is often challenging for radiologists due to the difficulty of distinguishing the features of malignant symptoms in images^{42–44}. Consequently, considerable research is being undertaken to develop computer-based applications including various classification models to overcome these challenges^{10,12,14,45–47}.

Microcalcifications are highly correlated with breast cancer^{2,3,7}, therefore, the aim of this investigation was to evaluate the performance of an innovative deep learning model for classifying breast lesions. The results demonstrated that deep learning not only enabled accurate segmentation of microcalcifications but also provided an efficient analysis of their characteristics, leading to a marked improvement in discriminating between benign and malignant breast lesions compared to more standard SVM, KNN and LDA methods. This may have particular significance for cases in which microcalcifications are the only indicator of malignant lesions^{4,11,48,49}.

Deep learning-based models employing large sample sets show greater discriminative performance in classifying microcalcifications through mammography compared to other machine learning methods. Compared to other methods, deep learning-based models provide a higher number of image segmentation features and help enhance the diagnostic accuracy through comprehensive characterization of these features. The discriminative power of deep learning can be increased by adopting a combinatorial approach to classify microcalcifications and masses simultaneously. Our results suggest that deep learning based-models on large datasets are promising in the earlier detection and treatment of breast cancer by identifying microcalcifications on mammograms.

Breast masses are also known to exhibit distinct features that vary from benign to malignant lesions²; however, machine-based methods are generally based on detecting microcalcifications or breast masses in isolation. In contrast, reports on methods that detect microcalcifications and masses simultaneously are scarce. In this study, we carried out a provisional and innovative trial using our deep learning-based model to distinguish both features in combination. The results showed that this combinatorial approach enhanced the diagnostic sensitivity of the model in patients presenting with both microcalcifications and masses. This implied that deep learning may offer an advanced statistical method for differentiating mammographic microcalcifications with greater accuracy and sensitivity, both in the presence or absence of breast masses. Not only could this facilitate earlier and more accurate classification of breast cancer, but also improve prognosis through timely treatment in malignant cases. It may also help avoid unnecessary surgical procedures, including total resection, and psychological and physiological pain in benign cases.

However, the current study suffered from the following limitations. First, the testing dataset should be expanded to provide more benign and malignant samples in order to achieve higher statistical power. In addition, by increasing the number of cases with breast masses, either alone or with microcalcifications, would allow deeper examination of the combinatorial approach and facilitate establishing the optimal diagnostic performance of our model and its potential value in future applications. Second, the features investigated in present study may not be sufficient enough to fully characterize microcalcifications, future studies will extract more. By selecting the most discriminative subset of them and optimizing the selection of various features, it helps improve the performance of deep learning in the classification stage. The current study was aiming to employ powerful deep learning based classifier to discriminate breast lesions by microcalcifications with or without the combined analysis of masses. With the settlement of problems addressed before, the nice performance of our trial in using deep learning opens a way to aid radiologist's diagnostic performance. It further facilitates the systematical investigation of breast cancer for early detection, diagnosis and clinical management.

Methods

Participant population. We retrospectively reviewed mammograms from 1204 female patients histopathologically diagnosed with benign or malignant breast lesions at the SunYat-sen University Cancer Center (Guangzhou, China) and Nanhai Affiliated Hospital of Southern Medical University (Foshan, China) between

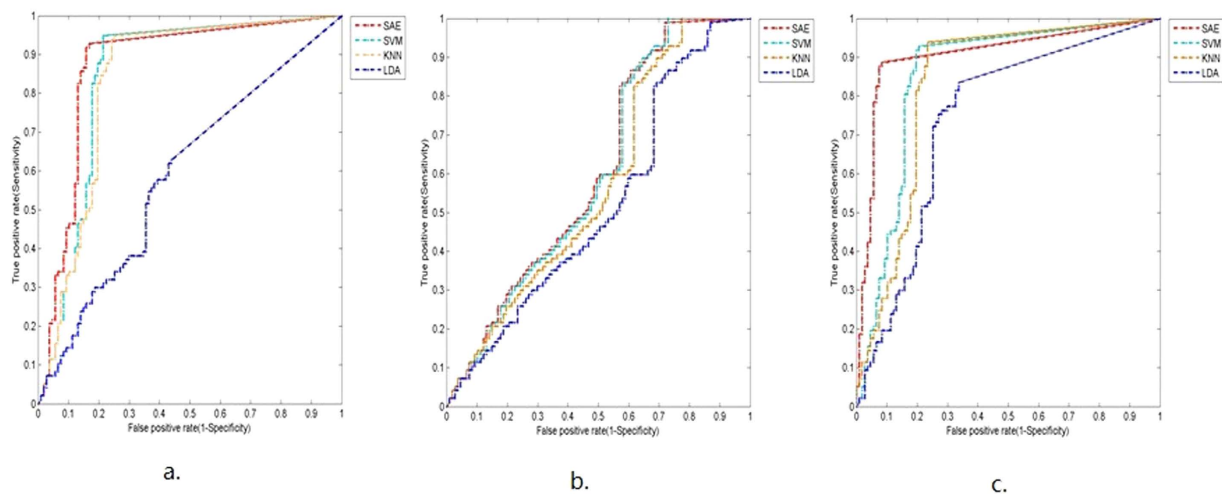


Figure 5. ROC curves comparing the discriminative performances of the four classification models. The three scenarios show the ROC curves based on the following classifications: (a) Microcalcifications alone (15 segmentation features). (b) Breast masses alone (26 segmentation features). (c) Microcalcifications plus breast masses (41 segmentation features).

	Test Dataset				Training Dataset
	accuracy	sensitivity	specificity	AUC	mean \pm std (Accuracy)
SVM	61.3%	1.00	0.26	0.60	0.68 \pm 0.11
KNN (N = 8)	58.8%	1.00	0.21	0.57	0.71 \pm 0.11
LDA	53.4%	0.99	0.12	0.52	0.67 \pm 0.13
SAE	61.3%	0.99	0.27	0.61	0.71 \pm 0.12

Table 3. Diagnostic performances of different classification models through mass features (26 features). The proposed SAE achieved superior performance in terms of the four measurements. The best measurements were highlighted in bold.

	Test Dataset				Training Dataset
	accuracy	sensitivity	specificity	AUC	mean \pm std (Accuracy)
SVM	85.8%	0.95	0.78	0.85	0.79 \pm 0.07
KNN (N = 6)	84.3%	0.94	0.76	0.83	0.77 \pm 0.06
LDA	74.0%	0.84	0.65	0.74	0.69 \pm 0.07
SAE	89.7%	0.89	0.90	0.90	0.85 \pm 0.06

Table 4. Diagnostic performances of different classification models through microcalcifications and mass features in combination (41 features). The proposed SAE achieved superior performance in terms of the four measurements. The best measurements were highlighted in bold.

May 2011 and March 2015. The sample comprised of 774 benign and 430 malignant breast lesions. All patients underwent molybdenum targeted mammography. Identified lesions were histopathologically confirmed as benign or malignant by performing open surgical biopsy or fine needle biopsy. The sample was divided into two groups: the training group comprised of images from 1000 randomly selected patients admitted between May 2011 and March 2015 (range, 26–75 years); the test group comprised of images from 204 randomly selected patients admitted between October 2013 and March 2015 (range, 28–75 years). All experimental protocols were approved by the Ethics Committee of the SunYat-sen University Cancer Center and the Ethics Committee of the Nanhai Affiliated Hospital of Southern Medical University, and were conducted in accordance with the Good Clinical Practice guideline. Informed consent was obtained from each patient for their consent to have their information used in research without affecting their treatment option or violating their privacy.

Imaging and analysis. Images were obtained on a GE Senographe DS mammography system and a Siemens Mammomat Inspiration mammography system. Craniocaudal (CC) and mediolateral oblique (MLO) projections were obtained for each breast. All images were digitized at a resolution of 1024×1024 pixels and at 8-bit gray scale level. Taking the raw image directly may bring in a large bias due to image deformation, uniform background

illumination, uneven imaging angle and position. Such problems may deteriorate the classification performance. To alleviate the problems, this study used various types of features that were widely used in researches on breast lesions as input data instead of original images^{34,40,41}. We not only considered the features invariant to rotation, but also the features invariant to rotation, scaling, and translation. A previously reported computerized segmentation approach²⁹ was used to extract any suspicious microcalcifications and masses from each image. Data about microcalcifications and suspicious breast masses were extracted through image segmentation. Both statistical and textural features were used to classify image features and obtain comprehensive characterization of microcalcifications and breast masses. A total of 41 quantitative measurements were recorded for each patient. Fifteen microcalcifications features and twenty-six breast masses features, estimated from the region of interests, were selected instead of original images as the input data for SAE model. The extracted features from mammograms aimed to provide comprehensive characterization of the image as much as possible. They consisted of intensity, statistic, shape and texture features. These features were extensively reported and tested widely in researches on breast lesions^{18,19,23,24,34,40,41}. The 15 microcalcifications features were selected to describe different dimensional aspects of microcalcifications, including one-dimensional shape features (average diameter), two-dimensional morphological features (microcalcifications area), fractal dimensional features (microcalcifications density, circularity proportion, solidity, sandy microcalcification, spiculation, volume ratio), gray level intensity statistics features (mean gray value), and statistic feature (microcalcifications number, circularity, linear microcalcification). The 26 breast masses features also characterized different aspects of masses, including morphological features (breast masses area), fractal dimensional features (solidity, elongation, axis ratio, heterogeneity, spiculation, volume ratio, convexity), texture features (mean gray, maximum gray, gray relativity, entropy, inverse difference entropy, difference entropy, correlation, difference variance, sum average, sum variance, energy, mutual information). Detailed information about the features was provided in the Appendix File S1. Once the comprehensive characterization for each lesion done, its feature description was feed into the deep learning model to classify its type into benign or malignant.

Deep learning model. Deep learning is a machine learning model with multiple hidden layers that learns inherent rules and features of large data sets. A stacked autoencoder (SAE) creates a deep network by stacking multiple autoencoders hierarchically^{31,34,35}. Each autoencoder is a neural network (NN) that attempts to reproduce its input; the output of each autoencoder is used as the training set for the next autoencoder. More specifically, in an SAE with n layers, the first layer is trained as an autoencoder to obtain the first hidden layer, and the output of the k^{th} hidden layer is used as the input of the $(k+1)^{\text{th}}$ hidden layer.

In this study, 15 microcalcifications features and 26 breast masses features were selected instead of original images as the input data for SAE model, respectively. The SAE model was trained in a layer-wise greedy fashion to learn low-level features of microcalcifications from input data according to the following mathematical procedures:

Training samples were denoted as $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$; an autoencoder encoded input $x^{(i)}$ to a hidden representation $y(x^{(i)})$ through a deterministic mapping function:

$$y(x) = f(W_1x + b_1). \quad (1)$$

Conversely, the autoencoder decoded the representation $y(x^{(i)})$ back into a reconstruction through a second deterministic mapping function:

$$z(x) = g(W_2y(x) + b_2) \quad (2)$$

where W_1 is a weight matrix, W_2 is a decoding matrix, b_1 is an encoding bias vector, and b_2 is a decoding bias vector.

A logistic sigmoid function: $f(x) = \frac{1}{1+e^{-x}}$ and $g(x) = \frac{1}{1+e^{-x}}$ was used in this study.

The objective of an autoencoder was to minimize the reconstruction error by applying the following formula:

$$\min_{W_1, W_2, b_1, b_2} \frac{1}{2} \sum_{i=1}^m \|x^{(i)} - z(x^{(i)})\|^2. \quad (3)$$

The encoding procedure was carried out from the first layer to the last layer by the following formulas:

$$a^{(k)} = f(z^{(k)}) \quad (4)$$

$$z^{(k+1)} = W^{(k,1)}a^{(k)} + b^{(k,1)}. \quad (5)$$

The decoding procedure was calculated from the last layer to the first layer by the following formulas:

$$a^{(n+k)} = f(z^{(n+k)}) \quad (6)$$

$$z^{(n+k+1)} = W^{(n-k,2)}a^{(n+k)} + b^{(n-k,2)} \quad (7)$$

where $W^{(k,1)}$ is a weight matrix of the k^{th} autoencoder, $W^{(k,2)}$ is a decoding matrix of the k^{th} autoencoder, $b^{(k,1)}$ is an encoding bias vector of the k^{th} autoencoder, and $b^{(k,2)}$ is a decoding bias vector of the k^{th} autoencoder, $f(x)$ is sigmoid function, $a^{(k)}$ is sigmoid value.

We added a softmax classifier on the top layer of the SAE network to create the deep learning model for analyzing breast lesions⁵⁰⁻⁵⁵.

References

- Cady, B. & Chung, M. Mammographic screening: no longer controversial. *American Journal of Clin Oncol* **28**(1), 1–4 (2005).
- American College of Radiology (ACR). Breast imaging reporting and data system (BI-RADS), breast imaging atlas. 4th ed., Reston, Va, Am College Radiology, 1–259 (2003).
- Fletcher, S. W. & Elmore, J. G. Clinical Practice: mammographic screening for breast cancer. *New Engl J Med* **348**(17), 1672–1680 (2003).
- Winchester, D. P., Jeske, J. M. & Goldschmidt, R. A. The diagnosis and management of ductal carcinoma *in situ* of the breast. *Am Cancer J Clin* **50**(3), 184 (2000).
- Schreer, I. & Luttgies, J. Breast cancer: early detection. *Eur J Radiol* **11** (Suppl 2), S307–S314 (2001).
- Stephen, A. & Feig, M. D. Ductal carcinoma *in situ*. Implications for screening mammography. *Radiol Clin N Am* **38**(4), 653–668 (2000).
- Yunus, M., Ahmed, N. & Masroor, I. Mammographic criteria for determining the diagnostic value of microcalcifications in the detection of early breast cancer. *J Pak Med Assoc* **4**(1), 24–29 (2004).
- Muttarak, M., Kongmebho, IP. & Sukhamwang, N. Breast calcifications: which are malignant. *Singap Med J* **50**(9), 907–914 (2009).
- Goergen, S. K., Evans, J. & Cohen, G. P. Characteristics of breast carcinomas missed by screening radiologists. *Radiology* **204**(1), 131–135 (1997).
- Barlow, W. E., Chi, C. & Carney, P. A. Accuracy of screening mammography interpretation by characteristics of radiologists. *J Natl Cancer I* **96**(24), 1840–1850 (2004).
- Muttarak, M., Pojchamarnwiputh, S. & Chaiwan, B. Breast carcinomas: why are they missed. *Singap Med J* **47**(8), 851–857 (2006).
- Astley, S. & Hutt, I. Automation in mammography: computer vision and human perception London. *World Scientific Publishing Co* 1–25 (1994).
- Burhene, L. J. W., Wood, S. A. & D’Orsi, C. J. Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology* **215**(2), 554–562 (2000).
- Jiang, Y., Nishikawa, R. M., Schmidt, R. A. & Metz, C. E. Comparison of independent double readings and computer-aided diagnosis (CAD) for the diagnosis of breast calcifications. *Acad Radiol* **13**(1), 84–94 (2006).
- Malich, A., Fischer, D. R. & B(o)ttcher, J. CAD for mammography: the technique, results, current role and further developments. *Eur Radiol* **16**(7), 1449–1460 (2006).
- Cai, H. M., Peng, Y. X., Ou, C. W., Chen, M. S. & Li, L. Diagnosis of breast masses from dynamic contrast-enhanced and diffusion-weighted MR: a machine learning approach. *PLoS One* **9**(1), e87387 (2014).
- Dartois, L. *et al.* A comparison between different prediction models for invasive breast cancer occurrence in the French E3N cohort. *Breast Cancer Res Tr* **150**(2), 415–426 (2015).
- Krishnan, M. *et al.* Statistical analysis of mammographic features and its classification using support vector machine. *Expert Systems with Applications* **37**(1), 470–478 (2010).
- Wang, D., Shi, L. & Heng, P. A. Automatic detection of breast cancers in mammograms using structured support vector machines. *Neurocomputing* **72**(13), 3296–3302 (2009).
- Akay, M. F. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert systems with applications* **36**(2), 3240–3247 (2009).
- Holsbach, N., Fogliatto, F. S. & Anzanello, M. J. A data mining method for breast cancer identification based on a selection of variables. *Ciência & Saúde Coletiva* **19**(4), 1295–1304 (2014).
- Sahan, S., Polat, K., Kodaz, H. & Güneş, S. A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis. *Comput BiolMed* **37**(3), 415–423 (2007).
- Pérez, N. *et al.* Improving the performance of machine learning classifiers for Breast Cancer diagnosis based on feature selection. Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on. IEEE.209–217 (2014).
- Pérez, N., Guevara, M. A. & Silva, A. Improving breast cancer classification with mammography, supported on an appropriate variable selection analysis. *SPIE medical imaging*. 867022–867022 (2013).
- Ibrahim, R. *et al.* Multi-level gene/MiRNA feature selection using deep belief nets and active learning. Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE.3957–3960 (2014).
- Güçlü, U. & Gerven, A. J. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J Neurosci* **35**(27), 10005–10014 (2015).
- Alipanahi, B., Delong, A., Weirauch, M. T. & Frey, B. J. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol* **33**(8), 831–838 (2015).
- Song, Y. Y. *et al.* Automatic vaginal bacteria segmentation and classification based on superpixel and deep learning. *J Med Imag Health In* **4**(5), 781–786 (2014).
- Shao, Y. Z. *et al.* Characterizing the clustered microcalcifications on mammograms to predict the pathological classification and grading: A mathematical modeling approach. *Journal of digital imaging* **24**(5), 764–771 (2011).
- Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**(5786), 504–507 (2006).
- Shin, H. C. *et al.* Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE T on Medical Imaging* **99**, 1–1 (2016).
- Suk, H. I. *et al.* Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Structure and Function* **220**(2), 841–859 (2015).
- Suk, H. I. *et al.* Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* **101**, 569–582 (2014).
- Arevalo, J., González, F. A. & Ramos-Pollán, R. Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer Methods and Programs in Biomedicine* **127**, 248–257 (2016).
- Petersen, K., Nielsen, M. & Diao, P. Breast tissue segmentation and mammographic risk scoring using deep learning. *Breast Imaging, Springer International Publishing* **2014**, 88–94 (2014).
- Collobert, R. & Weston, J. A unified architecture for natural language processing: Deep neural networks with multitask learning In Proc. 25th ICML 160–167 (2008).
- Huval, B., Coates, A. & Ng, A. Deep learning for class-generic object detection. *arXiv preprint arXiv* **1312**, 6885 (2013).
- Shin, H. C., Orton, M. R., Collins, D. J., Doran, S. J. & Leach, M. O. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. *IEEE T Pattern Anal* **35**(8), 1930–1943 (2013).
- Gao, S. H., Zhang, Y., Jia, K., Lu, J. & Zhang, Y. Single sample face recognition via learning deep supervised auto-Encoders. *IEEE T Inf Foren Sec* **10**(10), 2108–2118 (2015).
- Moura, D. C. & López, M. A. G. An evaluation of image descriptors combined with clinical data for breast cancer diagnosis. *International journal of computer assisted radiology and surgery* **8**(4), 561–574 (2013).
- Pérez, N. P. *et al.* Improving the Mann–Whitney statistical test for feature selection: An approach in breast cancer diagnosis on mammography. *Artificial intelligence in medicine* **63**(1), 19–31 (2015).
- Brem, R. F. *et al.* Impact of breast density on computer-aided detection for breast cancer. *AJR Am J Roentgenol* **184**(2), 439–444 (2005).
- Malich, A. *et al.* Tumor detection rate of a new commercially available computer-aided detection system. *Eur Radiol* **12**(10), 2454–2459 (2001).

44. Li, L. H., Wu, Z. B. & Salem, A. F. Computerized analysis of tissue density effect on missed cancer detection in digital mammography. *Comput Med Imag Grap* **30**(5), 291–297 (2006).
45. Sankar, D. & Thomas, T. Fast fractal coding method for the detection of microcalcification in mammograms. *Piscataway, NJ, IEEE* 368–373 (2008).
46. Yu, S. & Guan, L. A. CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films. *IEEE T Med Imaging* **19**(2), 115–126 (2000).
47. Jiang, J., Yao, B. & Wason, A. M. A genetic algorithm design for microcalcification detection and classification in digital mammograms. *Comput Medical Imag Grap* **31**(1), 49–61 (2007).
48. Stomper, P. C., Geradts, J. & Edge, S. B. Mammographic predictors of the presence and size of invasive carcinomas associated with malignant microcalcification lesion without a mass. *AJR Am J Roentgenol* **181**(6), 1679–1684 (2003).
49. Egan, R. L., McSweeney, M. B. & Sewell, C. W. Intramammary calcifications without an associated mass in benign and malignant diseases. *Radiology* **137**(1), 1–7 (1980).
50. Bengio, Y., Lamblin, P., Popovici, D. & Larochelle, H. Greedy layer-wise training of deep networks. *Advances in neural information processing systems* **19**, 153–160 (2007).
51. Palm, R. B. Prediction as a candidate for learning deep hierarchical models of data. Master thesis, Technical University of Denmark (2012).
52. He, H. & Garcia, E. A. Learning from imbalanced data. *Knowledge and Data Engineering, IEEE Transactions on* **21**(9), 1263–1284 (2009).
53. Guan, S. *et al.* Deep learning with MCA-based instance selection and bootstrapping for imbalanced data classification. The First IEEE International Conference on Collaboration and Internet Computing (CIC). 288–295 (2015).
54. Berry, J. *et al.* Training deep nets with imbalanced and unlabeled data. *Interspeech* 1756–1759 (2012).
55. Masko, D. & Hensman, P. The impact of imbalanced training data for convolutional neural networks. Bachelor thesis, KTH, School of Computer Science and Communication (2015).

Acknowledgements

This work was funded by the National Nature Science Foundation of China (Nos 81471711; 61372141; 81271622), the Guangdong Province of Higher School “Thousand Hundred Ten Talents Project” (No. 84000-52010004), the Nature Science Foundation of Guangdong Province, China (No. 2014A030311036), BGI-SCUT Innovation Fund Project (No. SW20130803), the Fundamental Research Fund for the Central Universities (No. 2015ZZ025) and the Science and Technology Planning Project of Guangdong Province, China (No. 2013B021800161).

Author Contributions

L.L. and H.C. designed and directed the experiments. J.W. and X.Y. analyzed the data, drafted and revised the paper. W.T. and C.J. contributed to data acquisition and carried out clinical studies. All authors read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wang, J. *et al.* Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning. *Sci. Rep.* **6**, 27327; doi: 10.1038/srep27327 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>