PAPER

A deep learning framework for supporting the classification of breast lesions in ultrasound images

IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology

Phys. Med. Biol. **62** (2017) 7714–7728 https://doi.org/10.1088/1361-6560/aa82ec

A deep learning framework for supporting the classification of breast lesions in ultrasound images

Seokmin Han[1], Ho-Kyung Kang[2], Ja-Yeon Jeong[2], Moon-Ho Park[2], Wonsik Kim[2], Won-Chul Bang[2] and Yeong-Kyeong Seong[2]

[1] Korea National University of Transportation, Uiwang-si, Kyunggi-do, Republic of Korea
[2] Samsung Electronics, Suwon-si, Kyunggi-do, Republic of Korea

E-mail: yk.seong@samsung.com

Received 14 April 2017, revised 17 July 2017
Accepted for publication 28 July 2017
Published 15 September 2017

Abstract

In this research, we exploited the deep learning framework to differentiate the distinctive types of lesions and nodules in breast acquired with ultrasound imaging.

A biopsy-proven benchmarking dataset was built from 5151 patients cases containing a total of 7408 ultrasound breast images, representative of semi-automatically segmented lesions associated with masses. The dataset comprised 4254 benign and 3154 malignant lesions. The developed method includes histogram equalization, image cropping and margin augmentation. The GoogLeNet convolutionary neural network was trained to the database to differentiate benign and malignant tumors.

The networks were trained on the data with augmentation and the data without augmentation. Both of them showed an area under the curve of over 0.9. The networks showed an accuracy of about 0.9 (90%), a sensitivity of 0.86 and a specificity of 0.96.

Although target regions of interest (ROIs) were selected by radiologists, meaning that radiologists still have to point out the location of the ROI, the classification of malignant lesions showed promising results. If this method is used by radiologists in clinical situations it can classify malignant lesions in a short time and support the diagnosis of radiologists in discriminating malignant lesions. Therefore, the proposed method can work in tandem with human radiologists to improve performance, which is a fundamental purpose of computer-aided diagnosis.

Keywords: deep learning, CADx, classification, breast cancer

(Some figures may appear in colour only in the online journal)

1. Introduction

Breast cancer is the most common cancer in women worldwide and the second leading cause of female cancer deaths (CDC 2017). Though mammography is the primary imaging modality for screening, ultrasound (US) imaging also plays a significant role in breast imaging as a diagnostic tool (Kornecki 2011). Double reading, where two radiologists independently read the same mammogram, has been reported to reduce the occurrence of missed cancers and most screening programs include it (Tabr *et al* 2011). Because double reading requires additional work and costs, a computer-aided diagnosis (CADx) system can alternatively provide a single radiologist with support for his or her decision. CADx is a computerized procedure to provide a second objective opinion on the interpretation of a medical image and diagnosis (Chen *et al* 2003, Joo *et al* 2004, Awai *et al* 2006, McCarville *et al* 2006, Doi 2007, Giger *et al* 2008, Drukker *et al* 2009, Cheng *et al* 2010, van Ginneken 2011, Giger *et al* 2013). One of the major applications of CADx is to differentiate between malignant and benign tumors and lesions (Sluimer *et al* 2003, Armato and Sensakovic 2004, Way *et al* 2006, Giger *et al* 2008, Way *et al* 2009, Cheng *et al* 2010, Sun *et al* 2013, Wang *et al* 2016). In addition to assistance with interpretation and diagnosis, computer-based methods can minimize intra-observer and inter-observer variability (Armato and Sensakovic 2004, Way *et al* 2006, 2009, Sun *et al* 2013) and provide quantitative support for clinical decisions, such as biopsy recommendations (Giger *et al* 2013), by efficiently analyzing large numbers of images; this helps to reduce unnecessary thoracotomy (McCarville *et al* 2006) and false-positive biopsies (Joo *et al* 2004). Therefore, these systems can be used to provide radiologists with a second opinion in a cost-effective way; they play a key role in the early detection of breast cancer and help to reduce the death rate among women with breast cancer (Ayer *et al* 2010).

A wide variety of machine learning methods have been suggested for early detection of breast cancer (Cruz and Wishartl 2006, Krishnan *et al* 2010, Vishrutha and Ravishankar 2014). Conventional CADx often includes feature extraction, feature selection (Tourassi *et al* 2001, Newell *et al* 2010, Gmez *et al* 2012, Yang *et al* 2013) and classification. These steps should be well designed and then integrated for overall performance. In conventional machine learning methods one of the most important issues (Chen *et al* 2003, Joo *et al* 2004, Armato and Sensakovic 2004, Way *et al* 2006, 2009, Cheng *et al* 2010, Newell *et al* 2010, Sun *et al* 2013) is the extraction of an effective feature for each problem. The extraction of effective features could potentially alleviate the burden of feature selection and classification. Nevertheless, the development of useful features is dependent on problems and should be assisted by the latter steps of feature selection and classification to achieve better lesion classification.

In general, the image features can be categorized into morphological and textural features. The extraction of effective features is a complicated task that involves many image processing steps. These image processing steps include morphological feature computing (Sahiner *et al* 2001, Chen *et al* 2003, Armato and Sensakovic 2004, Joo *et al* 2004, Way *et al* 2006, Cheng *et al* 2010), which is still difficult to solve (Arbelaez *et al* 2011), and image decomposition (Sun *et al* 2013, Yang *et al* 2013), followed by statistical summaries and presentations (Sorensen 2010, Gmez *et al* 2012, Sun *et al* 2013, Yang *et al* 2013) for the calculation of textural features (Tourassi 1999). In feature integration by classifier, the widely used techniques are based on the KNN (*k*-nearest neighbor) method (Sahan *et al* 2007, Holsbach *et al* 2014), LDA (linear discriminant analysis) (Perez *et al* 2013, 2014) and SVM (support vector machines) (Akay *et al* 2009, Wang *et al* 2009, Krishnan *et al* 2010). The extraction of meaningful features is highly dependent on the quality of each intermediate result in the image processing steps (Tourassi 1999, Sahiner *et al* 2001, Joo *et al* 2004, Armato and Sensakovic 2004, Cheng *et al* 2010), which often requires recursive trial and error to obtain satisfactory results.

Thus, it is time-consuming and very difficult to design and tune the overall performance of a conventional CADx framework to get a satisfactory result because many image processing steps need to be considered at the same time.

Recently, deep learning methods have been widely applied to address several perception-related problems (Bengio *et al* 2013). The main advantage of deep learning methods is to alleviate the burden of designing the specific features and classification framework, because deep learning models produce a set of transformation functions and image features directly from the data. Deep learning methods have been introduced to medical imaging with promising results in various applications, such as the computerized prognosis or diagnosis for Alzheimer's disease (Suk *et al* 2013, 2014, Suk *et al* 2015, Li *et al* 2015) and mild cognitive impairment (Suk *et al* 2015), organ segmentation (Zhang *et al* 2015) and detection (Shin *et al* 2013), ultrasound standard plane selection (Chen *et al* 2015), tissue classification in histological and histopathological images (Arevalo *et al* 2013, Cruz-Roa *et al* 2013) and knee cartilage segmentation (Prasoon *et al* 2013), among others. In the context of CADx, most works have focused on the problem of detection of abnormalities (CADe) (Seff *et al* 2015, Tajbakhsh *et al* 2015, Roth *et al* 2016).

Although some previous works have explored deep learning methods to address the automatic classification of identified lesions in breast images (Jalalian *et al* 2013, Cheng *et al* 2016), most of them utilized a small dataset, which may require additional evaluation. In this study we exploit the deep learning model of GoogLeNet to differentiate the distinctive types of lesions and nodules acquired with ultrasound imaging. The proposed framework is a component algorithm of S-Detect technology, which is implemented in RS80A (Samsung Medison, Inc.)

2. Methods and materials

In this research we apply a deep learning framework to a large dataset of ultrasound breast tumor images and provide a CADx method that can be helpful in real clinical situations. We present an illustration of the proposed architecture in figure 1.

A biopsy-proven benchmarking dataset was built from 5151 patient cases containing a total of 7408 ultrasound breast images, representative of semi-automatically segmented lesions associated with masses. The developed method comprises histogram equalization, image cropping and margin augmentation. Histogram equalization is performed for each region of interest (ROI) sample image, and the image is resized to match the input image size of the network. The GoogLeNet convolutional neural network (CNN) is trained to differentiate malignant tumors from benign ones. Of the 7408 ultrasound breast images, 6579 were used as the training set and 829 as the test set. The training dataset comprised 3765 images of benign masses and 2814 of malignant masses. Optimal parameters were chosen based on 10-fold cross-validation with the training data. Breast mass images were roughly cropped and extracted from the breast image.

2.1. Data preparation

All experimental protocols were approved by Samsung Medical Center, Seoul, South Korea. Informed consent was obtained from all patients for use of their information in the research without violating their privacy. At the Samsung Medical Center (Seoul, South Korea) 7408 breast sonogram images were scanned from 5151 patients; 5254 images were acquired with a iU22 system (Philips, Inc.) and 2154 with an RS80A (Samsung Medison, Inc.). All breast lesions were proved histopathologically by biopsy. In figure 2 we illustrate some examples of

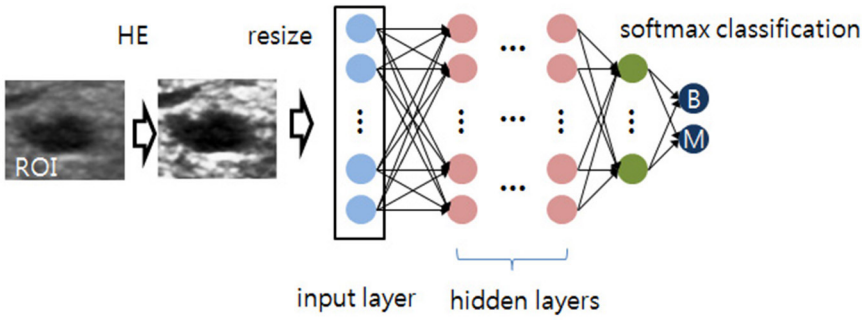

Figure 1. The conceptual architecture of the proposed deep learning CADx framework.

Table 1. Overview of the lesion size attributes of training data and test data.

	∼0.5	∼1.0	∼2.0	∼3.0	∼4.0	∼5.0	∼5.5 (cm)	Total
Training	672	2134	2433	846	355	132	6	6579
Test	88	343	314	58	16	9	1	829

malignant and benign lesions. In this research, we randomly selected 829 lesions (489 benign and 340 malignant) from the image dataset as a test set. We carefully selected test data so that the training and test sets were split on a patient level to prevent bias. In table 1 we present an overview of the lesion size attributes of the training and test data. In addition to the lesion size attributes, we also present the distribution of Breast Imaging Reporting and Data System (BIRADS) categories of training data and test data in figure 3.

The proposed CADx system aims to classify a previously identified ROI. This ROI can be obtained by a manual segmentation or automatically detected by a computer-aided detection system. In this research, we employed a semi-automatic segmentation system. Based on the location of the ROI provided by six radiologists, an automatic segmentation method was operated to draw the boundary of the lesion. Based on the boundary, the ROI was cropped with a margin.

2.2. Image cropping with a margin

In this research, the margin is defined as the distance between the lesion boundary and the boundary of the cropped image itself as shown in figure 4. To work out how much the margin affects the overall performance, we compared the performance as the margin increased from 0 to 240. With the margin, the images include the information in the background as well as the information in the breast lesion. Without the margin, the images include just the information about the lesion itself. Because the input image size of CNN is fixed to 255×255, the relative size of the lesion is affected by the size of the margin. In this research, a margin of 180 pixels showed the best performance in terms of accuracy. Thus, we set a margin of 180 pixels as the default margin for image cropping.

2.3. Data augmentation

For the training dataset, we added the cropped images with margins of 120 pixels and 150 pixels to the training dataset as data augmentation, as shown in figure 5. For the test dataset, the margin was fixed to 180 pixels. To facilitate the training, the ROI images were resized

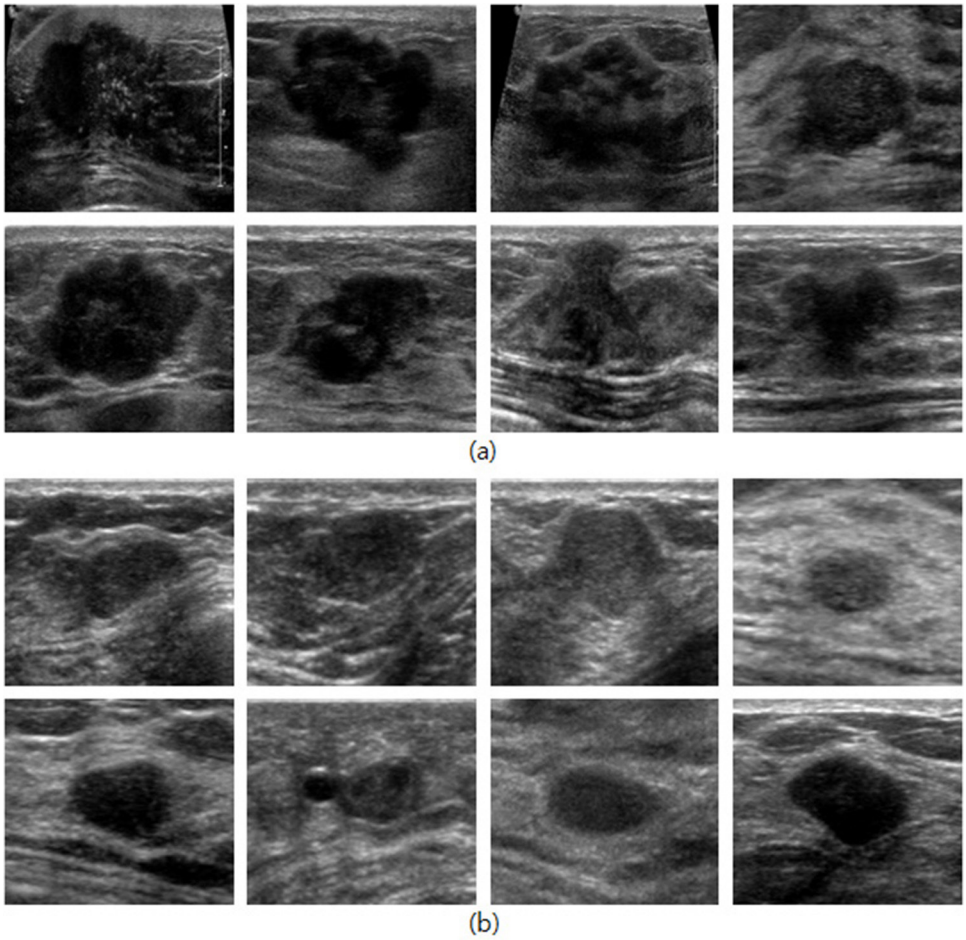

Figure 2. BIRADS distribution of (a) training data and (b) test data.

into patches of 255×255 size using bilinear interpolation. Because the input images for the network were resized to a fixed size of 255×255, this augmentation affects the relative size of the breast tumor in the images. The purpose of this augmentation is to make the network robust to the size variability of breast tumors. In addition to margin augmentation, we also employed mirroring augmentation, shuffling of training data (provided by Caffe framework (Jia *et al* 2014)) and translation $[-12, 12]$ at interval 4 from the center in horizontal and vertical directions. Translation augmentation was employed for the cases when tumor was not centered in the image. The total number of images used in training and testing are listed in table 2.

2.4. Network construction

We employed GoogLeNet, which was established in 2014, and modified the network for our purpose. Two auxiliary classifiers were removed in this research as shown in figure 6. The input layer was modified to deal with gray-scale images not color images. In this research, we have just a two-class problem, benign and malignant lesions. Because GoogLeNet has 1000 class outputs, we reduced the output to two class outputs. All pixels in each patch were treated as input neurons.

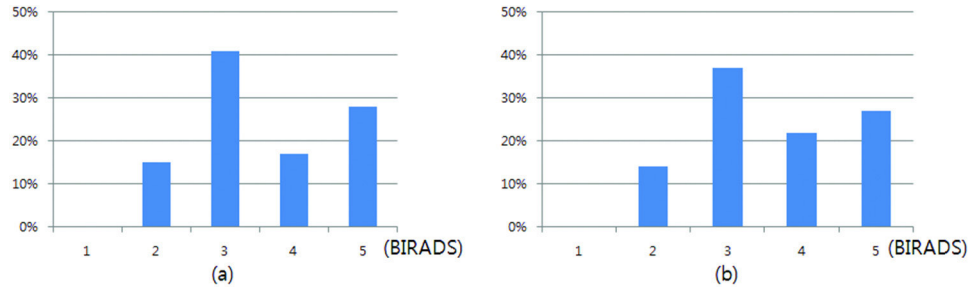

Figure 3. Examples of (a) malignant lesions and (b) benign lesions.

2.4.1. Pretraining networks on gray natural images. Because transfer learning followed by fine-tuning has a better performance than learning from scratch in many cases (Shin *et al* 2016), the network was initialized by the ImageNet pretraining model and fine-tuned to yield better performance. The idea of pretraining the networks has been successfully used to various applications. It involves initializing the weights of the networks by training them on other tasks (the source domain) such as a classification of ImageNet (Deng *et al* 2009) data. This strategy is especially useful when the number of data to be trained (the target domain) is not sufficient to train complex networks with a large number of parameters to learn. This pretraining of the networks has been successfully used in various domains, not only with natural images but also with medical images including mammograms (Lvy and Jain 2016), x-rays (Choi 2015), histological images (Xu *et al* 2016, Chen *et al* 2016) and retinal images (Maninis *et al* 2016, Worrall *et al* 2016).

When the pretraining technique is applied to ultrasound images, the statistics of natural images and ultrasound images are different in many ways. One of them is the channel size of input images. While most natural images have three values for each pixel, gray ultrasound images have only one intensity value. To apply a pretrained model (originally pretrained on natural images of ImageNet) to ultrasound images, the ultrasound images need to be reshaped to have three channels (Tajbakhsh *et al* 2016). This means repeating the same value for three channels of red, green and blue. After that, the pretrained model can be applied since the channel size of the first convolution kernel is three. Instead of tiling the input ultrasound images, the number of channels of the first convolution kernel can be reduced from three to one by summation across the channel.

Figure 7(a) shows the learned weights for the convolution kernels in the first layer of GoogLeNet. They are pretrained on ImageNet data. Some of them look like Gabor filters detecting edges in various orientations. Others look like blob detectors detecting color blobs. The problem appears when it is reduced to deal with gray input images. Figure 7(b) shows the reduced convolution kernels by the summation of weights across channels. While edge detectors remain valid, color blob detectors are reduced to invalid kernels with flattened patterns. As a consequence a certain number of kernels become redundant, losing their representational capacity. Instead, we pretrained the network only on the gray images from the beginning. We converted ImageNet images into gray ones and trained the network on them. The first layer of the network is composed of the convolutions with a channel size of one. Figure 7(c) shows the learned weights for the convolution kernels in the first layer. In comparison, pretraining on the gray images ends in learning much richer convolution kernels with fewer flattened kernels. This one-channel method may potentially lead to better performance. However, this method does not guarantee improved performance even though the learned weights in the first layer look much richer. We used this method to make the network work a little faster than the three-channel input case, which requires processing time to replicate the gray-scale images in the three channels.

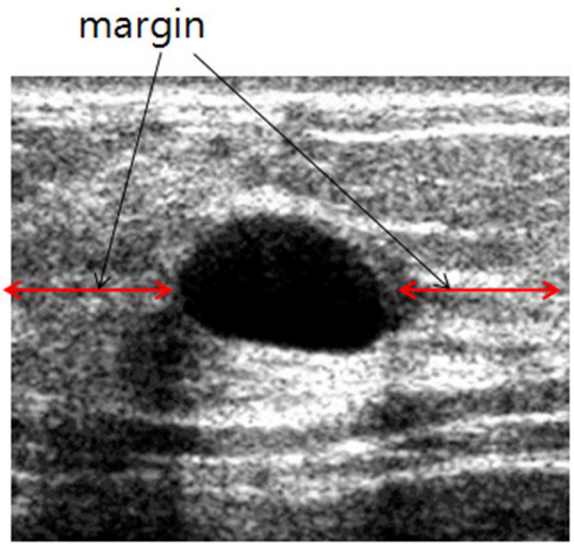

Figure 4. The definition of margin.

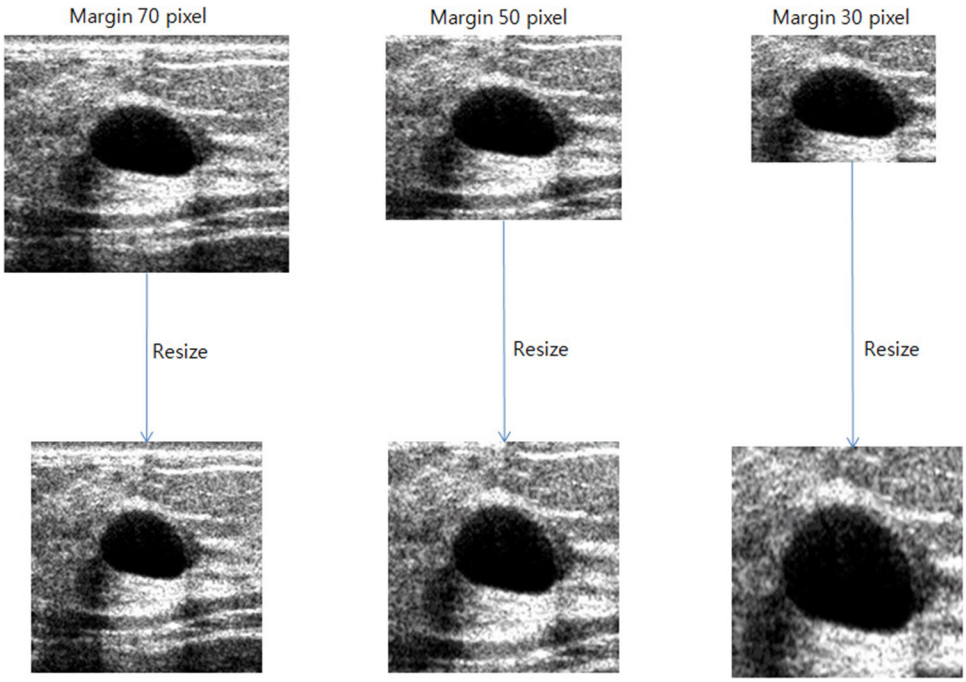

Figure 5. The concept of margin augmentation.

3. Experimental results

We compared the performance of the proposed deep learning framework for breast cancer classification with a conventional method in terms of accuracy, sensitivity, specificity and AUC (area under the curve). Optimal parameters were chosen based on 10-fold cross-

Table 2. Overview of the number of images.

	Training images	Augmented training images	Test images
Benign	3765	553 455	489
Malignant	2814	413 658	340

validation with the training data. Then, the optimized parameters were applied to evaluate the performance on the test dataset.

We present the performance of the one input channel and three input channel methods for comparison in table 3, when the margin was fixed at 180 pixels. As can be seen in table 3, the one-channel method does not guarantee a better performance than the three-channel method. However, we chose to use one-channel method in this research.

Figure 8 shows the effect of the margin size on the classification performance. The validation accuracy values at margin sizes of 0, 30, 50, 70, 90, 120, 150, 180, 210, 240, and 270 pixels are presented. As can be seen in figure 8, a margin of 180 pixels shows the best performance.

Because the validation accuracy has its best performance when the margin size is 180 pixels, we fixed the margin size as 180 pixels. Thus, we evaluated the performance of the proposed framework on the test set when the size of the margin was set to 180 pixels. For comparison, we also evaluated the performance when the size of the margin was set to zero pixels, and the ensemble of the networks trained with zero-pixel margin data and 180-pixel margin data. Those results are presented in figure 9. Table 4 presents the accuracy, sensitivity, specificity and AUC of the zero-pixel margin case, the 180-pixel margin case and the ensemble of the zero-pixel margin case and the 180-pixel margin case. For comparison, we also present the result of a conventional method which utilizes hand-crafted features and the SVM (support vector machine) classifier (Shi *et al* 2010) in figure 9 and table 4.

As we can see from table 4, the performance of CNN is very promising. The networks show an AUC of over 0.95, an accuracy of about 0.9(90%), a sensitivity of 0.83 and specificity of 0.95. Considering the large number of images in the training set and the test set, the proposed framework seems to be very helpful for classifying benign or malignant breast tumors in real clinical applications. In previous methods that use 100–300 images, the AUC ranges from 0.91 (Joo *et al* 2004) to 0.95 (Chen *et al* 2003, Yang *et al* 2013). The AUC range was reported to be from 0.78 to 0.87 in the previous research that used 400–500 images (Gmez *et al* 2012, Wang *et al* 2016), irrespective of the method (deep learning or other methods). In a method that used 2409 images (Drukker *et al* 2009), the AUC value was reported to be 0.92. Considering the AUC values of the previous methods, the result of the proposed method seems to be reasonable.

CNN training is implemented with the Caffe (Jia *et al* 2014) deep learning framework, using a NVidia K40 GPU on Ubuntu 12.04. A model snapshot with low validation loss is taken for the final model. The learning hyperparameters are set as follows: momentum 0.9, weight decay 0.0002, and a poly learning policy with base learning rate of 0.0001. The image batch size is 32, which is the maximum batch size that works with our system.

4. Discussion

The proposed deep learning-based method employing a large dataset seems to have great discriminating performance in the classification of benign and malignant tumors. Even though the location of the tumors had to be pointed out by radiologists, the proposed method could

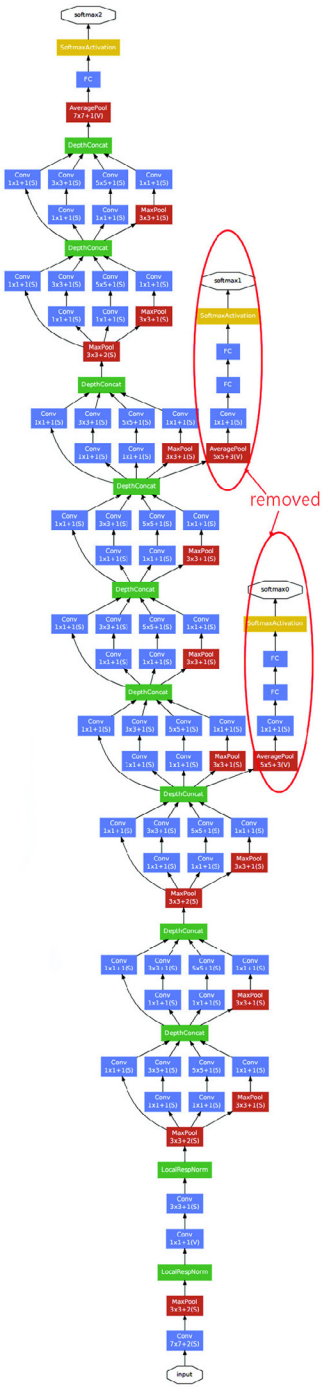

Figure 6. The modification of the original GoogLeNet (Szegedy *et al* 2015) for our research.

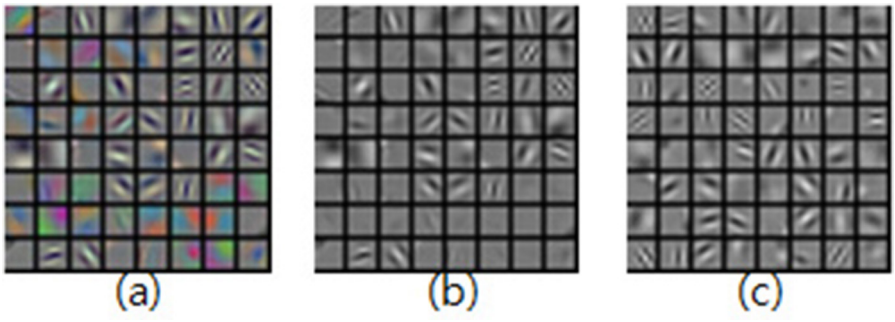

Figure 7. (a) The learned weights for a network composed of convolutions with three channels. (b) The reduced convolution kernels by the summation of weights across channels. (c) The learned weights for the network composed of the convolutions with one channel.

Table 3. Diagnostic performance of the one input channel and three input channel methods.

	Mean accuracy (%)	SD	Min	Max
One channel	90.21	0.3076	90.15	90.40
Three channels	90.24	0.3533	89.91	90.40

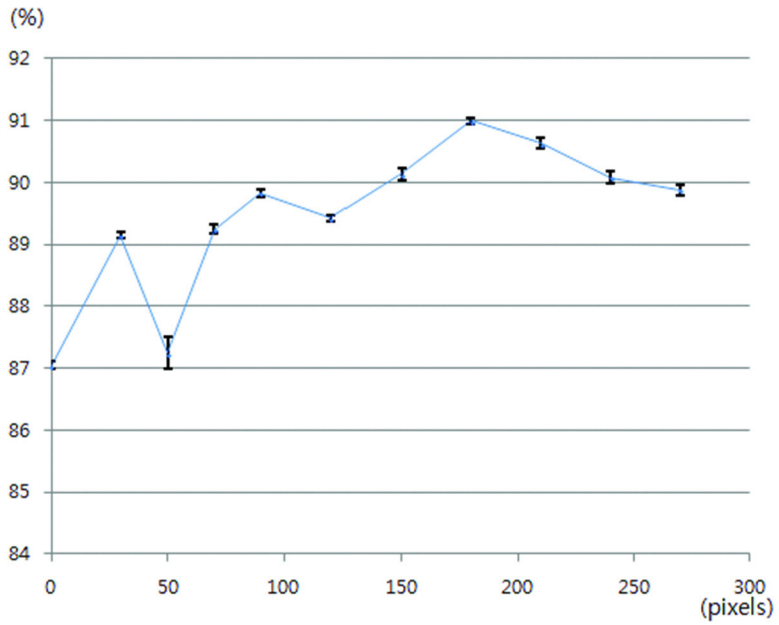

Figure 8. The maximum validation accuracy as the margin size increases. The *X* axis indicates the size of margin while the *Y* axis indicates the validation accuracy.

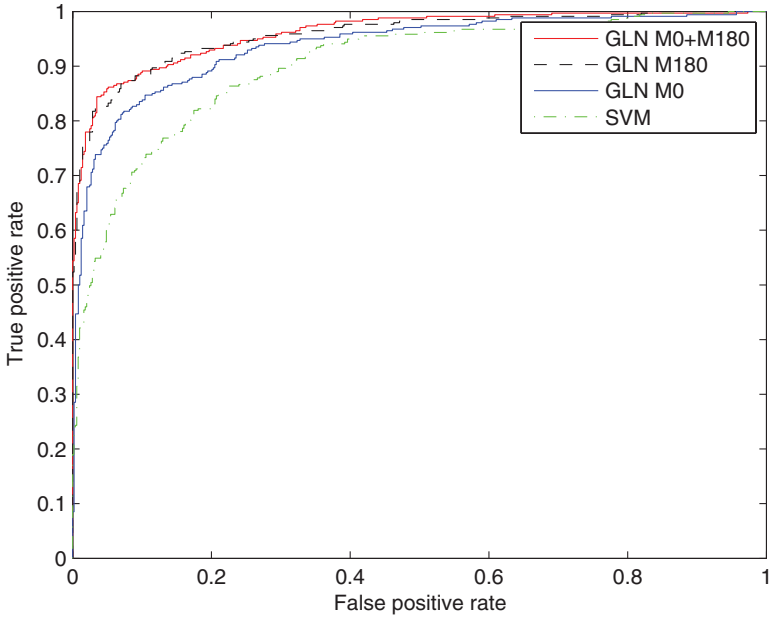

Figure 9. ROC curves of the GoogLeNet results for a zero-pixel margin, a 180-pixel margin, an ensemble of both of them, and SVM.

Table 4. Diagnostic performance of CNNs and ensemble networks.

	Accuracy (%)	Sensitivity	Specificity	AUC
M0	87.85 ± 0.2354	0.7965 ± 0.0244	0.9357 ± 0.0150	0.9315 ± 0.0038
M180	90.21 ± 0.3076	0.8332 ± 0.0096	0.9501 ± 0.0099	0.9575 ± 0.0010
M0 + M180	91.23 ± 0.1087	0.8429 ± 0.0052	0.9607 ± 0.0041	0.9601 ± 0.0004
SVM	82.87	0.7151	0.9065	0.9003

classify the tumors as benign or malignant accurately. Therefore, the proposed framework can help radiologists make accurate decisions on the following procedure. For a very experienced radiologist the proposed method may not be helpful in improving the early detection rate. However, it can be very helpful for radiologists who are not fully trained in ultrasound imaging to make better decisions.

As mentioned above, we have more training data for benign lesions than for malignant ones, which made the deep learning network sensitive to benign tumors, and showed good specificity.

Considerable research has been undertaken to develop CADx systems for radiologists. Unfortunately, because it is very hard to gather a large number of images in the medical imaging field, previous methods often have dealt with small datasets. In this research, we applied a deep learning method to a large dataset obtained from more than 1000 patients, and showed that a deep learning method has the potential for clinical application. We investigated the effect of margin size, which is the distance between the boundary of the tumor and the boundary of the input image itself, on the performance of classifying a tumor as benign or malignant. We increased the size of margin from 0 to 270 pixels. In our database, a 180-pixel margin showed the best validation accuracy. Thus, we set the margin size as 180 pixels, and applied the proposed framework to our test dataset. It showed about 89.86% accuracy and an AUC of

0.9578. To enhance the performance of the network, an ensemble network was also used. We used a zero-pixel margin for the ensemble network. This could produce a better result than a single neural network. Interestingly, though the zero-pixel margin network had 2% lower accuracy than the 180-pixel margin network, an ensemble of the two networks showed a better result than the network trained with the margin size of 180 pixels. This could be because an image with a zero-pixel margin has different information from an image of with a 180-pixel margin, which improved the classification performance.

In this research, target ROIs were selected by the radiologist, which means that radiologists should help to point out the location of ROIs. If this method is used by radiologists in clinical situations it can classify malignant lesions in a short time and support the diagnosis of radiologists in discriminating malignant lesions. Therefore, the proposed method can support human radiologists to have better classification performance, which is accordance with the fundamental purpose of CADx.

5. Conclusion

In this research, we exploited the deep learning framework to differentiate the distinctive types of beast lesions and nodules acquired with ultrasound imaging. A biopsy-proven benchmarking dataset was built and used to evaluate the proposed method. Optimal parameters were chosen based on 10-fold cross-validation with the training data. The networks showed an AUC over 0.9, an accuracy of about 0.91 (91%), a sensitivity of 0.86 and a specificity of 0.93. Although target ROIs were selected by radiologists, which means that radiologists will still have to point out the location of the ROI, the classification of malignant lesions showed promising results and outperformed the conventional method.

ORCID iDs

Seokmin Han https://orcid.org/0000-0001-7243-0657

References

Akay M F 2009 Support vector machines combined with feature selection for breast cancer diagnosis *Expert Syst. Appl.* **36** 3240–7

Arbelaez P *et al* 2011 Contour detection and hierarchical image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **33** 898–916

Arevalo J, Cruz-Roa A and González F A 2013 Hybrid image representation learning model with invariant features for basal cell carcinoma detection *Proc. SPIE* **8922** 89220M

Armato S G III and Sensakovic W F 2004 Automated lung segmentation for thoracic CT: impact on computer-aided diagnosis *Acad. Radiol.* **11** 1011–21

Awai K *et al* 2006 Pulmonary nodules: estimation of malignancy at thin-section helical CT—effect of computer-aided diagnosis on performance of radiologists *Radiology* **239** 276–84

Ayer T *et al* 2010 Computer-aided diagnostic models in breast cancer screening *Imaging Med.* **2** 313–23

Bengio Y *et al* 2013 Representation learning: a review and new perspectives *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 1798–828

Centers for Disease Control and Prevention (CDC) 2017 Cancer among women www.cdc.gov/cancer/dcpc/data/women.htm

Chen C M *et al* 2003 Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks *Radiology* **226** 504–14

Chen H *et al* 2015 Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer) pp 507–14

Chen H *et al* 2016 Mitosis detection in breast cancer histology images via deep cascaded network *Proc. 30th AAAI Conf. on Artificial Intelligence* (Palo Alto, CA: AAAI Press) pp 1160–6
Cheng J Z *et al* 2010 Computer-aided US diagnosis of breast lesions by using cell-based contour grouping *Radiology* **255** 746–54
Cheng J Z *et al* 2016 Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans *Sci. Rep.* **6**
Choi S 2015 X-ray image body part clustering using deep convolutional neural network: SNUMedinfo at ImageCLEF 2015 medical clustering task CLEF (working notes) http://ceur-ws.org/Vol-1391/117-CR.pdf
Cruz J A and Wishartl D S 2006 Applications of machine learning in cancer prediction and prognosis *Cancer Inform.* **2** 59–77
Cruz-Roa A A *et al* 2013 A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (*Lecture Notes in Computer Sciences* vol 8150) (Berlin: Springer) pp 403–10
Deng J *et al* 2009 Imagenet: a large-scale hierarchical image database *IEEE Conf. on Computer Vision and Pattern Recognition*
Doi K 2007 Computer-aided diagnosis in medical imaging: historical review, current status and future potential *Comput. Med. Imaging Graph.* **31** 198–211
Drukker K *et al* 2009 Automated method for improving system performance of computer-aided diagnosis in breast ultrasound *IEEE Trans. Med. Imaging* **28** 122–8
Getreuer P 2012 Chan-vese segmentation *Image Process. Line* **2** 214–24
Giger M L *et al* 2008 Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM *Med. Phys.* **35** 5799–820
Giger M L *et al* 2013 Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer *Annu. Rev. Biomed. Eng.* **15** 327–57
Gmez W *et al* 2012 Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound *IEEE Trans. Med. Imaging* **31** 1889–99
Holsbach N *et al* 2014 A data mining method for breast cancer identification based on a selection of variables *Cienc. Saude Coletiva* **19** 1295–304
Jalalian A *et al* 2013 Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review *Clin. Imaging* **37** 420–6
Jia Y *et al* 2014 Caffe: convolutional architecture for fast feature embedding *Proc. 22nd ACM Int. Conf. Multimedia* pp 675–8
Jiang Y *et al* 2006 Comparison of independent double readings and computer-aided diagnosis (CAD) for the diagnosis of breast calcifications *Acad. Radiol.* **13** 84–94
Joo S *et al* 2004 Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features *IEEE Trans. Med. Imaging* **23** 1292–300
Kornecki A 2011 Current status of breast ultrasound *Can. Assoc. Radiol. J.* **62** 31–40
Krishnan M *et al* 2010 Statistical analysis of mammographic features and its classification using support vector machine *Expert Syst. Appl.* **37** 470–8
Li F *et al* 2015 A robust deep learning for improved classification of AD/MCI patients *IEEE J. Biomed. Health Inform.* **19** 1610–6
Lvy D and Jain A 2016 Breast mass classification from mammograms using deep convolutional neural networks (arXiv:1612.00542)
Maninis K K *et al* 2016 Deep retinal image understanding *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016* (*Lecture Notes in Computer Sciences* vol 9901) (Berlin: Springer) pp 140–8
McCarville M B *et al* 2006 Distinguishing benign from malignant pulmonary nodules with helical chest CT in children with malignant solid tumors *Radiology* **239** 514–20
Newell D *et al* 2010 Selection of diagnostic features on breast MRI to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement *Eur. Radiol.* **20** 771–81
Perez N *et al* 2014 Improving the performance of machine learning classifiers for breast cancer diagnosis based on feature selection *IEEE Federated Conf. Computer Science and Information Systems* pp 209–17
Perez N *et al* 2013 Improving breast cancer classification with mammography, supported on an appropriate variable selection analysis *Proc. SPIE* **8670** 867022

Prasoon A *et al* 2013 Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (*Lecture Notes in Computer Sciences* vol 8150) (Berlin: Springer) 246–53
Roth H *et al* 2016 Improving computer-aided detection using convolutional neural networks and random view aggregation *IEEE Trans. Med. Imaging* **35** 1170–81
Sahiner B *et al* 2001 Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization *IEEE Trans. Med. Imaging* **20** 1275–84
Sahan S *et al* 2007 A new hybrid method based on fuzzy-artificial immune system and *k*-nn algorithm for breast cancer diagnosis *Comput. Biol. Med.* **37** 415–23
Seff A *et al* 2015 Leveraging mid-level semantic boundary cues for automated lymph node detection *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (*Lecture Notes in Computer Sciences* vol 9350) (Berlin: Springer) pp 53–61
Shi X *et al* 2010 Detection and classification of masses in breast ultrasound images *Digit. Signal Process.* **20** 824–36
Shin H C *et al* 2013 Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 1930–43
Shin H C *et al* 2016 Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning *IEEE Trans. Med. Imaging* **35** 1285–98
Sluimer I C *et al* 2003 Computer-aided diagnosis in high resolution CT of the lungs *Med. Phys.* **30** 3081–90
Sorensen L 2010 Quantitative analysis of pulmonary emphysema using local binary patterns *IEEE Trans. Med. Imaging* **29** 559–69
Sun T *et al* 2013 Computer-aided diagnosis for early-stage lung cancer based on longitudinal and balanced data *PLoS One* **8** e63559
Suk H I and Shen D 2013 Deep learning-based feature representation for AD/MCI classification *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (*Lecture Notes in Computer Sciences* vol 8150) (Berlin: Springer) pp 583–90
Suk H I *et al* 2015 Latent feature representation with stacked auto-encoder for AD/MCI diagnosis *Brain Struct. Funct.* **220** 841–59
Suk H I *et al* 2014 Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis *NeuroImage* **101** 569–82
Szegedy C *et al* 2015 Going deeper with convolutions *IEEE Conf. on Computer Vision and Pattern Recognition*
Tabr L *et al* 2011 Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades *Radiology* **260** 658–63
Tajbakhsh N *et al* 2015 Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (*Lecture Notes in Computer Sciences*) vol 9350 (Berlin: Springer) pp 62–9
Tajbakhsh N *et al* 2016 Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans. Med. Imaging* **35** 1299–312
Tourassi G D 1999 Journey toward computer-aided diagnosis: role of image texture analysis *Radiology* **213** 317–20
Tourassi G D *et al* 2001 Application of the mutual information criterion for feature selection in computer-aided diagnosis *Med. Phys.* **28** 2394–402
van Ginneken B 2011 Computer-aided diagnosis: how to move from the laboratory to the clinic *Radiology* **261** 719–32
Vishrutha V and Ravishankar M 2014 Early detection and classification of breast cancer *Proc. 3rd Int. Conf. on Frontiers of Intelligent Computing: Theory and Applications* pp 413–9
Wang J *et al* 2016 Discrimination of breast cancer with microcalcifications on mammography by deep learning *Sci. Rep.* **6**
Wang D *et al* 2009 Automatic detection of breast cancers in mammograms using structured support vector machines *Neurocomputing* **72** 3296–302
Way T W *et al* 2009 Computer-aided diagnosis of pulmonary nodules on CT scans: improvement of classification performance with nodule surface features *Med. Phys.* **36** 3086–98
Way T W *et al* 2006 Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours *Med. Phys.* **33** 2323–37

Worrall D E, Wilson C M and Brostow G J 2016 Automated retinopathy of prematurity case detection with convolutional neural networks *Int. Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis* pp 68–76

Xu Y *et al* 2016 Gland instance segmentation by deep multichannel side supervision *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016* (*Lecture Notes in Computer Sciences* vol 9901) (Berlin: Springer) pp 496–504

Yang M *et al* 2013 Robust texture analysis using multi-resolution gray-scale invariant features for breast sonographic tumor diagnosis *IEEE Trans. Med. Imaging* **32** 2262–73

Zhang W *et al* 2015 Deep convolutional neural networks for multi-modality isointense infant brain image segmentation *NeuroImage* **108** 214–24