ORIGINAL PAPER

Computer-Aided Diagnosis Scheme for Distinguishing Between Benign and Malignant Masses on Breast DCE-MRI Images Using Deep Convolutional Neural Network with Bayesian Optimization

Akiyoshi Hizukuri[1] · Ryohei Nakayama[1] · Mayumi Nara[2] · Megumi Suzuki[2] · Kiyoshi Namba[2]

Abstract

Although magnetic resonance imaging (MRI) has a higher sensitivity of early breast cancer than mammography, the specificity is lower. The purpose of this study was to develop a computer-aided diagnosis (CAD) scheme for distinguishing between benign and malignant breast masses on dynamic contrast material-enhanced MRI (DCE-MRI) by using a deep convolutional neural network (DCNN) with Bayesian optimization. Our database consisted of 56 DCE-MRI examinations for 56 patients, each of which contained five sequential phase images. It included 26 benign and 30 malignant masses. In this study, we first determined a baseline DCNN model from well-known DCNN models in terms of classification performance. The optimum architecture of the DCNN model was determined by changing the hyperparameters of the baseline DCNN model such as the number of layers, the filter size, and the number of filters using Bayesian optimization. As the input of the proposed DCNN model, rectangular regions of interest which include an entire mass were selected from each of DCE-MRI images by an experienced radiologist. Three-fold cross validation method was used for training and testing of the proposed DCNN model. The classification accuracy, the sensitivity, the specificity, the positive predictive value, and the negative predictive value were 92.9% (52/56), 93.3% (28/30), 92.3% (24/26), 93.3% (28/30), and 92.3% (24/26), respectively. These results were substantially greater than those with the conventional method based on handcrafted features and a classifier. The proposed DCNN model achieved high classification performance and would be useful in differential diagnoses of masses in breast DCE-MRI images as a diagnostic aid.

Keywords Breast magnetic resonance imaging · Mass · Deep convolutional neural network · Bayesian optimization · Computer-aided diagnosis

Introduction

Breast cancer is the most common cancer in women. In the USA, 40,610 women died due to breast cancer in 2017 [1]. Early detection and treatment are very important for breast cancer patients since 95% of patients with early breast cancer can be cured completely [2]. Dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) has become an established clinical imaging modality for diagnosis and staging of breast cancer [3, 4]. The sensitivity of breast DCE-MRI is higher than that of mammography, which is the standard screening modality. Especially in dense breast, the sensitivity has been improved from 33–59% with mammography to 71–94% with DCE-MRI [5–9]. However, the specificity of DCE-MRI, which is typically between 30–70%, is lower than that of mammography [10–14].

A computer-aided diagnosis (CAD) scheme [15] is one of the solutions to improve the specificity of breast MRI. The CAD scheme presents the likelihood of malignancy for a lesion on medical image as a "second opinion" in order to assist radiologists' diagnosis [15–18]. In our previous study, we developed the CAD scheme for distinguishing between benign and malignant masses on breast DCE-MRI [19]. The CAD scheme was based on the conventional method with the handcrafted features and a classifier. In the CAD scheme,

✉ Akiyoshi Hizukuri
hizukuri@fc.ritsumei.ac.jp

1 Department of Electronic and Computer Engineering, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577,, Japan

2 Department of Breast Surgery, Hokuto Hospital, 7-5 Kisen, Inada-cho, Obihiro-shi, Hokkaido 080-0833,, Japan

mass region was segmented from DCE-MRI images. Objective features were then extracted from the segmented mass region to distinguish between benign and malignant masses. Therefore, there was a problem that classification performance was strongly dependent on the segmented mass region.

Recently, deep convolutional neural networks (DCNNs) such as AlexNet, ZFNet, VGG16, and GoogLeNet have been applied to classification tasks [20–23]. The DCNNs have shown greater classification performances than the conventional methods with the handcrafted features and a classifier in an ImageNet Large Scale Visual Recognition Competition (ILSVRC) [24]. The DCNN can extract complex multi-level objective features from input images due to self-learning ability without the segmentation of target [25–27]. Therefore, there is not a problem that the classification performance is influenced from the segmented region. As with many studies [18, 28–31], the classification accuracy of our previous CAD scheme based on the conventional method might be improved by using the DCNN. The well-known DCNN models such as AlexNet, ZFNet, VGG16, and GoogLeNet were constructed for general images with RGB channels. Therefore, those models can be inadequacy in the classification task for medical images with grayscale channel.

The purpose of this study was to determine the optimum architecture of the DCNN models for distinguishing benign from malignant masses on breast DCE-MRI using Bayesian optimization [32–35]. In this study, we first determined a baseline DCNN model from well-known DCNN models in term of classification performance. The optimum architecture of the DCNN model was clarified by changing the hyperparameters of the baseline DCNN model such as the number of layers, the filter size, and the number of filters using Bayesian optimization. The usefulness of the optimized DCNN model was evaluated by comparing with conventional DCNN models.

Materials and Methods

The use of the following database was approved by the Institutional Review Board at Ritsumeikan University. The database was stripped of all patient identifiers.

Materials

Our database consisted of 56 DCE-MRI examinations—each of which contained five sequential phase images—that were obtained from 56 patients (mean age: 55.8 years, age range: 20–82 years). These DCE-MRI images were acquired with a 3.0 T MR scanner at Hokuto Hospital (Obihiro, Japan) from October 2009 and July 2015. The patients were excluded if they had undergone breast surgery in the past, size of mass

was more than 5 cm. It included 30 malignant and 26 benign breast masses. Table 1 shows the patients' clinical information. All masses underwent 10 G vacuum-assisted biopsy and/or surgical specimen. After the injection of a contrast agent, four post-contrast series of 3D MRI scans and data acquisitions were sequentially performed after a duration of 0 min, 1 min, 2 min, and 4 min. The one pre-contrast and the four post-contrast series generated images with a spatial resolution of $0.7 \times 0.7 \times 1.2$ mm^3, with a data matrix of 512×512 pixels. Figure 1 shows an example of pre-contrast and four post-contrast DCE-MRI images. Each of five image scan series consisted of 150 image slices.

### Determination of Baseline DCNN Model

To optimize the architecture of DCNN model for distinguishing between benign and malignant masses on DCE-MRI images, we first determined a baseline DCNN model from AlexNet, ZFNet, VGG16, and GoogLeNet in terms of area under the receiver operating characteristic (ROC) curve [36]. Here, we briefly described those DCNN models below.

AlexNet consists of five convolutional layers, three max-pooling layers, and three fully connected layers including cross channel normalization layer, rectified linear unit (ReLU) function, and dropout. The convolutional layer and the max-pooling layer play a role of feature extractor, whereas the fully connected layer plays a role of classifier. The first convolutional layer has 96 filters of size 11×11 with a stride of four pixels and padding with two pixels. The second convolutional layer has 256 filters of size 5×5. The third, fourth, and fifth convolutional layers have 384, 384, and 256 filters with

Table 1 Patients' clinical information

Patients characteristics	
# of patients	56
Age, mean ± SD	55.8 ± 15.3
Age, range	20–82
Age group	
20–30	3
30–40	8
40–50	12
50–60	11
60–70	11
70–82	11
# of malignant masses	30
# of benign masses	26
Mass characteristics	
Size of masses, mean ± SD [cm]	1.7 ± 0.6
# of masses (≦ 2 cm)	34
# of masses (2–5 cm)	22

Fig. 1 Example of DCE-MRI images with a malignant mass before injection of a contrast agent, and after a duration of 0, 1, 2, and 4 min

Before injection of a contrast agent

0 min

1 min

2 min

4 min
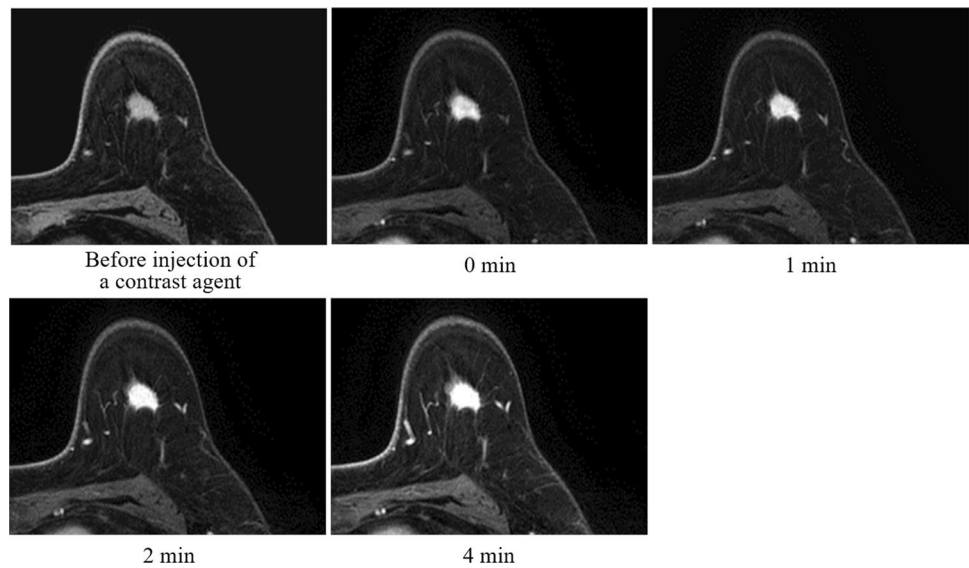
size of 3 × 3, respectively. The number of the units in the first and second fully connected layer is 4,096, whereas that of the units in the third fully connected layer is same as the number of the classes. In this study, the number of units in the third fully connected layer is two (Benign or Malignant).

ZFNet is similar architecture to AlexNet. The difference between ZFNet and AlexNet is only filter size and stride in first convolutional layer. The filter size of the first convolutional layer in ZFNet is 7 × 7, whereas the stride is two pixels. The other parameters of ZFNet are same as AlexNet.

VGG16 consists of 13 convolutional layers with filter size of 3 × 3, and three fully connected layers including ReLU function and dropout. The configurations of fully-connected layers in VGG16 are the same with AlexNet.

GoogLeNet consists of 22 layers with nine inception layers and one fully connected layer. The inception layer has multiple convolution filters [23].

Optimization of DCNN Architecture with Bayesian Optimization

After the determination of baseline DCNN model with the highest area under the ROC curve (AUC), the hyperparameters such as the number of the convolutional layer, the number of filters, and the filter size were optimized in the baseline DCNN model using Bayesian optimization with Gaussian process [32–35]. The Bayesian optimization is an algorithm for optimizing hyperparameters in a machine learning. Table 2 shows the search range for each hyperparameter in the DCNN model. When the number of the convolutional layer was − 4, it means to remove the final and the fourth from the last convolutional layer in the baseline DCNN model. On the

other hand, when the number of the convolutional layer was + 4, it means to add four convolutional layers after the final convolutional layer. With the number of the convolutional layer, the number of the max-pooling layer, and the number of the fully connected layer were 0, the configuration of the DCNN model was same as the baseline DCNN model.

In the Bayesian optimization, four different combinations of hyperparameters were first determined by selecting search value randomly in each hyperparameter. The DCNN model with each combination was then trained independently. With a Gaussian process based on the classification errors for each DCNN model, the combination of hyperparameters was updated to reduce the classification error. By repeating this process 100 times, the optimal combination of hyperparameters was founded efficiently.

Table 2 Candidate values for each hyperparameter

Hyperparameter	Values
Number of convolutional layers	− 4~ + 4
Number of max-pooling layers	− 3~ + 3
Number of fully connected layers	− 2~ + 1
Filter size	3~9
Number of filters	48~576
Number of units in fully connected layer	2048~6144
Minibatch size	4~20
Dropout	0.25~0.75
Initial learning rates	$0.5e^{-4}$~$1.5e^{-4}$
Momentum	0.8~0.95
L2 Regularization	$1e^{-10}$~$1e^{-2}$

Training and Testing of DCNN

The DCNN model was developed and evaluated using MATLAB 2019a on a workstation (CPU: Intel Core i7-7820X processor, RAM: 128 GB, and GPU: NVIDIA GeForce GTX 1080Ti).

A k-fold cross validation method [37] with $k = 3$ was utilized for the training and testing of the DCNN model. In the method, 56 patients were randomly divided into three groups (A, B, C) so that the number of patients was approximately equal. A group was used for test dataset, whereas the remaining two groups were used for training dataset. This process was repeated three times until every group had been used for test dataset.

The ROIs which include an entire mass were selected from each DCE-MRI image by an experienced radiologist (12 years of experience devoted in breast image diagnosis). For augmenting each training data, each ROI was flipped horizontally and cropped [38] randomly eight times. Thus, the total number of training samples was increased by 16 times. A stochastic gradient descent (SGD) was employed to minimize the loss between the output of the proposed DCNN model and the corresponding teacher signal.

Evaluation of Classification Performance

The classification accuracy, the sensitivity, the specificity, the positive predictive value (PPV), and the negative predictive value (NPV) [39] of the DCNN model were evaluated by using the ensemble average from the testing datasets over the threefold cross validation method. ROC analysis [36] was also used for evaluation of classification performance.

Results

The baseline DCNN model was first determined from AlexNet, ZFNet, VGG16, and GoogLeNet in terms of the AUC. Here, the learning rate, the mini-batch size, and the number of epochs were given by 0.0001, 3, and 15, respectively. The AUC for ZFNet was 0.889, showing to be greater than those for AlexNet (0.867, P = 0.302), VGG16 (0.800, P = 0.050), and GoogLeNet (AUC = 0.827, P = 0.080). Therefore, ZFNet was determined as the baseline DCNN model.

The hyperparameters in the determined baseline DCNN model were optimized using Bayesian optimization. Tables 3 and 4 show the optimal architecture and the parameters

Table 3 Optimized DCNN architecture determined by Bayesian optimization

Layer	Test dataset A	Test dataset B	Test dataset C
Input	$224 \times 224 \times 5$	$224 \times 224 \times 5$	$224 \times 224 \times 5$
Conv. + ReLu	Filter size: 7×7 Num. of filter: 193	Filter size: $7 \times 7x$ Num. of filter: 214	Filter size: 7×7 Num. of filter: 160
Normalization	Size of the channel window: 5	Size of the channel window: 5	Size of the channel window: 5
Max-pooling	Filter size: 3×3	Filter size: 3×3	Filter size: 3×3
Conv. + ReLu	Filter size: 5×5 Num. of filter: 49	Filter size: 5×5 Num. of filter: 442	Filter size: Num. of filter: 148
Normalization	Size of the channel window: 5	Size of the channel window: 5	Size of the channel window: 5
Max-pooling	Filter size: 3×3	Filter size: 3×3	Filter size: 3×3
Conv. + ReLu	Filter size: 5×5 Num. of filter: 440	Filter size: 5×5 Num. of filter: 533	Filter size: Num. of filter: 122
Conv. + ReLu	Filter size: 5×5 Num. of filter: 440	Filter size: 5×5 Num. of filter: 533	Filter size: Num. of filter: 122
Conv. + ReLu	Filter size: 5×5 Num. of filter: 440	Filter size: 5×5 Num. of filter: 533	Filter size: Num. of filter: 122
Max-pooling	Filter size: 3×3	Filter size: 3×3	Filter size: 3×3
Conv. + ReLu	Filter size: 3×3 Num. of filter: 123	Filter size: 3×3 Num. of filter: 96	Filter size: 3×3 Num. of filter: 209
Conv. + ReLu	Filter size: 3×3 Num. of filter: 134	Filter size: 3×3 Num. of filter: 104	Filter size: 3×3 Num. of filter: 228
Conv. + ReLu	Filter size: 3×3 Num. of filter: 146	Filter size: 3×3 Num. of filter: 113	Filter size: 3×3 Num. of filter: 247
Conv. + ReLu	Filter size: 3×3 Num. of filter: 157	Filter size: 3×3 Num. of filter: 122	Filter size: 3×3 Num. of filter: 266
FC	2	2	2
Output	Benign/Malignant	Benign/Malignant	Benign/Malignant

Table 4 Optimized hyperparameters relating to learning determined by Bayesian optimization

|	Test dataset A	Test dataset B	Test dataset C
Minibatch size	20	10	19
Initial learning rates	9.461e−05	5.631e−05	7.550e−05
Momentum	0.945	0.919	0.949
L2 regularization	1.529e−09	1.953e−10	1.087e−06

determined by Bayesian optimization when each group was used for testing dataset, respectively. The average AUC of the determined DCNN models for three test datasets was 0.945.

Figure 2 compares the ROC curve for the proposed DCNN model with those for AlexNet, ZFNet, VGG16, and GoogLeNet. The AUC for the proposed DCNN model (0.945) was significantly higher than those for the other four DCNN models (AlexNet: 0.867, $P = 0.015$; ZFNet: 0.889, $P = 0.026$; VGG16: 0.800, $P = 0.002$, GoogLeNet: 0.827, $P = 0.006$). Table 5 shows average classification accuracy among five different DCNN models. All evaluation indices for the proposed DCNN model was the highest among the five different DCNN models.

Figure 3 shows example of correctly classified cases and incorrectly classified cases in DCE-MRI images at 1-min post-contrast. The characteristics of masses incorrectly classified by the proposed DCNN model were as follows: (1) small masses (2 cm or lower in size), (2) malignant masses with regularity in shape, and (3) benign masses with irregularity in shape.

Discussion

To investigate the usefulness of the proposed DCNN model, we compared with the classification performance with our previous method based on the handcrafted features and a classifier [19]. In the previous method, ROI which included an entire mass was selected manually on the DCE-MRI image. The mass region was determined automatically by applying Otsu's method [40]. Quadratic discriminant analysis (QDA) was employed to distinguish between benign and
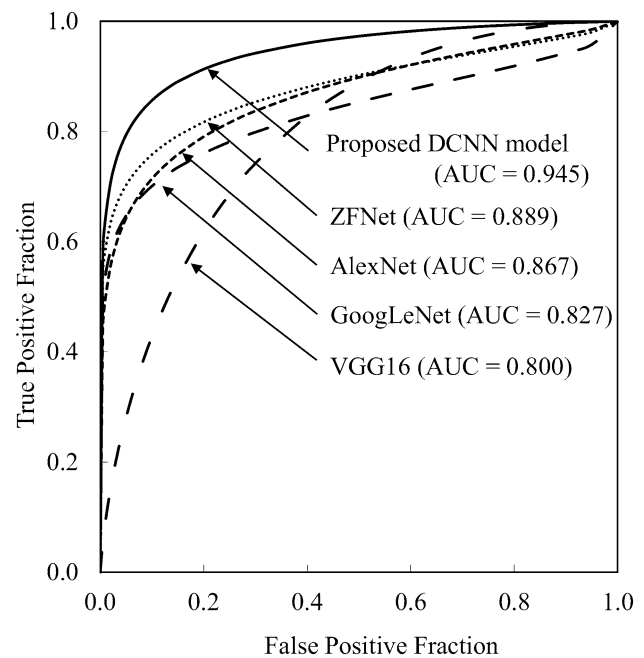
Fig. 2 Comparison of the receiver operating characteristic (ROC) curves between proposed DCNN model, AlexNet, ZFNet, VGG16, and GoogLeNet
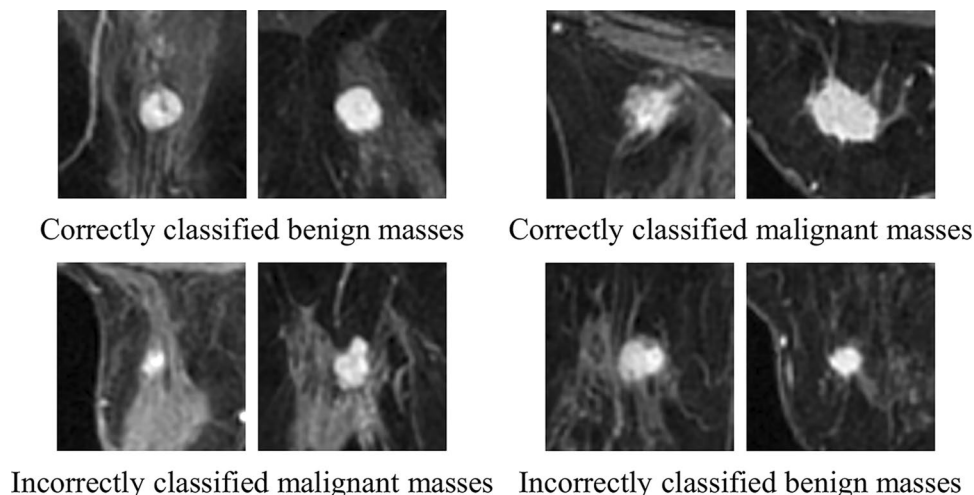
malignant masses. The four handcrafted features were used for the input of the QDA. With the previous method, the average classification accuracy, the sensitivity, the specificity, the PPV, and the NPV were 75.0% (42/56), 76.7% (23/30), 73.1% (19/26), 76.7% (23/30), and 73.1% (19/26), respectively. Figure 4 compares the ROC curve for the proposed DCNN model with that for the previous method. The AUC for the proposed DCNN model was significantly greater than that for the previous method (0.810, $P = 0.01$). These results would imply that the features extracted automatically from the proposed DCNN model were more useful for distinguishing between benign and malignant masses when compared with handcrafted features manually obtained in our previous method.

With the proposed DCNN model, the classification accuracy, sensitivity, specificity, PPV, and NPV for 34 masses smaller than 2 cm were 88.2% (30/34), 84.6% (11/13), 90.5% (19/21), 84.6% (11/13), and 90.5% (19/21), whereas

Table 5 Comparison of classification accuracy among five different DCNN models

Index	AlexNet	ZFNet	VGG16	GoogLeNet	Proposed method
Classification accuracy	78.6% (44/56)	76.8% (43/56)	76.8% (43/56)	75.0% (42/56)	92.9% (52/56)
Sensitivity	76.7% (23/30)	76.7% (23/30)	83.3% (25/30)	73.3% (22/30)	93.3% (28/30)
Specificity	80.8% (21/26)	76.9% (20/26)	69.2% (18/26)	76.9% (20/26)	92.3% (24/26)
PPV	82.1% (23/28)	79.3% (23/29)	75.8% (25/33)	78.6% (22/28)	93.3% (28/30)
NPV	75.0% (21/28)	74.1% (20/27)	78.3% (18/23)	71.4% (20/28)	92.3% (24/26)

Fig. 3 Example of correctly classified cases and incorrectly classified cases in DCE-MRI images at 1-min post-contrast

Correctly classified benign masses   Correctly classified malignant masses

Incorrectly classified malignant masses   Incorrectly classified benign masses

those for 22 masses larger than 2 cm were 100% (22/22), 100% (17/17), 100% (5/5), 100% (17/17), and 100% (5/5). We believe that the proposed DCNN model can reduce the number of unnecessary biopsies for masses larger than 2 cm. Patients with masses smaller than 2 cm which are early lesions will be able to undergo follow-up at a short interval. The classification accuracy for those masses might be improved by introducing the differences in growth speed between benign and malignant cases into the proposed method[41,42]. However, further studies are required by use of large data sets to evaluate the computerized method for analysis of changes over time.
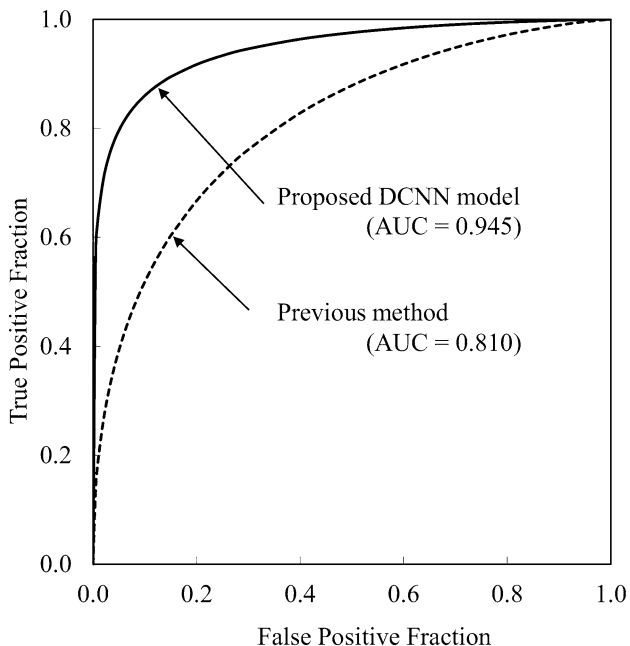

Fig. 4 Comparison of the receiver operating characteristic (ROC) curves between the proposed DCNN model and the previous method

In this study, a strategy for optimizing the hyperparameters in the DCNN model was proposed. It is very difficult to construct suitable DCNN model for classification tasks from scratch, because the combination of hyperparameters in the DCNN is infinite. In the proposed strategy, we determined baseline DCNN model from well-known DCNN models, and then optimized architecture and the other parameters of the baseline DCNN model for breast mass classification task using Bayesian optimization. With this strategy, we could easily construct suitable DCNN architecture according to tasks. In the fact, this study has shown that the proposed DCNN model achieved better classification accuracy than that of our previous method and well-known DCNN model. Therefore, we believe that this strategy is effective for optimizing DCNN architecture and its parameters in classification task.

There are some limitations in this study. We used only 56 patient data in this study. Thus, we need to evaluate the strategy for optimizing the hyperparameters and the proposed DCNN model in the further study by using larger dataset. Other limitation is that the proposed DCNN model used two-dimensional (2D) ROIs as the input. In clinical practice, radiologists usually diagnose by considering three-dimensional (3D) information on DCE-MRI. Therefore, 3D-based DCNN model might be more appropriate than 2D. However, 3D-based DCNN model has a lot of parameters to train and requires a large number of training data. In this study with a small dataset, 2D-based DCNN model was employed to distinguish between benign and malignant masses. Finally, the DCNN model determined by the Bayesian optimization might not yield the best classification accuracy because the Bayesian optimization does not evaluate all combinations of hyperparameters. However, we believe that Bayesian optimization is useful for effectively determining appropriate hyperparameters of the DCNN model from infinite combinations.

Conclusion

In this study, we developed the CAD scheme to distinguish benign from malignant masses on breast DCE-MRI images by use of the optimal DCNN model determined with Bayesian optimization. The proposed DCNN model achieved high classification accuracy for masses and would be useful in differential diagnoses of masses as a diagnostic aid.

References

1. Siegel RL, Miller KD, Jemal A, et al: Cancer statistics, 2017. CA: a cancer journal for clinicians 2017; 67(1): 7–30
2. Reynolds HE, Jackson VP: Self-referred mammography patients: analysis of patients' characteristics. American Journal of Roentgenology 1991; 157: 481-484
3. Kuhl CK, Schild HH: Dynamic image interpretation of MRI of the breast. Journal of Magnetic Resonance Imaging 2000; 12(6): 965-974
4. Gubern-Mérida A, Martí R, Melendez J, Hauth JL, Mann RM, Karssemeijer N, Platel B, et al: Automated localization of breast cancer in DCE-MRI. Medical image analysis 2015; 20(1): 265-274
5. Leach MO, Boggis CR, Dixon AK, et al: Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS), LANCET 2005; 365: 1769-1778
6. Peters NH, Borel Rinkes IH, Zuithoff NP, Mali WP, Moons KG, Peeters PH, et al: Meta-analysis of MR imaging in the diagnosis of breast lesions. Radiology 2008; 246(1): 116-124
7. Sardanelli F, Podo F, D'Agnolo G, et al: Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRIT study): interim results. Radiology 2007; 242(3): 698-715
8. Kuhl CK, Schrading S, Leutner CC, et al: Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. Journal of Clinical Oncology 2005; 23: 8469–8476
9. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, D'Orsi CJ, et al: Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology 2008; 246(2): 376-383
10. Nie K, Chen JH, Yu HJ, Chu Y, Nalcioglu O, Su MY: Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. Academic Radiology 2008; 15(12): 1513–1525
11. Siegmann KC, Krämer B, Claussen C, et al: Current status and new developments in breast MRI. Breast Care (Basel) 2011; 6(2): 87–92
12. Fusco R, Sansone M, Filice S, Carone G, Amato DM, Sansone C, Petrillo A, et al: Pattern recognition approaches for breast cancer DCE-MRI classification: a systematic review. Journal of medical and biological engineering 2016; 36(4): 449-459
13. Berg WA, Blume JD, Cormack JB, et al.: Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. The Journal of the American Medical Association 2008; 299(18): 2151-2163
14. Chae EY, Kim HH, Cha JH, Shin HJ, Kim H, et al: Evaluation of Screening Whole-Breast Sonography as a Supplemental Tool in Conjunction with Mammography in Women With Dense Breasts. Journal of Ultrasound in Medicine 2013; 32(9): 1573-1578
15. Doi K, MacMahon H, Katsuragawa S, Nishikawa RM, Jiang Y, et al: Computer-aided diagnosis in radiology: potential and pitfalls. European journal of Radiology 1999; 31(2): 97–109
16. Doi K: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Computerized medical imaging and graphics 2007; 31(4–5): 198–211
17. Hizukuri A, Nakayama R, Kashikura Y, Takase H, Kawanaka H, Ogawa T, Tsuruoka S, et al: Computerized determination scheme for histological classification of breast mass using objective features corresponding to clinicians' subjective impressions on ultrasonographic images. Journal of digital imaging 2013; 26(5): 958-970
18. Hizukuri A, Nakayama R: Computer-Aided Diagnosis Scheme for Determining Histological Classification of Breast Lesions on Ultrasonographic Images Using Convolutional Neural Network. Diagnostics 2018; 8(3): 1-9
19. Honda E, Nakayama R, Koyama H, Yamashita A, et al: Computer-aided diagnosis scheme for distinguishing between benign and malignant masses in breast DCE-MRI. Journal of digital imaging 2016; 29(3): 388-393
20. Krizhevsky A, Sutskever I, Hinton GE: Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 2012; 1097–1105
21. Zeiler MD, Fergus R: Visualizing and understanding convolutional networks. In European conference on computer vision, Springer, Cham, 2014; 818-833
22. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. Int. Conf. on Learning Representations 2015; San Diego, CA
23. Christian S, et.al.: Going deeper with convolutions, Proceedings of the IEEE conference on computer vision and pattern recognition. 2015
24. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L, et al: ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision, 2015; 115(3); 211-252
25. LeCun Y, Bengio Y, Hinton G, et al: Deep learning. nature 2015; 521(7553): 436
26. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Sánchez CI, et al: A survey on deep learning in medical image analysis, Medical image analysis 2017; 42: 60-88
27. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM, Giger ML, et al: Deep learning in medical imaging and radiation therapy, Medical Physics 2019; 46(1)
28. Bayramoglu N, Juho K, Janne H, et al: Deep learning for magnification independent breast cancer histopathology image classification, 23rd International conference on pattern recognition (ICPR), 2016; 2440–2445
29. Jadoon MM, Zhang Q, Haq IU, Butt S, Jadoon A, et al: Three-class mammogram classification based on descriptive CNN features, BioMed Research International, 2017, 1–11
30. Li Q, Cai W, Wang X, Zhou Y, Feng DD, Chen M, et al: Medical image classification with convolutional neural network, 13th International Conference on Control Automation Robotics & Vision (ICARCV), 2014, 844–848
31. Nejad EM, Affendey LS, Latip RB, Bin Ishak, et al: Classification of histopathology images of breast into benign and malignant using a single-layer convolutional neural network, Proceedings of the International Conference on Imaging, Signal Processing and Communication, 2017, 50–53
32. Rasmussen CE: Gaussian processes in machine learning. In Summer School on Machine Learning, Springer, Berlin, Heidelberg, 2003, 63-71

33. Le HT, Phung SL, Bouzerdoum A, Tivive FHC, et al: Human motion classification with micro-doppler radar and bayesian-optimized convolutional neural networks. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) , 2018, 2961–2965

34. Snoek J, Larochelle H, Adams RP: Practical Bayesian optimization of machine learning algorithms. In Advances in Neural Information Processing Systems, 2012, 2951–2959

35. Dernoncourt F, Lee JY: Optimizing neural network hyperparameters with Gaussian processes for dialog act classification. In 2016 IEEE Spoken Language Technology Workshop (SLT), 2016, 406–413

36. Metz CE: ROC methodology in radiologic imaging. Investigative Radiology 1986; 21: 720–733

37. Peng Y, Jiang Y, Antic T, Giger ML, Eggener SE, Oto A, et al: Validation of quantitative analysis of multiparametric prostate MR images for prostate cancer detection and aggressiveness assessment: a cross-imager study. Radiology 2014; 271(2): 461-471

38. Cireşan DC, Meier U, Gambardella LM, Schmidhuber J, et al: Deep, big, simple neuralnets for handwritten digit recognition. Neural computation 2010; 22(12): 3207-3220

39. Langlotz CP: Fundamental measures of diagnostic examination performance: Usefulness for clinicaldecision making and research. Radiology 2003; 228: 3–9

40. Otsu N: A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics 1979; 9(1): 62–66

41. Nakayama R, Uchiyama Y, Watanabe R, Katsuragawa S, Namba K, Doi K, et al: Computer-aided diagnosis scheme for histological classification of clustered microcalcifications on magnification mammograms. Medical physics 2004; 31(4): 789–799

42. Nakayama R, Watanabe R, Namba K, Takeda K, Yamamoto K, Katsuragawa S, Doi K, et al: Computer-aided diagnosis scheme for identifying histological classification of clustered microcalcifications by use of follow-up magnification mammograms. Academic radiology 2006; 13(10): 1219–1228