Diagnostic Accuracy Study

Medicine

OPEN

Differentiation of breast lesions on dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) using deep transfer learning based on DenseNet201

Mingzhu Meng, MD[a], Ming Zhang, MD[a], Dong Shen, MB[a], Guangyuan He, MB[a,*]

Abstract

In order to achieve better performance, artificial intelligence is used in breast cancer diagnosis. In this study, we evaluated the efficacy of different fine-tuning strategies of deep transfer learning (DTL) based on the DenseNet201 model to differentiate malignant from benign lesions on breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). We collected 4260 images of benign lesions and 4140 images of malignant lesions of the breast pertaining to pathologically confirmed cases. The benign and malignant groups was randomly divided into a training set and a testing set at a ratio of 9:1. A DTL model based on the DenseNet201 model was established, and the effectiveness of 4 fine-tuning strategies (S0: strategy 0, S1: strategy; S2: strategy; and S3: strategy) was compared. Additionally, DCE-MRI images of 48 breast lesions were selected to verify the robustness of the model. Ten images were obtained for each lesion. The classification was considered correct if more than 5 images were correctly classified. The metrics for model performance evaluation included accuracy (Ac) in the training and testing sets, precision (Pr), recall rate (Rc), f1 score (*f*1), and area under the receiver operating characteristic curve (AUROC) in the validation set. The Ac of the 4 fine-tuning strategies reached 100.00% in the training set. The S2 strategy exhibited good convergence in the testing set. The Ac of S2 was 98.01% in the testing set, which was higher than those of *S*0 (93.10%), *S*1 (90.45%), and *S*3 (93.90%). The average classification Pr, Rc, *f*1, and AUROC of S2 in the validation set were (89.00%, 80.00%, 0.81, and 0.79, respectively) higher than those of S0 (76.00%, 67.00%, 0.69, and 0.65, respectively), S1 (60.00%, 60.00%, 0.60, 0.66, and respectively), and S3 (77.00%, 73.00%, 0.74, 0.72, respectively). The degree of coincidence between S2 and the histopathological method for differentiating between benign and malignant breast lesions was high ($\kappa = 0.749$). The S2 strategy can improve the robustness of the DenseNet201 model in relatively small breast DCE-MRI datasets, and this is a reliable method to increase the Ac of discriminating benign from malignant breast lesions on DCE-MRI.

Abbreviations: Ac = accuracy, AI = artificial intelligence, AUROC = area under the receiver operating characteristic curve, CT = computed tomography, DCE-MRI = dynamic contrast enhanced magnetic resonance imaging, DTL = deep transfer learning, f1 = f1 score, Pr = precision, Rc = recall rate.

Keywords: breast lesions, deep transfer learning, fine-tuning, magnetic resonance imaging

1. Introduction

With the continuous development of artificial intelligence (AI), its application in medical imaging has become increasingly widespread, particularly in the extraction and analysis of medical image data. Due to this, medical imaging has transformed from a modality relying on subjective perception skills to an objective science. Current research on AI in breast imaging is mainly focused on detecting[1] and identifying benign and malignant lesions,[2,3] predicting molecular typing,[4] assessing risks, segmenting images,[5] formulating radiotherapy plans, and monitoring efficacy.[6,7] However, the impact of these AI technologies on the classification of benign and malignant breast lesions using dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is still limited, and this area needs to be explored further. Currently, no large-scale open-source datasets pertaining to DCE-MRI are available for such studies. DenseNet is a technological innovation that involves introducing shortcut connections to overcome the training problems of deeper networks.[8] In this model, all previous layers are concatenated to

MM and MZ contributed equally to this work.

The authors have no funding and conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

[a] *Department of Radiology, The Affiliated Changzhou No. 2 People's Hospital of Nanjing Medical University, Changzhou, China.*

** Correspondence: Guangyuan He, Department of Radiology, The Affiliated Changzhou No. 2 People's Hospital of Nanjing Medical University, No.68 Gehuzhong Rd, Changzhou 213164, Jiangsu Province, China (e-mail: zwhj123zwjn@163.com).*

How to cite this article: Meng M, Zhang M, Shen D, He G. Differentiation of breast lesions on dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) using deep transfer learning based on DenseNet201. Medicine 2022;101:45(e31214).

Received: 14 April 2022 / Received in final form: 7 September 2022 / Accepted: 19 September 2022

http://dx.doi.org/10.1097/MD.0000000000031214

form the input for each layer and to connect each layer to all the previous layers.[9] This design can alleviate the problem of gradient disappearance in deep neural networks.[10] However, another issue that must be addressed is the overfitting problem. Deep learning is a subfield of machine learning, and deep transfer learning (DTL) is the process of transferring knowledge from a task that was already learned to a new task (usually on a large dataset). According to the literature, DTL based on DenseNet has achieved good success in the classification of lung diseases on computed tomography (CT)[11–16] [such as the diagnosis of coronavirus disease (COVID-19) on chest CT images]; however, overfitting is often related to a small sample size.[11] The computing cost required to achieve reliable and state-of-the-art performance is high for vision-based models, and the datasets used must be robust.[17] In other words, deep supervised models are prone to over-fitting because they contain a large number of parameters, particularly in the absence of large training sets.[18]

In this study, different strategies for fine-tuning DenseNet201 were used to explore the identification efficiency of this model with respect to differentiating between benign and malignant breast lesions on DCE-MRI. The aim was to find a more precise DTL model for the classification and diagnosis of breast lesions.

2. Materials and methods

2.1. Database and patient population

This study was a retrospective analysis and was approved by the Second Hospital of Changzhou Affiliated to Nanjing Medical University of Chinese Medicine Ethics Review Committee. The requirement for obtaining informed consent from the patients was waived (Ethics Number: [2020]KY234-01). Data pertaining to a total of 310 patients with complete breast DCE-MRI and pathological data were collected between January 2017 and December 2020; 17 patients with bilateral lesions were included. The primary lesions were pathologically confirmed in all patients. Lesions were categorized into benign and malignant groups. The patient age, pathological type, and lesion diameter are presented in Table 1. There were significant differences in age and lesion diameter between the 2 groups (P = .029 and .000, respectively). The inclusion criteria were as follows; patients who did not receive any preoperative chemotherapy or chemoradiotherapy before the MRI; No puncture or surgical procedure was performed before the MRI examination.

2.2. DCE-MRI acquisition

DCE-MRI scans were performed on 2 3T MRI scanners using a dedicated breast coil, with the patient in a prone position. Gadolinium diethylenetriaminepentaacetic acid (0.1 mmol/kg, 2.50 mL/s) was administered via elbow vein injection. The detailed scan parameters are listed in Table 2.

2.3. Data preparation

The images were obtained at 6 phases (1 pre-contrast phase and 5 post-contrast phases), and a series of 12 to 54 images were selected for each lesion. To eliminate interference signals pertaining to other tissues (such as the aorta), the images were cropped (using Photoshop) and the image fragments containing the breast tissue (effort was made to ensure the inclusion of the axillary) were retained. In total, 8400 breast DCE-MRI images were collected (on average, 27 images per patient), including 4260 images of benign lesions (benign group) and 4140 images of malignant lesions (malignant group). Each group was randomly divided into a training set (benign group, 3840 images; malignant group, 3726 images) and a testing set (benign group, 420 images; malignant group, 414 images) using self-programmed instructions in a 9:1 ratio. A further 48 unilateral lesions (25 benign and 23 malignant) were included in a validation set that was used to estimate the robustness of the DTL model. Ten DCE-MRI images were selected for each lesion, and the classification was categorized as correct if > 5 images were correctly classified.

2.4. Computer configuration

The configuration of the computer that was used for the analysis was as following: 64-bit versions of the Windows operating system (Windows 10), Intel Core i7-10700F processor, NVIDIA GeForce GTX 2060 GPU, and 6 GB. Python programming language (Python Software Foundation, version 3.6, https://www.python.org/) was used for analysis, and Keras (version 2.2.4, https://github.com/keras-team/keras) with TensorFlow (version 2.0, www.tensorflow.org) was used in the backend. All other

Table 1

Clinical data of the patients in the training and testing sets.

Pathological diagnosis	Cases	Percentage (%)	Age (yr)	Lesion diameter (mm)
Malignant lesions			48.2 ± 11.4	24.00 ± 11.09
Invasive ductal carcinoma	124	80.52		
Intraductal carcinoma	19	12.34		
Invasive lobular carcinoma	4	2.60		
Mucinous carcinoma	4	2.60		
Lymphoma	1	0.65		
Papillary carcinoma	2	1.30		
Total	154	100.00		
Benign lesions			45.0 ± 10.5	32.89 ± 16.45
Cyst	17	9.83		
Adenosis	26	15.03		
Fibroadenoma	111	64.16		
Chronic inflammation	4	2.31		
Intraductal papilloma	13	7.51		
Lobular tumor	2	1.16		
Total	173	100.00		
F^*			4.807	32.068
P^*			.029	.000

* P < .05 was considered to be statistically significant.

Table 2

Dynamic contrast-enhanced magnetic resonance imaging acquisition parameters.

Parameter	Philips Achieva	GE Healthcare
Field strength	3.0T	3.0T
No. of coil channels	8	8
Acquisition plane	Axial	Axial
Pulse sequence	3D gradient echo (Thrive)	Enhanced fastgradient echo 3D
Repetition time (ms)	5.5	9.6
Echo time (ms)	2.7	2.1
Flip angle	10°	10°
No. of post-contrast images	5	5
Fat suppression	Yes	Yes
Scan time	9 min 30 s	8 min 20 s

3D = three dimensional, ms = millisecond, No = number, s = second.

processes were turned off while running the analysis program. The training and testing processes of the DTL model were recorded by a computer.

2.5. DTL diagram

First, the images were randomly shuffled using a set program. Data augmentation was performed before model training; the relevant parameters and values are listed in Table 3.

2.6. Densely connected convolutional networks

DenseNet consists of a dense block, transition layer, and bottleneck layer.[19] The DenseNet block shown in Figure 1 revises the sequential concatenation of all the feature maps $[x_0, x_1, \ldots, x_{l-1}]$ in the model, instead of connecting the output feature maps from all previous layers[20]; it can be expressed as formula 1:

$$\text{DenseNet}: \; x_l = N_l\left(concat\left[x_0, x_1, \ldots x_{l-1}\right]\right) \quad (1)$$

where l is the layer index and N is the nonlinear operation. x_l represents the feature of the l_{th} layer. DenseNet confers several advantages such as the ability to reuse features, reduce exploding features; this model is also associated with fewer gradient disappearance problems.[8]

In this study, we chose DenseNet201 as the backbone for developing a breast lesion diagnosis system because it provides the best performance based on the ImageNet classification task. We used binary cross entropy as our loss function and a stochastic gradient descent optimizer to minimize the loss. The loss function of the binary cross-entropy can be expressed as follows:

$$loss = -\sum_{i=1}^{n} \hat{y}_i \log y_i + (1 - \hat{y}_i) \log (1 - \hat{y}_i) \quad (2)$$

Table 3

Parameters pertaining to data augmentation.

Parameter	Value
Rotation range	60
Shear range	0.2
Zoom range	0.2
Horizontal flip	True
Fill mode	Nearest

where y is the classification label (0 or 1), the term y_i is the predictive probability of the model output, and n is the number of images per batch.

Binary cross-entropy is a loss function and is a measure of the accuracy with which a prediction model can predict the expected outcome. Expressed in simpler terms, in the case of $y = 1$, if y_i is close to 1, then the loss value will be close to 0. Conversely, if y_i is close to 0 at this time, then the loss value will be remarkably large, which is very much in line with the nature of the log (complex) function.

The DTL model architecture is divided into 3 parts: feature extraction, data training and testing, and model validation. The main hyperparameters were set as follows: binary cross-entropy was used as the loss function, optimization was based on the Adam optimizer, learning rate was set to 0.0001, dropout (neuron random loss function) was set to 0.5, the number of Epochs was set to 40, ReLu was used as the activation function, and the Sigmoid function was used as the classification function (as shown in formulas 3 and 4, respectively).

$$Relu\,(x) = f\,(x) = \begin{cases} \max\,(0, x)\,, x \geq 0 \\ 0, \qquad\quad x < 0 \end{cases} \quad (3)$$

$$\text{Sigmoid}\,(x) = f\,(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

The DTL model based on DenseNet201 is shown in Figure 2.

2.7. Fine-tuning strategies

There has been widespread use of the pre-trained DenseNet201 model in medical image analysis because of its high discrimination power derived from millions of natural (non-medical) images. However, this model is very time-consuming, and high-performance computers are required for model training. It has been reported in the literature that the performance of DenseNet can be improved through fine-tuning.[21,22] In this study, we sought to improve the performance of the DenseNet201 model by devising 4 fine-tuning strategies, as follows: Strategy 0 (S0), Strategy 1 (S1), Strategy 2 (S2), and Strategy 3 (S3). The parameters of the neural network were activated and used in the model training process, whereas the parameters of the layers that were kept frozen were not involved in training the model (Fig. 3). The feature extraction network was not altered in the freezing layers, but the number of parameters that were required to be trained was reduced; this feature can save training time and space resources.

2.8. Statistical analysis

Statistical analysis was performed using SPSS 23.0 statistical software (IBM, Chicago). The age of the patients and the diameters of the lesions are represented as mean ± standard deviation ($\bar{x}$± s). One-way analysis of variance (ANOVA) was used to analyze the variance between the 2 groups. The *kappa* test was used to determine the degree of agreement between methods. Statistical significance was set at $P < .05$.

2.9. Performance evaluation of the networks

Four performance indices were calculated and used to compared the performance of the models. These included accuracy (Ac), precision (Pr), recall rate (Rc), f_1 score (f_1), and the area under the receiver operating characteristic curve (AUROC). The calculations are presented in the following equations:

$$\text{Ac} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

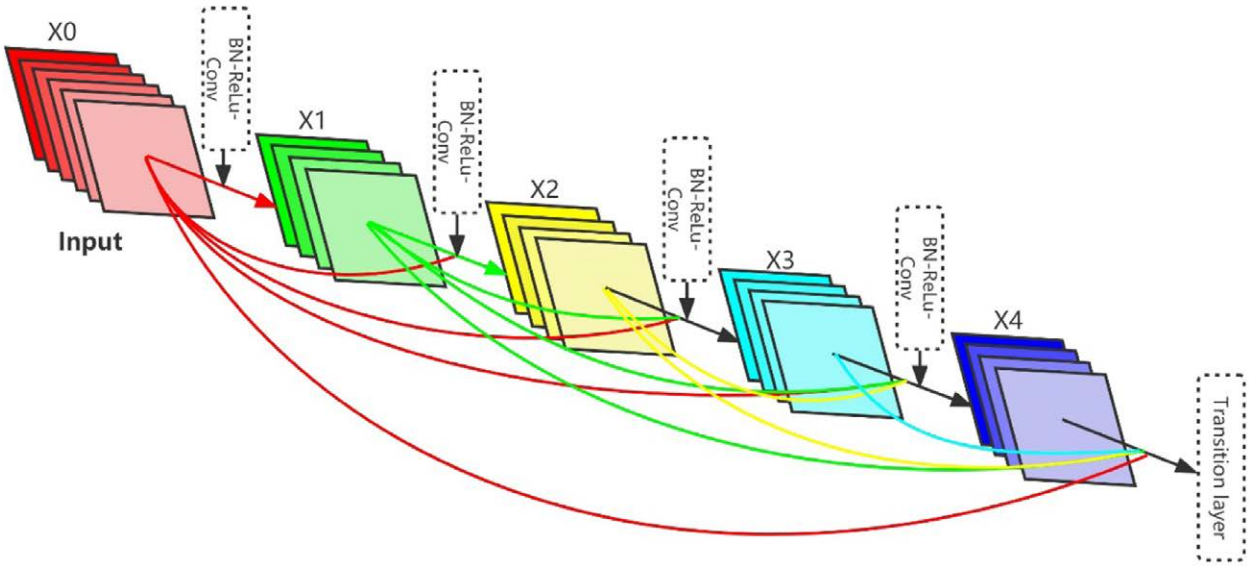

Figure 1. A 5-layer dense block with a growth rate of k = 4. Each layer takes all the preceding feature-maps as input.

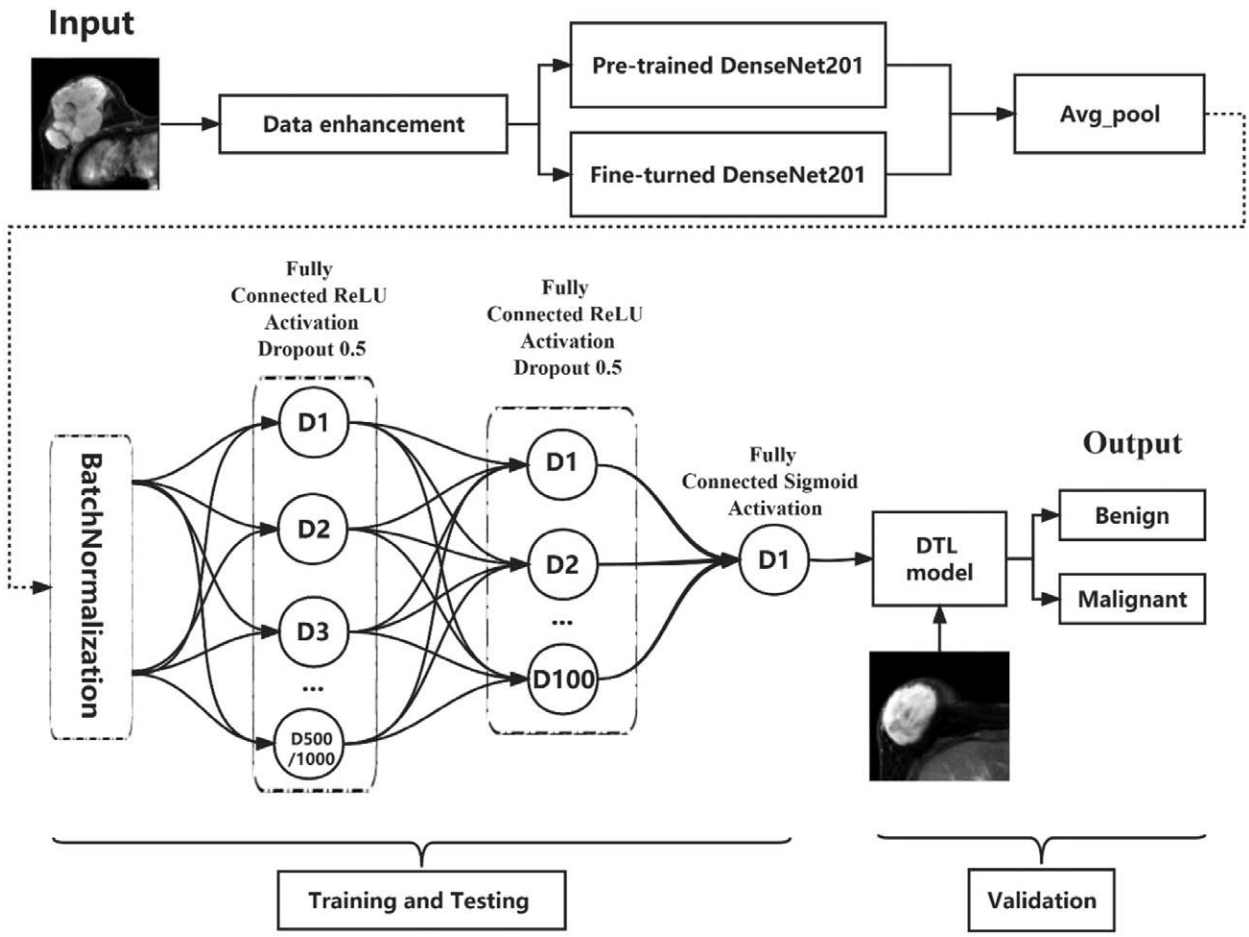

Figure 2. Illustration of the deep transfer learning (DTL) architecture. The input images were supplied in the BMP format. The process is divided into 3 parts: image neural network feature extraction, model training and testing, and model validation; the results are output after this process. DTL = deep transfer learning.

$$\mathrm{Pr} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (6)$$

$$\mathrm{Rc} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (7)$$

$$f_1 = \frac{2 \times \mathrm{Ac} \times \mathrm{Rc}}{\mathrm{Ac} + \mathrm{Rc}} \quad (8)$$

In our study, the positive and negative cases were assigned to the malignant and benign groups, respectively. TP, TN, FP, and

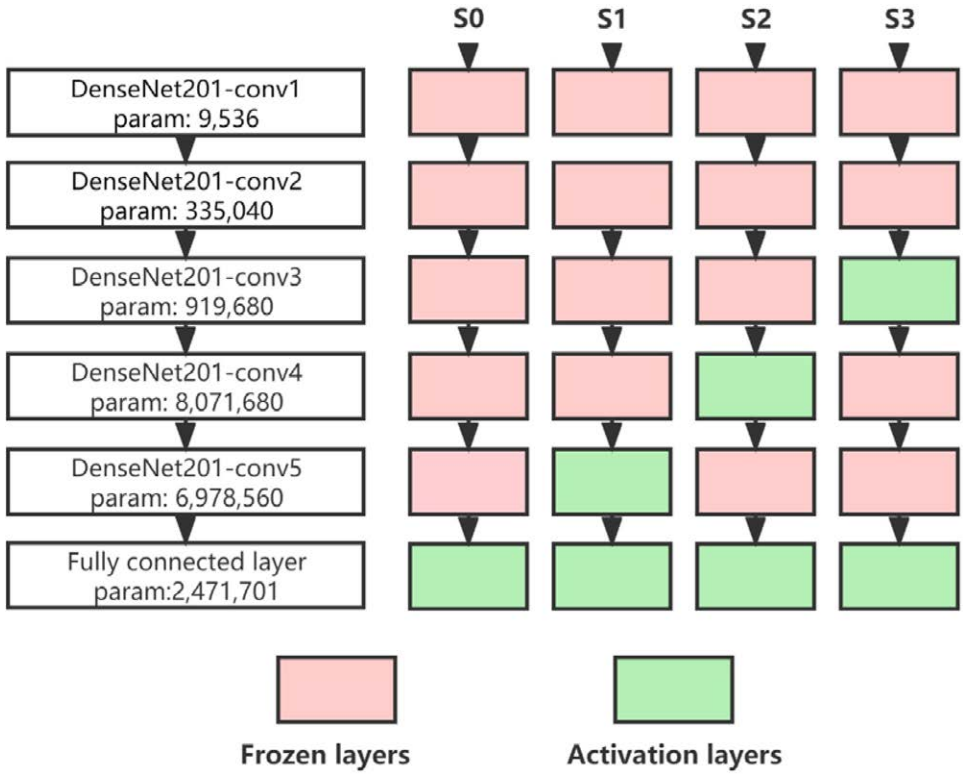

Figure 3. Schematic diagram of the 4 pre-set fine-tuning strategies param: the trainable parameters of the activation layers.

FN represent the true positives, true negatives, false positives, and false negatives, respectively.

Recall (also known as sensitivity) refers to the ratio of the correctly predicted positive observations to all the positive observations. f_1 is a measure of classification accuracy in statistics, it provides a reliable measure of the relationship between precision and recall. Therefore, a high score signifies an equilibrium between precision and recall.

3. Results

3.1. The results for the training and testing sets

The analysis results showed that the accuracy of the training set reached 100.00% for all fine-tuning strategies after 17 epochs, while S2 achieved the best test accuracy at 98.01%. With increasing epochs in the training set, the train loss value decreased for all the fine-tuning strategies. During the testing process, the test loss tended to increase for all the fine-tuning strategies, except for S2 (Fig. 4). This implies that among all the fine-tuned models, only S2 achieved convergence. These data suggest that the S2 model was a better fit than the other strategies. The time consumed for the training using S3 was 12.93% higher than that using S2; all the saved models were identical in size at 84.30 MB (Fig. 5).

3.2. Cross validation

By comparing the results of the 4 fine-tuned models in the training and testing sets, we determined that the S2 model was the best candidate model. Next, the S2 model was evaluated through 10-fold cross-validation (Fig. 6) using the dataset. The results are summarized in Table 4.

3.3. Visualization of the activated breast MRI images of the DTL model

A class activation map was composited by combining the input image and heat map (Fig. 7). Such a map can help in identifying the parts of the image on which the model was focusing while making the final prediction and hence can provide insights into the working of the model. The heat map is a coarse localization map that highlights the import regions for the classification target. Such an analysis can further help in hyperparameter tuning and aids in gaining an understanding of the reason underlying the failure of a model.

3.4. Validation results of the fine-tuning strategies

The classification report of the 4 strategies in the benign (250 images) and malignant (230 images) groups can be summarized as follows: overall Pr, Rc, $f1$, and AUROC of the *S*2 model in the validation set were 89.00%, 80.00%, 0.81, and 0.79, respectively, and were higher than those of the *S*0 (76.00%, 67.00%, 0.69, and 0.65), *S*1(60.00%, 60.00%, 0.60, and 0.66, respectively), and *S*3 (77.00%, 73.00%, 0.74, and 0.72) models. The accuracy for discriminating between benign and malignant breast lesions was as follows: *S*0, 60.41% (29/48); S1, 54.17% (26/48); *S*2, 75.00% (36/48), and S3, 70.83% (34/48). The degree of coincidence between the S2 model and the histopathology method for differentiating between benign and malignant breast lesions was high (κ = 0.749). Further details are provided in Tables 5 and 6. The AUROC of the *S*0, *S*1, *S*2, and *S*3 strategies in the validation set were 0.65, 0.66, 0.79, and 0.72, respectively (Fig. 8).

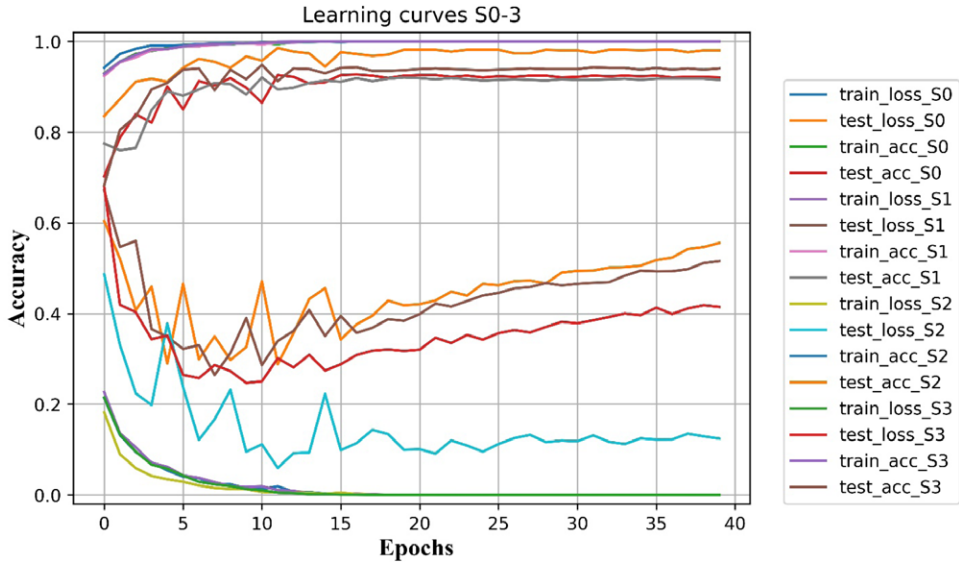

Figure 4. Learning curves of the fine-tuning strategies. As is evident from the figure, the accuracy of S2 was higher than that of the other strategies in the testing set, and this strategy was associated with a relatively lower loss value. train_loss: loss of the training set; train_acc: accuracy of the training set; test_loss: loss of the testing set; test_acc: accuracy of the testing set; *S0-S3*: strategy 0-3.

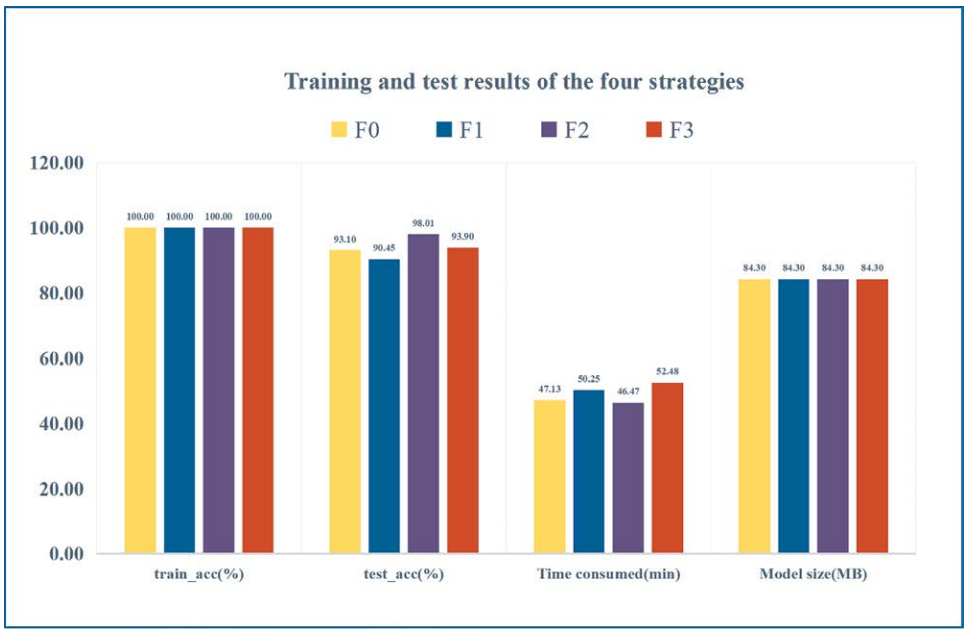

Figure 5. Comparison of the training results of the fine-tuning strategies train_acc: accuracy of the training set; val_acc: accuracy of the testing set. *S0-S3*: strategy 0-3.

4. Discussion

AI has shown advantages in diagnosing breast lesions, and has been used for the detection of calcification on mammography and classification of breast masses. However, there is significant intra-and inter-class heterogeneity owing to the diversity of imaging modes and clinicopathological characteristics.[23] Choosing a proper model to start a specific learning task for breast DCE-MRI remains a challenge.

DenseNet201 is an intensive convolutional neural network that connects each layer to every other layer in a feed-forward fashion. Foreach layer, the feature maps of all the preceding layers are used as inputs, and their own feature maps are used as inputs for all the subsequent layers.[8] In a study by Jaiswal et al, DenseNet201 was used to identify COVID-19 on chest CT.[11] Their data revealed that the DenseNet201 model achieved accuracies of 99.82%, 96.25%, and 97.4% for the training, testing, and validation sets, respectively. The Pr was 0.9629 in the testing set, which was higher than that of the other models such as VGG16 (0.9574), Inception ResNet (0.9015), and Resnet152V2 (0.9212). Using a densely connected convolutional neural network truncated with partial layer freezing and feature fusion, Montalbo et al[17] developed a method for diagnosing COVID-19 from chest X-rays using partial layer freezing. Their study showed that the performance-to-parameter size ratio of this method demonstrates its effectiveness in training DenseNet with fewer parameters compared to traditional deep convolutional neural networks; the results obtained with this method were promising.

Figure 6. Ten-fold cross-validation of the *S*2 model. We split our dataset into 10 parts. Then, 1 part was selected for each test, and the remaining 9 parts were used for training.

Table 4

The results of the 10-fold cross-validation.

Folds	Ac1 (%)	Loss1 ($\times 10^{-5}$)	Ac2 (%)	Loss2 (10^{-2})
Fold1	100.00	2.12	98.01	11.16
Fold2	100.00	2.23	97.93	11.09
Fold3	100.00	2.31	97.88	12.44
Fold4	100.00	2.30	97.88	11.59
Fold5	100.00	2.12	98.03	11.16
Fold6	100.00	1.91	98.14	13.80
Fold7	100.00	2.04	98.08	11.61
Fold8	100.00	2.21	97.75	10.84
Fold9	100.00	2.18	97.92	11.70
Fold10	100.00	2.15	97.99	12.99

Ac1 = accuracy of the training set; loss1: loss value of the training set, Ac2: accuracy of the testing set; loss2: loss value of the testing set.

There have been no literature reports on the ability of DenseNet201 to differentiate between benign and malignant breast lesions on DCE-MRI. In this study, 4 fine-tuning strategies were applied to improve the accuracy of DenseNet201. The fine-tuned S2 model was evaluated using the 10-fold cross-validation method, and its performance was found to be stable. The average classification Ac, Rc, $f1$, and AUROC of S2 in the validation set were higher than those of the other strategies. Rc (also known as sensitivity) is the ratio of the correctly predicted positive observations to the total number of observations in a class. $f1$ is a measure of classification accuracy, with a maximum value of 1 and a minimum value of 0. It is a robust metric that calculates the relationship between Pr and Rc; hence, a high $f1$score indicates a balance between Pr and Rc.

Theoretically, Ac increases as the training parameters increase in the network for the same dataset and computer environment, and the same is true for the training time. Surprisingly, our results only validated the former theoretical contention. Our data show that *S*2 has the most parameters and achieved the highest Ac; however, the time consumed was 12.93% higher for *S*3 compared to that for *S*2. This may be because the principal function of the third convolution layer (conv3 layer) which was activated by *S*3 was feature extraction, and a longer time period was required than that for the 4th convolution layer activated by *S*2 (conv4 layer).

Two major factors influence the use of deep learning technology in the medical imaging field. First, a significant limitation is the lack of medical image datasets that are publicly available for training. Because of the emphasis on privacy and confidentiality, medical images are difficult to obtain from the internet, even with advanced web crawler techniques. The dataset

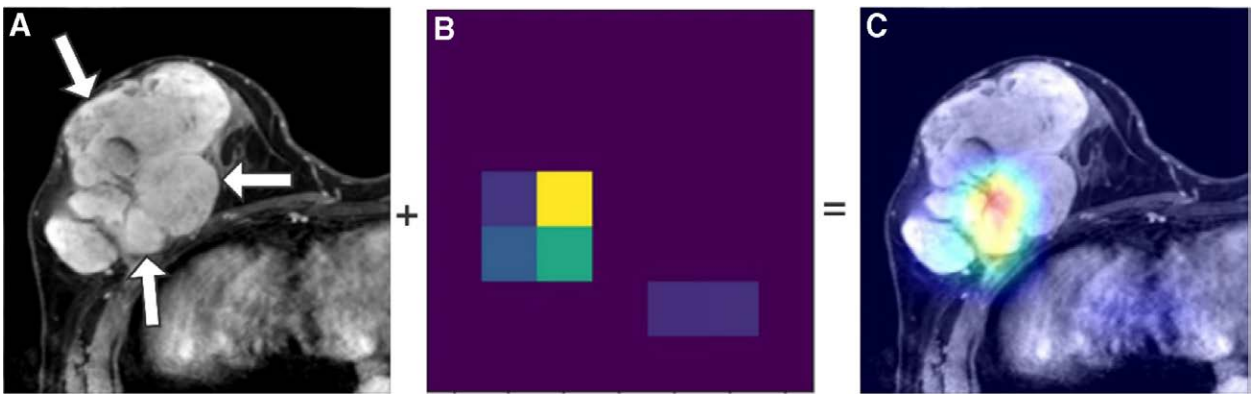

Figure 7. Class activation map (CAM) of a breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) image in DenseNet201. A: input DCE-MRI image, white arrow shows the benign breast lesion; B: heat map of the image. The yellow square is the most sensitive region of the input image found by the convolutional neural network; C: composite image obtained by fusing images A and B, which is easier to analyze visually. The lesion was pathologically confirmed as a fibroadenoma. CAM = class activation map, DCE-MRI = dynamic contrast enhanced magnetic resonance imaging.

Table 5

Classification report of the 4 strategies in the validation set.

	Support images				*S0*			*S1*			*S2*			*S3*		
Group	***S0***	***S1***	***S2***	***S3***	**Pr**	**Rc**	***f1***	**Pr**	**Rc**	***f1***	**Pr**	**Rc**	***f1***	**Pr**	**Rc**	***f1***
Group 1	352	256	352	320	88.00	64.00	0.74	62.00	62.00	0.62	100.00	73.00	0.84	88.00	70.00	0.78
Group 2	128	224	128	160	43.00	75.00	0.55	57.00	57.00	0.57	57.00	100.00	0.73	57.00	80.00	0.67
Avg/total	480	480	480	480	76.00	67.00	0.69	60.00	60.00	0.60	89.00	80.00	0.81	77.00	73.00	0.74

1: = f1 score, avg = average, Group 2 = malignant group, Group1 = benign group, Pr = precision, Rc = recall rate.

Table 6

Comparison of the different fine-tuning strategies and histopathological diagnosis in the validation set.

	S0					*S1*					*S2*					*S3*				
His	**B**	**M**	**T**	**κ**	***P****	**B**	**M**	**T**	**κ**	***P****	**B**	**M**	**T**	**κ**	***P****	**B**	**M**	**T**	**κ**	***P****
B	20	5	25			16	9	25			23	2	25			20	5	25		
M	14	9	23			13	10	23			4	19	23			9	14	23		
T	34	14	48	0.194	.145	29	19	48	0.059	.683	27	21	48	0.749	.000	29	19	48	0.417	.003

B = benign, His = histopathological diagnosis, M = malignant, T = total.

*, *kappa* test $\kappa \geq 0.7$ indicated a strong correlation; $0.7 > \kappa \geq 0.4$ indicated a relatively strong correlation; $\kappa < 0.4$ indicated a poorer correlation.

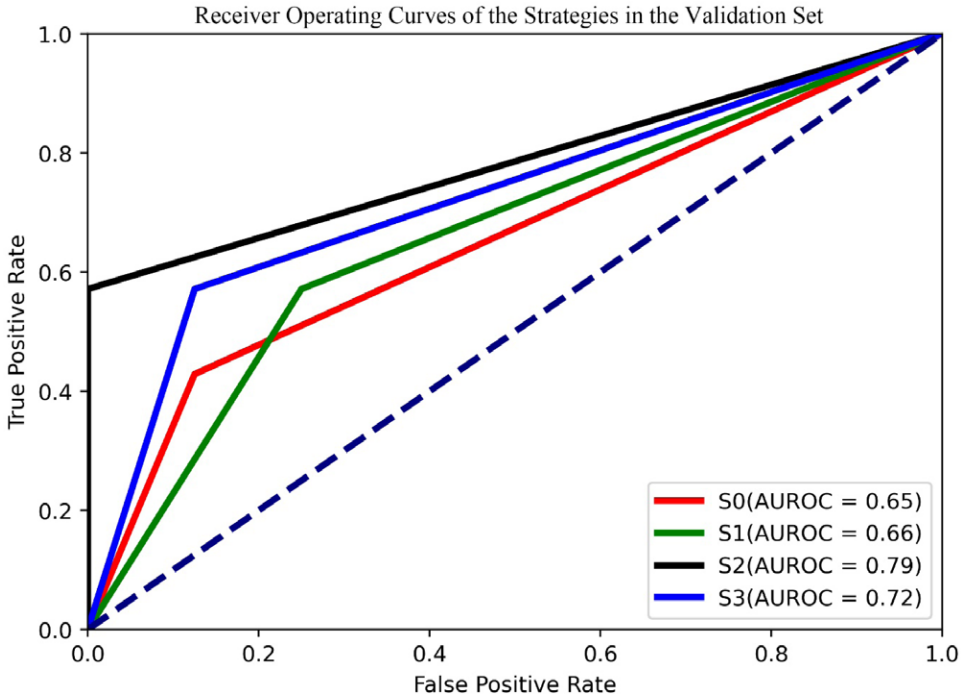

Figure 8. The receiver operator characteristic (ROC) curve and area under the receiver operator characteristic curve (AUROC) pertaining to the performance of the 4 strategies in the validation set. We can see that the AUROC of S2 (0.79) was higher than that of S0 (0.65), S1 (0.66), and S3 (0.72). ROC = receiver operator characteristic, AUROC = area under the receiver operator characteristic curve.

used in this study was from a single center and was relatively small. Thus, the robustness of the prediction model was poor. As our data shows, although 8400 images were included in the study, there were only 310 patients (27 images per patient). Multicenter studies should be conducted in the future to resolve this issue. Second, there is no widely applicable deep-learning algorithm. Depending on the dataset, different hyperparameters and optimizers may be required for the optimal performance of a deep learning algorithm. If a DTL model is proposed without fine-tuning, the results may not be ideal, as demonstrated by our results. In this study, we designed 4 fine-tuning strategies to improve the accuracy of DenseNet201. We found that the *S2* strategy performed better than the other strategies in differentiating benign lesions from malignant lesions on breast DCE-MRI. This capability will be of value and will be of interest as a topic for future studies.

However, this study has several limitations. First, the sample size was relatively small due to the small number of images included in our study, particularly there was a lack of rare breast lesions for training; this dataset may not sufficiently represent the full class of breast lesions, significantly reducing the reliability of the prediction results of the DTL models. Therefore, multicenter large-sample studies are required. Second, routine breast DCE-MRI diagnostics require a combination of history findings, breast ultrasound, or mammography, while our study was based on simple breast DCE-MRI images only. Third, the input data in

our study were 2D cropped images; it remains to be investigated whether 3D imagery can increase the performance of the DTL model. There has been considerable debate over 2D and 3D representation learning for 3D medical images. 2D approaches could benefit from large-scale 2D pretraining; however, they are generally weak in capturing large 3D contexts. 3D approaches are natively strong in 3D contexts; however, few publicly available 3D medical datasets are large and sufficiently diverse for universal 3D pretraining.[24] Finally, there are other classification convolutional neural networks that are suitable for the analysis of breast DCE-MRI images such as VGG19 and MobileNetV2. Thus, a DTL model that is more accurate and robust for the classification of breast DCE-MRI images is needed, which is the goal of our further research.

5. Conclusions

Our study further demonstrates that the performance of DenseNet201 can be improved through fine-tuning in transfer learning. We identified an optimal fine-tuning strategy (*S2*) to improve the robustness of the DenseNet201 model in analyzing relatively small breast DCE-MRI datasets. The study findings have important implications for future research, and thus it is necessary to conduct a multicenter study with a large sample size in the near future.

Acknowledgments

The authors wish to thank Ge Shiquan and Lu Jing for their technical assistance in operating the Python programming code and Li Li for the help with data collection in this study.

Author contributions

Conceptualization: Mingzhu Meng.
Data curation: Mingzhu Meng, Dong Shen, Guangyuan He.
Formal analysis: Mingzhu Meng, Ming Zhang.
Methodology: Mingzhu Meng, Ming Zhang.
Visualization: Dong Shen.
Project administration: Mingzhu Meng.
Resources: Mingzhu Meng, Dong Shen.
Writing – original draft: Mingzhu Meng, Ming Zhang.
Writing – review & editing: Mingzhu Meng, Ming Zhang.

References

[1] Mohiyuddin A, Basharat A, Ghani U, et al. Breast tumor detection and classification in mammogram images using modified YOLOv5 network. Comput Math Methods Med. 2022;2022:1359019.
[2] Niu J, Li H, Zhang C, et al. Multi-scale attention-based convolutional neural network for classification of breast masses in mammograms. Med Phys. 2021;48:3878–92.
[3] Wang Y, Choi EJ, Choi Y, et al. Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. Ultrasound Med Biol. 2020;46:1119–32.
[4] Reig B, Heacock L, Geras KJ, et al. Machine learning in breast MRI. J Magn Reson Imaging. 2020;52:998–1018.
[5] Zhang Y, Chan S, Chen JH, et al. Development of U-net breast density segmentation method for fat-sat MR images using transfer learning based on non-fat-sat model. J Digit Imaging. 2021;34:877–87.
[6] Sutton EJ, Onishi N, Fehr DA, et al. A machine learning model that classifies breast cancer pathologic complete response on MRI post-neoadjuvant chemotherapy. Breast Cancer Res. 2020;22:57–68.
[7] Zerouaoui H, Idri A. Reviewing machine learning and image processing based decision-making systems for breast cancer imaging. J Med Syst. 2021;45:8.
[8] Huang G, Liu Z, Maaten Lvd, et al. Densely connected convolutional networks. 2017 IEEE conference on computer vision and pattern recognition (CVPR). 2017;2261–69.
[9] Xu Z, Guo X, Zhu A, et al. Using deep convolutional neural networks for image-based diagnosis of nutrient deficiencies in rice. Comput Intell Neurosci. 2020;2020:7307252.
[10] Zhang Z, Liang X, Dong X, et al. A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. IEEE Trans Med Imaging. 2018;37:1407–17.
[11] Jaiswal A, Gianchandani N, Singh D, et al. Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. J Biomol Struct Dyn. 2020;39:5682–9.
[12] Yang S, Jiang L, Cao Z, et al. Deep learning for detecting corona virus disease 2019 (COVID-19) on high-resolution computed tomography: a pilot study. Ann Transl Med. 2020;8:450–57.
[13] Zhang Q, Chen Z, Liu G, et al. Artificial intelligence clinicians can use chest computed tomography technology to automatically diagnose coronavirus disease 2019 (COVID-19) pneumonia and enhance low-quality images. Infect Drug Resist. 2021;14:671–87.
[14] Wang S, Dong L, Wang X, et al. Classification of pathological types of lung cancer from CT images by deep residual neural networks with transfer learning strategy. Open Med (Wars). 2020;15:190–7.
[15] Zhang G, Yang Z, Gong L, et al. Classification of benign and malignant lung nodules from CT images based on hybrid features. Phys Med Biol. 2019;64:125011.
[16] Jangam E, Annavarapu CSR. A stacked ensemble for the detection of COVID-19 with high recall and accuracy. Comput Biol Med. 2021;135:104608.
[17] Montalbo FJP. Truncating a densely connected convolutional neural network with partial layer freezing and feature fusion for diagnosing COVID-19 from chest X-rays. MethodsX. 2021;8:101408.
[18] Tajbakhsh N, Jeyaseelan L, Li Q, et al. Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. Med Image Anal. 2020;63:101693.
[19] Li Z, Lin Y, Elofsson A, et al. Protein contact map prediction based on ResNet and DenseNet. Biomed Res Int. 2020;2020:7584968.
[20] Zhang YD, Satapathy SC, Zhang X, et al. COVID-19 diagnosis via DenseNet and optimization of transfer learning setting. Cognit Comput. Preprint posted online January 18, 2021. doi:10.1007/s12559-020-09776-8.
[21] Riasatian A, Babaie M, Maleki D, et al. Fine-tuning and training of DenseNet for histopathology image representation using TCGA diagnostic slides. Med Image Anal. 2021;70:102032.
[22] Tan T, Li Z, Liu H, et al. Optimize transfer learning for lung diseases in bronchoscopy using a new concept: sequential fine-tuning. IEEE J Transl Eng Health Med. 2018;6:1800808.
[23] Zhang J, Xie Y, Wu Q, et al. Medical image classification using synergic deep learning. Med Image Anal. 2019;54:10–9.
[24] Yang J, Huang X, He Y, et al. Reinventing 2D convolutions for 3D images. IEEE J Biomed Health Inform. 2021;25:3009–18.