electronics

MDPI

Article

Diagnosis of Breast Cancer with Strongly Supervised Deep Learning Neural Network

Haitian Gui [1], Tao Su [2], Zhiyong Pang [2,*], Han Jiao [2], Lang Xiong [3], Xinhua Jiang [3,*], Li Li [3] and Zixin Wang [2,*]

[1] School of Biomedical Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, China
[2] School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China
[3] Department of Medical Imaging, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, China
* Correspondence: stspzy@mail.sysu.edu.cn (Z.P.); jiangxh@sysucc.org.cn (X.J.); wangzix@mail.sysu.edu.cn (Z.W.); Tel.: +86-13925121769 (Z.W.)

Abstract: The strongly supervised deep convolutional neural network (DCNN) has better performance in assessing breast cancer (BC) because of the more accurate features from the slice-level precise labeling compared with the image-level labeling weakly supervised DCNN. However, manual slice-level precise labeling is time consuming and expensive. In addition, the slice-level diagnosis adopted in the DCNN system is incomplete and defective because of the lack of other slices' information. In this paper, we studied the impact of the region of interest (ROI) and lesion-level multi-slice diagnosis in the DCNN auxiliary diagnosis system. Firstly, we proposed an improved region-growing algorithm to generate slice-level precise ROI. Secondly, we adopted the average weighting method as the lesion-level diagnosis criteria after exploring four different weighting methods. Finally, we proposed our complete system, which combined the densely connected convolutional network (DenseNet) with the slice-level ROI and the average weighting lesion-level diagnosis after evaluating the performance of five DCNNs. The proposed system achieved an AUC of 0.958, an accuracy of 92.5%, a sensitivity of 95.0%, and a specificity of 90.0%. The experimental results showed that our proposed system had a better performance in BC diagnosis because of the more precise ROI and more complete information of multi-slices.

Keywords: deep learning; breast cancer diagnosis; lesion ROI; DCNN; multi-slice weighting

check for updates

Citation: Gui, H.; Su, T.; Pang, Z.; Jiao, H.; Xiong, L.; Jiang, X.; Li, L.; Wang, Z. Diagnosis of Breast Cancer with Strongly Supervised Deep Learning Neural Network. *Electronics* **2022**, *11*, 3003. https://doi.org/10.3390/electronics11193003

Academic Editor: Hyun Jae Baek

Received: 23 August 2022
Accepted: 20 September 2022
Published: 22 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

According to the Global Cancer data reported in 2020 by the World Health Organization, BC has already surpassed lung cancer and became the most commonly diagnosed cancer worldwide. One in every eight cancers diagnosed was BC in 2020 and BC was the fifth-leading cause of cancer mortality worldwide, with 685,000 deaths in 2020 [1]. The fact that BC can be cured if it is detected and diagnosed at an early stage indicates that screening is the key to reducing the mortality of this disease [2].

According to the female's individualized risk, a different diagnostic approach is customized to screen for BC, such as breast ultrasound, mammography, and breast magnetic resonance imaging (MRI) [3]. Among them, mammography is the mainstay of BC screening and has been proven to reduce mortality by 20–35% [4]. However, it is an imperfect examination. Its sensitivity to BC diagnosis would be reduced from the overall 85% [5] to 45–65% in women with dense breasts [6]. Breast MRI is extremely sensitive in the detection of invasive BC and is unlimited by breast tissue density. In particular, dynamic contrast-enhanced MRI (DCE-MRI), which can describe physiologic alterations as well as morphologic changes, has been used for several clinical situations, including high-risk screening, evaluation for an unknown primary carcinoma, preoperative evaluation in patients with known BC, evaluating response to neoadjuvant therapy, and suspected recurrence [7–11].

Conventional diagnoses made by radiologists are mainly based on subjective evaluation, which varies according to the experience of radiologists. Computer-aided methods contribute to reducing the deviation of the observers, which are reproducible and widely used in lesion detection and location, lesion segmentation, the definite of ROI, feature extraction and screening, and discrimination of malignant lesions.

Many previous studies have contributed to the diagnosis of malignant lesions automatically. A summary of related studies is presented in Table 1. Usually, traditional machine learning auxiliary diagnostic approaches adopt a classifier, such as the support vector machine (SVM) or random forest, combined with radiomics features to make the diagnosis [12–14]. Cai et al. [15] employed SVM combined with the apparent diffusion coefficient (ADC) and DCE-MRI features of the lesion extracted by the image algorithm to diagnose tumors. Dalmiş et al. [16] combined the ADC and DCE-MRI features with random forest for classifying lesions. However, features of the lesion input to the classifier were screened from the radiomics features, diffusion-weighted imaging (DWI) features, and DCE-MRI features. They were all generated by using algorithms on the segmented lesions. The complex process of features generated made it difficult to enlarge the dataset. In addition, it was possible to bring in subjective bias and lose some hidden inherent features of the lesion when screening the features. It was also possible to overfit when the dataset was small.

Table 1. Summary of related studies on BC diagnosis.

References	Dataset	Classification Method Used	ROI	Features	Diagnosis Metrics	Limitations
[12]	Ultrasound images: 184 benign 264 malignant	SVM	Whole Ultrasound	Radiomics features	Lesion-level (one slice)	Professional instruments required Features chosen with subjective bias Discarded the other slices
[13]	DCE-MRI: 42 benign 78 malignant	SVM	Whole DCE-MRI	Radiomics features	Lesion-level (one slice)	Small dataset Professional instruments required Features chosen with subjective bias Discarded the other slices
[14]	DCE-MRI: 40 benign 73 malignant	SVM	Whole DCE-MRI	Radiomics features	Lesion-level (one slice)	Small dataset Professional instruments required Discarded the other slices
[15]	DCE-MRI: 85 benign 149 malignant	SVM	Whole DCE-MRI	DWI Features + Radiomics features	Lesion-level (one slice)	Small dataset Professional instruments required Discarded the other slices
[16]	DCE-MRI: 149 benign 368 malignant	Random Forest	Whole DCE-MRI	Radiomics features + Feature maps extracted by DCNN	Lesion-level (one slice)	Rough ROI affect the performance (the highest value was 0.85) Discarded the other slices
[17]	DCE-MRI: 75 lesions	Unet++ and Faster RCNN	Segmented breast	Feature maps extracted by DCNN	Lesion-level (one slice)	Small dataset Without diagnosis
[18]	DCE-MRI: 438 patients, 3167 slices	ResNet101	Whole DCE-MRI	Feature maps extracted by DCNN	Slice-level	Rough ROI DCNN training and testing with slice-level data, can not stand for the performance in lesion-level diagnosis
[19]	DCE-MRI: 506 benign 1031 malignant	3D DenseNet	Segmented breast	Feature maps extracted by DCNN	3D lesion-level (one slice)	Rough ROI Discarded the other slices

Table 1. *Cont.*

References	Dataset	Classification Method Used	ROI	Features	Diagnosis Metrics	Limitations
[20]	DCE-MRI: 88 benign 139 malignant	ResNet50/Random Forest	Lesion level smallest bounding box	Feature maps extracted by DCNN	3D lesion-level (slice with max score)	Lesion-level ROI Lesion-level diagnosis with the max score slice (Discarded the other slices)

With the development of the DCNN and big data, research based on the DCNN in BC detection and diagnosis has increased rapidly. Jiao et al. [17] applied the Unet++ network to segment the breast and faster RCNN to detect the lesion. Liu et al. [18] located the lesion with a weakly supervised approach and classified the benign and malignant lesions at the slice level by adopting ResNet50 and achieved an AUC of 0.92. Zhou et al. [19] applied the weakly supervised 3d Resnet to make the diagnosis with seven phases of MRI. Zhou et al. [20] evaluated the diagnostic accuracy of lesions by applying ResNet101 to the ROI-based radiomics features, which considered the peri-tumor tissues.

There are several challenges for the DCNN approaches to diagnosing malignant lesions. Firstly, it is difficult to obtain a large enough dataset required for training the network with the situation of limited patient cases. Secondly, strongly supervised precisely labeling the lesions at the slice level by professional radiologists is costly, while the performance of the DCNN under the weakly supervised approach with picture-level labeling is poor. Thirdly, the variety of shapes and types of the lesion, as well as the complex background with the noise of human organs, make it difficult to detect, locate, and discriminate the malignant lesion. At last, since a lesion consists of several slices of various sizes and shapes, it is difficult to decide which slice represents the entire lesion accurately when adopting the slice-level diagnosis in the studies.

To retain the burrs of the lesion, which are the most important features in a malignant lesion, and minimize the impact of surrounding organs and tissues without the costly slice-level labeling at the same time, we proposed an improved region-growing algorithm to generate the slice-level precise ROI required by the strongly supervised DCNN and compared it with the other two kinds of ROI. To obtain results close to the real doctors' diagnosis, which involves comprehensive assessment of the information for all sections of the lesion, we evaluated four lesion-level diagnosis methods and proposed the average weighting of diagnostic scores of multi-slices as the final diagnosis criterion. In addition, we evaluated the performance of five different DCNNs, including ResNet50 [21], DenseNet [22], VGG16 [23], GoogLeNet [24], and AlexNet [25], in BC diagnosis under strongly supervised learning. Finally, a strongly supervised DCNN system combined with the slice-level precise ROI and the lesion-level diagnosis was proposed to diagnose BC.

2. Materials and Methods

2.1. Materials

The dataset included a total of 487 lesions from 166 patients with malignant lesions and 171 patients with benign lesions, of which 277 were lesions for training the network, 90 lesions for validation, and 120 lesions for testing. The total number of benign slices was 1866 and the total number of malignant slices was 1847. The malignant lesion included invasive cancer grade I, grade II, grade III, ductal carcinoma in situ, semi-invasive carcinoma in situ, and the other 9 kinds of lesions. The benign lesion included fibroadenoma, fibrocystic breast disease, intraductal papilloma, and the other 19 kinds of lesions. These lesions covered all the common types of breast tumors. All lesions were confirmed by the pathological biopsy or surgical pathology. The consecutive patients (mean age 49.1 years; range 18 to 78 years) enrolled in the study were collected in our cooperative hospital of Sun Yat-sen University Cancer Center (Guangzhou, China) and verified by an experienced doctor between January 2007 and December 2020. The details of the dataset split are in Table 2. The details of the

lesion types and numbers are in Table 3. This study was approved by the Ethics Committee of Sun Yat-sen University Cancer Center and informed patient consent was waived due to the nature of the retrospective analysis.

Table 2. Dataset split.

Dataset	Training	Validation	Testing	Subtotal
Benign lesions	162	45	60	267
Malign lesions	115	45	60	220
Benign slices	1278	259	329	1866
Malign slices	1053	341	453	1847

Table 3. Lesion types and numbers.

Lesion Type	Lesion Name	Number
Malignant		
	Invasive cancer	142
	Ductal carcinoma in situ	53
	Semi-invasive carcinoma in situ	8
	Premium Executive Internal Cancer	5
	Invasive lobular carcinoma	5
	low-grade intraductal carcinoma	1
	Intermediate-grade intraductal carcinoma	1
	Intraductal papillary carcinoma	1
	Invasive intraductal carcinoma	1
	Squamous cell carcinoma	1
	Eczematous breast cancer	1
	Calcified intraductal carcinoma	1
Benign		
	Fibroadenoma	160
	Fibrocystic breast disease	39
	Intraductal papilloma	33
	Adenopathy	7
	Papillary hyperplasia	4
	Hyperplastic nodule	3
	Lymphoma infiltration	3
	Lymphoma	3
	Chronic inflammation	2
	Fibroadenosis with lymphoid tissue	2
	Chronic granulomatous inflammation	1
	Complex sclerosing lesions	1
	Glandular hyperplasia with myoepithelial hyperplasia,	1
	Interstitial fibrous collagen hyperplasia	1
	Myoepithelial tumor	1
	Calcified nodules	1
	Dermal fibroma	1
	Compliant with BI-RADS Class II	1
	Compliant with BI-RADS Class III	1
	Fibroadenoma with tubular adenoma	1
	Epithelial hyperplasia	1

All MR images were acquired by MR scan equipment of MR1 to MR6 (MR1 was GE medical systems SIGNA EXCITE; MR2 was GE medical systems SIGNA HDx; MR3 was SIEMENS Trio Tim; MR5 was Philips medical systems achieve and MR6 as GE medical systems discovery MR750) at the 1.5T or 3.0 T field. Every DCE scan consisted of 1 unenhanced and 6 to 14 contrast-enhanced sequences.

Taking MR6 as an example, the patients were scanned in the prone position with a bilateral 8-channel phased-array breast-specific surface coil. The standard imaging of axial fast spin-echo (FSE) T1WI and axial and sagittal FSE T2WI A 3.0T scan was performed by using a superconductive magnetic system. The DWI was acquired on the axial plane. The DCE-MRI data were acquired using the VI-BRANT-FLEX technique in axial orientation or sagittal orientation after injecting 0.1 mmol/kg body weight contrast medium (gadopentetate dimeglumine; Magnevist, Bayer Schering Pharma, Berlin, Germany), with a hand venipuncture technique at a rate of 3 mL/s by using MRI-specific automatic power injector (Medrad Inc., Pittsburgh, PA, USA) after acquiring one set of unenhanced baseline images. Saline (10 mL at 3 mL/s) was then injected to wash the tube. Dynamic scanning was initiated by simultaneously pushing the high-pressure syringe button and the dynamic scan button. Eight postcontrast sets were acquired under the following scanning conditions: matrix, 320 × 320, slice thickness = 1.4 mm; repetition time = 3.896 ms, echo time = 1.674 ms, and flip angle = 5 degrees.

2.2. Methods

The proposed system consisted of three parts as shown in Figure 1. Part (a) realized slice-level ROI extraction from the input MRI with lesion-level annotations. Part (b) was slice-level training and classification based on DCNNs. Part (c) achieved lesion-level diagnosis by weighting multi slices of a lesion.

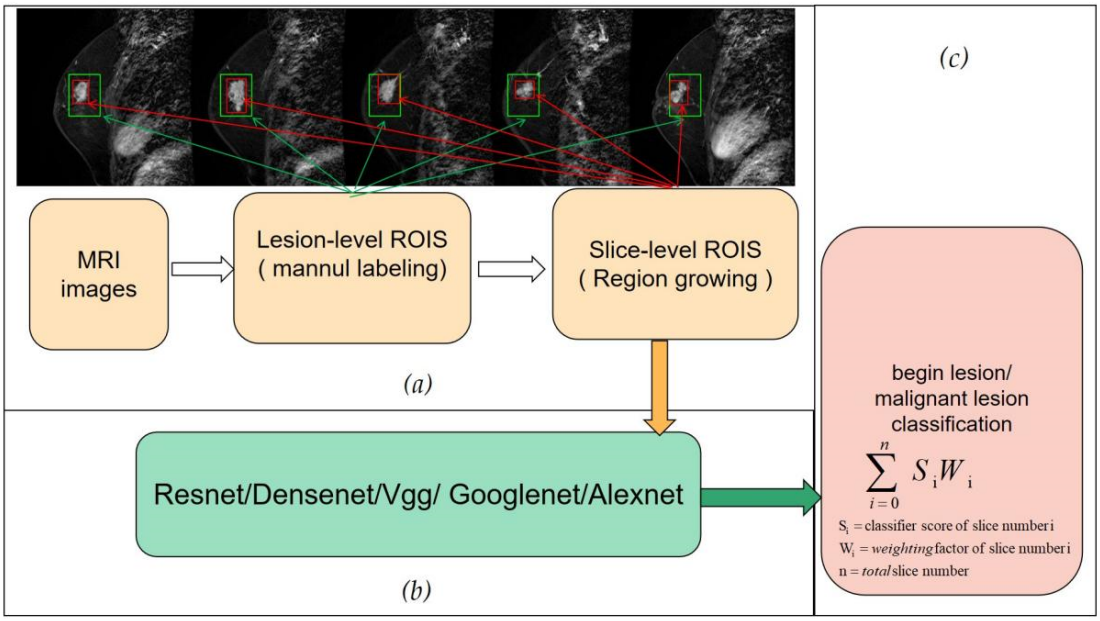


Figure 1. System architecture: (**a**) the extraction of the ROI as the input data of the DCNNs, the green box was the lesion-level ROI, and the red box was the slice-level ROI; (**b**) the DCNNs evaluated in this study including standard ResNet50, DenseNet, VGG16, GoogLeNet, and AlexNet for slice-level training and prediction; (**c**) lesion-level diagnosis layer.

2.2.1. Slice-Level ROI

To evaluate the role of different kinds of ROI in malignant lesion diagnostics, three kinds of ROI were studied in this paper. The radiologist manually labeled the smallest bounding box of the lesion ROI (SBBL-ROI) (green boxes in Figure 1a). It is a lesion-level ROI by finding the smallest bounding box of all slices of the lesion. One lesion was labeled once and all slices of the lesion had the same ROI. The region-growing algorithm [26] was used to segment all slices of the lesion as shown in Figure 2. First, set the original seed points by finding the coordinate of maximum pixel value after 3 × 3 block filtering (MaxPixel (3 × 3)) and the coordinate of maximum pixel value after 5 × 5 block filtering (MaxPixel (5 × 5), and set the thresh as the max_pixel (3 × 3) divided by ten. Second, the

values of the eight neighbor pixels of the seed point were compared with the threshold, the pixel was marked as one and pushed into the seed points stack if it was less than the threshold and greater than three-fifths of the MaxPixel (3 × 3). Otherwise, it was marked as zero. Took the first point out of the stack and recurred the above steps until the stack was empty. The segmented lesion ROI (SL-ROI) in each slice was generated by cropping all the pixels marked as 1 from the SBBL-ROI. The smallest bounding box of the slice ROI (SBBS-ROI) (red boxes in Figure 1a) was a rectangular region generated by finding the minimum x coordinate, maximum x coordinate, minimum y coordinate, and maximum y coordinate of the segmented lesion. The SBBS-ROI included the lesion and a few surrounding pixels. The ROI in each slice was different. The SBBL-ROI contained a slice of the lesion and peri-lesion, the SBBS-ROI contained the lesion and a little peri-lesion, and the SL-ROI contained the lesion only as shown in Figure 3.

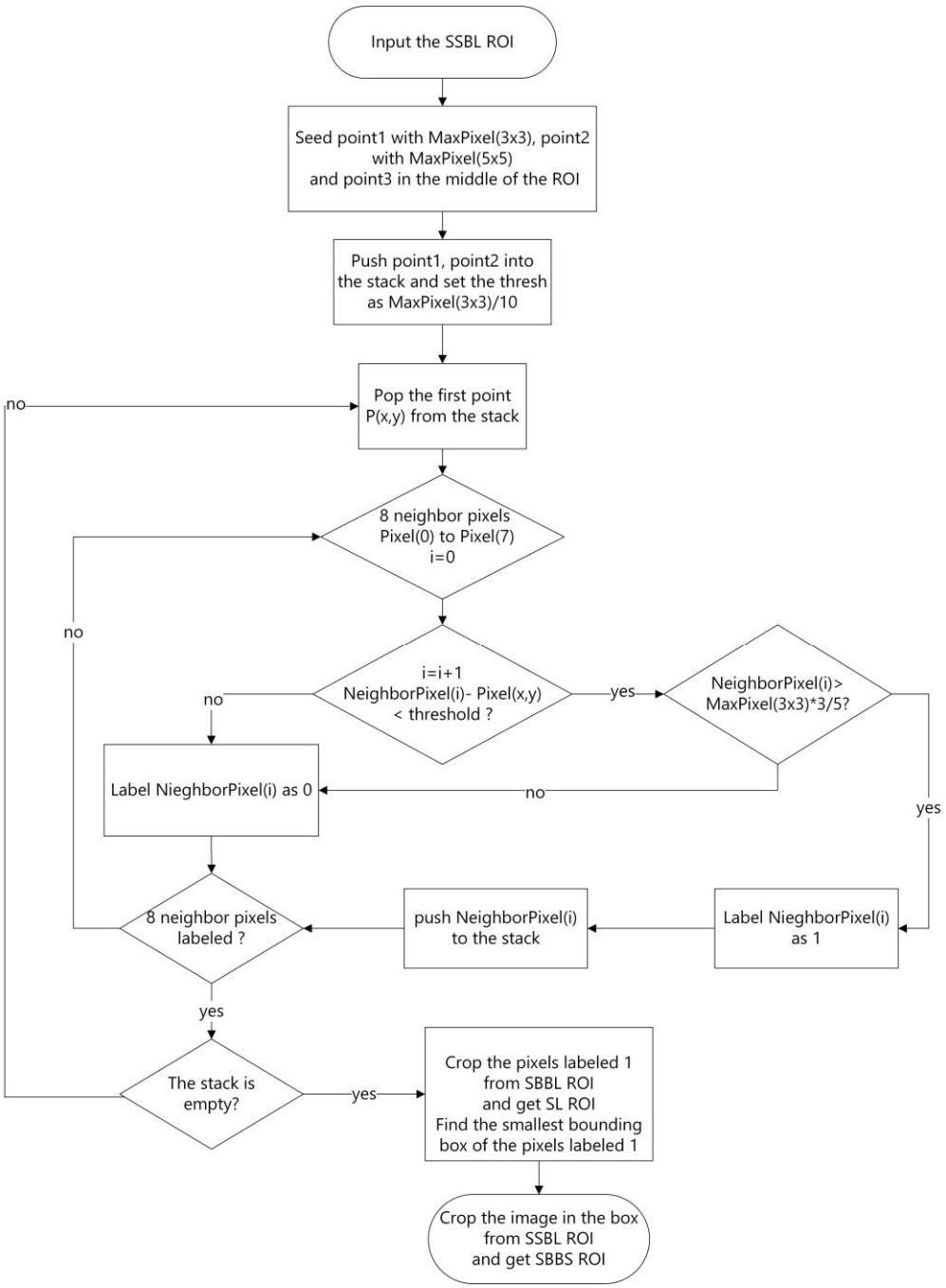


Figure 2. The flow of the SBBS-ROI extraction. * represents multiplication operation.

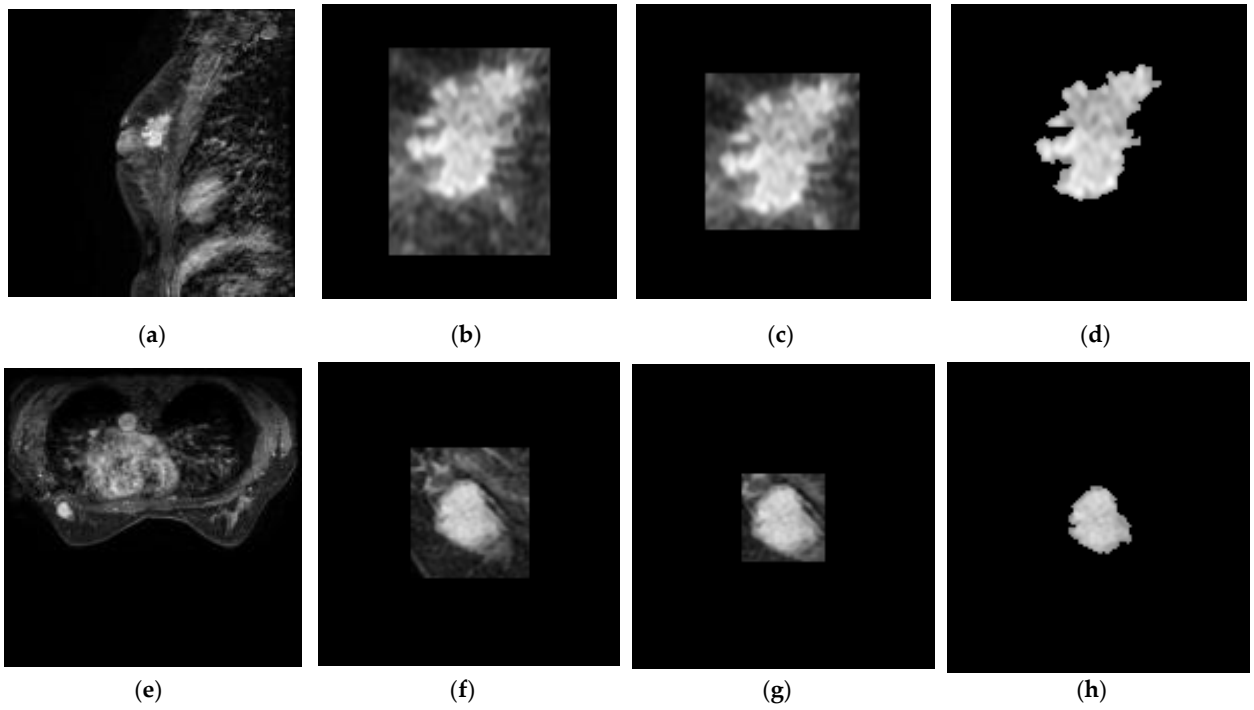


Figure 3. Original MRI and three kinds of ROI. (**a**,**e**) belong to the original. MRI images. (**b**,**f**) were SBBL-ROI. (**c**,**g**) were SBBS-ROI. (**d**,**h**) were SL-ROI. (**a**–**d**) belonged to a malignant lesion and (**e**–**h**) belonged to a benign lesion.

2.2.2. Slice-Level Training and Classification

DCNNs including classic ResNet50, DenseNet, GoogLeNet, VGG 16, and AlexNet were implemented on the training dataset with the Adam optimizer and cross-entropy loss function for slice-level training. The validation dataset was to evaluate the network and adjust the training parameters such as epoch, batch size, learning rate, and so on. Each slice of the malignant lesion was labeled to 1 and each slice of the benign lesion was labeled to 0. Training occurs at a batch size of 32, a learning rate of 0.0002, and epochs of 120. All the DCNN parameters such as the kernel size, stride, padding, depth, and so on, were kept as their default value. All the SBBS-ROIs were enlarged to 128 × 128 by padding with zero. The best model with the highest accuracy on the validation dataset was saved. The software environment was Python 3.8 with the open-source PyTorch library on a GPU-optimized workstation with a single NVIDIA GeForce RTX 3080Ti. After training the network, the best model was loaded to test all slices in the testing dataset. Each slice was standalone data. The model output two scores for each slice, representing the possibility of benign and malignant, respectively.

2.2.3. Lesion-Level Diagnosis

The lesion-level diagnosis was a comprehensive judgment. The benign score of the lesion was weighted by Equation (1). The malignant score of the lesion was weighted with the weighting parameters as Equation (2).

$$ScoreOfBenign = \sum_{i=1}^{N} SB_i \times W_i \quad (1)$$

$$ScoreOfMalign = \sum_{i=1}^{N} SM_i \times W_i \quad (2)$$

where SB_i and SM_i were the benign score and the malignant score of the ith slice, respectively, $ScoreOfBenign$ and $ScoreOfMalign$ were the benign score and the malignant score

of all slices of the lesion, respectively, N was the total number of the lesion's slices, and W_i was the weight parameter of the ith slice. It was possible to be the average weighted parameters $Wavg_i$ or the area ratio weighting parameters $Warea_i$, the perimeters ratio weighting parameters $Wperi_i$, the highest score weighting parameters $Whighestscore_i$ and so on.

In the average weighting method, all the lesion's slices shared the same weighting parameters and the sum was 1 as Equation (3).

$$Wavg_i = \frac{1}{N} \tag{3}$$

where $Wavg_i$ was the weighting parameters of the ith slice and N was the total number of slices of the lesion.

In the area ratio weighting method, the weighting parameter of each slice of the lesion was the area of the segmented lesion in the slice divided by the total area of the lesion in all slices as Equations (4) and (5).

$$Warea_i = \frac{Area_i}{SumArea} \tag{4}$$

$$SumArea = \sum_{i}^{N} Area_i \tag{5}$$

Where $Warea_i$ was the perimeter ratio weighting parameters of the ith slice, N was the total number of slices of the lesion, and $SumArea$ was the total area of the lesion in all slices.

In the perimeter ratio weighting method, the weighting parameter of each slice of the lesion was the perimeter of the segmented lesion in the slice divided by the total perimeter of the lesion in all slices as Equations (6) and (7).

$$Wperi_i = \frac{Peri_i}{SumPeri} \tag{6}$$

$$SumPeri = \sum_{i=1}^{N} Peri_i \tag{7}$$

where $Wperi_i$ was the area ratio weighting parameters of the ith slice, N was the total number of slices of the lesion, and $SumPeri$ was the total area of the lesion in all slices.

In the highest score weighting method, the weighting parameter of the highest score of the slice was set to 1 and the weighting parameters for the other slices were set to 0 as Equation (8).

$$Whightestscore_i = \begin{cases} 1 \text{ the ith slice was the highest score} \\ 0 \text{ the other slices} \end{cases} \tag{8}$$

At last, the larger of $ScoreOfBenign$ and $ScoreOfMalign$ were the final diagnosis result as Equation (9).

$$y = \begin{cases} benign \; ScoreOfBenign > ScoreOfMalign \\ \\ malign \; ScoreOfBenign \leq ScoreOfMalign \end{cases} \tag{9}$$

where y was the last output in the lesion-level diagnosis.

3. Results

3.1. Performance Metrics

We evaluated the performance of the system with Sensitivity *(Sen)*, Specificity (*Spec*), Accuracy (*Acc*), True-Positive Rate (*TPR*), False-Positive Rate (*FPR*), Receiver Operating

Characteristic (ROC), and the Area Under Curve (AUC). The *Sen*, *Spec*, *Acc*, *TPR*, and *FPR* were computed by Equations (10)–(14), respectively.

$$Sen(\%) = \frac{TP}{TP + FN} \tag{10}$$

$$Spec(\%) = \frac{TN}{TN + FP} \tag{11}$$

$$Acc(\%) = \frac{TN + TP}{TN + FN + TP + FP} \tag{12}$$

$$TPR(\%) = \frac{TP}{TP + FN} \tag{13}$$

$$FPR(\%) = \frac{FP}{FP + TN} \tag{14}$$

where the *TP*, *TN*, *FP*, and *FN* are described as follows:

True positive (*TP*): predicted positive in positive samples (malignant lesions).
True negative (*TN*): predicted negative in negative samples (benign lesions).
False positive (*FP*): predicted positive in negative samples.
False negative (*FN*): predicted negative in positive samples.

The Sen was the *TPR*, which represented the true-positive rate. The Spec represented the true-negative rate. The *ACC* represented the total accurate rate. The ROC was a curve drawn in the coordinate system with *TPR* as the *Y*-axis and *FPR* as the *X*-axis. It was the visual comparative plots of true-positive rates and false-positive rates. The AUC was the area under the ROC curve, representing the performance of the classification. The closer the value was to 1, the better the performance.

3.2. Slice-Level ROI

We experimented with five DCNNs, including ResNet50, DenseNet, VGG16, GoogLeNet, and AlexNet, for each type of ROI with the average weighting of all slices of the lesion on the testing subset and took the average AUC, average accuracy, average sensitivity, and average specificity of the five DCNNs as the measure of the performance of the three kinds of ROI. The DCNNs achieved an average AUC of 0.951, 0.813, and 0.882 for the SBBS-ROI, SBBL-ROI, and SL-ROI, respectively. By taking the average threshold of 0.484, 0.502, and 0.502 for malignant prediction, the DCNNs obtained average accuracy of 90.3%, 77.7%, and 81.7%, average sensitivity of 94.0%, 74.0%, and 78.7%, specificity of 86.7%, 81.3%, and 85.7% with SBBS-ROI, SBBL-ROI, and SL-ROI, respectively. The details of the AUC, accuracy, sensitivity, and specificity for each DCNN are shown in Table 4. The receiver operating characteristic (ROC) curves of each DCNN with different kinds of ROI are shown in Figure 4. The SBBS-ROI with a little peri-lesion was good for the DCNNs to extract more accurate features than the other two kinds of ROI from the experiments.

Table 4. Performance of BC classification of five DCNNs on testing subsets with different kinds of ROIs based on the average weighting.

DCNN	ROI Type	Threshold	Sensitivity	Specificity	Accuracy	AUC
DenseNet	SBBS	0.520	0.950	0.900	0.925	0.958
	SBBL	0.500	0.700	0.767	0.733	0.742
	SL	0.500	0.800	0.867	0.833	0.909
VGG16	SBBS	0.480	0.917	0.867	0.892	0.961
	SBBL	0.510	0.717	0.783	0.750	0.804
	SL	0.470	0.867	0.767	0.817	0.907
ResNet50	SBBS	0.460	0.950	0.850	0.900	0.934
	SBBL	0.520	0.700	0.867	0.783	0.800
	SL	0.490	0.767	0.900	0.833	0.883

Table 4. *Cont.*

DCNN	ROI Type	Threshold	Sensitivity	Specificity	Accuracy	AUC
GoogLeNet	SBBS	0.490	0.967	0.817	0.892	0.952
	SBBL	0.450	0.883	0.767	0.825	0.883
	SL	0.500	0.867	0.883	0.850	0.914
AlexNet	SBBS	0.470	0.917	0.900	0.909	0.948
	SBBL	0.530	0.700	0.883	0.792	0.837
	SL	0.550	0.633	0.867	0.750	0.797
Average	SBBS	0.484	0.940	0.867	0.903	0.951
	SBBL	0.502	0.740	0.813	0.777	0.813
	SL	0.502	0.787	0.857	0.817	0.882

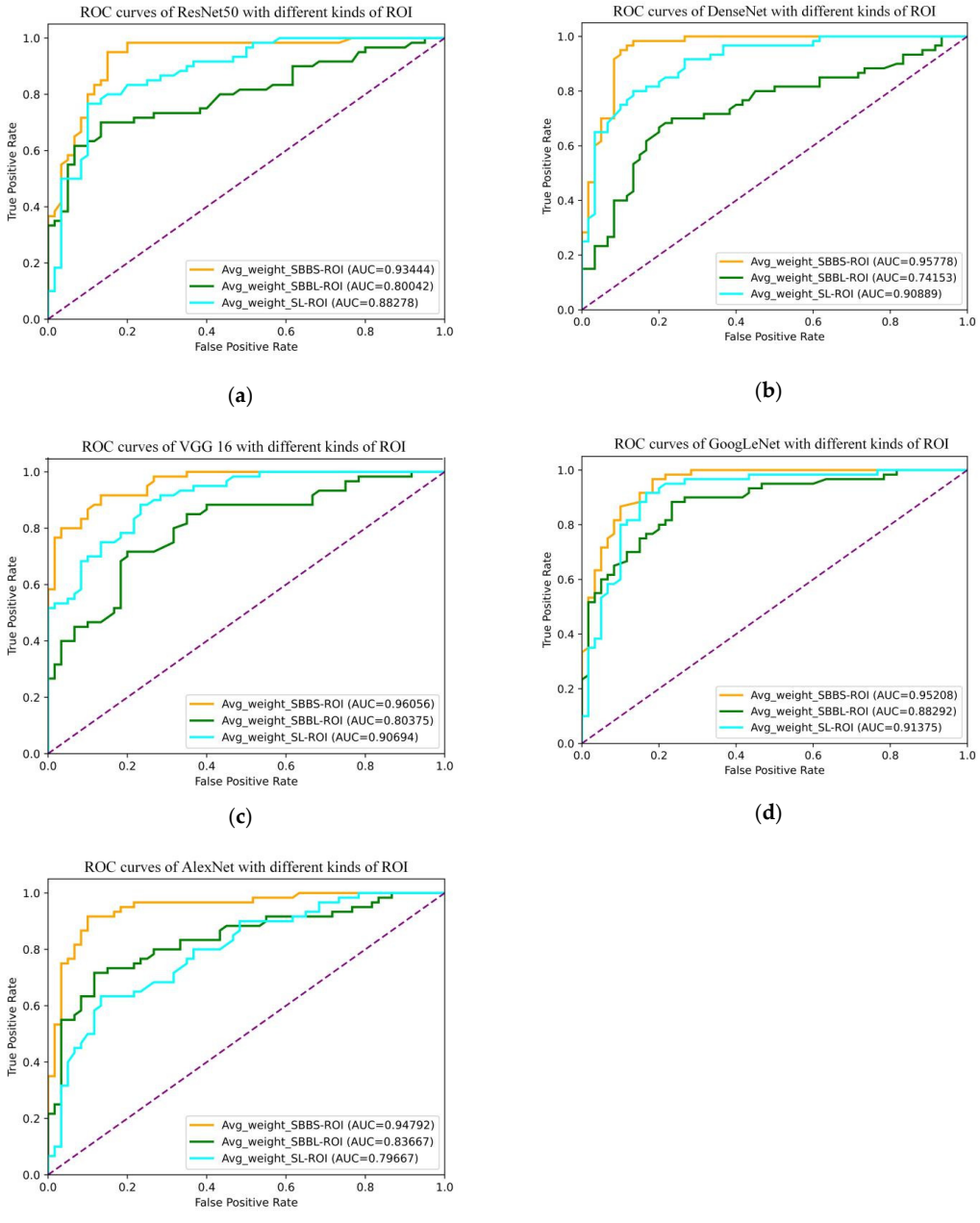


Figure 4. Lesion-level receiver operating characteristic (ROC) curves and AUC value of DCNNs with three kinds of ROI and average weighting method. (**a**) ResNet50. (**b**) DenseNet. (**c**) VGG16. (**d**) GoogLeNet. (**e**) AlexNet.

3.3. Lesion-Level Diagnosis

We also experimented with the five DCNNs, ResNet50, DenseNet, VGG16, GoogLeNet, and AlexNet, for each weighting method on the testing subset. Based on the SBBS-ROI, the DenseNet with the highest score weighting and area ratio weighting method achieved the highest accuracy of 93.3% and the VGG16 with the average weighting method achieved the highest AUC of 0.961. When taking the average AUC, the average accuracy, the average sensitivity, and the average specificity of the five DCNNs as the measure of the performance of the weighting method, the DCNNs achieved average AUCs of 0.951, 0.930, 0.935, and 0.934 with the average weighting, highest score weighting, area ratio weighting, and perimeter ratio weighting method, respectively. We took the average 0.484, 0.524, 0.480, and 0.480 as the threshold for malignant prediction for the average weighting, the highest score weighting, area ratio weighting, and perimeter ratio weighting, respectively. The DCNNs obtained average accuracy of 90.3%, 89.0%, 88.7%, and 88.8%, average sensitivity of 94.0%, 86.9%, 91.7%, and 92.3%, specificity of 86.7%, 91.0%, 85.7%, and 85.3% for the average weighting, highest score weighting, area ratio weighting, and perimeter ratio weighting, respectively. The details of the AUC, the accuracy, the sensitivity, and the specificity for each DCNN are shown in Table 5. The ROC curves of each DCNN with different weighting methods are shown in Figure 5.

Table 5. Performance of BC classification of five DCNNs with different weighting methods based on SBBS-ROI on testing subset.

DCNN	Weighting Methods	Threshold	Sensitivity	Specificity	Accuracy	AUC
ResNet50	Average	0.460	0.950	0.850	0.900	0.934
	Highest score	0.440	0.800	0.900	0.850	0.887
	Area ratio	0.440	0.917	0.833	0.875	0.919
	Perimeter ratio	0.440	0.917	0.833	0.875	0.918
DenseNet	Average	0.520	0.950	0.900	0.925	0.958
	Highest score	0.530	0.967	0.900	0.933	0.955
	Area ratio	0.520	0.950	0.917	0.933	0.939
	Perimeter ratio	0.520	0.950	0.900	0.925	0.940
VGG16	Average	0.480	0.917	0.867	0.892	0.961
	Highest score	0.570	0.830	0.950	0.892	0.949
	Area ratio	0.480	0.900	0.867	0.883	0.945
	Perimeter ratio	0.480	0.900	0.867	0.883	0.947
GoogLeNet	Average	0.490	0.967	0.817	0.892	0.952
	Highest score	0.600	0.900	0.883	0.892	0.959
	Area ratio	0.490	0.933	0.800	0.867	0.937
	Perimeter ratio	0.490	0.967	0.780	0.875	0.937
AlexNet	Average	0.470	0.917	0.900	0.909	0.948
	Highest score	0.480	0.850	0.917	0.883	0.904
	Area ratio	0.470	0.883	0.867	0.875	0.934
	Perimeter ratio	0.470	0.883	0.883	0.883	0.929
Average Score	Average	0.484	0.940	0.867	0.903	0.951
	Highest score	0.524	0.869	0.910	0.890	0.930
	Area ratio	0.480	0.917	0.857	0.887	0.935
	Perimeter ratio	0.480	0.923	0.853	0.888	0.934

3.4. DCNNs Combined with Slice-Level ROI and Lesion-Level Diagnosis

In this study, the AUC of lesion-level diagnosis and the AUC of slice-level diagnosis had a strong correlation. However, the AUC of the lesion-level diagnosis was higher than the AUC of the slice-level diagnosis. Taking DenseNet, for example, the AUC was 0.913 for slice-level diagnosis and was 0.958 for lesion-level diagnosis. The details of the performance of each DCNN are shown in Figure 6.

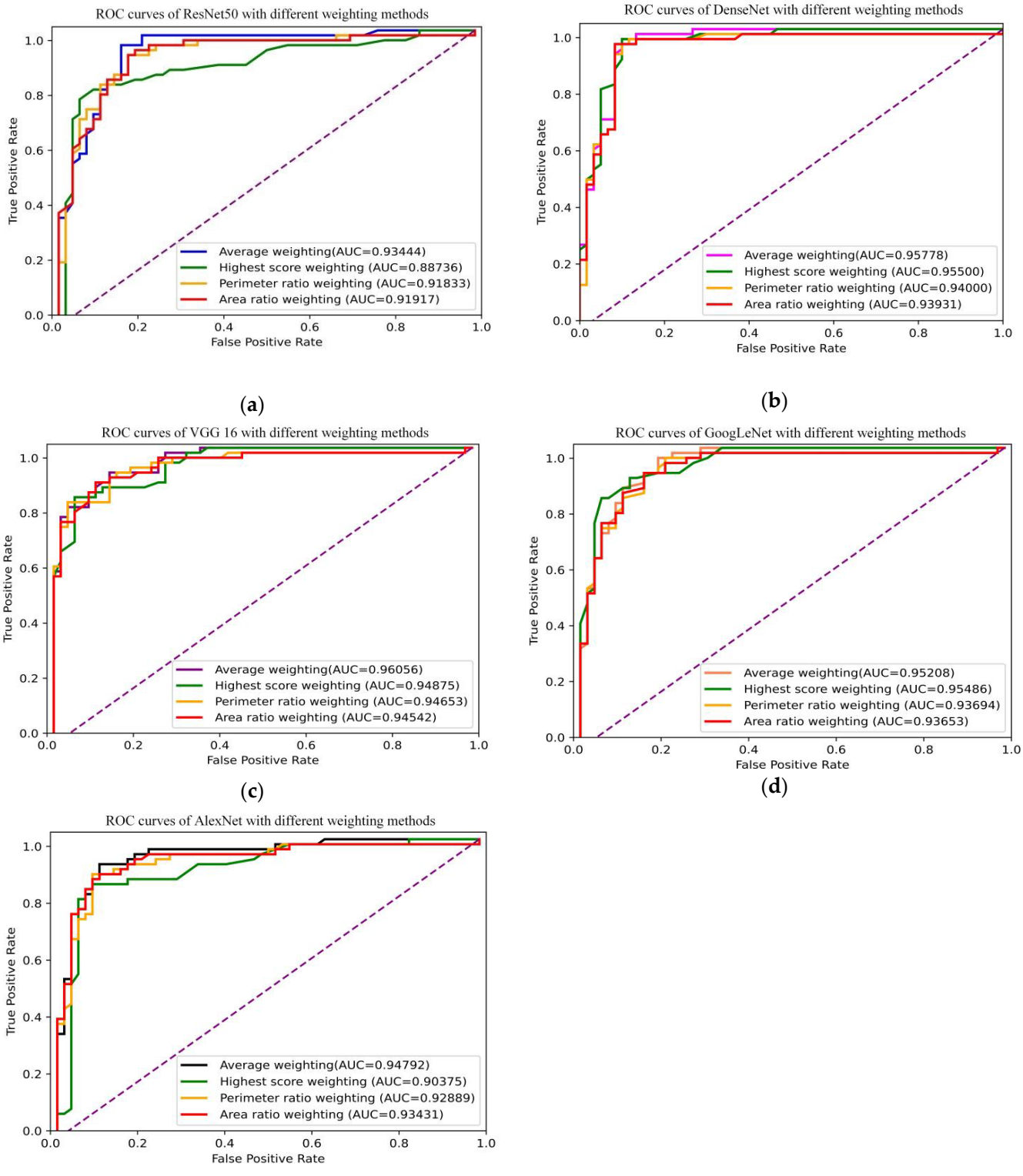
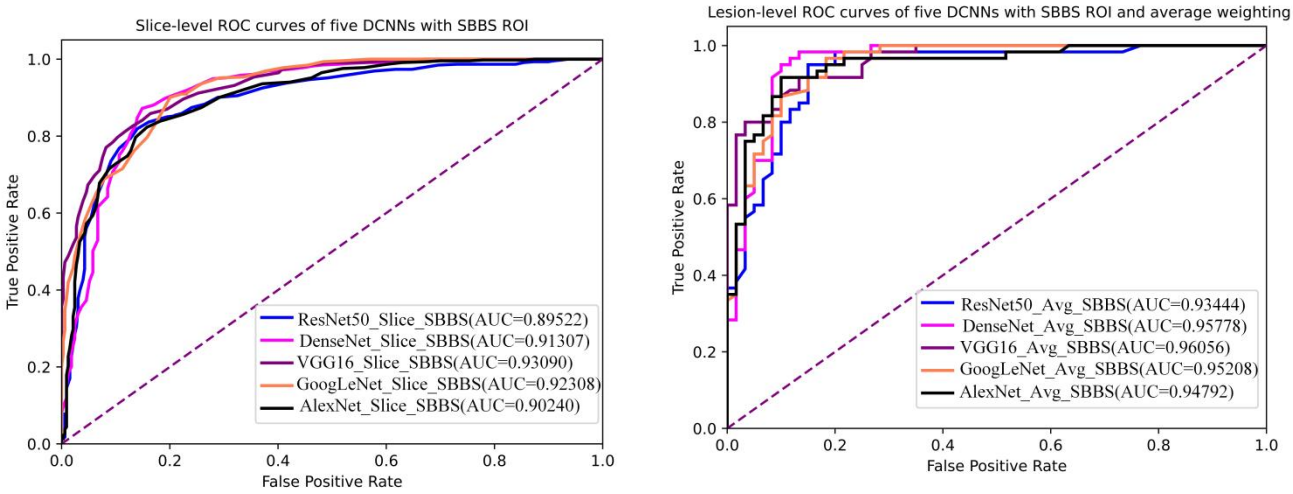


Figure 5. Lesion-level receiver operating characteristic (ROC) curves and (area under curve) AUC value of DCNNs with different weighting methods and SSBS ROI. (**a**) ResNet50. (**b**) DenseNet. (**c**) VGG16. (**d**) GoogLeNet. (**e**) AlexNet.



(**a**) (**b**)

Figure 6. ROC curves and AUC values of five DCNNs slice-level classification and lesion-level diagnosis. (**a**) Slice-based classification. (**b**) Lesion-level diagnosis.

4. Discussion

DCNN accelerates the process of BC automatic diagnosis. Features, DCNN, and the final diagnosis criterion are the three important parts to improve the performance of the DCNN automatic diagnosis system. Features mainly include radiomics features and feature maps extracted by deep networks. The radiomics features, such as texture features and morphological features, are generated by professional software from the segmented lesions. It requires professional radiologists and professional software algorithms and it is possible to bring personal subjective bias into the selection of features. In addition, these features are quantified and cannot present all the intrinsic characteristics of the lesions.

The feature maps are extracted by DCNN from the labeled ROI. The accuracy of the labeled ROI affects the performance of the DCNN. In the weakly supervised DCNN, the entire MR image is labeled as the ROI generally. Since there are multiple tissues and organs, such as the heart and blood vessels, especially in dense breasts, which cause DCNN to be confused when extracting the feature maps, the performance of weakly supervised DCNN is poor. In the strongly supervised DCNN, the entire lesion is usually labeled as the ROI by a radiologist with the smallest bounding box, which encompasses all slices of the lesion; this is the SBBL-ROI in our study. All slices of the lesion share the same ROI. Since the size and the shape of each slice of the lesion are different, the smallest bounding box of the lesion is not the smallest bounding box of each slice of the lesion. Although the surrounding organs are much less than in the whole MRI, there is still too much background information, which lowers the performance of the DCNN, especially in dense breasts. The SL-ROI is another kind of ROI that is the lesion segmented from the SBBL ROI. Its advantage is that the interference of surrounding tissues is completely removed. Its disadvantage is that the strength of enhancement and the enhancement approach cannot be extracted without the contrast of the background information. In addition, it is possible to remove the glitches of the lesions during segmentation.

Our proposed slice-level SBBS-ROI addressed the shortcomings of SBBL-ROI and SL ROI in strongly supervised DCNNs. The smallest bounding box of the lesion's slice ensured there were only a few background pixels around the lesion's slice for comparison. This minimized the interference of the surrounding organs of the lesion in each slice ROI and retained all the glitches of the lesion's slice. The approach that generated the SBBS-ROI with the region-growing algorithm was fully automatic and labor saving. The DCNNs could extract accurate feature maps for classification. The experimental results showed that the SBBS ROI improved all the DCNNs' performance. Although the ROI is very important

for the DCNN automatic diagnosis system, few articles study it. A related study [20] compared three different kinds of ROI at the lesion level and the SBBL-ROI achieved the best performance in their experiments. In our experiments, the SBBS-ROI outperforms the SBBL-ROI in all five DCNNs. In addition, the SBBS ROI can apply to all DCNN networks and be extended to the application of the other cancer diagnoses.

As for the DCNN, we evaluated five popular DCNNs in our study. We adopted the same multi-slice weighting approach when conducting the comparative experiments with different ROIs. DenseNet showed better performance in the SBBS-ROI experiment and SL-ROI experiment. We adopted the same SBBS-ROI when performing the comparative experiments with different weighting approaches. DenseNet showed the best performance in all the experiments, mostly because DenseNet established a dense connection between all the front layers and the back layers and achieved feature reuse through the connection of features on the channel. These features made DenseNet achieve good overall performance in classifying malignant and benign breast lesions of various sizes. For lesions with multiple slices, the criterion that adopts the classification results of these slices to make the final diagnosis of the lesions is worthy of study. Most of the studies only adopted one single slice of the lesion information, such as Hu et al. [27] who selected the max area slice of the lesion and Zhou et al. [20] who selected the slice of the highest classified score. Because each slice was only a section of the lesion with multi slices, the assessment result of a single slice cannot be used as the final diagnosis for a multi-slice lesion. It was more convincing to adopt the assessment results of multi slices. Therefore, we proposed the multi-slice weighting approach to diagnose a lesion in our study. In our experiments, we evaluated four multi-slice weighting methods and the result of the DCNNs on the testing subset showed that the DCNNs with the average weighting method had a better performance than other methods. In addition, it will be possible to train the weighted parameters using a deep learning method in the future.

There are several limitations to our study. First, all lesions in our dataset were labeled manually by experts, which increased the difficulty in expanding the dataset and the possibility of subjective prejudice. Second, the sagittal and coronal lesions are not distinguished in our study. Third, we did not use parameters, such as ADC and time intensity curve (TIC), which could assist radiologists in the diagnosis of lesions. Finally, there is room for improvement in accuracy and AUC. In our future work, to mitigate the first limitation, we will adopt and improve the DCNN, such as Faster RCNN, to detect and diagnose the BC automatically and with high accuracy. As for the second limitation, we will reorganize the dataset by case; as long as one of the sagittal lesions or coronal lesions is assessed as malignant, it is considered malignant, which may further improve the accuracy of the entire case. For the third limitation, we will combine the ADC and TIC with the features extracted from the ROI as the input of the DCNN classifier to improve the accuracy of the final diagnosis.

5. Conclusions

Under strongly supervised learning, the SBBS- ROI retained all the glitches and some background information around the lesion and it did not only keep the important features of the lesions but also minimized the interference of surrounding organs and tissues. It improved the performance of the DCNNs. In the five DCNNs we evaluated, DenseNet showed the best performance in SBBS-ROI, SL-ROI experiments, and the four multi-slice weighting experiments. The lesion-level diagnosis that adopted the information of multi slices of the lesion was closer to the real diagnosis and improved the performance of the DCNNs a lot. At last, we proposed the strongly supervised DCNN BC diagnosis system, which combined DenseNet SBBS-ROI, DenseNet, and the average weighting lesion-level diagnostic method had the best performance in BC diagnosis. The system achieved an AUC of 0.958, an accuracy of 92.5%, a sensitivity of 95.0%, and a specificity of 90.0%. It is the best combination for BC diagnosis in our experiments. It took 3.647 seconds to diagnose the 782 slices in the test dataset.

Author Contributions: Conceptualization, H.G. and Z.P.; methodology, H.G.; software, H.G. and H.J.; validation, H.G., Z.P. and L.X.; formal analysis, H.G.; investigation, H.G. and Z.W.; resources, L.L.; data curation, L.X. and X.J.; writing—original draft preparation, H.G.; writing—review and editing, T.S., Z.P., L.X. and Z.W.; visualization, H.J.; supervision, X.J., T.S. and Z.W.; project administration, X.J. and Z.W.; funding acquisition, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded, in part, by the National Key Research and Development Program of China under Grant 2021YFF0701000 and, in part, by the Science and Technology Program of Guangdong Province under Grant 2021B1101270007.

Data Availability Statement: DCE-MRI data used to support the findings of this study were supplied by the Sun Yat-sen University Cancer Center (Guangzhou, China) under license and have not been made freely available because of patient privacy. If our dataset is helpful to you, please contact the corresponding author by e-mail.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Latest Global Cancer Data: Cancer Burden Rises to 19.3 Million New Cases and 10.0 Million Cancer Deaths in 2020. Available online: https://www.iarc.fr/fr/news-events/ (accessed on 15 December 2020).
2. DeSantis, C.E.; Ma, J.; Goding Sauer, A.; Newman, L.A.; Jemal, A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA A Cancer J. Clin.* **2017**, *67*, 439–448. [CrossRef] [PubMed]
3. Jaglan, P.; Dass, R.; Duhan, M. Breast Cancer Detection Techniques: Issues and Challenges. *J. Inst. Eng. India Ser. B* **2019**, *100*, 379–386. [CrossRef]
4. Elmore, J.G.; Armstrong, K.; Lehman, C.D.; Fletcher, S.W. Screening for breast cancer. *JAMA* **2005**, *293*, 1245–1256. [CrossRef] [PubMed]
5. Zeeshan, M.; Salam, B.; Khalid, Q.S.B.; Sayani, R. Diagnostic Accuracy of Digital Mammography in the Detection of Breast Cancer. *Cureus* **2018**, *10*, e2448. [CrossRef] [PubMed]
6. Pisano, E.D.; Gatsonis, C.; Hendrick, E.; Yaffe, M.; Baum, J.K.; Acharyya, S.; Conant, E.F.; Fajardo, L.L.; Bassett, L.; D'Orsi, C.; et al. Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening. *N. Engl. J. Med.* **2005**, *353*, 1773–1783. [CrossRef] [PubMed]
7. Türkbey, B.; Thomasson, D.; Pang, Y.; Bernardo, M.; Choyke, P.L. The role of dynamic contrast-enhanced MRI in cancer diagnosis and treatment. *Diagn. Interv. Radiol.* **2010**, *16*, 186–192. [PubMed]
8. Liu, P.-F.; Krestin, G.P.; Huch, R.A.; Göhde, S.C.; Caduff, R.F.; Debatin, J.F. MRI of the uterus, uterine cervix, and vagina: Diagnostic performance of dynamic contrast-enhanced fast multiplanar gradient-echo imaging in comparison with fast spin-echo T2-weighted pulse imaging. *Eur. Radiol.* **1998**, *8*, 1433–1440. [CrossRef]
9. Jager, G.J.; Ruijter, E.T.; Van De Kaa, C.A.; De La Rosette, J.J.; Oosterhof, G.O.; Thornbury, J.R.; Ruijs, S.H.; Barentsz, J.O. Dynamic TurboFLASH subtraction technique for contrast-enhanced MR imaging of the prostate: Correlation with histopathologic results. *Radiology* **1997**, *203*, 645–652. [CrossRef]
10. Ocak, I.; Bernardo, M.; Metzger, G.; Barrett, T.; Pinto, P.; Albert, P.S.; Choyke, P.L. Dynamic contrast-enhanced MRI of prostate cancer at 3 T: A study of pharmacokinetic parameters. *AJR Am. J. Roentgenol.* **2007**, *189*, 192–201. [CrossRef]
11. Tartar, M.; Comstock, C.E.; Kipper, M.S. CHAPTER 2—Evaluation of the Symptomatic Patient: Diagnostic Breast Imaging. *Breast Cancer Imaging* **2008**, *2008*, 38–75.
12. Wei, M.; Du, Y.; Wu, X.; Su, Q.; Zhu, J.; Zheng, L.; Lv, G.; Zhuang, J. A benign and Malignant Breast Tumor Classification Method via Efficiently Combining Texture and Morphological Features on Ultrasound Images. *Comput. Math. Methods Med.* **2020**, *2020*, 5894010. [CrossRef] [PubMed]
13. Pang, Z.; Zhu, D.; Chen, D.; Li, L.; Shao, Y. A computer-aided diagnosis system for dynamic contrast-enhanced MR images based on level set segmentation and ReliefF feature selection. *Comput. Math. Methods Med.* **2015**, *2015*, 450531. [CrossRef] [PubMed]
14. Liu, H.; Wang, J.; Gao, J.; Liu, S.; Liu, X.; Zhao, Z.; Guo, D.; Dan, G. A comprehensive hierarchical classification based on multi-features of breast DCE-MRI for cancer diagnosis. *Med. Biol. Eng. Comput.* **2020**, *58*, 2413–2425. [CrossRef]
15. Cai, H.; Peng, Y.; Ou, C.; Chen, M.; Li, L. Diagnosis of Breast Masses from Dynamic Contrast-Enhanced and Diffusion-Weighted MR: A Machine Learning Approach. *PLoS ONE* **2014**, *9*, e87387. [CrossRef] [PubMed]
16. Dalmiş, M.U.; Gubern-Mérida, A.; Vreemann, S.; Bult, P.; Karssemeijer, N.; Mann, R.; Teuwen, J. Artificial Intelligence–Based Classification of Breast Lesions Imaged with a Multiparametric Breast MRI Protocol with Ultrafast DCE-MRI, T2, and DWI. *Investig. Radiol.* **2019**, *54*, 325–332. [CrossRef] [PubMed]
17. Jiao, H.; Jiang, X.; Pang, Z.; Lin, X.; Huang, Y.; Li, L. Deep Convolutional Neural Networks-Based Automatic Breast Segmentation and Mass Detection in DCE-MRI. *Comput. Math. Methods Med.* **2020**, *2020*, 2413706. [CrossRef] [PubMed]
18. Liu, M.Z.; Swintelski, C.; Sun, S.; Siddique, M.; Desperito, E.; Jambawalikar, S.; Ha, R. Weakly Supervised Deep Learning Approach to Breast MRI Assessment. *Acad. Radiol.* **2021**, *29*, 166–172. [CrossRef]

19. Zhou, J.; Luo, L.; Dou, Q.; Chen, H.; Chen, C.; Li, G.; Jiang, Z.; Heng, P.A. Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. *J. Magn. Reson. Imaging* **2019**, *50*, 1144–1151. [CrossRef]
20. Zhou, J.; Zhang, Y.; Chang, K.; Lee, K.E.; Wang, O.; Li, J.; Lin, Y.; Pan, Z.; Chang, P.; Chow, D.; et al. Diagnosis of Benign and Malignant Breast Lesions on DCE-MRI by Using Radiomics and Deep Learning with Consideration of Peritumor Tissue. *J. Magn. Reson. Imaging* **2019**, *51*, 798–809. [CrossRef]
21. He, K.; Zhang, Y.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
22. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
23. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
24. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Andrew Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
25. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *NIPS* **2012**, *60*, 84–90.
26. Hojjatoleslami, S.A.; Kittler, J. Region growing: A new approach. *IEEE Trans. Image Process. A Publ. IEEE Signal Proces. Soc.* **1998**, *7*, 1079–1084. [CrossRef] [PubMed]
27. Hu, Q.; Whitney, H.M.; Giger, M.L. A deep learning methodology for improved breast cancer diagnosis using multiparametric MRI. *Sci. Rep.* **2020**, *10*, 10536. [CrossRef] [PubMed]