Original Article

End-to-end deep learning radiomics: development and validation of a novel attention-based aggregate convolutional neural network to distinguish breast diffuse large B-cell lymphoma from breast invasive ductal carcinoma

Wen Chen[1,2,3,4,5,6#], Fei Liu[1,3,4,5,6#], Rui Wang[1,3,4,5,6], Ming Qi[1,3,4,5,6], Jianping Zhang[1,3,4,5,6], Xiaosheng Liu[1,3,4,5,6], Shaoli Song[1,3,4,5,6]

[1]Department of Nuclear Medicine, Fudan University Shanghai Cancer Center, Shanghai, China; [2]Academy for Engineering and Technology, Fudan University, Shanghai, China; [3]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China; [4]Center for Biomedical Imaging, Fudan University, Shanghai, China; [5]Shanghai Engineering Research Center of Molecular Imaging Probes, Shanghai, China; [6]Key Laboratory of Nuclear Physics and Ion-beam Application (MOE), Fudan University, Shanghai, China

Contributions: (I) Conception and design: W Chen, F Liu; (II) Administrative support: S Song, X Liu, W Chen, F Liu; (III) Provision of study materials or patients: W Chen, F Liu, R Wang; (IV) Collection and assembly of data: W Chen, F Liu, M Qi, J Zhang; (V) Data analysis and interpretation: W Chen, F Liu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Xiaosheng Liu, PhD; Shaoli Song, PhD. Department of Nuclear Medicine, Fudan University Shanghai Cancer Center, No. 270 Dong'an Road., Xuhui District, Shanghai 200032, China. Email: drliuxiaosheng@163.com; shaoli-song@163.com.

Background: Apart from invasive pathological examination, there is no effective method to differentiate breast diffuse large B-cell lymphoma (DLBCL) from breast invasive ductal carcinoma (IDC). In this study, we aimed to develop and validate an effective deep learning radiomics model to discriminate between DLBCL and IDC.

Methods: A total of 324 breast nodules from 236 patients with baseline ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT) were retrospectively analyzed. After grouping breast DLBCL and breast IDC patients, external and internal datasets were divided according to the data collected by different centers. Preprocessing was then used to process the original PET/CT images and an attention-based aggregate convolutional neural network (AACNN) model was designed. The AACNN model was trained using patches of CT or PET tumor images and optimized with an improved loss function. The final ensemble predictive model was built using distance weight voting. Finally, the model performance was evaluated and statistically verified.

Results: A total of 249 breast nodules from Fudan University Shanghai Cancer Center (FUSCC) and 75 breast nodules from Shanghai Proton and Heavy Ion Center (SPHIC) were selected as internal and external datasets, respectively. On the internal testing, our method yielded an area under the curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), and harmonic mean of precision and sensitivity (F1) of 0.886, 83.0%, 80.9%, 85.0%, 84.8%, 81.2%, and 0.828, respectively. Meanwhile on the external testing, the results were 0.788, 71.6%, 61.4%, 84.7%, 84.0%, 62.6%, and 0.709, respectively.

Conclusions: Our study outlines a deep learning radiomics method which can automatically, noninvasively, and accurately differentiate breast DLBCL from breast IDC, which will be more in line with the needs and strategies of precision medicine, individualized diagnosis, and treatment.

Keywords: Breast diffuse large B-cell lymphoma (breast DLBCL); breast invasive ductal carcinoma (IDC); positron emission tomography/computed tomography (PET/CT); deep learning radiomics; precision medicine

Submitted Dec 01, 2022. Accepted for publication Jul 24, 2023. Published online Aug 15, 2023.
doi: 10.21037/qims-22-1333
View this article at: https://dx.doi.org/10.21037/qims-22-1333

Introduction

Breast cancer is the most common cancer and the leading cause of death in women worldwide, of which invasive ductal carcinoma (IDC) is the most common pathological type. In China, newly diagnosed breast cancer cases accounts for 16.72% (n=306,000) of all new cancer cases and the number of deaths accounts for 8.12% (n=71,700) of the total cancer-related deaths in women (1,2). The current treatment for breast cancer includes surgery, chemotherapy, hormonal therapy, radiotherapy, targeted therapy, and immunotherapy, among which surgery is the first choice for early-stage breast cancers (3). Lymphoma represents a diverse group of diseases caused by clonal proliferation of lymphocytes (4). Lymphomas usually involve lymphoid organs and tissues. However, about 40% of lymphomas can also be located outside the lymph nodes (5). Breast lymphoma is a rare disease, the incidence of which is less than 0.5% of all breast malignancies, and diffuse large B-cell lymphoma (DLBCL) is the most common type of breast lymphoma (6). Breast lymphoma usually presents as a painless palpable mass, which is indistinguishable from breast cancer, and the common imaging examinations include mammographic, ultrasound, computed tomography (CT), and breast magnetic resonance imaging (MRI), which are unable to distinguish lymphoma from breast cancer (7,8). The most common diagnostic methods used to differentiate breast lymphoma from breast cancer are physical examination or image-guided needle biopsy (9). However, needle biopsy is an invasive technique which can damage normal breast tissue; misdiagnosis can occur if necrotic tissue is pierced when puncturing a large breast nodule containing necrotic tissue. Furthermore, as breast lymphoma is a rare disease with a low incidence, many breast nodules are misdiagnosed as breast cancer and are surgically removed directly (10). Unlike for breast cancer, which is primarily treated with surgery, radiation and chemotherapy are the main treatment modalities for breast lymphoma (9,11). Therefore, it is clinically important and necessary to explore a non-invasive technique that can differentially diagnose breast DLBCL from breast IDC in advance.

^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT) has the dual advantages of anatomical and functional imaging, which plays an increasingly important role in the diagnosis and treatment of lymphoma (12,13). ^{18}F-FDG PET/CT imaging has been used for the initial staging, restaging, early treatment response, efficacy evaluation, prognosis prediction, and follow-up of lymphoma patients (14,15). Radiomics refers to the high-throughput extraction of a large number of image features that describe the characteristics of tumors, which is non-invasive and has been widely used for auxiliary diagnosis and classification for tumors (16). Ou *et al.* showed that radiomics features extracted from PET/CT and PET parameters can differentiate breast lymphoma from breast carcinoma using machine learning (17,18). However, their study was conducted within a single institution and the sample was very small, with only 44 patients recruited. Moreover, machine learning is a manual and labor-intensive technique with poor feature stability and repeatability (19,20). Despite these insufficiencies, no other non-invasive methods for distinguishing the 2 diseases have been reported so far. Deep learning is one of the hottest branches of machine learning, the biggest advantage of which is that it is deep enough and the network capacity is large enough. The most obvious benefit of a deep enough network is that it can accommodate richer semantic information (21,22).

Hence, this study was conducted to develop and validate an end-to-end deep learning radiomics method through ^{18}F-FDG PET/CT images that could accurately and noninvasively distinguish breast DLBCL from breast IDC. We present this article in accordance with the TRIPOD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-22-1333/rc).

Methods

The outline of the workflow from radiomic analysis with

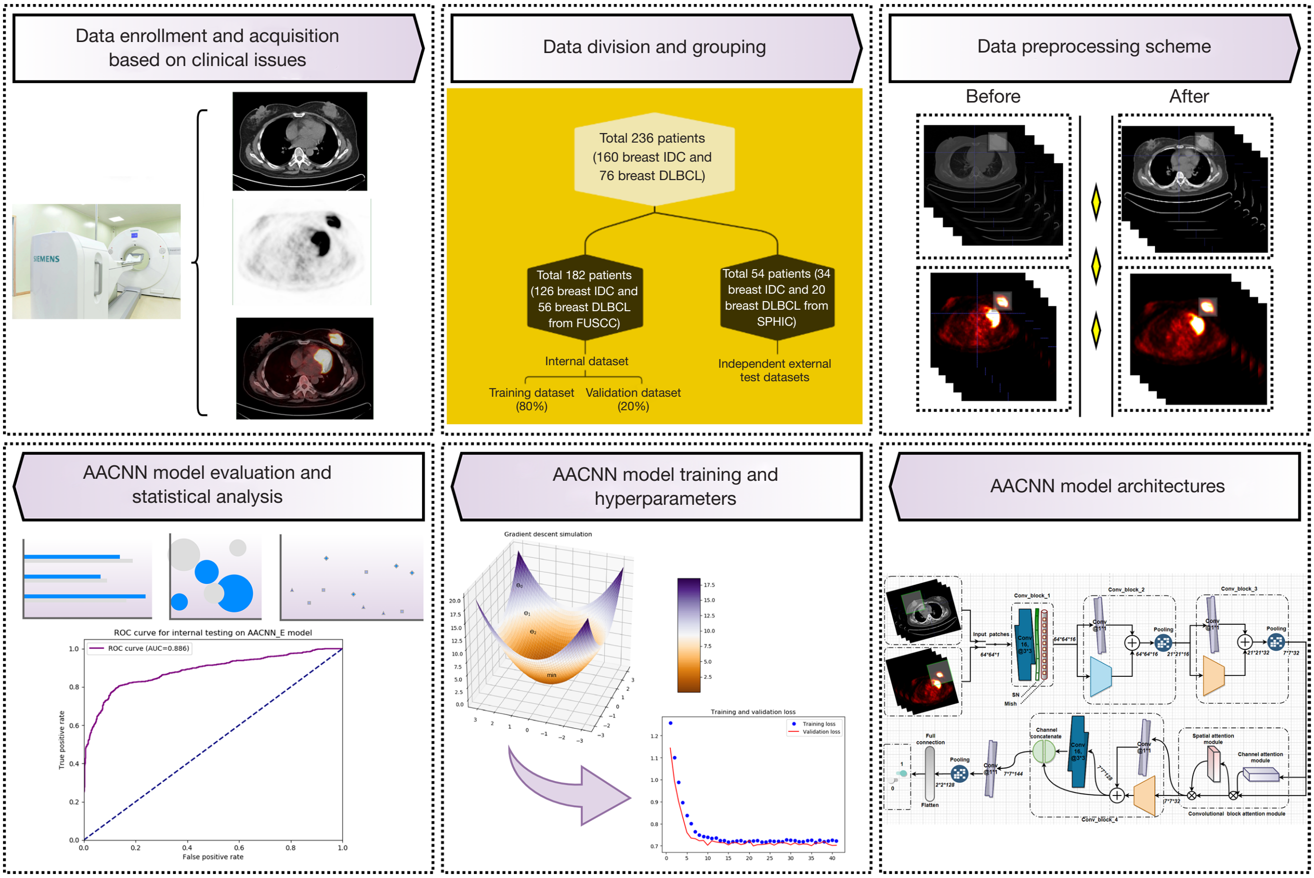
Data enrollment and acquisition based on clinical issues
SIEMENS
Data division and grouping
Total 236 patients (160 breast IDC and 76 breast DLBCL)
Total 182 patients (126 breast IDC and 56 breast DLBCL from FUSCC)
Total 54 patients (34 breast IDC and 20 breast DLBCL from SPHIC)
Internal dataset
Independent external test datasets
Training dataset (80%)
Validation dataset (20%)
Data preprocessing scheme
Before
After
AACNN model evaluation and statistical analysis
ROC curve for internal testing on AACNN_E model
ROC curve (AUC=0.886)
True positive rate
False positive rate
AACNN model training and hyperparameters
Gradient descent simulation
min
Training and validation loss
Training loss
Validation loss
AACNN model architectures
Input patches
Conv_block_1
Conv_block_2
Conv_block_3
Conv_block_4
Pooling
SN
Mish
Full connection
Flatten
Channel concatenate
Spatial attention module
Channel attention module
Convolutional block attention module

Figure 1 Outline of the workflow from radiomic analysis with deep learning. IDC, invasive ductal carcinoma; DLBCL, diffuse large B-cell lymphoma; FUSCC, Fudan University Shanghai Cancer Center; SPHIC, Shanghai Proton and Heavy Ion Center; AACNN, attention-based aggregate convolutional neural network; ROC, receiver operating characteristic; AUC, area under the ROC curve.

deep learning is illustrated in *Figure 1*, which mainly included the following steps: (I) data enrollment and ^{18}F-FDG PET/CT acquisition; (II) patient division and grouping with IDC and DLBCL; (III) tumor annotation and PET/CT image preprocessing; (IV) attention-based aggregate convolutional neural network (AACNN) model architectures building; (V) model training and hyperparameter optimization; (VI) model evaluation based on prediction results and statistical analysis.

Study population

In our study, we analyzed 324 breast nodules from 236 patients (160 breast IDC and 76 breast DLBCL patients) who underwent baseline ^{18}F-FDG PET/CT from January 2009 to December 2021 at Fudan University Shanghai Cancer Center (FUSCC) or Shanghai Proton and Heavy Ion Center (SPHIC). All patients aged over 18 years with a pathological diagnosis of breast IDC or DLBCL were included. The exclusion criteria were as follows: (I) other pathological types apart from IDC or DLBCL; (II) surgery, chemotherapy, radiotherapy, or other treatments had been administered before ^{18}F-FDG PET/CT imaging; (III) other types of cancers; (IV) incomplete clinical data.

A total of 249 breast nodules (182 patients) from FUSCC were selected as the internal dataset and were used to train and build the AACNN model. At a ratio of 4:1, these participants were randomly divided into the training and testing datasets based on stratified sampling (23-25). A further 75 nodules (54 patients) from SPHIC were selected as an independent external dataset to further evaluate the robustness and generalization ability of the AACNN model. Details of patient selection and grouping are shown in *Figure 2*. We reviewed the hospital medical

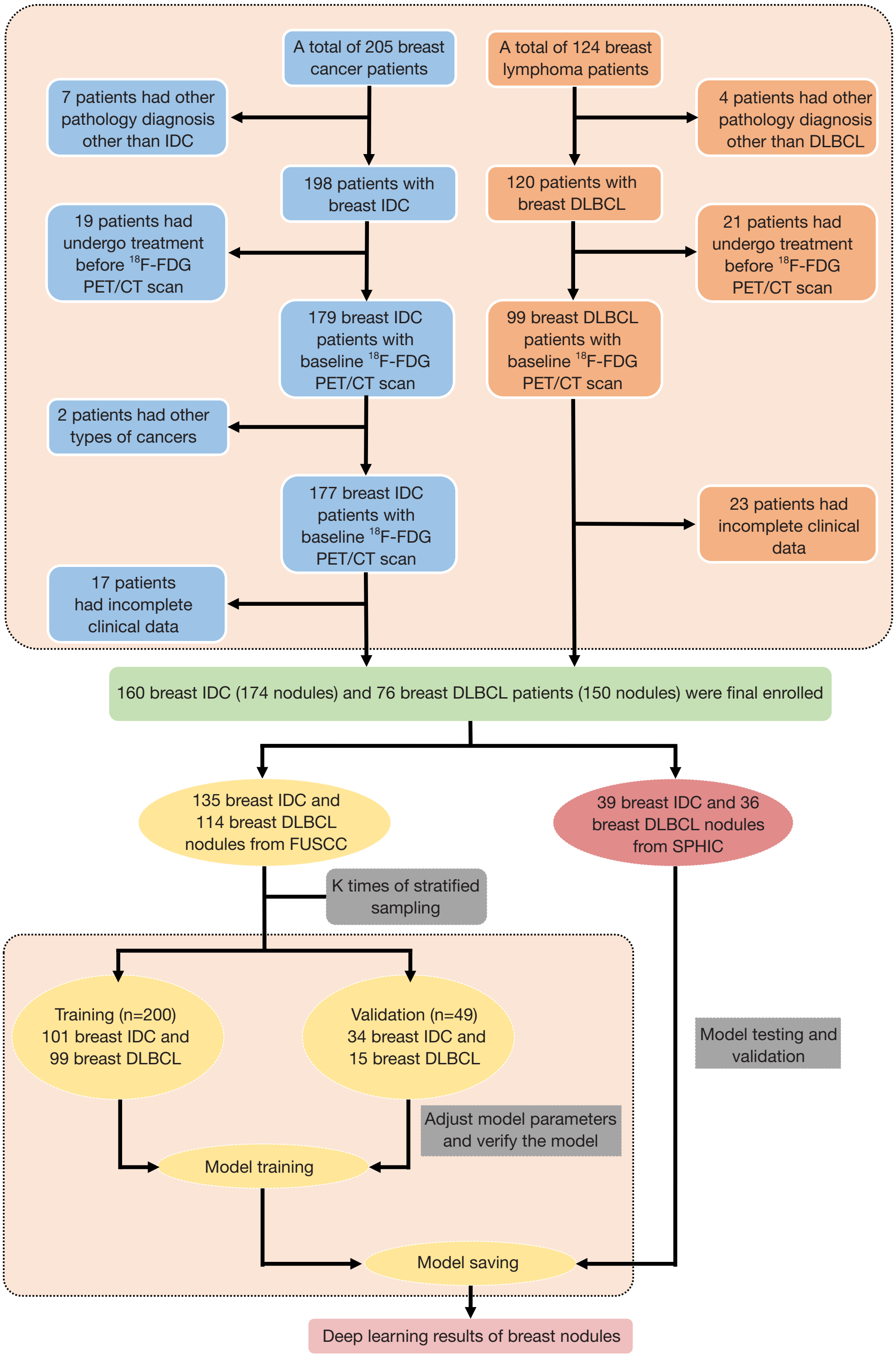
A total of 205 breast cancer patients
7 patients had other pathology diagnosis other than IDC
198 patients with breast IDC
19 patients had undergo treatment before ¹⁸F-FDG PET/CT scan
179 breast IDC patients with baseline ¹⁸F-FDG PET/CT scan
2 patients had other types of cancers
177 breast IDC patients with baseline ¹⁸F-FDG PET/CT scan
17 patients had incomplete clinical data
A total of 124 breast lymphoma patients
4 patients had other pathology diagnosis other than DLBCL
120 patients with breast DLBCL
21 patients had undergo treatment before ¹⁸F-FDG PET/CT scan
99 breast DLBCL patients with baseline ¹⁸F-FDG PET/CT scan
23 patients had incomplete clinical data
160 breast IDC (174 nodules) and 76 breast DLBCL patients (150 nodules) were final enrolled
135 breast IDC and 114 breast DLBCL nodules from FUSCC
39 breast IDC and 36 breast DLBCL nodules from SPHIC
K times of stratified sampling
Training (n=200) 101 breast IDC and 99 breast DLBCL
Validation (n=49) 34 breast IDC and 15 breast DLBCL
Adjust model parameters and verify the model
Model training
Model testing and validation
Model saving
Deep learning results of breast nodules

Figure 2 Flow chart of patient selection and study design. IDC, invasive ductal carcinoma; DLBCL, diffuse large B-cell lymphoma; ^{18}F-FDG, ^{18}F-fluorodeoxyglucose; PET/CT, positron emission tomography/computed tomography; FUSCC, Fudan University Shanghai Cancer Center; SPHIC, Shanghai Proton and Heavy Ion Center.

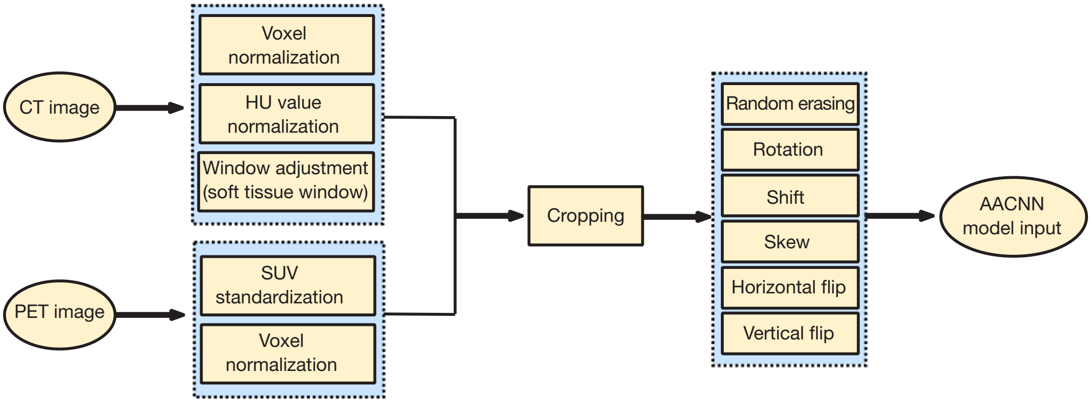

Figure 3 Flowchart of image preprocessing and preparation. SUV, standardized uptake value; CT, computed tomography; PET, positron emission tomography; HU, heat unit; AACNN, attention-based aggregate convolutional neural network.

records of all 236 patients, and the baseline characteristics including gender, age, height, weight, and clinical stage of the patients, and the size, volume, and PET parameters of nodules were summarized. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committees of FUSCC (No. 1909207-14-1910) and SPHIC (No. 200217EXP-01) and the requirement for individual consent for this retrospective analysis was waived.

^{18}F-FDG PET/CT image acquisition

The baseline ^{18}F-FDG PET/CT imaging were performed on the Biograph 16HR and mCT Flow Siemens PET/CT scanners (Siemens, Erlangen, Germany) in FUSCC and a Biograph 16 Siemens PET/CT scanner (Siemens, Germany) in SPHIC. Before the scan, all patients were fasted for at least 6 hours with the blood glucose levels maintained below 8 mmol/L. After receiving an intravenous injection with 3.7 MBq/kg of ^{18}F-FDG, the patients rested for approximately 1 hour, and then a whole-body scan from the head to the mid-thigh was performed for each patient. For PET/CT scans, the CT scan was performed firstly using a low-dose technique (120 kV and 140 mA for Biograph 16HR and mCT Flow scanners, 120 kV and 150 mA for Biograph 16 scanner). PET scans were acquired with 2 minutes allocated for each table position. The PET imaging dataset was reconstructed using the ordered subsets expectation maximum (OSEM) iterative reconstruction algorithm with CT data attenuated.

Image preprocessing and preparation

As described above, PET/CT images were acquired by 3 different machines in 2 different centers. On the one hand, the principles of CT and PET imaging are different, and on the other hand, the image information contained in these 2 image modalities is different, so we adopted different image preprocessing methods accordingly.

A diagram displaying the specifics of PET/CT image preprocessing is provided in *Figure 3*. The slice thickness and in-plane resolution of CT images are different. Firstly, all CT images were resampled into isotropic voxels of unit dimension through the bilinear interpolation algorithm to ensure comparability, and the voxel volume was set to 1 mm^3 (26). Secondly, we normalized the heat unit (HU) maximum and minimum values and adjusted the window range to (−200 to 300). Similarly, the slice thickness and in-plane resolution of PET images were not consistent. Firstly, we converted PET scanning intensity from count units (CNTS) with absolute activity concentration (Bq/mL) to standard uptake value (SUV) values according to patient weight (27). Secondly, the PET image was resampled to 1 mm^3 voxel using the bilinear interpolation algorithm to ensure comparability and reduce the impact on model classification (28).

The ITK-SNAP software (http://www.itksnap.org/) was used to segment all breast nodules on PET images with anatomical information provided by CT images. The semi-automatic segmentation was interpreted and delineated by 2 experienced nuclear medicine radiologists who were blinded to the patients' pathological information. If the radiologists

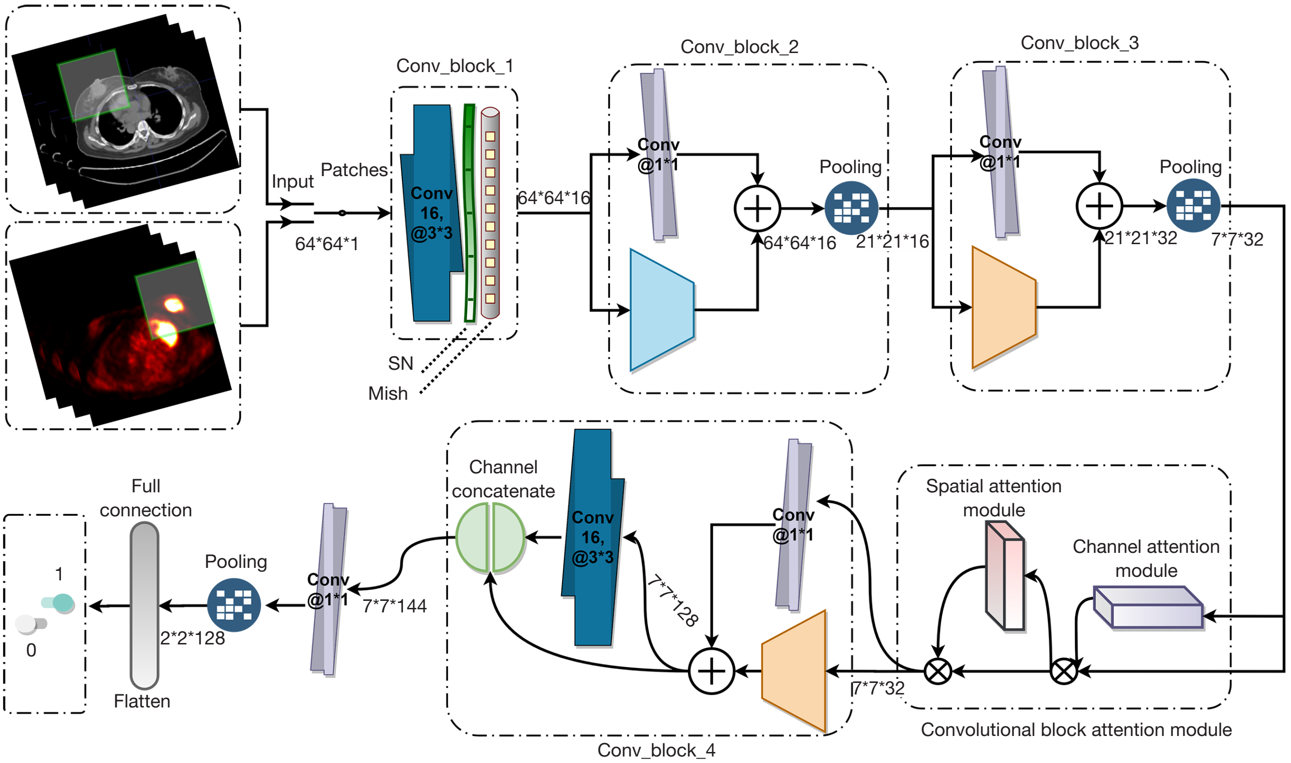

Figure 4 Schematic diagram of the network structure of the AACNN model. SN, switchable normalization; AACNN, attention-based aggregate convolutional neural network.

had different opinions, a consensus was reached through discussion. According to the segmentation results confirmed by nuclear medicine radiologists, we cut each tumor region of interest (ROI) into 2-dimensional (2D) patches of 128×128 (PET images) and 64×64 (CT images). Finally, a series of data balance and enhancement techniques (rotation, translation, horizontal and vertical flip, etc.) were applied to increase the data size and data diversity of the training image, so as to prepare for the data input of the AACNN model (29).

For quantitative analysis, the 2D and 3-dimensional (3D) nodule size, the volume, and PET parameters including the minimum, mean, and maximum SUV (SUV_{min}, SUV_{mean}, and SUV_{max}, respectively) were calculated using python programs by delineating the ROI of the lesions. Metabolic tumor volume (MTV) was recorded at the threshold of 41% SUV_{max} in hypermetabolic regions, and total lesion glucose (TLG) was calculated following the equation below:

$$TLG = SUV_{mean} \times MTV \qquad [1]$$

AACNN model architectures

Figure 4 shows the network structure of our proposed AACNN model for discriminating breast lymphoma from breast cancer. The AACNN model takes PET tumor region patches and CT tumor region patches as input, respectively. It trains different independent models to identify disease classes based on the above images. The network architecture contains and extends bottleneck modules for dimensionality or feature reduction, residual modules that facilitate gradient propagation, and includes element-wise addition and channel-wise concatenation operation. Besides, it uses hierarchical and iterative skip connections to merge class-feature pyramid networks for improved resolution (30,31).

The first block of the network consists of a simple 3×3 convolutional layer (32), a switchable normalization layer (33), and the Mish activation function (34). The convolution calculation principle is given in Eq. [2], the adaptive normalization principle is given in Eq. [3], and the Mish activation function principle is given in Eq. [4].

$$N = \frac{(W - K + 2 \times P)}{S} + 1 \qquad [2]$$

N represents the output size, W represents the input size, K represents the convolution kernel size, P represents the size of the padding value, and S represents the step size.

$$\hat{h}_{ncij} = \gamma \frac{h_{ncij} - \sum_{k\in\Omega} w_k \mu_k}{\sqrt{\sum_{k\in\Omega} w'_k \delta_k^2 + \varepsilon}} + \beta \qquad [3]$$

Each pixel of the image is denoted as h_{ncij}, and the normalized pixel is denoted as. $\hat{h}_{ncij}$ n,c,i,j are the minibatch size, the number of channels, the width of channels, and the height of channels, respectively. The statistics of switchable normalization (mean μ, variance σ2) are weighted and averaged by a suitable normalization method in the set $\Omega=[bn, in, \ln]$. w_k and $w_k^{'}$ are the weight coefficients of the corresponding statistics.

$$f(x) = x \times \tanh\left(\ln\left(1 + e^x\right)\right) \qquad [4]$$

x represents the input feature and $f(x)$ represents the activated feature.

The second block of the network includes 2 branches. The main branch contains a bottleneck module and a pooling layer to reduce the feature dimension. 1×1 convolutions were inserted to learn more non-sparse features. The feature information of the 2 branches is superimposed. Inspired by the Inception V3 network structure (35), the third and fourth blocks of the network used asymmetric convolutions with 1×n and n×1 filters to replace the traditional n×n symmetric convolution on the basis of the second part, where channel ∈ [8, 32]. The purpose of this structural design is to reduce the parameter size and computational cost. In order to better highlight the main features and suppress irrelevant features, we also added a convolutional block attention module (CBAM) between the third and fourth blocks (36). In the fourth block, channel-wise concatenation was added to further enhance deep features. The equation for calculating the CBAM is as follows:

$$\begin{aligned}
&F^{'} = M_c(F) \otimes F, \\
&F^{''} = M_s(F^{'}) \otimes F^{'}, \\
&M_c(F) = \sigma\left(MLP\left(AvgPool(F)\right) + MLP\left(MaxPool(F)\right)\right), \\
&M_s(F) = \sigma\left(f^{7\times7}\left(\left[AvgPool(F); MaxPool(F)\right]\right)\right)
\end{aligned} \qquad [5]$$

Where F represents intermediate features, M_c represents channel attention features, M_s represents spatial attention features, ⊗ represents element-wise multiplication, $F^{'}$ represents features computed by the channel attention mechanism, $F^{''}$ is the output of final refinement, MLP is multi-layer perceptron, σ is sigmoid function, *AvgPool* stands for average pooling, *MaxPool* stands for maximum pooling, and $f^{7\times7}$ is a 7×7 convolution kernel.

Model training and hyperparameters

The input size of the AACNN model was set to 64×64×1. During model training, the convolution weights were initialized using the 'He' uniform variance scaling initialization method (37). The L2 regularization technique (weight decay coefficient =0.0001) was used to limit the squared magnitude of the kernel weights to avoid overfitting. The Dropout technique was applied to the fully connected layer to make the fully connected layer randomly select some nodes to not play a role during the learning process (38). The Adam optimizer was used to update the network gradient by minimizing the loss function to train the model. In order to ensure the robustness of the prediction model and improve the computational cost efficiency, an early termination condition was introduced, and the training was automatically stopped when the accuracy of the validation set had still not improved after 20 repetitions. Other hyperparameter settings included a dynamic learning rate of 0.001, a number of iterations of 1,000, and a batch size of 64.

In addition, to alleviate the impact of erroneous labels, based on the idea of label smoothing (39), we improved the binary cross-entropy loss function, which is formulated as follows:

$$Loss = \begin{cases} (1-\varepsilon) \times \left(-\sum_{i=1}^{2} p_i \log q_i\right), & \text{if } i = y \\ \varepsilon \times \left(-\sum_{i=1}^{2} 0.2 \times p'_i \log q_i\right), & \text{otherwise} \end{cases} \qquad [6]$$

Where i represents 1 of the 2 classes, y represents the actual class, p_i represents the actual label, ε represents constant 0.1, q_i represents the predicted label, and p'_i represents an array of 1 (which has the same shape and type as the predicted label array).

Performance evaluation

In this study, multiple repeated random subsampling validation was used to evaluate the model, which has the same purpose as the fixed split in k-fold cross validation, and can better control the number of model training and validation, as well as the ratio of training set to testing set, and obtain more accurate results (37). The performance of the AACNN model was investigated by 7 evaluation indexes, namely, accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), harmonic mean of precision and sensitivity (F1), and the area under the receiver operating characteristic (ROC) curve (AUC). The equations are expressed as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad [7]$$

$$SEN = \frac{TP}{TP + FN} \quad [8]$$

$$SPE = \frac{TN}{TN + FP} \quad [9]$$

$$PPV = \frac{TP}{TP + FP} \quad [10]$$

$$NPV = \frac{TN}{TN + FN} \quad [11]$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad [12]$$

TP represents the number of true positives, *FP* represents the number of false positives, *TN* represents the number of true negatives, and *FN* represents the number of false negatives. SEN represents the proportion of all DLBCL lesions that have been correctly predicted by the AACNN model, which measures AACNN's ability to recognize breast lymphoma. SPE represents the proportion of all IDC lesions that have been correctly predicted by the AACNN model, which measures AACNN's ability to recognize breast cancer. PPV indicates the proportion of DLBCL lesions predicted by the AACNN model that are indeed DLBCL. NPV indicates the proportion of IDC lesions predicted by that AACNN model that are indeed IDC.

Platform and software

A computer running Windows 10 with an Intel Core I7-8750H @ 2.2 GHz × 2.2 GHz CPU, 8 GB memory, and an Nvidia GeForce GTX 1060 graphics processing unit (Nvidia, San Jose, CA, USA) was used for the analysis. The software used in this paper included PyCharm (version 2018.2, http://www.jetbrains.com), ITK-SNAP (version 3.8.0), 3D Slicer (version 4.11; http://www.slicer.org), MATLAB (version R2018a; http://www.mathworks.com), and SPSS (version 26.0; http://www.spss.com.cn). To build the AACNN model and process PET/CT images, we utilized several publicly available software packages, including Keras, SimpleITK, NumPy, skimage, os, pandas, SciPy, Scikit-Learn, and Pydicom.

Statistical analysis

The SPSS software (IBM Corp., Armonk, NY, USA) was used for the statistical analysis. Statistical methods included independent samples *t*-test for the comparison between 2 groups, Mann-Whiney *U* test for continuous characteristics, and χ^2 test or Fishers exact test for categorical characteristics. DeLong's method using PyCharm was used to test the statistical significance between AUCs. A P value of below 0.05 was considered statistically significant.

Results

Study population

A total of 236 patients (235 females and 1 male; average age, 51.31 years; range 24 to 86 years) were included in our multi-center study (*Table 1*). Among the patients, the average height, weight, and body mass index (BMI) was 1.59±0.05 m, 58.90±8.69 kg, and 23.30±3.24 kg/m^2, respectively. The clinical features of the patients were not significantly different between the groups. Of the 236 patients, 160 (67.80%) had breast IDC and 76 (32.20%) had breast DLBCL (Table S1). Among the breast IDC patients, 12 (7.50%), 66 (41.25%), 31 (19.38%), and 51 (31.88%) patients were with tumor-node-metastasis (TNM) stage I, II, III, and IV, and among the breast DLBCL patients, 24 (31.58%), 30 (39.47%), 1 (1.32%), and 21 (27.63%) patients were with Ann Arbor stage I, II, III, and IV, respectively. The clinical stage between breast IDC and DLBCL was statistically significant (P<0.001).

As shown in *Table 2*, a total of 249 (135 IDC and 114 DLBCL) and 75 (39 IDC and 36 DLBCL) breast nodules were included in the internal and external datasets, respectively. The clinical characteristics of the nodules had no differences between the internal and external datasets. The nodule size (containing 2D and 3D size and nodule volume) and PET parameters (containing SUV_{mean}, SUV_{max}, MTV, and TLG) were significantly different between the

Table 1 Patient characteristics in the internal and external datasets

Characteristics	Total (n=236)	Internal dataset (n=182)	External dataset (n=54)	P value
Sex				0.585
Female	235 (99.58)	181 (99.45)	54 (100.00)	
Male	1 (0.42)	1 (0.55)	0 (0.00)	
Age (year)	51.31±11.96	51.15±12.06	51.85±11.72	0.707
Height (m)	1.59±0.05	1.59±0.05	1.58±0.05	0.343
Weight (kg)	58.90±8.69	59.46±8.91	57.04±7.69	0.072
BMI (kg/m^2)	23.30±3.24	23.46±3.29	22.74±3.01	0.154
Stage				0.241
I	36 (15.25)	24 (13.19)	12 (22.22)	
II	96 (40.68)	79 (43.41)	17 (31.48)	
III	32 (13.56)	23 (12.64)	9 (16.67)	
IV	72 (30.51)	56 (30.77)	16 (29.63)	

The data are represented as means ± standard deviation or number (percentage). BMI, body mass index.

Table 2 Nodule characteristics in the internal and external datasets

Characteristics	Internal dataset (n=249)	External dataset (n=75)	P value
Nodule size			
2D size (cm)	4.59±2.93	4.21±2.59	0.328
3D size (cm)	5.22±3.30	4.94±2.88	0.519
Nodule volume (cm^3)	65.50±195.04	46.92±154.02	0.450
PET parameters			
SUV$_{min}$	0.59 (0.00, 1.83)†	0.60 (0.01, 1.50)†	0.783
SUV$_{mean}$	5.08 (1.33, 9.78)†	5.43 (0.91, 17.72)†	0.498
SUV$_{max}$	12.75 (2.47, 59.86)†	15.49 (1.44, 50.73)†	0.062
MTV	29.21 (0.20, 815.07)†	15.99 (0.45, 446.06)†	0.256
TLG	256.84 (0.28, 8,249.31)†	192.38 (0.80, 7,905.91)†	0.618

†, values refer to mean (range), other data are represented as means ± standard deviation. 2D, 2-dimensional; 3D, 3-dimensional; PET, positron emission tomography; SUV, standard uptake value; MTV, metabolic tumor volume; TLG, total lesion glycolysis.

IDC and DLBCL groups in the internal dataset testing according to the univariate analysis (P<0.001) (*Table 3*). Meanwhile, in the external dataset testing, only PET parameters (SUV$_{min}$, SUV$_{mean}$, and SUV$_{max}$) were statistically significant between IDC and DLBCL (Table S2). The representative 2D and 3D schematic diagrams of breast nodules are shown in *Figure 5*.

AACNN model prediction results

The AACNN model was trained based on CT (AACNN_CT) and PET (AACNN_PET) patches respectively, and the external dataset was adopted to further test the robustness of the model. As shown in *Table 4*, AACNN_CT was tested on the internal testing dataset, the highest AUC was 0.886, the highest prediction ACC was 82.2%, the highest SEN

Table 3 Nodule characteristics between IDC and DLBCL in the internal dataset

Characteristics	IDC (n=135)	DLBCL (n=114)	P value
Nodule size			
2D size (cm)	3.89±1.86	5.41±3.67	<0.001
3D size (cm)	4.34±1.95	6.25±4.17	<0.001
Nodule volume (cm^3)	19.88±41.65	119.53±275.67	<0.001
PET parameters			
SUV$_{min}$	0.59 (0.03, 1.32)†	0.59 (0.01, 1.83)†	0.978
SUV$_{mean}$	3.81 (1.42, 8.89)†	6.59 (1.33, 24.45)†	<0.001
SUV$_{max}$	9.65 (2.47, 35.30)†	16.41 (2.49, 59.86)†	<0.001
MTV	8.89 (0.64, 164.63)†	53.27 (0.20, 815.07)†	<0.001
TLG	38.08 (1.11, 687.35)†	515.91 (0.28, 8,249.31)†	<0.001

†, values refer to mean (range), other data are represented as means ± standard deviation. IDC, invasive ductal carcinoma; DLBCL, diffuse large B-cell lymphoma; 2D, 2-dimensional; 3D, 3-dimensional; PET, positron emission tomography; SUV, standard uptake value; MTV, metabolic tumor volume; TLG, total lesion glycolysis.

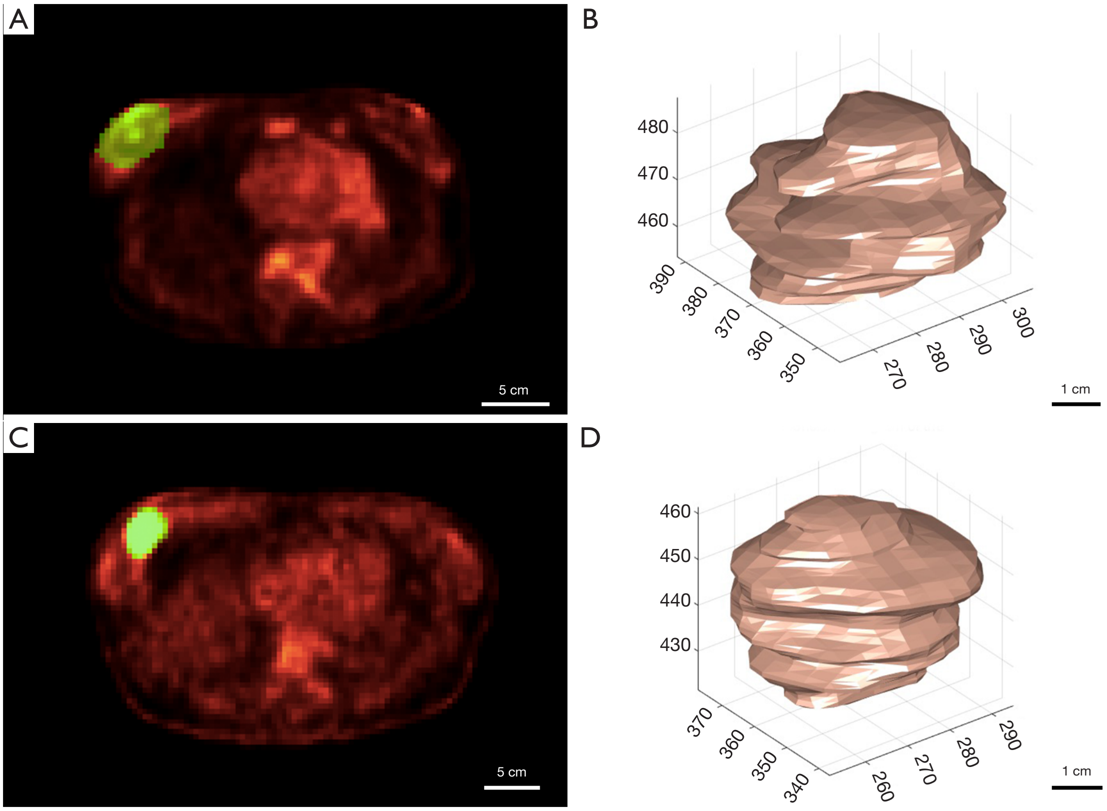

Figure 5 2D and 3D schematic diagram of breast nodules. 2D (A) and 3D (B) schematic diagram of a breast cancer lesion. 2D (C) and 3D (D) schematic diagram of a breast lymphoma lesion. 2D schematic diagram of a breast lymphoma lesion. 2D, 2-dimensional; 3D, 3-dimensional.

Table 4 The prediction performance of AACNN models

Model	AUC (95% CI)	ACC (95% CI)	SEN (95% CI)	SPE (95% CI)	PPV (95% CI)	NPV (95% CI)	F1 (95% CI)
AACNN_CT							
Internal testing	0.886 (0.849–0.882)	0.822 (0.803–0.840)	0.824 (0.797–0.848)	0.820 (0.793–0.845)	0.826 (0.799–0.850)	0.818 (0.791–0.843)	0.825 (0.806–0.842)
External testing	0.784 (0.769–0.798)	0.716 (0.698–0.733)	0.663 (0.634–0.687)	0.785 (0.760–0.808)	0.802 (0.778–0.823)	0.640 (0.615–0.665)	0.726 (0.709–0.742)
AACNN_PET							
Internal testing	0.831 (0.812–0.848)	0.765 (0.744–0.784)	0.781 (0.752–0.808)	0.747 (0.716–0.776)	0.762 (0.733–0.789)	0.768 (0.737–0.796)	0.771 (0.751–0.790)
External testing	0.726 (0.704–0.746)	0.655 (0.637–0.673)	0.575 (0.549–0.599)	0.761 (0.735–0.785)	0.759 (0.733–0.783)	0.577 (0.552–0.602)	0.654 (0.636–0.672)
AACNN_E							
Internal testing	0.886 (0.870–0.900)	0.830 (0.811–0.847)	0.809 (0.782–0.834)	0.850 (0.824–0.873)	0.848 (0.822–0.872)	0.812 (0.784–0.836)	0.828 (0.810–0.846)
External testing	0.788 (0.772–0.803)	0.716 (0.698–0.733)	0.614 (0.589–0.638)	0.847 (0.825–0.867)	0.840 (0.817–0.861)	0.626 (0.602–0.650)	0.709 (0.692–0.726)

AUC, area under the curve; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; F1, harmonic mean of precision and sensitivity; AACNN, attention-based aggregate convolutional neural network; CT, computed tomography; PET, positron emission tomography; E, ensemble; CI, confidence interval.

was 82.4%, the highest SPE was 82.0%, the highest PPV was 82.6%, the highest NPV was 81.8%, and the highest F1 value was 0.825, respectively. The highest predictions of AACNN_CT on the external testing dataset achieved AUC, ACC, SEN, SPE, PPV, NPV, and F1 of 0.784, 71.6%, 66.3%, 78.5%, 80.2%, 64.0%, and 0.726, respectively. When PET patches were used as the input of the AACNN model, the best internal testing performance showed that the AUC, ACC, SEN, SPE, PPV, NPV, and F1 were 0.831, 76.5%, 78.1%, 74.7%, 76.2%, 76.8%, and 0.771, respectively. Similarly, on the external testing dataset, the best AUC, ACC, SEN, SPE, PPV, NPV, and F1 were 0.726, 65.5%, 57.5%, 76.1%, 75.9%, 57.7%, and 0.654, respectively. In general, the prediction performance of AACNN_CT model was better than AACNN_PET.

As shown in *Table 4*, the AACNN_ensemble (AACNN_E) model prediction results (AUC, ACC, SEN, SPE, PPV, NPV, and F1) were 0.886, 83.0%, 80.9%, 85.0%, 84.8%, 81.2%, and 0.828 (internal testing) and 0.788, 71.6%, 61.4%, 84.7%, 84.0%, 62.6%, and 0.709 (external testing), respectively, which was improved in 5 performance metrics (AUC, ACC, SPE, PPV, and F1) compared to all models, and the remaining 2 metrics (SEN, NPV) was also close to the highest (82.4%, 81.8%). The model results of the AACNN_E was a back-end fusion based on distance weight voting between the model prediction probability values of AACNN_CT and AACNN_PET (40), the Eq. [13] for the back-end fusion based on distance weights voting is displayed below.

$$P_{\text{fusion}}=\begin{cases}\max\left(P_{AACNN\_CT},P_{AACNN\_PET}\right), & \text{if } P_{AACNN\_CT}\geq 0.5 \text{ and } P_{AACNN\_PET}\geq 0.5\\ \min\left(P_{AACNN\_CT},P_{AACNN\_PET}\right), & \text{if } P_{AACNN\_CT}<0.5 \text{ and } P_{AACNN\_PET}<0.5\\ \max\left(\left|P_{AACNN\_CT}-0.5\right|,\left|P_{AACNN\_PET}-0.5\right|\right), & \text{otherwise}\end{cases} \quad [13]$$

P_{fusion} represents fusion prediction results, P_{AACNN_CT} represents prediction probability values of AACNN_CT, and P_{AACNN_PET} represents prediction probability values of AACNN_PET.

Further, the AUCs between AACNN_CT and AACNN_PET, AACNN_CT and AACNN_E, AACNN_PET, and AACNN_E were statistically significant ($P<0.05$, DeLong's test). The ROC curves of all models are shown in *Figure 6*.

The proposed AACNN models can extract deep features with different locations, sizes, contours, and texture abstractions in convolutional layers. To more clearly demonstrate the ability of different layers of the model to extract features for lesions, the class activation maps from shallow to deep is represented in *Figure* 7.

Finally, as shown in *Table 5*, the prediction probability values of the 3 models AACNN_CT, AACNN_PET, and AACNN_E were analyzed by univariate analysis and found to be statistically significant ($P<0.001$). The effectiveness of the AACNN models proposed in this paper was further validated statistically.

Discussion

Breast nodules are overwhelmingly seen in women, of which breast IDC is the most common malignancy. Although breast DLBCL is a rare disease, its incidence rate is increasing with the changes of people's living environment and other factors. The ineffectiveness of conventional imaging in distinguishing these 2 diseases leads to unnecessary surgery in patients with breast lymphoma (8). Artificial intelligence technology represented by deep learning is a widely used intelligent analysis technology in medical data or image analysis, which can be used in various image analysis tasks involved in clinical diagnosis, treatment, and prognosis prediction. Deep learning radiomics does not require precise segmentation of tumor regions and pre-definition of high-throughput feature sets based on prior knowledge. Key features relevant to clinical questions can be automatically extracted and optimized through different deep learning network architectures. Automatic feature extraction does not involve human interaction, and the extracted features are the most implicit and advanced. In our study, we proposed an end-to-end AACNN model which can effectively differentiate breast DLBCL from breast IDC, and the radiomic features learned from the model had significant differences between the 2 diseases. Meanwhile, we used the multi-center data for preliminary verification. As most breast DLBCL patients show multiple lesions, most breast IDC patients show single lesion, and the AACNN model we established was to make accurate differentiation of the lesions, 76 breast DLBCL patients (150 lesions) and 160 breast IDC patients (174 lesions) were recruited finally, and the number of lesion samples was balanced at almost 1:1. To the best of our knowledge, this is the first deep learning model to identify these 2 diseases.

In previous studies, researchers usually only considered the 2D diameter of the lesion. Contrastingly, in our study, we added the 3D diameter and volume of the lesion, and the results showed that the spatial data of the lesions had statistical significance in differentiating the 2 diseases in the internal dataset testing, which can provide more comprehensive spatial distribution information for the lesion and may provide more effective information for the selection of treatment and prognosis evaluation of patients (41). Representative 2D and 3D images of the lesion are shown in *Figure 5*. PET parameters (including SUV, MTV, TLG, etc.) derived from PET/CT are closely related to tumor metabolism. Ou *et al.*'s study revealed that the SUV metrics from PET/CT images had potential utility in differentiating breast lymphoma from carcinoma. In our study, the PET parameters (including SUV_{mean}, SUV_{max}, MTV, and TLG) after the above image preprocessing were significant to differentiate the 2 diseases on the internal dataset testing, which was consistent with a previous finding (17). However, only 3 PET parameters (SUV_{min}, SUV_{mean}, and SUV_{max}) showed significance in the external dataset testing; the other parameters showed no statistical differences. Firstly, the reason may be due to the small sample size with limited diversity included in the external testing. Secondly, the clinical parameters may not be specific indicators of breast disease, which may lead to difficulty in discriminating between breast IDC and DLBCL. Similarly, the sensitivity of the external dataset for our model was low, with 0.663 for AACNN_CT, 0.575 for AACNN_PET, and 0.614 for AACNN_E, which further verified our conjecture (*Table 4*). Thirdly, because of the instability and non-specificity of these clinical factors, the potential of radiomic analysis in this field is urgently needed.

In order to effectively and noninvasively distinguish breast DLBCL and breast IDC patients, we designed a new AACNN model structure (*Figure 4*). The structure combines the Mish activation function, the improved loss function based on label smoothing, and the convolutional attention mechanism module. We also proposed a back-end voting strategy basing on distance weight voting, and built an ensemble model named AACNN_E. The experimental

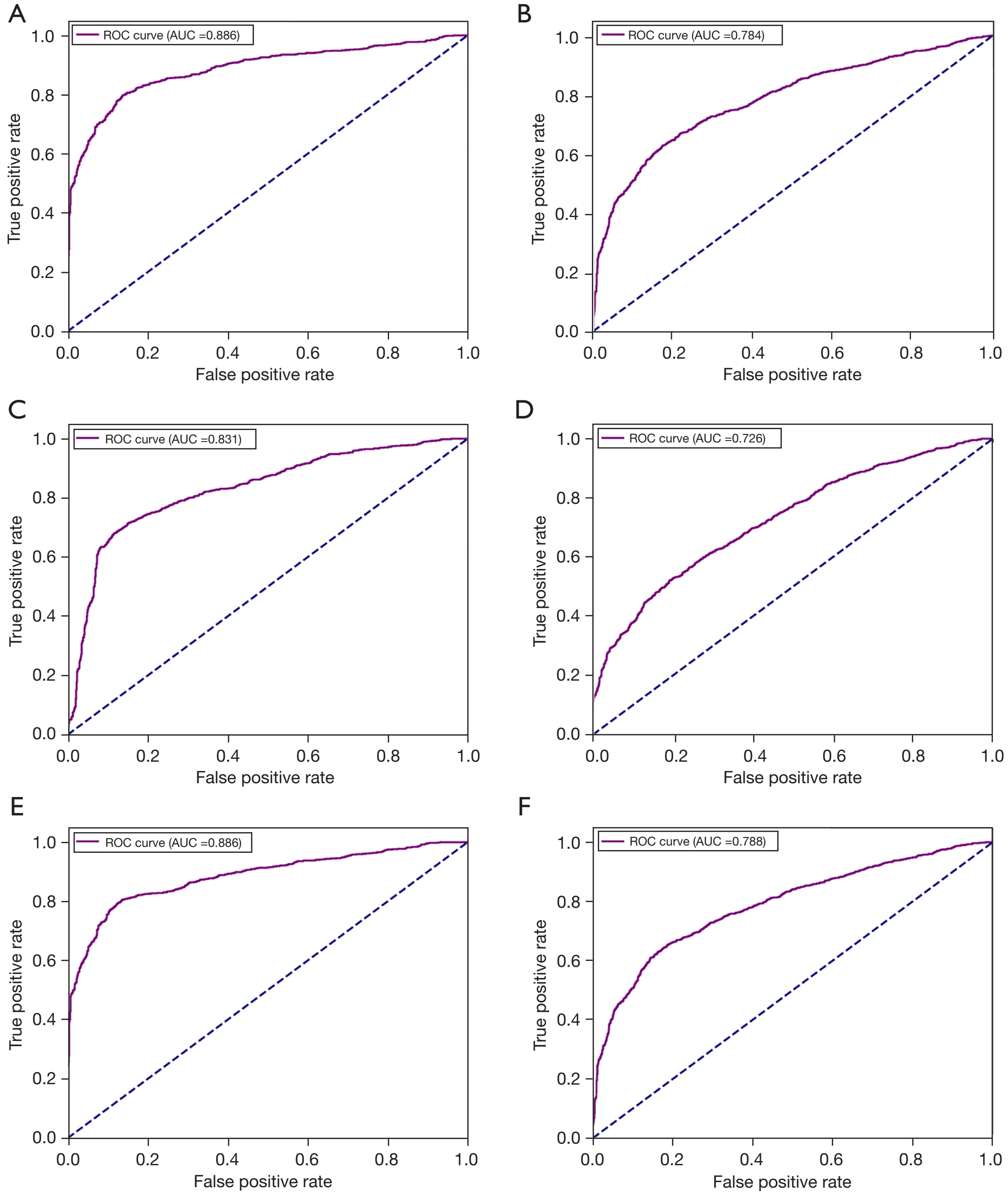

Figure 6 ROC curves of the best AACNN applied to the internal/external testing. ROC curve for internal (A) and external (B) testing on AACNN_CT model. ROC curve for internal (C) and external (D) testing on AACNN_PET model. ROC curve for internal (E) and external (F) testing on AACNN_E model. ROC, receiver operating characteristic; AUC, area under the ROC curve; AACNN, attention-based aggregate convolutional neural network; CT, computed tomography; PET, positron emission tomography; E, ensemble.

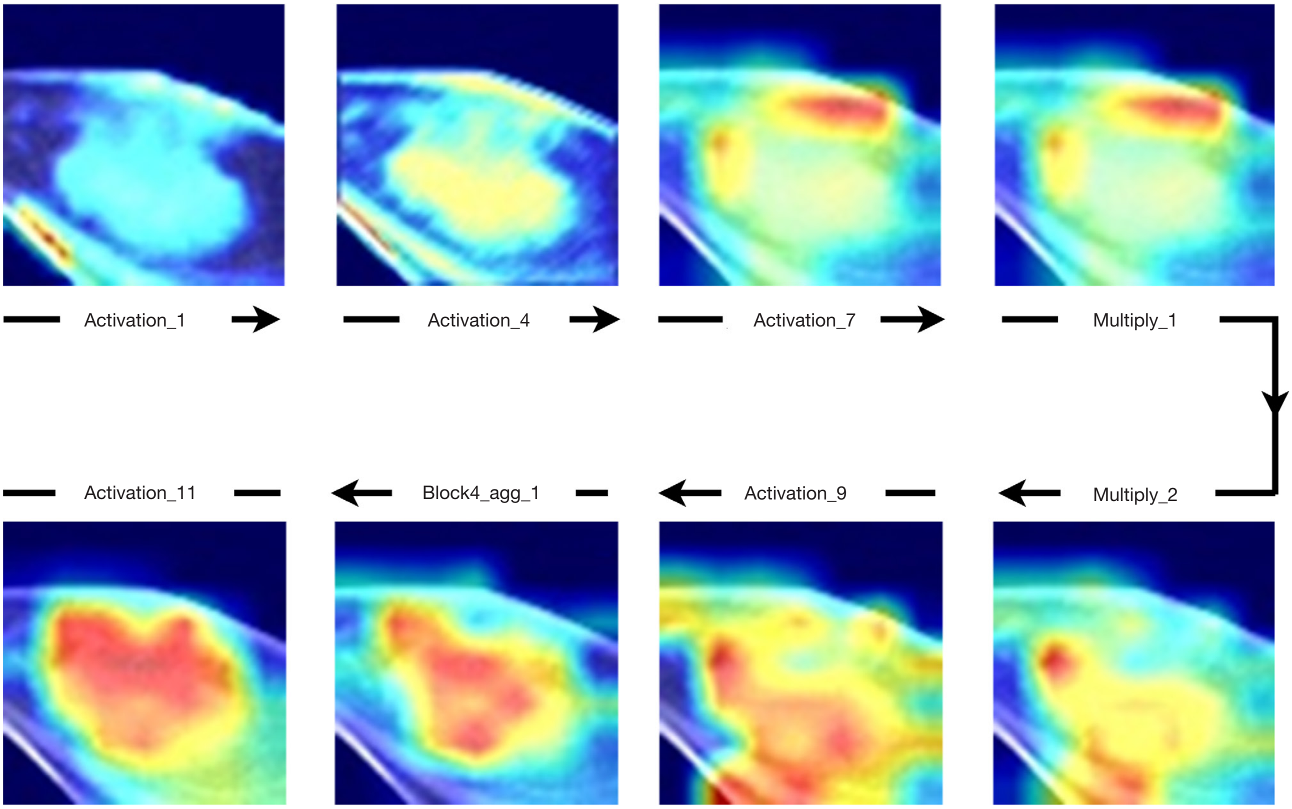

Figure 7 Class activation maps visualization of AACNN layer by layer. AACNN, attention-based aggregate convolutional neural network.

Table 5 The predicted probability values of AACNN models

Model	Internal testing			External testing		
	IDC	DLBCL	P value	IDC	DLBCL	P value
AACNN_CT	0.36±0.16	0.73±0.22	<0.001	0.39±0.17	0.63±0.24	<0.001
AACNN_PET	0.40±0.12	0.56±0.12	<0.001	0.39±0.10	0.49±0.11	<0.001
AACNN_E	0.34±0.16	0.72±0.23	<0.001	0.36±0.17	0.61±0.25	<0.001

The data are represented as means ± standard deviation. IDC, invasive ductal carcinoma; DLBCL, diffuse large B-cell lymphoma; AACNN, attention-based aggregate convolutional neural network; CT, computed tomography; PET, positron emission tomography; E, ensemble.

results (AUC =0.886; ACC =83.0%) indicated that the ensemble model could improve the discrimination ability of breast DLBCL from breast IDC to a certain extent.

In addition, compared to existing studies, data from 3 PET/CT machines of different types in 2 centers with longer time spans were included in our study, which increased the diversity and reliability of the samples (Table S3). Independent AACNN models based on PET and CT were trained and tested by internal cohorts, and independent external cohorts were used to validate the generalization and robustness of the model (*Table 4* and *Figure 6*) (42). Meanwhile, we do not need to manually extract information such as lesion size, dimension, texture, and so on. Our model can automatically learn these features and make predictions directly. Further, we have demonstrated that the prediction result of the AACNN models can be directly used as a significant factor for discriminating breast IDC from DLBCL. The AUC of our model was improved by nearly 5% compared to existing models.

Last but not least, the class activation map visualization (from shallow to deep) (*Figure 7*) realized the interpretability of AACNN prediction results, revealed a corner of the mystery of the "black box" problem in deep learning, and further reflected the effectiveness in extracting disease radiomics features of the AACNN model proposed in this paper. Class activation map visualization refers to the generation of a heat map of class activations on the input image, indicating the importance of each position to the category. It helps to understand which part of a picture causes the AACNN to make the final decision. We have presented the class activation map of the largest tumor area patch of a patient with breast lymphoma. The brighter (redder) pixel regions in the maps indicate the greater contribution of the classification. On the contrary, the bluer and darker pixel regions account for a smaller proportion of the contribution to the classification. It is worth noting that the obtained class activation maps are consistent with the lesion regions that clinicians pay attention to, indicating that the AACNN classification model trained in this paper can well locate image features with clinical diagnostic value.

Our study still had some limitations. Our study was based on a relatively small sample of breast DLBCL, and participants with incomplete clinical data were omitted, which may affect the accuracy and reliability of the research results. Although the data enhancement method was used to alleviate the problem of insufficient sample diversity, there are still uncertainties and inauthenticity compared with the real data. Combined with the actual clinical application and promotion, PET/CT fusion images may have better effects as the input of the model. This study did not combine deep learning with traditional radiomics methods, and the method of fusing deep features and traditional omics features may further improve the performance of the current AACNN model.

Conclusions

This study developed a deep learning radiomics strategy based on AACNN model and ensemble prediction which may automatically achieve remarkable prediction performance at the classification task of distinguishing breast DLBCL from breast IDC. The proposed model relied on data from 2 centers and was validated through an external dataset, demonstrating that deep features learned through convolutional layers could distinguish between the 2 diseases. The experimental results showed that our proposed method could learn more lesion information, significantly improving the ability to differentiate between breast IDC and breast DLBCL. Our proposed model is expected to become a non-invasive auxiliary tool for precision diagnosis in the future.

Acknowledgments

Funding: This study was supported by the National Natural Science Foundation of China (Nos. 82102097 and 81971648) and the Talent Introduction Fund of Fudan University Shanghai Cancer Center (No. YJRC202104).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-22-1333/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-22-1333/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committees of FUSCC (No. 1909207-14-1910) and SPHIC (No. 200217EXP-01) and the requirement for individual consent for this retrospective analysis was waived.

References

1. Zheng R, Zhang S, Zeng H, Wang S, Sun K, Chen R, Li L, Wei W, He J. Cancer incidence and mortality in China, 2016. J Natl Cancer Cent 2022;2:1-9.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
3. Rodrigues-Ferreira S, Nahmias C. Predictive biomarkers for personalized medicine in breast cancer. Cancer Lett 2022;545:215828.
4. Matasar MJ, Zelenetz AD. Overview of lymphoma diagnosis and management. Radiol Clin North Am 2008;46:175-98, vii.
5. Gurney KA, Cartwright RA. Increasing incidence and descriptive epidemiology of extranodal non-Hodgkin lymphoma in parts of England and Wales. Hematol J 2002;3:95-104.
6. Sousaris N, Barr RG. Sonoelastography of Breast Lymphoma. Ultrasound Q 2016;32:208-11.
7. Surov A, Holzhausen HJ, Wienke A, Schmidt J, Thomssen C, Arnold D, Ruschke K, Spielmann RP. Primary and secondary breast lymphoma: prevalence, clinical signs and radiological features. Br J Radiol 2012;85:e195-205.
8. Picasso R, Tagliafico A, Calabrese M, Martinoli C, Pistoia F, Rossi A, Zaottini F, Derchi L. Primary and Secondary Breast Lymphoma: Focus on Epidemiology and Imaging Features. Pathol Oncol Res 2020;26:1483-8.
9. Nicholson BT, Bhatti RM, Glassman L. Extranodal Lymphoma of the Breast. Radiol Clin North Am 2016;54:711-26.
10. Breese RO, Friend K. Thirteen-Centimeter Breast Lymphoma. Am Surg 2022;88:1891-2.
11. Gluskin J, D'Alessio D, Kim AC, Morris EA, Chiu A, Noy A. Primary lymphoma of the breast: A report of two cases. Clin Imaging 2020;68:295-9.
12. Cheson BD. Staging and response assessment in lymphomas: the new Lugano classification. Chin Clin Oncol 2015;4:5.
13. Meignan M, Itti E, Gallamini A, Younes A. FDG PET/CT imaging as a biomarker in lymphoma. Eur J Nucl Med Mol Imaging 2015;42:623-33.
14. Thanarajasingam G, Bennani-Baiti N, Thompson CA. PET-CT in Staging, Response Evaluation, and Surveillance of Lymphoma. Curr Treat Options Oncol 2016;17:24.
15. Eertink JJ, Burggraaff CN, Heymans MW, Dührsen U, Hüttmann A, Schmitz C, et al. Optimal timing and criteria of interim PET in DLBCL: a comparative study of 1692 patients. Blood Adv 2021;5:2375-84.
16. Lee JW, Lee SM. Radiomics in Oncological PET/CT: Clinical Applications. Nucl Med Mol Imaging 2018;52:170-89.
17. Ou X, Wang J, Zhou R, Zhu S, Pang F, Zhou Y, Tian R, Ma X. Ability of (18)F-FDG PET/CT Radiomic Features to Distinguish Breast Carcinoma from Breast Lymphoma. Contrast Media Mol Imaging 2019;2019:4507694.
18. Ou X, Zhang J, Wang J, Pang F, Wang Y, Wei X, Ma X. Radiomics based on (18) F-FDG PET/CT could differentiate breast carcinoma from breast lymphoma using machine-learning approach: A preliminary study. Cancer Med 2020;9:496-506.
19. Jarrett D, Stride E, Vallis K, Gooding MJ. Applications and limitations of machine learning in radiation oncology. Br J Radiol 2019;92:20190001.
20. Le NQK, Kha QH, Nguyen VH, Chen YC, Cheng SJ, Chen CY. Machine Learning-Based Radiomics Signatures for EGFR and KRAS Mutations Prediction in Non-Small-Cell Lung Cancer. Int J Mol Sci 2021;22:9254.
21. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, Tridandapani S, Auffermann WF. Deep Learning in Radiology. Acad Radiol 2018;25:1472-80.
22. Saba L, Biswas M, Kuppili V, Cuadrado Godia E, Suri HS, Edla DR, Omerzu T, Laird JR, Khanna NN, Mavrogeni S, Protogerou A, Sfikakis PP, Viswanathan V, Kitas GD, Nicolaides A, Gupta A, Suri JS. The present and future of deep learning in radiology. Eur J Radiol 2019;114:14-24.
23. Chen W, Hou X, Hu Y, Huang G, Ye X, Nie S. A deep learning- and CT image-based prognostic model for the prediction of survival in non-small cell lung cancer. Med Phys 2021;48:7946-58.
24. Lu Z, Xu S, Shao W, Wu Y, Zhang J, Han Z, Feng Q, Huang K. Deep-Learning-Based Characterization of Tumor-Infiltrating Lymphocytes in Breast Cancers From Histopathology Images and Multiomics Data. JCO Clin Cancer Inform 2020;4:480-90.
25. Mohanty SP, Hughes DP, Salathé M. Using Deep Learning for Image-Based Plant Disease Detection. Front Plant Sci 2016;7:1419.
26. Gong J, Liu J, Li H, Zhu H, Wang T, Hu T, Li M, Xia X, Hu X, Peng W, Wang S, Tong T, Gu Y. Deep Learning-Based Stage-Wise Risk Stratification for Early Lung Adenocarcinoma in CT Images: A Multi-Center Study. Cancers (Basel) 2021.
27. Liu Q, Sun D, Li N, Kim J, Feng D, Huang G, Wang L, Song S. Predicting EGFR mutation subtypes in lung adenocarcinoma using (18)F-FDG PET/CT radiomic features. Transl Lung Cancer Res 2020;9:549-62.
28. Eertink JJ, Brug T, Wiegers SE, Zwezerijnen G, Zijlstra

JM. 18F-FDG PET/CT baseline rdiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma patients. Blood 2020;136:27-8.

29. Zhang J, Xia Y, Zeng H, Zhang Y. NODULe: Combining constrained multi-scale LoG filters with densely dilated 3D deep convolutional neural network for pulmonary nodule detection. Neurocomputing 2018;317:159-67.
30. He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016 June 27-30; Las Vegas, NV, USA. Piscataway: Proceedings of the IEEE; 2016:770-8.
31. Lin TY, Dollar P, Girshick R, He KM, Hariharan B, Belongie S. Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017 July 21-26; Honolulu, HI, USA. Piscataway: Proceedings of the IEEE; 2017:936-44.
32. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR), 2015 May 7-9; San Diego, CA. Boston: OpenReview.net; 2015:1-14.
33. Luo P, Zhang R, Ren J, Peng Z, Li J. Switchable Normalization for Learning-to-Normalize Deep Representation. IEEE Trans Pattern Anal Mach Intell 2021;43:712-28.
34. Misra D. Mish: A self-regularized non-monotonic neural activation function. arXiv 2020, August:1-13. doi: 10.48550/arXiv.1908.08681.
35. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 June 27-30; Las Vegas, NV, USA. Piscataway: Proceedings of the IEEE; 2016:2818-26.
36. Wei W, Wong Y, Du Y, Hu Y, Kankanhalli M, Geng W. A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface. Pattern Recognit Lett 2019;109:131-8.
37. Christopher AR, Timothy AW, Aaron EM. Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. Remote Sensing 2019;11:185.
38. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929-58.
39. Müller R, Kornblith S, Hinton G. When does label smoothing help? NeurIPS 2019;422:4694-703.
40. Liao, T, Lei Z, Zhu T, Zeng S, Li Y, Yuan C. Deep Metric Learning for K Nearest Neighbor Classification. IEEE Trans Knowl Data Eng 2023;35:264-75.
41. Wang X, Bao N, Xin X, Tan J, Li H, Zhou S, Liu H. Automatic evaluation of endometrial receptivity in three-dimensional transvaginal ultrasound images based on 3D U-Net segmentation. Quant Imaging Med Surg 2022;12:4095-108.
42. Lam NFD, Sun H, Song L, Yang D, Zhi S, Ren G, Chou PH, Wan SBN, Wong MFE, Chan KK, Tsang HCH, Kong FS, Wáng YXJ, Qin J, Chan LWC, Ying M, Cai J. Development and validation of bone-suppressed deep learning classification of COVID-19 presentation in chest radiographs. Quant Imaging Med Surg 2022;12:3917-31.

Cite this article as: Chen W, Liu F, Wang R, Qi M, Zhang J, Liu X, Song S. End-to-end deep learning radiomics: development and validation of a novel attention-based aggregate convolutional neural network to distinguish breast diffuse large B-cell lymphoma from breast invasive ductal carcinoma. Quant Imaging Med Surg 2023;13(10):6598-6614. doi: 10.21037/qims-22-1333

Supplementary

Table S1 Patient characteristics between IDC and DLBCL

Characteristics	IDC (n=160)	DLBCL (n=76)	P value
Sex			0.146
Female	160 (100.00)	75 (98.68)	
Male	0 (0.00)	1 (1.32)	
Age (year)	50.66±11.51	52.70±12.82	0.221
Height (m)	1.59±0.05	1.59±0.04	0.859
Weight (kg)	58.55±8.41	59.64±9.27	0.384
BMI (kg/m^2)	23.16±3.04	23.58±3.62	0.391
Stage			<0.001
I	12 (7.50)	24 (31.58)	
II	66 (41.25)	30 (39.47)	
III	31 (19.38)	1 (1.32)	
IV	51 (31.88)	21 (27.63)	

The data are represented as means ± standard deviation or number (percentage). IDC, invasive ductal carcinoma; DLBCL, diffuse large B-cell lymphoma; BMI, body mass index.

Table S2 Nodule characteristics between breast carcinoma and lymphoma in the external dataset

Characteristics	IDC (n=39)	DLBCL (n=36)	P value
Nodule size			
2D size (cm)	3.89±1.48	4.57±3.40	0.279
3D size (cm)	4.57±1.66	5.35±3.78	0.255
Nodule volume (cm^3)	15.40±17.60	81.06±217.97	0.080
PET parameters			
SUV$_{min}$	0.50 (0.03, 1.27)[†]	0.71 (0.01, 1.50)[†]	0.021
SUV$_{mean}$	3.57 (0.99, 7.88)[†]	7.44 (0.91, 17.72)[†]	<0.001
SUV$_{max}$	10.65 (1.44, 24.80)[†]	20.73 (1.65, 50.73)[†]	<0.001
MTV	5.45 (0.45, 40.84)[†]	27.40 (0.53, 446.06)[†]	0.108
TLG	22.50 (0.80, 175.67)[†]	376.41 (0.96, 7,905.91)[†]	0.127

[†], values refer to mean (range), other data are represented as means ± standard deviation. IDC, invasive ductal carcinoma; DLBCL, diffuse large B-cell lymphoma; 2D, 2-dimensional; 3D, 3-dimensional; SUV, standard uptake value; MTV, metabolic tumor volume; TLG, total lesion glycolysis.

https://dx.doi.org/10.21037/qims-22-1333

Table S3 Comparison of AACNN_E model's performance with existing models

Models	Patients [nodules]	Datasets numbers	External dataset testing	Model performance evaluation metrics						
				AUC	ACC	SEN	SPE	PPV	NPV	F1
Ou *et al.* (17)	44 [67]	1	No	0.845	0.762	0.861	0.556	–	–	–
Ou *et al.* (18)	44 [65]	1	No	0.806	0.808	0.806	0.842	–	–	–
AACNN_E	324 [236]	2	Yes	0.886	0.830	0.809	0.850	0.848	0.812	0.828

AUC, area under the curve; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; F1, harmonic mean of precision and sensitivity; AACNN, attention-based aggregate convolutional neural network ensemble; –, not mentioned.

https://dx.doi.org/10.21037/qims-22-1333