



# Fully automated classification of pulmonary nodules in positron emission tomography-computed tomography imaging using a two-stage multimodal learning approach

Tongtong Li<sup>1,2#</sup>, Junfeng Mao<sup>3,4#</sup>, Jiandong Yu<sup>1,2#</sup>, Ziyang Zhao<sup>1,2</sup>, Miao Chen<sup>1,2</sup>, Zhijun Yao<sup>1,2</sup>, Lei Fang<sup>5</sup>, Bin Hu<sup>1,2,6,7,8</sup>

<sup>1</sup>School of Information Science and Engineering, Lanzhou University, Lanzhou, China; <sup>2</sup>Gansu Provincial Key Laboratory of Wearable Computing, Lanzhou University, Lanzhou, China; <sup>3</sup>Department of Nuclear Medicine, The 940th Hospital of Joint Logistics Support Force of Chinese People's Liberation Army, Lanzhou, China; <sup>4</sup>School of Basic Medical Sciences, Gansu University of Traditional Chinese Medicine, Lanzhou, China; <sup>5</sup>Department of Nuclear Medicine, Taikang Tongji (Wuhan) Hospital, Wuhan, China; <sup>6</sup>School of Medical Technology, Beijing Institute of Technology, Beijing, China; <sup>7</sup>CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; <sup>8</sup>Joint Research Center for Cognitive Neurosensor Technology of Lanzhou University & Institute of Semiconductors, Chinese Academy of Sciences, Lanzhou, China

**Contributions:** (I) Conception and design: T Li, J Mao, J Yu; (II) Administrative support: Z Yao, B Hu; (III) Provision of study materials or patients: J Mao, L Fang; (IV) Collection and assembly of data: J Mao, L Fang; (V) Data analysis and interpretation: T Li, J Mao, J Yu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

**Correspondence to:** Zhijun Yao, CS, PhD. School of Information Science and Engineering, Lanzhou University, No. 222 South Tianshui Road, Lanzhou, China; Gansu Provincial Key Laboratory of Wearable Computing, Lanzhou University, Lanzhou 730000, China. Email: yaozj@lzu.edu.cn; Lei Fang, MD, PhD. Department of Nuclear Medicine, Taikang Tongji (Wuhan) Hospital, No. 322 Sixin North Road, Hanyang District, Wuhan 430050, China. Email: 116fanglei@163.com; Bin Hu, CS, PhD. School of Information Science and Engineering, Lanzhou University, No. 222 South Tianshui Road, Lanzhou 730000, China; Gansu Provincial Key Laboratory of Wearable Computing, Lanzhou University, Lanzhou 730000, China; School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China; CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; Joint Research Center for Cognitive Neurosensor Technology of Lanzhou University & Institute of Semiconductors, Chinese Academy of Sciences, Lanzhou 730000, China. Email: bh@lzu.edu.cn.

**Background:** Lung cancer is a malignant tumor, for which pulmonary nodules are considered to be significant indicators. Early recognition and timely treatment of pulmonary nodules can contribute to improving the survival rate of patients with cancer. Positron emission tomography-computed tomography (PET/CT) is a noninvasive, fusion imaging technique that can obtain both functional and structural information of lung regions. However, studies of pulmonary nodules based on computer-aided diagnosis have primarily focused on the nodule level due to a reliance on the annotation of nodules, which is superficial and unable to contribute to the actual clinical diagnosis. The aim of this study was thus to develop a fully automated classification framework for a more comprehensive assessment of pulmonary nodules in PET/CT imaging data.

**Methods:** We developed a two-stage multimodal learning framework for the diagnosis of pulmonary nodules in PET/CT imaging. In this framework, Stage I focuses on pulmonary parenchyma segmentation using a pretrained U-Net and PET/CT registration. Stage II aims to extract, integrate, and recognize image-level and feature-level features by employing the three-dimensional (3D) Inception-residual net (ResNet) convolutional block attention module architecture and a dense-voting fusion mechanism.

**Results:** In the experiments, the proposed model's performance was comprehensively validated using a set

of real clinical data, achieving mean scores of 89.98%, 89.21%, 84.75%, 93.38%, 86.83%, and 0.9227 for accuracy, precision, recall, specificity, F1 score, and area under curve values, respectively.

**Conclusions:** This paper presents a two-stage multimodal learning approach for the automatic diagnosis of pulmonary nodules. The findings reveal that the main reason for limiting model performance is the nonsolitary property of nodules in pulmonary nodule diagnosis, providing direction for future research.

**Keywords:** Pulmonary nodule classification; multimodal; positron emission tomography-computed tomography (PET/CT); two-stage; deep learning

Submitted Feb 03, 2024. Accepted for publication Jun 17, 2024. Published online Jul 22, 2024.

doi: 10.21037/qims-24-234

View this article at: <https://dx.doi.org/10.21037/qims-24-234>

## Introduction

Lung cancer is a life-threatening malignant tumor with high mortality rate, accounting for 11.6% of total cancer cases and 18.4% of total cancer deaths according to global cancer statistics 2018 (1). Pulmonary nodules are considered to be significant indicators of primary lung cancer (2), and it has been demonstrated that the early detection and timely treatment of pulmonary nodules can significantly improve the 5-year survival rate of patients (3).

In recent years, the rapid advancement of medical imaging technology has enabled the realization of noninvasive imaging of the pulmonary region. For instance, computed tomography (CT) (4) is a structural imaging technology and is usually used to obtain detailed anatomical information of organs and tissues through the transmission and absorption of X-rays. Meanwhile, positron emission tomography (PET) (5) is a nuclear medical functional imaging technique that is commonly used to obtain information on activity, metabolism, and function by detecting the decay of positron emissions from a radioactive tracer in obtaining images. The metabolic and anatomical information of pulmonary region can be simultaneously obtained using PET and CT (PET/CT) imaging (6). However, traditional diagnostic methods for pulmonary nodules rely heavily on manual slice-by-slice screening by physicians, as pulmonary nodules typically exhibit various shapes and multiscale characteristics, resulting in significant healthcare burdens in practice (7).

Artificial intelligence technology, especially deep learning, has generated new possibilities in the application of smart healthcare. Recent studies indicate significant progress in the deep learning-based diagnosis of lung nodules (8,9). Shao *et al.* (10) proposed dual-stream three-dimensional (3D) convolutional neural network to

distinguish benign and invasive adenocarcinoma nodules based on 18F-fluorodeoxyglucose (<sup>18</sup>F-FDG) PET/CT. Apostolopoulos *et al.* (11) proposed a transfer learning-based VGG-16 to classify solitary pulmonary nodules (SPNs) in PET/CT imaging with 94% accuracy (Acc). Liu *et al.* (12) proposed a 3D multimodal ensemble learning architecture (i.e., multiscale ensemble model) that can be well adapted to the heterogeneity of SPNs in CT imaging for diagnosing benign and malignant of lung nodules. Furthermore, several studies (13-15) have designed vision transformer-based deep learning models to recognize benign and malignant pulmonary nodules.

Although the above studies have achieved acceptable Acc, the research related to the diagnosis of lung nodules has mainly focused on the nodule level due to the diversity and complexity of pulmonary nodules, with a heavy reliance on the manual screening by physicians. Under these circumstances, the means to establishing a fully automated framework from PET/CT imaging for pulmonary nodule identification has become a particularly intense area of interest.

In this study, we developed a novel fully automated classification framework for the diagnosis of pulmonary nodules in PET/CT imaging using a two-stage multimodal learning approach. Specifically, we first employed pretrained U-Net and PET/CT registration to extract the region of interest (ROI; i.e., segmentation of pulmonary parenchyma region from PET/CT imaging), referred to as *Stage I ROI segmentation*. We then used a 3D Inception-residual net (ResNet) convolutional block attention module (CBAM) and a dense-voting mechanism to extract, integrate, and classify multimodal features for pulmonary nodule diagnosis, which we referred to as *Stage II nodule classification*. The main contribution of this paper can be

**Table 1** The demographic characteristics of the study participants

Characteristic	Benign	Malignant
Number	305	194
Gender		
Male	187	116
Female	118	78
Age (years)		
Range	24–84	20–91
Mean	60.931	61.2474
Median	63	62

summarized as follows:

- ❖ We propose a novel two-stage paradigm for fully automated identification of pulmonary nodules.
- ❖ We design a feature fusion strategy by integrating image-level, feature-level, and score-level information.
- ❖ The proposed model achieved state-of-the-art (SOTA) performance compared with classical models.
- ❖ Our findings highlight the critical role of solitary nodule detection in the diagnosis of pulmonary nodules.

The rest of this paper is structured as follows: Section “Methods” reports the details of the experimental data, data preprocessing, and the proposed model architecture, respectively. Section “Results” presents the experimental results and analysis. Section “Discussion” briefly discusses the experimental findings, misclassification analysis, limitations and future directions. Finally, we conclude this study in Section “Conclusions”. We present this article in accordance with the TRIPOD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-234/rc>).

## Methods

### Dataset

The PET/CT imaging data were collected from June 2015 to July 2020 from The 940th Hospital of Joint Logistics Support Force of Chinese People’s Liberation Army. A total of 1893 participants underwent PET/CT scans with the Biograph True Point 64 (Siemens

Healthineers, Erlangen, Germany). Of note, in the actual data collection, PET and CT images were acquired from a dedicated PET/CT scanner, where both PET and CT were completed simultaneously with the following parameters: tube voltage, 120 kV; tube current, 21–318 mAs; volume CT dose index, 1.43–21.44 mGy; layer thickness, 3 mm; layer spacing, 0.8 mm, and contrast agent, ioversol injection. Moreover, patients were instructed to fast for a minimum of 6 hours and with a glucose level lower than 11.1 mmol/L before intravenous administration of  $^{18}\text{F}$ -FDG (18F-fluorodeoxyglucose). PET/CT images were collected after an activity of 3.7–7.4 mBq/kg of  $^{18}\text{F}$ -FDG was injected. Meanwhile, patients remained in a supine resting position for a duration of  $60 \pm 5$  minutes. The image resolution of CT and PET were  $512 \times 512$  pixels at  $0.9766 \text{ mm} \times 0.9766 \text{ mm}$  (x–y axes) and  $128 \times 128$  pixels at  $4.0728 \text{ mm} \times 4.0728 \text{ mm}$  (x–y axes), respectively. The CT and PET images had the same resolution of 1 mm in the z-direction.

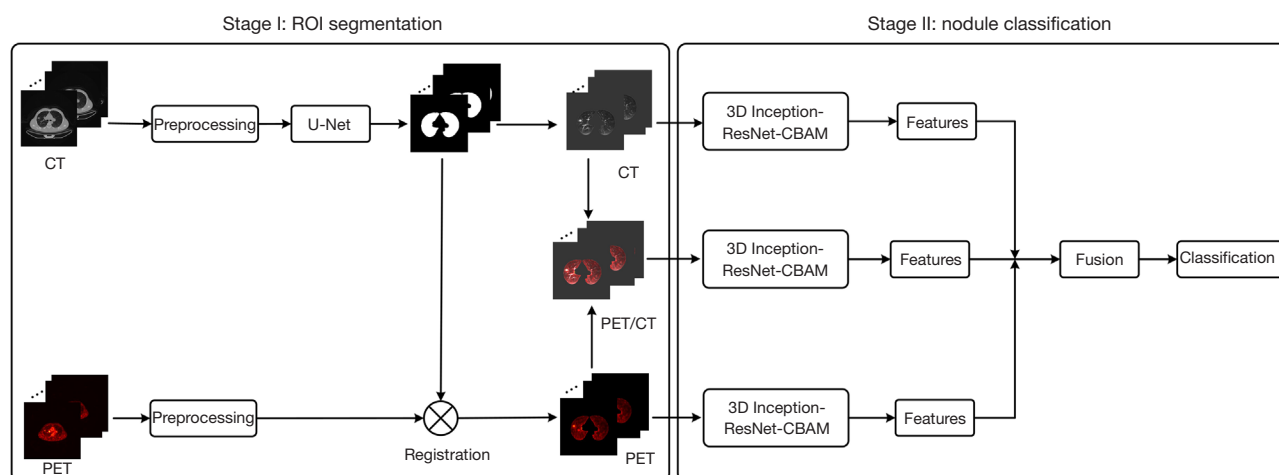
The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of The 940th Hospital of Joint Logistics Support Force of Chinese People’s Liberation Army (No. 2014-06). Informed consent was obtained from all individual participants.

In this study, the inclusion criteria were as follows: (I)  $\geq 18$  years old, (II) completion of preoperative PET/CT scans and follow-up via biopsy within 1 month and subsequent follow-up within 1 year postoperatively (pathological confirmation), (III) a nodule diameter  $\geq 10$  and  $\leq 30$  mm, and (IV) no history of previous surgery or chemotherapy. Finally, 499 participants were included for further study, and their demographic characteristics are reported in *Table 1*.

### Data preprocessing

The objective of data preprocessing is to improve the adaptability between the data and the model. The preprocessing steps for CT images are as follows: (I) the pixel values of CT image were converted to Hounsfield units (HU); (II) voxel dimensions were resampled to  $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$  using trilinear interpolation; (III) a fixed size of  $350 \times 350$  was achieved via cropping, with the central region being preserved; and (IV) the window width and window level were adjusted to 1,500 and –400, respectively.

For the FDG-PET images, they were first resampled to  $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$  using trilinear interpolation and then were cropped to the size of  $350 \times 350$ , with the central region



**Figure 1** An overview of the proposed two-stage classification framework for accurately recognizing pulmonary nodules in PET/CT images. ROI, region of interest; CT, computed tomography; PET, positron emission tomography; 3D, three-dimensional; CBAM, convolutional block attention module.

being preserved. Finally, PET images were registered to their corresponding CT images, allowing for the analysis of PET/CT images from an anatomical-structural and functional-metabolic perspective.

Of note, during the data preprocessing phase, data annotations were obtained by three radiologists (medical practice >10 years) from our collaborating institutions in response to our request. Specifically, a nodule was classified as malignant only if deemed so by a consensus of at least two radiologists. Nodules with uncertain classifications were excluded from the analysis in this study. Although this process was time-consuming, it was necessary, and a valuable ground truth helps to improve the performance and reliability of the model (16). Finally, we selected 10 adjacent slices of CT for each nodule as input data for the segmentation model.

### Two-stage classification framework

In this section, we described our proposed a novel fully automated two-stage multimodal learning approach for the automatic diagnosis of pulmonary nodules in PET/CT imaging, as shown in *Figure 1*, which includes (I) Stage I ROI segmentation and (II) Stage II nodule classification.

#### Stage I ROI segmentation

In order to remove the irrelevant information in PET/CT images, such as stents, inner wall of chest cavity, and other information outside the lung parenchyma, we extracted the

pulmonary parenchyma region from the selected 10 adjacent slices as the ROI. Specifically, the lung parenchyma region masks of CT images were extracted slice by slice using the U-Net classical medical image segmentation network (17) in the ROI segmentation stage. In addition, we invited the radiologists from our collaborating institutions to evaluate the effectiveness of the segmented lung parenchyma images, and the segmented masks for data with poor quality were manually calibrated by the radiologists.

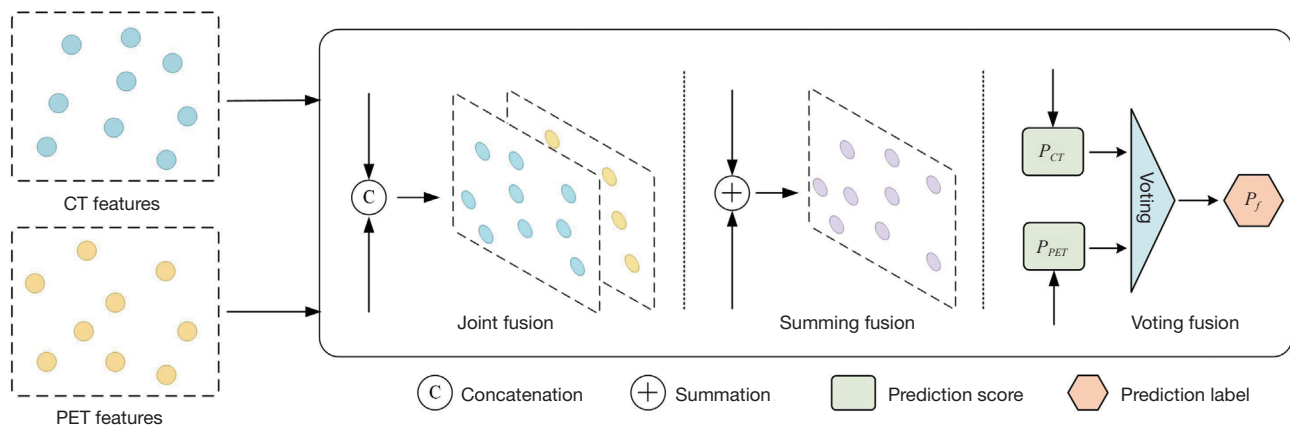
Notably, the pre-trained U-Net was used to create masks for CT data. Hofmanninger *et al.* (18) developed a readily available tool for the segmentation of pathological lungs that allows for the direct acquisition of lung parenchymal masks. The segmentation model parameter settings are available online (<https://github.com/JoHof/lungmask>).

#### Stage II nodule classification

The nodule classification stage involves automatic feature extraction, feature classification, and feature fusion.

##### Feature extraction

Deep learning models have shown strong competitiveness compared to traditional machine learning models in several computer vision tasks by virtue of their ability to automatically extract higher-order features. In our study, three parallel inputs were fed into the 3D Inception-ResNet CBAM module: CT images of lung parenchyma, PET images of lung parenchyma, and PET/CT fusion images of lung parenchyma. Of note, the PET/CT fusion images were obtained through joint fusion or summing fusion with



**Figure 2** Overview of the fusion strategies. CT, computed tomography; PET, positron emission tomography.

the dimensions of channel, dim\_x, dim\_y, and dim\_z being 2, 350, 350, and 10, respectively, or 1, 350, 350, and 10, respectively (see the Feature fusion section and Figure 2). As presented in Figures 3,4, the segmented lung parenchyma was first fed into stem module to extract low-order features. Subsequently, the stacked 3D Inception-ResNet module (19), CBAM (20), and 3D reduction module were used to automatically extract higher-order features. It is worth noting that the architecture of 3D Inception-ResNet module not only improves the model's learning capability by increasing the “width” and “depth” of the feature extractor but also preserves the interlayer information via 3D convolution.

### Feature fusion

As shown in Figure 2, three fusion strategies, including joint fusion, summing fusion, and voting fusion, were used for feature integration. Joint fusion is a feature-level integration strategy achieved by concatenating various feature maps along channel dimensions. Summing fusion is an image-level integration strategy achieved through image registration and pixel-wise summation, which highlights lesion information to some extent. In the voting fusion approach, the highest probability sum with each class from all voting results is output as the prediction after the probability scores of each input are fed into the soft-voting module and summed together, which is also known as *dense-voting fusion*.

### Feature classification

Figure 5 illustrates the architecture of feature classification method. The extracted high-order features were identified using the stacked global average pooling (GAP), full connection (FC), and dropout layer. The predicted class information was mapped to range (0, 1) using the softmax

activation function (21). Of note, the FC layer allows for the combination of features from different regions, while the GAP and dropout layers help to prevent overfitting and improve generalization.

The output results of the Stage II represent the final prediction for the classification of pulmonary nodules as benign or malignant.

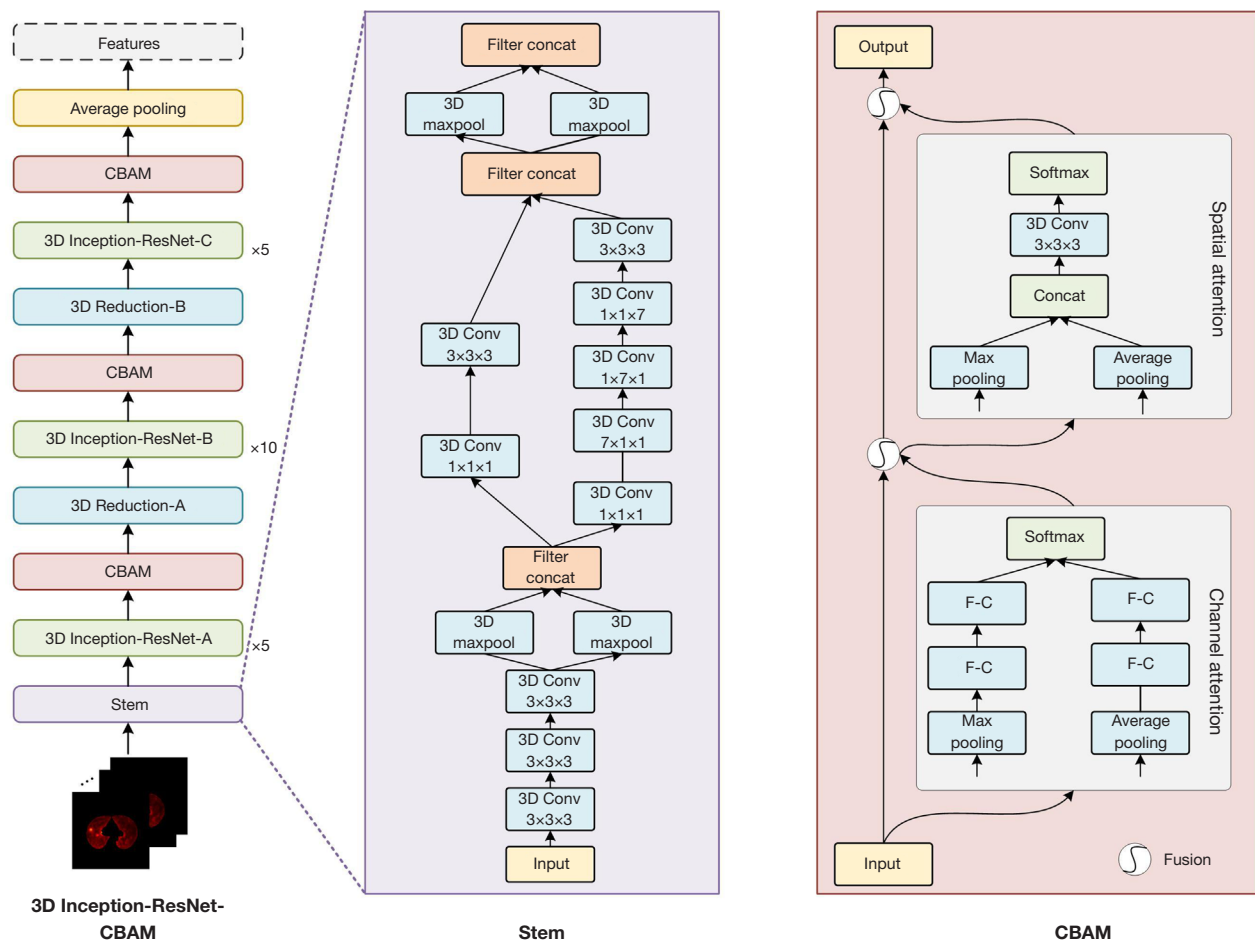
## Results

### Experimental setup

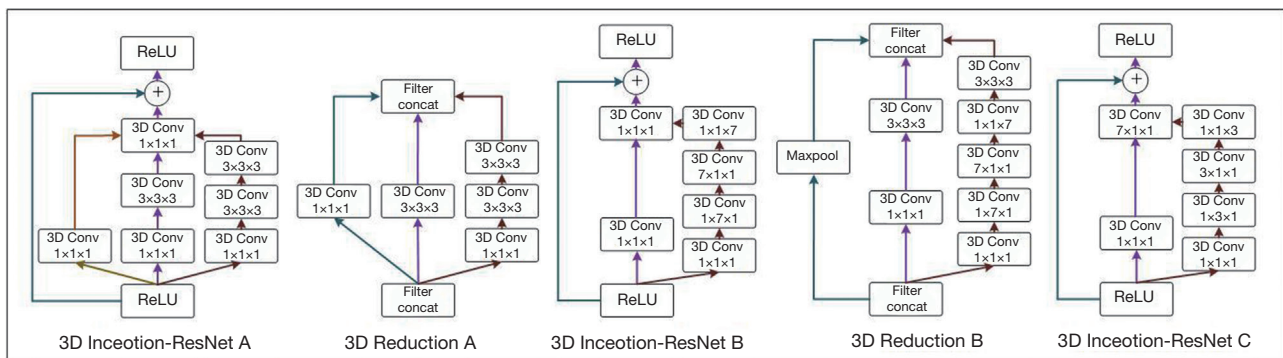
The reported experimental results in this study were the mean values of the stratified fivefold cross validation. The experimental dataset was divided into a disjoint training set and test set at a ratio of 4:1. One-fold of the training set was designated as the validation set to fine-tune the model weights. Of note, we set the epoch to 500 and implemented an early stopping strategy to prevent model overfitting. The model would stop training when the loss of the validation set no longer decreased, with a patience of 10. In addition, the cross-entropy loss was employed to calculate the distance between predictions and ground truth labels, as presented in Eq. [1]. The adaptive moment estimation (Adam) optimizer with an initial learning rate of 1e-3 and cosine annealing decay was used to optimize the network. The detailed parameter settings of the proposed model are reported in Table 2.

$$l_n = -w_{y_n} \log \frac{\exp(x_n, y_n)}{\sum_{c=1}^C \exp(x_n, c)} \quad [1]$$

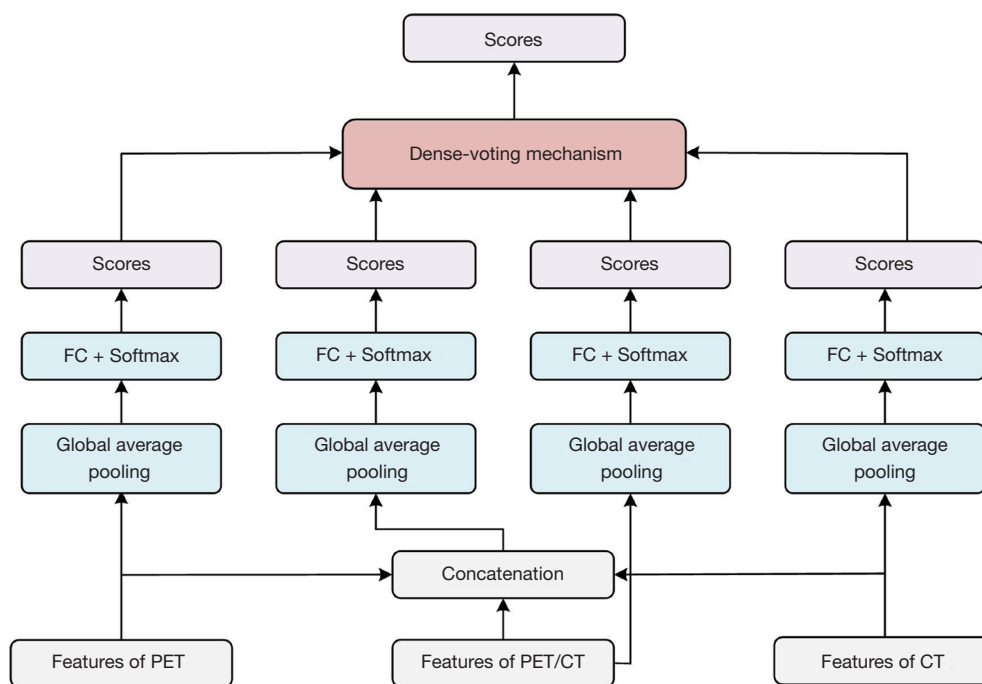
In Eq. [1],  $l_n$  is the loss of each sample,  $w$  is the weight of each class, and  $C$  indicates that there are  $c$  classes.



**Figure 3** The architecture of the 3D Inception-ResNet CBAM. 3D, three-dimensional; ResNet, residual net; CBAM, convolutional block attention module; Conv, convolutional layer; Concat, concatenation; F-C, fully connected layer.



**Figure 4** The architecture of the 3D Inception-ResNet A, B, C, and reduction A and B. ReLU, rectified linear unit; 3D, three-dimensional; Conv, convolutional layer; ResNet, residual net; Concat, concatenation.



**Figure 5** The architecture of the feature classification and score fusion. FC, fully connected layer; PET, positron emission tomography; CT, computed tomography.

**Table 2** Parameter settings of the proposed model

Parameter	Setting	Other
Learning rate	1e-3	Cosine annealing restarts
Epoch	500	Early stopping
Optimizer	Adam	–
Batch size	24	–

Adam, adaptive moment estimation; –, no settings.

The experiments in this study were compiled using Python version 3.8.16 (Python Software Foundation, Wilmington, DE, USA) and PyTorch-1.8.0 with Compute Unified Device Architecture (CUDA) version 10.2.89 and an A100 GPU (Nvidia Corp., Santa Clara, CA USA) running on Ubuntu 20.04 (Canonical, London, UK). If readers are interested in our work, the relevant core code can be provided upon request.

### Evaluation criteria

In this study, six classical evaluation metrics were used to

measure the performance of the proposed model, including *Acc*, precision (*Prec*), recall (*Rec*), specificity (*Spec*), F1 score, and the number of trainable parameters (*Para*). These metrics were calculated using the following formulae:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \quad [2]$$

$$Prec = \frac{TP}{TP + FP} \quad [3]$$

$$Rec = \frac{TP}{TP + FN} \quad [4]$$

$$Spec = \frac{TN}{TN + FP} \quad [5]$$

$$F1 = \frac{2 \times Prec \times Rec}{Prec + Rec} \quad [6]$$

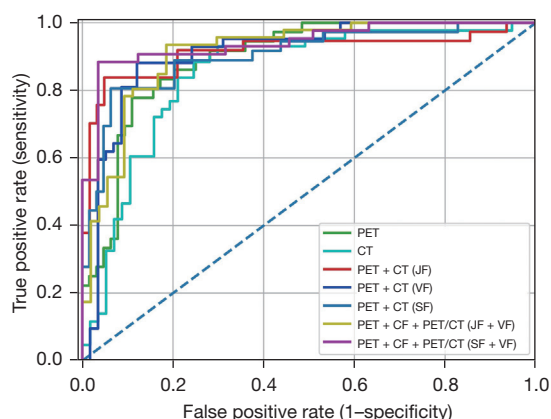
where TP, FP, TN, and FN are true positive, false positive, true negative, and false negative, respectively.

The receiver operating characteristic (ROC) curve and area under the curve (AUC) value are also frequently used to validate the performance of models in medical imaging analysis. The ROC curve tends to approach the upper-left

**Table 3** Comparison of different inputs of the proposed model

Modality	Fusion	Acc (%) (mean $\pm$ SD)	Prec (%) (mean $\pm$ SD)	Rec (%) (mean $\pm$ SD)	Spec (%) (mean $\pm$ SD)	F1 (%) (mean $\pm$ SD)	AUC	Para (M)
PET	–	87.58 $\pm$ 2.26	85.52 $\pm$ 4.74	81.92 $\pm$ 2.49	91.15 $\pm$ 2.93	83.65 $\pm$ 3.14	0.9012	44.768
CT	–	81.56 $\pm$ 2.38	73.80 $\pm$ 3.17	80.50 $\pm$ 7.22	81.90 $\pm$ 2.23	76.96 $\pm$ 4.88	0.8473	44.768
PET + CT	JF	88.58 $\pm$ 2.77	87.26 $\pm$ 3.92	82.30 $\pm$ 5.77	92.47 $\pm$ 1.74	84.67 $\pm$ 4.70	0.9153	97.091
PET + CT	VF	88.38 $\pm$ 3.02	87.21 $\pm$ 6.30	82.14 $\pm$ 5.55	92.08 $\pm$ 4.39	84.48 $\pm$ 4.69	0.9083	97.100
PET + CT	SF	88.38 $\pm$ 2.67	87.05 $\pm$ 5.89	82.31 $\pm$ 3.13	92.14 $\pm$ 3.72	84.56 $\pm$ 3.93	0.9081	44.768
PET + CT + PET/CT	JF + VF	88.78 $\pm$ 3.76	89.05 $\pm$ 4.79	81.08 $\pm$ 8.32	93.04 $\pm$ 3.08	84.79 $\pm$ 4.35	0.9206	134.304
PET + CT + PET/CT	SF + VF	89.98 $\pm$ 2.28	89.21 $\pm$ 2.52	84.75 $\pm$ 6.51	93.38 $\pm$ 1.17	86.83 $\pm$ 3.82	0.9227	150.670

Acc, accuracy; SD, standard deviation; Prec, precision; Rec, recall; Spec, specificity; AUC, area under curve; Para, number of trainable parameters; M, Mega; PET, positron emission tomography; CT, computed tomography; JF, joint fusion; VF, dense-voting fusion; SF, summing fusion.



**Figure 6** Receiver operating characteristic curves for the comparison of different inputs to the proposed model. PET, positron emission tomography; CT, computed tomography; JF, joint fusion; VF, dense-voting fusion; SF, summing fusion.

corner when the model exhibits excellent performance and results, and its AUC value correspondingly increases.

### Experimental results and analysis

The performance of several classical models, fusion strategies, and parameter settings were compared to validate the robustness of the proposed model.

Table 3 reports the performance of different inputs and fusion strategies of the proposed model in PET/CT images. Experimental results showed that the model achieved the best performance when all three channels, including CT, PET, and PET/CT, were simultaneously fed to the model

**Table 4** Confusion matrix of the proposed model

Confusion matrix	Predicted label	
	Benign	Malignant
Ground truth		
Benign	56	3
Malignant	7	34

(Acc =89.98%; AUC =0.9012). In the unimodal input experiment, PET imaging as an input exhibited higher Acc and Spec (Acc =87.58%, Spec =91.15%) compared to CT imaging (Acc =81.56%; Spec =81.90%). Moreover, the fusion of PET and CT imaging using the joint fusion strategy achieved better performance (Acc =88.58%; AUC =0.9153), slightly outperforming PET input alone (Acc =87.58%; AUC =0.9012). The ROC curves and confusion matrix of the proposed model are shown in Figure 6 and Table 4, respectively.

Table 5 provides a comparison of different 3D encoders between the proposed model and classical deep models, including AlexNet (22), LeNet (23), ResNet (24), Inception-v4 (19), DenseNet (25), and vision transformer (ViT) (26) [two-dimensional (2D) convolution and 2D pooling were restructured as 3D convolution and 3D pooling, respectively], with PET, CT, and PET/CT being used as inputs. The results clearly demonstrate that the proposed model still achieved SOTA performance compared with classical deep learning models. Interestingly, the ResNet and Inception-v4 models exhibited similar performance, ranking second only to our method.

**Table 5** Comparison of different 3D encoders between the proposed model and classical deep models with PET, CT, and PET/CT being used as inputs

Models	Acc (%) (mean ± SD)	Prec (%) (mean ± SD)	Rec (%) (mean ± SD)	Spec (%) (mean ± SD)	F1 (%) (mean ± SD)	AUC	Para (M)
AlexNet*	86.37±4.47	85.69±7.62	78.24±7.36	91.25±4.89	81.62±6.11	0.9034	15.643
LeNet*	87.37±3.78	85.55±6.44	81.91±8.31	90.59±6.05	83.34±4.24	0.9075	1.299
ResNet*	88.37±2.81	86.22±4.01	83.35±7.12	91.32±3.51	84.61±4.23	0.9059	199.717
Inception-v4*	88.96±2.63	87.36±3.95	84.11±2.88	91.93±4.07	85.62±1.82	0.9166	148.728
DenseNet*	87.37±3.05	87.81±5.54	78.51±6.91	92.58±3.93	82.67±3.95	0.9102	51.506
VIT*	76.95±2.93	67.39±4.34	77.40±12.9	76.03±6.46	71.66±6.64	0.8078	281.929
Proposed	89.98±2.28	89.21±2.52	84.75±6.51	93.38±1.17	86.83±3.82	0.9227	150.670

\*, a classical deep learning model. 3D, three-dimensional; PET, positron emission tomography; CT, computed tomography; Acc, accuracy; SD, standard deviation; Prec, precision; Rec, recall; Spec, specificity; AUC, area under curve; Para, number of trainable parameters; M, Mega; ResNet, residual net; VIT, visual transformer.

**Table 6** Ablation experiments with the CBAM module

Models	CBAM module	Acc (%) (mean ± SD)	Prec (%) (mean ± SD)	Rec (%) (mean ± SD)	Spec (%) (mean ± SD)	F1 (%) (mean ± SD)	AUC	Para (M)
Inception-ResNet		88.18±3.26	84.31±4.15	85.87±2.97	89.14±5.17	85.04±2.92	0.9156	143.898
Inception-ResNet CBAM	√	89.98±2.28	89.21±2.52	84.75±6.51	93.38±1.17	86.83±3.82	0.9227	150.670

CBAM, convolutional block attention module; Acc, accuracy; SD, standard deviation; Prec, precision; Rec, recall; Spec, specificity; AUC, area under curve; Para, number of trainable parameters; M, Mega; ResNet, residual net.

**Table 7** Effect of hyperparameter setting classification on performance

Optimizer	LR	DS	Acc (%) (mean ± SD)	Prec (%) (mean ± SD)	Rec (%) (mean ± SD)	Spec (%) (mean ± SD)	F1 (%) (mean ± SD)	AUC	Para (M)
SGD	0.01 <sup>#</sup>		87.80±3.42	84.77±6.18	84.33±8.35	90.05±4.72	84.18±4.31	0.9051	150.670
	0.0001	CAR	85.98±3.28	84.58±6.27	78.38±5.83	90.85±4.07	81.22±4.71	0.8939	
Adam	0.001 <sup>#</sup>		87.60±1.13	86.69±3.18	80.10±5.31	92.1±2.19	83.16±3.13	0.9109	
	0.0001	CAR	89.98±2.28	89.21±2.52	84.75±6.51	93.38±1.17	86.83±3.82	0.9227	

<sup>#</sup>, default values. LR, learning rate; DS, decay strategy; Acc, accuracy; SD, standard deviation; Prec, precision; Rec, recall; Spec, specificity; AUC, area under curve; Para, number of trainable parameters; M, Mega; SGD, stochastic gradient descent; CAR, cosine annealing restarts; Adam, adaptive moment estimation.

Tables 6,7 present the impact of the CBAM module and the hyperparameter setting on the proposed model, respectively. It is apparent that incorporating attention modules and resetting hyperparameters improved the performance of the model. Tables 8,9 present the performance of the proposed model under different data preprocessing strategies. Table 8 demonstrates that the no-normalization dataset was more suitable for PET/

CT preprocessing, which may be attributed to the PET/CT imaging process. Table 9 summarizes the performance of the proposed model that used the augmented data obtained through data rotation and translation. Figure 7 illustrates the loss curve of the training and validation sets using the early stopping strategy during training. A detailed discussion can be found in section “Interpretation of experiments”.

**Table 8** Effect of data normalization on the proposed model

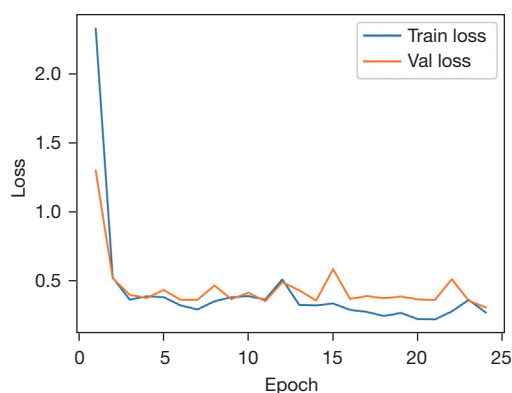
Model	Normalization	Acc (%) (mean $\pm$ SD)	Prec (%) (mean $\pm$ SD)	Rec (%) (mean $\pm$ SD)	Spec (%) (mean $\pm$ SD)	F1 (%) (mean $\pm$ SD)	AUC	Para (M)
Inception-ResNet	√	88.00 $\pm$ 4.13	86.37 $\pm$ 6.31	82.43 $\pm$ 4.14	91.16 $\pm$ 4.83	84.32 $\pm$ 4.89	0.9046	150.670
CBAM		89.98 $\pm$ 2.28	89.21 $\pm$ 2.52	84.75 $\pm$ 6.51	93.38 $\pm$ 1.17	86.83 $\pm$ 3.82	0.9227	

Acc, accuracy; SD, standard deviation; Prec, precision; Rec, recall; Spec, specificity; AUC, area under curve; Para, number of trainable parameters; M, Mega; ResNet, residual net; CBAM, convolutional block attention module.

**Table 9** Effect of data augmentation on the proposed model

Models	Augmentation	Acc (%) (mean $\pm$ SD)	Prec (%) (mean $\pm$ SD)	Rec (%) (mean $\pm$ SD)	Spec (%) (mean $\pm$ SD)	F1 (%) (mean $\pm$ SD)	AUC	Para (M)
Inception-ResNet	√	88.97 $\pm$ 1.95	88.12 $\pm$ 2.81	81.99 $\pm$ 8.82	91.58 $\pm$ 4.27	85.02 $\pm$ 2.83	0.9223	150.670
CBAM		89.98 $\pm$ 2.28	89.21 $\pm$ 2.52	84.75 $\pm$ 6.51	93.38 $\pm$ 1.17	86.83 $\pm$ 3.82	0.9227	

Acc, accuracy; SD, standard deviation; Prec, precision; Rec, recall; Spec, specificity; AUC, area under curve; Para, number of trainable parameters; M, Mega; ResNet, residual net; CBAM, convolutional block attention module.



**Figure 7** The loss curve for the use of the early stopping strategy during training. Train\_loss, loss of the training set; Val\_loss, loss of validation set.

## Discussion

In this section, we provide a brief interpretation of the experimental results and explore the causes of misclassification. Furthermore, we point out the limitations and future research directions of this study. The discussion encompasses three main aspects: (I) interpretation of experiments, (II) misclassification analysis, and (III) limitations and future directions.

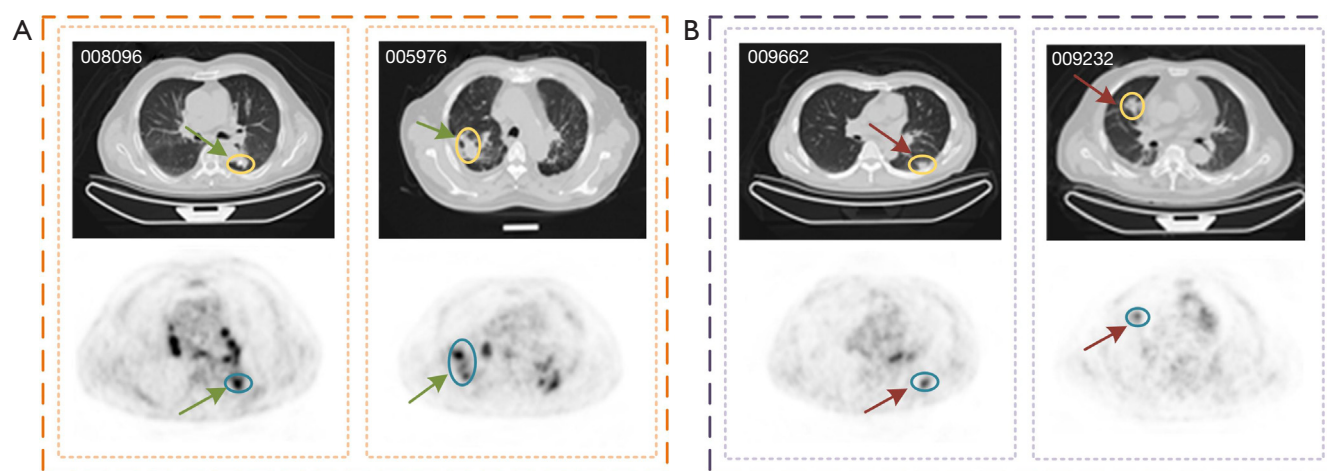
### Interpretation of experiments

This study focused on fully automated recognition of

pulmonary nodules by narrowing the gap of modal on PET and CT images. We proposed a novel two-stage multimodal framework to automatically segment ROIs and identified pulmonary nodules. Specifically, the objective of Stage I is to obtain pulmonary parenchyma masks on CT images using a classical medical image segmentation model, U-Net. Stage II involves pulmonary nodule recognition using 3D Inception-ResNet CBAM and the dense-voting integration strategy.

Experimental results indicated that the best performance was obtained when three channels, CT, PET, and PET/CT, were used as parallel inputs. This means that high-order features can be extracted using 3D Inception-ResNet CBAM, and these features can be integrated using the dense-voting fusion strategy. Furthermore, the Inception block improves the adaptability of the network width and multiscale characteristics (27). The residual architecture enhances network depth using skip connection (24), while the CBAM module helps the network focus on the ROI (20). This implies that the feature extraction structure can be constructed with a combination of both deep and wide structures (28). In summary, pulmonary nodules typically exhibit various shapes and multiscale characteristics (29), as shown in *Figure 8*, and the high-order features of PET/CT pulmonary nodules images can be effectively extracted using 3D Inception-ResNet CBAM architecture.

It is worth noting that use of PET as the input achieved higher Acc and Spec compared to use of CT in unimodal



**Figure 8** Instances of misclassified images of the proposed model. (A) The benign nodules (green arrows) were incorrectly predicted as malignant, as shown in the orange dashed box, and (B) the malignant nodules (red arrows) were incorrectly predicted as benign, as shown in the purple dashed box. The top row is the CT images, and the bottom row is the PET images. The nodules on the CT images and PET images are highlighted with yellow and blue circles, respectively. CT, computed tomography; PET, positron emission tomography.

experiments using the 3D Inception-ResNet CBAM architecture, indicating both the high sensitivity of PET images in detecting pulmonary nodules, which is consistent with previous findings (10), and the valuable role it plays in multimodal fusion classification. However, this also resulted in a high false-positive rate (30), which was confirmed in the modal fusion experiments. Specifically, the fusion strategy that used PET and CT images as input achieved a performance similar to the strategy that used PET alone. Meanwhile, the various feature integration strategies did not show satisfactory results (only 1% improvement), as shown in *Tables 3,5* and in *Figure 6*, which may be attributed to data bottlenecks (31). Nevertheless, the experimental results of the proposed model demonstrated the potential benefits of integrating multimodal features to improve the performance, which provides direction for future research work.

The effect of different fusion strategies and hyperparameter settings were also evaluated, as shown in *Tables 3,7*. Image-level fusion contributes to localizing and highlighting lesion regions in PET/CT images (30,32), while the feature-level integration, especially late fusion, helps to ensure the individual discriminative capabilities of each modality, greatly improving model robustness (33). Moreover, resetting the hyperparameters has been proven to improve the performance of models (34). Additionally, it is important to note that data normalization leads to a negative increase in the model performance, as shown in *Table 8*. This may be related to the imaging principle of PET/CT, in which

the imaging value range is  $(0, N)$ , and thus the excessive imaging values led to negative effects during weight training after data normalization (35). *Figure 7* presents the decreasing trend of the loss function during training, indicating the model's acquisition of feature knowledge from PET/CT data. Moreover, the implementation of the early stopping strategy helped to mitigate model overfitting. *Table 9* indicates that the application of geometric transformation-based data augmentation techniques did not result in performance improvement, implying that the augmentation of the dataset does not significantly increase the effectiveness and diversity of data. This results could be attributed to the choice of data augmentation approaches, which represent a direction for future research.

In summary, we developed a novel two-stage framework, consisting of a pulmonary parenchyma mask segmentation stage and a pulmonary nodule identification stage to achieve the fully automated diagnosis of pulmonary nodules in PET/CT imaging.

### Misclassification analysis

The instances of misclassifications are shown in *Figure 8* and *Table 10*. Below, we outline the reasons for misclassification by visualizing misidentified images and integrating them with imaging findings, medical history, and clinical diagnosis. The potential factors contributing to misclassification are as follows:

**Table 10** Imaging findings, medical history, and clinical diagnosis related to the images in *Figure 8*

ID	Imaging findings	Medical history	Clinical diagnosis
008096	CT: Nodular shadow in the dorsal segment of the left lower lung, lesion size of 1.6×1.4×2.1 cm, and hairy edge of the lesion  PET: Abnormally increased nodular radioactivity uptake in the dorsal segment of the left lower lung. SUVmax 6.73	Nodule in the dorsal segment of the left lower lung with visible calcifications; tuberculosis in his youth	Benign nodule
005976	CT: Thickening and disorganization of the texture of both lungs, with multiple nodules and milia in the lungs  PET: Multiple nodules in both lungs and abnormally high cornucopia of radioactivity uptake. SUVmax 7.59	Untreated tuberculosis of both lungs	Benign nodule
009662	CT: Multiple nodular foci in the left lower lung, with the largest measuring 1.3 cm in diameter  PET: Stenosis of the bronchial opening in the anterior segment of the left upper lobe with abnormally high radioactivity uptake. SUVmax 6.02. Mildly elevated radioactivity uptake present in both lungs with multiple striated shadows. SUVmax 1.61	Multiple nodules in the left lower lung and hemangioma of the liver on previous examination	Malignant nodule
009232	CT: A nodular focus visible in the right middle lung, a lesion size of 2.7×2.5×1.2 cm, an irregular margin of the lesion, and the signs of long burrs and shallow lobulation  PET: Abnormally increased nodular radioactivity uptake in the right middle lung. SUVmax 7.73	Elevated levels of malignant tumor markers CEA and CA199	Malignant nodule

CT, computed tomography; PET, positron emission tomography; SUVmax, maximum standardized uptake value; CEA, carcinoembryonic antigen; CA199, carbohydrate antigen 199.

- ❖ The primary cause for the misclassification between the benign and malignant nodules is the nonsolitary property. Nonsolitary pulmonary nodules are characterized by their overlapping with surrounding tissues or organs (36), making it challenging for the model to accurately locate and identify them (see 008096, 009662, and 009232 in *Figure 8*). Additionally, these nodules typically exhibit diverse morphological features, which may lead to misclassification from the use of generic knowledge weights during model training.
- ❖ Another potential cause of misclassification could be the irregular nodules and abnormal standardized uptake value (SUV) in both lungs. As has been widely acknowledged, nodules often present irregular shapes, such as elliptical or lobular, in both lungs (37). Additionally, the irregular or abnormal SUVs of different tissues in distinct patients (38) due to varying uptake levels (see 005976 in *Figure 8*) might have led to the inaccurate feature representation of the model.

In summary, the experimental results indicated that the primary reason for misclassification was the nonsolitary

property of nodules, which also implies that it is necessary to improve the classification performance of the proposed model by detecting and isolating the nodules from the lung parenchyma.

### *Limitations and future directions*

Although acceptable performance was achieved using the proposed model, there were several limitations. First, real clinical data are critically needed to validate the proposed model across a broader spectrum of datasets. Second, the integration strategies at both the image- and feature-levels could weaken the Spec of the modality. Third, manual hyperparameter settings rely heavily on the experience and subjective judgment of the researchers, which to some extent limits the application of the model. Finally, the effectiveness of segmentation may be affected by uncertainties, such as differences in image resolution and spatial resolution during image coregistration, internal motion introduced by respiration and heartbeat rate, etc.

In the future, we will focus on improving the performance and automation of our classification model

in several directions: (I) it is essential to localize and detect nodules from the pulmonary parenchyma, and we will construct a detection model to isolate and identify them. (II) We will design a tailored data fusion strategy to explore the information complementarity of intermodal and intramodal features, respectively. (III) We will attempt to introduce a network architecture search approach to automatically extract features and fine-tune hyper-parameters. (IV) Another effective method that should be considered is the integration of images, medical history, and electronic diagnostic reports to improve the performance of model. (V) We will introduce generative adversarial networks to augment the semantic information of data (39). (VI) Finally, uncertainty metrics will be employed to validate the performance of model by providing a more objective measure of the robustness of the model.

## Conclusions

We developed a two-stage multimodal learning framework for the automatic classification of pulmonary nodules in PET/CT imaging. Stage I involves segmenting pulmonary parenchyma masks using the pretrained U-Net model, and Stage II identifies pulmonary nodules using 3D Inception-ResNet CBAM architecture and dense-voting feature fusion mechanism. The proposed model was evaluated on a set of clinical test sets and achieved outstanding performance, with average scores of 89.98%, 89.21%, 84.75%, 93.38%, 86.83%, and 0.9227 for Acc, Prec, Rec, Spec, F1, and AUC, respectively. In addition, our findings reveal that the nonsolitary property of pulmonary nodules is the primary cause of reduced improvement of model performance, representing a direction for future research.

## Acknowledgments

**Funding:** This work was supported in part by the Science and Technology Program of Gansu Province (No. 23YFGA0004, No. 24JRRA506), the National Key Research and Development Program of China (No. 2019YFA0706200), and the Fundamental Research Funds for the Central Universities (No. lzujbky-2024-it16, No. lzujbky-2022-it24).

## Footnote

**Reporting Checklist:** The authors have completed the TRIPOD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-24-234/rc>

[amegroups.com/article/view/10.21037/qims-24-234/rc](https://qims.amegroups.com/article/view/10.21037/qims-24-234/rc)

**Conflicts of Interest:** All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-234/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. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of The 940th Hospital of Joint Logistics Support Force of Chinese People's Liberation Army (No. 2014-06). Informed consent was obtained from all individual participants.

**Open Access Statement:** This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
2. Tarver T. Cancer Facts & Figures 2012. American Cancer Society (ACS). Atlanta, GA: American Cancer Society, 2012. 66 p., pdf. Available online: <http://www.cancer.org/Research/CancerFactsFigures/CancerFactsFigures/cancer-facts-figures-2012>
3. Blandin Knight S, Crosbie PA, Balata H, Chudziak J, Hussell T, Dive C. Progress and prospects of early detection in lung cancer. *Open Biol* 2017;7:170070.
4. Roos JE, Paik D, Olsen D, Liu EG, Chow LC, Leung AN, Mindelzun R, Choudhury KR, Naidich DP, Napel S, Rubin GD. Computer-aided detection (CAD) of lung nodules in CT scans: radiologist performance and reading time with incremental CAD assistance. *Eur Radiol* 2010;20:549-57.

5. Bar-Shalom R, Valdivia AY, Blaufox MD. PET imaging in oncology. *Semin Nucl Med* 2000;30:150-85.
6. Kapoor V, McCook BM, Torok FS. An introduction to PET-CT imaging. *Radiographics* 2004;24:523-43.
7. Al Mohammad B, Hillis SL, Reed W, Alakhras M, Brennan PC. Radiologist performance in the detection of lung cancer using CT. *Clin Radiol* 2019;74:67-75.
8. Jin H, Yu C, Gong Z, Zheng R, Zhao Y, Fu Q. Machine learning techniques for pulmonary nodule computer-aided diagnosis using CT images: A systematic review. *Biomed Signal Process Control* 2023;79:104104.
9. Huang S, Yang J, Shen N, Xu Q, Zhao Q. Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective. *Semin Cancer Biol* 2023;89:30-7.
10. Shao X, Niu R, Shao X, Gao J, Shi Y, Jiang Z, Wang Y. Application of dual-stream 3D convolutional neural network based on (18)F-FDG PET/CT in distinguishing benign and invasive adenocarcinoma in ground-glass lung nodules. *EJNMMI Phys* 2021;8:74.
11. Apostolopoulos ID, Pintelas EG, Livieris IE, Apostolopoulos DJ, Papathanasiou ND, Pintelas PE, Panayiotakis GS. Automatic classification of solitary pulmonary nodules in PET/CT imaging employing transfer learning techniques. *Med Biol Eng Comput* 2021;59:1299-310.
12. Liu H, Cao H, Song E, Ma G, Xu X, Jin R, Liu C, Hung CC. Multi-model Ensemble Learning Architecture Based on 3D CNN for Lung Nodule Malignancy Suspiciousness Classification. *J Digit Imaging* 2020;33:1242-56.
13. Liu M, Li L, Wang H, Guo X, Liu Y, Li Y, Song K, Shao Y, Wu F, Zhang J, Sun N, Zhang T, Luan L. A multilayer perceptron-based model applied to histopathology image classification of lung adenocarcinoma subtypes. *Front Oncol* 2023;13:1172234.
14. Chen K, Lai YC, Vanniarajan B, Wang PH, Wang SC, Lin YC, Ng SH, Tran P, Lin G. Clinical impact of a deep learning system for automated detection of missed pulmonary nodules on routine body computed tomography including the chest region. *Eur Radiol* 2022;32:2891-900.
15. Niu C, Wang G. Unsupervised contrastive learning based transformer for lung nodule detection. *Phys Med Biol* 2022;67:10.1088/1361-6560/ac92ba.
16. Wang F, Cheng C, Cao W, Wu Z, Wang H, Wei W, Yan Z, Liu Z. MFCNet: A multi-modal fusion and calibration networks for 3D pancreas tumor segmentation on PET-CT images. *Comput Biol Med* 2023;155:106657.
17. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A. editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*, Springer, 2015;9351:234-41.
18. Hofmanninger J, Prayer F, Pan J, Röhrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *Eur Radiol Exp* 2020;4:50.
19. Szegedy C, Ioffe S, Vanhoucke V, Alemi A, editors. Inception-v4, inception-resnet and the impact of residual connections on learning. *Proceedings of the AAAI Conference on Artificial Intelligence 2017*. doi: <https://doi.org/10.1609/aaai.v31i1.11231>.
20. Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018:3-19.
21. Dubey SR, Singh SK, Chaudhuri BB. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing* 2022;503:92-108.
22. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Part of *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012.
23. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998;86:2278-324.
24. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016:770-8.
25. Huang G, Liu Z, van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017:4700-8.
26. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv: 2010.11929*, 2020.
27. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015:1-9.
28. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, editors. *Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference*

- on Computer Vision and Pattern Recognition (CVPR), 2016:2818-26.
29. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441-6.
  30. Li Y, Su M, Li F, Kuang A, Tian R. The value of 18F-FDG-PET/CT in the differential diagnosis of solitary pulmonary nodules in areas with a high incidence of tuberculosis. *Ann Nucl Med* 2011;25:804-11.
  31. Li S, Zhao B, Wang X, Yu J, Yan S, Lv C, Yang Y. Overestimated value of (18)F-FDG PET/CT to diagnose pulmonary nodules: Analysis of 298 patients. *Clin Radiol* 2014;69:e352-7.
  32. Li T, Lin Q, Guo Y, Zhao S, Zeng X, Man Z, Cao Y, Hu Y. Automated detection of skeletal metastasis of lung cancer with bone scans using convolutional nuclear network. *Phys Med Biol* 2022. doi: 10.1088/1361-6560/ac4565.
  33. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit Med* 2020;3:136.
  34. Koutsoukas A, Monaghan KJ, Li X, Huan J. Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. *J Cheminform* 2017;9:42.
  35. Lin Q, Li T, Cao C, Cao Y, Man Z, Wang H. Deep learning based automated diagnosis of bone metastases with SPECT thoracic bone images. *Sci Rep* 2021;11:4223.
  36. Cruickshank A, Stieler G, Ameer F. Evaluation of the solitary pulmonary nodule. *Intern Med J* 2019;49:306-15.
  37. Gavrielides MA, Li Q, Zeng R, Myers KJ, Sahiner B, Petrick N. Minimum detectable change in lung nodule volume in a phantom CT study. *Acad Radiol* 2013;20:1364-70.
  38. Miwa K, Inubushi M, Wagatsuma K, Nagao M, Murata T, Koyama M, Koizumi M, Sasaki M. FDG uptake heterogeneity evaluated by fractal analysis improves the differential diagnosis of pulmonary nodules. *Eur J Radiol* 2014;83:715-9.
  39. Chen Y, Yang XH, Wei Z, Heidari AA, Zheng N, Li Z, Chen H, Hu H, Zhou Q, Guan Q. Generative Adversarial Networks in Medical Image augmentation: A review. *Comput Biol Med* 2022;144:105382.

**Cite this article as:** Li T, Mao J, Yu J, Zhao Z, Chen M, Yao Z, Fang L, Hu B. Fully automated classification of pulmonary nodules in positron emission tomography-computed tomography imaging using a two-stage multimodal learning approach. *Quant Imaging Med Surg* 2024;14(8):5526-5540. doi: 10.21037/qims-24-234