

PAPER

## Multi-scale discriminative regions analysis in FDG-PET imaging for early diagnosis of Alzheimer's disease

To cite this article: Jin Zhang *et al* 2022 *J. Neural Eng.* **19** 046030

View the [article online](#) for updates and enhancements.

### You may also like

- [Predictive modeling of outcomes following definitive chemoradiotherapy for oropharyngeal cancer based on FDG-PET image characteristics](#)  
Michael R Folkert, Jeremy Setton, Aditya P Apte *et al.*
- [Multiparametric cardiac  \$^{18}\text{F}\$ -FDG PET in humans: pilot comparison of FDG delivery rate with  \$^{82}\text{Rb}\$  myocardial blood flow](#)  
Yang Zuo, Javier E López, Thomas W Smith *et al.*
- [A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities](#)  
M Vallières, C R Freeman, S R Skamene *et al.*



**EDINBURGH INSTRUMENTS**

WORLD LEADING MOLECULAR SPECTROSCOPY SOLUTIONS

[edinst.com](http://edinst.com)

The advertisement features a red background with the Edinburgh Instruments logo on the left, which consists of a stylized sunburst of white dots. To the right, several pieces of scientific equipment are displayed, including a large white and black spectrometer labeled 'FLS 1000' and a smaller instrument labeled 'FSS'. The text 'EDINBURGH INSTRUMENTS' is prominently displayed in white, bold, uppercase letters. Below the logo, the text 'WORLD LEADING MOLECULAR SPECTROSCOPY SOLUTIONS' is written in white, bold, uppercase letters. In the bottom right corner, the website 'edinst.com' is shown in white text inside a red rounded rectangle.



## PAPER

## Multi-scale discriminative regions analysis in FDG-PET imaging for early diagnosis of Alzheimer's disease

Jin Zhang<sup>1</sup>, Xiaohai He<sup>1,\*</sup> , Linbo Qing<sup>1</sup>, Yining Xu<sup>1</sup>, Yan Liu<sup>2</sup> and Honggang Chen<sup>1</sup><sup>1</sup> College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, Sichuan, People's Republic of China<sup>2</sup> Department of Neurology, The Affiliated Hospital of Southwest Jiaotong University, The Third People's Hospital of Chengdu, Chengdu 610014, Sichuan, People's Republic of China

\* Author to whom any correspondence should be addressed.

E-mail: [hxh@scu.edu.cn](mailto:hxh@scu.edu.cn)**Keywords:** Alzheimer's disease, fluorodeoxyglucose positron emission tomography, artificial intelligence, medical image processing, mild cognitive impairmentRECEIVED  
31 March 2022REVISED  
6 July 2022ACCEPTED FOR PUBLICATION  
26 July 2022PUBLISHED  
9 August 2022**Abstract**

*Objective.* Alzheimer's disease (AD) is a degenerative brain disorder, one of the main causes of death in elderly people, so early diagnosis of AD is vital to prompt access to medication and medical care. Fluorodeoxyglucose positron emission tomography (FDG-PET) proves to be effective to help understand neurological changes via measuring glucose uptake. Our aim is to explore information-rich regions of FDG-PET imaging, which enhance the accuracy and interpretability of AD-related diagnosis. *Approach.* We develop a novel method for early diagnosis of AD based on multi-scale discriminative regions in FDG-PET imaging, which considers the diagnosis interpretability. Specifically, a multi-scale region localization module is discussed to automatically identify disease-related discriminative regions in full-volume FDG-PET images in an unsupervised manner, upon which a confidence score is designed to evaluate the prioritization of regions according to the density distribution of anomalies. Then, the proposed multi-scale region classification module adaptively fuses multi-scale region representations and makes decision fusion, which not only reduces useless information but also offers complementary information. Most of previous methods concentrate on discriminating AD from cognitively normal (CN), while mild cognitive impairment, a transitional state, facilitates early diagnosis. Therefore, our method is further applied to multiple AD-related diagnosis tasks, not limited to AD vs. CN. *Main results.* Experimental results on the Alzheimer's Disease Neuroimaging Initiative dataset show that the proposed method achieves superior performance over state-of-the-art FDG-PET-based approaches. Besides, some cerebral cortices highlighted by extracted regions cohere with medical research, further demonstrating the superiority. *Significance.* This work offers an effective method to achieve AD diagnosis and detect disease-affected regions in FDG-PET imaging. Our results could be beneficial for providing an additional opinion on the clinical diagnosis.

**1. Introduction**

Alzheimer's disease (AD) is a degenerative and irreversible brain disorder that is the most common type of dementia [1]. Approximately 90 million people worldwide are currently diagnosed with AD, and it is predicted that 300 million people will develop AD worldwide by 2050 [2, 3]. Mild cognitive impairment (MCI) is generally regarded as the transition from cognitively normal (CN) to AD, and over 33% of MCI patients would progress to AD within five or

more years [4]. AD brings heavy economic and spiritual burdens to patients, families, and society [5]. Presently, there is no curative treatment for AD, while the progression may be slowed through treatment, such as medications, exercise, and memory training [6]. In this regard, early detection and accurate diagnosis of AD are critical for delaying the disease progression and improving the general well-being of patients [7].

Improvements within medical imaging have created new opportunities regarding both diagnosing

and understanding many types of neurodegenerative diseases [8, 9], one of which is AD. Since various neuroimaging technologies can differentiate neuropathological alterations, they have been widely used for AD diagnosis [10–15]. Fluorodeoxyglucose positron emission tomography (FDG-PET) imaging is one of the effective functional biomarkers for AD diagnosis by indicating glucose metabolism activity and distribution. Since patients have been shown to have severe abnormalities in glucose metabolism [16–19], FDG-PET can be used to detect or locate regions particularly affected by the disease, which in turn helps physicians to diagnose AD/MCI. Manual evaluation of brain FDG-PET images remains qualitative and operator-dependent, which is time-consuming and expensive.

Recently, various computer-aided diagnosis techniques have been proposed for the AD diagnosis based on FDG-PET imaging to improve the accuracy. These approaches could be categorized into two groups according to the feature representation. One is based on machine learning [20–24] that requires hand-crafted features, which is usually divided into two stages: feature extraction and classification. Pan *et al* [25] devised a multilevel feature representation containing the features of regions and their connectivity to diagnose AD. Gray *et al* [26] proposed a regional feature extraction method that utilized clinical data and FDG-PET imaging to improve the performance of AD diagnosis. The other is based on deep learning [27–30], which automatically learns the features to represent FDG-PET data. Deep learning-based approaches can be roughly classified into three types based on the input type: two-dimensional (2D) slice-based methods, three-dimensional (3D) subject-based methods, and region of interest (ROI)/patch-based methods.

### 1.1. 2D slice-based methods

This kind of method generally extracts 2D slices from 3D FDG-PET scans [31], which are then provided into the 2D convolutional neural network (CNN) model. Thanks to the success of 2D CNN in natural image classification, the existing CNN is utilized in a transfer learning fashion [32]. Pan *et al* [33] developed a separable convolution network to extract complement information from slices of FDG-PET scans in three views for AD diagnosis. Ding *et al* [32] divided the FDG-PET scan into 16 evenly spaced sections and then passed them into the InceptionV3 architecture to learn the general features for AD classification. The limitation of these algorithms is that FDG-PET scan is 3D while slices are processed independently, which may lead to the loss of 3D information and data leakage.

### 1.2. 3D subject-based methods

With the development of high-performance computing resources, more and more researches focus on

3D subject-based methods [28, 30]. Islam and Zhang [34] proposed a 3D deep network using whole-brain PET scans as input and employed various visualization techniques to assist in understanding the network behavior in AD diagnosis. Yee *et al* [28] presented a 3D CNN model with residual connections that achieved superior performance in AD diagnosis. The superiority of this method is the complete integration of 3D information, since the input is the whole FDG-PET scan. However, there are more parameters to be learned in the training process, which increases the risk of overfitting.

### 1.3. ROI/patch-based methods

Methods belonging to this group typically divide the FDG-PET scan into smaller regions as input [27, 35]. The whole FDG-PET scan includes parts of the brain unaffected by AD, which are not informative. This problem can be overcome by selecting and constructing a set of 3D patches/regions from the FDG-PET scan to focus on regions that are known to be information-rich. By extracting ROIs from 3D PET images by atlas, Guo *et al* [36] designed a hierarchical graph convolution network to conquer its drawback on the Euclidean grid in AD diagnosis. The main disadvantage of these methodologies is that they consider only a few patches/regions, whereas abnormal changes in patients span multiple brain areas.

There are still limitations in the performance and interpretability of AD diagnosis based on FDG-PET imaging. First, AD is believed to affect different brain regions discriminatively. Therefore, existing studies based on the full-volume FDG-PET images for AD diagnosis may extract the redundant information, leading to overfitting of the model and thus reducing diagnostic performance. This problem can be avoided by focusing on the AD-affected regions of FDG-PET images. However, the FDG-PET scans from the public dataset Alzheimer's Disease Neuroimaging Initiative (ADNI) [37] only provide image-level labels, without the pixel-level labels that are manually labeled. Thus, it remains a challenging task to automatically identify and analyze information-rich regions of FDG-PET scans without prior information. Also, how to effectively fuse the information of discriminative regions for AD diagnosis is also one of the current research issues. Second, there is a lack of clear understanding of why and how deep neural networks perform well in AD diagnosis tasks. So, the application of AD diagnosis methods based on deep learning requires enhanced interpretability. Third, most of the previous research only pay attention to AD vs. CN classification, whereas MCI is recognized as a transitional state that is critical for AD diagnosis. Further investigation is clearly needed for other two-class classification tasks (AD vs. MCI and MCI vs. CN) and the multi-class classification task (AD vs. MCI vs. CN).

To overcome the above limitations, we propose a novel framework in this paper for the early diagnosis

of AD based on multi-scale discriminative regions in FDG-PET imaging. Specifically, a multi-scale region localization (MSRL) module is proposed to automatically locate the disease-affected regions in full-volume FDG-PET images without access to pixel-level labels, and select the optimal regions according to the designed confidence score. Moreover, to further mine the effective information of discriminative regions, a multi-scale region classification (MSRC) module is designed, in which region-level feature representations are jointly learned and adaptively fused and decision fusion is further performed to improve the performance of AD diagnosis. To fully evaluate the proposed framework and consider all disease states during the AD development, we apply it in multiple AD-related classification problems, including multi-class and three two-class classification tasks.

To sum up, the main contributions of this study are summarized as follows:

- To improve the performance and interpretability of AD diagnosis, we propose a novel framework to take account of multi-scale discriminative regions jointly with an expectation to be complement mutually, thus offering more information. To our best knowledge, this is the first report of solving AD vs. CN, AD vs. MCI, MCI vs. CN and AD vs. MCI vs. CN classification tasks based on FDG-PET images.
- An MSRL module is proposed to automatically locate disease-related discriminative regions in full-volume FDG-PET images based on saliency maps without prior information. On this basis, a confidence score is designed to evaluate the confidence and prioritization of the extracted regions, further selecting the optimal regions.
- To reduce the redundant information of disease-affected regions and provide complementary information, an MSRC module is proposed that dynamically controls the contributions of different regions through an adaptive fusion gate.
- We comprehensively evaluate the proposed method on the public dataset ADNI for multiple AD-related classification tasks, and experimental results show that it yields superior performance compared with state-of-the-art (SOTA) methods. Furthermore, some disease-affected regions of the cerebral cortex highlighted by extracted regions cohere with medical research, further demonstrating its superiority.

The remaining of this paper is organized as follows. Section 2 describes the dataset used for evaluation and the proposed method in detail. In section 3, we report and analyze the experimental results. Section 4 investigates the influence of several key components of the proposed method. Finally, a conclusion is given in section 5.

## 2. Materials and methods

### 2.1. Dataset

#### 2.1.1. Data acquisition

FDG-PET scans used in this paper come from the ADNI [37] dataset, which is the largest publicly available dataset for neuroimaging-based AD research. Totally, 1732 FDG-PET scans of 873 subjects constitute the experimental dataset. The disease state of collected FDG-PET scans is categorized into three classes: AD, MCI, and CN. Each subject is scanned on at least one visit, which means that the scans from one subject have a unique subject ID and different session ID. The scans are split by subject ID into training, validation and testing sets, and their demographic and clinical information is shown in table 1. The data is split before preprocessing to prevent the data leakage mentioned in AD diagnosis studies based on MRI scans [38], which has been demonstrated to improve the diagnosis performance significantly [39, 40]. Data leakage is discussed for the first time, to the best of our knowledge, for AD diagnosis based on FDG-PET scans. Mini-mental state examination (MMSE), global clinical dementia rating (CDR), and clinical dementia rating sum of boxes (CDR-SOB) scores are clinician-rated methods for cognitive dysfunction. These scores are reported in terms of mean (standard deviation), mainly utilized for the balanced data partitioning with similar distributions of scales between the training, validation and testing sets.

#### 2.1.2. Data preprocessing

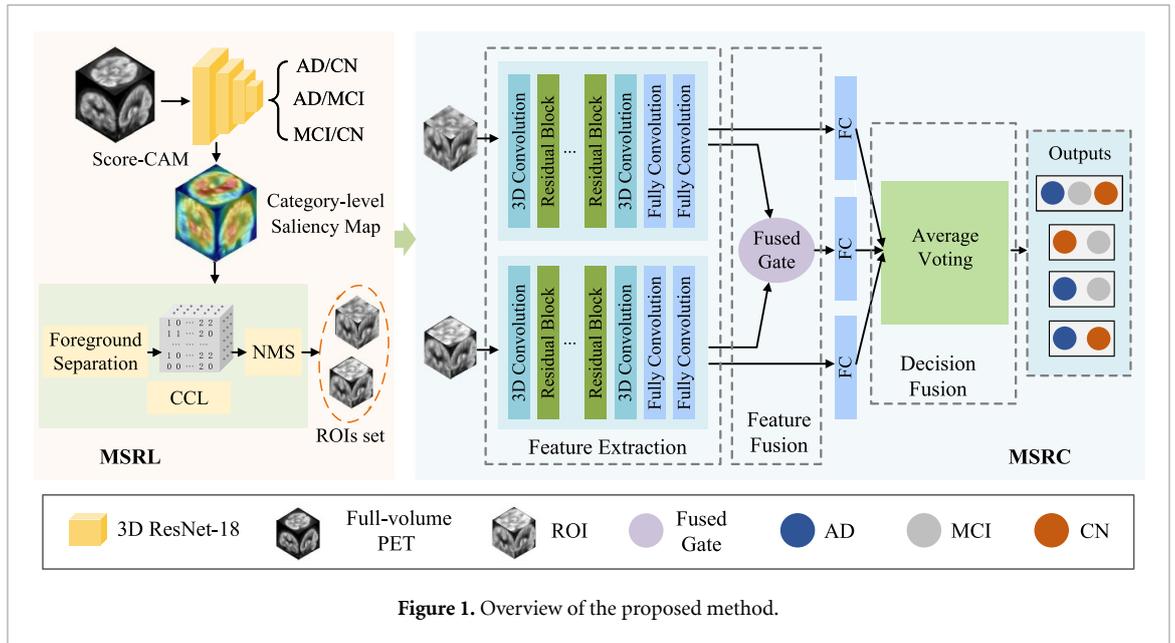
For this study, the selected FDG-PET scans are preprocessed by performing spatial normalization, resampling, intensity normalization and data augmentation. We use the Clinica software platform [41] developed by ARAMIS Lab to achieve the data preprocessing, which supports Functional MRI of the Brain Software Library (FSL), statistical parametric mapping (SPM), and FreeSurfer [42]. The input is the FDG-PET images transformed to the Brain Imaging Data Structure format. The FDG-PET scans are spatially normalized into Montreal Neurological Institute space using the Diffeomorphic Anatomical Registration using Exponentiated Lie algebra (DARTEL) deformation model of SPM, and then are resized to  $128 \times 128 \times 128$  voxels along sagittal, coronal, and axial dimensions respectively to reduce computational cost. Moreover, the intensity normalization is done by using the min-max normalization. Data augmentation on FDG-PET scans is completed by Gaussian blurring with  $\sigma$  having a uniform range of 0–1.5.

### 2.2. Overview of the proposed framework

In this paper, we present a novel framework to tackle the problem of FDG-PET-based AD diagnosis. As shown in figure 1, our proposed method

**Table 1.** Summary of demographics, MMSE, CDR, and CDR-SOB of the dataset.

Split	Diagnosis	Subjects	Scans	Age	MMSE	CDR	CDR-SOB
Training	CN	184	411	76.01(6.37)	28.98(1.88)	0.02(0.17)	0.10(0.40)
	MCI	347	673	73.53(7.84)	27.51(2.75)	0.49(0.13)	1.58(1.06)
	AD	146	292	75.24(7.12)	21.72(4.20)	0.90(0.42)	5.30(2.43)
Validation	CN	27	56	76.98(5.50)	29.03(1.21)	0.00(0.16)	0.11(0.31)
	MCI	45	101	74.28(7.74)	27.61(3.40)	0.48(0.19)	1.61(1.03)
	AD	13	18	74.00(7.34)	21.78(3.90)	0.83(0.37)	5.63(2.88)
Testing	CN	34	64	74.82(6.11)	29.12(1.14)	0.03(0.12)	0.08(0.31)
	MCI	54	82	73.37(6.47)	27.35(2.22)	0.48(0.20)	1.49(1.02)
	AD	23	35	72.16(7.99)	22.34(3.46)	0.85(0.31)	5.04(2.04)



mainly consists of two parts: the MSRL module and the MSRC module. Briefly, multi-scale discriminative regions are extracted by the MSRL module (section 2.3) from the full-volume PET imaging in an unsupervised manner based on category-level saliency maps obtained by score-weighted class activation mapping (Score-CAM) [43]. Particularly, a confidence score is designed to measure the density distribution of anomalies in the regions and further evaluate the confidence and prioritization of the extracted regions to select the optimal regions, which are then fed into the MSRC module (section 2.4) to extract feature representations and classification outputs for these regions. At the same time, multi-scale regions are fused by the adaptive fusion method to aggregate useful information together and then are processed to obtain the fusion feature representation and classification output. Finally, multiple classification outputs are performed by the decision-level fusion to obtain the final result.

### 2.3. MSRL

Commonly, an original image contains only some regions perceived as important, which are called ROIs. In AD diagnosis based on medical imaging,

ROIs are the disease-related regions that provide effective information. The goal of ROI extraction is to preserve specific brain regions that contain discriminative information, thereby reducing redundant information. Even if experts can manually label the ground-truth regions within nuclear medicine, it would be very time-consuming. Therefore, we propose the MSRL module based on unsupervised saliency-guided localization for region identification without the requirement of manual labeling. The MSRL module consists of four sequential components, i.e. (a) category-level saliency map acquisition, (b) separation of foreground and background, (c) location proposal, and (d) region selection.

#### 2.3.1. Category-level saliency map acquisition

Since the visualization results could indicate regions in the input image that the network focuses on, they are calculated in this paper by the Score-CAM [43] algorithm to highlight regions responsible for AD diagnosis using a backbone network. The backbone network is 3D ResNet-18, which is a deep-learning model for the classification of the original full-volume FDG-PET scans. Score-CAM performs a linear combination of activation maps with its forward passing score

as the weight to obtain the saliency maps, thus eliminating the dependence on gradients.

The highlighted areas of activation map  $A^k$  are projected on the input space by multiplying each normalized activation map with the input image  $I$  to obtain a masked image. The masked image are then passed to the backbone network  $f$  with softmax to get the scores for each class as follows:

$$S_k = \text{softmax}(f(A^k \circ I)) \quad (1)$$

where  $k$  denotes the  $k$ th channel of the activation map.

The score  $S_k^c$  of the target class  $c$  used as weight  $\alpha_k^c$  is therefore described as:

$$\alpha_k^c = S_k^c \quad (2)$$

where  $\alpha_k^c$  captures the importance of the  $k$ th activation map for the target class  $c$ .

The final saliency map is the sum across all the activation maps for the linear combination between the target class score and each activation map with rectified linear unit (ReLU), which becomes:

$$m^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right). \quad (3)$$

For correctly predicted FDG-PET images of different categories in the testing set, there are saliency maps and the prediction probabilities from the backbone network. Based on them, the category-level saliency map for each category is represented as:

$$M_c = \frac{\sum_{i=1}^n p_i^c \times m_i^c}{n} \quad (4)$$

where  $m_i^c$  is the saliency map of sample  $i$  from category  $c$ ,  $p_i^c$  is the prediction probability of sample  $i$  from category  $c$ ,  $n$  is the total number of samples belonging to category  $c$ , and  $c \in \{\text{AD}, \text{MCI}, \text{CN}\}$ .

### 2.3.2. Separation of foreground and background

Based on the category-level saliency map, we can further analyze the clues to achieving correct classification, that is, the glucose metabolism of brain regions in different categories. In order to highlight the difference, the separation of foreground and background helps to focus on highly activated regions in the category-level saliency map, which is affected by AD/MCI. The saliency map  $M_c$  for category  $c$  is performed by intensity normalization to obtain the normalized saliency map  $M_c'$  with a range of  $[0, 1]$ . Given a quantile threshold  $0 < \tau_s < 1$ , the normalized category-level saliency map  $M_c'$  is divided into the background (pixels with values less than  $\tau_s$ ) and foreground  $M_f$  (pixels with values greater than  $\tau_s$ ). The threshold value is chosen appropriately with the purpose of making  $M_f$  concentrated in the disease-related regions, which is associated with the varying intensities of each  $M_c'$ . Therefore, the threshold  $\tau_s$

can be determined based on the number and size of regions for each  $M_c'$ . The separation of the foreground  $M_f$  is described as follows:

$$M_f(x, y, z) = \begin{cases} 1, & M_c'(x, y, z) \geq \tau_s \\ 0, & M_c'(x, y, z) < \tau_s \end{cases} \quad (5)$$

where  $x$ ,  $y$  and  $z$  represent the coordinates of the axial, coronal, and sagittal planes, respectively.

### 2.3.3. Location proposal

After obtaining the foreground  $M_f$ , it is clustered with connected component labeling (CCL), thus the highly activated regions are segmented in an unsupervised manner. CCL groups together pixels belonging to the same connected component, which is the collection of all pixels in an image that is connected relative to the defined connectivity. Connected components, in a 3D image, are clusters of pixels with the same value that share edge or vertex. In this paper, two neighboring pixels in  $M_f$  are called connected, if they both have values of 1, forming a highly activated region, while background pixels with values of 0 are usually ignored. All connected components in  $M_f$  are extracted and labeled as  $R'$  to represent the highly activated regions, which are the discriminative regions associated with AD/MCI.

### 2.3.4. Region selection

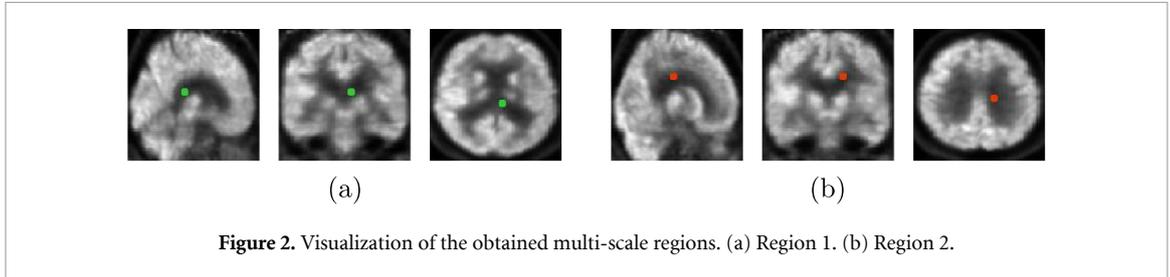
The regions  $R'$  obtained above might have duplicative information with overlap, so we need to perform automatic region selection to get the best ROIs with different scales. Besides, too many regions increase computational complexity and memory cost, and cause information redundancy, affecting diagnostic performance. Non-maximum suppression (NMS) [44] is widely used to filter the predictions of an object detector in the object detection task, we thus utilize it to filter out the overlapping regions combined with the confidence of regions to obtain the optimal ROIs  $R$ . First, the intersection over union (IoU) is calculated to measure the overlap between two regions. A larger IoU represents more overlap between the two regions. Second, we design a confidence score measuring the density distribution of high activation points in a region to represent ROI confidence and prioritization. A higher score denotes the denser distribution of discriminative features. The score  $s_k$  for the region  $k$  is calculated as follows:

$$s_k = \alpha \cdot \lg \sum_{i=1}^m b_i^k / (\text{dis}(i_k - u_k) + \beta) \quad (6)$$

where  $i_k$  and  $u_k$  represent the high activation point and central point of region  $k$ .  $\text{dis}()$  refers to the distance.  $b_i^k$  is the value of the high activation point  $i$  in region  $k$ .  $\alpha$  is the normalized coefficient as 0.25.  $m$  is the total number of pixels in region  $k$ .  $\beta$  is used to avoid dividing by zero, set as 0.1.

**Table 2.** Details of the obtained multi-scale discriminative regions.

ROIs	Coordinates (central point)	Size (length, width and height)	Confidence score
Region 1	78, 60, 81	60, 60, 60	0.8650
Region 2	70, 55, 66	100, 100, 100	0.8675



NMS discards the regions with low confidence score or large overlap while the reserved regions cover the discriminating information for classification. The IoU threshold  $\tau_{nms}$  determines that the regions whose IoU is greater than  $\tau_{nms}$  are removed. The procedure is repeated for each item in the regions until no more available regions. In this paper, the confidence threshold is 0.70 and the IoU threshold is 0.30. Finally, the details of the obtained multi-scale discriminative regions with high confidence are shown in table 2, and the central locations of these regions are marked with colored dots in axial, coronal, and sagittal planes, respectively, as displayed in figure 2.

## 2.4. MSRC

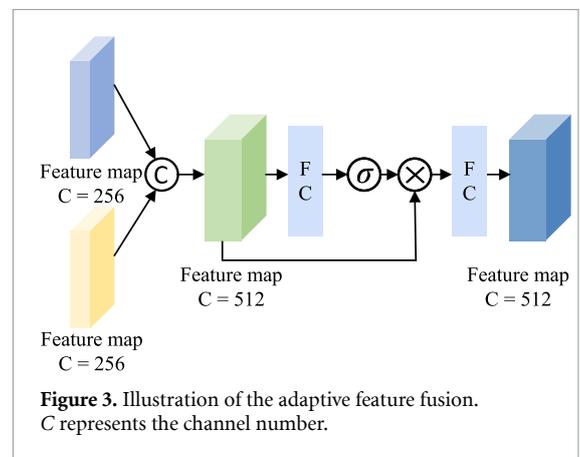
To improve the performance of AD diagnosis by organically integrating the complementary information of the multi-scale regions, we propose MSRC, which consists of three branches: feature extraction branch, adaptive feature fusion branch, and decision fusion branch.

### 2.4.1. Feature extraction branch

This branch is to learn the representation among the multi-scale regions. Assuming the set of multi-scale regions is  $R = \{R_i\}_{i=1}^k$  for an individual FDG-PET scan, where  $k$  is the number of regions. 3D ResNet-18 [45] is currently a widely used network in natural image classification and has proven to be suitable for medical image classification, so it is adopted in this study to extract region-level features. Through 3D ResNet-18, each region  $R_i$  is embedded into a  $b$ -dimensional vector space ( $b = 256$ ) to obtain region representation  $v_i$ .

### 2.4.2. Adaptive feature fusion branch

To strengthen distinctive AD-related features by integrating the region information adaptively, we introduce a fusion gate to control information flow as shown in figure 3. The adaptive feature fusion dynamically controls the contribution of multi-scale regions and accumulates the distinctiveness derived



from each scale as follows, thus improving the classification performance

$$\text{gate} = \sigma(W_1[v_1, v_2]) \quad (7)$$

$$\tilde{v} = W_2(\text{gate} \circ [v_1, v_2]) \quad (8)$$

where  $v_1$  and  $v_2$  stand for region representations, gate is the control gate.  $\sigma$  is sigmoid function and  $\circ$  denotes element-wise product.  $\tilde{v}$  represents the output of the feature fusion branch.  $W_1$  and  $W_2$  are weight matrices.

### 2.4.3. Decision fusion branch

After fusing multi-scale regions, the feature representation of each single region and the fused feature representation are independently used as the input of classification layer to generate prediction results  $y_i^c$ , which is called decision-level fusion. In real applications, the importance of regions varies from individual to individual. So, it is necessary to fuse each of the outputs so that the final integration reflects more accurate results. One simple way to adjust the contribution of each output is to apply the average voting, in which each output contributes equally to the classification results. The class label, which has maximum

score in the above fusion process is selected as the final prediction, as follows:

$$\hat{y} = \arg \max_{c=1}^C \left( \frac{1}{n} \sum_{i=1}^n y_i^c \right) \quad (9)$$

where  $C$  is the number of classes,  $y_i^c$  represents the prediction result for the  $c$ th class, in which  $i \in \{1, 2, 3\}$  indicates different input.

### 2.5. Loss function

To assist the training of the MSRC module, we introduce a loss function based on multi-feature learning. The prediction distributions of region feature representation  $v_i$  and fused feature representation  $\tilde{v}$  are substituted into the cross entropy loss function as components  $L_i$  and  $\tilde{L}$  of the final loss function, respectively. The learning algorithm is optimized by minimizing the following loss function:

$$L_{\text{all}} = \lambda \left( \sum_i^k L_i + \tilde{L} \right) \quad (10)$$

where  $k$  is the number of discriminative regions. Tuning parameter  $\lambda$  controls the influences of region-level and fusion-level training losses, empirically set as  $1/(k+1)$ .

## 3. Results

### 3.1. Implementation details

All the experiments are conducted using the Pytorch library on a computer with Ubuntu 18.04 operating system. The proposed framework is implemented by using Python 3.6 with an NVIDIA GeForce GTX 3090 GPU with 24 GB of memory. We perform training by utilizing the Adam [46] optimizer with a batch size of 8 for 200 epochs. The initial learning rate is 0.0001, which is then adjusted according to the cosine annealing strategy.

### 3.2. Evaluation criteria

To quantitatively evaluate the results of two-class classification tasks, five metrics, namely, accuracy (ACC), sensitivity (SEN), specificity (SPE), F1 score (F1) and the area under the curve (AUC) are utilized as evaluation criteria. These evaluation metrics are defined as follows:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (11)$$

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{SPE} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (12)$$

$$\text{F1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FN} + \text{FP}} \quad (13)$$

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

For the multi-class classification task, nine metrics, namely, accuracy (ACC), macro-sensitivity ( $\text{SEN}_a$ ), micro-sensitivity ( $\text{SEN}_i$ ), macro-specificity ( $\text{SPE}_a$ ), micro-specificity ( $\text{SPE}_i$ ), macro-F1 ( $\text{F1}_a$ ), micro-F1 ( $\text{F1}_i$ ), macro-AUC ( $\text{AUC}_a$ ) and micro-AUC ( $\text{AUC}_i$ ) are utilized as evaluation criteria, which are defined as follows:

$$B_a = \frac{1}{C} \sum_{i=1}^C B(\text{TP}_i, \text{FP}_i, \text{TN}_i, \text{FN}_i) \quad (14)$$

$$B_i = B \left( \sum_{i=1}^C \text{TP}_i, \sum_{i=1}^C \text{FP}_i, \sum_{i=1}^C \text{TN}_i, \sum_{i=1}^C \text{FN}_i \right) \quad (15)$$

where  $\text{TP}_i$ ,  $\text{FP}_i$ ,  $\text{FN}_i$  and  $\text{TN}_i$  are the true positive, false positive, false negative and true negative of class  $i$  samples, respectively and  $C$  is the number of classes.  $B$  stands for SEN, SPE, F1 and AUC.

### 3.3. Comparison with other methods

To justify the performance of the proposed method, the experimental results are compared with SOTA FDG-PET-based AD diagnosis approaches, as well as the baseline method using the same experimental data as our study. Majority of previous studies [27, 28, 33, 47] only focus on discriminating AD from CN, whereas MCI as a transitional state is crucial for AD diagnosis. Further investigation is discussed in this paper for the two-class classification tasks (AD vs. CN, MCI vs. AD, MCI vs. CN) and the multi-class classification task (AD vs. CN vs. MCI).

The SOTA methods for AD vs. CN classification are divided into feature-based classification methods and deep learning-based classification methods. Since the FDG-PET scans and preprocessing progress adopted in these studies are not identical and it is not known whether the data leakage occurred, the comparisons are made for reference only. 3D ResNet-18, a classical CNN for solving classification problems, is further utilized with the same FDG-PET scans and preprocessing process as in this paper to make a fair comparison with our proposed method. The comparison results of AD vs. CN classification are summarized in table 3. As can be seen, the proposed method, with an accuracy of 97.83%, is significantly better (almost 4% higher) than currently available SOTA methods. Further, our method outperforms ResNet-18 using the same data with gains of 3.26%, 8.00%, 1.50% and 6.20% regarding ACC, SEN, SPE and F1, which suggests the proposed framework considering multi-scale discriminative regions is effective and reasonable.

The comparison results for MCI vs. AD and MCI vs. CN are shown in tables 4 and 5, respectively. For MCI vs. AD classification, the accuracy of our method is 6.02% higher when compared to the best previously published method [50]. Compared to ResNet-18, the proposed framework can improve the diagnostic accuracy of MCI vs. CN, with an enhancement of

**Table 3.** Performance comparison with SOTA methods for AD vs. CN classification. Bold fonts highlight the best performance.

Category	Method	Subjects	AD vs. CN classification (%)				
			ACC	SEN	SPE	AUC	F1
Feature-based methods	Padilla <i>et al</i> * [20]	53AD + 52CN	86.59	87.50	85.36	—	—
	Gray <i>et al</i> * [26]	50AD + 54CN	88.40	83.20	93.60	—	—
	Li <i>et al</i> * [48]	25AD + 30CN	89.10	92.00	86.00	97.00	—
	Pan <i>et al</i> * [49]	237AD + 242CN	94.20	91.45	96.76	97.42	—
Deep learning methods	Lu <i>et al</i> * [27]	226AD + 304CN	93.58	91.54	95.06	—	—
	Liu <i>et al</i> * [31]	93AD + 100CN	91.20	91.40	91.00	95.30	—
	Yee <i>et al</i> * [28]	237AD + 359CN	93.50	92.30	94.20	97.60	—
	BMNet* [50]	198AD + 263CN	89.80	89.28	91.20	92.81	91.11
	Huang <i>et al</i> * [35]	465AD + 480CN	89.11	90.24	87.77	92.69	—
	Shen <i>et al</i> * [47]	—	86.60	89.50	85.20	—	—
	MiSePyNet* [33]	237AD + 242CN	93.13	90.32	95.49	97.11	—
	ResNet-18	146AD + 184CN	94.57	88.00	97.01	<b>98.51</b>	89.80
	<b>Ours</b>	146AD + 184CN	<b>97.83</b>	<b>96.00</b>	<b>98.51</b>	96.78	<b>96.00</b>

The symbol \* means just as a reference since the selected experimental data and data preprocessing processes in these methods are different.

**Table 4.** Performance comparison with SOTA methods for MCI vs. AD classification. Bold fonts highlight the best performance.

Method	Subjects	MCI vs. AD classification (%)				
		ACC	SEN	SPE	AUC	F1
Li <i>et al</i> * [48]	25AD + 29MCI	80.20	80.00	80.00	85.00	—
BMNet* [50]	198AD + 147MCI	81.18	77.53	84.41	81.67	—
ResNet-18	146AD + 347MCI	84.00	50.00	<b>92.93</b>	85.74	56.52
<b>Ours</b>	146AD + 347MCI	<b>87.20</b>	<b>80.77</b>	88.89	<b>87.06</b>	<b>72.41</b>

The symbol \* means just as a reference since the selected experimental data and data preprocessing processes in these methods are different.

**Table 5.** Performance comparison with SOTA methods for MCI vs. CN classification. Bold fonts highlight the best performance.

Method	Subjects	MCI vs. CN classification (%)				
		ACC	SEN	SPE	AUC	F1
Li <i>et al</i> * [48]	29MCI + 30CN	65.80	66.00	<b>65.00</b>	72.00	—
ResNet-18	347MCI + 184CN	68.39	<b>84.27</b>	46.97	69.97	<b>75.38</b>
<b>Ours</b>	347MCI + 184CN	<b>69.03</b>	77.53	57.58	<b>74.45</b>	74.19

The symbol \* means just as a reference since the selected experimental data and data preprocessing processes in these methods are different.

0.64%, and further achieve a notable improvement for AUC, i.e. 4.48%. It can be seen that the challenge in the disease diagnosis is to distinguish adjacent stages. The results report that the diagnostic accuracy of MCI is signally lower than other states. Since MCI is an intermediate stage between normal aging and AD, which does not significantly affect the daily life of patients, it is difficult to diagnose MCI.

The comparison results of AD vs. CN vs. MCI classification are shown in table 6. The accuracy of our method using FDG-PET is about 4% higher than that of ResNet-18. We can also observe that our classification network performs well in almost all criteria, with  $SPE_a$  of 78.32% and  $SPE_i$  of 81.49%, which are 0.94% and 1.93% higher than that of ResNet-18. SPE is an important index used to evaluate the performance of accurate diagnosis of AD, which can be improved by identifying the location and signature features of AD-related abnormalities. Therefore, the

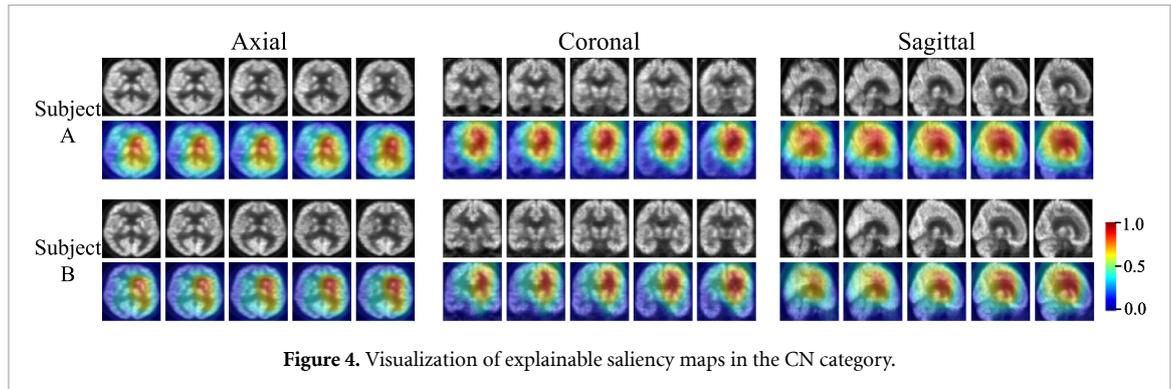
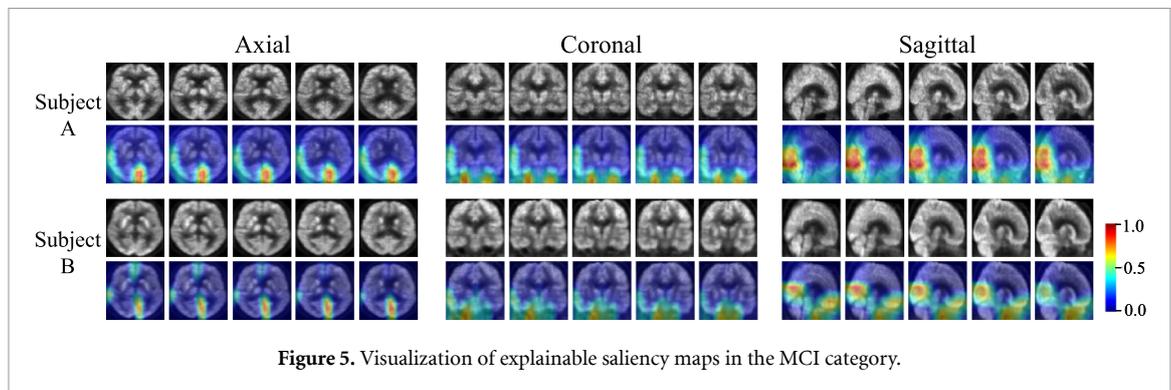
proposed approach by fusing multi-scale regions generally outperforms using full-volume imaging with redundant information in specific performance. In summary, our experimental results indicate that the proposed approach using multi-scale discriminative regions achieves higher performance in AD diagnosis as compared to past studies, irrespective of two-class classification tasks or the multi-class classification task.

### 3.4. Visualization of explainable saliency maps

In AD diagnosis, abnormal areas could be more meaningful for clinicians and researchers. Hence, we use the Score-CAM [43] to generate explainable category-level saliency maps for visualization by weighted summation of saliency maps of various categories in the testing set. Category-level saliency maps could help to more intuitively display the typical regions that contribute to disease

**Table 6.** Performance comparison for AD vs. MCI vs. CN classification. Bold fonts highlight the best performance.

Method	AD vs. MCI vs. CN classification (%)								
	ACC	SEN <sub>a</sub>	SEN <sub>i</sub>	SPE <sub>a</sub>	SPE <sub>i</sub>	AUC <sub>a</sub>	AUC <sub>i</sub>	F1 <sub>a</sub>	F1 <sub>i</sub>
ResNet-18	59.12	<b>59.37</b>	59.12	77.38	79.56	74.60	76.80	59.34	59.12
<b>Ours</b>	<b>62.98</b>	59.28	<b>62.98</b>	<b>78.32</b>	<b>81.49</b>	<b>76.55</b>	<b>78.82</b>	<b>62.67</b>	<b>62.98</b>

**Figure 4.** Visualization of explainable saliency maps in the CN category.**Figure 5.** Visualization of explainable saliency maps in the MCI category.

prediction, simultaneously improving interpretability. Figures 4–6 present the explainable saliency maps grouped by category respectively. To facilitate observation, 3D saliency maps from different subjects are sectioned on three different planes: axial, coronal, and sagittal planes. As there are too many slices, a dimension on the plane is selected as the central dimension and five slices are cut at equal intervals for display.

The saliency maps highlight regions that the classification network is most susceptible to, which show the relevance of each voxel for contributing to the AD diagnosis. The color scale is blue to red, with high values getting the red color and low values getting the blue color. The higher intensity (red) in the saliency maps indicates the higher importance of the region, and some differences between the categories can be represented by comparing the intensity. As can be seen from figure 4, for CN subjects, the regions with high activation are mainly located in the cingulate gyrus, frontal lobe and parietal lobe. The saliency maps for MCI patients indicate that the network pays more attention to the occipital and parietal lobes, which show relatively specificity for differentiating MCI, as presented in figure 5. It is interesting to notice that the regions with high activation in figure 6

are the parietal lobe, temporal lobe, and frontal lobe, suggesting that these regions could serve as distinguishing features in the AD diagnosis. By comparing the explainable saliency maps grouped by category, we can find that the saliency maps in the AD category have a wide range of high activation regions, indicating that metabolic reduction occurs in multiple brain areas. The regions with high activation in the CN category are concentrated and obvious, which is significantly different from MCI and AD. In contrast, the highly activated regions in the MCI category are relatively small, which represents that the metabolic changes are not significant. The findings are consistent with some medical research on AD diagnosis, which lays a good foundation for the proposed method based on multi-scale discriminative regions.

### 3.5. Ablation study for ROI selection

In the proposed method, the adaptive feature fusion of multi-scale regions jointly is employed to improve the diagnosis performance by organically integrating the complementary information. To compare the performance of each region and verify the effectiveness of region fusion, the experimental results achieved from

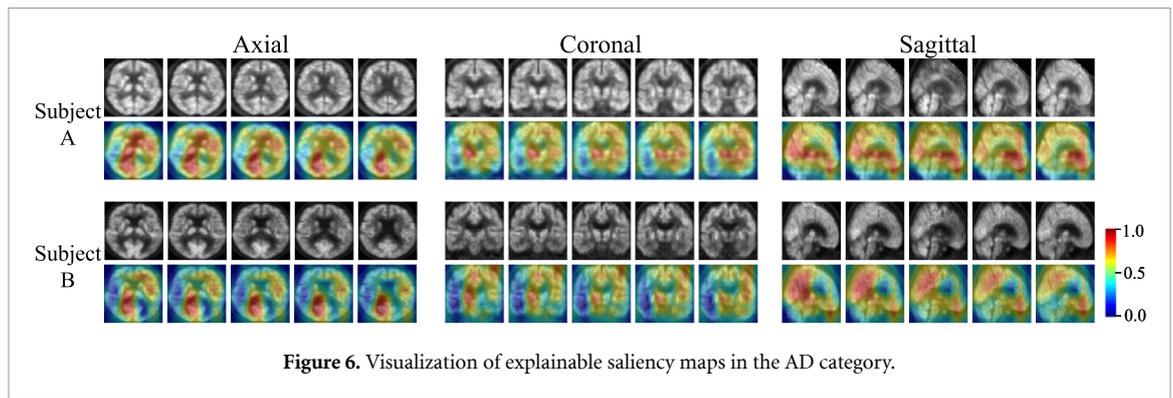


Figure 6. Visualization of explainable saliency maps in the AD category.

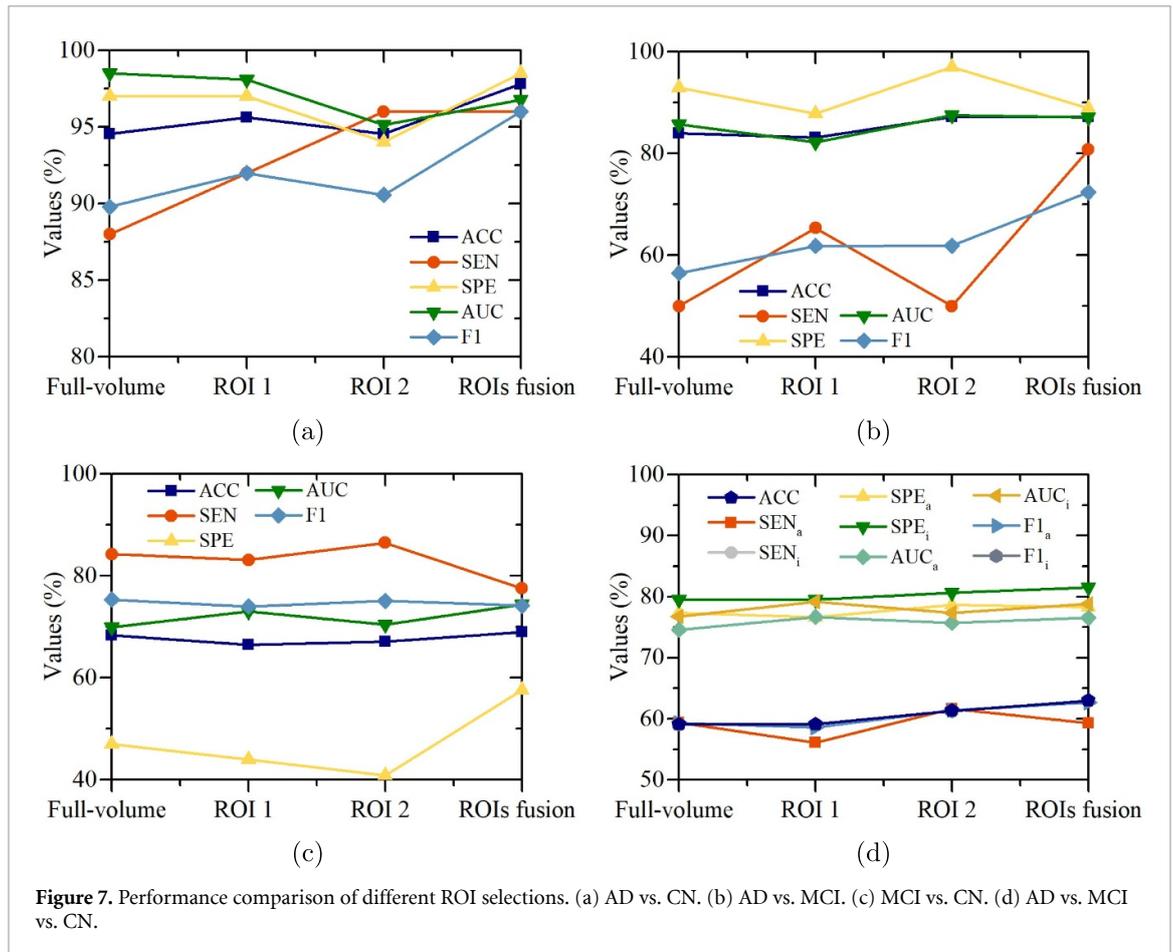


Figure 7. Performance comparison of different ROI selections. (a) AD vs. CN. (b) AD vs. MCI. (c) MCI vs. CN. (d) AD vs. MCI vs. CN.

the full-volume, ROI 1, ROI 2, and ROIs fusion of FDG-PET imaging are reported in figure 7.

For AD vs. CN classification in figure 7(a), our proposed method that fuses multi-scale region representations of FDG-PET images obtains the best performance. Substantial upgrades lie in ACC and F1, particularly with F1, which is gained by 5.43% compared to ROI 2. For AD vs. MCI presented in figure 7(b), the performance of ROIs fusion is significantly enhanced concerning ACC, SEN, and F1. SEN is clearly increased from 50.00% (full-volume) to 80.77% (ROIs fusion), and achieves an improvement of more than 30%. This means that the region fusion approach focuses on positive samples (AD), resulting in an overall performance enhancement. As can

be observed from figure 7(c), for MCI vs. CN classification, the method fusing multi-scale ROIs yields consistently better results than other methods. For example, the ACC and AUC values achieved by our proposed model are 69.03% and 74.45%, respectively, which are far superior to the method using ROI 1 (ACC = 66.45% and AUC = 73.05%). From figure 7(d), we notice that for the multi-class classification, the proposed method reaches optimal performance in almost all criteria. These results confirm that exploiting multi-scale discriminative regions with metabolic reduction can reduce the interference of useless information and thus improve the performance of AD diagnosis. In general, the method using full-volume FDG-PET images yields the worse

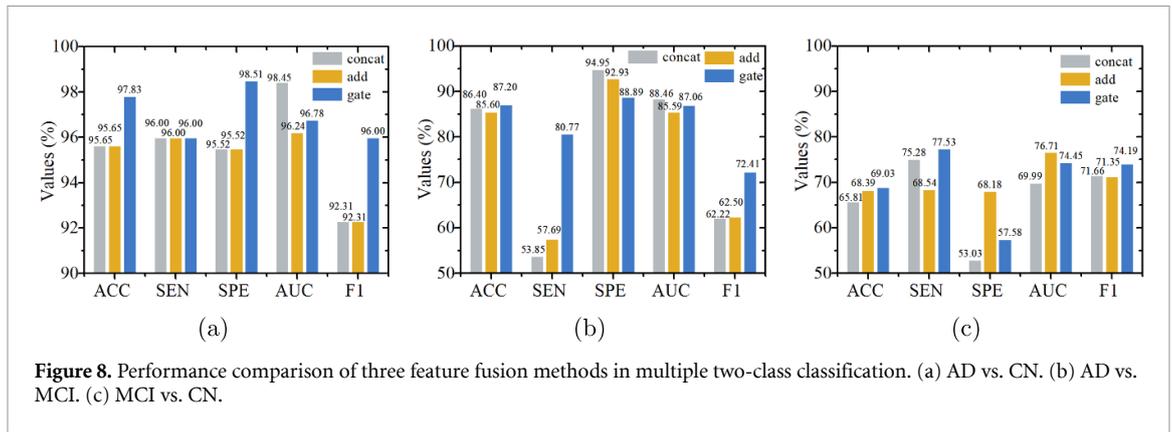


Figure 8. Performance comparison of three feature fusion methods in multiple two-class classification. (a) AD vs. CN. (b) AD vs. MCI. (c) MCI vs. CN.

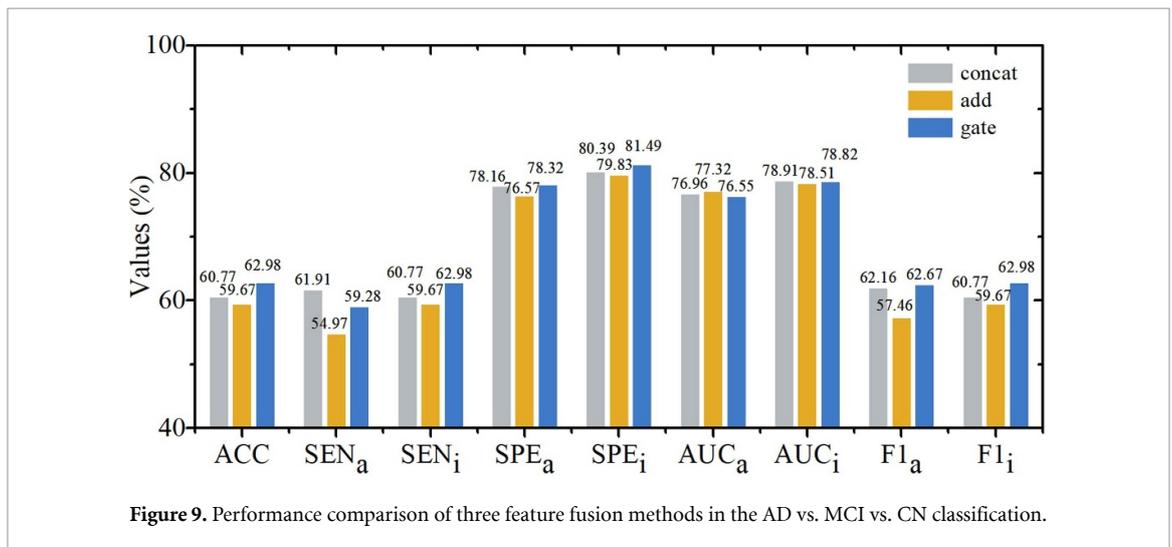


Figure 9. Performance comparison of three feature fusion methods in the AD vs. MCI vs. CN classification.

performance in terms of various criteria. The likely reason is that the network learns disturbing features instead of discriminative features, thus causing overall performance degradation. Furthermore, we can observe that the superiority of using ROI 1 over ROI 2 is particularly obvious in AD vs. CN, while ROI 2 outperforms ROI 1 in most cases for other classification tasks. The results further prove that the parietal lobe and frontal lobe have distinguishing representations for AD diagnosis, and the temporal lobe also has certain specificity. And the attention to the occipital lobe and parietal lobe is valuable for the MCI diagnosis.

### 3.6. Ablation study for region fusion methods

We further investigate the influence of the proposed adaptive feature fusion method on the performance of AD diagnosis. Besides the fusion gate strategy utilized in this paper, namely ‘gate’, there are also two widely applied methods for feature fusion, i.e. (a) the concatenation operation that designates a consolidated dimension, and (b) the addition operation that focuses on the addition of feature maps. Here, we compare the proposed framework with its two variants, i.e. ‘concat’ and ‘add’ that employ the concatenation operation and the addition operation for feature fusion, respectively.

The performance achieved by these feature fusion methods on two-class classification tasks is reported in figure 8. We can find that ‘gate’ is superior to ‘add’ in terms of several criteria, which demonstrates the effectiveness of our adaptive feature fusion for AD diagnosis. Specifically, ‘gate’ rises by 2.18% (ACC), 2.99% (SPE), and 3.69% (F1) in AD diagnosis and 8.99% (SEN), and 2.84% (F1) in MCI diagnosis, compared with ‘add’. Furthermore, the proposed ‘gate’ can achieve satisfactory improvements in overall performance compared to ‘concat’. For instance, in AD vs. CN classification, the ACC and SPE values obtained by ‘gate’ are 97.83% and 98.51%, respectively, which outperform that of ‘concat’ by a large margin (ACC = 95.65%, SPE = 95.52%). The experimental results of AD vs. MCI vs. CN classification are illustrated in figure 9. Among the three fusion methods, the performance of ‘gate’ arrives at the peak point in most metrics, for example, it holds diversities of 2.21% and 1.10% to the corresponding slightly inferior metrics concerning ACC and SPE<sub>i</sub>. The results of ablation experiments clearly demonstrate that the proposed classification method can fuse discriminant information effectively based on the adaptive gate fusion, thereby improving the overall performance of AD diagnosis.

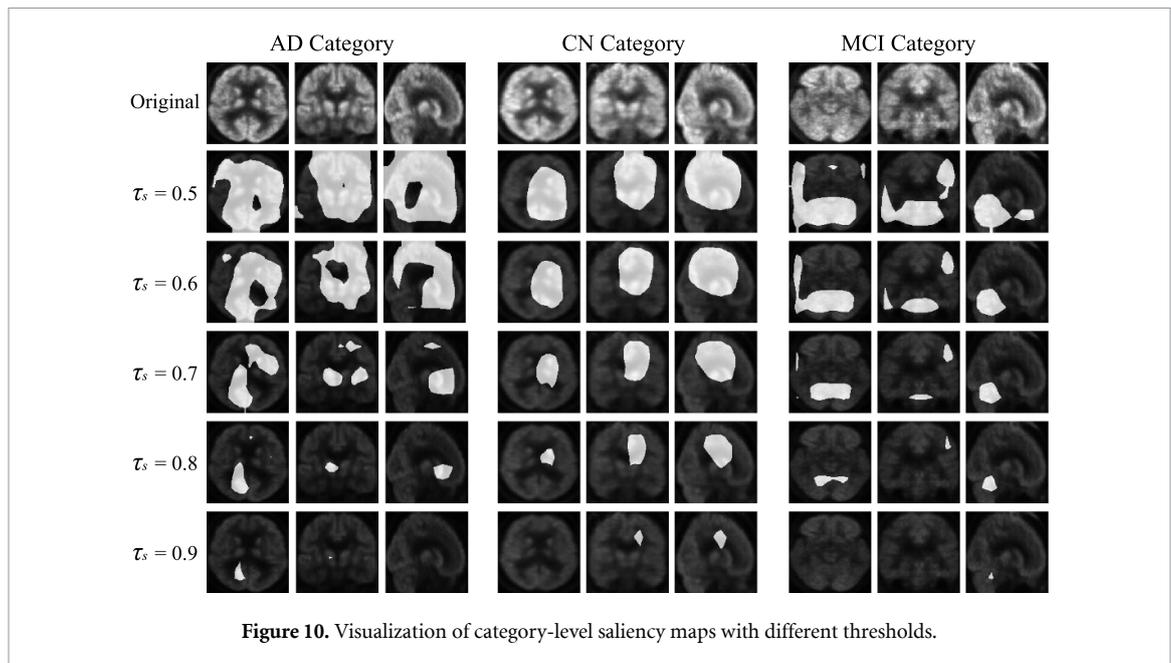


Figure 10. Visualization of category-level saliency maps with different thresholds.

Table 7. Influence of the NMS algorithm for two-class classification tasks. Bold fonts highlight the best performance.

Method	AD vs. CN classification (%)					AD vs. MCI classification (%)					MCI vs. CN classification (%)				
	ACC	SEN	SPE	AUC	F1	ACC	SEN	SPE	AUC	F1	ACC	SEN	SPE	AUC	F1
Full-volume	94.57	88.00	97.01	<b>98.51</b>	89.80	84.00	50.00	92.93	85.74	56.52	68.39	<b>84.27</b>	46.97	69.97	<b>75.38</b>
w/o NMS	94.57	88.00	97.01	96.06	89.80	86.40	61.54	<b>92.93</b>	86.48	65.31	68.39	75.28	<b>59.09</b>	71.43	73.22
<b>with NMS</b>	<b>97.83</b>	<b>96.00</b>	<b>98.51</b>	96.78	<b>96.00</b>	<b>87.20</b>	<b>80.77</b>	88.89	<b>87.06</b>	<b>72.41</b>	<b>69.03</b>	77.53	57.58	<b>74.45</b>	74.19

## 4. Discussion

### 4.1. Analysis of the threshold value $\tau_s$

In the above-mentioned experiments, we exploit a fixed threshold ( $\tau_s = 0.7$ ) for our proposed MSRL method. Given a threshold value  $\tau_s$ , the category-level saliency map that contains values ranging from 0 to 1 can be divided into background and foreground. In order to investigate the selection of the appropriate threshold value aiming to make the foreground regions concentrated in disease-related regions, multiple values uniformly distributed from 0.5 to 0.9 with a step size of 0.1 are employed for comparative experiments.

We report the impact of adopting different threshold values to category-level saliency maps as illustrated in figure 10. With  $\tau_s$  raising, the foreground areas (in white) become lesser and more pronounced, while the background areas (in black) increase substantially at the expense of splitting the foreground areas and losing more pixels. As can be seen from figure 10, at  $\tau_s = 0.5$  and  $\tau_s = 0.6$ , the foreground area is large and still contains a lot of useless information. Subtle anomalies within the large areas may be overwritten by uninformative normal areas. Also, utilizing large regions could bring a huge computational burden resulting in affecting the practicability of the proposed framework in applications.

In addition, when  $\tau_s$  is small, the distribution of foreground area is scattered, which is detrimental to the localization of discriminant regions. Moreover, at  $\tau_s = 0.8$  and  $\tau_s = 0.9$ , the foreground area is too small, which means it does not have enough ability to capture discrimination information from the brain. Therefore, we select  $\tau_s = 0.7$  as the threshold value, which is conducive to the further extraction of discriminative regions with appropriate sizes.

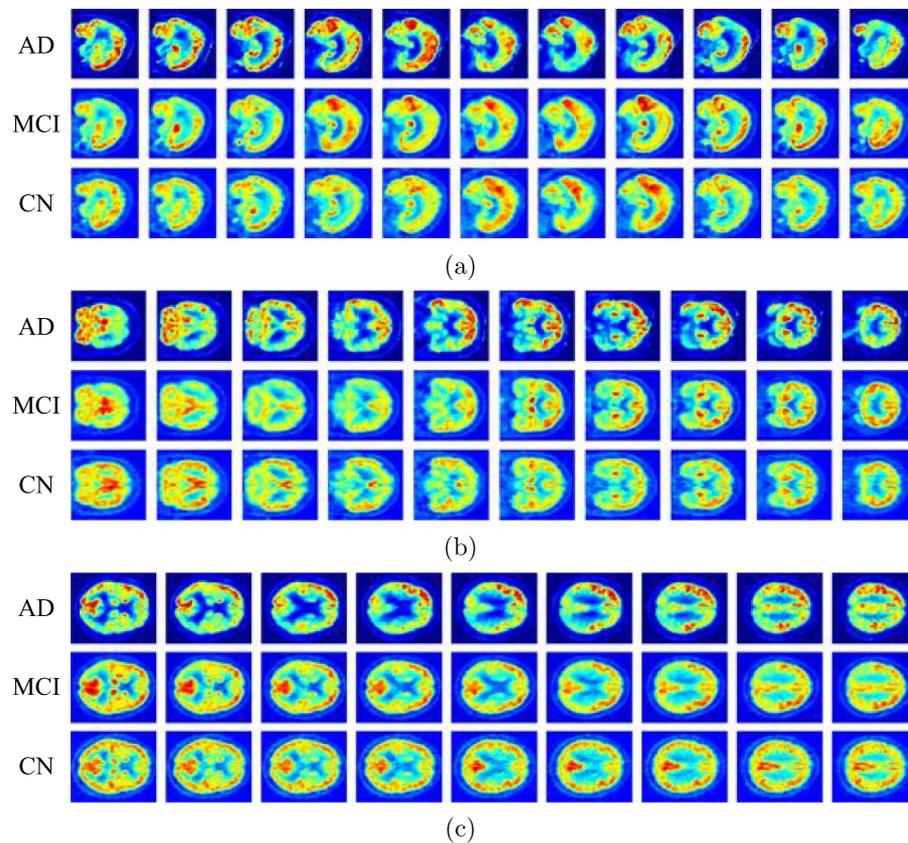
### 4.2. Influence of the NMS algorithm

We further investigate the influence of using the NMS algorithm on the diagnosis performances achieved by our proposed method. The proposed method using NMS indicated ‘with NMS’, is compared with its variant without NMS, denoted ‘w/o NMS’.

The experimental results as shown in table 7 demonstrate that in two-class classification tasks, ‘w/o NMS’ yields consistently worse performance compared to ‘with NMS’ in almost all criteria, potentially because (a) there is redundancy in the learned information, which is easy to lead to overfitting, and (b) regions with relatively low scores could affect the overall performance. For instance, for AD vs. CN classification, ACC and SEN values achieved by ‘with NMS’ are 97.83% and 96.00%, which are much better than that of ‘w/o NMS’. Furthermore, we conduct comparative experiments on AD vs. MCI

**Table 8.** Influence of the NMS algorithm for AD vs. MCI vs. CN classification. Bold fonts highlight the best performance.

Method	AD vs. MCI vs. CN classification (%)								
	ACC	SEN <sub>a</sub>	SEN <sub>i</sub>	SPE <sub>a</sub>	SPE <sub>i</sub>	AUC <sub>a</sub>	AUC <sub>i</sub>	F1 <sub>a</sub>	F1 <sub>i</sub>
Full-volume	59.12	59.37	59.12	77.38	79.56	74.60	76.80	59.34	59.12
w/o NMS	60.77	<b>63.25</b>	60.77	<b>78.86</b>	80.39	<b>79.28</b>	<b>80.61</b>	62.09	60.77
<b>with NMS</b>	<b>62.98</b>	59.28	<b>62.98</b>	78.32	<b>81.49</b>	76.55	78.82	<b>62.67</b>	<b>62.98</b>

**Figure 11.** The FDG-PET images in three planes of different categories. From top to bottom are from the AD patient, MCI patient, and CN subject. (a) Sagittal plane. (b) Coronal plane. (c) Axial plane.

vs. CN classification task to comprehensively verify the effectiveness of the NMS algorithm, with results shown in table 8. We can observe that the proposed method generally outperforms ‘w/o NMS’ in multi-class classification. The ACC value achieved by ‘with NMS’ is 62.98%, which is higher than that obtained by ‘w/o NMS’. The possible reason is that the NMS algorithm in this study selects regions with high scores (dense distribution of discriminative features) and suppresses redundant regions with low scores. Besides, the NMS algorithm can reduce irrelevant and redundant features and decrease the computational complexity and memory cost during the training. This further implies that using more ROIs cannot drastically enhance the performance of AD diagnosis, but brings about a growth in network parameters. In summary, these experimental results further prove that the NMS algorithm in this study could yield a positive influence on the AD diagnosis.

### 4.3. Analysis of FDG-PET imaging in AD diagnosis

Recently, FDG-PET imaging has made considerable contributions to the development of clinical and methodological research about AD diagnosis. It demonstrates the level of glucose metabolism in the brain, which enables high efficiency in forecasting the occurrence and development of AD as an independent influencing factor. To verify the effectiveness of the proposed MSRL method, slices sectioned on three planes (axial, coronal, and sagittal planes) of FDG-PET images from different categories after pseudo-color processing are displayed in figure 11 to compare the glucose metabolic level in different states.

It is generally believed that normal elderly people have normal glucose metabolism, and 18F-FDG is symmetrically distributed in bilateral frontal, temporal, and parietal lobes, as shown in figure 11. Reductions of cerebral metabolism are well established in AD [16, 17, 51]. Compared with CN

subjects, the metabolism in the temporal-parietal lobe and posterior cingulate gyrus recedes in typical patients with AD. And as AD progresses, the frontal lobe might also be affected [52]. Changes in cerebral metabolism also have been detected in MCI in many studies [18, 19]. As can be seen from figure 11, MCI patients have a decrease in cerebral glucose metabolism compared to CN subjects, with a characteristic regional pattern of the occipital lobe and temporal-parietal lobe hypometabolism. In comparison with MCI patients, the range and degree of reduced FDG metabolism are obviously ascended in AD patients. The multi-scale discriminative regions we selected are consistent with previous medical studies, which further confirms the effectiveness of the proposed method.

## 5. Conclusion

In this paper, we propose a novel deep learning method based on multi-scale discriminative regions in FDG-PET imaging to deal with the early diagnosis of AD. First, the disease-affected regions are located in an unsupervised manner, the priority of which is then evaluated by the designed confidence score to further select the optimal multi-scale discriminative regions. Then, the proposed MSRC module is applied to adaptively fuse feature representations of discriminative regions, which greatly reduces the interference information to a great extent and boosts the interpretability of AD diagnosis. Next, decision-level fusion is introduced to further improve diagnostic performance. We evaluate the proposed framework on multiple AD-related classification tasks, and the experimental results on the public ADNI dataset reveal that it is with the great generalization ability and superior performance. The resultant multi-scale regions from our approach highlight some disease-relevant regions of the cerebral cortex, consistent with the commonly affected regions during AD development, further demonstrating the effectiveness of our proposed method.

## Data availability statement

The data that support the findings of this study is available in <http://adni.loni.usc.edu/>, available upon reasonable request from the authors.

## Acknowledgments

The research in our paper is sponsored by Chengdu Major Technology Application Demonstration Project (2019-YF09-00120-SN), the Sichuan Science and Technology Program (No. 2021YFS0239), and the Fundamental Research Funds for the Central Universities (No. 2021SCU12061). The authors thank

the ADNI investigators for providing access to the data. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. For more information about ADNI, see [www.adni-info.org](http://www.adni-info.org). Data collection and sharing for this project was funded by the ADNI (National Institutes of Health, Grant No. U01 AG024904) and DOD ADNI (Department of Defense, Award No. W81XWH-12-2-0012).

## ORCID iD

Xiaohai He  <https://orcid.org/0000-0001-8399-3172>

## References

- [1] Tiwari S, Atluri V, Kaushik A, Yndart A and Nair M 2019 Alzheimer's disease: pathogenesis, diagnostics and therapeutics *Int. J. Nanomed.* **14** 5541–54
- [2] Zhan L *et al* 2015 Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease *Front. Aging Neurosci.* **7** 48
- [3] Zhu X, Suk H-I, Zhu Y, Thung K-H, Wu G and Shen D 2015 Multi-view classification for identification of Alzheimer's disease *Int. Workshop on Machine Learning in Medical Imaging* (Springer) pp 255–62
- [4] Silveira M and Marques J 2010 Boosting Alzheimer disease diagnosis using pet images *2010 20th Int. Conf. on Pattern Recognition* (IEEE) pp 2556–9
- [5] Anonymous 2021 2021 Alzheimer's disease facts and figures *Alzheimer's Dementia* **17** 327–406
- [6] Zhao L 2020 2020 Alzheimer's disease facts and figures *Alzheimer's Dementia* **16** 391–460
- [7] Liu M, Li F, Yan H, Wang K, Ma Y, Shen Li, Shen L and Xu M (Alzheimer's Disease Neuroimaging Initiative) 2020 A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease *NeuroImage* **208** 116459
- [8] Zhang Li, Wang M, Liu M and Zhang D 2020 A survey on deep learning for neuroimaging-based brain disorder analysis *Front. Neurosci.* **14** 779
- [9] Zhang Q, Du Q and Liu G 2021 A whole-process interpretable and multi-modal deep reinforcement learning for diagnosis and analysis of Alzheimer's disease *J. Neural Eng.* **18** 066032
- [10] Zhou T, Liu M, Thung K-H and Shen D 2019 Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data *IEEE Trans. Med. Imaging* **38** 2411–22
- [11] Lian C, Liu M, Zhang J and Shen D 2020 Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI *IEEE Trans. Pattern Anal. Mach. Intell.* **42** 880–93
- [12] Brand L, Nichols K, Wang H, Shen L and Huang H 2019 Joint multi-modal longitudinal regression and classification for Alzheimer's disease prediction *IEEE Trans. Med. Imaging* **39** 1845–55
- [13] Ebrahimighahnavieh M A, Luo S and Chiong R 2020 Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review *Comput. Methods Programs Biomed.* **187** 105242
- [14] Li J, Bian C, Luo H, Chen D, Cao L and Liang H (Alzheimer's Disease Neuroimaging Initiative) 2021 Multi-dimensional persistent feature analysis identifies connectivity patterns of resting-state brain networks in Alzheimer's disease *J. Neural Eng.* **18** 016012

- [15] Liu Z, Lu H, Pan X, Xu M, Lan R and Luo X 2022 Diagnosis of Alzheimer's disease via an attention-based multi-scale convolutional neural network *Knowl.-Based Syst.* **238** 107942
- [16] Jack C R Jr et al 2018 NIA-AA research framework: toward a biological definition of Alzheimer's disease *Alzheimer's Dementia* **14** 535–62
- [17] Ou Y-N et al 2019 FDG-PET as an independent biomarker for Alzheimer's biological diagnosis: a longitudinal study *Alzheimer's Res. Ther.* **11** 1–11
- [18] Landau S M, Harvey D, Madison C M, Koeppe R A, Reiman E M, Foster N L, Weiner M W and Jagust W J (Alzheimer's Disease Neuroimaging Initiative) 2011 Associations between cognitive, functional and FDG-PET measures of decline in AD and MCI *Neurobiol. Aging* **32** 1207–18
- [19] Apostolova I, Lange C, Suppa P, Spies L, Klutmann S, Adam G, Grothe M J and Buchert R 2018 Impact of plasma glucose level on the pattern of brain FDG uptake and the predictive power of FDG PET in mild cognitive impairment *Eur. J. Nucl. Med. Mol. Imaging* **45** 1417–22
- [20] Padilla P, López M, Górriz J M, Ramirez J, Salas-Gonzalez D and Alvarez I 2011 NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease *IEEE Trans. Med. Imaging* **31** 207–16
- [21] Cheng B, Liu M, Zhang D, Munsell B C and Shen D 2015 Domain transfer learning for MCI conversion prediction *IEEE Trans. Biomed. Eng.* **62** 1805–17
- [22] Gray K R, Aljabar P, Heckemann R A, Hammers A and Rueckert D (Alzheimer's Disease Neuroimaging Initiative) 2013 Random forest-based similarity measures for multi-modal classification of Alzheimer's disease *NeuroImage* **65** 167–75
- [23] Cabral C, Morgado P M, Costa D C and Silveira M (Alzheimer's Disease Neuroimaging Initiative) 2015 Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages *Comput. Biol. Med.* **58** 101–9
- [24] Garali Iène, Adel M, Bourennane S and Guedj E 2018 Histogram-based features selection and volume of interest ranking for brain PET image classification *IEEE J. Trans. Eng. Health Med.* **6** 1–12
- [25] Pan X, Adel M, Fossati C, Gaidon T and Guedj E 2018 Multilevel feature representation of FDG-PET brain images for diagnosing Alzheimer's disease *IEEE J. Biomed. Health Inf.* **23** 1499–506
- [26] Gray K R, Wolz R, Heckemann R A, Aljabar P, Hammers A and Rueckert D (Alzheimer's Disease Neuroimaging Initiative) 2012 Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease *NeuroImage* **60** 221–9
- [27] Lu D, Popuri K, Ding G W, Balachandar R and Beg M F (Alzheimer's Disease Neuroimaging Initiative) 2018 Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease *Med. Image Anal.* **46** 26–34
- [28] Yee E, Popuri K and Beg M F (Alzheimer's Disease Neuroimaging Initiative) 2020 Quantifying brain metabolism from FDG-PET images into a probability of Alzheimer's dementia score *Human Brain Mapp.* **41** 5–16
- [29] Singh S, Srivastava A, Mi L, Caselli R J, Chen K, Goradia D, Reiman E M and Wang Y 2017 Deep-learning-based classification of FDG-PET data for Alzheimer's disease categories *Proc. SPIE* **10572** 143–58
- [30] Etmnani K et al 2021 A 3D deep learning model to predict the diagnosis of dementia with Lewy bodies, Alzheimer's disease and mild cognitive impairment using brain 18F-FDG PET *Eur. J. Nucl. Med. Mol. Imaging* **49** 1–22
- [31] Liu M, Cheng D and Yan W (Alzheimer's Disease Neuroimaging Initiative) 2018 Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images *Front. Neuroinform.* **12** 35
- [32] Ding Y et al 2019 A deep learning model to predict a diagnosis of Alzheimer disease by using <sup>18</sup>F-FDG PET of the brain *Radiology* **290** 456–64
- [33] Pan X, Phan T-L, Adel M, Fossati C, Gaidon T, Wojak J and Guedj E 2020 Multi-view separable pyramid network for AD prediction at MCI stage by <sup>18</sup>F-FDG brain PET imaging *IEEE Trans. Med. Imaging* **40** 81–92
- [34] Islam J and Zhang Y 2019 Understanding 3D CNN behavior for Alzheimer's disease diagnosis from brain PET scan (arXiv:1912.04563)
- [35] Huang Y, Xu J, Zhou Y, Tong T and Zhuang X (Alzheimer's Disease Neuroimaging Initiative (ADNI)) 2019 Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network *Front. Neurosci.* **13** 509
- [36] Guo J, Qiu W, Li X, Zhao X, Guo N and Li Q 2019 Predicting Alzheimer's disease by hierarchical graph convolution from positron emission tomography imaging 2019 *IEEE Int. Conf. on Big Data (Big Data)* (IEEE) pp 5359–63
- [37] Jack C R Jr et al 2008 The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods *J. Magn. Reson. Imaging* **27** 685–91
- [38] Wen J et al 2020 Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation *Med. Image Anal.* **63** 101694
- [39] Fung Y R, Guan Z, Kumar R, Wu J Y and Fiterau M 2019 Alzheimer's disease brain MRI classification: challenges and insights (arXiv:1906.04231)
- [40] Bäckström K, Nazari M, Gu I Y-H and Jakola A S 2018 An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images 2018 *IEEE 15th Int. Symp. on Biomedical Imaging (ISBI 2018)* (IEEE) pp 149–53
- [41] Routier A et al 2021 Clinica: an open-source software platform for reproducible clinical neuroscience studies *Front. Neuroinform.* **15** 689675
- [42] Fischl B 2012 Freesurfer *NeuroImage* **62** 774–81
- [43] Wang H, Wang Z, Du M, Yang F, Zhang Z, Ding S, Mardziel P and Hu X 2020 Score-cam: score-weighted visual explanations for convolutional neural networks *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops* pp 24–25
- [44] Felzenszwalb P F, Girshick R B, McAllester D and Ramanan D 2009 Object detection with discriminatively trained part-based models *IEEE Trans. Pattern Anal. Mach. Intell.* **32** 1627–45
- [45] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8
- [46] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)
- [47] Shen T, Jiang J, Lu J, Wang M, Zuo C, Yu Z and Yan Z 2019 Predicting Alzheimer disease from mild cognitive impairment with a deep belief network based on 18F-FDG-PET images *Mol. Imaging* **18** 1536012119877285
- [48] Li R, Perneczky R, Yakushev I, Foerster S, Kurz A, Drzezga A and Kramer S (Alzheimer's Disease Neuroimaging Initiative) 2015 Gaussian mixture models and model selection for [18F] fluorodeoxyglucose positron emission tomography classification in Alzheimer's disease *PLoS One* **10** e0122731
- [49] Pan X, Adel M, Fossati C, Gaidon T, Wojak J and Guedj E (Alzheimer's Disease Neuroimaging Initiative) 2019 Multiscale spatial gradient features for <sup>18</sup>F-FDG PET image-guided diagnosis of Alzheimer's disease *Comput. Methods Programs Biomed.* **180** 105027
- [50] Cui W, Yan C, Yan Z, Peng Y, Leng Y, Liu C, Chen S and Jiang Xi 2021 Bilinear pooling and metric learning network for early Alzheimer's disease identification with FDG-PET images (arXiv:2111.04985)

- [51] Demey I, Falasco G, Mendez P A C, Urrutia L, Bérnago Y, de Leon M P, Sevlever G, Allegri R F and Vazquez S 2021 A comparative study of 18F-FDG PET brain metabolism and cognitive performance between normal controls and amnesic Alzheimer's disease patients *Alzheimer's Dementia* **17** e050083
- [52] Jing J, Zhang F, Zhao Li, Xie J, Chen J, Zhong R, Zhang Y and Dong C 2021 Correlation between brain <sup>18</sup>F-AV45 and 18F-FDG PET distribution characteristics and cognitive function in patients with mild and moderate Alzheimer's disease *J. Alzheimer's Dis.* **79** 1317–25