Computers in Biology and Medicine 157 (2023) 106788

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Multimodal cross enhanced fusion network for diagnosis of Alzheimer's disease and subjective memory complaints

Yilin Leng [a,b], Wenju Cui [b,c], Yunsong Peng [d], Caiying Yan [e], Yuzhu Cao [b,c], Zhuangzhi Yan [a], Shuangqing Chen [e,*], Xi Jiang [f,*], Jian Zheng [b,c,**], the Alzheimer's Disease Neuroimaging Initiative

[a] Institute of Biomedical Engineering, School of Communication and Information Engineering, Shanghai University, Shanghai, 200444, China
[b] Department of Medical Imaging, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, 215163, China
[c] School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, 230026, China
[d] Department of Medical Imaging, International Exemplary Cooperation Base of Precision Imaging for Diagnosis and Treatment, Guizhou Provincial People's Hospital, Guizhou, 550002, China
[e] Department of Radiology, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, 211103, China
[f] Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, 611731, China

ARTICLE INFO

Keywords:
Alzheimer's disease (AD) diagnosis
Multiscale long-range receptive field
Cross enhanced fusion
Subjective memory complaints (SMC) diagnosis

ABSTRACT

Deep learning methods using multimodal imagings have been proposed for the diagnosis of Alzheimer's disease (AD) and its early stages (SMC, subjective memory complaints), which may help to slow the progression of the disease through early intervention. However, current fusion methods for multimodal imagings are generally coarse and may lead to suboptimal results through the use of shared extractors or simple downscaling stitching. Another issue with diagnosing brain diseases is that they often affect multiple areas of the brain, making it important to consider potential connections throughout the brain. However, traditional convolutional neural networks (CNNs) may struggle with this issue due to their limited local receptive fields. To address this, many researchers have turned to transformer networks, which can provide global information about the brain but can be computationally intensive and perform poorly on small datasets. In this work, we propose a novel lightweight network called MENet that adaptively recalibrates the multiscale long-range receptive field to localize discriminative brain regions in a computationally efficient manner. Based on this, the network extracts the intensity and location responses between structural magnetic resonance imagings (sMRI) and 18-Fluoro-Deoxy-Glucose Positron Emission computed Tomography (FDG-PET) as an enhancement fusion for AD and SMC diagnosis. Our method is evaluated on the publicly available ADNI datasets and achieves 97.67% accuracy in AD diagnosis tasks and 81.63% accuracy in SMC diagnosis tasks using sMRI and FDG-PET. These results achieve state-of-the-art (SOTA) performance in both tasks. To the best of our knowledge, this is one of the first deep learning research methods for SMC diagnosis with FDG-PET.

1. Introduction

Alzheimer's Disease (AD) is a common neurodegenerative condition that eventually leads to irreversible neuronal injury. Due to its irreversible nature [1], it is increasingly important to identify AD in its early stages. Subjective Memory Complaints (SMC) are often the first stage of AD, and are therefore an important focus of research [2]. Many studies [3–5] have found that a range of neuroimaging biomarkers can be used to diagnosis both AD and SMC. Structural magnetic resonance imaging (sMRI) is a noninvasive method that can detect high-resolution structural changes in the brain caused by atrophy, such as changes in thickness, volume, shape, and texture [6]. Similarly, 18-Fluoro-Deoxy-Glucose Positron Emission computed Tomography (FDG-PET) uses radioactive tracers to track cerebral metabolic rate of glucose reflecting hemodynamic and detect changes in brain function [7].

Recently, Deep learning methods have demonstrated excellent performance on high-dimensional complex data, and have been widely

* Corresponding authors.
** Corresponding author at: School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, 230026, China.
E-mail addresses: keai77lyl@shu.edu.cn (Y. Leng), sznaonao@163.com (S. Chen), xijiang@uestc.edu.cn (X. Jiang), zhengj@sibet.ac.cn (J. Zheng).

https://doi.org/10.1016/j.compbiomed.2023.106788
Received 29 December 2022; Received in revised form 9 February 2023; Accepted 11 March 2023
Available online 15 March 2023

applied to the diagnosis of AD [6]. These methods typically involve three basic steps: (1) identifying regions of interest (ROI), (2) extracting features, and (3) constructing classification models. In the ROI identifying step, brain imaging data is often automatically divided into multiple brain regions using the generic template AAL (Automated Anatomical Labeling) [6], or determined manually based on relevant brain regions in anatomy [8]. For feature extraction, existing networks can be classified into 2D-based, 3D-based, and transformer-based networks according to their architecture. 2D-based networks perform classification by extracting voxel-level and slice-level features with a small number of network parameters, but may result in a loss of spatial information [9,10]. 3D-based networks construct 3D convolutions to natively extract spatial contextual information from original images, but are limited by the local receptive field of convolutions, which can make it challenging to establish long-range dependencies on inter-regional connections [11,12]. Transformer-based networks can obtain arbitrary receptive field through patch embedding and attention mechanisms, but require significant computational effort and may not perform optimally on small datasets [12,13]. For model construction, most existing deep learning methods combine feature extraction and classification into an end-to-end network. In the following, we will present deep learning methods for single-modality and multi-modality based on sMRI and FDG-PET for the diagnosis of AD and SMC.

Based on sMRI, Chen et al. [10] extracted multi-view slice features and global structural features using multiple slice-level and subject-level subnetworks, which ignored complex spatial information. Liu et al. [9] proposed a joint learning multi-task network that performed hippocampus segmentation and then extracted corresponding features for classification. This manual extraction of regional features reduces the number of learnable parameters in the network, but ignores global information, such as minor atrophy throughout the brain. In addition, Li et al. [14] divided 3D images into patches and then clustered them into multiple densenets for training. Lian et al. [15,16] extracted discriminative regions using weak supervision and combined them with a hybrid multi-level network. These methods use 3D raw images to obtain rich contextual information with certain interpretability, but either rely heavily on prior knowledge or perform localization and diagnosis independently. Therefore, we aim to provide an end-to-end task-oriented network that expands the receptive field to establish long-range dependencies for AD and SMC classification and the localization of discriminative regions.

As for FDG-PET, Cui et al. [17] extracted the radiological features of brain regions delineated by the AAL template, and enhanced the connections between these regions by incorporating a bilinear pooling mechanism into their network. Guo et al. [18] designed a hierarchical graph convolutional network (GNN) to overcome the limitation of Euclidean distance, and used the corresponding features of brain regions delineated by the AAL template as graph nodes for clustering and classification. These methods rely on prior knowledge to simplify high-dimensional images with a small number of model parameters, but this approach can lead to the loss of a significant amount of original spatial information and can be difficult as methods on sMRI to interpret for localizing discriminative regions. To address these issues, Pan et al. [11] used separable convolution to learn representations from axial, coronal, and sagittal views from slice-wise to spatial-wise successively, which preserves spatial information and reduces the number of training parameters compared to 2D and 3D networks. Islam et al. [19] used a simple 3D network to classify PET images and visualize the whole brain. Yee et al. [20] proposed a 3D network with residual connections. However, these networks often have large numbers of parameters, which make them prone to overfitting.

In addition to using a single modality, several studies [21,22] have demonstrated that combining multi-modality data, such as sMRI and FDG-PET can help diagnosis of AD. Huang et al. [21] selected the hippocampal region as the ROI in both sMRI and FDG-PET modalities, and trained them on two separate VGG-11 networks which then flattened and directly concatenated. Liu et al. [23] developed a cascaded deep CNN to learn multi-level and multimodal features, dividing the original images into multiple equal-sized patches for feature extraction, and then concatenating them for classification. These approaches enable interaction between the two modalities through the use of a shared feature extractor or simple dimensionality reduction and concatenation. However, these methods make it challenging to visualize features for interpretation and lead to adequately consider the correlation, complementarity, and heterogeneity between different modalities by treating the two modalities equally and simply fusing the features.

To sum up, we propose a novel multimodal cross enhanced fusion network with a multiscale long-range receptive field for the diagnosis of AD and SMC. Specifically, we uniformly divide the 3D image into patches and operate directly on them as input. We then extract features using a multiscale long-range receptive fields and separate the mixture of spatial and channel dimensions guided by channel-attention. Finally, the spatial and channel responses between the features of sMRI and FDG-PET are complementarily enhanced and fused by our efficient Cross Enhanced Fusion mechanism. In addition, we use patch-level localization of discriminative information for clinical diagnosis.

Our main contributions are as follows:

1. We propose a patch-based efficient 3D Multimodal cross Enhanced fusion Network (MENet) that can localize discriminative regions and diagnose AD and SMC without prior knowledge, and it performs highly competitively. To the best of our knowledge, this is one of the first studies to classify SMC vs. NC with FDG-PET using deep a learning approach.
2. We design a multiscale long-range reception module (MLR) that dynamically recalibrates the cross-dimensional feature weights guided by channel attention. This mechanism allows us to obtain features at various scales without increasing the channel dimension and computational complexity.
3. We propose a cross enhanced fusion mechanism to emphasize the correlation and complementarity between features of sMRI and FDG-PET, which helps to enhance local discriminative capability on the patch-level.

The rest of this paper is organized as follows. Section 2 introduces the materials we used. Section 3 introduces the proposed method in detail. In Section 4, we compare our method with previous studies, conduct the ablation study, verify the effect of region localization and discuss the limitations. Finally, this paper is concluded in Section 5.

2. Materials

2.1. Data acquisition

The data we used in this work are from the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) database. We select the 1.5T and 3T T1-weighted sMRI data and the 18-F FDG-PET data which are obtained in a state of rest with 30–35 min with 185 ± 18.5 MBq FDG. These data are from 536 subjects, including 254 NC subjects, 98 SMC subjects, and 184 AD subjects. Note that for subjects that appear in multiple datasets of ADNI, we only keep one of them. Table 1 summarizes the demographic and clinical information of the subjects in the dataset.

2.2. Image preprocessing

In our work, the sMRI and FDG-PET data are from various sites, in which the sMRI data are collected using a variety of scanners with

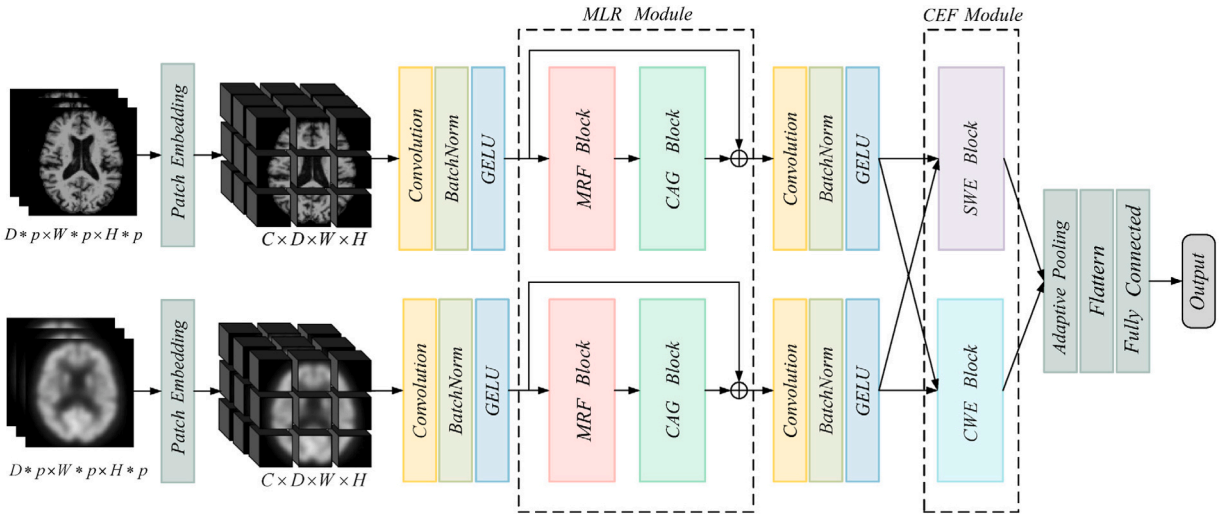

Fig. 1. The architecture of MENet. The MENet consists of two symmetrical branches and a fusion branch. The sMRI and FDG-PET images are input into the MLR module to adaptively extract features in the multiscale receptive field, the CEF module extracts the response across sMRI and PDG-PET modalities for cross enhanced fusion to obtain the final classification result.

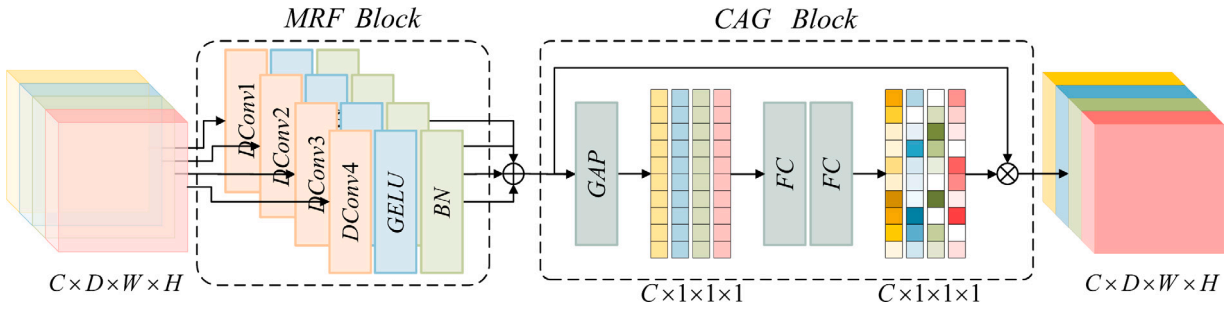

Fig. 2. The architecture of Multiscale Long-range Receptive Field Module. This module consists of an MRF and a CAG block. The MRF block is responsible for efficiently acquiring multi-scale receptive field information, while the CAG block is responsible for adaptive calibrating channel-wise information to skew available computing resources to relatively important information. DConv$i$ denotes depth-wise convolutions with kernel size k_i.

Table 1
Demographic information of the subjects. The gender is presented as male/female, Age, Mini-Mental State Examination (MMSE) scores, and education years are presented as mean ± standard deviation.

Type	Gender(M/F)	Age(years)	MMSE	Education
NC	168/87	74.48 ± 6.48	26.60 ± 3.78	15.78 ± 3.02
SMC	70/31	74.94 ± 6.11	26.39 ± 3.48	15.19 ± 2.98
AD	116/70	74.10 ± 7.15	26.77 ± 3.64	16.19 ± 2.99

protocols customized to each scanner, while the same patient has more than one FDG-PET data. ADNI reviews the sMRI data and corrects them by B1 field inhomogeneity and gradient nonlinearity. We process these data as follows, for sMRI: (1) motion correction and conform; (2) Non-Uniform intensity normalization by N3 algorithm [24]; (3) Talairach transform computation; (4) Intensity normalization; (5) Skull stripping and Affine registration by using FreeSurfer (https://fsl.fmrib.ox.ac.uk/); then (6) spatial normalization to the Montreal Neurological Institute (MNI) space with the resolution of 3mm×3mm×3mm using the Statistical Parametric Mapping (SPM) [25]; (7) smooth with 8 mm by MATLAB2020a. For each FDG-PET data, we (1) remove the data with head movement over 2 mm, and average the multiple PET images of the same patient; (2) linearly align with its corresponding sMRI scan; (3) intensity normalize by min–max scaling; (5) spatial normalize to the MNI space; (6) convert to uniform isotropic resolution by SPM12 with an 8 mm full-width half-maximum (FWHM) Gaussian filter. Finally, the size of sMRI and FDG-PET images is 91 × 109 × 91.

3. Methods

3.1. Architecture

As shown in Fig. 1, the MENet mainly consists of a Multiscale Long-range Reception (MLR) module and a Cross Enhanced Fusion (CEF) module. Specifically, the preprocessed data are first divided into non-overlapping patches as inputs, then our MENet extracts features in different receptive fields and uses attention to recalibrate feature weights to obtain information that contributes to the diagnosis. Finally, the structural and metabolic features of sMRI and FDG-PET are complementarily enhanced and fused by the CEF module for the diagnosis of AD and SMC.

3.2. Multiscale long-range reception module

The module described in Fig. 2 is inspired by ConvmixerNet [14] and consists of a patch embedding layer followed by two novel blocks: the Multiscale Receptive Field (MRF) block and the Channel-Attention-Guided (CAG) block. Unlike other methods that pre-divide images into patches, we use a simple convolution to embed the images into patches of size p, embedding dimension d. This is achieved by using a convolution with c input channels, d output channels, kernel size p, and stride p:

$$\begin{aligned} X_0 &= F_{patch}(X_{in}) \\ &= BN(\sigma\{Conv_c(X_{in}, sride = p, kernel\_size = p)\}) \end{aligned} \tag{1}$$

3.2.1. Multiscale receptive field (MRF) block

MLPs [26] and self-attention [27] mechanisms are able to mix distant spatial location information, indicating that they have receptive fields of arbitrary size. While these approaches may be more flexible and structurally more suitable for establishing long-range dependencies to obtain a large receptive field, CNN-based approaches with inductive bias tend to show higher accuracy when training models on smaller datasets [28]. Recent studies [29,30] have shown that integrating learning mechanisms into the network can help capture spatial correlations among features and enhance the representation capability of networks.

Motivated by these findings, we introduce a multiscale receptive field (MRF) block in our network to extract rich multiscale spatial features from different receptive field maps and overcome the limitation of local information in establishing long-range dependencies. Specifically, we build a subnetwork with four branches, each containing a depthwise convolution with various kernels. The input channels in each branch have a dimension that is one-fourth of the total dimension. The features obtained by the different convolution kernels are defined as the information in the corresponding receptive field. This allows us to obtain features at various scales without increasing the channel dimension and computational complexity.

However, the network is three-dimensional and has a large channel dimension, which leads to an increase in parameters by the second power of the convolution kernel size (parameters $= D \times W \times H \times \frac{C_{in}}{2} \times \frac{C_{out}}{2} \times group$). To solve this issue, we use depthwise convolution where the convolution group is equal to the channel dimension. This allows us to separate channels and regions by considering both channel and region variations. The multiscale feature map generation function is given by

$$x_i = Conv_c(X_0, k_i \times k_i \times k_i, C') \tag{2}$$

Where the k_i and $C' = \frac{C}{4}$ are the kernel size and the group size of the ith branch, separately. $x_i \in \mathbb{R}^{C' \times D \times H \times W}$ denotes the feature map at different scales generated by convolution with different kernels. Each feature map with a different scales of x_i has a common channel dimension C', and $i = 0, 1, 2, 3$. Each branch independently learns multiscale spatial information and establishes cross-channel interactions locally.

3.2.2. Channel-attention-guided (CAG) block

We propose an efficient Channel-Attention-Guided (CAG) block that skews available computational resources towards relatively important information by introducing input-conditional dynamics inherently in each branch of the feature map. This can be viewed as a self-attention function on the channel, and these relationships are not restricted to the local receptive fields that convolutional filters respond to.

As shown in Fig. 2, the CAG block consists of four independent channel attention mechanisms in parallel and eventually in series. Each channel attention mechanism consists of a global average pool (GAP), two fully connected (FC) layers, and their corresponding activation functions. These elements encode global information and adaptively recalibrate the relationships between channels, respectively. The global average pooling operation is as follows,

$$\widetilde{x_i} = F_{GAP}(x_i) = \frac{1}{D \times W \times H} \sum_{d=1}^{D} \sum_{w=1}^{W} \sum_{h=1}^{H} x_i(d, w, h) \tag{3}$$

Where F_{GAP} represents the average pooling operation. $x_i \in \mathbb{R}^{C' \times D \times H \times W}$ represents the feature maps in four scale receptive fields. D, W, H, and C represent the depth, width, height, and channel number of feature maps, respectively. The global average pooling operation generates channel-related statistics, which help embed global spatial information into channel descriptors. Each $\widetilde{x_i} \in \mathbb{R}^{C' \times 1 \times 1 \times 1}$ is processed by the following formula to obtain the channel attention features with scaling activation after squeezing:

$$a_i = \sigma(W_1 \delta(W_0(\widetilde{x_i}))) \tag{4}$$

Where $a_i \in \mathbb{R}^{C' \times 1 \times 1 \times 1}$ represents the channel-wise attention of each branch. $W_0 \in \mathbb{R}^{C' \times \frac{C'}{r}}$ and $W_1 \in \mathbb{R}^{\frac{C'}{r} \times C'}$ represent two fully-connected layers (FC). δ refers to the Rectified Linear Unit function (ReLU). σ represents the excitation function, and we choose the commonly used sigmoid function as the excitation function here.

The two FC layers effectively combine linear information between channels, which promotes the interaction of channel information in high and low dimensions. The excitation function assigns weights to the channels after interacting channel-wise, fully capturing channel-wise dependencies. After that, the recalibrated multiscale receptive field feature map is obtained by concatenating the features of different branches as follows,

$$X_{out} = Cat([X_1, X_2, X_3, X_4]) \tag{5}$$

Where X_{out} represents the feature in the multiscale receptive field. The pseudocode is as follows.

Algorithm 1: Multiscale Long-range Reception Mechanism

Input: A image X_{in} of size $C \times D \times W \times H$
Output: A multiscale stacked receptive field feature map X_{out} with adaptive correction

```
Split X_in into 4 parts the channel-wise;
for i in range (0, 4) do
    Calculate feature in different scale receptive field
      x_i = F_patch(X_in);
    Generate channel-related information x̃_i = F_GAP(x_i);
    Calculate channel attention a_i ;
    Replace the feature x_i by X_i = x_i ⊗ a_i ;
end
Concatenate the X_out = Cat([X_1, X_2, X_3, X_4])
```

3.3. Cross enhanced fusion module

The motivation of this module is to improve the fusion efficiency of sMRI and FDG-PET modalities using a small number of parameters while retaining important information and reducing redundancy. Inspired by fusion methods of non-local [31] and the edge-guided attention mechanism of CDFRegNet [32], we propose a Cross Enhanced Fusion module including a space-wise enhancement (SWE) block, a channel-wise enhancement (CWE) block, and a concatenate operation. These blocks allow us to capture and exploit the spatial and channel response between the sMRI and FDG-PET features to fuse local discriminative information at the patch level.

Specifically, the discriminative features extracted from the two subnetworks are input to the SWE block and CWE block to realize the spatial-wise enhancement from sMRI to FDG-PET and channel-wise enhancement from FDG-PET to sMRI. The enhanced features are then concatenated and processed through two FC layers, a ReLU activation, and Batch Normalization to learn a multilevel feature representation with improved discriminative power. Finally, we use a softmax classification layer to diagnose each subject.

3.3.1. Spatial-wise enhancement block

In order to improve the accuracy of FDG-PET images, which can be blurry and prone to registration bias, we utilize the location information from sMRI features to enhance the FDG-PET features spatially. To do this, we first reduce the dimension of the features using convolution, then determine the spatial relationship between the FDG-PET and sMRI features by multiplying the transposed FDG-PET features with the sMRI features. The resulting spatial relationship is then compressed to obtain an enhancement function, which is then applied to the original FDG-PET features to yield the final, spatially enhanced FDG-PET features.

Table 2
Quantitative comparison of different architecture on the ADNI dataset for AD vs. NC and SMC vs. NC classification. The best results are marked in bold.

Modality	Methods	AD vs. NC				SMC vs. NC				Param	FLOPs
		ACC(%)	AUC	SPE(%)	SEN(%)	ACC(%)	AUC	SPE(%)	SEN(%)		
sMRI	2DResnet18	81.40	0.8628	86.27	75.00	43.48	0.6926	95.64	22.52	4.9M	76.79G
	3DResnet10	90.70	0.9478	**94.29**	86.11	78.26	0.7747	88.73	52.63	14.36M	34.32G
	3DViT	89.53	0.9294	92.30	86.00	81.40	**0.9061**	**96.00**	61.11	88.57M	45.12G
	Ours	**91.86**	**0.9550**	92.65	**91.67**	**81.98**	0.8492	85.78	**72.15**	**2.97M**	20.19G
FDG-PET	2DResnet18	90.70	0.9328	92.65	88.89	55.07	0.6421	58.43	47.37	4.9M	76.79G
	3DResnet10	**94.19**	0.9183	98.24	**89.12**	75.36	0.6705	82.68	**57.89**	14.36M	34.32G
	3DViT	91.86	0.9389	94.12	88.51	**78.26**	0.5832	**97.48**	21.05	88.57M	45.12G
	Ours	93.02	**0.9437**	**98.48**	83.33	74.56	**0.7021**	82.84	53.16	**2.97M**	20.19G
sMRI & FDG-PET	2DResnet18	85.87	0.9543	94.23	75.00	63.77	0.7147	68.63	52.63	9.8M	153.58G
	3DResnet10	93.02	**0.9894**	90.00	**97.22**	75.36	0.7242	88.22	47.37	28.71M	68.64G
	3DViT	92.61	0.9704	96.08	87.84	78.26	0.7147	**91.17**	42.11	176.28M	90.23G
	Ours	**97.67**	0.9855	**98.21**	**97.22**	**81.63**	**0.8737**	85.29	**72.15**	**8.5M**	14.44G

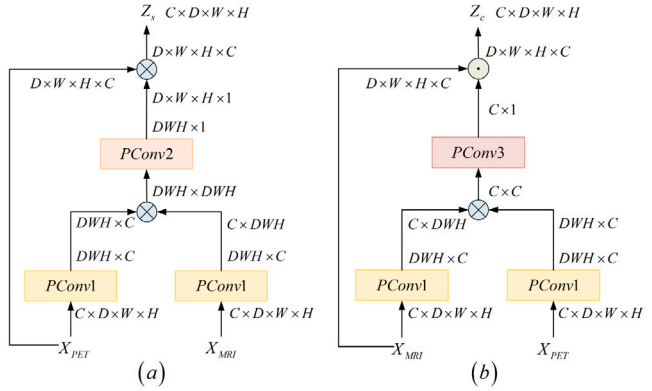

Fig. 3. The cross enhanced fusion mechanism includes (a) SWE and (b) CWE block. PConv1, PConv2, and PConv3 denote point-wise convolutions with 1 × 1 × 1.

As illustrated in Fig. 3, unifying the number of channels of sMRI and FDG-PET features can be implemented as convolution with kernel size 1, c_{out} output channels:

$$f(X) = \sigma\{BN(Conv_{c_{in}\to c_{out}}(X, kernel\_size = 1))\} \tag{6}$$

Where f consists of a convolution, a batch normalization, and a sigmoid function. To further reduce the computational complexity and center on the spatial-wise features, we flatten the features in the dimension of D, W, and H to obtain the two-dimensional vector X_{PET} and X_{MRI}. The spatial-wise relationship r_s of spatial response can be implemented as,

$$r_s = f(X_{PET}) \circ f(X_{MRI})^{\mathrm{T}} \tag{7}$$

Where X_{PET} and X_{MRI} represent the input from the MLR module of sMRI and FDG-PET, respectively. $\circ$ represents matrix multiplication. Convolution is used again to reduce the relationship of spatial response to one dimension,

$$R_s = Conv_{D_2W_2H_2\to 1}(r_s, kernel\_size = 1) \tag{8}$$

Where R_s represents the relationship vector of spatial response enhancement. Finally, R_s is multiplied by the original FDG-PET feature to obtain the final enhanced feature,

$$Z_s = X_{PET} \otimes R_s \tag{9}$$

Where $\otimes$ denotes the spatial-wise multiplication.

3.3.2. Channel-wise enhancement block

Similarly, considering that the sMRI merely reflects the basic structural changes and treats all features equally, we aim to enhance the sMRI features by weighting them based on the metabolism intensity information of FDG-PET features. This is achieved through a channel-wise enhancement of the original sMRI features. Like in the SWE block, we first use convolution with a kernel size of 1 to unify the D, W, and H dimensions of the sMRI and FDG-PET features, then flatten the features in the D, W, and H dimensions to obtain the two-dimensional vectors X'_{PET} and X'_{MRI}. The channel-wise relationship r_c of channel response can be implemented as follows:

$$r_c = f(X'_{PET}) \circ f(X'_{MRI})^{\mathrm{T}} \tag{10}$$

Where X'_{PET} and X'_{MRI} represent the input from the MLR module of sMRI and FDG-PET, separately. Convolution is used again to reduce the relationship of spatial response,

$$R_c = Conv_{c_1\to 1}(r_c, kernel\_size = 1) \tag{11}$$

Where R_c represents the relationship vector of channel response enhancement. Finally, the vector R_c is multiplied by the original sMRI feature to obtain the final enhanced feature,

$$Z_c = X_{MRI} \odot R_c \tag{12}$$

Where $\odot$ denotes the channel-wise multiplication.

After applying these two enhancement blocks to the original features, we obtain PET-enhanced features Z_{PET} based on spatial-wise location information and MRI-enhanced features Z_{MRI} based on channel-wise intensity information. Finally, we concatenate them using the following formula,

$$Z_{output} = f(Z_{MRI} \oplus Z_{PET}) \tag{13}$$

Where the Z_{output} represents the final fusion features. $\oplus$ denotes the concatenate operation.

3.4. Implementation

3.4.1. Training parameter setting

The proposed MENet is implemented on a single GPU (*i.e.*, NVIDIA GeForce 3090 24 GB), using Python based on the PyTorch package. The stochastic gradient descent (SGD) optimizer is used with the momentum of 0.9, weight decay of 0.001, and batch size of 8. Training is conducted for 500 epochs using the cross-entropy loss with an initial learning rate of 0.01 that is decreased by a factor of 10 after every 100 epochs.

3.4.2. Baseline training strategy

The baseline training strategy consists of two parts: feature extraction and cross enhanced fusion. In the first part, we train the two subnetworks separately using features extracted from sMRI and FDG-PET. Each subnetwork consists of two convolutional layers, a ReLU activation, and batch normalization. Once these subnetworks have converged from scratch by minimizing the cross-entropy loss, we concatenate them and freeze the parameters to train the remaining CEF module and part of the classification prediction. Finally, the network obtains the multiscale stacked receptive field features guided by channel attention.

Table 3
Quantitative comparison of state-of-the-art methods on the ADNI dataset for AD vs. NC classification.

Modality	Methods	Year	Data(AD/NC)	ACC(%)	AUC	SPE(%)	SEN(%)
sMRI	Lian et al [15]	2020	199/229	90.30	0.9510	**96.50**	82.40
	Guan et al [33]	2021	367/436	89.92	0.9401	91.70	87.65
	Han et al [34]	2022	408/773	91.60	**0.9660**	93.50	89.20
	Ours	2022	186/253	**91.86**	0.9550	92.00	**91.67**
FDG-PET	Hao et al [6]	2020	211/160	92.66	0.9300	84.04	87.65
	Liu et al [22]	2021	93/100	91.20	0.9530	91.00	**91.40**
	Cui et al [17]	2022	198/263	89.36	**0.9574**	89.47	89.29
	Ours	2022	186/253	**93.02**	0.9437	**98.48**	83.33
sMRI & FDG-PET	Huang et al [21]	2019	465/480	89.11	0.9269	87.77	90.24
	Lin et al [35]	2021	362/308	89.26	0.9098	96.48	82.69
	Pan et al [36]	2021	440/368	93.05	0.9723	90.91	94.74
	Ours	2022	186/253	**97.67**	**0.9855**	**98.21**	**97.22**

Table 4
Quantitative comparison of state-of-the-art methods on the ADNI dataset for SMC vs. NC classification.

Modality	Methods	Year	Data(SMC/NC)	ACC(%)	AUC	SPE(%)	SEN(%)
sMRI	Lin et al [37]	2022	113/112	65.93	0.6300	64.48	76.92
	Chen et al [38]	2022	35/36	78.87	**0.8600**	80.00	77.78
	Jia et al [39]	2022	26/50	73.04	0.7559	38.75	**91.33**
	Ours	2022	101/253	**81.98**	0.8492	**85.78**	72.15
AV45PET	Ilker et al [40]	2019	100/100	52.90	–	45.70	**60.00**
FDG-PET	**Ours**	2022	101/253	**74.56**	**0.7021**	**82.84**	53.16
sMRI&FDG-PET	**Ours**	2022	101/253	**81.63**	**0.8737**	**85.29**	**72.15**

3.4.3. Data augmentation

As the amount of publicly available ADNI data is small, we perform some data augmentation to mitigate the overfitting problem as follows. 50% of the training set is augmented online by a combination of (1) randomly affine transforming (angle range in radians is $\pi/36$, $\pi/18$, $\pi/18$), (2) randomly flipping in the axial plane, and (3) randomly rescaling in the range of 0.9 and 1.1.

4. Results and discussion

To evaluate the performance of our proposed method, we compare it with several different architectures and the state-of-the-art (SOTA) methods. We also validate the effectiveness of the key components of our method, including the MLR module and the CEF module, and verify the automatically identified multiscale discriminative locations. All models are trained using 5-fold cross-validation, and four common metrics are used to quantify the performance for both AD vs. NC and SMC vs. NC classification tasks: area under the receiver operating characteristic curve (AUC), accuracy (ACC), specificity (SPE), and sensitivity (SEN).

4.1. Competing methods

4.1.1. Comparison with different architecture

In this set of experiments, we compare our MENet with several mainstream classification architectures, including (1) 2D-CNN-based models, *i.e.*, Resnet18, (2) 3D-CNN-based models, *i.e.*, Resnet10, and (3) Transformer models, *i.e.*, ViT. To accommodate the different input requirements of these architectures, we use the following approaches: for 2D-CNN-based networks, we split and splice the data according to the slicing direction for training; for 3D-CNN-based networks, we use the original images; for the ViT-based model, we apply 3D patch embedding of size 7*7*7 with a projection dimension of 1024. For single-modality experiments, we use the features extraction subnetwork described in Section 3.4.2, followed by two convolutional layers, a ReLU activation, and batch normalization. For multi-modality experiments, we replicate this subnetwork to create a parallel network, concatenate them, and use a softmax to obtain the classification results. We compare the performance of these models with ours in sMRI, FDG-PET, and multi-modality, and the results are shown in Table 2.

Overall, our proposed method achieves the highest values of the evaluation metrics compared to the other methods. Firstly, 3D-CNN-based, ViT-based, and our method significantly improve the performance of both AD vs. NC and SMC vs. NC classification tasks compared to 2D-CNN-based networks, suggesting that high-dimensional information directly extracted from the original 3D image can better reflect its complete information. Secondly, the parameters of our MENet are always the smallest among the compared methods, yet it still performs well. While the accuracy of 3DResnet is better than ours for FDG-PET single-modality classification of AD vs. NC, it has much more parameters. Finally, all the networks perform worse in the SMC vs. NC tasks based on FDG-PET single-modality, which we believe may be due to the early functional metabolic situation being less obvious or even absent. Jiang et al. [41] similarly found that FDG-PET may not be helpful for SMC vs. NC classification, which aligns with our suspicion.

4.1.2. Comparison with SOTA methods

In this set of experiments, we compare our MENet with the state-of-the-art comparison methods. While the datasets and data volumes used in each study are different, and the network parameters are not available for comparison, we can broadly compare the classification metrics of each method across different models. The results for the AD group and SMC group are shown in Tables 3 and 4.

For the AD vs. NC classification tasks, our network's results are generally comparable to the most advanced results in both single-modality and multi-modality. This suggests that the idea of patch embedding in VIT can be integrated with CNNs to extract stronger feature representation than manually extracted features and establish long-range dependencies.

For the SMC vs. NC classification tasks, to the best of our knowledge, there are no deep learning methods for NC and SMC classification using FDG-PET, and our work fills the gap in research. Our proposed method outperforms existing work in both single-modality and multi-modality terms, indicating that it is sensitive to subtle brain changes. We also observe that the results for FDG-PET are worse than those of sMRI, possibly because there is no significant decrease in brain metabolism in SMC subjects [41–43].

Table 5
Quantitative comparison of ablation experiments for each module on the ADNI dataset for AD vs. NC AND SMC vs. NC classification.

Modality	Methods	Kernel	AD vs. NC				SMC vs. NC				Param	FLOPs
			ACC(%)	AUC	SPE(%)	SEN(%)	ACC(%)	AUC	SPE(%)	SEN(%)		
sMRI	BL	3	81.40	0.8383	88.39	72.22	71.01	0.7232	78.08	52.63	1.9M	39.37G
	BL	5	86.05	0.9089	**92.37**	77.78	79.15	0.8346	79.41	**78.48**	2.1M	9.66G
	BL	7	87.21	0.9167	84.35	**91.67**	78.09	0.8348	80.88	70.89	2.55M	4.78G
	BL	9	84.88	0.9006	82.53	88.89	75.36	0.7158	**90.20**	36.84	3.34M	3.18G
	BL	3+5+7+9	88.37	0.9383	86.76	**91.67**	78.45	**0.8590**	79.90	74.68	2.96M	20.18G
	BL+MLR	3+5+7+9	**90.70**	**0.9439**	90.69	**91.67**	**81.98**	0.8492	85.78	72.15	2.97M	20.19G
FDG-PET	BL	3	83.24	0.9135	75.68	89.22	70.32	0.5775	**94.12**	8.86	1.9M	39.37G
	BL	5	89.53	0.9278	96.15	80.56	71.73	0.7141	81.37	46.84	2.1M	9.66G
	BL	7	90.70	0.8770	95.59	65.79	70.32	0.7016	73.53	62.03	2.55M	4.78G
	BL	9	93.18	0.9798	94.12	91.89	72.79	0.6657	84.80	41.77	3.34M	3.18G
	BL	3+5+7+9	93.02	0.9439	**98.53**	83.33	73.85	**0.8077**	72.06	**78.48**	2.96M	20.18G
	BL+MLR	3+5+7+9	**94.19**	**0.9877**	96.64	**94.10**	**74.56**	0.7021	82.84	53.16	2.97M	20.19G
sMRI & FDG-PET	BL	3	81.40	0.8528	82.35	80.56	72.46	0.7600	80.88	52.63	3.28M	76.59G
	BL	5	89.53	0.9367	92.16	86.11	75.27	0.7781	90.69	35.44	3.68M	18.86G
	BL	7	92.61	0.9681	91.22	93.63	76.81	0.7284	90.69	42.11	4.57M	9.35G
	BL	9	92.33	0.9660	92.57	92.16	75.36	0.7463	92.65	31.58	6.16M	6.29G
	BL	3+5+7+9	94.19	0.9850	92.16	97.21	76.68	0.8266	72.15	**78.43**	5.93M	34.16G
	BL+MLR	3+5+7+9	95.35	**0.9894**	94.61	96.17	79.71	0.7579	**96.15**	36.84	5.95M	34.17G
	BL+MLR+CEF(Ours)	3+5+7+9	**97.67**	0.9855	**98.21**	**97.22**	**81.63**	**0.8737**	85.29	72.15	8.05M	35.35G

4.2. Ablation study on the network

4.2.1. Baseline

To verify the effectiveness of each module, we propose two baseline models: a single-model and a multi-model. The single-model consists of two ordinary convolutions, a depthwise convolution, and a sigmoid function. The multi-model consists of two single-models in parallel that are concatenated at the end. We compare the results of these models using different kernel sizes for the depthwise convolution, which corresponds to different receptive fields.

According to Table 5, we observe that the results for single-model based on sMRI are better with convolution kernels of size 5 and 7, while larger convolution (*i.e.*, kernel = 9) perform better for single-model based on FDG-PET. This supports the effectiveness of our MRF and CAG blocks, which automatically select the more important features of each mode based on their importance and allocate computing resources to relatively important information. In the multi-model, we find that simply concatenating the two branches does not always result in better performance than the single-model, which confirms that simple concatenate is not sufficient to achieve optimal mode fusion.

4.2.2. Multiscale long-range reception module

To verify the effectiveness of the MRF block, we replace the depthwise convolution with our proposed parallel multiscale convolution in both single-model and multi-model baseline models. As illustrated in Table 5, in both AD and SMC classification tasks, the model with the MRF block outperforms the baseline model with a single kernel convolution, indicating that it can perceive subtle changes and establish long-range dependencies when it has local and large-scale receptive fields.

In the experiment of verify the CAG block, we add the CAG block to the model with the MRF block. As illustrated in Table 5, simple stacking operation different scales of information performs worse tha channel-attention-guided stacking. The CAG block consumes few parameters and exhibits better performance, demonstrating that it effectively filters the rich feature information obtained from the MRF block and reduces feature redundancy while improving classification performance.

4.2.3. Cross enhanced fusion module

To verify the effectiveness of the CEF module, the experiment is divided into two parts: (1) applying the feature enhancement (SWE and CWE blocks) after the original multimodel and inputting the sigmoid function to obtain classification scores; (2) performing feature enhancement and concatenating after joining the network with the MLR module to obtain classification scores.

As shown in Table 5, the classification of AD and SMC are improved after feature enhancement in the original multi-model. The ACC are improved by 2.34% in the AD task, and by 1.92% in the SMC task. The results of the network with the CEF module after the MLR module are superior to all other experiments, indicating that the SWE and CWE blocks can effectively capture the spatial and channel response between sMRI and FDG-PET modalities to fuse local discriminative information at the patch-level.

4.3. Results on one-vs-all tasks

Given the presence of the three classes of NC, SMC and AD are mixed in clinical populations, the performance of our proposed method in real-world scenarios is explored through three one-vs-all downstream tasks. The experimental procedure is consistent with that described above and the results are presented in Fig. 4. Generally, the results indicate that the ACC of all groups improved following modality fusion as compared to their single modal counterparts, with better performance in the distinction of AD-vs-all. However, the performance of the PET-based results is inferior to that of the MRI results, which may be attributed to the introduction of difficult samples in the SMC category. The distinction between NC-vs-all performs better than the distinction between SMC-vs-all, which aligns with the typical progression of AD, with early Symptomatic Mild Cognitive Impairment patients being the most challenging to differentiate.

4.4. Localization of discriminative region on the patch-level

Our proposed method can automatically identify brain structural and metabolic anomalies in whole-brain images in sMRI and FDG-PET using a classification task-oriented approach. As illustrated in Fig. 5, we generate the Gradient-weighted Class Activation Mapping (Grad-CAM) [44] for the prediction returned by the model in the classification tasks for AD vs. NC and SMC vs. NC groups, respectively. Grad-CAM uses the global average of gradients to calculate the weight of the feature map and obtains the pixels with a positive correlation to the category. We interpret the heat map generated by gradcam as a patch-level localization that affects network decisions. Each gradient mapping is displayed in three-dimensional form among multi-planes (coronal, sagittal, and axial).

We compare the discriminative regions localized by subjects at different patch sizes, which are not pre-delineated but learned adaptively by the network. In addition, we compare the differences in discriminative brain regions among different subjects from the same

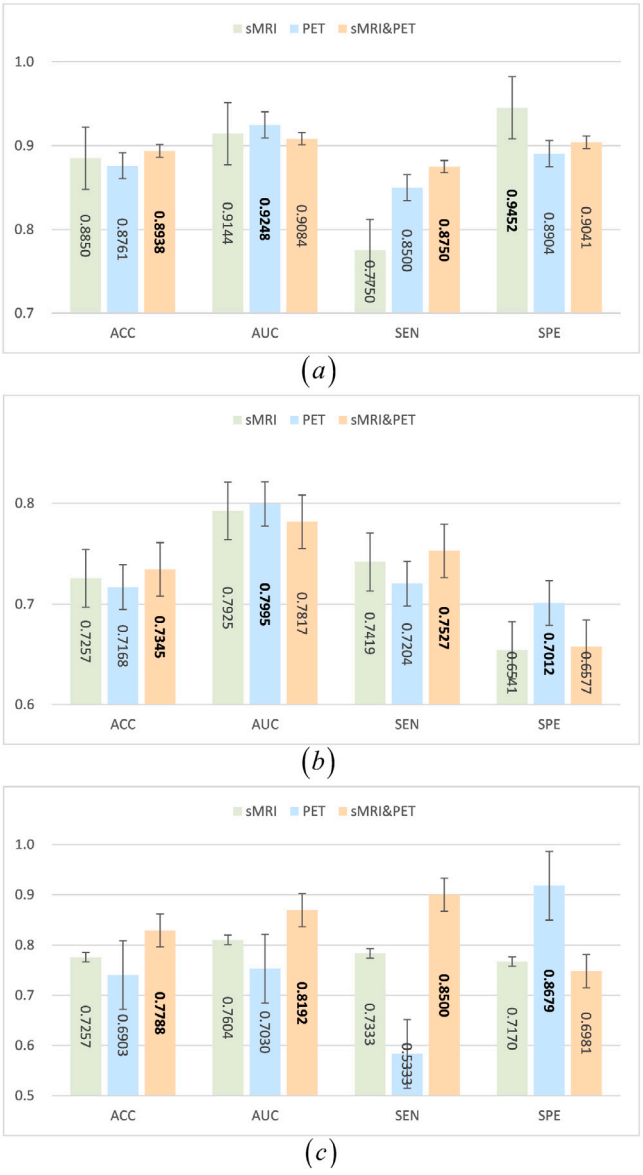

Fig. 4. Results of (a) AD-vs-all, (b) SMC-vs-all and (c) NC-vs-all diagnosis, obtained by the modal including MRI, PET and MRI&PET respectively.

group based on inspection of the entire dataset. For different AD subjects, our method co-locates atrophy in the ventricle regions, lingual, amygdala, fusiform, hippocampus, superior temporal gyrus, medial and paracingulate gyrus by sMRI, while locating atrophy in the superior cerebellum, lingual, calcarine, hippocampus, thalamus, middle temporal gyrus, middle frontal gyrus by FDG-PET. For different SMC subjects, our method co-locates atrophy in the ventricle regions, inferior frontal gyrus, rolandic operculum, precuneus, lingual, hippocampus, and caudate by sMRI, while locating atrophy in the inferior frontal gyrus, middle frontal gyrus, superior parietal cortex by FDG-PET. The classification performance of these regions in AD and SMC are consistent with previous clinical studies [45,46]. Among these, the discriminative ventricle regions, lingual, hippocampus, and middle frontal gyrus are common to both AD and SMC subjects, indicating that our method can be useful in recommending clinical interventions for patients when atrophy occurs in these regions.

Overall, our results show that the localization of discriminative regions differs at different input scales, supporting the idea of the effectiveness of our adaptive recalibration method for different scale features. Additionally, the localization of discriminative regions differs between the sMRI and FDG-PET modalities, and the dual localization of structural atrophy and metabolic discriminative regions is helpful for classification. Furthermore, the identification of discriminative regions varies among different subjects within the same group, demonstrating the adaptability of our MENet to the characteristics of discriminative regions. Finally, the localization of our network for AD and SMC patients is consistent to some extent, indicating a strong correlation and recurrence between the two tasks.

4.5. Limitations and future work

Although our proposed method achieves well performance in the diagnosis of AD and SMC, it still has many limitations at present. By considering the following points, we hope to improve the performance of our method. Firstly, based on the consideration of model parameters, we fix four branches in the MRF block to adaptively obtain a multi-scale receptive field, which means that we fixed the range of optional receptive field in advance, lacking comparison with other combinations of the receptive field. Extending the branches of the block or changing the combination of branches to fully automate the selection of receptive fields may be a solution for this problem. Secondly, due to the setting of patch embedding at the first layer, the size of network localization is limited by the size of patches, which indicates that we cannot obtain more fine-grained information. We can refer to the pyramid-shaped module to obtain the multiscale input information, and use the coarse localization to guide more accurate discriminative localization. Thirdly, the images lose some details after multilayer preprocessing, especially after spatial normalization to MNI, the use of a single template leads to potential bias which means the feature representation generated from a single template may not be enough to reveal the potentially complex differences between the patient group and the normal control group. It is a worthwhile direction to compare feature extraction under different MNI templates.

5. Conclusion

In this paper, we propose an efficient patch-based 3D convolutional neural network named MENet for diagnosis of AD and SMC, which can automatically localize the discriminative structural and metabolic regions by using raw 3D data from sMRI and FDG-PET. Notably, our MENet requires no predefined landmarks or additional location modules (*e.g.*, hippocampus segmentation). It requires fewer training parameters, but has superior performance compared to existing state-of-the-art deep-learning-based methods for diagnostic tasks in AD and SMC.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is fund by the Key Project of Health Commission of Jiangsu Province, China under Grant ZDB2020011, the Technology Project of Diagnosis and treatment of key diseases in Suzhou, China under Grant LCZX201909, and the Suzhou Science and Technology Bureau, China under Grant SJC2021023. Data used in the preparation of this article are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

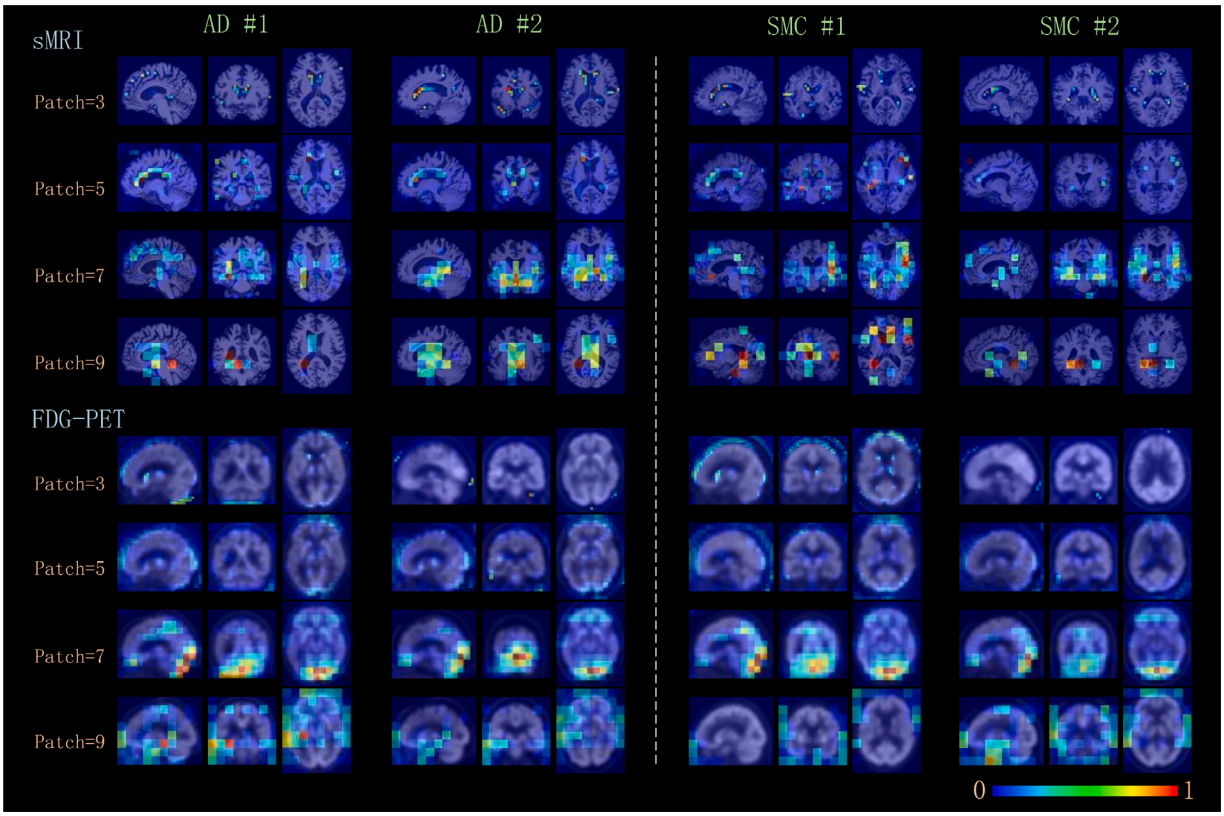

Fig. 5. Illustration of the patch-level grad-cam for two AD and two SMC subjects. Each subject has two modalities (sMRI and FDG-PET), each modality is displayed in 3 planes (coronal, sagittal, and axial), and each plane has 4 types of patch-based regions.

References

[1] 2022 Alzheimer's disease facts and figures, Alzheimer's & Dementia 18 (4) (2022) 700–789.

[2] P. Scheltens, et al., Alzheimer's disease, Lancet 397 (10284) (2021) 1577–1590.

[3] G.B. Frisoni, et al., The clinical use of structural MRI in Alzheimer disease, Nat. Rev. Neurol. 6 (2) (2010) 67–77.

[4] L. Du, et al., Identifying associations among genomic, proteomic and imaging biomarkers via adaptive sparse multi-view canonical correlation analysis, Med. Image Anal. 70 (2021) 102003.

[5] L. Du, et al., Detecting genetic associations with brain imaging phenotypes in Alzheimer's disease via a novel structured SCCA approach, Med. Image Anal. 61 (2020) 101656.

[6] X. Hao, et al., Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease, Med. Image Anal. 60 (2020) 101625.

[7] S. Minoshima, et al., 18F-FDG PET Imaging in Neurodegenerative Dementing Disorders: Insights into Subtype Classification, Emerging Disease Categories, and Mixed Dementia with Copathologies, p. 11.

[8] J. Ashburner, K.J. Friston, Voxel-based morphometry—The methods, NeuroImage 11 (6) (2000) 805–821.

[9] M. Liu, et al., A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease, NeuroImage 208 (2020) 116459.

[10] L. Chen, H. Qiao, F. Zhu, Alzheimer's disease diagnosis with brain structural MRI using multiview-slice attention and 3D convolution neural network, Front. Aging Neurosci. 14 (2022) 871706.

[11] X. Pan, et al., Multi-view separable pyramid network for AD prediction at MCI stage by 18 F-FDG brain PET imaging, IEEE Trans. Med. Imaging 40 (1) (2021) 81–92.

[12] I. Tolstikhin, et al., MLP-Mixer: An all-MLP architecture for vision, 2021, arXiv:2105.01601, [cs], (Accessed: Mar. 01, 2022). [Online]. Available: http://arxiv.org/abs/2105.01601.

[13] A. Dosovitskiy, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2021, arXiv, (Accessed: Sep. 14, 2022). [Online]. Available: http://arxiv.org/abs/2010.11929.

[14] F. Li, M. Liu, Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks, Comput. Med. Imaging Graph. 70 (2018) 101–110.

[15] C. Lian, et al., Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural MRI, IEEE Trans. Pattern Anal. Mach. Intell. 42 (4) (2020) 880–893.

[16] C. Lian, et al., Attention-guided hybrid network for dementia diagnosis with structural MR images, IEEE Trans. Cybern. 52 (4) (2022) 1992–2003.

[17] W. Cui, et al., Bmnet: A new region-based metric learning method for early Alzheimer's disease identification with FDG-PET images, Front. Neurosci. 16 (2022) 831533.

[18] J. Guo, et al., Predicting alzheimer's disease by hierarchical graph convolution from positron emission tomography imaging, in: 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 5359–5363.

[19] J. Islam, Y. Zhang, Understanding 3D CNN behavior for Alzheimer's disease diagnosis from brain PET scan, 2019, arXiv, (Accessed: Sep. 28, 2022). [Online]. Available: http://arxiv.org/abs/1912.04563.

[20] E. Yee, et al., Quantifying brain metabolism from FDG-PET images into a probability of Alzheimer's dementia score, Hum. Brain Mapp. 41 (1) (2020) 5–16.

[21] Y. Huang, et al., Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network, Front. Neurosci. 13 (2019) 509.

[22] Y. Liu, et al., Assessing clinical progression from subjective cognitive decline to mild cognitive impairment with incomplete multi-modal neuroimages, Med. Image Anal. 75 (2021) 102266.

[23] M. Liu, et al., Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis, Neuroinform 16 (3–4) (2018) 295–308.

[24] J.G. Sled, A.P. Zijdenbos, A.C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, IEEE Trans. Med. Imaging 17 (1) (1998) 87–97.

[25] J. Ashburner, K.J. Friston, Unified segmentation, NeuroImage 26 (3) (2005) 839–851.

[26] A. Trockman, J.Z. Kolter, Patches are all you need?, 2022, arXiv, (Accessed: May 23, 2022). [Online]. Available: http://arxiv.org/abs/2201.09792.

[27] A. Vaswani, et al., Attention is all you need, 2017, arXiv:1706.03762, [cs], (Accessed: Mar. 02, 2022). [Online]. Available: http://arxiv.org/abs/1706.03762.

[28] N. Tajbakhsh, et al., Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation, Med. Image Anal. 63 (2020) 101693.

[29] C. Szegedy, et al., Going deeper with convolutions, 2014, arXiv, (Accessed: Sep. 15, 2022). [Online]. Available: http://arxiv.org/abs/1409.4842.

[30] C. Szegedy, et al., Inception-v4, inception-ResNet and the impact of residual connections on learning, 2016, arXiv, (Accessed: Sep. 15, 2022). [Online]. Available: http://arxiv.org/abs/1602.07261.

[31] X. Wang, et al., Non-local neural networks, 2018, arXiv, (Accessed: Sep. 15, 2022). [Online]. Available: http://arxiv.org/abs/1711.07971.
[32] Y. Cao, et al., CDFRegNet: A cross-domain fusion registration network for CT-to-CBCT image registration, Comput. Methods Programs Biomed. 224 (2022) 107025.
[33] H. Guan, et al., Multi-site MRI harmonization via attention-guided deep domain adaptation for brain disorder identification, Med. Image Anal. 71 (2021) 102076.
[34] K. Han, et al., Multi-task multi-level feature adversarial network for joint Alzheimer's disease diagnosis and atrophy localization using sMRI, Phys. Med. Biol. 67 (8) (2022) 085002.
[35] W. Lin, et al., Bidirectional mapping of brain MRI and PET with 3D reversible GAN for the diagnosis of Alzheimer's disease, Front. Neurosci. 15 (2021) 646013.
[36] Y. Pan, et al., Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data, IEEE Trans. Pattern Anal. Mach. Intell. (2021) 1.
[37] H. Lin, et al., Identification of subjective cognitive decline due to Alzheimer's disease using multimodal MRI combining with machine learning, Cerebral Cortex (2022) bhac084.
[38] H. Chen, et al., Machine learning based on the multimodal connectome can predict the preclinical stage of Alzheimer's disease: a preliminary study, Eur. Radiol. 32 (1) (2022) 448–459.
[39] H. Jia, H. Lao, Deep learning and multimodal feature fusion for the aided diagnosis of Alzheimer's disease, Neural Comput. Appl. (2022).
[40] I. Ozsahin, B. Sekeroglu, G.S.P. Mok, The use of back propagation neural networks and 18F-Florbetapir PET for early detection of Alzheimer's disease using Alzheimer's Disease Neuroimaging Initiative database, PLoS ONE 14 (12) (2019) e0226577.
[41] J. Jiang, et al., Using radiomics-based modelling to predict individual progression from mild cognitive impairment to Alzheimer's disease, Eur. J. Nucl. Med. Mol. Imaging 49 (7) (2022) 2163–2173.
[42] A. Brugnolo, et al., Metabolic correlates of rey auditory verbal learning test in elderly subjects with memory complaints, JAD 39 (1) (2014) 103–113.
[43] J.A. Matías-Guiu, et al., Neural basis of cognitive assessment in alzheimer disease, amnestic mild cognitive impairment, and subjective memory complaints, Am. J. Geriatr. Psychiatry 25 (7) (2017) 730–740.
[44] R.R. Selvaraju, et al., Grad-CAM: Visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2) (2020) 336–359.
[45] Q. Dong, et al., Glucose metabolism in the right middle temporal gyrus could be a potential biomarker for subjective cognitive decline: a study of a han population, Alz. Res. Therapy 13 (1) (2021) 74.
[46] L. Scheef, et al., Glucose metabolism, gray matter structure, and memory decline in subjective memory impairment, Neurology 79 (13) (2012) 1332–1339.