

Improving Alzheimer's Disease Diagnosis With Multi-Modal PET Embedding Features by a 3D Multi-Task MLP-Mixer Neural Network

Zi-Chao Zhang^{1b}, Xingzhong Zhao, Guiying Dong, and Xing-Ming Zhao^{1b}, *Senior Member, IEEE*

Abstract—Positron emission tomography (PET) with fluorodeoxyglucose (FDG) or florbetapir (AV45) has been proved effective in the diagnosis of Alzheimer's disease. However, the expensive and radioactive nature of PET has limited its application. Here, employing multi-layer perceptron mixer architecture, we present a deep learning model, namely 3-dimensional multi-task multi-layer perceptron mixer, for simultaneously predicting the standardized uptake value ratios (SUVRs) for FDG-PET and AV45-PET from the cheap and widely used structural magnetic resonance imaging data, and the model can be further used for Alzheimer's disease diagnosis based on embedding features derived from SUVR prediction. Experiment results demonstrate the high prediction accuracy of the proposed method for FDG/AV45-PET SUVRs, where we achieved Pearson's correlation coefficients of 0.66 and 0.61 respectively between the estimated and actual SUVR and the estimated SUVRs also show high sensitivity and distinct longitudinal patterns for different disease status. By taking into account PET embedding features, the proposed method outperforms other competing methods on five independent datasets in the diagnosis of Alzheimer's disease and discriminating between stable and progressive mild cognitive impairments, achieving the area under receiver operating characteristic curves of 0.968 and 0.776 respectively on ADNI dataset, and generalizes better to other external datasets. Moreover, the top-weighted patches extracted from the trained model involve important brain regions related to Alzheimer's disease, suggesting good biological interpretability of our proposed method."

Index Terms—Alzheimer's disease, positron emission tomography, magnetic resonance imaging, incomplete multi-modal data, multi-layer perceptron mixer.

I. INTRODUCTION

ALZHEIMER'S disease (AD) is a leading cause of dementia [1] with an estimated prevalence of 10-30% for populations aged over 65 [2]. The characteristics of AD include cortical and hippocampus atrophy, the accumulation of beta-amyloid peptides and abnormal phosphorylation of tau protein in the brain, and the decline of brain glucose metabolism [1], [2], [3]. Thanks to advances in imaging technologies, e.g., structural magnetic resonance imaging (sMRI) and positron emission tomography (PET), it is becoming possible to detect AD more effectively. With sMRI we can accurately detect the brain regions with decreased volume. By using PET, such as fluorodeoxyglucose (FDG)-PET or florbetapir (AV45)-PET, we can have a look at the glucose metabolism or beta-amyloid burden in the brain [4], [5], [6]. Accordingly, given the accumulation of imaging data, some machine learning or deep learning based methods have been developed to predict AD status, e.g., the diagnosis of AD or the prognosis of its prodromal stage mild cognitive impairment (MCI), as AD or progressive MCI (pMCI) individuals tend to have atrophic brain regions, a lower level of glucose metabolism, and higher beta-amyloid burden compared with cognitive normal (CN) or stable MCI (sMCI) individuals [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]. Methods using PET are preferred since PET is able to directly detect the pathological biomarkers for AD, e.g., beta-amyloid, and shows superior diagnostic accuracy compared to sMRI, especially for the early detection of AD [5], [6], [21]. Besides, PET can also be integrated with sMRI to further enhance AD diagnosis performance [7], [8], [10], [15], [20], [22], [23], [24] as these modalities can provide complementary information, such as morphology and metabolism. For example, Ning et al. developed a machine learning framework to integrate sMRI and PET in the diagnosis of AD, where using both sMRI and PET achieved higher performance compared to the sMRI-only or PET-only model [15]. Recently, Goel et al. utilize wavelet transform to fuse sMRI and PET for AD diagnosis, which showed superior performance to the single modal methods [22].

Manuscript received 21 November 2022; revised 15 April 2023; accepted 18 May 2023. Date of publication 29 May 2023; date of current version 7 August 2023. This work was supported in part by the National Key R&D Program of China under Grant 2020YFA0712403, in part by the National Natural Science Foundation of China under Grants T2225015 and 61932008, and in part by the Greater Bay Area Institute of Precision Medicine (Guangzhou) under Grant IPM21C008. (Corresponding author: Xing-Ming Zhao.)

Zi-Chao Zhang, Xingzhong Zhao, and Guiying Dong are with the Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai 200433, China (e-mail: zczhang19@fudan.edu.cn; 18210850006@fudan.edu.cn; dongguiying2017@163.com).

Xing-Ming Zhao is with the Department of Neurology, Zhongshan Hospital, Institute of Science and Technology for Brain-Inspired Intelligence, State Key Laboratory of Medical Neurobiology, MOE Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China, and also with International Human Phenome Institutes, Shanghai 200433, China (e-mail: xmzhao@fudan.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JBHI.2023.3280823>, provided by the authors.

Digital Object Identifier 10.1109/JBHI.2023.3280823

However, compared to sMRI, PET is costly, and people might have radiative concerns about using radioactive tracers, which resulted in a much smaller amount of available PET data compared to sMRI in many public imaging databases for AD. Although some previous studies [7], [8], [10] employed both sMRI and PET data for AD diagnosis, the subjects with missing PET images were usually excluded, which significantly reduced the number of subjects of interest. To mitigate this problem, some computational methods have been developed to synthesize PET images from the matched sMRIs [25], [26], [27] to increase the number of usable samples. For instance, Li et al. developed a convolutional neural network (CNN) based model for directly synthesizing PET images from sMRI [25], which also improved the AD diagnosis performance. Recently, the generative adversarial network (GAN) was employed for PET image synthesis [9], [27], [28], where a generator was used for PET image synthesis and a discriminator was employed for discriminating the synthesized from real PET images. For instance, Hu et al. proposed a bidirectional GAN model to synthesize the FDG-PET images, which also demonstrated some sensitivity of the synthesized images for AD diagnosis [27]. Rather than directly predicting the entire PET image, a few studies have investigated predicting descriptive indices of PET images [29], [30], which hold stronger clinical significance. For example, Whitwell et al. developed a machine learning model to predict the PET based binary beta-amyloid status from MRI and other clinical features [29]. Reith et al. used a deep learning model to predict the progression of the AV45-PET based biomarker in the future to improve clinical trial patient selection [30].

Despite the promising results achieved in image synthesizing or indices prediction for PET, there is still much room to improve. Firstly, it can be hard and unstable to synthesize the whole PET image with GAN, given a relatively small amount of matched sMRI and PET data [31], [32]. Secondly, only a single modality of PET images was usually considered in the existing methods, whereas different modalities of PETs can provide complementary information for AD diagnosis [20], [29], [33]. For example, FDG-PET and AV45-PET can respectively detect the glucose and beta-amyloid peptide in the brain. Thirdly, most previous methods for AD diagnosis or PET image/indices prediction employed CNN [25], [26], [27], [30]. Since CNN uses local operators such as convolution and pooling [34], models based on CNN need to stack more convolutional layers to capture the longer-distant relationship lying in input, which leads to larger and deeper models and difficulty for training. Moreover, large models are inappropriate choices given a relatively small amount of training data.

In this work, by employing the recently emerging multi-layer perceptron mixer (MLP-Mixer) architecture [35], we establish a deep learning model, namely 3-dimensional multi-task multi-layer perceptron mixer (3D-Mixer),¹ consisting of a regression module and a classification module. Different from previous works that only focus on a single modality or synthesizing the whole PET images, our proposed 3D-Mixer can simultaneously

predict standardized uptake value ratios (SUVRs) for both FDG-PET and AV45-PET via the regression module. Besides, to the best of our knowledge, this is the first work that employs MLP-Mixer architecture for PET data prediction which is able to capture long-distant relationships with a few layers and avoid the inherent disadvantages of CNN and GAN. Evaluations on independent datasets demonstrate that our 3D-Mixer can predict FDG/AV45-PET SUVRs effectively, and the estimated SUVRs also exhibit strong discrimination capacity between CN and AD subjects. Moreover, by combining sMRI with PET embedding features, i.e., the output of the second last layer of the regression module, the classification module can accurately predict AD status, i.e., either the diagnosis of AD or the prognosis prediction of MCI. Benchmarking on five independent datasets demonstrates the proposed 3D-Mixer model outperforms other competing CNN and non-CNN methods.

II. METHODS

A. Overview

Fig. 1 shows the framework of the proposed 3D multi-layer perceptron mixer model (3D-Mixer) that is able to simultaneously predict FDG/AV45-PET SUVRs and AD status, which consists of a regression module and a classification module. The input patches of each module were first embedded into a feature matrix via the linear layer and there were N Mixer blocks along each branch in the regression and classification modules. We first trained the regression module to simultaneously predicts FDG/AV45-PET SUVRs via the FDG/AV45-PET branch, which took 3D sMRI patches as inputs. And the classification module was then used for AD status prediction, i.e., discrimination of CN subjects from AD subjects, or the discrimination of sMCI subjects from pMCI subjects, based on the concatenated embedding features derived from the regression module and classification module, where the classification module was pretrained on the discrimination of CN subjects from AD subjects taking input as sMRI patches extracted from pre-identified landmarks. The backbone of both regression and classification modules were Mixer blocks [35], and each Mixer block contains two multi-layer perceptrons (MLPs) as feature mixing layers that were sequentially applied to the columns and the rows of the input matrix of the Mixer block.

B. Datasets

Five datasets were used in this work, including the Alzheimer's disease neuroimaging initiative (ADNI) database,² the National Alzheimer's Coordinating Center (NACC) database [36], the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) database [37], the Open Access Series of Imaging Studies 3 (OASIS) dataset [38], and the Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) dataset [39]. In total, as summarized in Table I, 1821 subjects from ADNI (including the subsets ADNI-1/2/GO/3), 4381 subjects from NACC, 586 subjects from AIBL, 733 subjects

¹The source code is available at <https://github.com/ZhaoXM-Lab/3D-Mixer>.

²[Online]. Available: <https://adni.loni.usc.edu/>

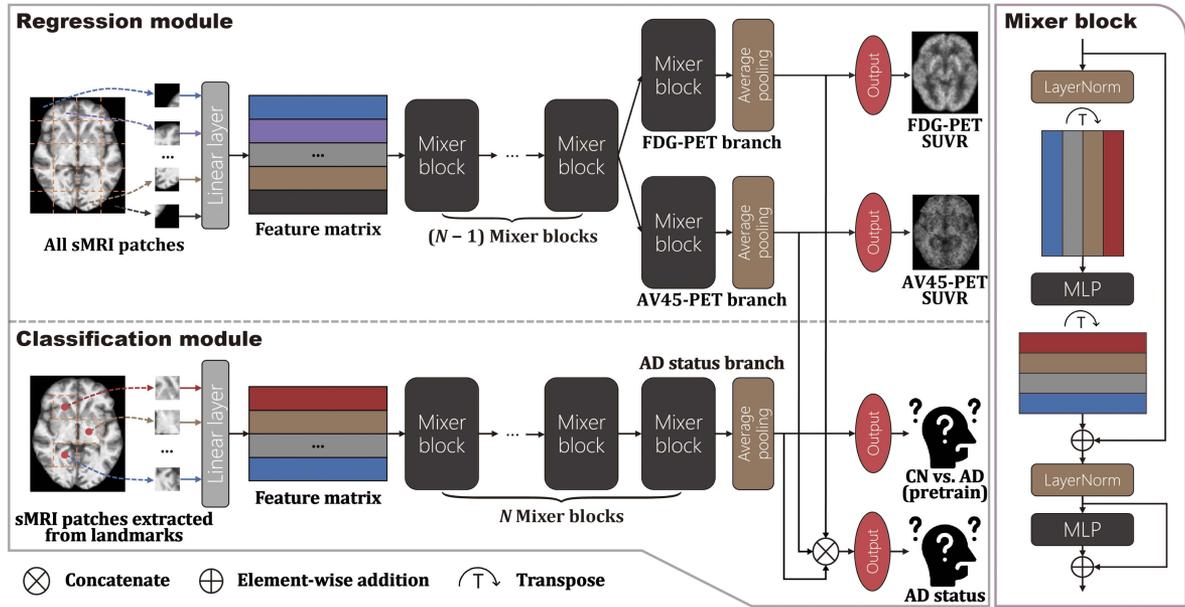


Fig. 1. Framework of 3D-Mixer that consists of a regression module and a classification module. Input sMRI patches of each module were embedded into a feature matrix by the linear layer and there were N Mixer blocks along each branch as shown in regression and classification modules. Regression module took 3D sMRI patches as input and predicted FDG-PET and AV45-PET SUVRs simultaneously as outputs via the FDG-PET branch and AV45-PET branch. Classification module was pretrained independently on the classification of CN and AD subjects taking input as sMRI patches extracted from pre-identified landmarks. And the ultimate AD status was predicted based on the concatenated embedding features derived from the pretrained regression module and the pretrained classification module. Mixer block was used to build the backbones of 3D-Mixer, where LayerNorm denotes layer normalization. Each Mixer block contains two multi-layer perceptrons (MLPs) as feature mixing layers, which were sequentially applied to the columns and the rows of the input matrix of the Mixer block.

from OASIS, and 69 subjects from MIRIAD were drawn. All participants gave informed consent in accordance with the local Institutional Review Board of each study. (Detailed ethic statements for each study can respectively be found on their web pages.) ADNI, NACC, AIBL, and OASIS datasets contain CN, sMCI, pMCI, and AD subjects, while MIRIAD dataset contains only CN and AD subjects. The sMCI individuals were those diagnosed as MCI at baseline or the earliest available visits and did not progress to AD in the follow-ups, while pMCI subjects were those diagnosed as MCI but progressed to AD in the follow-ups. It is worth noting that during training, the data at all available visits were used to increase sample size, while only the data at baseline or the earliest available visit were used for the subjects in the validation and test set.

C. Preprocessing

For the sMRIs of all datasets we used in this article, a unified preprocessing pipeline was built with the Statistical Parametric Mapping software (SPM12; www.fil.ion.ucl.ac.uk/spm) [40]. Each sMRI was spatially registered to the MNI152 template and segmented into gray matter, white matter, and cerebrospinal fluid, resulting in 3D preprocessed images of the size of $121 \times 145 \times 121$. In this work, only gray matter was used.

The FDG-PET SUVR (UCBERKELEYFDG_05_28_20) and AV45-PET SUVR data (UCBERKELEYAV45_01_14_21) were directly retrieved from the ADNI website (<http://adni.loni.usc.edu/>), where the frames of each original PET image for a visit were firstly averaged to create a single frame, and then the

mean SUVR for multiple brain regions of each subject were computed [41].

D. Patch Extraction and Landmark Identification

For the regression module of 3D-Mixer, it took input as a sequence of non-overlapping 3D sMRI patches. Following previous studies [16], [18], we use a patch size of $25 \times 25 \times 25$, and a window with the size of $100 \times 125 \times 100$ centered in the preprocessed sMRI was uniformly divided into 80 patches.

Similar to previous works [16], [17], [18], anatomical landmarks were identified for selecting patches from sMRI as input of the classification module. Firstly, we sampled an equal number of CN and AD subjects as much as possible in the training set. We extracted 80 patches without overlapping for each sMRI as mentioned above, and the mean of voxel intensities for each patch was used to characterize the patch. Subsequently, those patches that were differentially distributed between the CN and AD groups were identified with t -test, and a fixed number (50 in this article) of patches with the top smallest p -values were used as the anatomical landmarks. It is worth noting that the anatomical landmarks were identified in the training data, and we used the same patch size and landmarks in training, validation, and testing for the proposed 3D-Mixer and other competing methods.

E. 3D-Mixer Model

The proposed 3D-Mixer (Fig. 1) consists of a regression module and a classification module. For each module, the input 3D sMRI patches were embedded into a 2D feature matrix

TABLE I
SUMMARY OF THE DATASETS

Dataset	AD status	Subjects	#FDG	#AV45	#PET
ADNI-1	CN	209	125	47	147
ADNI-1	sMCI	168	101	17	104
ADNI-1	pMCI	192	117	21	121
ADNI-1	AD	175	86	2	87
ADNI-2	CN	300	295	290	296
ADNI-2	sMCI	343	335	335	338
ADNI-2	pMCI	120	117	118	118
ADNI-2	AD	149	145	146	148
ADNI-GO/3	CN	77	6	77	77
ADNI-GO/3	sMCI	63	60	33	63
ADNI-GO/3	pMCI	1	1	1	1
ADNI-GO/3	AD	14	12	7	14
ADNI-GO/3	unknown	10	8	8	10
NACC	CN	2594	/	/	/
NACC	sMCI	632	/	/	/
NACC	pMCI	341	/	/	/
NACC	AD	814	/	/	/
AIBL	CN	465	/	/	/
AIBL	sMCI	29	/	/	/
AIBL	pMCI	12	/	/	/
AIBL	AD	80	/	/	/
OASIS	CN	503	/	/	/
OASIS	sMCI	21	/	/	/
OASIS	pMCI	5	/	/	/
OASIS	AD	204	/	/	/
MIRIAD	CN	23	/	/	/
MIRIAD	AD	46	/	/	/

ADNI: Alzheimer's disease neuroimaging initiative database; NACC: National Alzheimer's Coordinating Center database; AIBL: Australian Imaging Biomarkers and Lifestyle flagship study of ageing database; OASIS: Open Access Series of Imaging Studies 3 dataset; MIRIAD: Minimal Interval Resonance Imaging in Alzheimer's Disease dataset; CN: Cognitive normal; sMCI: Stable mild cognitive impairment; pMCI: Progressive mild cognitive impairment; AD: Alzheimer's disease; ADNI-1, ADNI-2, ADNI-3, and ADNI-GO are the subsets of ADNI. #FDG/AV45 or #PET denotes the number of subjects that have sMRI matched FDG/AV45-PET SUVRS or have any sMRI matched PET SUVRS.

$X \in \mathbb{R}^{k \times c}$ via a linear layer, where k denotes the number of input patches and c denotes the dimensions of features for each patch. Subsequently, for each branch, the feature matrix was further sequentially transformed by N Mixer blocks [35] and fed to an output layer with sigmoid activation. In this article, N and c were set to 4 and 64 respectively, and k was 80 or 50 for the regression module or the classification module respectively.

1) Mixer Block: In 3D-Mixer, instead of the widely applied 3D CNN, we adopt Mixer blocks to build the backbone of both classification and regression modules. A Mixer block consists of two feature mixing layers implemented by two MLPs that were sequentially applied to the columns and rows of the input matrix, which leads to a much larger receptive field compared to CNN architectures. Given the input matrix X , the output of the Mixer block $Y = \text{Mixer}(X)$ can be formulated as follows:

$$U_{*,i} = X_{*,i} + W_2^T \sigma \left(W_1 \text{LayerNorm}(X)_{*,i} + \mathbf{b}_1 \right) + \mathbf{b}_2, \quad (1)$$

for $i = 1 \dots c$,

$$Y_{j,*} = U_{j,*} + \sigma \left(\text{LayerNorm}(U)_{j,*} W_3^T + \mathbf{b}_3 \right) W_4 + \mathbf{b}_4^T, \quad (2)$$

for $j = 1 \dots k$,

where $X, Y \in \mathbb{R}^{k \times c}$, $X_{*,i} = [x_{1i} \ x_{2i} \ \dots \ x_{ki}]^T$ and $Y_{j,*} = [y_{j1} \ y_{j2} \ \dots \ y_{jc}]$ are the i th column of X and the j th row of Y , $W_1, W_2 \in \mathbb{R}^{rk \times k}$, $W_3, W_4 \in \mathbb{R}^{rc \times c}$, $\mathbf{b}_1 \in \mathbb{R}^{rk}$, $\mathbf{b}_2 \in \mathbb{R}^k$, $\mathbf{b}_3 \in \mathbb{R}^{rc}$, and $\mathbf{b}_4 \in \mathbb{R}^c$ are the learnable weights, k denotes the number of input patches, c is the dimensions of features for each patch, r represents the expansion factor (set to 4 in this article) and $\sigma(\cdot)$ is the element-wise activation function Gaussian Error Linear Unit. LayerNorm denotes layer normalization. A dropout operation was also employed after each dense layer with a dropout rate of 0.1 in pretraining and 0.3 in the final training.

2) Regression Module: The regression module took input as a sequence of non-overlapping 3D sMRI patches and simultaneously predicted FDG/AV45-PET SUVRS via the FDG-PET branch and AV45-PET branch, where N Mixer blocks were employed along each branch.

Let $\mathbf{p} \in \mathbb{R}^m$ ($m = 25 \times 25 \times 25 = 15625$) denote a flattened sMRI patch. The input 3D sMRI patches were first embedded into a feature matrix $X_{reg}^{(0)} \in \mathbb{R}^{k_r \times c}$ by a linear layer:

$$X_{reg}^{(0)} = \begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \dots \\ \mathbf{p}_{k_r}^T \end{bmatrix} W_{reg} + \begin{bmatrix} \mathbf{b}_{reg}^T \\ \mathbf{b}_{reg}^T \\ \dots \\ \mathbf{b}_{reg}^T \end{bmatrix}, \quad (3)$$

where $W \in \mathbb{R}^{m \times c}$ and $\mathbf{b} \in \mathbb{R}^c$ are the learnable weights, and k_r denotes the number of input patches for the regression module. Subsequently, $N - 1$ shared Mixer blocks and a task-specific Mixer block were sequentially built for each output branch:

$$X_{reg}^{(n)} = \text{Mixer}_{reg}^{(n)} \left(X_{reg}^{(n-1)} \right), \quad \text{for } n = 1 \dots N - 1, \quad (4)$$

$$X_{reg}^{(FDG)} = \text{Mixer}_{reg}^{(FDG)} \left(X_{reg}^{(N-1)} \right), \quad (5)$$

$$X_{reg}^{(AV45)} = \text{Mixer}_{reg}^{(AV45)} \left(X_{reg}^{(N-1)} \right), \quad (6)$$

where $\text{Mixer}_{reg}^{(n)}(\cdot)$ denotes the n th shared Mixer block in the regression module, $\text{Mixer}_{reg}^{(FDG)}(\cdot)$ and $\text{Mixer}_{reg}^{(AV45)}(\cdot)$ respectively represents the Mixer block in the FDG-PET SUVR branch and the AV45-PET SUVR branch. Next, the average pooling was performed to extract the PET embedding features:

$$\mathbf{v}_{FDG} = \text{AvgPool} \left(X_{reg}^{(FDG)} \right), \quad (7)$$

$$\mathbf{v}_{AV45} = \text{AvgPool} \left(X_{reg}^{(AV45)} \right), \quad (8)$$

where $\mathbf{v} \in \mathbb{R}^c$, $v_j = \frac{1}{k} \sum_i x_{ij}$ for $X \in \mathbb{R}^{k \times c}$. Finally, the FDG-PET SUVR (o_{FDG}) and AV45-PET SUVR (o_{AV45}) were predicted via two output layers with sigmoid activation:

$$o_{FDG} = g \left(\mathbf{v}_{FDG}^T \mathbf{w}_{FDG} + b_{FDG} \right), \quad (9)$$

$$o_{AV45} = g \left(\mathbf{v}_{AV45}^T \mathbf{w}_{AV45} + b_{AV45} \right), \quad (10)$$

where $\mathbf{w}_{FDG}, \mathbf{w}_{AV45} \in \mathbb{R}^c, b_{FDG}$, and b_{AV45} are the learnable weights, and $g(\cdot)$ is the sigmoid function.

3) Classification Module: The classification module is similar to the regression module in structure but has only one output branch, which predicted the AD status based on the concatenated embedding features derived from the regression module and the pretrained classification module.

Let $\tilde{\mathbf{p}} \in \mathbb{R}^m$ denote a flattened sMRI patch extracted from the pre-identified landmark. Then the input patches of the classification module were embedded into a feature matrix $X_{clf}^{(0)} \in \mathbb{R}^{k_c \times c}$:

$$X_{clf}^{(0)} = \begin{bmatrix} \tilde{\mathbf{p}}_1^T \\ \tilde{\mathbf{p}}_2^T \\ \dots \\ \tilde{\mathbf{p}}_{k_c}^T \end{bmatrix} W_{clf} + \begin{bmatrix} \mathbf{b}_{clf}^T \\ \mathbf{b}_{clf}^T \\ \dots \\ \mathbf{b}_{clf}^T \end{bmatrix}, \quad (11)$$

where k_c represents the number of the pre-identified landmarks. Afterward, N Mixer blocks were sequentially applied:

$$X_{clf}^{(n)} = \text{Mixer}_{clf}^{(n)} \left(X_{clf}^{(n-1)} \right), \text{ for } n = 1 \dots N, \quad (12)$$

Finally, the AD status (o_{clf}) were predicted based on the concatenated embedding features:

$$o_{clf} = g \left([\mathbf{v}_{FDG}^T; \mathbf{v}_{AV45}^T; \mathbf{v}_{clf}^T] \mathbf{w}_{clf} + b_{clf} \right), \quad (13)$$

Where $\mathbf{v}_{clf} = \text{AvgPool}(X_{clf}^{(N)})$, and $\mathbf{w}_{clf} \in \mathbb{R}^{3c}$ and b_{clf} are the learnable weights.

F. Loss Function

When training the regression module for FDG/AV45-PET SUVR prediction, we used mean square error (MSE) as the loss function:

$$l_{reg} = \sum_{i \in S_{FDG}} (y_{FDG}^i - o_{FDG}^i)^2 / |S_{FDG}| + \sum_{i \in S_{AV45}} (y_{AV45}^i - o_{AV45}^i)^2 / |S_{AV45}| \quad (14)$$

where o is the output of the regression model (e.g., o_{FDG}^i is the predicted FDG-PET SUVR for subject i), y is the true PET SUVR (e.g., y_{FDG}^i is the true FDG-PET SUVR for subject i), and S is the index set (e.g., $S_{FDG} = \{i | y_{FDG}^i \text{ exist}\}$).

For the classification tasks, we used the binary cross-entropy loss, which can be formulated as:

$$l_{clf} = -1/|S_{clf}| \sum_{i \in S_{clf}} [y_{clf}^i \log o_{clf}^i + (1 - y_{clf}^i) \log (1 - o_{clf}^i)] \quad (15)$$

where y_{clf}^i is the binary classification label for subject i (e.g., 0 for CN and 1 for AD), and o_{clf}^i is the predicted AD status.

III. EXPERIMENTS

A. Experiment Settings

In this work, the proposed 3D-Mixer was verified on FDG/AV45-PET SUVR prediction tasks and two AD status

prediction tasks, where AD status prediction tasks include the classification of CN and AD subjects and the classification of sMCI and pMCI subjects. Specifically, the regression module was trained for predicting FDG/AV45-PET SUVRs. For the AD status prediction tasks, the classification module was independently pretrained for the classification of CN and AD subjects. Subsequently, we froze the weights of the trained regression module and trained the whole model for each classification task using the concatenated embedding features derived from the regression module and the classification module.

To evaluate the effectiveness of the proposed 3D-Mixer for AD status prediction, we employed several recently proposed deep learning based methods for comparison, including landmark-based deep multi-instance learning (LDMIL) [17], dual attention multi-instance deep learning network (DA-MIDL) [18], ResNet-18 [42], and Vision Transform (ViT) [43], where LDMIL, DA-MIDL, and ResNet-18 are based on CNN while ViT is a popular non-CNN model. The same hyperparameters as the original works for the model architectures were used for LDMIL, DA-MIDL, and ResNet-18. And for ViT, the depth (the number of transformer blocks), hidden size, MLP size, and heads were set to 4, 64, 256, and 8 respectively. Specifically, here we also used batch normalization layers for LDMIL model to improve its training stability. Besides, for the competing methods, transfer learning [16] was applied for the classification of sMCI and pMCI subjects by initializing the weights of the models with the weights learned in the classification of CN and AD subjects.

To train the models, we used the optimizer Adam. The learning rate was initially set to 1×10^{-4} and was multiplied by 0.5 once the validation loss had not decreased for 10 epochs. And only the model with the smallest validation loss was saved.

B. Simultaneous Prediction of FDG/AV45-PET SUVRs From sMRIs

For FDG/AV45-PET SUVR prediction, the regression module took input of 3D sMRI patches at each time point and predicted the matched FDG/AV45-PET SUVRs via two output branches (Fig. 1). All subjects from the Alzheimer's disease neuroimaging initiative (ADNI) database (including ADNI-1, ADNI-2, ADNI-GO, and ADNI-3) were pooled together and then randomly divided into a training set (70%), validation set (15%), and test set (15%). We trained the regression module on the training set and validated it on the validation set, and further evaluated its performance on the test set. In addition, we tested our regression module on four independent datasets, including OASIS, NACC, AIBL, and MIRIAD.

We first evaluated the performance of our 3D-Mixer on predicting either FDG-PET SUVRs or AV45-PET SUVRs based on sMRIs with the regression module. Fig. 2(a) shows the correlation between the estimated and actual FDG/AV45-PET SUVRs over the ADNI test set, from which we observed that our 3D-Mixer can efficiently predict the SUVRs with correlation coefficients above 0.6. As expected, our estimated SUVRs can characterize AD status as effectively as the actual SUVRs, and we further show that the predicted FDG/AV45-PET SUVRs have

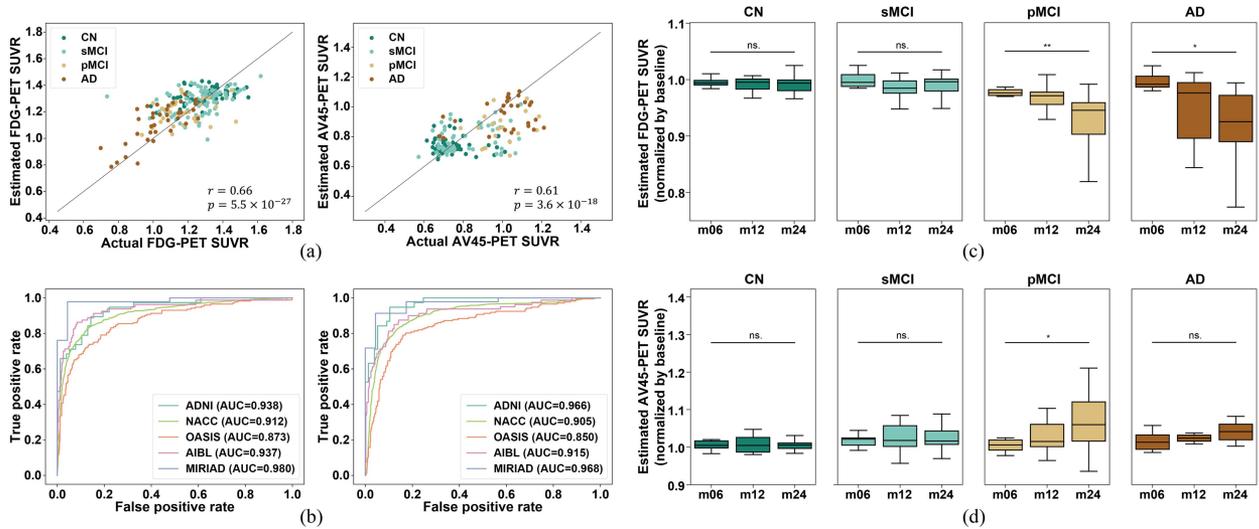


Fig. 2. Prediction of FDG/AV45-PET SUVRs using the regression module of 3D-Mixer. (a) The comparison of the estimated and the actual SUVRs on the ADNI test set, where each dot denotes a subject, and r denotes the Pearson's correlation coefficients between the estimated and actual SUVRs. (b) The receiver operating characteristic curves respectively for the classification of CN and AD subjects from the ADNI test set, OASIS, and MIRIAD datasets using the estimated FDG-PET SUVRs (left) and estimated AV45-PET SUVRs (right). (c) and (d) The estimated FDG/AV45-PET SUVRs (normalized by baseline) at different time points (bl, m06, m12, and m24 respectively denote baseline, the visits 6 months, 12 months, and 24 months after baseline) in the ADNI test set. **: $p < 0.01$, *: $p < 0.05$, ns.: Not significant.

a similar pattern to the reals across different AD status groups (Supplementary Fig. S1), where the subjects with the status closer to AD had lower estimated FDG-PET SUVRs and higher estimated AV45-PET SUVRs, indicating lower metabolism levels and higher beta-amyloid deposition levels in the brains. The above results indicate that the brain sMRIs can indeed help with predicting FDG/AV45-PET SUVRs, in agreement with the previous conclusion that brain atrophy detected by sMRI is associated with hypometabolism and beta-amyloid deposition that can be respectively detected by FDG-PET and AV45-PET [44], [45].

We then showed the Pearson's correlation coefficients between the estimated FDG/AV45-PET SUVRs and the Mini-Mental State Examination (MMSE) scores along four time points over the ADNI test set (Supplementary Fig. S2), where MMSE is a widely used measure of cognitive impairment and a higher MMSE score indicates better cognitive status. Firstly, we observed that the estimated FDG-PET SUVRs positively correlated with MMSE scores along the time points while negative correlations were observed for the estimated AV45-PET SUVRs, indicating that the predicted SUVRs can indeed characterize cognitive impairment status. Secondly, both estimated FDG-PET and AV45-PET SUVRs show reasonable correlations with MMSE scores at later time points, implying the potential for prognosis analysis. For example, the estimated FDG-PET SUVRs at bl (baseline) correlated with the MMSE scores at m12 (12 months after baseline) ($r = 0.65$, $p = 1.2 \times 10^{-5}$), while AV45-PET SUVRs at m06 correlated with MMSE scores at m12 ($r = -0.63$, $p = 1.0 \times 10^{-4}$).

Subsequently, We evaluated the diagnostic power of the estimated FDG/AV45-PET SUVRs by discriminating AD from CN subjects. Fig. 2(b) shows the receiver operating characteristic (ROC) curves for the classification of CN and AD subjects on the five datasets. We can observe that both the estimated

FDG-PET SUVRs and estimated AV45-PET SUVRs show promising abilities for discriminating AD from CN subjects on all five datasets with AUC (area under ROC curve) scores ≥ 0.85 . Especially, the estimated AV45-PET SUVRs achieved great performances on the ADNI test set and MIRIAD with AUC scores greater than 0.96. The AUC scores for the OASIS dataset were relatively lower, which might be due to the differences introduced by the inconsistencies of sMRI acquisition and the population that the subjects were from. The above observations demonstrate that the estimated FDG/AV45-PET SUVRs have a promising power for AD diagnosis on independent datasets.

To further explore the diagnostic potential of the estimated FDG/AV45-PET SUVRs, we looked at the longitudinal SUVR patterns estimated at three time points after baseline for subjects in the ADNI test set. As shown in Fig. 2(c)–(d), the pMCI and AD groups demonstrated a substantial decline in glucose metabolism, i.e., FDG-PET SUVRs, while the CN and sMCI groups had relatively stable glucose metabolism in the 2-year follow-up, in agreement with the conclusion drawn from the previous studies [41], [46] with actual FDG-PET SUVR. On the other hand, the continuously increasing beta-amyloid burden, i.e., AV45-PET SUVRs, was also specifically observed for the pMCI and AD groups, while it was more stable for the CN and sMCI groups. These observations imply the potential of the estimated FDG/AV45-PET SUVRs for the AD diagnosis and the prognosis analysis of MCI individuals.

C. AD Status Prediction

Inspired by the above findings, we developed the 3D-Mixer model to predict AD status based on the concatenated embedding features derived from both the regression module (the PET embedding features) and the classification module, where the classification module takes input as sMRI patches extracted from

TABLE II
THE PERFORMANCE OF 3D-MIXER AND OTHER COMPETING METHODS ON CN VS. AD CLASSIFICATION

Dataset	Metric	LDMIL	DA-MIDL	ViT	ResNet-18	3D-Mixer ⁻	3D-Mixer
ADNI-2	AUC	0.950	0.880	0.933	0.949	0.959	0.968
	AUPR	0.920	0.798	0.908	0.927	0.937	0.954
	ACC	0.868	0.787	0.887	0.880	0.900	0.906
NACC	AUC	0.905	0.840	0.896	0.908	0.912	0.926
	AUPR	0.783	0.677	0.741	0.807	0.811	0.837
	ACC	0.871	0.833	0.864	0.844	0.882	0.891
OASIS	AUC	0.896	0.817	0.868	0.880	0.886	0.893
	AUPR	0.756	0.632	0.692	0.763	0.764	0.785
	ACC	0.857	0.805	0.816	0.808	0.854	0.861
AIBL	AUC	0.931	0.848	0.916	0.921	0.951	0.957
	AUPR	0.798	0.552	0.733	0.766	0.831	0.857
	ACC	0.924	0.860	0.900	0.837	0.919	0.934
MIRIAD	AUC	0.988	0.939	0.982	0.967	0.987	0.988
	AUPR	0.994	0.958	0.992	0.985	0.994	0.994
	ACC	0.901	0.780	0.925	0.896	0.933	0.910

The highest score in each row is bolded. AUC: Area under receiver operating characteristic curve; AUPR: Area under precision-recall curve; ACC: Accuracy.

pre-identified landmarks. By following previous works [16], [17], ADNI-1 was used for training and validating, and ADNI-2 (the subjects that overlapped with ADNI-1 were excluded) was used as the test set, where ADNI-1 was further divided into a training set (85% of the subjects) and a validation set (15% of the subjects). Additionally, we also use the data split we used in the SUVR prediction tasks, where all ADNI subjects were pooled together and split into training, validation, and test set. We denote this test set as ADNI-held-out set. Four independent datasets NACC, OASIS, AIBL, and MIRIAD were also used for testing. We repeated the experiments five times with different random states and reported the averaged performance.

We first evaluated the proposed 3D-Mixer and other competing methods on the classification of CN and AD subjects as summarized in Table II. We observed that our proposed 3D-Mixer outperformed the other competing methods on ADNI-2 in terms of AUC, area under precision-recall curve (AUPR), and accuracy (ACC), achieving an AUC of 0.968, an AUPR of 0.954, and an ACC of 0.906. And the proposed method also achieved the best performance on the ADNI-held-out set (Supplementary Table SI). Without any fine-tuning, we applied the models trained on ADNI-1 to the independent external datasets (i.e., NACC, OASIS, AIBL, and MIRIAD), and our 3D-Mixer still achieved the overall optimal performance (Table II), which implies the strong robustness of our 3D-Mixer model. Besides, we also checked the impact of the PET embedding features on the performance of 3D-Mixer with a variant of 3D-Mixer (denoted by 3D-Mixer⁻) without using the PET embedding features. From Table II and Supplementary Table SI, we observed that PET embedding features can indeed boost the performance of classification, where 3D-Mixer achieved higher AUC and AUPR scores for the discrimination of CN subjects from AD subjects on all datasets compared to 3D-Mixer⁻. Interestingly, however, even with simple architecture, 3D-Mixer⁻ outperformed all competing CNN models and ViT over the ADNI and NACC datasets and

is comparable over other independent datasets, indicating the potential of the MLP-Mixer architecture for 3D brain sMRI analysis.

Except for the diagnosis of AD, we also evaluated the performance of 3D-Mixer and other competing methods for the prognosis prediction of MCI, i.e., the discrimination between sMCI and pMCI subjects, where pMCI will ultimately convert to AD. Table III shows the performance of 3D-Mixer and other competing methods over the ADNI-2 and other independent datasets. Our proposed 3D-Mixer outperformed all other methods on the two largest datasets (i.e., the ADNI-2 and NACC dataset) in terms of AUC, AUPR, and ACC and was superior to other methods over the OASIS dataset in terms of AUC and ACC. Specifically, 3D-Mixer achieved an AUPR of 0.560 on the ADNI-2 dataset which was improved by 6% compared with the second-best LDMIL (AUPR = 0.498). Moreover, the proposed method also overall outperformed other competing methods on the ADNI-held-out set (Supplementary Table SII). Besides, without considering the PET embedding features, the performance of 3D-Mixer⁻ degraded significantly compared with 3D-Mixer. We also noticed that the PET embedding features were more useful for MCI prognosis prediction than the AD diagnosis. For example, with the benefit of PET embedding features, the AUC of 3D-Mixer for the classification of sMCI and pMCI subjects was improved respectively by more than 2% and 3% on ADNI-2 and OASIS compared with 3D-Mixer⁻, and the AUPR was improved by nearly 4% on NACC.

D. Interpretability of 3D-Mixer

Our proposed 3D-Mixer can simultaneously predict FDG/AV45-PET SUVRs and AD status with high accuracy based on sMRI patches. To explore the interpretability of our proposed 3D-Mixer model, for each prediction task,

TABLE III
THE PERFORMANCE OF 3D-MIXER AND OTHER COMPETING METHODS ON sMCI vs. pMCI CLASSIFICATION

Dataset	Metric	LDMIL	DA-MIDL	ViT	ResNet-18	3D-Mixer ⁻	3D-Mixer
ADNI-2	AUC	0.715	0.676	0.724	0.708	0.754	0.776
	AUPR	0.498	0.445	0.478	0.447	0.554	0.560
	ACC	0.708	0.689	0.708	0.716	0.749	0.755
NACC	AUC	0.676	0.638	0.652	0.664	0.698	0.721
	AUPR	0.534	0.497	0.430	0.507	0.515	0.553
	ACC	0.624	0.625	0.592	0.622	0.664	0.674
OASIS	AUC	0.555	0.515	0.632	0.604	0.630	0.663
	AUPR	0.223	0.236	0.238	0.268	0.245	0.254
	ACC	0.577	0.646	0.562	0.592	0.692	0.715
AIBL	AUC	0.737	0.692	0.669	0.692	0.699	0.707
	AUPR	0.567	0.537	0.542	0.469	0.570	0.540
	ACC	0.707	0.673	0.654	0.668	0.673	0.673

The highest score in each row is bolded. AUC: Area under receiver operating characteristic curve; AUPR: Area under precision-recall curve; ACC: Accuracy.

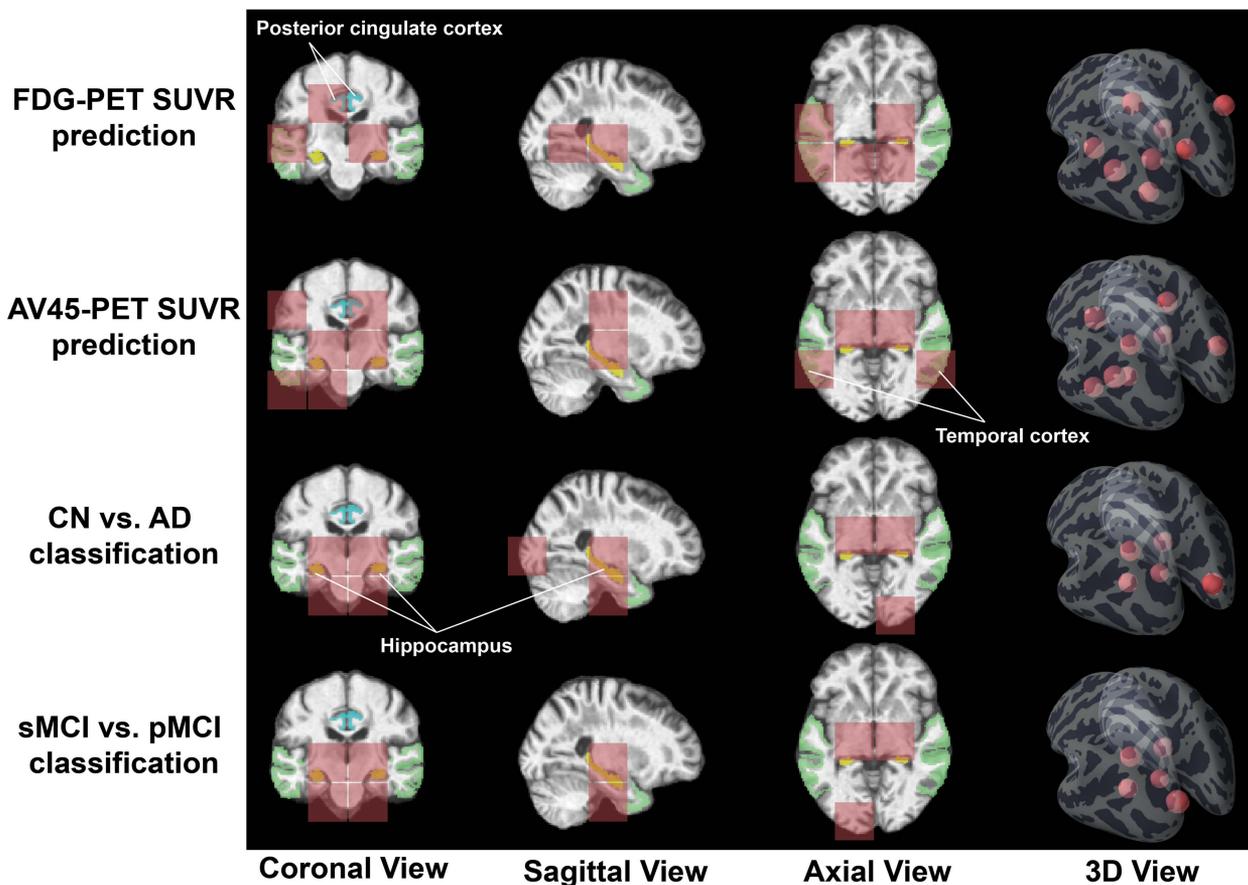


Fig. 3. Top 10% most weighted patches for FDG-PET SUVR prediction (the first row), AV45-PET SUVR prediction (the second row), the classification of CN and AD subjects (the third row), and the classification of sMCI and pMCI subjects (the fourth row) in the corresponding output branch, which were marked by red blocks and balls.

we extracted the top 10% most weighted patches in the corresponding output branch as shown in Fig. 3.

The reason of using sMRI to predict PET SUVRs is that brain morphology is associated with glucose metabolism and beta-amyloid burden in the brain. For example, the atrophy of

the hippocampus detected by sMRI has been associated with the disruption of the cingulum bundle, which could further cause hypometabolism in multiple brain regions related to AD, including the posterior cingulate cortex and middle temporal gyrus [47], [48]. Besides, it has been shown that a high level of

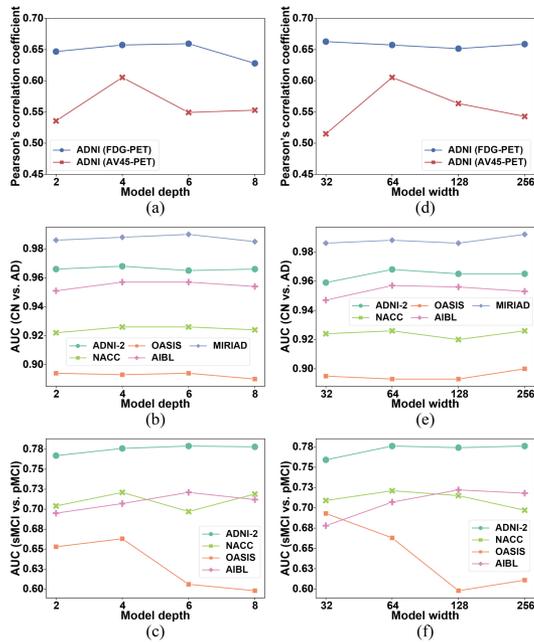


Fig. 4. Influence of the hyper-parameters on the performance. (a) Pearson's correlation coefficients between the estimated and actual FDG/AV45-PET SUVRs achieved with different depths. (b) and (c) The AUC of 3D-Mixer trained with different depths for the classification of CN and AD subjects and the classification of sMCI and pMCI subjects. (d) Pearson's correlation coefficients between the estimated and actual FDG/AV45-PET SUVRs achieved with different widths. (e) and (f) The AUC scores of 3D-Mixer trained with different widths for the classification of CN and AD subjects and the classification of sMCI and pMCI subjects.

beta-amyloid burden might accelerate cortical atrophy, and that beta-amyloid modulates the association between neurofilament light chain and brain atrophy in Alzheimer's disease. And the volume loss of the hippocampus, cingulate cortex, and temporal cortex has been reported to be associated with beta-amyloid burden in multiple brain regions, especially the precuneus [4], [49], [50]. In line with the above findings, we observed that for both FDG-PET SUVR and AV45-PET SUVR predictions, there are common top-weighted patches involved in the regions of the hippocampus, posterior cingulate cortex, and temporal cortex (Fig. 3). Then we compare the FDG-PET images used in this study from CN and AD subjects, as we showed in Supplementary Fig. S3, the posterior cingulate cortex, and temporal cortex had lower levels of glucose metabolism in AD subjects. And we also noticed higher beta-amyloid burden in multiple brain regions, particularly in the precuneus, in AD subjects by comparing AV45-PET images of CN and AD subjects (Supplementary Fig. S4). Taken together, the highly weighted patches in our model for predicting FDG/AV45-PET SUVRs can effectively characterize the metabolism level and beta-amyloid burden in the brain, which brings insights into the mechanism of the proposed 3D-Mixer on effectively predicting the FDG/AV45-PET SUVRs.

Moreover, for the classification of CN and AD subjects and the classification of sMCI and pMCI subjects, the top 10% most weighted patches in the classification branch were similar,

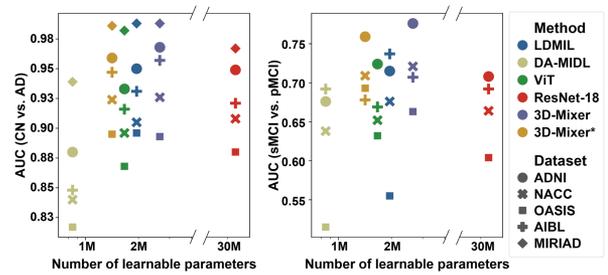


Fig. 5. Performance in term of AUC of 3D-Mixer and other competing methods with various model sizes, i.e., the number of learnable parameters, on the classification of CN and AD subjects and the classification of sMCI and pMCI subjects.

which were enriched around the region of the hippocampus. Many previous studies [13], [51], [52] have shown that the hippocampus is associated with the decreasing memory of AD, and the decreasing volume of the hippocampus is also often used as a biomarker for AD. Specifically, the atrophy of the right hippocampus was also reported significantly associated with the conversion from MCI to dementia [53]. From the above results, we can see that the proposed 3D-Mixer model is able to effectively capture the important features for both AD diagnosis and MCI prognosis prediction, and those features can help interpret the mechanisms underlying AD.

E. Influence of the Hyper-Parameters of 3D-Mixer

In the 3D-Mixer model, there are two hyper-parameters, including the depth, i.e., the number of Mixer blocks along each branch (N), and the width, i.e., the dimensions of embedding features for each patch. We chose 4 and 64 as the depth and width in our implementation based on the performance on validation set (Supplementary Fig. S5). In this section, we evaluated the influence of these parameters on the performance across different datasets by training the model with different depths and widths.

Firstly, we trained 3D-Mixer with a different number (N) of MLP-Mixer blocks selected from $\{2, 4, 6, 8\}$. As we show in Fig. 4(a)–(c), the depth had a relatively small influence on FDG-PET SUVR prediction while the highest correlation coefficient was achieved at the depth of 4 for AV45-PET SUVR prediction. And for AD status prediction, the model depth only had a relatively evident impact on OASIS dataset on the classification of sMCI and pMCI subjects. Specifically, the performance dropped from the peak of 0.663 to 0.606 in term of AUC when the depth grew from 4 to 6. This was reasonable since the deeper model was less easy for training and had a larger number of learnable parameters which usually leads to less robustness.

To evaluate the influence of width, we trained 3D-Mixer with different widths selected from $\{32, 64, 128, 256\}$. Fig. 4(d)–(f) show that our model reached the best performance at the width of 64 for AV45-PET SUVR prediction while the width had a weak impact on the FDG-PET SUVR prediction. Similar to the depth, the width had a more significant impact on the performance of the classification of sMCI and pMCI subjects compared to the classification of CN and AD subjects. As we showed in Fig. 4(f), for the classification of sMCI and pMCI subjects, the AUC score

on ADNI-2, NACC, and AIBL grew with the width when it was small, while large width can have a negative impact on the performance on the three independent datasets (i.e., NACC, OASIS and AIBL dataset), especially the OASIS dataset.

IV. DISCUSSION

The existing PET prediction methods usually synthesize the whole image of FDG-PET and use CNN based GAN architectures, while our proposed method can simultaneously make predictions for the quantitative descriptions of both FDG-PET and AV45-PET. Although compared to the existing GAN based method of synthesizing the entire image, our method only predicts overall quantitative descriptions of PET images, we actually use the embedded features of PET when the model is further used for predicting AD status, which is similar to the GAN based method when it was used for AD status prediction [9], [28]. Moreover, compared to GAN synthesizing the entire PET image, the predicted SUVRs are simpler and more intuitive, and our method also avoids the potential instability induced by the GAN architecture. Besides, different modalities of PET may contain complementary information and have different pathological implications, hence multi-modal PET is preferable. The experimental results also demonstrate the effectiveness of our multi-modal method for AD status prediction, while it only needs sMRI as input after training. In addition, to some extent, the proposed model can also provide interpretation for AD status prediction tasks by simultaneously predicting the FDG/AV45-PET SUVRs to better assist clinicians in diagnosis.

The good performance of 3D-Mixer may be attributed to our employment of MLP-Mixer architecture instead of the extensively used CNN [17], [18], [25], [54]. For CNN based models, the sizes of receptive fields are restricted by the number of stacked layers. Specifically, for 3D MRI images, the convolutional layers use 3D convolution kernels. Hence, model sizes grow heavily to obtain sufficiently large receptive fields. However, a single Mixer block will be theoretically sufficient for obtaining a receptive field that covers whole input images. This allows us to better control the model size. And our experiment results also show that our MLP-Mixer architecture model outperformed other CNN models in AD status prediction. Here, for each competing method, we showed its model size, i.e., the number of learnable parameters, and its AD status prediction performance in term of AUC in Fig. 5, where 3D-Mixer* denotes the model with the halved width of 3D-Mixer. We observed that the proposed models outperformed other competing methods with similar model sizes for each independent dataset. Hence, the proposed 3D-Mixer showed higher parameter efficiency. Higher parameter efficiency can be beneficial for both academic research and real-world applications, for reducing the risk of overfitting and the consumption of computing and storage resources while keeping good performance.

We also noticed that all competing methods reached a relatively lower performance in the prognosis prediction of MCI as well as in migrating the ADNI-trained model to some independent datasets (e.g., OASIS). Firstly, the relatively

lower performance in the MCI prognosis prediction might be attributed to the difficulty of the task itself. For example, there are many possible causes of MCI (e.g., AD, vascular dementia, depression), under which the mechanism may be complex [55]. Moreover, despite strong heterogeneity among patients with MCI, imaging differences between sMCI and pMCI may not be as significant as those between CN and AD [56]. Concerning the performance decline when migrating the ADNI-trained model to some other datasets, this is to be expected due to the batch effect and other differences (e.g., subject inclusion criteria) across different datasets [57], [58], [59]. For example, sMRIs from other independent datasets could be acquired with various parameters. In particular, for the OASIS datasets, the performance decline could also be attributed to the inconsistency of diagnostic criteria compared to ADNI [60].

Despite the effectiveness of the proposed model, we acknowledge several limitations. Firstly, we show the effectiveness of the MLP-Mixer architecture, however, we only used a simple strategy (concatenation) to integrate the features of different modalities, where the interaction between different modalities is not fully considered. Secondly, the heterogeneity of the sMRIs from different datasets is not taken into account in this work. For instance, the performance usually declined when the ADNI-trained model was applied to other external datasets (Tables II and III). Besides, currently, we only consider sMRI, FDG-PET, and AV45-PET while other modalities [61], [62] of data might also be sensitive to AD (e.g., tau PET). In our future work, we planned to design better architecture to model the interaction of features of different modalities and utilize the harmonization methods [63], [64] for MRI data to reduce the heterogeneity of them to further improve our prediction. Moreover, other modalities of data will also be considered in our future works.

V. CONCLUSION

In this article, we proposed a model based on MLP-Mixer architecture, i.e., 3D-Mixer, consisting of a regression module for simultaneously predicting FDG/AV45-PET SUVRs and a classification module for AD status prediction. By benchmarking on independent datasets, we showed that our proposed model can effectively predict FDG/PET-PET SUVRs, which also exhibit good discrimination capacity between CN and AD subjects. Moreover, by integrating sMRIs with the multi-modal PET embedding features derived from the regression module, 3D-Mixer outperforms other competing methods for AD status prediction and shows better generalization to external independent datasets.

ACKNOWLEDGMENT

Data collection and sharing of ADNI used in this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the

following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Neil Kowall, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).

Data used in the preparation of this article were partly obtained from the MIRIAD database and OASIS-3. The MIRIAD investigators did not participate in analysis or writing of this report. The MIRIAD dataset is made available through the support of the U.K. Alzheimer's Society (Grant RF116). The original data collection was funded through an unrestricted educational grant from GlaxoSmithKline (Grant 6GKC). The Principal Investigators for OASIS-3: T. Benzinger, D. Marcus, J. Morris; NIH P50

AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.

Data used in the preparation of this article were partly obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (www.loni.usc.edu/ADNI). The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at www.aibl.csiro.au.

REFERENCES

- [1] A. Association, "2019 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, vol. 15, no. 3, pp. 321–387, Mar. 2019.
- [2] C. L. Masters, R. Bateman, K. Blennow, C. C. Rowe, R. A. Sperling, and J. L. Cummings, "Alzheimer's disease," *Nature Rev. Dis. Primers*, vol. 1, pp. 1–18, 2015.
- [3] R. J. Perrin, A. M. Fagan, and D. M. Holtzman, "Multimodal techniques for diagnosis and prognosis of Alzheimer's disease," *Nature*, vol. 461, no. 7266, pp. 916–922, 2009.
- [4] G. Chetelat et al., "Accelerated cortical atrophy in cognitively normal elderly with high β -amyloid deposition," *Neurology*, vol. 78, no. 7, pp. 477–484, Feb. 2012.
- [5] D. F. Wong et al., "In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18 F-AV-45 (Flobetapir F 18)," *J. Nucl. Med.*, vol. 51, no. 6, pp. 913–920, Jun. 2010.
- [6] C. Marcus, E. Mena, and R. M. Subramaniam, "Brain PET in the diagnosis of Alzheimer's disease," *Clin. Nucl. Med.*, vol. 39, no. 10, pp. e413–e426, Oct. 2014.
- [7] L. Brand, K. Nichols, H. Wang, L. Shen, and H. Huang, "Joint multi-modal longitudinal regression and classification for Alzheimer's disease prediction," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1845–1855, Jun. 2020.
- [8] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *Neuroimage*, vol. 55, no. 3, pp. 856–867, Apr. 2011.
- [9] Y. Pan, M. Liu, C. Lian, Y. Xia, and D. Shen, "Spatially-constrained fisher representation for brain disease identification with incomplete multi-modal neuroimages," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2965–2975, Sep. 2020.
- [10] X. Zhu, H.-I. Suk, and D. Shen, "Multi-modality canonical feature selection for Alzheimer's disease diagnosis," in *Proc. Med. Image Comput. Comput.-Assist. Interv.*, 2014, pp. 162–169.
- [11] S. Qiu et al., "Development and validation of an interpretable deep learning framework for Alzheimer's disease classification," *Brain*, vol. 143, no. 6, pp. 1920–1933, Jun. 2020.
- [12] Y. Ding et al., "A deep learning model to predict a diagnosis of Alzheimer disease by using 18 F-FDG PET of the brain," *Radiology*, vol. 290, no. 3, pp. 456–464, 2019.
- [13] W. B. Scoville and B. Milner, "Loss of recent memory after bilateral hippocampal lesions," *J. Neurol., Neurosurgery, Psychiatry*, vol. 20, no. 1, pp. 11–21, Feb. 1957.
- [14] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, "Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease," *Med. Image Anal.*, vol. 46, pp. 26–34, May 2018.
- [15] Z. Ning, Q. Xiao, Q. Feng, W. Chen, and Y. Zhang, "Relation-induced multi-modal shared representation learning for Alzheimer's disease diagnosis," *IEEE Trans. Med. Imag.*, vol. 40, no. 6, pp. 1632–1645, Jun. 2021.
- [16] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.
- [17] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, 2018.
- [18] W. Zhu, L. Sun, J. Huang, L. Han, and D. Zhang, "Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2354–2366, Sep. 2021.

- [19] T. Abuhmed, S. El-Sappagh, and J. M. Alonso, "Robust hybrid deep learning models for Alzheimer's progression detection," *Knowl.-Based Syst.*, vol. 213, 2021, Art. no. 106688.
- [20] X. Liu, K. Chen, T. Wu, D. Weidman, F. Lure, and J. Li, "Use of multi-modality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease," *Transl. Res.*, vol. 194, pp. 56–67, 2018.
- [21] A. Morinaga et al., "A comparison of the diagnostic sensitivity of MRI, CBF-SPECT, FDG-PET and cerebrospinal fluid biomarkers for detecting Alzheimer's disease in a memory clinic," *Dement. Geriatr. Cogn. Disord.*, vol. 30, no. 4, pp. 285–292, 2010.
- [22] T. Goel, R. Sharma, M. Tanveer, P. N. Suganthan, K. Maji, and R. Pilli, "Multimodal neuroimaging based Alzheimer's disease diagnosis using evolutionary RVFL classifier," *IEEE J. Biomed. Health Inform.*, early access, Feb. 06, 2023, doi: [10.1109/JBHI.2023.3242354](https://doi.org/10.1109/JBHI.2023.3242354).
- [23] Y. Huang, J. Xu, Y. Zhou, T. Tong, and X. Zhuang, "Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network," *Front. Neurosci.*, vol. 13, May 2019, Art. no. 509.
- [24] F. Yang et al., "Combining PET with MRI to improve predictions of progression from mild cognitive impairment to Alzheimer's disease: An exploratory radiomic analysis study," *Ann. Transl. Med.*, vol. 10, no. 9, pp. 513–513, May 2022.
- [25] R. Li et al., "Deep learning based imaging data completion for improved brain disease diagnosis," in *Proc. Med. Image Comput. Comput.-Assist. Interv.*, 2014, pp. 305–312.
- [26] Y. Pan, M. Liu, C. Lian, Y. Xia, and D. Shen, "Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multi-modal neuroimages," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2019, pp. 137–145.
- [27] S. Hu, B. Lei, S. Wang, Y. Wang, Z. Feng, and Y. Shen, "Bidirectional mapping generative adversarial networks for brain MR to PET synthesis," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 145–157, Jan. 2022.
- [28] Y. Pan, M. Liu, Y. Xia, and D. Shen, "Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6839–6853, Oct. 2022.
- [29] J. L. Whitwell et al., "Clinical and MRI models predicting amyloid deposition in progressive aphasia and apraxia of speech," *NeuroImage Clin.*, vol. 11, pp. 90–98, 2016.
- [30] F. H. Reith, E. C. Mormino, and G. Zaharchuk, "Predicting future amyloid biomarkers in dementia patients with machine learning to improve clinical trial patient selection," *Alzheimer's Dement. Transl. Res. Clin. Interv.*, vol. 7, no. 1, pp. 1–8, Jan. 2021.
- [31] C. Chu, K. Minami, and K. Fukumizu, "Smoothness and stability in GANs," in *Proc. Int. Conf. Learn. Representations*, 2020. [Online]. Available: <https://openreview.net/forum?id=HJeOekHKwr>
- [32] H. Thanh-Tung, T. Tran, and S. Venkatesh, "Improving generalization and stability of generative adversarial networks," in *Proc. Int. Conf. Learn. Representations*, 2019. [Online]. Available: <https://openreview.net/forum?id=ByxPYjC5KQ>
- [33] W. J. Jagust et al., "Relationships between biomarkers in aging and dementia," *Neurology*, vol. 73, no. 15, pp. 1193–1199, Oct. 2009.
- [34] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, vol. 29. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2016/file/c8067ad1937f728f51288b3eb986afaa-Paper.pdf
- [35] I. O. Tolstikhin et al., "MLP-mixer: An all-MLP architecture for vision," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 24261–24272.
- [36] D. L. Beekly et al., "The national Alzheimer's coordinating center (NACC) database: The uniform data set," *Alzheimer Dis. Assoc. Disord.*, vol. 21, no. 3, pp. 249–258, Jul. 2007.
- [37] K. A. Ellis et al., "The Australian imaging, biomarkers and lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease," *Int. Psychogeriatrics*, vol. 21, no. 4, pp. 672–687, Aug. 2009.
- [38] P. LaMontagne et al., "OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease," *medrxiv*, 2019, doi: [10.1101/2019.12.13.19014902](https://doi.org/10.1101/2019.12.13.19014902).
- [39] I. B. Malone et al., "MIRIAD—Public release of a multiple time point Alzheimer's MR imaging dataset," *Neuroimage*, vol. 70, pp. 33–36, Apr. 2013.
- [40] W. Penny, K. Friston, J. Ashburner, S. Kiebel, and T. Nichols, *Statistical Parametric Mapping*. Amsterdam, The Netherlands: Elsevier, 2007.
- [41] S. M. Landau et al., "Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI," *Neurobiol. Aging*, vol. 32, no. 7, pp. 1207–1218, 2011.
- [42] K. Hara, H. Kataoka, and Y. Satoh, "Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet?," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6546–6555.
- [43] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2021. [Online]. Available: <https://openreview.net/forum?id=YicbFdNTTy>
- [44] A. L. W. Bokde et al., "The effect of brain atrophy on cerebral hypometabolism in the visual variant of alzheimer disease," *Arch. Neurol.*, vol. 58, no. 3, pp. 480–486, Mar. 2001.
- [45] D. Tosun, N. Schuff, W. Jagust, and M. W. Weiner, "Relationship between regional brain amyloid- β deposition and brain atrophy rates in mild cognitive impairment," *Alzheimer's Dement.*, vol. 6, no. 4S_Part_20, Jul. 2010, Art. no. e15.
- [46] R. Y. Lo, "Longitudinal change of biomarkers in cognitive decline," *Arch. Neurol.*, vol. 68, no. 10, pp. 1257–1266, Oct. 2011.
- [47] N. Villain et al., "Relationships between hippocampal atrophy, white matter disruption, and gray matter hypometabolism in Alzheimer's disease," *J. Neurosci.*, vol. 28, no. 24, pp. 6174–6181, Jun. 2008.
- [48] G. Chetelat et al., "Direct voxel-based comparison between grey matter hypometabolism and atrophy in Alzheimer's disease," *Brain*, vol. 131, no. 1, pp. 60–71, Dec. 2007.
- [49] L. G. Apostolova et al., "3D PIB and CSF biomarker associations with hippocampal atrophy in ADNI subjects," *Neurobiol. Aging*, vol. 31, no. 8, pp. 1284–1303, Aug. 2010.
- [50] M. S. Kang et al., "Amyloid-beta modulates the association between neurofilament light chain and brain atrophy in Alzheimer's disease," *Mol. Psychiatry*, vol. 26, no. 10, pp. 5989–6001, Oct. 2021.
- [51] L. A. Van De Pol, A. Hensel, F. Barkhof, H. J. Gertz, P. Scheltens, and W. M. Van Der Flier, "Hippocampal atrophy in Alzheimer disease: Age matters," *Neurology*, vol. 66, no. 2, pp. 236–238, Jan. 2006.
- [52] N. Schuff et al., "MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers," *Brain*, vol. 132, no. 4, pp. 1067–1077, May 2008.
- [53] L. G. Apostolova et al., "3D comparison of low, intermediate, and advanced hippocampal atrophy in MCI," *Hum. Brain Mapping*, vol. 31, no. 5, pp. 786–797, Feb. 2010.
- [54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [55] S. Gauthier et al., "Mild cognitive impairment," *Lancet*, vol. 367, no. 9518, pp. 1262–1270, 2006.
- [56] C.-Y. Wee, P.-T. Yap, and D. Shen, "Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns," *Hum. Brain Mapping*, vol. 34, no. 12, pp. 3411–3425, Dec. 2013.
- [57] J.-P. Fortin, E. M. Sweeney, J. Muschelli, C. M. Crainiceanu, and R. T. Shinohara, "Removing inter-subject technical variability in magnetic resonance imaging studies," *Neuroimage*, vol. 132, pp. 198–212, May 2016.
- [58] J. Radua et al., "Increased power by harmonizing structural MRI site differences with the Combat batch adjustment method in ENIGMA," *Neuroimage*, vol. 218, Sep. 2020, Art. no. 116956.
- [59] E. E. Bron et al., "Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease," *NeuroImage Clin.*, vol. 31, Jan. 2021, Art. no. 102712.
- [60] J. Wen et al., "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Med. Image Anal.*, vol. 63, 2020, Art. no. 101694.
- [61] E. Westman, J. S. Muehlboeck, and A. Simmons, "Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion," *Neuroimage*, vol. 62, no. 1, pp. 229–238, 2012.
- [62] D. J. Selkoe and J. Hardy, "The amyloid hypothesis of Alzheimer's disease at 25 years," *EMBO Mol. Med.*, vol. 8, no. 6, pp. 595–608, 2016.
- [63] H. Guan, Y. Liu, E. Yang, P. T. Yap, D. Shen, and M. Liu, "Multi-site MRI harmonization via attention-deep guided domain adaptation for brain disorder identification," *Med. Image Anal.*, vol. 71, 2021, Art. no. 102076.
- [64] R. Pomponio et al., "Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan," *Neuroimage*, vol. 208, Mar. 2020, Art. no. 116450.