frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 19 May 2025
DOI 10.3389/fnins.2025.1576931

OPEN ACCESS

EDITED BY
Shuqiang Wang,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Yongxiang Wang,
Shandong Provincial Hospital, China
Qiankun Zuo,
Hubei University of Economics, China

*CORRESPONDENCE
Raymond Kai-Yu Tong
✉ kytong@cuhk.edu.hk
Sadia Shakil
✉ sadiashakil@cuhk.edu.hk

RECEIVED 14 February 2025
ACCEPTED 17 April 2025
PUBLISHED 19 May 2025

CITATION
Ren Z, Zhou M, Shakil S and Tong RK-Y (2025) Alzheimer's disease recognition via long-range state space model using multi-modal brain images.
Front. Neurosci. 19:1576931.
doi: 10.3389/fnins.2025.1576931

Alzheimer's disease recognition via long-range state space model using multi-modal brain images

Ziyin Ren, Meng Zhou, Sadia Shakil* and Raymond Kai-Yu Tong*

Department of Biomedical Engineering, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China

As a persistent neurodegenerative abnormality, Alzheimer's disease (AD) is affecting an increasing number of elderly people. The early identification of AD is critical for halting the disease progression at an early stage. However, the extraction and fusion of multi-modal features at different scales from brain images remains a challenge for effective AD recognition. In this work, a novel feature fusion long-range state space model (FF-LSSM) model is suggested for effective extraction and fusion of multi-level characteristics from scannings of MRI and PET. The FF-LSSM can extract whole-volume features at every scale and effectively decide their global dependencies via adopted 3D Mamba encoders. Moreover, a feature fusion block is employed to consolidate features of different levels extracted by each encoder to generate fused feature maps. A classifier is cascaded at the end, using the fused features to produce the predicted labels. The FF-LSSM model is optimized and evaluated using brain images of subjects from the ADNI dataset. The inference result on the testing set reveals the FF-LSSM accomplishes a classification ACC of 93.59% in CN vs. AD and 79.31% in sMCI vs. pMCI task, proving its effectiveness in disease classification. Finally, the introduction of the Grad-CAM method illustrates that the implied FF-LSSM can detect AD- and MCI-related brain regions effectively.

KEYWORDS

Alzheimer's disease, long-range sequential modeling, mild cognitive impairment, multi-modal brain images, multi-modality integration

1 Introduction

Alzheimer's disease (AD) is an irreversibly progressive central nervous system degenerative disorder, which has become the major cause of most dementia among old citizens: there are over fifty million AD sufferers worldwide, according to statistics (Dadar et al., 2017; Zou et al., 2024). As patients progress to AD, their cognitive abilities and memory gradually decline. Mild cognitive impairment (MCI) is the initial stage of this decline. Depending on its severity and progression, MCI can be further classified as either the earlier stable MCI (sMCI) stage or the later progressive MCI (pMCI) stage. After the MCI stage, as the continuous neurodegeneration, patients will eventually progress to AD, leading to a complete loss of self-care capabilities (Scheltens et al., 2021; Colom-Cadena et al., 2020). The governments and families of AD patients will face significant financial burdens due to the disease's high incidence. Nowadays, there is no effective treatment for AD; current medications only work to lessen symptoms and halt the disease's progression (Pawar et al., 2025). However, early diagnosis of Alzheimer's remains essential as it can provide early intervention and thus prohibit AD progression at an early stage. Recently, the development of artificial intelligence (AI) has made it possible to detect AD using deep learning (DL) methods.

Brain imaging can provide rich information about patients' pathological and anatomical status. Currently, many kinds of brain imaging methods have been used in AD detection. The ability of magnetic resonance imaging (MRI) to detect subtle structural alterations in the brain, including AD biomarkers like atrophy in the hippocampus, frontal lobe, and temporal lobe, makes it a common diagnostic tool for AD (Hunter et al., 2024). Due to this capability, several tools have been proposed using structural MRI for AD detection in past years (Lin et al., 2018; Mofrad et al., 2021; Inan et al., 2024). However, structural AD biomarkers often appear after the disease has progressed; when AD progression is just getting started, these biomarkers are usually undetectable. Another common neuroimaging method for identifying AD is Positron Emission Tomography (PET). During PET imaging, a radiotracer is administered into the patient's body, and metabolic activity can be measured by observing the accumulation of radio substances. In practice, different radiotracers will be chosen according to distinct imaging purposes. Among them, Fluorine-18 fluorodeoxyglucose (18F-FDG) is one of the most popular radiopharmaceuticals. As a glucose analog, the uptake of 18F-FDG is a marker for glucose consumption and thus can highlight active brain regions during PET imaging (Li and Tang, 2015). Recent research has shown that brain 18F-FDG PET can detect signs of AD neuropathy in people with MCI earlier than brain MRI, making it an informative tool for early AD recognition (Nobili et al., 2018). Therefore, there are an increasing number of methods based on 18F-FDG PET modality for AD screening have been proposed (Chen et al., 2022; Duan et al., 2023; Rogeau et al., 2024). However, PET imaging has limitations in terms of resolution and signal-to-noise ratio, which makes it difficult to obtain information on small scales. A potential improvement is to fuse features from two imaging modalities. It is possible to combine PET's superior early-stage AD detection capabilities with the high spatial resolution of MRI.

The fusion of multi-modal features for computer-aided diagnosis has become increasingly popular, many approaches have been suggested in recent years for detecting Schizophrenia (Kanyal et al., 2024), ADHD (Sethu and Vyas, 2020; Yao et al., 2021), ASD (Wang Q. et al., 2023; Abbas et al., 2023), and also AD (Zuo et al., 2023a,b, 2024a,b; Zong et al., 2024). The current multi-modal approaches for AD identification can be broadly divided into two types: those that rely on conventional machine learning (ML) and those that depend on DL. ML-based methods usually use the constructed classifier for disease recognition after extracting features from the brain's pre-defined regions of interest (ROIs). For example, a linear support vector machine is adopted by Zhang et al. (2011) to incorporate multi-modal features such as cerebrospinal fluid (CSF) biomarkers and 93 ROIs' MRI tissue volumes and PET intensity values for AD classification. The work in Tong et al. (2017) utilized a framework based on non-linear graph fusion to integrate biomarkers from different modalities and adopted a random forest algorithm to distinguish MCI, AD, and cognitive normal (CN) control. A multi-modal progressive graph-based transductive learning method is suggested by Wang et al. (2017) that can gradually learn latent intrinsic representations from MRI and PET imaging to achieve optimized dementia classification. The work in Shi et al. (2019) suggested diagnosing AD and MCI using the interaction of coupled representations from MRI and PET features. DL methods become more and more prevalent in recent years because, in contrast to more conventional ML approaches, DL-based approaches can not only automatically learn representations from the input without manually designed features, but they can also extract characteristics from the latent space to capture the abstract expression of data. The work in Lu et al. (2018), for example, suggested a multi-scale deep neural network (DNN) that uses MRI and PET modalities to predict the conversion to MCI and AD. By identifying the interaction between the multi-modal images, Zhang and Shi (2020) proposed a feature fusion model based on the residual network and attention mechanism. The work by Fang et al. (2020) designed an ensemble classifier for AD prediction by fusing each modality's slice-wise probabilistic score generated by three different deep convolution neural network (CNN) models. The work by Hu et al. (2023) utilized VGG-16-based CNN and multi-head self-attention mechanism to build a VGG-TSwinformer model using T1- and T2-weighted MRI for MCI transition prediction. However, many models concentrate solely on modality-shared representations or modality-specific information while neglecting feature integration. Additionally, most previous methods divide input images into patches without considering the extraction and amalgamation of features at various scales from the voxel to the global level.

In the analysis of high-dimensional medical images, as the input size increases, the computational complexity and convergence time of traditional architectures, such as Transformer, grow drastically, which limits their application potentials. Recently, a novel DL framework, Mamba, has been proposed to tackle the challenge of modeling long sequences (Gu and Dao, 2023). The basic structure of Mamba is based on the state space model (SSM) (Kalman, 1960), by introducing a selection mechanism and a hardware optimization method, Mamba can capture global long-range dependencies and achieve higher training and inference efficiency. Although it was originally designed to handle long sequence modeling problems, the linear complexity of SSM can effectively handle the computational challenges brought by high-dimensional medical images. Many studies have incorporated Mamba frameworks into computer vision and medical image applications, such as the Vision Mamba (Zhu et al., 2024), which introduces the Vim block for enhanced location-aware visual understanding. The work by Ma et al. (2024) proposed the U-Mamba that integrates the Mamba layer into the nnUNet encoder for 2D medical image segmentation. SegMamba is an innovative architecture incorporating a 3D Mamba encoder for segmenting 3D medical images (Xing et al., 2024). The 3D Mamba encoder can convert an input 3D feature map to three long sequences in different orders and thus enable the SSM to model 3D features. Like its extraordinary performance in long-sequence modeling of large language models, it is observed that the SSM can not only extract whole-volume 3D features at every scale but also effectively decide their global dependencies. All these works together demonstrate the great potential of the Mamba architecture in DL based medical imaging analysis. Using high-resolution brain imaging data, such as MRI and PET, to diagnose AD often requires capturing the most subtle status changes over long sequences and processing the relationship between multiple potential degeneration locations, which is a task suitable for introducing SSM-based models.

To tackle the challenge of global dependencies extraction and multi-scale feature fusion, in this research, we propose a Feature Fusion Long-range State Space Model (FF-LSSM) for AD detection and MCI conversion prognosis using MRI and PET images. According to our awareness, it is the first DL model that incorporates the SSM into AD classification. The FF-LSSM takes MRI and PET images of each subject as input. Features from two modalities are extracted separately by two Mamba encoders based on the Tri-orientated Spatial Mamba (TSMamba) block (Xing et al., 2024). A feature fusion block based on the dynamic feature fusion (DFF) module (Yang et al., 2024) is introduced to integrate multi-scale features from distinct modalities. Then, a classifier is introduced to generate the predicted labels from fused feature maps. For better model explainability, we also employ the gradient-weighted class activation mapping (Grad-CAM) method to plot the brain areas that contribute the most to dementia classification.

The remaining paragraphs of this paper are expanded in the sequence outlined below. In Section 2, we describe subjects and image pre-processing methods applied in our research. In Section 3, we thoroughly illustrate the suggested DL model. In Section 4, we clarify the model implementation detail and experimental outcomes. In Section 5, we collate the obtained results with those reported in previous papers and discuss the proposed method's advantages and restrictions. In Section 6, a summary conclusion is given by us.

2 Materials

2.1 Dataset

All data applied in this research is acquired from the public Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, which is a multi-center research program with the goal of discovering and validating biomarkers for AD (Jack et al., 2008). In ADNI, thousands of people from throughout North America were brought together for clinical assessments and brain imaging. This study uses MP-RAGE T1-weighted MRI and 18F-FDG PET images from ADNI studies for model training and evaluation. Data is obtained from the baseline MRI and PET acquisition of 656 individuals, which encompasses 197 from CN group, 162 from sMCI group, 104 from pMCI group, and 193 from AD group. We obtained labels for sMCI and pMCI groups from published data of one previous study (Gao et al., 2021), which are defined based on whether a subject progressed to AD within 36 months after the initial assessment. Since not all volunteers in the ADNI database have data from all modalities, we only screened out those who had both MRI and PET brain images as subjects for this study. The detailed specifics of the selected subjects, including their gender, age, clinical dementia rating (CDR), and mini-mental state examination (MMSE), are itemized in Table 1.

TABLE 1 Demographic information and clinical scores of studied groups (mean ± std).

Group	Gender (F/M)	Age (Years)	CDR	MMSE
CN	102/95	74.1 ± 5.6	0.0 ± 0.1	29.1 ± 1.1
sMCI	55/107	74.6 ± 7.3	1.4 ± 0.8	27.5 ± 1.8
pMCI	43/61	74.7 ± 6.6	1.9 ± 1.0	26.8 ± 1.7
AD	96/97	74.9 ± 8.0	4.5 ± 1.7	23.2 ± 2.2

2.2 Image pre-processing

The 18F-FDG PET images provided by ADNI are all in a slice-wise DICOM format, while the MP-RAGE MRI are in 3D NIFTI format. To ensure the uniformity of the file type and ease subsequent steps, we first use the SPM12 toolbox (Penny et al., 2011) to alter PET files from DICOM to NIFTI format. Subsequently, both MRI and PET images are read and manipulated by script commands of FSL 6.0.7 (Smith et al., 2004; Woolrich et al., 2009; Jenkinson et al., 2012). The pre-processing steps follow a standard pipeline. First, the brain extraction is performed using the BET algorithm (Smith, 2002) to strip the skull from the native space of each subject. Then, all MRI images are linearly registered to the MNI152 standard space by utilizing the FLIRT algorithm (Jenkinson et al., 2002). Afterward, the PET images are first aligned to their corresponding MP-RAGE files in the individual space and then transferred into the template space using the transformation matrix generated in the previous step. After all scannings are registered to the standard space, we remove zero-valued voxels close to the edges and leave the center of the volume with a size of 153 × 180 × 150 mm. The remaining volumes are then down-sampled and resized to a cubic shape of 128 × 128 × 128 mm to reduce the computational complexity. At last, all voxel values in each volume I are normalized between 0 and 1 to facilitate the model training and evaluation. The image pre-processing pipelines are summarized in Figure 1.

3 Method

After the data pre-processing steps are finished, an FF-LSSM framework with four parts is proposed for multi-modal AD classification, as demonstrated in Figure 2. Our suggested DL framework is comprised of the subsequent four elements: (1) two 3D Mamba encoders constructed with down-sampling layers and TSMamba blocks to obtain and combine multi-level features from the voxel to the global level, (2) one feature fusion block based on the DFF module, and (3) one classifier to generate desired labels. The following subsections describe each block's specifics in detail.

3.1 Mamba encoder

As the backbone of the proposed model, two Mamba encoders (Xing et al., 2024) are adopted separately to obtain latent features from the pre-processed MRI and PET volumes. In the Mamba encoder, the input 3D volume first passes a stem 3D convolution layer that has a large kernel size of 7, a stride length of 2, and a padding of 3. In this research, we empirically set the output channel number to 16 for this stem layer. Thus, the stem layer can extract an initially down-sampled feature map $z_0 \in 16 \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}$ for each input volume $I \in 1 \times D \times H \times W$.

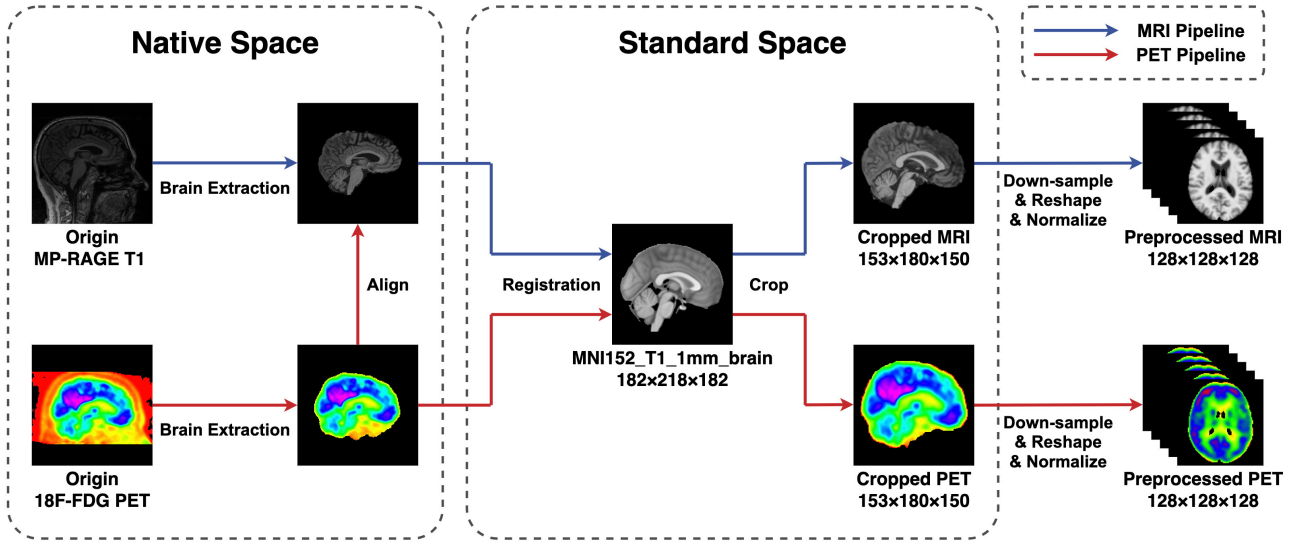

FIGURE 1
The pipeline of image preprocessing applied in this research. The BET method is applied for extracting the brain, and the FLIRT algorithm is used for linear registration. To reduce the computation complexity, all images are first cropped and then down-sampled and reshaped to a size of 128 × 128 × 128 mm. At last, all voxel values are normalized between 0 and 1.

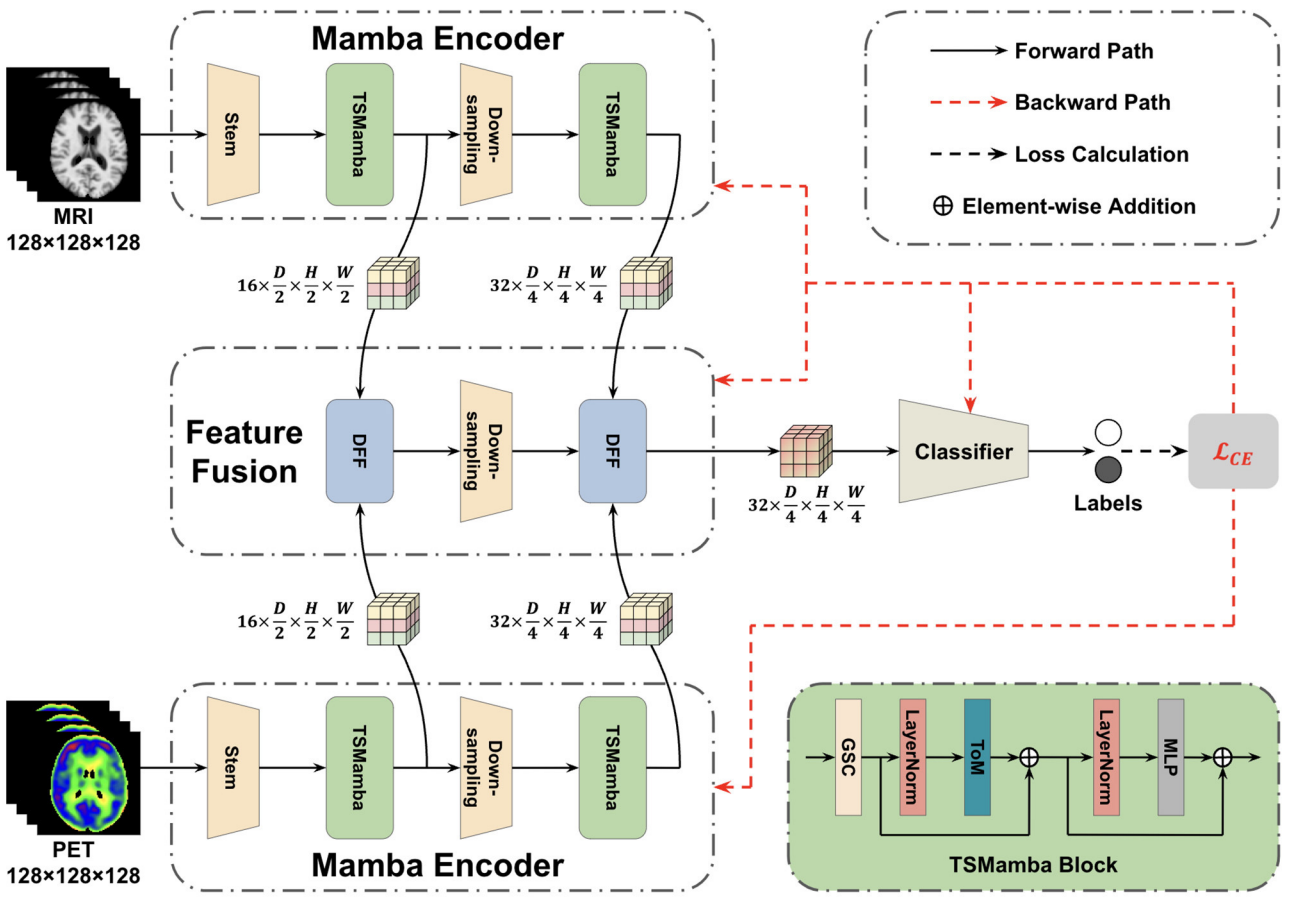

FIGURE 2
Architecture of the suggested FF-LSSM framework. The FF-LSSM contains two 3D Mamba encoders for feature extraction, one feature fusion block to fuse multi-modal features, and one classifier to generate predicted labels.

Following the stem layer, the scaled feature map z_0 is fed into a TSMamba block. As demonstrated in Figure 2, the TSMamba block comprises a Gated Spatial Convolution (GSC) module followed by a Tri-orientated Mamba (ToM) module and a multi-layer perceptron (MLP). To facilitate the training and convergence of the model, residual connections and the layer normalization (LN) operation are also introduced. The GSC module in the TSMamba block is applied to preliminarily process spatial features and relationships in input feature maps. The output of GSC is divided into two parts, one is retained as the residual link, and the other one is fed into a ToM module after the LN operation. Unlike the original Mamba module that is proposed for modeling long time-series 1D sequences, the ToM module is a modified Mamba layer that can calculate feature dependencies from three directions, making it applicable for modeling 3D volumes. For a given 3D input feature map x, the ToM module first flattens it into voxel sequences in three directions, representing the forward x_f, reverse x_r, and inter-slice x_s sequences. Each sequence is processed by an independent Mamba module to extract features and relationships at each level. Finally, the outputs of three Mamba modules are added to achieve the feature integration, which computational descriptions can be denoted as (Xing et al., 2024):

$$ToM(x) = M(x_f) + M(x_r) + M(x_s) \tag{1}$$

where M means the 1D Mamba module. Output of the ToM module is also split into two sections, one is fed into an MLP after the LN operation, and the other one is used as the residual connection. For a given input feature map y_i, the computational expression of the TSMamba block is as follows (Xing et al., 2024):

$$\hat{y}_i = GSC(y_i) \tag{2}$$

$$\tilde{y}_i = ToM(LN(\hat{y}_i)) + \hat{y}_i \tag{3}$$

$$y_{i+1} = MLP(LN(\tilde{y}_i)) + \tilde{y}_i \tag{4}$$

where y_{i+1} represents the output. In this work, we set the input and output channels of all layers in a TSMamba block to be the same, so for an input $z_0 \in 16 \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}$, an output $z_1 \in 16 \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}$ of the same size is returned. Details of the TSMamba block can be found in the salient work of Xing et al. (2024) and will not be repeated here.

For further down-sampling and feature extraction, the feature extracted by the first TSMamba block then passes through a down-sampling 3D convolution layer of kernel 3, stride 2, and padding 1. For the extracted feature of the first TSMamba module $z_1 \in 16 \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}$, the scaled feature $z_2 \in 32 \times \frac{D}{4} \times \frac{H}{4} \times \frac{W}{4}$ is generated by this down-sampling layer. Then, the scaled feature passes through the second TSMamba module for further feature interaction and feature fusion. The output of the second TSMamba module can be denoted as $z_3 \in 32 \times \frac{D}{4} \times \frac{H}{4} \times \frac{W}{4}$.

3.2 Feature fusion block

The DFF module (Yang et al., 2024) is introduced in the feature fusion block to fuse features extracted by the Mamba encoders. The principle of DFF is using global information from the input itself as a guide to adaptively fuse local features at multi-scales. The progress of feature fusion involves the dynamic selection of essential features by considering their global information. The outputs of the first TSMamba block in both MRI and PET Mamba encoders are used as the input of the first DFF module. We can denote two input features of the first DFF module as $F_1^1 \in 16 \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}$ and $F_2^1 \in 16 \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}$. As shown in Figure 3A, a concatenation operation is performed on the channel dimension of inputs F_1^1 and F_2^1, resulting in the feature with a dimension of $32 \times \frac{D}{2} \times \frac{H}{2} \times \frac{W}{2}$. By letting the concatenated input pass through an average pooling operation, a convolution layer with 32 output channels, and a Sigmoid activation in sequence, the global channel information w_{ch}^1 that describes the importance of features can be extracted (Yang et al., 2024):

$$w_{ch}^1 = Sigmoid(c^{1\times1\times1}(AVGPool[F_1^1; F_2^1])) \tag{5}$$

where c stands for the convolution layer. The product of w_{ch}^1 and the concatenated input then passes through a $1 \times 1 \times 1$ convolution layer with 16 output channels to make the channel number the same as the original one. In this way, the global channel information guides the retaining of highlighted features while removing useless interference. The following equation can represent the process (Yang et al., 2024):

$$F^1 = c^{1\times1\times1}(w_{ch}^1 \cdot [F_1^1; F_2^1]) \tag{6}$$

Besides channel information, spatial information is also crucial in the DFF process. While generating global channel information, F_1^1 and F_2^1 also pass through two $1 \times 1 \times 1$ convolution layers with 16 output channels, respectively. The summation of these two convolution layers' outputs then passes through a Sigmoid activation to generate the global spatial information w_{sp} (Yang et al., 2024):

$$w_{sp}^1 = Sigmoid(c^{1\times1\times1}(F_1^1) + c^{1\times1\times1}(F_2^1)) \tag{7}$$

The global spatial information is then multiplied with F^1, which integrates the spatial and channel information, thereby highlighting critical locations on the feature map. Thus, the first DFF module's output can be calculated by Yang et al. (2024):

$$\hat{F}^1 = w_{sp}^1 \cdot F^1 \tag{8}$$

The output of the first DFF module then passes through the same down-sampling 3D convolution layer. This convolution layer resizes the feature map to make it have a dimension of $32 \times \frac{D}{4} \times \frac{H}{4} \times \frac{W}{4}$, the same size as the outputs of the second TSMamba module in both Mamba encoders. The outputs of each Mamba encoder's second TSMamba module are then supplied into the second DFF module together with the fused feature map generated by the first

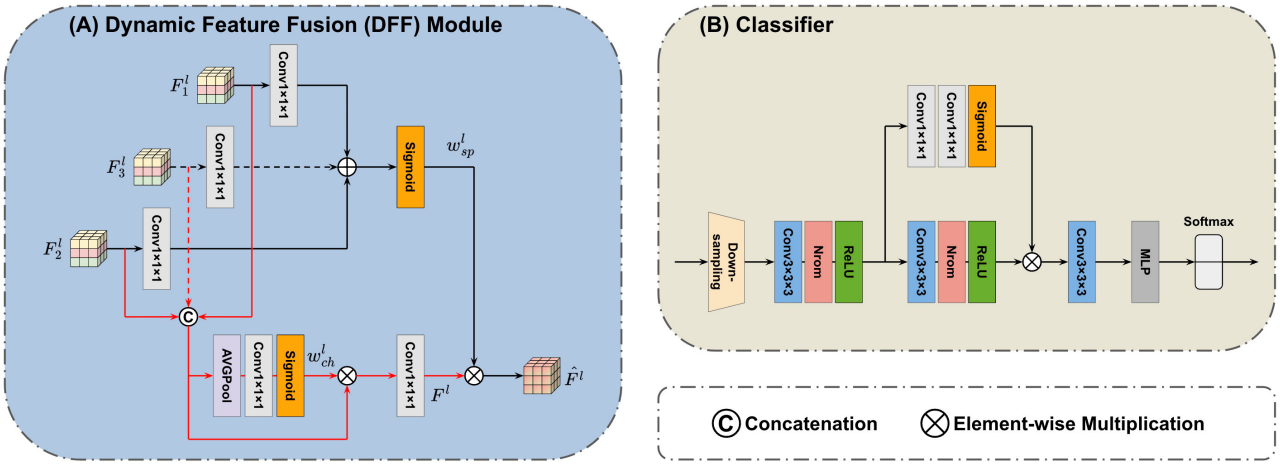

FIGURE 3
Architecture of the applied **(A)** DFF module (Yang et al., 2024) and **(B)** classifier. The DFF consists of two paths. In one path, the global channel information w_{ch}^{l} is extracted from concatenated input features. In another path, the global spatial information w_{sp}^{l} is extracted from input features by separate convolution layers. The final output $\hat{F}^{l}$ integrates both the channel and spatial information. The superscript l indicates the l-th DFF module. The classifier first adopts a down-sampling layer to reshape the feature map. The scaled features are then processed by three convolution layers, one self-attention branch, and an MLP to generate the predicted label.

DFF module. For the second DFF module, which has three inputs, Equations 5–8 can be replaced by:

$$w_{ch}^{2} = Sigmoid(c^{1\times1\times1}(AVGPool[F_1^2; F_2^2; F_3^2])) \tag{9}$$

$$F^{2} = c^{1\times1\times1}(w_{ch}^{2} \cdot [F_1^2; F_2^2; F_3^2]) \tag{10}$$

$$w_{sp}^{2} = Sigmoid(c^{1\times1\times1}(F_1^2) + c^{1\times1\times1}(F_2^2) + c^{1\times1\times1}(F_3^2)) \tag{11}$$

$$\hat{F}^{2} = w_{sp}^{2} \cdot F^{2} \tag{12}$$

Since the second DFF module has an additional input path compared with the first one, the channel number after concatenation becomes 96. Therefore, in the branch that calculates the global channel information, the convolution layers' input and output channels are adjusted appropriately to make sure the size of the output stays constant. The final output for the feature fusion block can be denoted as $z_4 \in 32 \times \frac{D}{4} \times \frac{H}{4} \times \frac{W}{4}$.

3.3 Classifier

After the feature fusion block, a classifier is accessed to generate the predicted result from combined and learned feature maps. As demonstrated by Figure 3B, the input feature map of the classifier first goes through another down-sampling 3D convolution layer, which produces the further rescaled feature map denoted as $z_5 \in 64 \times \frac{D}{8} \times \frac{H}{8} \times \frac{W}{8}$. Subsequently, this feature map asses through two convolution blocks (each consisting of one 3 × 3 × 3 convolution, one normalize, and one ReLU layer) and then one last 3 × 3 × 3 convolution layer followed by an MLP. The fully connected MLP outputs two classes with a Softmax function to produce output labels. Moreover, the classifier adopts a self-attention branch with two cascaded 1 × 1 × 1 convolution layers followed by the Sigmoid activation. This attention branch can highlight important features and thus facilitate modeling multi-level and long-range correlations.

4 Experiments and results

4.1 Implementation

We deploy the suggested DL model leveraging the Pytorch toolkit in a Python 3.10.14 environment. The experimental environment is built on an Ubuntu 18.04 platform with an NVIDIA GeForce RTX3090 GPU accelerated by the CUDA framework. All models are trained and evaluated on two tasks of binary classification: AD detection (CN vs. AD) and MCI prognosis (sMCI vs. pMCI). In both tasks, subjects from the corresponding groups are randomly assigned to the training and testing sets in a 4:1 ratio. The same training and testing sets are used for all models' optimization and inference processes. In the model training, we adopt an Adam optimizer, its learning rate is set to 1×10^{-4}, and a weight decay parameter of 0.02 is added. The training batch size is set to 4, i.e., the brain images of 4 subjects. We define the cross-entropy between the model's output and the subject's real label as the loss function:

$$L_{CE}(p, q) = -\sum_{i=1}^{n} p_i \log q_i \tag{13}$$

where n implies the class number, p_i refers to the real label, and q_i indicates the label predicted by the model. To quantitatively

TABLE 2 Comparison of results for classification of CN vs. AD and sMCI vs. pMCI tasks using uni- and multi-modal images as input.

Modality	CN vs. AD (%)				sMCI vs. pMCI (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
MRI	88.46	86.84	90.00	92.96	72.41	69.56	74.29	69.07
PET	91.03	89.47	92.50	95.10	77.59	73.91	**80.00**	74.91
PT-DCN	92.31	89.74	94.87	95.20	77.59	78.26	77.14	79.13
FF-LSSM	**93.59**	**92.11**	**95.00**	**96.18**	**79.31**	**86.96**	74.29	**80.62**

Bold values indicate the highest value of each metric.

TABLE 3 Results of the ablation study in CN vs. AD and sMCI vs. pMCI classification tasks.

Method	CN vs. AD (%)				sMCI vs. pMCI (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
CNN+Concat	84.62	79.49	89.74	90.86	70.69	60.87	77.14	63.73
CNN+DFF	85.90	84.62	87.18	91.49	72.41	69.57	74.29	71.30
GSC+Concat	87.18	89.74	84.62	93.16	74.14	73.91	74.29	70.56
GSC+DFF	89.74	87.18	92.31	93.03	75.86	65.22	**82.86**	72.17
TSMamba+Concat	92.31	**92.11**	92.50	**96.55**	77.59	73.91	80.00	77.39
FF-LSSM	**93.59**	**92.11**	**95.00**	96.18	**79.31**	**86.96**	74.29	**80.62**

Bold values indicate the highest value of each metric.

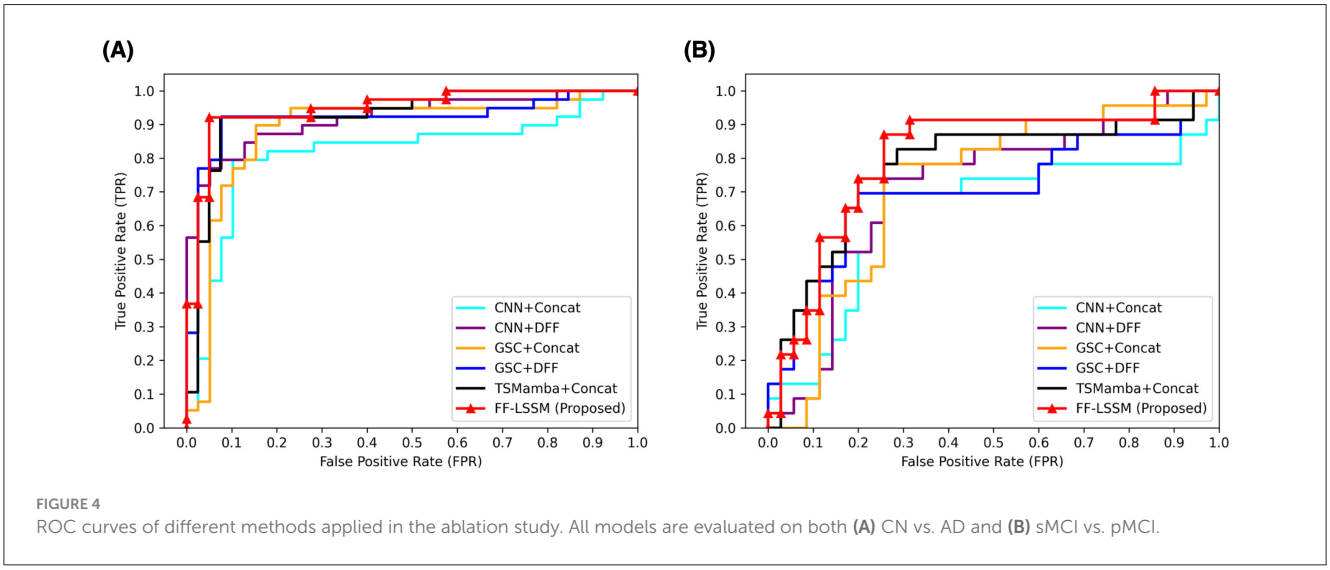

FIGURE 4
ROC curves of different methods applied in the ablation study. All models are evaluated on both **(A)** CN vs. AD and **(B)** sMCI vs. pMCI.

measure model performance, we calculate four commonly used indicators on the test set, which are accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC).

4.2 Effectiveness of disease classification

To investigate the efficacy of the suggested FF-LSSM in AD detection, we first examine its performance on CN vs. AD task. Meanwhile, to verify the advantages of multi-modality methods over uni-modality ones, we also evaluate variant models using only MRI or PET as input. In the variant uni-modality models, we remove one Mamba encoder from them. Since one input path is eliminated, the channel number of their DFF module's convolution layers is adjusted accordingly. Besides, we also implement a SOTA method, the pathwise transfer dense convolution network (PT-DCN) (Gao et al., 2021), on the same training and testing dataset for comparison. Since the input size of the original PT-DCN is different from ours, we change part of its structures and try our best effort to restore its deployment environment for a fair comparison. On the left side of Table 2, we list the results of the aforementioned experiments. Experimental results show that the MRI-based uni-modality model performs the worst, the PET variant performs slightly better, and the proposed FF-LSSM has the best classification performance, achieving 93.59% in ACC. Credit to better abilities of modeling high-dimension medical images and feature integration, the propsed FF-LSSM also outperforms the previous PT-DCN that based on traditional CNN.

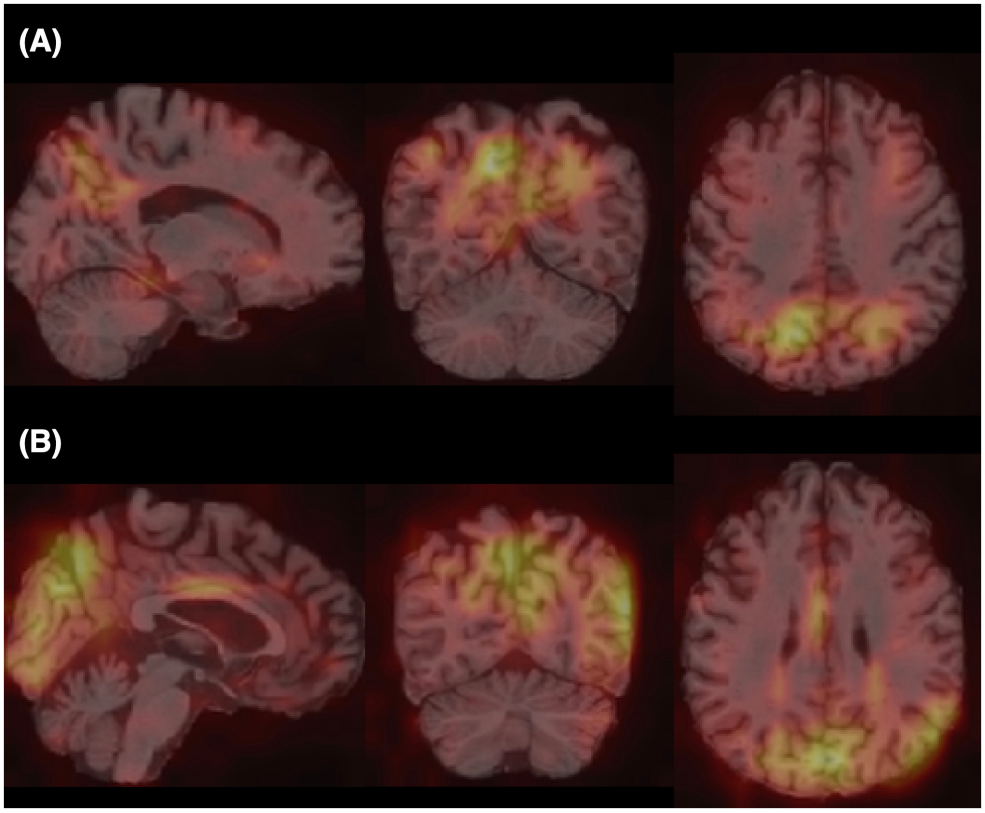

FIGURE 5
Grad-CAM heat maps obtained by **(A)** averaging over all subjects in CN vs. AD classification and **(B)** averaging over all subjects in sMCI vs. pMCI task. A brighter color represents larger gradients at the location.

Subsequently, we conducted the same experiment on the sMCI vs. pMCI task to evaluate the effectiveness of models in MCI prognosis, and the corresponding results are enumerated on the right side of Table 2. In addition to the PET-based variant showing better SPE, the experimental results on the sMCI vs. pMCI task are consistent with the previous CN vs. AD one: FF-LSSM shows the highest 79.31% ACC, followed by the PET-based model and PT-DCN, and then the MRI-based model. This demonstrates the effectiveness of FF-LSSM in the prognosis of MCI and further proves the significance of the multi-modality integration technique.

4.3 Ablation study

After verifying the efficacy of the suggested model, to validate the efficacy of the elements introduced in our model, we conducted ablation experiments. First, the effectiveness of ToM modules is tested by removing the ToM and only retaining GSC modules, resulting in the variant model GSC+DFF. Subsequently, the efficacy of the GSC module is tested by removing GSC modules and only retaining the down-sampling convolution layers, resulting in the variant model CNN+DFF. Furthermore, the efficacy of the DFF module is tested by replacing all DFF modules with concatenation operation in all models, which results in three more variants: TSMamba+Concat, GSC+Concat, and CNN+Concat. The experimental outcomes are demonstrated in Table 3, and corresponding ROC curves are illustrated in Figure 4. It is evident that the ToM module makes a substantial contribution to the model's overall efficiency when the DFF module is kept untouched. Since the ToM is the main feature extractor in the FF-LSSM and is responsible for extracting features at each level and modeling long-range dependencies, removing the ToM module leads to drastic decreases in all measures. Unsurprisingly, removing the GSC module further deteriorates the model performance since the GSC module also contributes to the feature extraction. While keeping the number of channels unchanged, the model with only down-sampling convolution layers performs the worst in both CN vs. AD and sMCI vs. pMCI tasks. This is because it is challenging to model relationships between multi-level features for a plain convolution layer. Furthermore, variant models with the simple concatenation operation always perform worse than models containing the DFF module. It can be observed that introducing the feature fusion mechanism can enhance the capability of all variant models on both classification tasks.

4.4 Feature visualization

Furthermore, for model explainability, the Grad-CAM technique is applied to figure brain areas that are most significant

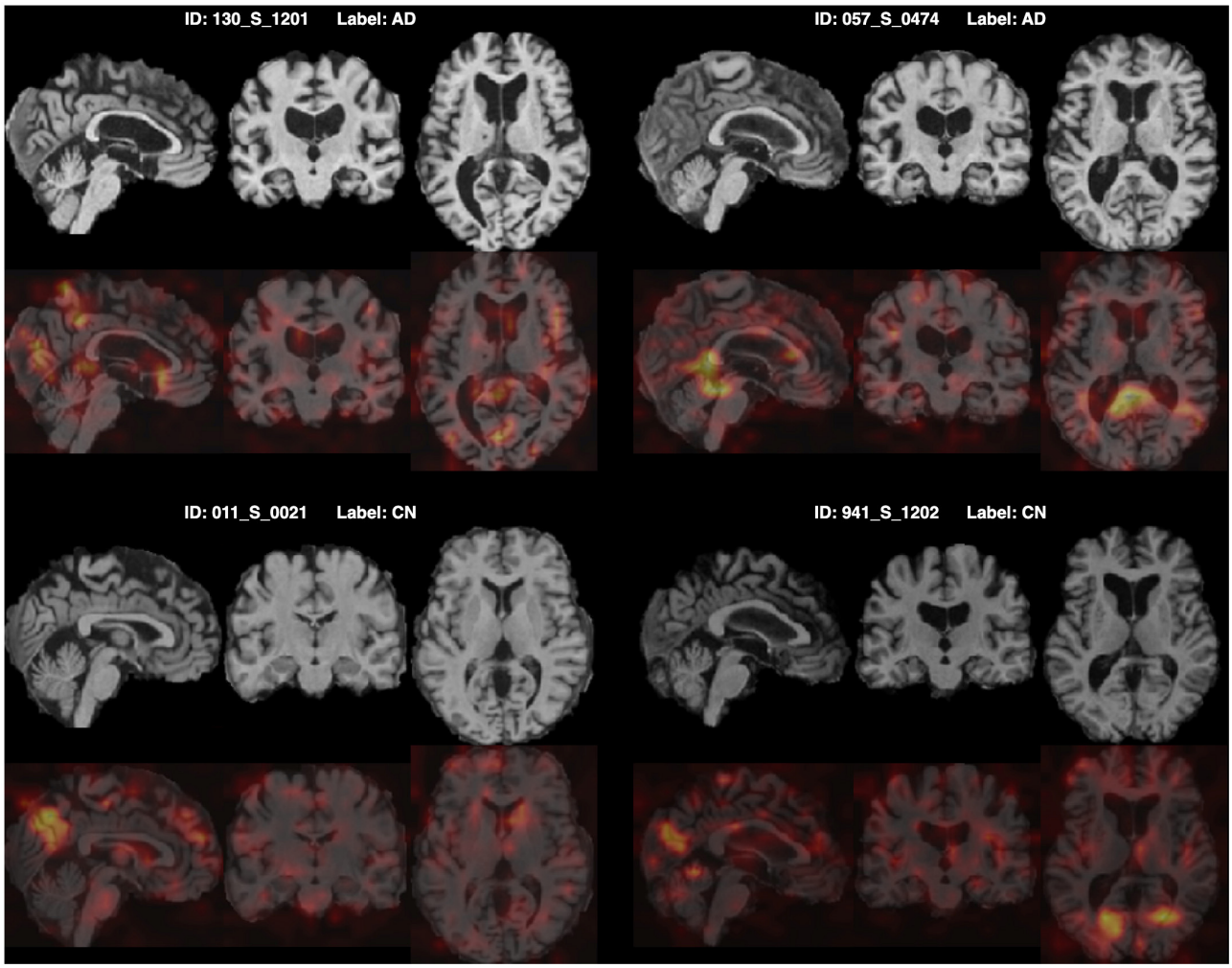

FIGURE 6
Origin MRI images and corresponding Grad-CAM heat maps of four subjects from the CN vs. AD classification.

to the disease classifications (Selvaraju et al., 2020). Grad-CAM uses gradients of the last convolution layer and the predicted results to give a heat map visualization of why a model made its decision. It is a valuable tool for model interpretation. The Grad-CAM heat map averaged on all subjects for CN vs. AD classification is displayed in Figure 5A. It can be noticed that the implied FF-LSSM highlights many regions closely related to AD, such as the Precuneus (Karas et al., 2007), Cuneus (Zheng et al., 2025), Hippocampus (Rao et al., 2022), Caudate (Persson et al., 2018), and Lingual cortex (Liu et al., 2017). Figure 5B shows the averaged Grad-CAM for the sMCI vs. pMCI task. It comes out that the middle Cingulate, Cuneus, Precuneus, and Occipital cortex show the most significant gradients in sMCI and pMCI subjects classification, consistent with previous studies that these regions show significant volume decrease and functional abnormality in MCI progressing (Choo et al., 2010; Pagani et al., 2017; Risacher et al., 2009). In Figures 6, 7, as a reference, we also provide several individuals' MRI images and Grad-CAM heat maps from the CN vs. AD and sMCI vs. pMCI classification, respectively. This result shows that in addition to classifying dementia stages, the proposed FF-LSSM also has great potential in detecting abnormal brain regions associated with AD and MCI. This has positive significance for its future practical applications because it can provide explainable auxiliary information for diagnoses.

5 Discussion

With MRI and PET brain images, we build an FF-LSSM framework to extract and fuse multi-modal features for disease classifications. In Table 4, we compare our results with those reported in previous literature that also applied multi-modal methods based on MRI and PET volumes from the ADNI. The compared methods include the hierarchical feature fusion classification algorithm (Liu et al., 2014), the multiple instance-graph method (Tong et al., 2014), the multi-scale DNN (Lu et al., 2018), the multi-modal DNN with random drop-out (Forouzannezhad et al., 2018), the latent feature representation learning method (Zhou et al., 2019), the dual-modality CNN (Huang et al., 2019; Lin et al., 2021), the PT-DCN (Gao

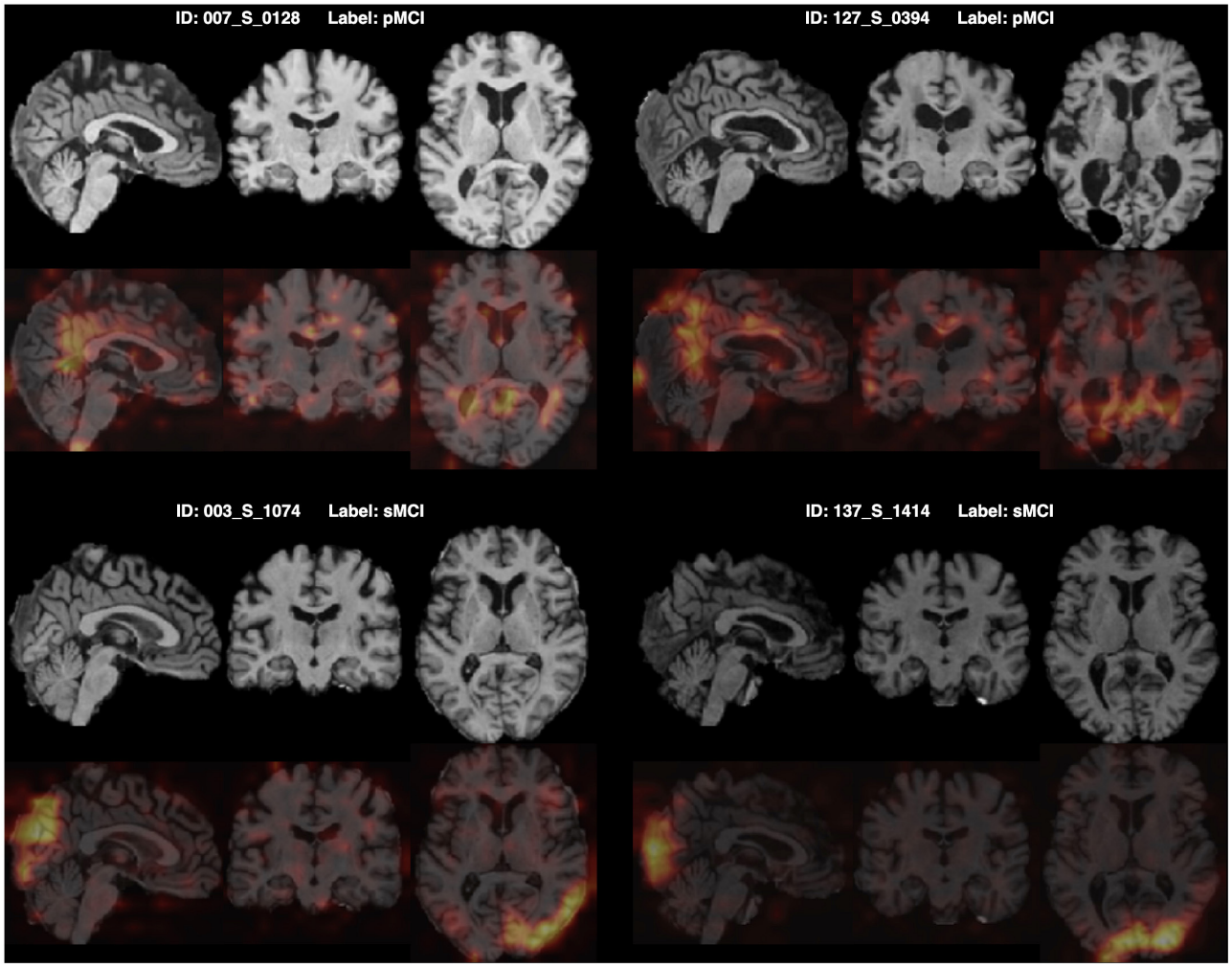

FIGURE 7
Origin MRI images and corresponding Grad-CAM heat maps of four subjects from the sMCI vs. pMCI classification.

TABLE 4 Comparison with classification performance of previous multi-modal methods.

Method	Subjects				CN vs. AD (%)				sMCI vs. pMCI (%)			
	CN	sMCI	pMCI	AD	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Liu et al. (2014)	229	-	-	198	82.2	77.4	86.1	88.1	-	-	-	-
Tong et al. (2014)	231	238	167	198	90.0	84.9	92.6	-	70.4	67.0	73.0	-
Lu et al. (2018)	360	409	217	238	82.9	79.7	83.8	-	-	-	-	-
Forouzannezhad et al. (2018)	248	296	193	159	89.1	87.4	92.1	-	68.2	78.1	57.5	-
Zhou et al. (2019)	204	205	157	171	-	-	-	-	74.3	-	-	75.5
Huang et al. (2019)	731	441	326	647	90.1	90.9	89.2	90.8	76.9	68.2	**83.9**	79.6
Lin et al. (2021)	308	233	183	362	92.3	90.4	94.4	92.8	74.1	75.0	73.1	76.6
Gao et al. (2021)	427	342	234	352	92.0	89.1	94.0	95.6	75.3	77.3	74.1	78.6
Zhang et al. (2023)	129	-	-	110	91.1	91.0	91.1	94.1	-	-	-	-
Ours	197	162	104	193	**93.6**	**92.1**	**95.0**	**96.2**	**79.3**	**87.0**	74.3	**80.6**

Bold values indicate the highest value of each metric.

et al., 2021), and the multi-modal cross-attention AD diagnosis framework (Zhang et al., 2023). Considering most of the above studies do not provide pre-trained models or attach source codes, for comparison, we straightly use their published results in the literature. According to the results, the model we suggested outperforms previous approaches in diagnosing AD and predicting the prognosis of MCI. This is because methods used in previous studies are primarily based on ML or convolution-based layers, which cannot handle the long-range dependencies between local features. The introduction of the TSMamba module successfully solves this problem and boosts classification accuracy.

However, the proposed model also has many limitations and there is room for further improvement in many aspects. First, only MP-RAGE MRI and 18F-FDG PET are applied in the current study. While the amalgamation of these two modalities can grant rich intelligence for dementia classification, the incorporation of other modalities, such as DTI images and genetic data, can enhance model performance to a greater degree. In addition, all experiments are performed solely on the ADNI dataset. Since different datasets usually have distinct imaging devices and acquisition parameters, migrating between them may encounter difficulties due to data diversities. Fortunately, domain adaptation and transfer learning technologies have largely solved the challenge of data heterogeneity (Zhou et al., 2023). In future studies, using multiple datasets for model training and evaluation can be more conducive to its clinical application since this scenario simulates the data heterogeneity of different medical centers. Third, due to the fine resolution of the MRI and PET images, the whole-volume feature extraction is very time-consuming. In the adopted ToM module, modeling a 64 $\times$ 64 $\times$ 64 input map can lead to a sequential length of about 260k, which is a large demand for both the time and GPU memory. Due to the limitation of GPU capacity, we only introduce two TSMamba layers in each Mamba encoder. Reducing the input resolution can reduce time and computational requirements, allowing us to stack more TSMamba layers in the encoder, but the reduced image resolution may also destroy small-scale features and worsen the model performance. Thus, the trade-off between input resolution and model scale is a direction that needs further experiments. Finally, acquiring PET modality is often more difficult than MRI due to its higher cost and potential radiation hazards. Even in large public datasets, such as ADNI, only a small fraction of the subjects have recorded PET scannings. Therefore, for multi-modality models, data sparsity is also a problem that needs to be overcome in the future. We exclusively study individuals who have both MRI and PET scans in the current study; however, to make the model more widely applicable, it is necessary to consider subjects with incomplete images. Fortunately, with the recent remarkable development of generative models (Wang S.-Q. et al., 2023; Wang et al., 2025), it has become possible to produce reliable PET images from other modalities using generative AIs. In future studies, using generative models to produce data for subjects with missing modalities will undoubtedly enhance the universality of the model.

6 Conclusion

For effective AD recognition, we suggest an innovative DL framework termed FF-LSSM in this study. The FF-LSSM contains two 3D Mamba encoders, one feature fusion block, and one classifier. The FF-LSSM uses MRI and PET volumes of every individual as input to conduct dementia classification by extracting and fusing multi-modal features at different scales. The experimental results show that, compared to the uni-modality model, FF-LSSM accomplishes higher classification accuracy in two binary classification tasks, which is 93.59% for the AD detection and 79.31% for the MCI prognosis, respectively. This result proves the advantage of multi-modality feature fusion and the feasibility of the suggested approach in AD detection and MCI prognosis. When juxtaposed with methods proposed by previous literature, the FF-LSSM suggested by us also achieves better classification performance. Finally, the extracted features are visualized using the Grad-CAM technique and show patterns consistent with previous studies, demonstrating the model's explainability in dementia recognition.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

ZR: Methodology, Writing – original draft, Writing – review & editing, Conceptualization, Visualization. MZ: Visualization, Writing – review & editing. SS: Funding acquisition, Supervision, Writing – review & editing. RT: Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Chinese University of Hong Kong under project ID 4055211.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

References

Abbas, S. Q., Chi, L., and Chen, Y-P. P. (2023). Deepmnf: deep multimodal neuroimaging framework for diagnosing autism spectrum disorder. *Artif. Intell. Med.* 136:102475. doi: 10.1016/j.artmed.2022.102475

Chen, Y., Wang, H., Zhang, G., Liu, X., Huang, W., Han, X., et al. (2022). Contrastive learning for prediction of alzheimer's disease using brain 18f-fdg pet. *IEEE J. Biomed. Health Inform.* 27, 1735–1746. doi: 10.1109/JBHI.2022.3231905

Choo, I. H., Lee, D. Y., Oh, J. S., Lee, J. S., Lee, D. S., Song, I. C., et al. (2010). Posterior cingulate cortex atrophy and regional cingulum disruption in mild cognitive impairment and alzheimer's disease. *Neurobiol. Aging* 31, 772–779. doi: 10.1016/j.neurobiolaging.2008.06.015

Colom-Cadena, M., Spires-Jones, T., Zetterberg, H., Blennow, K., Caggiano, A., DeKosky, S. T., et al. (2020). The clinical promise of biomarkers of synapse damage or loss in alzheimer's disease. *Alzheimers Res Ther.* 12, 1–12. doi: 10.1186/s13195-020-00588-4

Dadar, M., Pascoal, T. A., Manitsirikul, S., Misquitta, K., Fonov, V. S., Tartaglia, M. C., et al. (2017). Validation of a regression technique for segmentation of white matter hyperintensities in alzheimer's disease. *IEEE Trans. Med. Imaging* 36, 1758–1768. doi: 10.1109/TMI.2017.2693978

Duan, J., Liu, Y., Wu, H., Wang, J., Chen, L., and Chen, C. P. (2023). Broad learning for early diagnosis of alzheimer's disease using fdg-pet of the brain. *Front. Neurosci.* 17:1137567. doi: 10.3389/fnins.2023.1137567

Fang, X., Liu, Z., and Xu, M. (2020). Ensemble of deep convolutional neural networks based multi-modality images for alzheimer's disease diagnosis. *IET Image Process* 14, 318–326. doi: 10.1049/iet-ipr.2019.0617

Forouzannezhad, P., Abbaspour, A., Li, C., Cabrerizo, M., and Adjouadi, M. (2018). "A deep neural network approach for early diagnosis of mild cognitive impairment using multiple features," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)* (Orlando, FL: IEEE), 1341–1346. doi: 10.1109/ICMLA.2018.00218

Gao, X., Shi, F., Shen, D., and Liu, M. (2021). Task-induced pyramid and attention gan for multimodal brain image imputation and classification in alzheimer's disease. *IEEE J. Biomed. Health. Inform.* 26, 36–43. doi: 10.1109/JBHI.2021.3097721

Gu, A., and Dao, T. (2023). Mamba: linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*. doi: 10.48550/arXiv.2312.00752

Hu, Z., Wang, Z., Jin, Y., and Hou, W. (2023). Vgg-tswinformer: transformer-based deep learning model for early alzheimer's disease prediction. *Comput. Methods Programs Biomed.* 229:107291. doi: 10.1016/j.cmpb.2022.107291

Huang, Y., Xu, J., Zhou, Y., Tong, T., Zhuang, X., and Alzheimer's Disease Neuroimaging Initiative (ADNI). (2019). Diagnosis of alzheimer's disease via multi-modality 3d convolutional neural network. *Front. Neurosci.* 13:509. doi: 10.3389/fnins.2019.00509

Hunter, T. R., Santos, L. E., Tovar-Moll, F., and De Felice, F. G. (2024). Alzheimer's disease biomarkers and their current use in clinical research and practice. *Mol. Psychiatry* 30, 1–13. doi: 10.1038/s41380-024-02709-z

Inan, M. S. K., Sworna, N. S., Islam, A. M., Islam, S., Alom, Z., Azim, M. A., et al. (2024). A slice selection guided deep integrated pipeline for Alzheimer's prediction from structural brain MRI. *Biomed. Signal Process. Control* 89:105773. doi: 10.1016/j.bspc.2023.105773

Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27, 685–691. doi: 10.1002/jmri.21049

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Kalman, R. E. (1960). *A* new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 35–45. doi: 10.1115/1.3662552

Kanyal, A., Mazumder, B., Calhoun, V. D., Preda, A., Turner, J., Ford, J., et al. (2024). Multi-modal deep learning from imaging genomic data for schizophrenia classification. *Front. Psychiatry* 15:1384842. doi: 10.3389/fpsyt.2024.1384842

Karas, G., Scheltens, P., Rombouts, S., Van Schijndel, R., Klein, M., Jones, B., et al. (2007). Precuneus atrophy in early-onset alzheimer's disease: a morphometric structural MRI study. *Neuroradiology* 49, 967–976. doi: 10.1007/s00234-007-0269-2

Li, Z., and Tang, J. (2015). "Deep matrix factorization for social image tag refinement and assignment," in *2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)* (IEEE), 1–6. doi: 10.1109/MMSP.2015.7340796

Lin, W., Lin, W., Chen, G., Zhang, H., Gao, Q., Huang, Y., et al. (2021). Bidirectional mapping of brain MRI and pet with 3D reversible gan for the diagnosis of alzheimer's disease. *Front. Neurosci.* 15:646013. doi: 10.3389/fnins.2021.646013

Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., et al. (2018). Convolutional neural networks-based MRI image analysis for the alzheimer's disease prediction from mild cognitive impairment. *Front. Neurosci.* 12:777. doi: 10.3389/fnins.2018.00777

Liu, M., Zhang, D., Shen, D., and Initiative, A. D. N. (2014). Hierarchical fusion of features and classifier decisions for alzheimer's disease diagnosis. *Hum. Brain Mapp.* 35, 1305–1319. doi: 10.1002/hbm.22254

Liu, X., Chen, W., Hou, H., Chen, X., Zhang, J., Liu, J., et al. (2017). Decreased functional connectivity between the dorsal anterior cingulate cortex and lingual gyrus in alzheimer's disease patients with depression. *Behav. Brain Res.* 326, 132–138. doi: 10.1016/j.bbr.2017.01.037

Lu, D., Popuri, K., Ding, G. W., Balachandar, R., and Beg, M. F. (2018). Multimodal and multiscale deep neural networks for the early diagnosis of alzheimer's disease using structural mr and fdg-pet images. *Sci. Rep.* 8:5697. doi: 10.1038/s41598-018-22871-z

Ma, J., Li, F., and Wang, B. (2024). U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv*:2401.04722. doi: 10.48550/arXiv.2401.04722

Mofrad, S. A., Lundervold, A. J., Vik, A., and Lundervold, A. S. (2021). Cognitive and MRI trajectories for prediction of alzheimer's disease. *Sci. Rep.* 11:2122. doi: 10.1038/s41598-020-78095-7

Nobili, F., Arbizu, J., Bouwman, F., Drzezga, A., Agosta, F., Nestor, P., et al. (2018). European association of nuclear medicine and european academy of neurology recommendations for the use of brain 18f-fluorodeoxyglucose positron emission tomography in neurodegenerative cognitive impairment and dementia: Delphi consensus. *Eur. J. Neurol.* 25, 1201–1217. doi: 10.1111/ene.13728

Pagani, M., Nobili, F., Morbelli, S., Arnaldi, D., Giuliani, A., Öberg, J., et al. (2017). Early identification of mci converting to ad: a fdg pet study. *Eur. J. Nucl. Med. Mol. Imaging* 44, 2042–2052. doi: 10.1007/s00259-017-3761-x

Pawar, S., Rauf, M. A., Abdelhady, H., and Iyer, A. K. (2025). Tau-targeting nanoparticles for treatment of Alzheimer's disease. *Exploration* 5:20230137. doi: 10.1002/EXP.20230137

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Amsterdam: Elsevier.

Persson, K., Bohbot, V., Bogdanovic, N., Selbæk, G., Brækhus, A., and Engedal, K. (2018). Finding of increased caudate nucleus in patients with Alzheimer's disease. *Acta Neurol. Scand.* 137, 224–232. doi: 10.1111/ane.12800

Rao, Y. L., Ganaraja, B., Murlimanju, B., Joy, T., Krishnamurthy, A., and Agrawal, A. (2022). Hippocampus and its involvement in Alzheimer's disease: a review. *3 Biotech* 12:55. doi: 10.1007/s13205-022-03123-4

Risacher, S. L., Saykin, A. J., Wes, J. D., Shen, L., Firpi, H. A., and McDonald, B. C. (2009). Baseline mri predictors of conversion from MCI to probable ad in the adni cohort. *Curr. Alzheimer Res.* 6, 347–361. doi: 10.2174/156720509788929273

Rogeau, A., Hives, F., Bordier, C., Lahousse, H., Roca, V., Lebouvier, T., et al. (2024). A 3d convolutional neural network to classify subjects as Alzheimer's disease, frontotemporal dementia or healthy controls using brain 18f-fdg pet. *NeuroImage* 288:120530. doi: 10.1016/j.neuroimage.2024.120530

Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., et al. (2021). Alzheimer's disease. *Lancet* 397, 1577–1590. doi: 10.1016/S0140-6736(20)32205-4

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-cam: visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* 128, 336–359. doi: 10.1007/s11263-019-01228-7

Sethu, N., and Vyas, R. (2020). "Overview of machine learning methods in adhd prediction," in *Advances in Bioengineering* (Singapore: Springer Singapore), 51–71. doi: 10.1007/978-981-15-2063-1_3

Shi, Y., Suk, H-I., Gao, Y., Lee, S-W., and Shen, D. (2019). Leveraging coupled interaction for multimodal alzheimer's disease diagnosis. *IEEE Trans. Neural Netw. Learn. Syst*. 31, 186–200. doi: 10.1109/TNNLS.2019.2900077

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp*. 17, 143–155. doi: 10.1002/hbm.10062

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23, S208–S219. doi: 10.1016/j.neuroimage.2004.07.051

Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D., Initiative, A. D. N., et al. (2017). Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. *Pattern Recognit*. 63, 171–181. doi: 10.1016/j.patcog.2016.10.009

Tong, T., Wolz, R., Gao, Q., Guerrero, R., Hajnal, J. V., Rueckert, D., et al. (2014). Multiple instance learning for classification of dementia in brain MRI. *Med. Image Anal*. 18, 808–818. doi: 10.1016/j.media.2014.04.006

Wang, Q., Wu, M., Fang, Y., Wang, W., Qiao, L., and Liu, M. (2023). "Modularity-constrained dynamic representation learning for interpretable brain disorder analysis with functional MRI," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer: New York), 46–56. doi: 10.1007/978-3-031-43907-0_5

Wang, S., Zhou, T., Shen, Y., Li, Y., Huang, G., and Hu, Y. (2025). Generative AI enables EEG super-resolution via spatio-temporal adaptive diffusion learning. *IEEE Trans. Consum. Electron*. doi: 10.1109/TCE.2025.3528438. [Epub ahead of print].

Wang, S.-Q., Zhang, Z., He, F., and Hu, Y. (2023). Generative AI for brain imaging and brain network construction. *Front. Neurosci*. 17:1279470. doi: 10.3389/fnins.2023.1279470

Wang, Z., Zhu, X., Adeli, E., Zhu, Y., Nie, F., Munsell, B., et al. (2017). Multi-modal classification of neurodegenerative disease by progressive graph-based transductive learning. *Med. Image Anal*. 39, 218–230. doi: 10.1016/j.media.2017.05.003

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., et al. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage* 45, S173–S186. doi: 10.1016/j.neuroimage.2008.10.055

Xing, Z., Ye, T., Yang, Y., Liu, G., and Zhu, L. (2024). "Segmamba: long-range sequential modeling mamba for 3D medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer: New York), 578–588. doi: 10.1007/978-3-031-72111-3_54

Yang, J., Qiu, P., Zhang, Y., Marcus, D. S., and Sotiras, A. (2024). D-net: dynamic large kernel with dynamic feature fusion for volumetric medical image segmentation. *arXiv preprint arXiv:2403.10674*. doi: 10.2139/ssrn.5093171

Yao, D., Yang, E., Sun, L., Sui, J., and Liu, M. (2021). "Integrating multimodal MRIS for adult ADHD identification with heterogeneous graph attention convolutional network," in *Predictive Intelligence in Medicine: 4th International Workshop, PRIME 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 4* (Springer: New York), 157–167. doi: 10.1007/978-3-030-87602-9_15

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., Initiative, A. D. N., et al. (2011). Multimodal classification of alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867. doi: 10.1016/j.neuroimage.2011.01.008

Zhang, J., He, X., Liu, Y., Cai, Q., Chen, H., and Qing, L. (2023). Multi-modal cross-attention network for Alzheimer's disease diagnosis with multi-modality data. *Comput. Biol. Med*. 162:107050. doi: 10.1016/j.compbiomed.2023.107050

Zhang, T., and Shi, M. (2020). Multi-modal neuroimaging feature fusion for diagnosis of alzheimer's disease. *J. Neurosci. Methods* 341:108795. doi: 10.1016/j.jneumeth.2020.108795

Zheng, Y., Gu, H., Kong, Y., and Alzheimer's Disease Neuroimaging Initiative (ADNI). (2025). Statin is associated with higher cortical thickness in early Alzheimer's disease. *Exp. Gerontol*. 202:112698. doi: 10.1016/j.exger.2025.112698

Zhou, M., Xu, Z., and Tong, R. K-Y. (2023). Superpixel-guided class-level denoising for unsupervised domain adaptive fundus image segmentation without source data. *Comput. Biol. Med*. 162:107061. doi: 10.1016/j.compbiomed.2023.107061

Zhou, T., Liu, M., Thung, K.-H., and Shen, D. (2019). Latent representation learning for alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. *IEEE Trans. Med. Imaging* 38, 2411–2422. doi: 10.1109/TMI.2019.2913158

Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., and Wang, X. (2024). Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*. doi: 10.48550/arXiv.2401.09417

Zong, Y., Zuo, Q., Ng, M. K.-P., Lei, B., and Wang, S. (2024). A new brain network construction paradigm for brain disorder via diffusion-based graph contrastive learning. *IEEE Trans. Pattern Anal. Mach. Intell*. 46, 10389–10403. doi: 10.1109/TPAMI.2024.3442811

Zou, Y., Gao, B., Lu, J., Zhang, K., Zhai, M., Yuan, Z., et al. (2024). "Long non-coding RNA CASC15 enhances learning and memory in mice by promoting synaptic plasticity in hippocampal neurons," in *Exploration*, volume 4 (Wiley Online Library), 20230154. doi: 10.1002/EXP.20230154

Zuo, Q., Chen, L., Shen, Y., Ng, M. K.-P., Lei, B., and Wang, S. (2024a). BDHT: generative AI enables causality analysis for mild cognitive impairment. *IEEE Trans. Autom. Sci. Eng*. 22, 5601–5613. doi: 10.1109/TASE.2024.3425949

Zuo, Q., Shen, Y., Zhong, N., Chen, C. P., Lei, B., and Wang, S. (2023a). Alzheimer's disease prediction via brain structural-functional deep fusing network. *IEEE Trans. Neural Syst. Rehabil. Eng*. 31, 4601–4612. doi: 10.1109/TNSRE.2023.3333952

Zuo, Q., Wu, H., Chen, C. P., Lei, B., and Wang, S. (2024b). Prior-guided adversarial learning with hypergraph for predicting abnormal connections in alzheimer's disease. *IEEE Trans. Cybern*. 54, 3652–3665. doi: 10.1109/TCYB.2023.3344641

Zuo, Q., Zhong, N., Pan, Y., Wu, H., Lei, B., and Wang, S. (2023b). Brain structure-function fusing representation learning using adversarial decomposed-VAE for analyzing MCI. *IEEE Trans. Neural Syst. Rehabil. Eng*. 31, 4017–4028. doi: 10.1109/TNSRE.2023.3323432