Artificial Intelligence In Medicine 145 (2023) 102678

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

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

Research paper

Diagnosis of Alzheimer's disease by joining dual attention CNN and MLP based on structural MRIs, clinical and genetic data

Yan-Rui Qiang, Shao-Wu Zhang *, Jia-Ni Li, Yan Li, Qin-Yi Zhou, for the Alzheimer's Disease Neuroimaging Initiative

Key Laboratory of Information Fusion Technology, School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Multi-modality data
Attention mechanism
Convolutional neural network
Multilayer perceptron

ABSTRACT

Alzheimer's disease (AD) is an irreversible central nervous degenerative disease, while mild cognitive impairment (MCI) is a precursor state of AD. Accurate early diagnosis of AD is conducive to the prevention and early intervention treatment of AD. Although some computational methods have been developed for AD diagnosis, most employ only neuroimaging, ignoring other data (e.g., genetic, clinical) that may have potential disease information. In addition, the results of some methods lack interpretability. In this work, we proposed a novel method (called DANMLP) of joining dual attention convolutional neural network (CNN) and multilayer perceptron (MLP) for computer-aided AD diagnosis by integrating multi-modality data of the structural magnetic resonance imaging (sMRI), clinical data (i.e., demographics, neuropsychology), and APOE genetic data. Our DANMLP consists of four primary components: (1) the Patch-CNN for extracting the image characteristics from each local patch, (2) the position self-attention block for capturing the dependencies between features within a patch, (3) the channel self-attention block for capturing dependencies of inter-patch features, (4) two MLP networks for extracting the clinical features and outputting the AD classification results, respectively. Compared with other state-of-the-art methods in the 5CV test, DANMLP achieves 93% and 82.4% classification accuracy for the AD vs. MCI and MCI vs. NC tasks on the ADNI database, which is 0.2%~15.2% and 3.4%~26.8% higher than that of other five methods, respectively. The individualized visualization of focal areas can also help clinicians in the early diagnosis of AD. These results indicate that DANMLP can be effectively used for diagnosing AD and MCI patients.

1. Introduction

Alzheimer's disease (AD) is the most common irreversible primary degenerative disease of the central nervous system in middle or late life [1]. AD begins with a gradual loss of memory and cognitive function, and its pathologic characteristic is the degeneration of specific nerve cells, the presence of neuritis plaques and neurofibrillary tangles [2]. As a transient clinical stage from normal control (NC) to dementia, mild cognitive impairment (MCI) is a critical period for controlling AD progression. Until now, although there is no effective way to reverse the progression of AD, accurate and early diagnosis of AD/MCI is crucial for subsequent effective clinical intervention treatments, delaying the onset of cognitive symptoms, maintaining residual brain functions, and reducing complications [3].

Considering that structural magnetic resonance imaging (sMRI) can describe morphological changes in brain imaging, some machine learning-based methods have been developed to identify AD patients, MCI patients and normal controls (NC) through sMRI images [4–6]. However, these traditional machine learning-based methods rely heavily on the quality of handcrafted features (e.g., cortical thickness, hippocampal volume and gray matter densities) extracted from sMRI images, and also believe that the chosen features are the most discriminating information, which may neglect some important discriminative information inherent in sMRI images.

Given that deep learning (DL) methods, especially convolutional neural networks (CNNs), can generally automatically learn the informative features that have better representations of the data than the handcrafted features, they have been widely applied in various medical image analysis tasks [7]. It has been proved that CNNs have the excellent ability to learn high-level features from sMRI images for greatly improving the performance of brain disease diagnosis [8]. According to the brain partition with different scales, existing CNN-based AD diagnosis methods for sMRI images can be roughly categorized into 2D slice-level methods [9,10], 3D patch-level methods [11–14],

* Corresponding author.
E-mail address: zhangsw@nwpu.edu.cn (S.-W. Zhang).

https://doi.org/10.1016/j.artmed.2023.102678
Received 16 November 2022; Received in revised form 12 July 2023; Accepted 3 October 2023
Available online 5 October 2023
0933-3657/

region-level methods [15], and 3D subject-level methods [16,17]. In 2D slice-level methods [9,10], the 2D slices extracted from the 3D sMRI volume are inputted into 2D CNNs for NC/MCI/AD classification. 2D slice-level methods can increase the number of training samples by extracting more 2D slices from a single 3D sMRI image to alleviate the curse of dimensionality. However, these methods independently analyze all slices of a subject with the 2D convolutional filters, losing the 3D space dependence information of 3D sMRI images. That is, the 3D spatial information is not adequately modeled by 2D slice-level methods. In addition, there are many ways to select 2D slices, which will affect the robustness of classification models. In 3D patch-level methods [11–14], the 3D patches extracted from the 3D sMRI images are inputted into 3D CNNs for AD diagnosis. These methods can use more training samples of 3D patches to train the models to alleviate the curse of dimensionality. The lower number of parameters can be learned by using the same network for all patches, while more detailed features can be learned by using different networks for different patches. However, how to select 3D patches and combine these local patches to represent the whole brain structure well is still a challenge in 3D patch-level methods. In addition, most of these 3D patches are not informative for AD diagnosis because they are not affected by AD disease. In region-level methods [15], the regions of interest (ROI) are segmented from brain sMRI images and then fed into 2D/3D CNN models for AD diagnosis. However, these methods just focus on the ROIs (e.g., hippocampus), while AD alterations span over multiple brain areas, and segmenting these ROIs is resource-intensive. In 3D subject-level methods [16,17], the whole sMRIs are inputted into CNN models for NC/MCI/AD classification at the subject level. Although these methods fully integrate the spatial information of sMRI images, they risk overfitting due to the small number of samples compared to the size of the input sMRI [14]. In addition, 3D subject-level methods have higher computational complexity.

Although existing AD diagnosis methods have achieved better classification results, most of them focus on brain neuroimaging data, such as sMRI, PET and DTI, but rarely consider the potential impact of the clinical (i.e., demographic, neuropsychological, etc.) and genetic data, so the improvement of performance is constrained. Most large-scale genome-wide association studies (GWAS) have revealed associations between single nucleotide polymorphisms (SNPs) and the risk of AD, and SNPs in AD-related genes can profoundly induce significant degradation of certain brain functions [18], such as Apolipoprotein E (APOE) $\epsilon 4$ gene is a high-risk pathogenic gene of AD. APOE $\epsilon 4$ allele carriers are more prone to amyloid deposition, which increases the risk of AD by 3∼4 folds [19,20]. AD clinical data, such as demographic, neuropsychological and cognitive assessment, can measure and track AD processes to help clinicians diagnose [21]. Demographic factors (e.g., age and gender) can influence AD progression. Frequent neuropsychological assessments can easily detect within-subject changes. Cognitive assessment (e.g., mini-mental state examination, MMSE) can capture subtle clinical decline to discern the treatment effects among participants with earlier AD disease. In addition, most multi-modality DL methods adopt the 3D subject-level way. However, high-dimensional images may lead to the CNNs unable to effectively learn the detailed structural features. In contrast, most ROI- and patched-based methods fragment the connections between brain regions, ignoring the correlations between brain structures. Moreover, the results of most existing deep learning-based AD diagnosis methods are less interpretable because of the black-boxed learning procedure. Therefore, it is necessary to develop effective AD diagnosis methods to improve AD classification accuracy and interpretability simultaneously.

In this work, we proposed a novel method (called DANMLP) that joins dual attention CNN and MLP for AD diagnosis by integrating the sMRI, clinical and APOE genetic data. DANMLP consists of the Patch-CNN, position self-attention block, channel self-attention block and two MLPs. The Patch-CNN is used to learn the features within each sMRI patch. The position self-attention block emphasizes the feature pairs with positional correlation within a patch. The channel self-attention block emphasizes the features with channel correlation between patches and obtains output features of each brain region. One MLP is used to extract AD discriminant features from the clinical and genetic data, and the other MLP is adopted to fuse the features extracted from image, clinical and genetic data for NC/MCI/AD classification. Our work aims to improve AD classification accuracy and interpretability simultaneously. The experimental results on the ADNI database demonstrate that our DANMLP is effective in AD diagnosis, which can improve the interpretability of the results.

In summary, the main contributions of our work are as follows:

(1) A joint framework of dual attention CNN and MLP is proposed to integrate sMRI, clinical and genetic data to improve the diagnosis performance of AD.

(2) The Patch-CNN is designed to extract the discriminative features within each patch in 3D sMRI images.

(3) A 3D dual attention block is introduced to capture the inter-position and inter-channel dependencies to effectively extract the spatial structure information of the brain and obtain the output features of each brain region for improving the interpretability of AD diagnosis results.

The rest of this paper is organized as follows. Section 2 introduces the related works. Section 3 presents the materials of sMRI imaging, clinical and APOE genetic data used in this work and our proposed DANMLP method. Section 4 shows the experimental settings and results. Section 5 shows the discussion. Section 5 concludes this paper.

2. Related work

2.1. Alzheimer's disease diagnosis with single-modality data

Most research on AD diagnosis mainly focuses on neuroimaging, in which position emission tomography (PET) and magnetic resonance imaging (MRI) receive more attention. For example, Abramova et al. proposed a multi-view separable pyramid network (MiSePyNet) for AD diagnosis by learning the feature representations from axial, coronal and sagittal views of PET [22]. Although PET, as a good indicator of brain metabolism level, can capture cerebral glucose metabolic rate at resting state and reveal metabolic aberrations before structural brain changes, it is expensive and requires administration or inhalation of a radioisotope as a tracer. While as non-invasive medical imaging techniques, such as functional MRI (fMRI) [23], diffusion tensor imaging (DTI) [24] and structural MRI (sMRI) [9,15,16], use a strong magnetic field and radio frequency pulse to image the internal body structures, which are often used to study the pathological brain changes associated with AD in vivo. fMRI can demonstrate changes in blood sample levels of the brain and assesses brain activity in different states. DTI can show the direction of nerve conduction bundles in the brain's white matter, enabling fine imaging of central nerve fibres. sMRI is sensitive to morphological changes caused by brain atrophy, and it can capture changes in brain anatomy. For example, Gan et al. proposed a functional connectivity network (FCN) analysis framework to reveal the pathological basis of brain diseases based on fMRI images [23]. De and Chowdhury employed three VoxCNNs to separately train three types of 3D volumetric data of Echo Planar Imaging (EPT), Fractional Anisotropy (FA) and Mean Diffusivity (MD) in each DTI scan, and used a random forest (RF) classifier to classify the derived metadata in the form of region-averaged FA and MD values, then the outputs of three VoxCNNs and one RF are combined with a modulated rank averaging decision fusion approach to realize AD classification [24]. Folego et al. developed an ADNet method to realize the multiclass AD biomarker identification task by combining 3D CNN with domain adaptation and using the whole sMRI as input. ADNet is prone to overfitting due to excessive computational complexity [16].

2.2. Alzheimer's disease diagnosis with multi-modality data

AD is a complex and heterogeneous disease, and using single-modality data for AD diagnosis is often insufficient. In view that multi-modality data can provide complementary information to improve AD diagnostic accuracy, some methods of using multi-modality data (i.e., fMRI, DTI, sMRI) have been developed for AD diagnosis. Combining the information from different types of multi-modality data can help to improve AD diagnostic accuracy compared with single-modality methods. For example, Gupta et al. used multiple toolboxes (e.g., DPARSF, FSL, etc.) to extract features from sMRI, fMRI, and DTI, then adopted the multiple kernel learning (MKL) framework to perform AD classification [25]. The performance of this method largely depends on the quality of feature extraction. Huang et al. took both sMRI and PET images of the hippocampal area as the inputs of 3D VGG-variant CNNs to separately learn the features from sMRI and PET images, then concatenated these features to a fully connected network for AD diagnosis [26]. Kang et al. took both sMRI and DTI images as the inputs of the VGG-16 network to learn the slice features of subjects with transfer learning, then adopted the LASSO algorithm to perform feature selection for reducing the feature dimension and redundant information, finally fed the selected features into support vector machine (SVM) classifier to distinguish early MCI from NC [27]. Although existing multi-modality methods show encouraging performance on the AD classification task, most of them are restricted to using the multi-modality neuroimaging data of sMRI, fMRI and PET, and adopting the simple integration strategies, while these multi-modality neuroimaging data contain more redundant information. Simply integrating these multi-modality neuroimaging data would generate redundant noises, which is unfavorable to training the AD classification models and improving their performance. In addition, integrating multi-modality AD data simply by increasing the number of modalities does not increase AD diagnosis power. Considering that sMRI can sensitively capture the changes of brain anatomy and is often used in clinical diagnosis due to its non-invasive and low-cost, and there is more complementary information among clinical, genetic data and sMRI data, here we will design dual attention CNN and MLP to separately learn the AD discriminative features from sMRI images, clinical and genetic data, and then design another MLP to fuse these learned features to realize AD diagnoses.

2.3. Attentional mechanisms in medical imaging

The basic idea of the attentional mechanism comes from the animal's visual system, which is able to focus attention on critical areas when processing large amounts of visual input. The attention mechanism can not only be used as the judgment basis for validating deep learning models, but also can improve their performance by allowing the models to focus more on important features and ignore unimportant features. The attentional mechanisms used in medical images can be broadly classified into three types: hard attention [28],soft attention [29,30] and self-attention [31,32]. Guan et al. first used the hard attention mechanism in medical image processing to classify thorax disease based on chest X-ray images by designing a three-branch attention-guided CNN (AG-CNN) model [28]. However, since the hard attention mechanism takes the non-differentiable form of one-hot encoding, it cannot be trained by back-propagation algorithms commonly used in deep learning, and the training process is often done through reinforcement learning. Unlike the hard attention mechanism, the soft attention mechanism can be easily combined with deep learning, because its learning process is differentiable. For example, Schlemper et al. introduced the attention gates into the standard CNNs model to focus on the target features of varying shapes and sizes [29]. Abramova et al. introduced a squeeze-and-excitation (SE) blocks module into U-Net to reconcile the weight of each channel [30]. Although these methods have made some progress, they only emphasize the influence of a single feature on the classification result and ignore the correlation between features. The self-attention mechanism can solve this problem by making each output feature contain the relationship between the input features. For example, Li et al. developed a 3D self-attention CNN for Low-Dose CT denoising by employing self-attention to capture extensive spatial information within and between CT slices [31]. Shen et al. proposed a multiscale temporal self-attention and dynamic graph convolution hybrid network (MTS-DGCHN) for EEG-based stereogram recognition by using the temporal self-attention block to learn temporal continuity information of EEG signals [32]. Considering the superiority of the self-attention mechanism, and that AD is a complex and heterogeneous disease with numerous connections between brain regions during the progression of the disease, we will adopt a 3D dual self-attention mechanism to fully emphasize the correlation between features in terms of both positions and channels for AD diagnose in this work.

3. Material and methods

In this section, we first present the data used in this work. Then, we introduce how to preprocess these data. Finally, we show our DANMLP model in detail.

3.1. Data and preprocessing

sMRI, clinical and genetic data used in this work were obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.l oni.usc.edu/). In the ADNI database, the subjects are divided into three categories: Alzheimer's disease (AD), mild cognitive impairment (MCI) and normal control (NC) by the standard clinic criteria, such as mini-mental state examination (MMSE) scores, clinical dementia rating (CDR) scores, neuropsychiatric inventory-questionnaire (NPI-Q) scores, and geriatric depression scale (GDS) scores, functional assessment questionnaire (FAQ) scores. We selected 750 subjects from the ADNI database, including 250 AD, 250 MCI, and 250 NC subjects, who have all the magnetization-prepared rapid gradient-echo (MPRAGE) T1-weighted image (sMRI), clinical (i.e., age, gender, MMSE, CDR, etc.) and APOE genotyping data. Each subject's sMRI, clinical and APOE genotyping data were taken within ±6 months. The demographic, neuropsychological and cognitive assessment of 750 subjects from the ADNI database is shown in Table S1.

For subsequent better feature learning and AD diagnosis, we adopted the typical procedures of Anterior Commissure (AC)–Posterior Commissure (PC) correction that can eliminate noise introduced by subject movement during the sMRI scan for more accurate localization and comparison of brain structures in Statistical Parametric Mapping 12 (https://www.fil.ion.ucl.ac.u k/spm/software/spm12/), and skull-stripping, and cerebellum removal in Computational Anatomy Toolbox 12 (http://dbm.neuro.uni-jena.de/cat/) to preprocess the original sMRI images downloaded from ADNI. Then, the corrected sMRI images were segmented into Gray Matter (GM), White Matter (WM), and Cerebrospinal Fluid (CSF) according to the tissue probability map (TPM) template. The GM, WM and CSF images [were] normalized to the Montreal Neurological Institute standard space (MNI) using affine linear registration to generate images with $121 \times 145 \times 121$ voxels. Finally, we removed the border area without information to obtain the images with $100 \times 120 \times 100$ voxels. Only GM images were used in this work.

For the genetic data, the APOE locus contains three alleles, $\epsilon 2$, $\epsilon 3$ and $\epsilon 4$, which can generate three pure heterozygotes (i.e., $\epsilon 2/2$, $\epsilon 3/3$, $\epsilon 4/4$) and three heterozygotes (i.e., $\epsilon 2/3$, $\epsilon 2/4$, $\epsilon 3/4$) for a total of six common phenotypes. ADNI recorded the phenotypes of the subjects. We counted the number (i.e., 0, 1, 2) of APOE $\epsilon 2$ and APOE $\epsilon 4$ contained in each subject to indicate the APOE genotype.

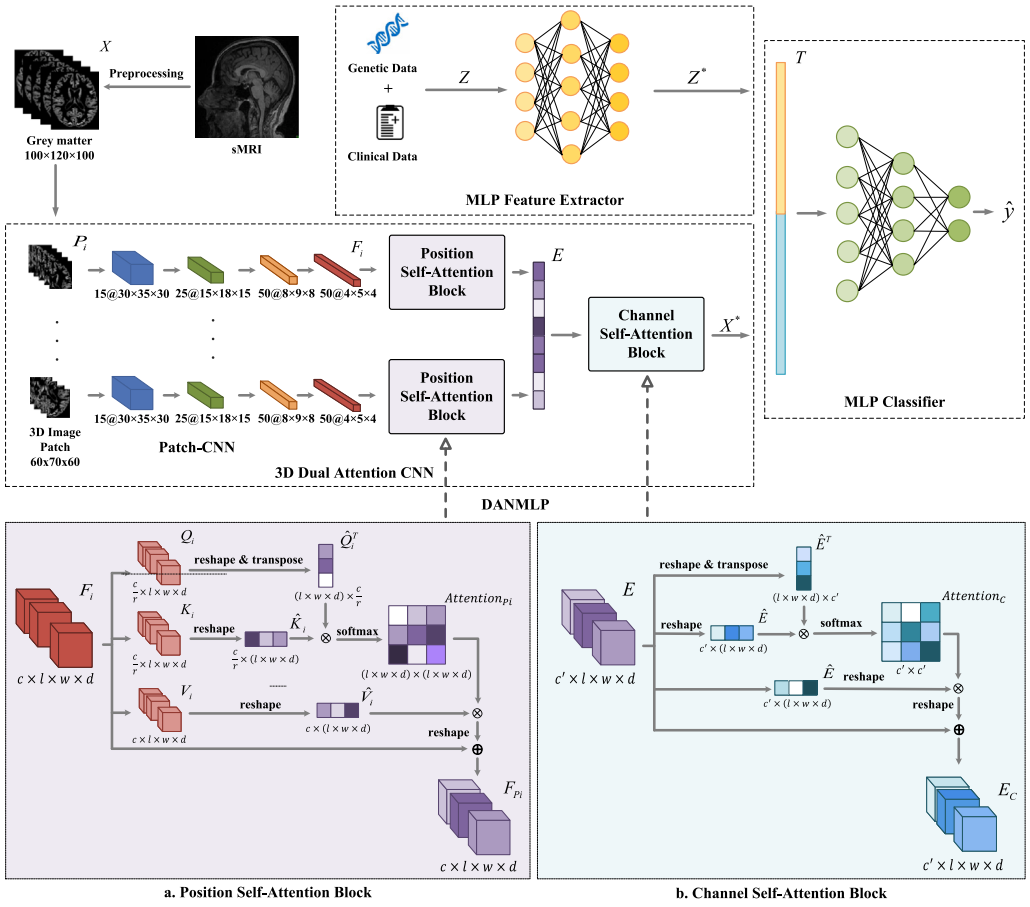

Fig. 1. Schematic of DANMLP model framework. DANMLP comprises three components: 3D dual attention CNN (3D-DACNN), MLP feature extractor and MLP classifier. 3D-DACNN is a patch-level network used to learn the features of sMRI images, which consists of a Patch-CNN with I CNN branches, I position self-attention blocks and a channel self-attention block. MLP feature extractor is used to learn the features from clinical and genetic data. MLP classifier is adopted to fuse the features learned from images, clinical and genetic data for NC/MCI/AD classification. (a) The architecture of the position self-attention block. This block is used to augment the position-dependent features within the patch. (b) The architecture of the channel self-attention block. This block is used to emphasize interdependent feature maps between patches.

3.2. Overall architecture of DANMLP model

We denote the subject dataset as $D = \{(X_n, Z_n, y_n) | n = 1, 2, \ldots, N\}$, where N is the total number of subjects; $X_n = \{P_{n,i} | i = 1, 2, \ldots, I\}$ is the sMRI image of nth subject, $P_{n,i} \in \mathbb{R}^{w_1 \times w_2 \times w_3}$ is the ith patch of sMRI image with $w_1 \times w_2 \times w_3$ size, and I is the total number of patches; Z_n is the clinical and APOE genotype data of nth subject; y_n is the label of nth subject, $y_n \in Y = \{y_1 = '' AD'', y_2 = '' MCI'', y_3 = '' NC''\}$. Our DANMLP model (Fig. 1) consists of three key components: the 3D dual attention CNN (3D-DACNN) to learn the features from sMRI images, the MLP feature extractor to learn the features from the clinical and APOE genotype data, the MLP classifier to realize AD diagnosis by concatenating the outputs of 3D-DACNN and MLP feature extractor. The main mathematical notations used in this study are listed in Table S2.

3.2.1. 3D dual attention CNN

To effectively learn the features from sMRI images, we designed 3D dual attention CNN (3D-DACNN) network that consists of a Patch-CNN with eight CNN branches, eight position self-attention blocks, and a channel self-attention block. The Patch-CNN is used to learn the features from different patches of sMRI. The position self-attention block emphasizes feature pairs with positional dependencies within a patch. The channel self-attention block emphasizes the features with channel dependencies between patches, and fuses the features from different CNN branches. In the following, we will describe each block of the 3D-DACNN in detail.

Patch-CNN The Patch-Net serves to learn more abstract features from the original patches $P_{n,i} (n = 1, 2, \ldots, N, i = 1, 2, \ldots, I)$, and to reduce the size of the feature map. In our DANMLP model framework, the deep Patch-CNN has I CNN branches, each of which processes the patch at the corresponding location. That is, the ith CNN branch processes patch P_i. Each CNN branch composed of four convolutional blocks stacked on top of each other has the same structure. Each convolutional block contains a 3D convolutional layer, a ReLU activation function, and a 3D max-pooling layer. Specifically, each input feature map H_i^l of the lth block from the ith branch is processed by a 3D convolution kernel, followed by a ReLU activation function. For the initial block of the ith branch, $H_i^{(1)} = P_i$. Formally, the value at position (x_1, x_2, x_3) on the jth feature map in the lth convolutional block is given by the following formula,

$$g_{i,j}^{l,x_1,x_2,x_3} = ReLU\left(b_{i,j}^l + \sum_m \sum_{a_1=0}^{A_1-1} \sum_{a_2=0}^{A_2-1} \sum_{a_3=0}^{A_3-1} k_{i,j,m}^{l,a_1,a_2,a_3} h_{i,m}^{l,x_1+a_1,x_2+a_2,x_3+a_3}\right) \quad (1)$$

where (A_1, A_2, A_3) is the size of the 3D convolution kernel K_i^l, $k_{i,j,m}^{l,a_1,a_2,a_3}$ is the (a_1, a_2, a_3)-th value of the kernel connected to the mth feature map in the previous layer. ReLU(·) is a non-linear activation function that outputs the input value when it is greater than or equal to zero, and outputs zero otherwise.

We then use the 3D max-pooling layer to reduce the dimensionality of the feature maps, and the output of lth block $H_{i,j}^{l+1}$ is given by the

following formula,

$$h_{i,j}^{l+1,x_1,x_2,x_3} = \max_{u_1=0}^{U_1-1} \max_{u_2=0}^{U_2-1} \max_{u_3=0}^{U_3-1} g_{i,j}^{l,s_1x_1+u_1,s_2x_2+u_2,s_3x_3+u_3} \quad (2)$$

where (U_1, U_2, U_3) is the filter size of the 3D max-pooling layer, (s_1, s_2, s_3) is the stride size.

Among them, the size of the convolution kernel is $3 \times 3 \times 3$. The pooling layer with a filter size of $2 \times 2 \times 2$ and a stride size of $(2, 2, 2)$ is used for down-sampling. The number of convolution kernels from block1 to block4 is set to 15, 25, 50, and 50 in order.

Position Self-Attention Block To better capture the position dependent features within each patch, we added the position self-attention block at the end of each branch in Patch-CNN. The output $F_i \in \mathbb{R}^{c\times l\times w\times d}$ of ith Patch-CNN's branch is fed into three different 3D convolution layers to generate three new feature maps Q_i, K_i and V_i, where $\{Q_i, K_i\} \in \mathbb{R}^{\frac{c}{r}\times l\times w\times d}$, $V_i \in \mathbb{R}^{c\times l\times w\times d}$, and r denotes the reduction ratio. Then, we computed the position self-attention matrix describing the feature similarity between positions. That is, we transform Q_i and K_i into $\hat{Q}_i$ and $\hat{K}_i$ through the reshaping operation of $\frac{c}{r} \times (l \times w \times d)$ size, and perform matrix multiplication on $\hat{Q}_i^T$ and $\hat{K}_i$ to obtain the position self-attention matrix $Attention_{Pi} \in \mathbb{R}^{(l\times w\times d)\times(l\times w\times d)}$ of ith CNN branch by a softmax layer.

$$Attention_{Pi} = Softmax(\hat{Q}_i^T \hat{K}_i) \quad (3)$$

where Softmax$(\cdot)$ is a probability distribution function that maps the input to the (0,1) range. Its formula is $Softmax\left(\delta_r\right) = \exp\left(\delta_r\right) / \sum_{j=1}^{J} \exp\left(\delta_j\right)$, here δ_r is the rth element of the input vector δ, and J represents the length of δ.

We use $Attention_{Pi}$ to re-weight V_i, and reshaped V_i (with $c\times l\times w\times d$ size) into $\hat{V}_i$(with $c \times (l \times w \times d)$ size). Multiply $Attention_{Pi}$ and $\hat{V}_i$ to obtain the corrected feature map $V^* \in \mathbb{R}^{c\times(l\times w\times d)}$.

$$V_i^* = \hat{V}_i Attention_{Pi} \quad (4)$$

To prevent the gradient vanishing, we adopt a skip connection between new feature map V_i^* and original feature map F_i. Meanwhile, we reshape the V_i^* with $c \times (l \times w \times d)$ size to $\hat{V}_i^*$ with $c \times l \times w \times d$ size.

Finally, we multiply $\hat{V}_i^*$ by a learnable parameter μ, and perform an element-wise sum operation with the features to obtain the position self-attention output $F_{Pi} \in \mathbb{R}^{c\times l\times w\times d}$ of ith patch.

$$F_{Pi} = \mu \hat{V}^* + F_i \quad (5)$$

These position self-attention blocks can capture the features with position dependency regardless of their distance in the position dimension. Thus, we can enhance the feature representation of regions where atrophy occurs together within the same patch.

According to the channel direction, we concatenate the position self-attention outputs F_{Pi} to obtain the position self-attention feature map $E \in \mathbb{R}^{c'\times l\times w\times d}$ $(c' = c \times I)$ of one subject.

$$E = concat\left(F_{P1}, F_{P2}, \ldots, F_{PI}\right) \quad (6)$$

Channel Self-Attention Block Considering that AD is a neurological disease that gradually triggers brain atrophy on a whole-brain scale, and the structure used in Patch-CNN disrupts the intrinsic connections of the brain, we employ the channel self-attention block to explicitly interdependencies between patches for enhancing interdependent feature maps between patches.

The position self-attention feature map $E \in \mathbb{R}^{c'\times l\times w\times d}$ is fed to the channel self-attention block. Then, we reshape E to a matrix $\hat{E}$ with $c' \times (l \times w \times d)$ size, and perform matrix multiplication between $\hat{E}$ and its transpose to generate the channel self-attention matrix $Attention_C \in \mathbb{R}^{c'\times c'}$ by a Softmax layer. Matrices $Attention_C$ and $\hat{E}$ are multiplied to obtain the calibration matrix $E^* \in \mathbb{R}^{c'\times(l\times w\times d)}$.

$$Attention_C = Softmax\left(\hat{E}\hat{E}^T\right) \quad (7)$$

$$E^* = Attention_C \hat{E} \quad (8)$$

By reshaping the size of E^* to produce matrix $\hat{E}^*$ with size $c' \times l \times w \times d$, we multiply $\hat{E}^*$ by a learnable parameter ϵ, and perform an element-wise sum operation to obtain the channel self-attention output E_C of one subject.

$$E_C = \epsilon \hat{E}^* + E \quad (9)$$

The channel self-attention block considers the connections between features in the channel dimension, and emphasizes the feature pairs with strong relevance, meanwhile preserving the connections between key features and unimportant features and avoiding the loss of potentially relevant features. Therefore, features between patches can be well fused in this way.

Finally, the channel self-attention output E_C is inputted into a simple fully connected layer to reduce the feature dimension for obtaining the embedding features X^* of sMRI images.

$$X^* = \mathrm{Re}\, LU\left(W_1 E_C + b_1\right) \quad (10)$$

where W_1 and b_1 are the weight and bias, respectively.

3.2.2. MLP feature extractor

Compared with neuroimaging data, the clinical and genetic data are one-dimension data, so we can represent subjects in the form of vectors. While multilayer perceptron (MLP) is a forward-structured artificial neural network, which maps the input vectors to the embedding output vectors. Therefore, we adopt MLP to learn the embedding features from clinical and APOE genotyping data. In our DANMLP model, the MLP feature extractor consists of alternately stacked twice fully connected layers, ReLU activation functions, and dropout layers used to mitigate overfitting. The feature matrix Z derived from clinical and APOE genotyping data of all subjects is inputted into the MLP feature extractor to learn the embedding feature matrix Z^*.

$$Z^* = ReLU(W_3\left(ReLU\left(W_2 Z + b_2\right)\right) + b_3) \quad (11)$$

where W_2 and W_3 are the trainable weight matrices, b_2 and b_3 are the trainable bias matrices.

3.2.3. MLP classifier

The output matrix X^* from 3D-DACNN and the output matrix Z^* from the MLP feature extractor are concatenated to form an embedding matrix T, which is fed into an MLP classifier to obtain the label $\hat{y}_n$ of nth subject.

$$T = concat\left(X^*, Z^*\right) \quad (12)$$

$$\hat{y} = LogSoftmax(W_5\left(ReLU\left(W_4 T + b_4\right)\right) + b_5) \quad (13)$$

where W_4 and W_5 are the trainable weight matrices, b_4 and b_5 are the trainable bias matrices.

The loss function used in DANMLP is defined as follows:

$$\mathcal{L}_{NLL} = -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log\left(\hat{y}_n\right) + \left(1 - y_n\right)\log\left(1 - \hat{y}_n\right)\right] \quad (14)$$

where N is the total number of subjects, y_n is the true label of nth subject, and $\hat{y}_n$ is the predictive result of DANMLP.

4. Results

In this section, we first describe the experimental settings and evaluation metrics, then present the experimental results of our DANMLP and other comparison methods, as well as the ablation results of DANMLP. Finally, in order to improve the interpretability of our DANMLP, we will examine the discriminative brain regions at overall and individual levels, respectively.

Table 1
Results of DANMLP and other five methods in 5CV test.

Method	AD vs. MCI				MCI vs. NC			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
VoxCNN	0.824	0.832	0.816	0.896	0.556	0.664	0.448	0.562
ResNet	0.886	0.896	0.876	0.937	0.614	0.640	0.588	0.671
Xing's method	0.778	0.768	0.788	0.867	0.568	0.552	0.584	0.593
3DAN	0.870	0.908	0.832	0.925	0.614	0.476	0.752	0.645
Spasov's method	0.928	**0.948**	0.908	**0.959**	0.790	0.740	0.840	0.878
DANMLP	**0.930**	0.940	**0.920**	0.953	**0.824**	**0.764**	**0.884**	**0.895**

Table 2
The ablation experimental results of DANMLP in 5CV test.

Method	AD vs. MCI				MCI vs. NC			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
DANMLP-MLPFE	0.880	0.912	0.848	0.934	0.676	0.752	0.600	0.710
DANMLP-PCSA	0.894	0.880	0.908	0.950	0.766	0.732	0.800	0.820
DANMLP-CSA	0.918	0.920	0.916	0.950	0.788	**0.792**	0.784	0.876
DANMLP-PSA	0.924	0.924	**0.924**	0.952	0.784	0.772	0.796	0.878
DANMLPsub	0.890	0.892	0.888	0.943	0.772	0.692	0.852	0.868
DANMLP	**0.930**	**0.940**	0.920	**0.953**	**0.824**	0.764	**0.884**	**0.895**

4.1. Experimental settings and evaluation metrics

Experimental Settings We evaluated our DANMLP on the classification tasks of AD subjects versus MCI subjects (AD vs. MCI) and MCI subjects versus NC subjects (MCI vs. NC) in five-fold cross validation (5CV) test. For the 5CV test (as shown in Fig. S1), all subjects are randomly divided into five blocks of approximately equal size. One of the five blocks is singled out in turn as the test sample set to evaluate the model performance, and the other four blocks are used as the training and validation sample sets (the sample ratio for training and validation sets is 3:1) to train the model. In each binary classification task, the sample size of each class is equal in training, validation, and test sets. This process is repeated for 5 iterations, each time setting aside a different test block. The average results from 5 folds (or models) are used to evaluate the performance of different methods.

Our DANMLP is trained using the Adam optimizer for 40 epochs. The initial learning rates for the MLP feature extractor and 3D-DACNN are set to 0.001 and 0.0001, respectively. Besides, we used the annealing algorithm to adjust the learning rate, that is, when the loss does not decrease in 5 epochs, the learning rate decreases by 0.1 times. Other hyperparameters of DANMLP are set as follows: batch size = 10, patch number = 8, patch size = 60 × 70 × 60, reduction ratio = 5, dropout rate = 0.5. All experiments run on Linux OS with 24G×4 NVIDIA RTX 3090 GPU, 40 × 2.4 GHz Intel Xeon CPU, and 128 GB RAM. PyTorch 3.7 was adopted to implement our DANMLP. Additionally, we designed the experiments of training DANMLP with different optimization algorithms (as shown in Table S3 and Fig. S2). We also designed the experiments to investigate the training time and inference speed of DANMLP with different parameter sizes (as shown in Table S4). Form Table S3 and Fig. S2, we can see that the Adam optimization technique outperforms the other three optimization techniques in terms of ACC, AUC and SPE. Thus, here we adopt the Adam optimization technique in our DANMLP model. From Table S4, we can see that the training time is almost independent of the model size, while the inference time increases with the increase of the model size. Considering our computer hardware settings, we selected the DANMLP model with 33M (millions) parameters in this work.

Evaluation Metrics We used four metrics to evaluate the classification performance of DANMLP, including accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under the receiver operating characteristic curve (AUC). These metrics are defined as follows: $ACC = \frac{TP+TN}{TP+TN+FP+FN}$, $SEN = \frac{TP}{TP+FN}$, $SPE = \frac{TN}{TN+FP}$, where TP, TN, FP, and FN are denoted as true positive, true negative, false positive, and false negative values, respectively. The ROC (Receiver Operating Characteristic) curve can be plotted by varying the threshold of the binary classifier and calculating the true positive rate ($TPR = SEN$) and false positive rate ($FPR = 1 - SPE$). AUC is the area under the ROC curve.

4.2. Performance comparison of DANMLP with other methods

To evaluate the performance of our DANMLP for diagnosing AD and MCI patients, we compared our DANMLP with other five state-of-the-art methods, such as VoxCNN [33], ResNet [33], Xing's methods [34], 3DAN [17] and Spasov's method [35]. VoxCNN [33] used a VGG-like network architecture to diagnose AD by designing four volumetric convolutional blocks for extracting features from sMRI images, two deconvolutional layers with the batch norm and dropout for regularization. ResNet [33] built 21 layers containing six VoxRes blocks, each with 64 3D filters, to realize AD diagnosis from sMRI data. Xing's method [34] used approximate rank pooling to transform the 3D sMRI image volume into a 2D image, then pre-trained 2D VGG 11 to extract the features that were sent to a small classifier with an attention module for AD diagnosis. 3DAN [17] built a 3D attention network by integrating an attention mechanism with a 3D CNN for AD diagnosis from sMRI data. Spasov's method [35] presented a deep learning architecture by using dual learning and 3D separable convolutions to identify MCI patients from sMRI, local Jacobian determinant image, demographic, neuropsychological, and APOE $\epsilon 4$ genetic data. Considering that the comparison methods did not publicly share their datasets, to ensure fairness in comparing our DANMLP with other methods, we implemented the codes of other methods on our dataset under the same classification task. The codes of VoxCNN, ResNet, and Xing's method are available online from the website they provided. The results of our DANMLP and the other five methods are shown in Table 1.

From Table 1, we can see that for the classification task of MCI vs. NC, the performance of our DANMLP is superior to the other five methods, and the accuracy of DANMLP achieves 0.824, which is 0.268, 0.210, 0.256, 0.210, and 0.034 higher than that of VoxCNN, ResNet, Xing's method, 3DAN and Spasov's method, respectively. For the classification task of AD vs. MCI, the accuracy of DANMLP achieves 0.930, which is 0.106, 0.044, 0.152, 0.060 and 0.002 higher than that of VoxCNN, ResNet, Xing's method, 3DAN and Spasov's method, respectively. In addition, we find that the results of our DANMLP and Spasov's method are obviously better than that of the other four methods based on sMRI data only, indicating that integration of the multi-modality imaging, clinical and genomic data can effectively improve the performance of AD diagnosis, realizing early diagnosis of MCI. Although the performance of Spasov's method using the whole sMRI images as input is very close to our DANMLP, Spasov's method additionally inputs the local Jacobian determinant image information. That is to say, Spasov's method inputs an additional data source than our DANMLP. If DANMLP also inputs the Jacobian determinant images, its performance should be superior to Spasov's method. The results in Table 1 show that our DANMLP has excellent performance in AD-related classification tasks, especially our DANMLP can effectively distinguish MCI patients from the NC population.

In addition, we also adopted the floating point operations (FLOPs) [36] to measure the computational complexity of DANMLP and the other five methods. The FLOPs of DANMLP and the other five methods are shown in Table S5, from which we can see that the computational complexity of DANMLP is lower than that of the other five methods.

4.3. Ablation studies for DANMLP

To evaluate the contributions of diverse architecture components in our DANMLP, we conducted the ablation experiments in the 5CV test. The ablation experimental results of DANMLP are shown in Table 2.

In Table 2, DANMLP-MLPFE denotes that we remove the MLP feature extractor in DANMLP framework; DANMLP-PCSA denotes that we remove the position self-attention blocks and the channel self-attention block in DANMLP framework; DANMLP-CSA denotes that we remove the channel self-attention block in DANMLP framework; DANMLP-PSA denotes that we remove the position self-attention blocks in DANMLP framework; $\text{DANMLP}_{\text{sub}}$ denotes that we utilize one CNN branch and the position self-attention mechanism to extract the features of sMRI image, while the rest of the DANMLP framework remains unchanged. Fig. S3 gives the schematic diagrams of DANMLP-MLPFE, ANMLP-PCSA, DANMLP-CSA, DANMLP-PSA, and $\text{DANMLP}_{\text{sub}}$.

As shown in Table 2, we can see that for MCI vs. NC task, the ACC of DANMLP is 0.148, 0.058, 0.036 and 0.040 higher than that of DANMLP-MLPFE, DANMLP-PCSA, DANMLP-CSA, DANMLP-PSA, respectively. For AD vs. MCI task, the ACC of DANMLP is 0.050, 0.036, 0.012 and 0.006 higher than that of DANMLP-MLPFE, DANMLP-PCSA, DANMLP-CSA, DANMLP-PSA, respectively. These results indicate that the position self-attention blocks, the channel self-attention block, and the MLP feature extractor used in our DANMLP can effectively improve the performance of AD diagnosis. Although the channel self-attention module does not contribute much for improving the performance of DANMLP in AD vs. MCI task (the ACC and AUC of DANMLP are 0.012, 0.003 higher than that of DANMLP-CSA, respectively), its contribution is significant for improving the performance of DANMLP in MCI vs. NC task (the ACC and AUC of DANMLP are 0.036, 0.019 higher than that of DANMLP-CSA), which is a more challenging classification task as the features between MCI subject and NC subject are more similar in this task. From Table 2, we can also find that for the classification task of MCI vs. NC, the ACC of DANMLP is 0.052 higher than that of $\text{DANMLP}_{\text{sub}}$; for the classification task of AD vs. MCI, the ACC of DANMLP is 0.04 higher than that of $\text{DANMLP}_{\text{sub}}$. These results show that when the training samples of AD/MCI subjects are small, the patch-based multi-channel CNNs approach performs better than the subject-based approach.

The contribution of joining the position self-attention blocks and the channel self-attention block for improving the performance of DANMLP is much more than that of using only the position/channel self-attention blocks, and MLP feature extractor, while the contribution of channel self-attention is much more than that of position self-attention. These results show that the position self-attention blocks can effectively capture the discriminative features within patches, while the channel self-attention block captures the discriminative features between patches by fusing the features from different branches. The feature information extracted by the channel self-attention block and the position self-attention blocks is complementary, which helps to improve the performance of AD-related classification tasks. In addition, from Table 2, we can also find that the contribution of the MLP feature extractor for the MCI vs. NC task is greater than that for AD vs. MCI task, indicating that the introduction of genetic and clinical information helps improve the accuracy of AD early diagnosis, and also show that for some MCI patients with slight brain atrophy, their genotype features may vary significantly different from those of NC population. Therefore, integrating genetic and clinical data in sMRI imaging data can effectively improve the accuracy of MCI vs. NC classification.

4.4. Discriminative brain regions for AD and MCI

To investigate the brain regions focused by DANMLP, we conducted the experiments on the output characteristics of each brain region. First, using the Anatomical Automatic Labeling (AAL) template [37] to segment the gray matter image to get 90 brain regions (ROIs) that are used as the inputs of DANMLP, we obtained the feature maps of every brain region from the channel self-attention block. Then, we summed the elements in each brain region feature map to get the ROI index $\Omega_{n,j}$ ($j = 1, 2, \ldots, 90$) of the brain region. Finally, we adopted the Spearman rank correlation analysis and t-test to select the significant brain regions that DANMLP focuses on.

The process of selecting the significant brain regions is as follows: (1) Build two variables X_j and Y. $X_j = [\Omega_{1j}, \Omega_{2j}, \ldots, \Omega_{nj}, \ldots, \Omega_{Nj}]$ is composed of the ROI indices of jth brain region from AD and MCI subjects (or MCI and NC subjects), $j = 1, 2, \ldots, \ldots, 90$, N=500, and $Y = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n, \ldots, \hat{y}_N]$ is composed of the DANMLP output values $\hat{y}_n$ of AD and MCI subjects (or MCI and NC subjects); (2) Calculate the Spearman rank correlation coefficient r_S between X_j and Y; (3) Employ t-test to determine the significance level of $|r_S|$ (i.e., absolute value of r_S).

Table S6 and Table S7 list the top 10 significant brain regions focused by DANMLP for AD vs. MCI and MCI vs. NC tasks, respectively. From Table S6, we can see that for AD vs. MCI task, the amygdala, hippocampus, and parahippocampal gyrus are all highly discriminatory brain regions. It is consistent with the conclusions of many existing AD-related studies [38–40]. The amygdala is closely related to emotion and memory [41]. The hippocampus has been proven to be the earliest damaged area of AD, which is associated with cognitive decline and memory impairment [15,42]. The parahippocampal gyrus is closely linked to the hippocampus, which is also affected by AD [43]. From Table S7, we can find that the supplementary motor areas, cuneus and precuneus, that induce MCI have been confirmed by other studies [44–47].

4.5. Individual level view of some AD and MCI subjects

To further analyze the brain regions focused by DANMLP, individual-level visualization was performed on AD and MCI subjects. In the classification task of AD vs. MCI, our goal is to visualize the brain regions with more severe lesions in AD subjects compared to MCI subjects. Therefore, we visualize the brain regions of AD subjects based on MCI subjects. In the classification task of MCI vs. NC, our goal is to visualize the brain regions with more severe lesions in MCI subjects compared to NC subjects. Therefore, we visualize the brain regions of MCI subjects based on NC subjects.

Taking the visualization process of AD subjects as an example, the calculation process is as follows: (1) Calculate the average ROI index $\bar{\Omega}_j^{MCI}$ of the jth brain region for all MCI subjects in test set using the formula $\overline{\Omega}_j^{MCI} = \sum_{n=1}^{N^{MCI}} \Omega_{n,j}^{MCI} / N^{MCI}$, where $\Omega_{n,j}^{MCI}$ denotes the ROI index of the jth brain region for the nth MCI subject, and N^{MCI} represents the total number of MCI subjects in the test set; (2) Calculate the absolute value $\rho_{n,j}^{AD} = \left| \Omega_{n,j}^{AD} - \overline{\Omega}_j^{MCI} \right|$ of the difference between $\Omega_{n,j}^{AD}$ and $\bar{\Omega}_j^{MCI}$,where $\Omega_{n,j}^{AD}$ is the ROI index of the jth brain region in the nth AD subject, and we take $\rho_{n,j}^{AD}$ as the visualization value of the jth brain region for the nth AD subject compared with the MCI subjects; (3) To eliminate the influence of outliers, the color bar range is set from 0 to the threshold value γ^{AD}. The threshold γ^{AD} can be calculated using the formula $\gamma^{AD} = \overline{\rho}^{AD} + k\sigma^{AD}$, where $\overline{\rho}^{AD}$ represents the mean visualization value of all brain regions for all AD subjects in the test set, σ^{AD} represents their standard deviation, k is an adjustment coefficient (here we set $k = 3$); (4) Display all the visualization values $\rho_{n,j}^{AD}$ of the nth AD subject using BrainNet [48]. Fig. 2a shows the visualization results of five AD subjects.

To visualize the brain region of the MCI subject in the classification task of MCI vs. NC, we can calculate the visualization value $\rho_{n,j}^{MCI}$ for the jth brain region of the nth MCI subject and threshold γ^{MCI} by using the same calculation steps as described above for AD subjects. Fig. 2b shows the visualization results of five MCI subjects.

From Fig. 2a, we can observe that compared to MCI subjects, the brain regions with lesions in AD subjects are mainly concentrated in the temporal lobe regions (including the hippocampus). When comparing the true positive and false negative subjects in Fig. 2a, all brain regions of false negative subjects are more similar to MCI subjects, especially with lower visualization values in the temporal lobe regions. Similar results are also found in MCI vs. NC classification tasks (Fig. 2b). In addition, by comparing Fig. 2a and Fig. 2b, we can also observe that

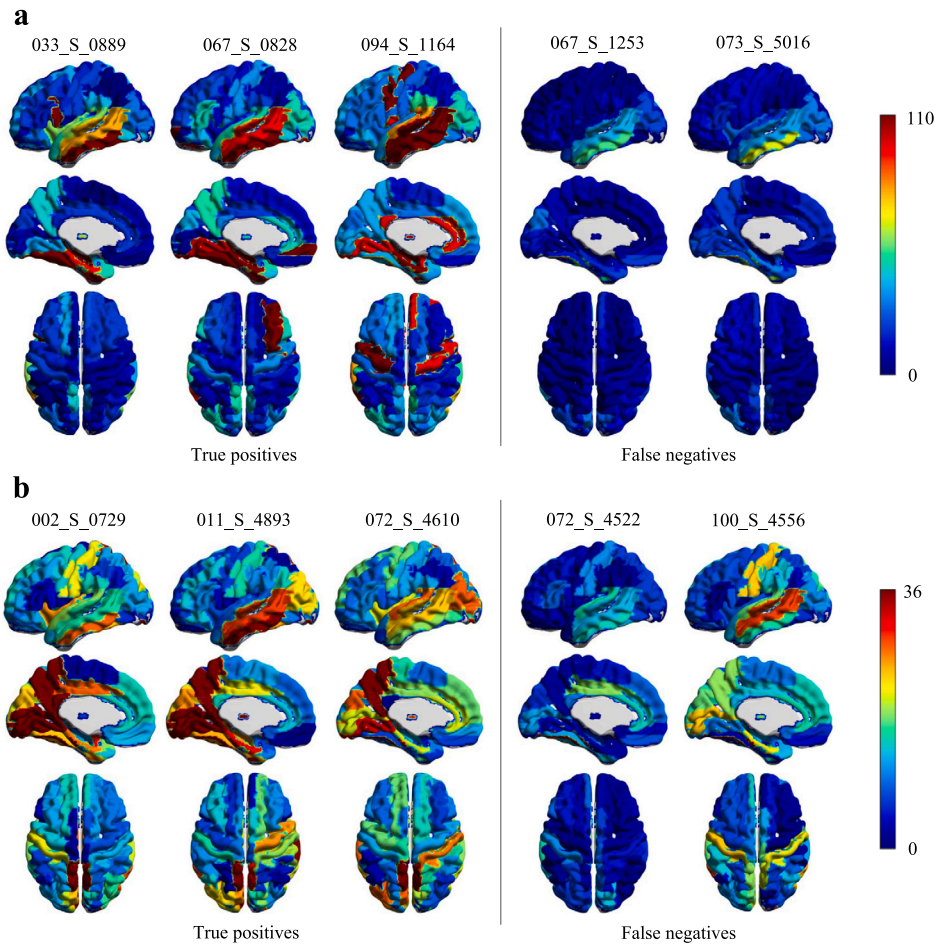

Fig. 2. Visualization results of discriminative brain regions identified by DANMLP for some AD and MCI subjects. (a) Visualization results of five AD subjects in the classification task of AD vs. MCI. The left panel (True positives) displays three correctly classified AD subjects, and the right panel (False negatives) shows two AD subjects misclassified as MCI subjects. (b) Visualization results of five MCI subjects in the classification task of MCI vs. NC. The left panel (True positives) displays three correctly classified MCI subjects, and the right panel (False negatives) shows two MCI subjects misclassified as NC subjects. The color bar represents the size of the visualization values $\rho_{n,j}^{AD}/\rho_{n,j}^{MCI}$ for each AD/MCI subject, which can reflect the degree of pathological changes in certain regions of the brain. The higher the visualization value $\rho_{n,j}^{AD}$ (or $\rho_{n,j}^{MCI}$) of a brain region, the greater the difference of this brain region between AD and MCI (or MCI and NC).

compared to the classification task of AD vs. MCI, the MCI vs. NC task is more difficult, resulting in smaller and more complex differences between their features. Overall, the discriminative brain region visualization results of DANMLP can display the variability of AD-related brain region features among subjects, which can help physicians accurately diagnose AD and MCI patients.

5. Discussion

In this section, we first present the experimental results of comparing our 3D-DACNN with other feature extraction methods, and the experimental results of comparing the position and channel self-attention mechanisms used in DANMLP with other attention mechanisms. We then give the experimental results of comparing the gray matter images used in DANMLP with other types of images. Finally, we analyze the limitations of DANMLP.

5.1. Effect of different feature extraction methods on DANMLP

To investigate the performance of DANMLP with different feature extraction methods, we compared 3D-DACNN used in our DANMLP with other four most popular feature extraction methods of wavelet transform [49], dictionary learning [50], deep neural network (DNN) [51], and recurrent neural network (RNN) [52] in 5CV test, by replacing the 3D-DACNN part of DANMLP with each of other four feature extraction methods. The results of DANMLP with different feature extraction methods are shown in Fig. 3 and Table S8.

From Fig. 3 and Table S8, we can see that the results of 3D-DACNN used in our DANMLP outperform the other four feature extraction methods in both binary classification tasks. For the MCI vs. NC classification task, the ACC of 3D-DACNN is 0.156, 0.130, 0.080, and 0.048 higher than that of the wavelet transform, dictionary learning, DNN, and RNN, respectively. For the AD vs. MCI classification task, the ACC of 3D-DACNN is 0.070, 0.068, 0.034, and 0.042 higher than that of the wavelet transform, dictionary learning, DNN, and RNN, respectively. These results indicate that the 3D CNN used in DANMLP is powerful in processing high-dimensional neuroimaging data such as sMRI. The reason may be that 3D CNN can effectively capture the three-dimensional spatial information contained in sMRI, and better detect the spatial structure and morphological features in neuroimaging, thus improving the classification performance. Compared to 3D CNN, wavelet transform, and dictionary learning may have limitations in extracting high-level features, while DNN and RNN may suffer from overfitting problems.

5.2. Effect of different attention mechanisms on DANMLP

To further discuss the performance of our position self-attention block and channel self-attention block, we compared them with the

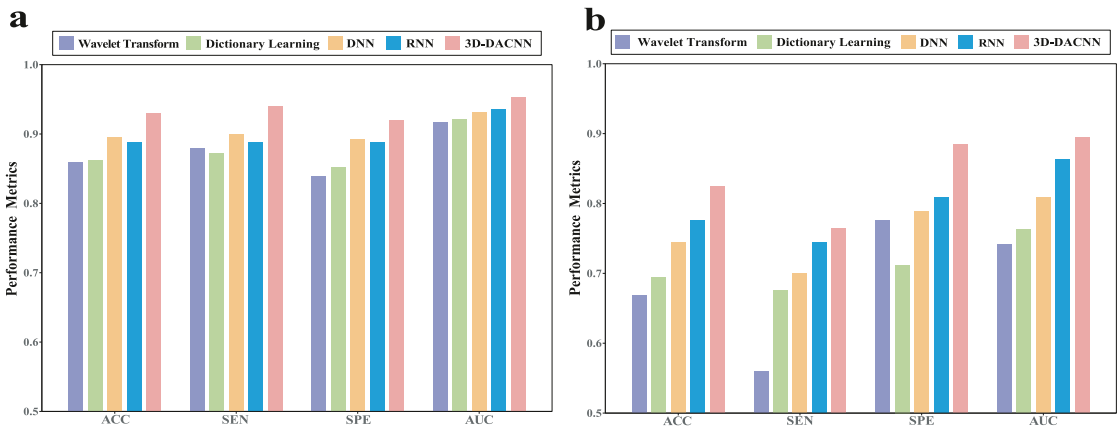

Fig. 3. Results of DANMLP with different feature extraction methods. (a) Results of AD vs. MCI. (b) Results of MCI vs. NC.

Table 3
Performance of DANMLP with different attention mechanisms in 5CV test.

Method	AD vs. MCI				MCI vs. NC			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
DANMLP-CBAM	0.910	0.904	0.916	0.950	0.786	0.692	0.880	0.886
DANMLP-scSE	0.900	0.904	0.896	0.950	0.782	0.700	0.864	0.886
DANMLP-simAM	0.898	0.936	0.860	0.948	0.742	0.696	0.788	0.850
DANNLP-CA	0.882	0.916	0.848	0.941	0.688	0.720	0.656	0.763
DANMLP	**0.930**	**0.940**	**0.920**	**0.953**	**0.824**	**0.764**	**0.884**	**0.895**

Table 4
Results of DANMLP using different images as inputs in 5CV test.

image	AD vs. MCI				MCI vs. NC			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
WM	0.864	0.900	0.828	0.934	0.776	0.648	0.904	0.876
CSF	0.874	0.884	0.864	0.933	0.778	0.640	0.916	0.858
GM+WM+CSF	0.882	0.900	0.864	0.937	0.770	0.692	0.848	0.873
GM	**0.930**	**0.940**	**0.920**	**0.953**	**0.824**	**0.764**	**0.884**	**0.895**

other four attention mechanisms. Considering that both the convolutional block attention (CBAM) [53] and the spatial and channel squeeze & excitation (scSE) [54] contain the position and channel attention blocks that can be used separately, we replaced the position self-attention block and channel self-attention block in DANMLP with the corresponding position attention block and channel attention block in CBAM and scSE, respectively. For the simple parameter-free attention (simAM) [55], and the coordinate attention (CA) [56], we adopted the attention structures of simAM and CA in the position and channel self-attention block of DANMLP. The experimental results are shown in Table 3.

From Table 3, we can see that the position and channel self-attention mechanism used in our DANMLP is powerful in improving the classification performance. For MCI vs. NC classification task, the ACC of DANMLP is 0.038, 0.042, 0.082, and 0.136 higher than that of DANMLP-CBAM, DANMLP-scSE, DANMLP-simAM, and DANMLP-CA, respectively. For AD vs. MCI classification task, the ACC of DANMLP is 0.020, 0.030, 0.032, and 0.048 higher than that of DANMLP-CBAM, DANMLP-scSE, DANMLP-simAM, and DANMLP-CA, respectively. These results show that our position and channel self-attention blocks can effectively improve the classification performance, the position attention block can further extract the important features within patches, and the channel attention block can effectively integrate the features across patches.

5.3. Effect of different images on DANMLP

Generally speaking, without prior knowledge, the features learned automatically by deep learning are superior to handcrafted features. However, it has been demonstrated that gray matter is associated with memory impairment and cognitive decline, which is an important biological marker for AD [57,58]. Many studies have used gray matter for AD-related research [59,60]. To investigate the effectiveness of using gray matter (GM) images for early diagnosis of AD, we conducted comparative experiments using white matter (WM) and cerebrospinal fluid (CSF) images, as well as unsegmented sMRI (i.e., GM+WM+CSF). These images are used as the inputs of DANMLP. The results of DANMLP using different images as inputs are shown in Table 4.

As shown in Table 4, we can see that GM images achieve better performance in both AD-related classification tasks. For the classification task of MCI vs. NC, the ACC of GM images is 0.048, 0.046, 0.054 higher that of WM images, CSF images, and unsegmented images (i.e., GM+WM+CSF), respectively. For AD vs. MCI task, the ACC of GM images is 0.066, 0.056, 0.048 higher that of WM images, CSF images, and unsegmented images, respectively. In fact, GM atrophy has been proven to be a primary marker of neurodegeneration, and it can serve as a biomarker for early diagnosis of AD [61,62]. While WM and CSF also have some discriminatory power for AD, their sensitivity is far less than that of GM [63,64]. We note that the performance of unsegmented images is significantly worse than GM images across all metrics, maybe that WM and CSF introduce the noises in AD vs. MCI and MCI vs. NC tasks.

5.4. Limitations and future work

Although our proposed DANMLP method has achieved good performance in AD-related diagnosis and discrimination of pathological regions, there are still the following limitations: (1) DANMLP uses patch-CNN with eight branches to extract features for each patch. Although it achieves better results, the network is relatively large with many parameters. In the future, we will consider designing a lightweight network. (2) The size of the input patch is fixed. However, AD-induced brain atrophy may occur in regions of multiple different sizes. The fixed-size patch is not conducive to extracting the features of large-scale atrophy. It is more reasonable to use multi-scale patches as inputs. (3) DANMLP just studied the classification task of NC, MCI and AD, and used the data from the ADNI database. More sub-types of AD-related diseases and related data should be considered. In future work, we should study more subdivisions of diagnosis of AD-related disorders, such as subjective cognitive decline (SCD), early mild cognitive impairment (EMCI), and late mild cognitive impairment (LMCI). In addition, considering the important role of genes in the development of AD and the generation of numerous gene expression data, we should develop a new method to integrate the neuroimaging and gene expression data for exploring the correlation between various brain regions and genes, so as to uncover the pathogenesis of AD-related disorders.

6. Conclusions

In this work, we proposed a novel AD/MCI diagnosis method of DANMLP by joining dual attention CNN and MLP based on the sMRI, clinical and genetic data. Patch-CNNs in DANMLP are used to extract image features of local patches. The position self-attention block in DANMLP is used to effectively capture the discriminative features within patches, while the channel self-attention block is used to capture the discriminative features between patches. MLP feature extractor is used to learn the embedding features of clinical and genetic data. The experimental results demonstrate that our DANMLP is superior to other methods in the classification tasks of AD vs. MCI and MCI vs. NC. DANMLP can successfully capture AD-related brain regions (i.e., hippocampus, amygdala), and the individualized visualization of focal areas can effectively help clinicians in the early accurate diagnosis of AD.

We would like to confirm that none of this work has been previously published, or been under review in any other journal. All authors were fully involved in this research and in preparation of the manuscript. The authors declare no conflict of interest.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, and there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript.

Acknowledgments

This work was partly supported by National Natural Science Foundation of China (62173271, 61873202).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.artmed.2023.102678.

References

[1] Gaugler J, James B, Johnson T, Reimer J, Solis M, Weuve J, et al. Alzheimer's disease facts and figures. Alzheimers Dementia 2022;18(2022):700–89.
[2] Mckhann G. Report of the NINCDS-ADRDA work group under the auspices of department of health and human service task force on Alzheimer's disease. Neurology 1984;34:939–44.
[3] Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drug development pipeline: 2019. Alzheimer's & Dementia: translational research & clinical interventions, 5. 2019, p. 272–93.
[4] Colliot O, Chételat G, Chupin M, Desgranges B, Magnin B, Benali H, et al. Discrimination between Alzheimer's disease, mild cognitive impairment and normal aging by using automated segmentation of the hippocampus. Radiology 2008;248:194–201.
[5] Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008;131:681–9.
[6] McDonald C, McEvoy L, Gharapetian L, Fennema-Notestine C, Hagler D, Holl D, et al. Regional rates of neocortical atrophy from normal aging to early Alzheimer disease. Neurology 2009;73:457–65.
[7] Bernal J, Kushibar K, Asfaw DS, Valverde S, Oliver A, Martí R, et al. Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: A review. Artif Intell Med 2019;95:64–81.
[8] Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural networks: A review. J Med Syst 2018;42:1–13.
[9] Kang W, Lin L, Zhang B, Shen X, Wu S, Initiative AsDN. Multi-model and multi-slice ensemble learning architecture based on 2D convolutional neural networks for Alzheimer's disease diagnosis. Comput Biol Med 2021;136:104678.
[10] Yagis E, Atnafu SW, García Seco de Herrera A, Marzi C, Scheda R, Giannelli M, et al. Effect of data leakage in brain MRI classification using 2D convolutional neural networks. Sci Rep 2021;11:1–13.
[11] Li F, Liu M, Initiative AsDN. Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. Comput Med Imag Graph 2018;70:101–10.
[12] Liu M, Zhang J, Adeli E, Shen D. Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. IEEE Trans Biomed Eng 2018;66:1195–206.
[13] Cheng D, Liu M, Fu J, Wang Y. Classification of MR brain images by combination of multi-CNNs for AD diagnosis. In: Ninth international conference on digital image processing. SPIE; 2017, p. 875–9.
[14] Lian C, Liu M, Zhang J, Shen D. Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE Trans Pattern Anal Mach Intell 2018;42:880–93.
[15] Liu M, Li F, Yan H, Wang K, Ma Y, Shen L, et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. Neuroimage 2020;208:116459.
[16] Folego G, Weiler M, Casseb RF, Pires R, Rocha A. Alzheimer's disease detection through whole-brain 3D-CNN MRI. Front Bioeng Biotechnol 2020;8:534592.
[17] Jin D, Zhou B, Han Y, Ren J, Han T, Liu B, et al. Generalizable, reproducible, and neuroscientifically interpretable imaging biomarkers for Alzheimer's disease. Adv Sci 2020;7:2000675.
[18] Li QS, Sun Y, Wang T. Epigenome-wide association study of Alzheimer's disease replicates 22 differentially methylated positions and 30 differentially methylated regions. Clin Epigenetics 2020;12:1–14.
[19] Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small G, et al. Gene dose of apolipoprotein E type 4 Allele and the risk of Alzheimer's disease in late onset families. Science 1993;261:921–3.
[20] Jo T, Nho K, Bice P, Saykin AJ, Initiative AsDN. Deep learning-based identification of genetic variants: Application to Alzheimer's disease classification. Briefings Bioinf 2022;23:bbac022.
[21] Varatharajah Y, Ramanan VK, Iyer R, Vemuri P. Predicting short-term MCI-to-AD progression using imaging, CSF, genetic factors, cognitive resilience, and demographics. Sci Rep 2019;9:1–15.
[22] Pan X, Phan T-L, Adel M, Fossati C, Gaidon T, Wojak J, et al. Multi-view separable pyramid network for AD prediction at MCI stage by 18 F-FDG brain PET imaging. IEEE Trans Med Imaging 2020;40:81–92.
[23] Gan J, Peng Z, Zhu X, Hu R, Ma J, Wu G. Brain functional connectivity analysis based on multi-graph fusion. Med Image Anal 2021;71:102057.
[24] De A, Chowdhury AS. DTI based Alzheimer's disease classification with rank modulated fusion of CNNs and random forest. Exp Syst Appl 2021;169:114338.
[25] Gupta Y, Kim J-I, Kim BC, Kwon G-R. Classification and graphical analysis of Alzheimer's disease and its prodromal stage using multimodal features from structural, diffusion, and functional neuroimaging data and the APOE genotype. Front Aging Neurosci 2020;12:238.
[26] Huang Y, Xu J, Zhou Y, Tong T, Zhuang X, Initiative AsDN. Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. Front Neurosci 2019;13:509.
[27] Kang L, Jiang J, Huang J, Zhang T. Identifying early mild cognitive impairment by multi-modality MRI-based deep learning. Front Aging Neurosci 2020;12:206.
[28] Guan Q, Huang Y, Zhong Z, Zheng Z, Zheng L, Yang Y. Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification. 2018, arXiv preprint arXiv:180109927.
[29] Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: Learning to leverage salient regions in medical images. Med Image Anal 2019;53:197–207.
[30] Abramova V, Clèrigues A, Quiles A, Figueredo DG, Silva Y, Pedraza S, et al. Hemorrhagic stroke lesion segmentation using a 3D U-net with squeeze-and-excitation blocks. Comput Med Imaging Graph 2021;90:101908.
[31] Li M, Hsu W, Xie X, Cong J, Gao W. SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network. IEEE Trans Med Imaging 2020;39:2289–301.
[32] Shen L, Sun M, Li Q, Li B, Pan Z, Lei J. Multiscale temporal self-attention and dynamical graph convolution hybrid network for EEG-based stereogram recognition. IEEE Trans Neural Syst Rehabil Eng 2022;30:1191–202.
[33] Korolev S, Safiullin A, Belyaev M, Dodonova Y. Residual and plain convolutional neural networks for 3D brain MRI classification. In: 2017 IEEE 14th international symposium on biomedical imaging. IEEE; 2017, p. 835–8.
[34] Xing X, Liang G, Blanton H, Rafique MU, Wang C, Lin A-L, et al. Dynamic image for 3D MRI image Alzheimer's disease classification. In: European conference on computer vision. Springer; 2020, p. 355–64.
[35] Spasov S, Passamonti L, Duggento A, Lio P, Toschi N, Initiative AsDN. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. Neuroimage 2019;189:276–87.
[36] Yang Z, Yu H, Fu Q, Sun W, Jia W, Sun M, et al. NDNet: Narrow while deep network for real-time semantic segmentation. IEEE Trans Intell Transp Syst 2020;22:5508–19.
[37] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 2002;15:273–89.
[38] Mattsson N, Insel PS, Donohue M, Jögi J, Ossenkoppele R, Olsson T, et al. Predicting diagnosis and cognition with 18F-AV-1451 tau PET and structural MRI in Alzheimer's disease. Alzheimer's Dementia 2019;15:570–80.

[39] Dai Z, Yan C, Wang Z, Wang J, Xia M, Li K, et al. Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). Neuroimage 2012;59:2187–95.
[40] Li J, Pan P, Huang R, Shang H. A meta-analysis of voxel-based morphometry studies of white matter volume alterations in Alzheimer's disease. Neurosci Biobehav Rev 2012;36:757–63.
[41] Tyng CM, Amin HU, Saad MN, Malik AS. The influences of emotion on learning and memory. Front Psychol 2017;8:1454.
[42] Pini L, Pievani M, Bocchetta M, Altomare D, Bosco P, Cavedo E, et al. Brain atrophy in Alzheimer's disease and aging. Ageing Res Rev 2016;30:25–48.
[43] van Hoesen GW, Augustinack JC, Dierking J, Redman SJ, Thangavel R. The parahippocampal gyrus in Alzheimer's disease: Clinical and preclinical neuroanatomical correlates. Ann New York Acad Sci 2000;911:254–74.
[44] Hsu CL, Best JR, Voss MW, Handy TC, Beauchet O, Lim C, et al. Functional neural correlates of slower gait among older adults with mild cognitive impairment. J Gerontol: Series A 2019;74:513–8.
[45] Ikonomovic M, Klunk W, Abrahamson E, Wuu J, Mathis C, Scheff S, et al. Precuneus amyloid burden is associated with reduced cholinergic activity in Alzheimer disease. Neurology 2011;77:39–47.
[46] Qi D, Wang A, Chen Y, Chen K, Zhang S, Zhang J, et al. Default mode network connectivity and related white matter disruption in type 2 diabetes mellitus patients concurrent with amnestic mild cognitive impairment. Curr Alzheimer Res 2017;14:1238–46.
[47] Mattioli P, Pardini M, Famà F, Girtler N, Brugnolo A, Orso B, et al. Cuneus/precuneus as a central hub for brain functional connectivity of mild cognitive impairment in idiopathic REM sleep behavior patients. Eur J Nucl Med Mol Imaging 2021;48:2834–45.
[48] Xia M, Wang J, He Y. BrainNet viewer: A network visualization tool for human brain connectomics. PLoS One 2013;8:e68910.
[49] El-Dahshan ESA, Mohsen HM, Revett K, Salem ABM. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. Exp Syst Appl 2014;41:5526–45.
[50] Mairal J, Bach F, Ponce J, Sapiro G. Online dictionary learning for sparse coding. In: Proceedings of the 26th annual international conference on machine learning. 2009, p. 689–96.
[51] Bengio Y. Learning deep architectures for AI. Found Trends® Mach Learn 2009;2:1–127.
[52] Dupond S. A thorough review on the current advance of neural network structures. Annu Rev Control 2019;14:200–30.
[53] Woo S, Park J, Lee J-Y, Kweon IS. Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision. 2018, p. 3–19.
[54] Roy AG, Navab N, Wachinger C. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. In: Medical image computing and computer assisted intervention–MICCAI 2018: 21st international conference, Granada, Spain, September (2018) 16-20, proceedings, Part I. Springer; 2018, p. 421–9.
[55] Yang L, Zhang R-Y, Li L, Xie X. Simam: A simple, parameter-free attention module for convolutional neural networks. In: International conference on machine learning. PMLR; 2021, p. 11863–74.
[56] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021, p. 13713–22.
[57] Graham WV, Bonito-Oliva A, Sakmar TP. Update on Alzheimer's disease therapy and prevention strategies. In: Caskey CT, editor. Annual review of medicine, vol. 682017. p. 413–30.
[58] Frisoni GB, Pievani M, Testa C, Sabattoli F, Bresciani L, Bonetti M, et al. The topography of grey matter involvement in early and late onset Alzheimer's disease. Brain 2007;130:720–30.
[59] Crossley NA, Mechelli A, Scott J, Carletti F, Fox PT, McGuire P, et al. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. Brain 2014;137:2382–95.
[60] Ortiz A, Munilla J, Gorriz JM, Ramirez J. Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. Int J Neural Syst 2016;26.
[61] Krajcovicova L, Klobusiakova P, Rektorova I. Gray matter changes in Parkinson's and Alzheimer's disease and relation to cognition. Curr Neurol Neurosci Rep 2019;19.
[62] Gauthier S, Patterson C, Gordon M, Soucy JP, Schubert F, Leuzy A. Commentary on recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. A Canadian perspective. Alzheimer's Dementia 2011;7:330–2.
[63] Schuff N, Woerner N, Boreta L, Kornfield T, Shaw LM, Trojanowski JQ, et al. MRI of hippocampal volume loss in early Alzheimers disease in relation to apoe genotype and biomarkers. Brain 2009;132:1067–77.
[64] Chetelat G, Landeau B, Eustache F, Mezenge F, Viader F, de la Sayette V, et al. Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. Neuroimage 2005;27:934–46.