Computers in Biology and Medicine 148 (2022) 105901

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Alzheimer's disease diagnosis via multimodal feature fusion

Yue Tu, Shukuan Lin *, Jianzhong Qiao *, Yilin Zhuang, Peng Zhang

School of Computer Science and Engineering, Northeastern University, Shenyang, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Computer-aided diagnosis
Multimodal information data
Feature transformation
Feature fusion

ABSTRACT

Alzheimer's disease (AD) is the most common neurodegenerative disorder in the elderly. Early diagnosis of AD plays a vital role in slowing down the progress of AD because there is no effective drug to treat the disease. Some deep learning models have recently been presented for AD diagnosis and have more satisfactory performance than classic machine learning methods. Nevertheless, most of the existing computer-aided diagnostic models used neuroimaging features for diagnosis, ignoring patients' clinical and biological information. This makes the AD diagnosis inaccurate. In this study, we propose a novel multimodal feature transformation and fusion model for AD diagnosis. The feature transformation aims to avoid the difference in feature dimensions between different modal data and further mine the significant features for AD diagnosis. A geometric algebra-based feature extension method is proposed to obtain different levels of high-dimensional features from patients' clinical and personal biological data. Then, an influence degree-based feature filtration algorithm is proposed to filtrate those features that have no apparent guiding significance for AD diagnosis. Finally, an ANN (Artificial Neural Network)-based framework is designed to fuse transformed features with neuroimaging features extracted by CNN (Convolutional Neural Network) for AD diagnosis. The more in-depth feature mining of patients' clinical information and biological information can significantly improve the performance of computer-aided AD diagnosis. The experiments are obtained on the ADNI dataset. Our proposed model can converge faster and achieves 96.2% accuracy in AD diagnostic task and 87.4% accuracy in MCI (Mild Cognitive Impairment) diagnostic task. Compared with other methods, our proposed approach has an excellent performance in AD diagnosis and surpasses SOTA (state-of-the-art) methods. Therefore, our model can provide more reasonable suggestions for clinicians to diagnose and treat disease.

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder that often appears in the aged. AD has unique characteristics in etiology and pathology. Its clinical manifestations are the decline of daily life and cognitive ability, accompanied by mental symptoms and behavioral disorders [1]. Researchers around the world have been exploring the causes of the disease [2,3]. Still, there is no clear etiology recognized by all researchers, and there is no treatment method that can prevent or reverse it. Therefore, early detection and early diagnosis are crucial for slowing down the development of the disease and reducing the incidence rate and morbidity.

According to the pathological features of AD [4], it can be separated into three phases, i.e., NC (Normal Control), MCI (Mild Cognitive Impairment), and AD. NC status is typical without any symptoms. MCI is a cognitive impairment state between average random cognitive ability and dementia. It has a high probability to become AD eventually [5]. MCI is divided into sMCI (stable MCI) and pMCI (progress MCI). Medically, two types of MCI are usually distinguished according to whether the patient has progressed 36 months after the first symptom onset.

Clinically, doctors usually diagnose AD by comprehensively analyzing the features of patients' multiple modalities. These features usually come from patients' multimodal information, including neuroimage data, gene sequence data, profile data, and clinical mental state scale data. Magnetic resonance imaging (MRI), an imaging method that utilizes magnetic resonance sensations to acquire electromagnetic signals from a human body and reconstruct it, is widely used to diagnose AD. Fig. 1 shows a sectional view of 3D MRI among three directions. In addition, a patient's personal attributes (such as age, gender, and weight), gene sequence data, and clinical mental state scale data are also commonly used to assist in AD diagnosis. Besides, the clinical mental state scale is an essential basis for the diagnosis, evaluation, and rehabilitation of progressive neurological impairment. The clinical mental state scales have a variety of classes. Two commonly used scales

* Corresponding authors.
E-mail addresses: linshukuan@cse.neu.edu.cn (S. Lin), qiaojianzhong@mail.neu.edu.cn (J. Qiao).

https://doi.org/10.1016/j.compbiomed.2022.105901
Received 21 March 2022; Received in revised form 26 June 2022; Accepted 16 July 2022
Available online 20 July 2022

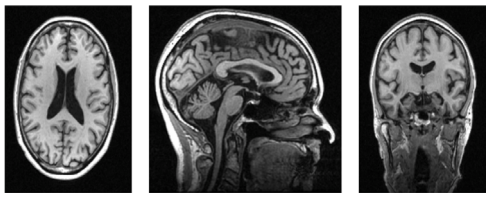

Fig. 1. The three views of a 3D MRI image. These views are axial view, sagittal view, and coronal view. Various views can depict different structural information of the brain.

are MMSE (Mini-mental State Examination) [6] and CDR (Clinical Dementia Rating) [7]. They are initially used to distinguish patients with neurological disorders (like AD) from patients with mental disorders.

With data mining and deep learning technology evolution, computer-aided diagnosis (CAD) is widely used in clinics. However, the existing studies usually used patients' MRI data to diagnose AD [8–11]. The ignorance of patients' multimodal features makes the computer-aided diagnostic model not wholly simulate the state of clinicians in AD diagnosis. Mainly, few existing models apply patients' profile data. Liu et al. [12] proposed a neural network diagnostic architecture which used patients' neuroimaging data and profile data to diagnose AD. Besides, Duc et al. [13] proposed a diagnostic model with joint MMSE with functional MRI for AD diagnosis. However, the profile data or MMSE data are directly placed in the fully connected layer as neurons without considering the different levels of high-dimensional features of the data. The neuroimaging features of patients extracted by neural networks are usually more complex and high-dimensional. The simple combination of high-dimensional neuroimaging features and low-dimensional profile features may cause dimension mismatch and cannot significantly improve accuracy. Although various models for diagnosing AD have been proposed, there are still some challenges for further enhancing the diagnostic accuracy. These challenges are summarized as follows.

Challenge1: Clinically, doctors often make a comprehensive diagnosis according to the patients' multimodal features. Using only the patients' neuroimaging features will make the diagnostic result inaccurate. How to design a model to fuse the multimodal features of patients is a challenge for computer-aided AD diagnosis.

Challenge2: Simply fusing the low-dimensional patients' clinical and biological information features and high dimensional neuroimaging features will give rise to the results being dominated by one or some of them. How to mine more noteworthy features from the low-dimensional features for AD diagnosis, and how to fuse different dimension features to give accurate diagnosis results is another challenge.

We propose a novel Alzheimer's disease diagnostic model based on multimodal feature transformation and fusion to tackle the aforementioned challenges. As far as we know, this is the first time that the low-dimensional features of patients' clinical and biological information are transformed to fuse with MRI data for diagnosing AD. The framework of our proposed multimodal feature fusion approach is shown in Fig. 2. In order to mine more momentous features for AD diagnosis from the low-dimensional patients' multimodal data, we propose a multimodal feature transformation method. This method contains two parts, feature dimension extension and feature filtration. Because the low-dimensional information data has linear indivisibility problems and lack of correlation between different features, we consider mapping low dimensional features to a high dimensional space through the extension of geometric algebra. Then, the proposed feature filtration algorithm can filtrate the features that have a necessary consequence on AD diagnosis. Other features are pruned. To make CADs better simulate the clinicians' process of diagnosing AD, we consider fusing these features with patients' neuroimaging features. To this end, a multimodal feature fusion diagnostic model is proposed to accurately diagnose AD and provide diagnostic advice to clinicians.

Our paper has the following contributions:

- A multimodal feature extension method based on geometric algebra is proposed to obtain higher and deeper features from multimodal information data. This method can extend original multimodal information data into different levels of high-dimensional features, which contains more information about AD diagnosis. It is proved theoretically that our proposed feature extension method is complete and correct for the expression of the original multimodal information data.
- A feature filtration algorithm based on influence degree is proposed to cut out those features extended by our proposed feature extension method. With the help of mathematical statistics, our proposed algorithm can filtrate features which have no apparent guiding significance for AD diagnosis and provide more valuable features for the diagnostic network.
- A neuroimaging feature extracting model based on a CNN (Convolutional Neural Network) is designed to extract features from MRI data. For the sake of better integrating neuroimaging features and multimodal information features, a multimodal feature fusion model based on an ANN (Artificial Neural Network) is designed to fuse multimodal features for diagnosing AD.
- Experimental results on ADNI indicate that our diagnostic model can achieve higher accuracy than other diagnostic methods. Besides, our proposed method can converge faster, which improves the efficiency of the diagnosis process.

2. Related work

Lately, deep learning has made remarkable achievements in computer vision, explicitly using CNN to obtain features. It supplies various ideas for diagnosing AD. Using CNN to acquire MR images' features may learn the features of different dimensions and reduce the complexity of manual feature extraction. Better extraction feature information from MRI can contribute to the accurate diagnosis of AD. Usually,

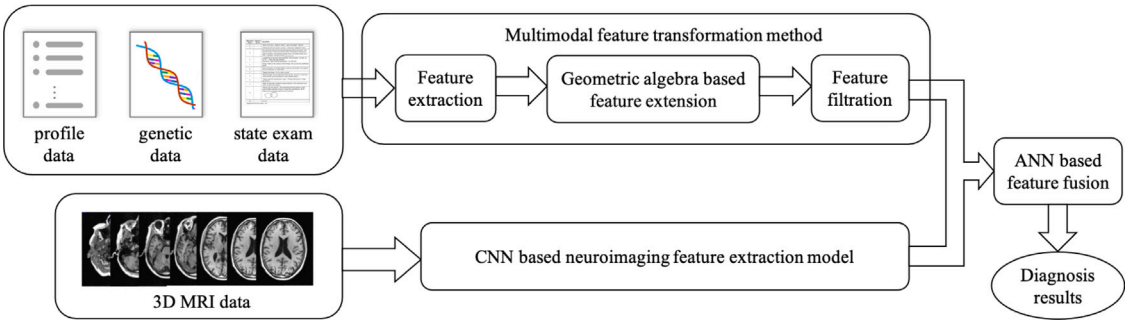

Fig. 2. The framework of our proposed multimodal feature fusion method. Our proposed architecture is divided into multimodal feature transformation, neuroimaging feature extraction, and feature fusion. The synergy of various modules jointly realizes the diagnosis of AD.

the existing feature extraction methods can be separated into four types (voxel-based methods [14–16], image patch-based [17–19], ROI (region of interest)-based [20–22], and whole image-based [23–25]).

These methods usually use 3D MR images [10,26,27] or 2D slice structures [28] to extract features and improve diagnosis effectiveness through different networks or image processing. Lian et al. [29] presented a weakly supervised attention network. By explicitly considering the individual dissimilarities of different subjects, multiple clinical scores and diagnostic categories can be predicted from MRI data, and location-sensitive areas can be learned without prior anatomical knowledge. Alinsaif et al. [8] presented a hybrid of 3D shearlet-based descriptors with in-depth features for diagnosing AD. Although these medical image diagnosis methods based on MRI have achieved good results to a certain extent, the single use of patients' MRI data for AD diagnosis can easily cause the loss of valuable features, thus affecting diagnosis results.

In order to comprehensively consider the neuroimaging diagnostic features of patients, researchers used various modal data for AD diagnosis. These various data include PET (Positron Emission Computed Tomography) [30,31] data, fMRI (functional MRI) [32,33] data, CSF (cerebrospinal fluid) [34] data, DTI (diffusion tensor imaging) [35] data, EEG (Electroencephalogram) [36] data, and etc. Zhou et al. [30] trained the feature extractors in different modes, extracted the features in different stages, and finally fused the extracted features to obtain the diagnostic results. Abrol et al. [32] proposed a parallel feature extraction model for MRI and fMRI, and realized AD diagnosis by constructing a support vector machine (SVM) classifier. Besides, Fiscon et al. [36] used Fourier and Wavelet analysis to extract features from EEG and used a supervised learning approach to diagnose AD. However, there is a problem of the high missing rate of some multimodal neuroimaging data, which makes the diagnosis accuracy often not good enough.

Although the existing computer-aided AD diagnosis methods have achieved good results, they mostly use a single way of feature extraction on neuroimages, ignoring patients' clinical and biological information. Researchers consider adding different feature data to imaging features for fusion diagnosis to better use patients' clinical and biological information data. Liu et al. [12] presented a neural network-based diagnostic model. It used patients' neuroimaging data and profile data to diagnose AD. The profile data is directly put in the fully connected layer as neurons. Again, Duc et al. [13] proposed a diagnostic model with joint MMSE with functional MRI for AD diagnosis. However, [13] only used MMSE data and carried out a pre-screening process, which affected the final diagnosis results and made the model lose reusability. Similarly, only using MMSE alone cannot significantly improve the accuracy of AD diagnosis. Moreover, the existing methods of fusion diagnosis based on multimodal information data usually extract shallow features and fuse them by simple splicing, so the diagnosis effect is poor. We are also based on such a research status.

3. Materials and methods

3.1. Datasets

The major imaging initiatives for Alzheimer's disease are ADNI [37] and AddNeuroMed [38]. ADNI is committed to collecting neuroimages, as well as AddNeuroMed is dedicated to dissecting biomarkers in AD patients. The dataset used in this paper is acquired from ADNI. The structural MRI dataset is used. The data information is listed in Table 1. 1461 subjects, including 767 men and 694 women, aged between 54 and 96 years old. The imaging protocol setting of these MR images is listed in Table 2. All MR images used in this paper are sagittal scanning images. Sagittal scanning can clearly show the shape and course of the sulcus and gyrus. It has more advantages for clinicians to diagnose brain-related diseases. For MR images, we unify the oversized images into a consistent size by cutting and scaling, which makes the data more convenient to train our network. The size of unified MRI data is $192 \times 192 \times 160$. In addition, APOE (Apolipoprotein E) is a Protein Coding gene, which is the most significant risk gene of AD [39]. It has three isoform alleles ($\epsilon 2$, $\epsilon 3$, and $\epsilon 4$). APOE $\epsilon 4$ is regarded as the primary genetic risk factor [40]. For APOE data, we convert it into an eigenvector form to facilitate the model's input.

Table 1
Data sample information.

CLASS	AD	sMCI	pMCI	NC
Number	218	448	328	467
Sex (M/F)	111/107	276/172	164/164	216/251
Age (years)	77.93 ± 6.85	76.61 ± 6.93	76.73 ± 7.41	77.89 ± 6.11
Weight (kg)	69.81 ± 11.47	74.57 ± 14.74	72.52 ± 14.81	72.79 ± 13.72
MMSE	21.28 ± 4.71	26.71 ± 2.71	23.41 ± 4.79	29.11 ± 1.15
CDR	0.94 ± 0.51	0.45 ± 0.13	0.78 ± 0.47	0.06 ± 0.17

Table 2
The imaging protocol setting of our acquired datasets.

Protocol parameter	Value
Acquisition plane	Sagittal
Acquisition type	3D
Field strength	1.5 tesla
Slice thickness	1.2 mm
Echo time	3.5–3.7 ms
Inversion time	1000.0 ms
Repetition time	3000.0 ms
Weighting	T1

3.2. Methodology

When clinicians diagnose Alzheimer's disease, they not only make a judgment according to patients' neuroimaging status but also make a comprehensive judgment according to patients' personal attributes, gene sequence data, and mental state examinations. The existing methods of computer-aided AD diagnosis usually construct different neural networks to extract features from MRI for diagnosis. However, using only a patient's neuroimaging features will make AD diagnosis inaccurate. Given this, we propose a multimodal feature fusion method for diagnosing AD, which contains three parts, i.e., multimodal feature transformation method, CNN neuroimaging feature extraction model, and ANN feature fusion diagnostic model. Our proposed multimodal feature transformation method transforms patients' clinical and biological information data, which aims to extract different levels of features that can guide AD diagnosis. The CNN neuroimaging feature extraction model extracted patients' MRI data. These features extracted from the two parts are jointly diagnosed AD by the ANN feature fusion module to improve the diagnostic accuracy. Fig. 3 shows the architecture of our proposed multimodal feature fusion diagnostic model.

Firstly, the design of a multimodal feature transformation module (Section 3.2.1) considers the low dimension of patients' multimodal information data, and the information contained is limited. To mine the potential deeper high-dimensional features, we consider using geometric algebraic feature transformation to extend the features of multimodal information data (Section 3.2.1.1). The feature extension method will generally introduce noise data into new feature groups. Therefore, a feature filtration algorithm (Section 3.2.1.2) is proposed to prune the features that are not significant for AD diagnosis. After filtrating, we can get the feature groups that significantly impact the diagnosis of AD. Secondly, researchers widely use CNN in image feature extraction because of its excellent performance in computer vision. We design a neuroimaging feature extraction model (Section 3.2.2) based on CNN to extract features from patients' MR images. Finally, a feature fusion diagnosis model (Section 3.2.3) based on ANN is designed to fuse the multimodal information features and neuroimaging features to obtain the diagnosis results. ANN has excellent representation ability and broad universality, so it can fuse various features and get diagnostic results. More details are introduced below.

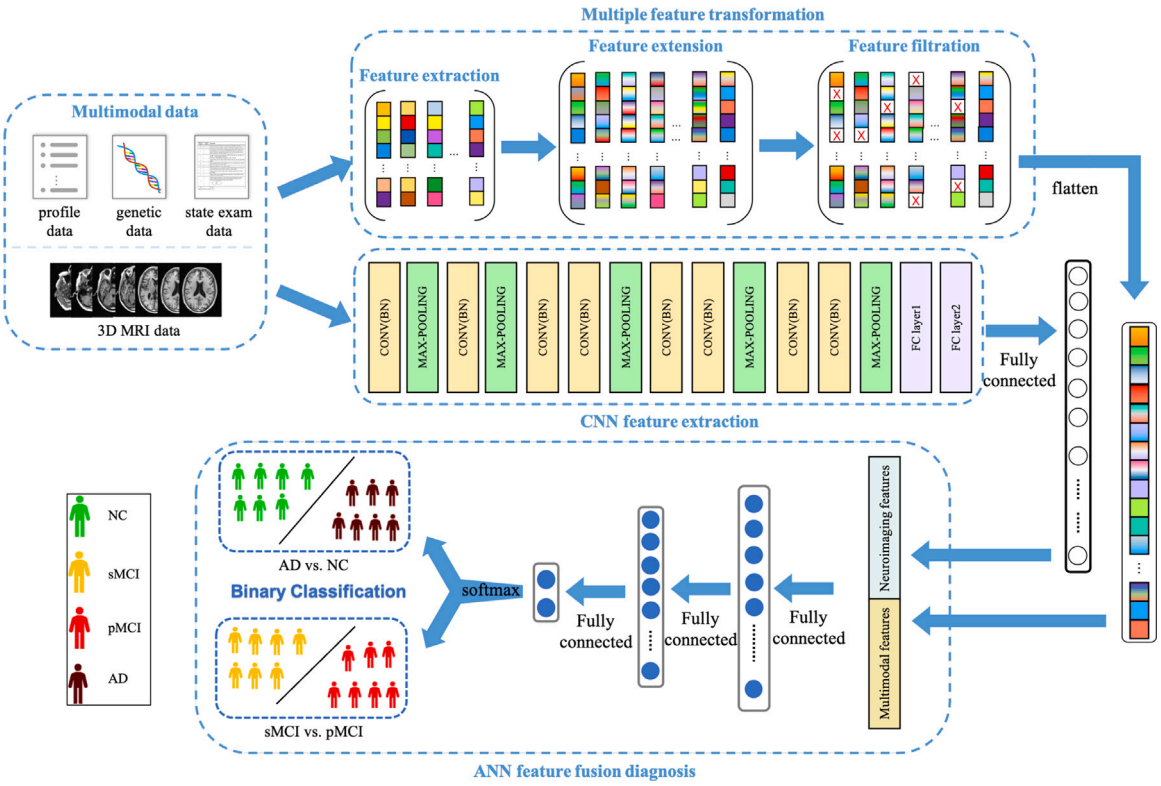

Fig. 3. The whole architecture of our proposed multimodal feature fusion diagnostic model, including a multiple feature transformation module, a CNN feature extraction module and an ANN feature fusion diagnosis module.

3.2.1. *Multimodal feature transformation method*

Clinically, when diagnosing AD, it is usually necessary to obtain patients' multimodal features in addition to neuroimaging features for diagnosis. These multimodal features include patients' personal attributes (such as gender, age, and weight), gene sequence data, and clinical mental state examinations data for patients' mental and neurological conditions. We propose a feature dimension extension method and a feature filtration algorithm. Patients' clinical information data and biological information data are low-dimensional. The low-dimensional data has the problems of linear indivisibility and lack of correlation between different features. Therefore, we consider mapping low dimensional features to a high dimensional space through the extension of geometric algebra. Our proposed feature dimension extension method can extend the feature vectors from Euclidean space to geometric algebraic space. After the dimension extension, the first-order features of patients' clinical and biological information data are expanded to multi-level high-order features in the geometric algebraic space. The extension uses the correlation details and high-level structure between different variables in patients' clinical and biological information data. At the same time, the transformed high-order features have better interpretability. In the process of obtaining multi-level features after feature dimension extension, a few features that are irrelevant to AD diagnosis (i.e., noise data) will inevitably be introduced. Hence, we present a feature filtration algorithm based on statistical hypothesis testing to prune noise features that have no significant impact on AD diagnosis. These two parts are introduced below.

3.2.1.1. Feature dimension extension method. AD patients' clinical and biological information data are multimodal and have diverse dimensions. We propose a universal method to extend the feature's dimension for different kinds of data. A feature vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ in $\Re^n$ space is used to represent a patient's feature. The set of vectors $\boldsymbol{E} = (\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n})$ where $\boldsymbol{e_i} = (0, 0, \ldots, 1, \ldots, 0)$ forms an orthonormal basis of $\Re^n$ space, where $\boldsymbol{e_i} \cdot \boldsymbol{e_j} = \delta_{ij}$ and

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{1}$$

For any two vectors $\boldsymbol{a}, \boldsymbol{b} \in \Re^n$, the geometric product is $\boldsymbol{ab} = \boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{a} \wedge \boldsymbol{b}$, where $\boldsymbol{a} \cdot \boldsymbol{b}$ is the dot product of $\boldsymbol{a}$ and $\boldsymbol{b}$, and $\boldsymbol{a} \wedge \boldsymbol{b}$ is the outer product of $\boldsymbol{a}$ and $\boldsymbol{b}$. In order to better represent the extension process from $\Re^n$ Euclidean space to $\Re^{2^n}$ geometric algebraic space, the following definitions are first given.

Definition 1 (*m-Order Derived Vector*)**.** For the set of vectors $\boldsymbol{E} = (\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n})$, m-order derived vector $\boldsymbol{e}(D)$ is defined as the geometric product of m vectors in $\boldsymbol{E}$, where the numbers of vectors are listed in D. That is

$$\boldsymbol{e}(D) = \prod_{i=1}^{m} \boldsymbol{e}_i \tag{2}$$

where $m = |D|$ is the cardinal number of set D. For example, $D = 2, 6, 3$, then we have $\boldsymbol{e}(\boldsymbol{D}) = \boldsymbol{e_2 e_6 e_3}$.

Definition 2 (*Metric Trace*)**.** For the set of vectors $\boldsymbol{E} = (\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n})$ and a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ in $\Re^n$, the trace of the m-order derived vector $\boldsymbol{e}(D)$ under vector $\boldsymbol{x}$ is defined as:

$$mtr(D, \boldsymbol{x}) = \prod_{i=1}^{m} x_i \tag{3}$$

where $m = |D|$ is the cardinal number of set D. For example, $D = 2, 6, 3$, then we have $mtr(D, \boldsymbol{x}) = x_2 x_6 x_3$ So, we can get that metric trace is a scalar and $mtr(D, \boldsymbol{x}) \in \Re$.

Definition 3 (*Power Ordered Set*)**.** Given a set $S = \{1, 2, \ldots, n\} \subseteq \mathbb{N}$, the potential of S is n. We use $P(S)$ to denote the power set of S, and the potential of $P(S)$ is 2^n. Power ordered set $PO(S)$ is an ordered set that satisfies the following four conditions:

(1) $PO(S)$ is a power set of S, where the potential is 2^n.
(2) All elements in $PO(S)$ are subsets of S and are arranged in the order of increasing potential.
(3) Equipotential elements are arranged in the order of the elements in the S set.
(4) For every element in $PO(S)$, the order of signals is arranged in ascending order.

Then we define $PO(S)_i^m$ to represent an element i with potential m in the power order set $PO(S)$. That is,

$$PO(S) = \left\{ PO(S)_1^0, PO(S)_2^1, \ldots, PO(S)_{n+1}^1, \ldots, PO(S)_i^m, \ldots, PO(S)_{2^n}^n \right\} \tag{4}$$

Definition 4 (*Geometric Product Basis*)**.** For the set of vectors $\boldsymbol{E} = \left(\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n}\right)$ and a set $S = \{1, 2, \ldots, n\} \subseteq \mathbb{N}$, geometric product basis $\boldsymbol{EP}$ is defined as a group of geometric product basic vectors. That is

$$\boldsymbol{EP} = \{\boldsymbol{e}(D) \,|\, D \in PO(S)\} \tag{5}$$

The order of elements in $\boldsymbol{EP}$ is the same as that in $PO(S)$, and there has $EP_i^m = \boldsymbol{e}\left(PO(S)_i^m\right)$. Geometric product basis transforms the normal vector basis in $\Re^n$ space into the vector basis in $\Re^{2^n}$ geometric algebraic space, which is helpful to feature transformation.

To adjust the influence level of different features on the AD diagnosis, a harmonic factor h_i called influence degree is introduced into the transformation process.

Definition 5 (*Influence Degree*)**.** Given a basis vector $\boldsymbol{e}(D)$, influence degree h_i is defined to indicate the probability that the hypothesis "$\boldsymbol{e}(D)$ has a significant impact on AD diagnosis" is true.

The details about influence degree h_i will be introduced in Section 3.2.1.2.

Established on the above definition, we present the feature extension method. Given a feature vector $\boldsymbol{x} = \left(x_1, x_2, \ldots, x_n\right)$ and an orthonormal basis $\boldsymbol{E} = \left(\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n}\right)$ in $\Re^n$ space, $S = \{1, 2, \ldots, n\} \subseteq \mathbb{N}$, we transform the feature vector to $\Re^{2^n}$ space by using Eq. (6):

$$\boldsymbol{ft}(\boldsymbol{x}) = \prod_{i=1}^{2^n} h_i \cdot mtr\left(PO(S)_i^m, \boldsymbol{x}\right) \cdot EP_i^m \tag{6}$$

EP_i^m is the element of $\boldsymbol{EP}$. By using Eq. (6), a feature vector $\boldsymbol{x}$ from $\Re^n$ Euclidean space can be extended to $\boldsymbol{ft}(\boldsymbol{x})$ in $\Re^{2^n}$ geometric algebraic space. The dimension of the original feature vector is n, and the dimension of the extended feature vector is 2^n, which is equal to the potential of $PO(S)$.

Above, we give the process of transforming vector $\boldsymbol{x}$ in $\Re^n$ space into vector $\boldsymbol{ft}(\boldsymbol{x})$ in $\Re^{2^n}$ space. Actually, there are usually multiple measure variables in $\Re^n$ space. Assuming that there are λ measure variables in $\Re^n$ space, we use matrix $\boldsymbol{X}$ to represent multiple vectors in $\Re^n$:

$$\boldsymbol{X} = \begin{bmatrix} x(1)_1 & x(2)_1 & \cdots & x(\lambda)_1 \\ x(1)_2 & x(2)_2 & \cdots & x(\lambda)_2 \\ \vdots & \vdots & \ddots & \vdots \\ x(1)_n & x(2)_n & \cdots & x(\lambda)_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}(1) \\ \boldsymbol{x}(2) \\ \vdots \\ \boldsymbol{x}(\lambda) \end{bmatrix}^T \tag{7}$$

where $\boldsymbol{x}(\cdot)$ is a single feature vector of $\Re^n$.

For the feature matrix $\boldsymbol{X}$, we have the following equation to transform it into $\Re^n$ space:

$$\boldsymbol{FT}(\boldsymbol{X}) = \boldsymbol{H} \circ \boldsymbol{MTR}(PO(S), \boldsymbol{X}) \circ \boldsymbol{EP} \tag{8}$$

where $\circ$ is the Hadamard product of the matrices. $\boldsymbol{H}$ is the influence degree matrix and $\boldsymbol{MTR}(PO(S), \boldsymbol{X})$ is the metric trace matrix. Specially, we have the following equation to calculate $\boldsymbol{FT}(\boldsymbol{X})$ (Eq. (9) is given in Box I). To theoretically prove the correctness and effectiveness of our proposed feature extension approach, we take the case of single vector as an example to give the prof process and some more theorems.

Lemma 1. *All the vectors in* $\boldsymbol{EP}$ *are linearly independent.*

Proof. Suppose $\boldsymbol{V}$ is a vector in $\Re^{2^n}$ space, $\boldsymbol{V}$ can be represented by vectors in $\boldsymbol{EP}$ as the following equation:

$$\boldsymbol{V} = \alpha_1^0 + \sum_{a=1}^{C_n^1} \alpha_a^1 \cdot EP_a^1 + \sum_{b=1}^{C_n^2} \alpha_b^2 \cdot EP_b^2 + \cdots + \sum_{i=1}^{C_n^m} \alpha_i^m \cdot EP_{j+\sum_{j=1}^{m-1} C_n^j}^m + \cdots + \alpha_1^n \cdot EP_{2^n}^n \tag{10}$$

where α_u^v represents the coefficient of the nth vector with potential v. Especially, when $v = 0$, the vector with potential 0 is a scalar.

Considering to prove that the vectors in $\boldsymbol{EP}$ are linearly independent, it is necessary to prove that $\boldsymbol{V} = \boldsymbol{0}$ if and only if all α_u^v are equal to 0. According to the principle of geometric algebra, only vectors with the same potential can be calculated. So, if we have the equation $\boldsymbol{V} = \boldsymbol{0}$, it is equivalent to the following equation set:

$$\begin{cases} \alpha_1^0 = 0 \\ \sum_{a=1}^{C_n^1} \alpha_a^1 \cdot EP_a^1 = 0 \\ \sum_{b=1}^{C_n^2} \alpha_b^2 \cdot EP_b^2 = 0 \\ \cdots \ \cdots \\ \sum_{i=1}^{C_n^m} \alpha_i^m \cdot EP_{j+\sum_{j=1}^{m-1} C_n^j}^m = 0 \\ \cdots \ \cdots \\ \alpha_1^n \cdot EP_{2^n}^n = 0 \end{cases} \tag{11}$$

It is evident that $alpha_1^0 = \alpha_1^n = 0$.

Next, we analyze the second equation in the equation set, i.e.,

$$\sum_{a=1}^{C_n^1} \alpha_a^1 \cdot EP_a^1 = \alpha_1^1 \cdot EP_1^1 + \alpha_2^1 \cdot EP_2^1 + \cdots + \alpha_n^1 \cdot EP_n^1 = 0 \tag{12}$$

Multiply both sides of the equation by EP_1^1, then we have

$$\alpha_1^1 \cdot \left| EP_1^1 \right|^2 + \left(\alpha_2^1 \cdot EP_2^1 + \cdots + \alpha_n^1 \cdot EP_n^1\right) \cdot EP_1^1 = 0 \tag{13}$$

Since $\left| EP_1^1 \right|^2$ is a scalar, we can get $\alpha_1^1 = 0$ and

$$\alpha_2^1 \cdot EP_2^1 + \cdots + \alpha_n^1 \cdot EP_n^1 = 0 \tag{14}$$

For Eq. (14), multiply both sides of the equation by EP_2^1, obtain that $\alpha_2^1 = 0$. By analogy, we can get

$$\alpha_1^1 = \alpha_2^1 = \cdots = \alpha_n^1 = 0 \tag{15}$$

For $\sum_{b=1}^{C_n^2} \alpha_b^2 \cdot EP_b^2 = \alpha_1^2 \cdot \boldsymbol{e_1 e_2} + \alpha_2^2 \cdot \boldsymbol{e_1 e_3} + \cdots + \alpha_{C_n^2}^2 \cdot \boldsymbol{e_{n-1} e_n} = 0$, multiply both sides of the equation by $\boldsymbol{e_2 e_1}$, then we can get $\alpha_1^2 = 0$. Repeat similar operations to get $\alpha_1^2 = \alpha_2^2 = \cdots = \alpha_n^2 = 0$.

The other equations in Eq. (11) can also be calculated according to the above principle, and all the $\alpha_u^v = 0$. Therefore, if and only if the coefficients in the equation of vector $\boldsymbol{V}$ represented by all vectors in $\boldsymbol{EP}$ are 0, $\boldsymbol{V} = \boldsymbol{0}$.

In summary, all the vectors in $\boldsymbol{EP}$ are linearly independent. □

Through Lemma 1, we can get that the extended feature vector group is linearly independent after the feature extension. We intend to ensure that the extended eigenvector group can cover the geometric algebraic space. That is, $\boldsymbol{EP}$ can represent any vector in the geometric algebraic space. In addition, if we cannot guarantee the uniqueness of the extension, the results will directly affect the accuracy of AD diagnosis, and our method will become inconsistent. For the sake of proving the correctness and uniqueness of our proposed approach, we provide Theorem 1.

Theorem 1. *After the feature extension, any vector in the* $\Re^{2^n}$ *space can be represented by Eq. (6), and the expression is unique.*

$$\boldsymbol{FT}(\boldsymbol{X}) = \begin{bmatrix} h(1)_1 \cdot mtr\left(PO(S)_1^0, \boldsymbol{x}(1)\right) \cdot EP(1)_1^0 & \cdots & h(\lambda)_1 \cdot mtr\left(PO(S)_1^0, \boldsymbol{x}(\lambda)\right) \cdot EP(\lambda)_1^0 \\ h(1)_2 \cdot mtr\left(PO(S)_2^1, \boldsymbol{x}(1)\right) \cdot EP(1)_2^1 & \cdots & h(\lambda)_2 \cdot mtr\left(PO(S)_2^1, \boldsymbol{x}(\lambda)\right) \cdot EP(\lambda)_2^1 \\ \vdots & \ddots & \vdots \\ h(1)_{2^n} \cdot mtr\left(PO(S)_{2^n}^n, \boldsymbol{x}(1)\right) \cdot EP(1)_{2^n}^n & \cdots & h(\lambda)_{2^n} \cdot mtr\left(PO(S)_{2^n}^n, \boldsymbol{x}(\lambda)\right) \cdot EP(\lambda)_{2^n}^n \end{bmatrix} \tag{9}$$

Box I.

Proof. Given a feature vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ in $\Re^n$, suppose that the expression obtained after transforming the vector $\boldsymbol{x}$ into $\Re^{2^n}$ space is not unique. So, there are at least two representation equations, both of which can represent the vector obtained after the vector x is transformed. The two equations are as follows.

$$\boldsymbol{V_1} = \alpha_1^0 + \sum_{a=1}^{C_n^1} \alpha_a^1 \cdot EP_a^1 + \sum_{b=1}^{C_n^2} \alpha_b^2 \cdot EP_b^2 + \cdots + \sum_{i=1}^{C_n^m} \alpha_i^m \cdot EP^m_{j+\sum_{j=1}^{m-1} C_n^j} + \cdots + \alpha_1^n \cdot EP_{2^n}^n \tag{16}$$

$$\boldsymbol{V_2} = \beta_1^0 + \sum_{a=1}^{C_n^1} \beta_a^1 \cdot EP_a^1 + \sum_{b=1}^{C_n^2} \beta_b^2 \cdot EP_b^2 + \cdots + \sum_{i=1}^{C_n^m} \beta_i^m \cdot EP^m_{j+\sum_{j=1}^{m-1} C_n^j} + \cdots + \beta_1^n \cdot EP_{2^n}^n \tag{17}$$

The feature vector after the extension has the same meaning, so $\boldsymbol{V_1} = \boldsymbol{V_2}$.

Let Eq. (16)–Eq. (17):

$$\boldsymbol{0} = \left(\alpha_1^0 - \beta_1^0\right) + \sum_{a=1}^{C_n^1} \left(\alpha_a^1 - \beta_a^1\right) \cdot EP_a^1 + \sum_{b=1}^{C_n^2} \left(\alpha_b^2 - \beta_b^2\right) \cdot EP_b^2 + \cdots + \sum_{i=1}^{C_n^m} \left(\alpha_i^m - \beta_i^m\right) \cdot EP^m_{j+\sum_{j=1}^{m-1} C_n^j} + \cdots + \left(\alpha_1^n - \beta_1^n\right) \cdot EP_{2^n}^n \tag{18}$$

According to Lemma 1, all the vectors in $\boldsymbol{EP}$ are linearly independent. So, if and only if the coefficients in the equation of vector $\boldsymbol{V}$ represented by all vectors in $\boldsymbol{EP}$ are 0, $\boldsymbol{V} = \boldsymbol{0}$. In Eq. (18), all the coefficients equal to 0, i.e.

$$\alpha_1^0 = \beta_1^0, \alpha_a^1 = \beta_a^1, \ldots, \alpha_i^m = \beta_i^m, \ldots, \alpha_1^n = \beta_1^n \tag{19}$$

$\boldsymbol{V_2}$ and $\boldsymbol{V_2}$ are the same vectors obtained by the feature extension method. The hypothesis is not valid, so the original proposition is true.

So, it can be sure that any vector in the $\Re^{2^n}$ space can be represented by Eq. (6), and the expression is unique. □

According to Lemma 1, the vectors in $\boldsymbol{EP}$ are linearly independent. That is, any vector in $\boldsymbol{EP}$ cannot be written in the linear combination form of other vectors. Following the proof of Theorem 1, it can be further concluded that $\boldsymbol{EP}$ can be regarded as a set of bases in the $\Re^{2^n}$ space after transforming the original features in Euclidean space into geometric algebraic space. Meanwhile, the representation of the base $\boldsymbol{EP}$ for any vector in the $\Re^{2^n}$ space is unique. In other words, when we map the low dimensional feature x in the original Euclidean space $\Re^n$ to the high dimensional space $\Re^{2^n}$, the basis $\boldsymbol{EP}$ can completely represent the features of any dimension in the $\Re^{2^n}$ space. Because of this completeness, the extension is feasible.

Through the above theorems, it can be proved that the proposed feature extension method is reasonable. The first-order features of patients' biological and clinical information are expanded to multi-level high-order features by the feature extension. These features can more comprehensively describe the patient from a variety of perspectives.

3.2.1.2. Feature filtration algorithm. After feature dimension extension, the first-order features of patients' biological information data and clinical information data are expanded to multi-level high-order features. Some of these multi-level high-order features play an essential role in AD diagnosis. However, some redundant features will interfere with AD diagnosis. Therefore, a feature filtration algorithm is presented in this paper to remove redundant features. To better mine the effect of different features on AD diagnosis from patients' biological information data and clinical information data, an algorithm is designed to calculate the influence degree h_i based on statistical test theory.

Since there are many eigenvectors represented by $\boldsymbol{e}\left(PO(S)_i^m\right)$ (the types here refer to vector types and variable types), different test methods are filtrated for various eigenvectors. In mathematical statistics, on the premise of determining the statistical model, the p-value is used to describe the compatibility between observed data and predicted/expected results. In particular, the differences between data and model predictions are usually measured by test statistics. Therefore, the p-value can be regarded as the possibility that the test statistics filtrated by the research are at least as significant as the observed values when all the model assumptions, including the original assumptions, are correct. Logically, the p-value tests all assumptions generated by the data. Based on the statistical meaning of p-value, $h_i = 1 - p$ is used to measure the influence of currently observed $\boldsymbol{e}\left(PO(S)_i^m\right)$ on AD diagnosis. Besides, due to the diversity of variables to be studied, different test methods (analytical models) should be filtrated for different variables.

Usually, the transformed features are mostly multi-vector. In response to such situations, we first analyzed the data and obtained the following assumptions: (1) the observed values of each group in the sample are independent of each other, (2) the number of samples is sufficient for statistical analysis, (3) there is no multicollinearity among the features, (4) There are no obvious outliers, leverage points and strong influence points in the sample. With the samples satisfying the above assumptions, we can construct a binary logistic regression model to evaluate the effect of the feature group on Alzheimer's disease. It should be noted that our null hypothesis is that the feature group has a significant effect on Alzheimer's disease. We evaluate the multi-vector's influence degree by analyzing the possibility that the hypothesis holds.

In addition, some of the transformed features are single-vector. When testing a single feature, we adopt the mechanism of double testing, that is, we set a two-step test. First, we analyze the influence of this feature on AD diagnosis, and then we add a judgment verification for the rationality of the test to enhance the reliability of the method. Likewise, for dichotomous and categorical variables, we assume (1) the observed values of each group in the sample are independent of each other, (2) the number of samples is sufficient for statistical analysis. For continuous variables, based on the above assumptions, we add (1) the assumption of sample normality and (2) the equal variance of Levene test. For different variable types, we choose different test methods. The process of calculating h_i is sorted into Algorithm 1.

For different types of feature $\boldsymbol{e}\left(PO(S)_i^m\right)$, different processing and calculation methods are designed to obtain influence degrees. We can know from the actual dataset that $\boldsymbol{e}\left(PO(S)_i^m\right)$ may be a single vector or a multi-vector. There are three possible variables for a single vector: dichotomous, categorical, and continuous variables. Based on the relevant theories of the mathematical–statistical hypothesis test, different test methods are designed to measure the impact of different types of variables on AD diagnosis. The algorithm describes this process in detail. Lines 4–6 are used to process dichotomous variable vectors. Lines 8–10 are used to process categorical variable vectors. Lines 12–16

Algorithm 1: calculation of the influence degree h_i

Input: feature vector $e\left(PO(S)_i^m\right)$
Output: influence degree h_i
Obtain $e\left(PO(S)_i^m\right)$;
if $e\left(PO(S)_i^m\right)$ *is a single vector* **then**
  if $e\left(PO(S)_i^m\right)$ *is a dichotomous variable* **then**
    Chi-square (χ^2) test is utilized to test the correlation between the feature and AD;
    $h_i = 1 - p$(where p is the p-value of the test);
    The odds ratio test is utilized to test the differences between the dichotomous variables;
  else
    if $e\left(PO(S)_i^m\right)$ *is a categorical variable* **then**
      $2 \times C$ χ^2 test is utilized to test the correlation between the feature and AD;
      $h_i = 1 - p$;
      Mantel–Haenszel test is utilized to test the differences between different classes;
    else
      Remove significant outliers from data;
      Both the Levene test and normality test are valid;
      Independent-Samples t-test is utilized to test the differences whether the feature effect AD;
      $h_i = 1 - p$;
      Kendall's tau b correlation analysis is utilized to test the correlation between the feature and AD;
    end
  end
else
  The sample has no multilinearity, no obvious outliers, leverage points, and strong influence points;
  Binary logistic regression analysis is utilized to test the correlation between the feature and AD;
  $h_i = 1 - p$;
end
if $h_i > \epsilon$ *(Note: ϵ is a threshold indicating the level of influence on AD diagnosis)* **then**
  Continue;
else
  $h_i = 0$;
end
Return h_i.

are used to process the single-vectors that have continuous variables or other types. Lines 18–20 are used to process multi-vectors.

After running Algorithm 1, the influence degrees of different features on AD diagnosis can be obtained. At the same time, the features with $h_i = 0$ will be cut off during the feature fusion process to eliminate the noise introduced by feature redundancy.

3.2.2. CNN feature extraction network

In Section 3.2.1, feature transformation and feature filtration algorithms are introduced. Besides, the neuroimaging features of patients are also essential for diagnosing AD. In order to extract features from patients' MR images, a feature extraction network structure is designed based on a convolutional neural network. As we all know, CNN is widely used for feature extraction of image data because of its excellent feature extraction ability. VGG [41] is a very typical CNN network model for extracting image features. Considering that this study is oriented to 3D MRI data, 3D CNN needs to be constructed, containing many parameters to be learned. In order to retain the advantages of VGG in image feature extraction and improve the training speed of the model, we optimize the VGG network, including the number of channels, convolution layers, and other hyperparameters. It can adapt to the feature extraction of 3D MR images, reducing the model's complexity, while improving its training speed. This process determines the number of channels and convolution layers through experiments. Our proposed CNN model consists of convolution layers with batch normalization (conv(bn)), max-pooling layers, and fully connected (FC) layers. The network structure and parameter settings are shown in Table 3. When the input 3D MRI has different sizes, the network structure only needs to adjust fewer parameters to adapt to this change.

In this CNN architecture, ReLU (Rectified Linear Unit) is selected as the activation function because it can avoid the disappearance of gradient and prevent overfitting [42]. In the rear of each convolution layer, a batch normalization (BN) layer is added to reduce data fluctuation and improve training speed [43]. To extract the significantly different features of patients, max-pooling is used for feature filtration [44]. In FC layers, Dropout is used to avoid overfitting [45].

3.2.3. ANN feature fusion diagnostic model

In order to fuse the features extracted from neuroimage data with patient profiles, genes, and other features, an ANN feature fusion model is designed in this paper. As shown in Fig. 3, this fusion diagnostic model contains three fully connected layers. Firstly, the neuroimaging features extracted by CNN and the features extracted from patients' clinical and biological information are spliced. Then, two fully connected (FC) layers are established to obtain more profound fusion features. Finally, the softmax layer is used to get the final diagnosis results. Similarly, Dropout is also used in FC layers to avoid overfitting. Meanwhile, Adam optimizer is used to train the whole network for reaching the optimal solution [46].

3.3. Experimental settings

The experiments are implemented on a computer with an Intel Core i9-9900k CPU and an NVIDIA GeForce RTX 2080Ti GPU. Adam [46] is used as an optimizer on Pytorch to train our proposed model. The initial learning rate and decay rate are set to 0.0001 and 0.9. Our proposed model is verified on both AD diagnostic task (i.e., AD versus NC) and MCI diagnostic task (i.e., sMCI versus pMCI). We divided the data set into training set, verification set and test set according to the ratio of 7:1:2. More specifically, four metrics, accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC, obtained by summing the area under the ROC curve) are the most commonly used in bioinformatics literature. We use these to evaluate the proposed model's diagnostic performance.

4. Results and discussions

This section obtains experimental results from our proposed feature fusion diagnostic model in dataset ADNI (Alzheimer's Disease Neuroimaging Initiative). Experimental results of our proposed feature transformation method are shown in Section 4.1. Also, we compared our proposed fusion diagnostic model with several SOTA (state-of-the-art) AD diagnostic models (see Section 4.2).

4.1. Performance of feature transformation method

To assess the performance of our presented feature transformation approach WH-T (with-transformation method), we compare our approach with the other two cases, i.e., WH-I and NO-I. WH-I (with-information) means that all multimodal information data given in this paper is available without feature transformation. They are only input into the diagnosis network as input neurons and fused with neuroimaging features for diagnosing AD. NO-I (no-information) means that only neuroimaging features are used for diagnosing AD. All three methods (NO-I, WH-I, and WH-T) use the same neuroimaging feature extraction architecture (described in Section 3.2) to ensure rationality.

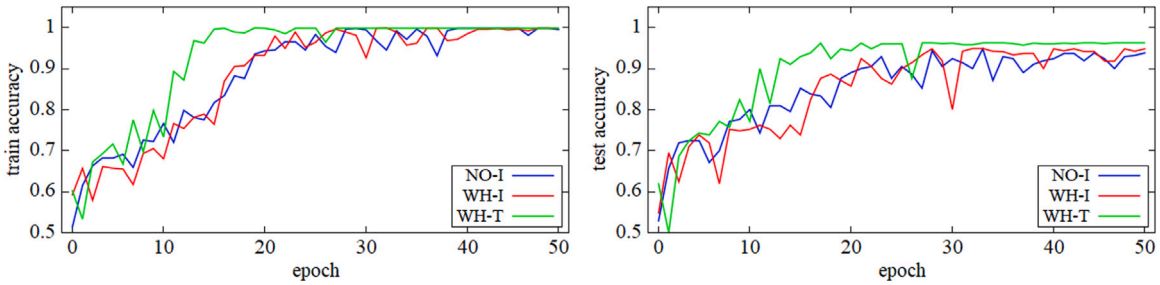

Fig. 4. Accuracy of AD vs. NC diagnostic task according to different epochs. The left one is the train accuracy and the right one is the test accuracy. The green line is our proposed method.

Table 3
The architecture of the CNN feature extraction network.

Layer	Stage	Input	Size	Output
1	conv(bn)	192 × 192 × 160 × 1	3 × 3 × 3 × 8	192 × 192 × 160 × 8
2	max-pooling	192 × 192 × 160 × 8	2 × 2 × 2	96 × 96 × 80 × 8
3	conv(bn)	96 × 96 × 80 × 8	3 × 3 × 3 × 16	96 × 96 × 80 × 16
4	maxpooling	96 × 96 × 80 × 16	2 × 2 × 2	48 × 48 × 40 × 16
5	conv(bn)	48 × 48 × 40 × 16	3 × 3 × 3 × 32	48 × 48 × 40 × 32
6	conv(bn)	48 × 48 × 40 × 32	3 × 3 × 3 × 32	48 × 48 × 40 × 32
7	max-pooling	48 × 48 × 40 × 32	2 × 2 × 2	24 × 24 × 20 × 32
8	conv(bn)	24 × 24 × 20 × 32	3 × 3 × 3 × 64	24 × 24 × 20 × 64
9	conv(bn)	24 × 24 × 20 × 64	3 × 3 × 3 × 64	24 × 24 × 20 × 64
10	max-pooling	24 × 24 × 20 × 64	2 × 2 × 2	12 × 12 × 10 × 64
11	conv(bn)	12 × 12 × 10 × 64	3 × 3 × 3 × 64	12 × 12 × 10 × 64
12	conv(bn)	12 × 12 × 10 × 64	3 × 3 × 3 × 64	12 × 12 × 10 × 64
13	max-pooling	12 × 12 × 10 × 64	2 × 2 × 2	6 × 6 × 5 × 64
14	FC	6 × 6 × 5 × 64	2048	2048
15	FC	2048	256	256

Table 4
Results of two diagnosis tasks in different methods.

Method	AD vs. NC				sMCI vs. pMCI			
	ACC	SPE	SEN	AUC	ACC	SPE	SEN	AUC
NO-I	0.938	0.842	0.974	0.972	0.795	0.755	0.726	0.819
WH-I	0.948	0.772	**0.993**	0.971	0.845	**0.902**	0.715	0.904
WH-T	**0.962**	**0.930**	0.974	**0.986**	**0.874**	0.881	**0.905**	**0.958**

Table 4 shows the performance of different methods. Hence one can see from the table that the diagnostic model with feature transformation can obtain the best accuracy. The feature transformation method can transform the low-dimensional features into high-dimensional features at different levels. The extension of features provides more information for AD diagnosis and reaches more accurately. Moreover, other proposed WH-T model has almost reached the optimal level during different evaluation indicators.

Comparing AD vs. NC task and sMCI vs. pMCI task on the dataset ADNI, it follows that the diagnosis results on AD vs. NC task are generally better than those on sMCI vs. pMCI task. The same results have also appeared in other related studies [8,47]. A reasonable explanation is given for these results. First, in clinical, the neuroimaging features of Alzheimer's disease are mainly manifested in two aspects: the atrophy of the hippocampus and medial temporal lobe; the slow metabolism of deoxyglucose in the inferior parietal lobule, praecuneus, and posterior cingulate cortex. The two main neuroimaging features are easier to distinguish for AD vs. NC but more difficult for sMCI vs. pMCI. In other words, it is more difficult for clinicians to distinguish sMCI vs. pMCI than AD vs. NC accurately. This diagnostic task often requires experienced doctors.

Similarly, it will be more difficult for the neural network to extract features for computer-aided AD diagnosis when diagnosing sMCI vs. pMCI. Therefore, the diagnostic performance of MCI is not more apparent than that of AD. Second, doctors often diagnose the two states of MCI through patients' clinical features (Note: The clinical definition of sMCI is that a patient's situation has not progressed or stabilized within three years after the patient was first diagnosed with MCI. The clinical definition of pMCI is that a patient's situation has changed significantly in recent three years.), rather than simply based on patients' neuroimaging features. The diagnosis often requires doctors to judge in combination with the previous medical history, but computer-aided diagnosis cannot achieve this. Even some patients with sMCI or pMCI have no significant differences in neuroimaging features. Therefore, it also brings some difficulties to computer-aided MCI diagnosis.

Figs. 4–5 show the accuracies values according to different epochs in the two tasks. We can acquire that all models are trained to the optimal state from these figures. It is worth noting that WH-T can achieve the optimal state in fewer epochs. That is, the model has a faster convergence speed. Meanwhile, WH-I can also partly accelerate the convergence speed of the model. Comparative experiments demonstrate that multimodal features can enhance the diagnostic accuracy of AD. At the same time, due to the introduction of more diagnostic information, the training speed of the model is improved, and the optimal state can be achieved in a shorter training time. In addition, the comparative experiments of the WH-I and WH-T indicate that the scientific and reasonable transformation of multimodal features can improve the diagnostic result and further accelerate the convergence.

To further demonstrate the performance of our proposed method, we present the ROC curves of two diagnostic tasks. As shown in Fig. 6, ROC curves can clearly show the diagnostic ability of the models for AD. It can be seen from these figures that the ROC curves of the WH-T group (the green curves in these figures) are closer to the upper left corner. Also, the ROC curves of WH-I and NO-I have smaller areas under the curves. Therefore, it can be concluded that our method is superior to other methods.

The above experimental results demonstrate that the proposed multimodal feature transformation method is effective. Based on the neuroimaging feature extraction network, multimodal features are added to obtain better diagnostic results. At the same time, the multimodal features obtained by using the feature transformation method are better

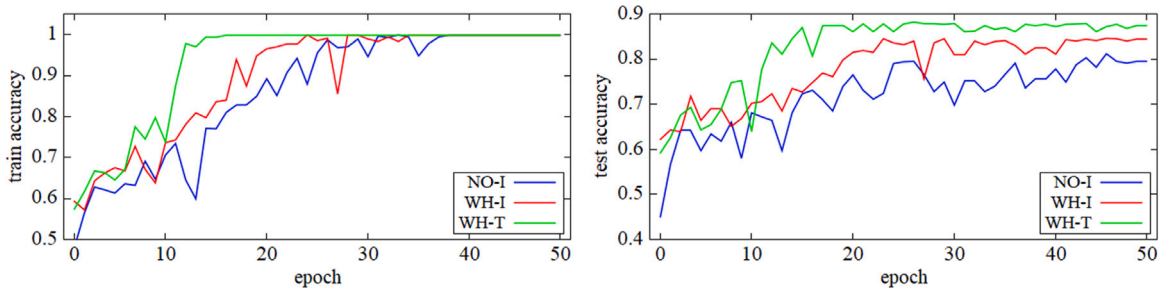

Fig. 5. Accuracy of sMCI vs. pMCI diagnostic task according to different epochs. The left one is the train accuracy and the right one is the test accuracy. The green line is our proposed method.

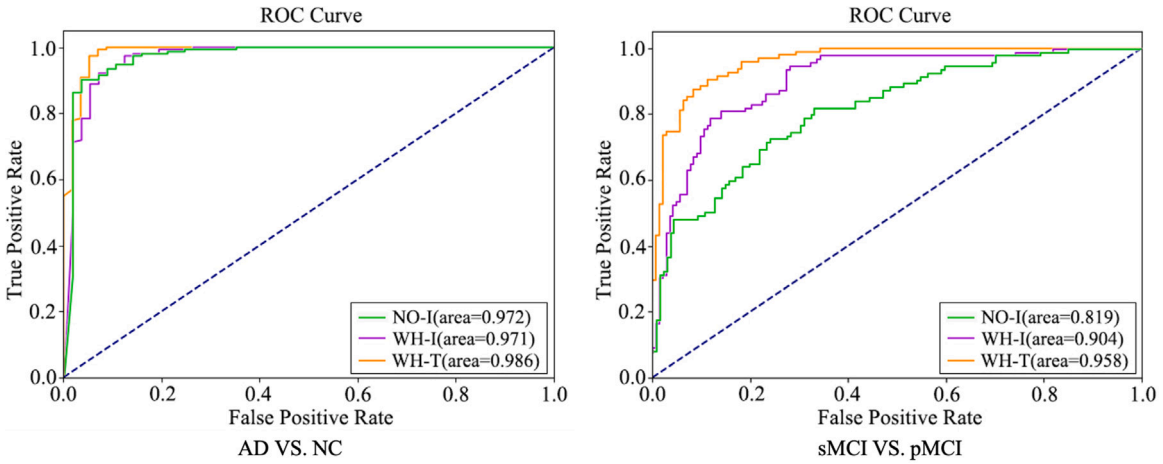

Fig. 6. ROC curves of two diagnostic tasks. The orange lines represent our proposed method. The dark blue dotted line is the boundary line of AUC = 0.5.

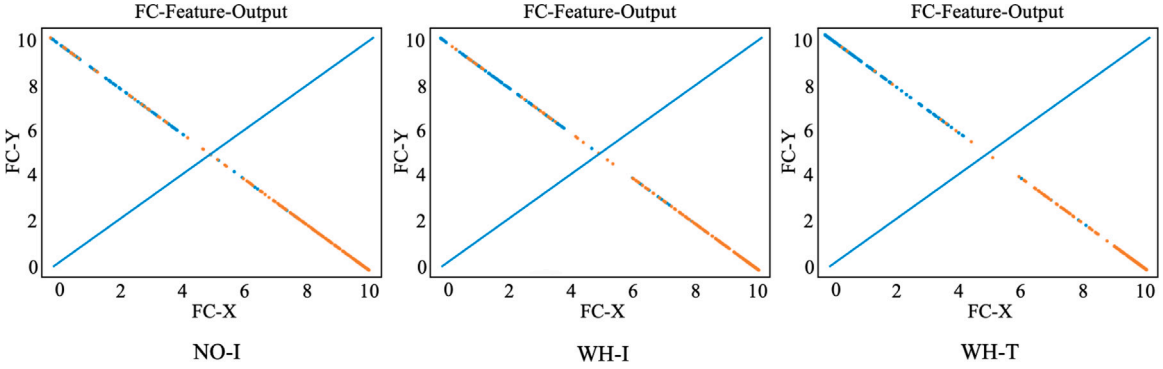

Fig. 7. Feature visualization of AD vs. NC task. Blue dots indicate samples with positive labels, and orange dots indicate samples with negative labels.

than those obtained by directly inputting multimodal information data without transformation. The reason is that the multimodal feature transformation method transforms the original low-dimensional features into different high-dimensional features. The transformation of features can better mine the potential information that impacts AD diagnosis. In addition, the feature filtrating algorithm reduces the input of noise features (i.e., those features that have no significant impact on AD diagnosis) to a certain extent and further improves the diagnostic effect.

Meanwhile, the overall diagnostic model has a faster convergence speed due to multimodal features in the diagnosis process. In addition, the introduction of multimodal features can make the diagnostic model have more references in diagnosing diseases and have a more comprehensive and detailed understanding of the situation of patients. In model training, many comprehensive and detailed patient features are bound to make it easier for the model to diagnose diseases than a single neuroimaging information. That is, the convergence speed of the model is faster than others.

In order to more intuitively show the effect of multimodal feature fusion on Alzheimer's diagnosis, a feature visualization method is used to show the feature output results of the fully connected layer in NO-I, WH-I, and WH-T. Figs. 7–8 respectively show the feature visualization results of three situations in different diagnostic tasks.

The output of a sample can be expressed by an ordered pair (FC-X, FC-Y) in a two-dimensional space, where FC represents the feature vector of FC layer of the network. Since FC layer finally uses the Softmax function for prediction, all samples are on the same straight line. In these figures, blue dots indicate samples with positive labels, and orange dots indicate samples with negative labels. When making a diagnosis, if Y > X, the test sample is judged to be a positive sample; otherwise, it is a negative sample. The straight-line Y = X can be regarded as the judgment boundary between positive and negative

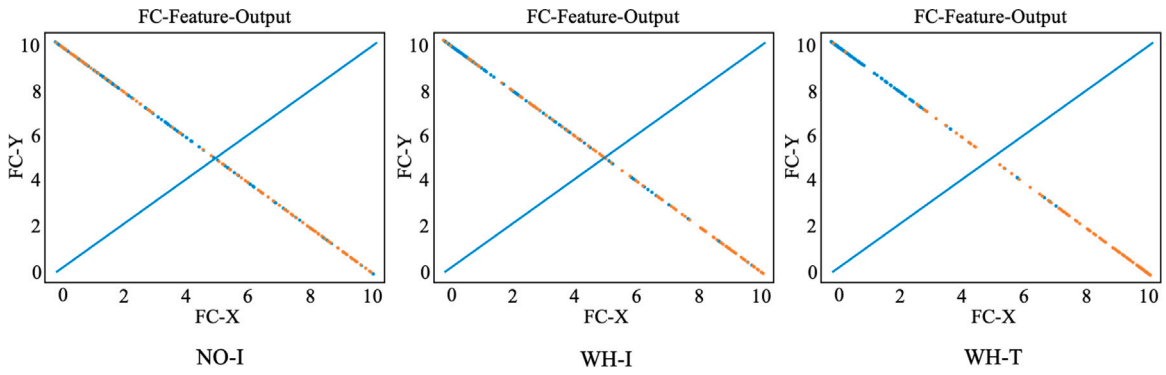

Fig. 8. Feature visualization of sMCI vs. pMCI task. Blue dots indicate samples with positive labels, and orange dots indicate samples with negative labels.

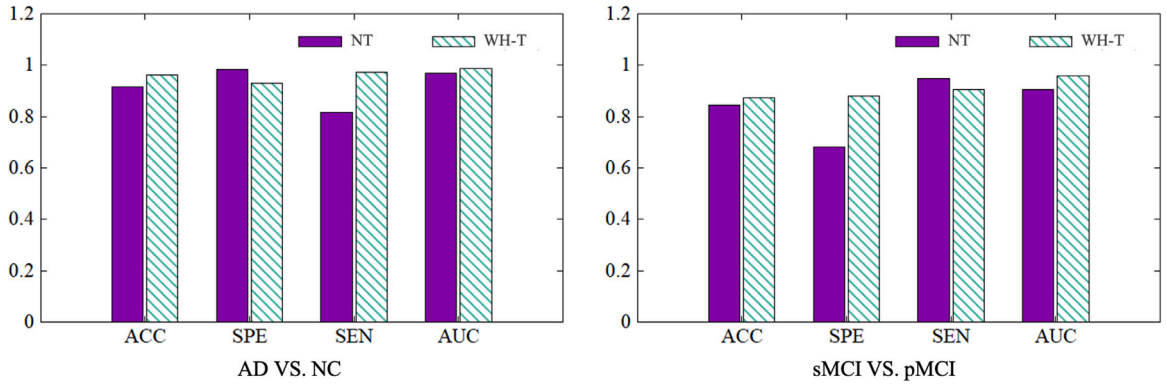

Fig. 9. The comparison results between the NT method and our proposed WH-T method in the two diagnostic tasks. Four evaluation indexes are used to measure diagnostic performance.

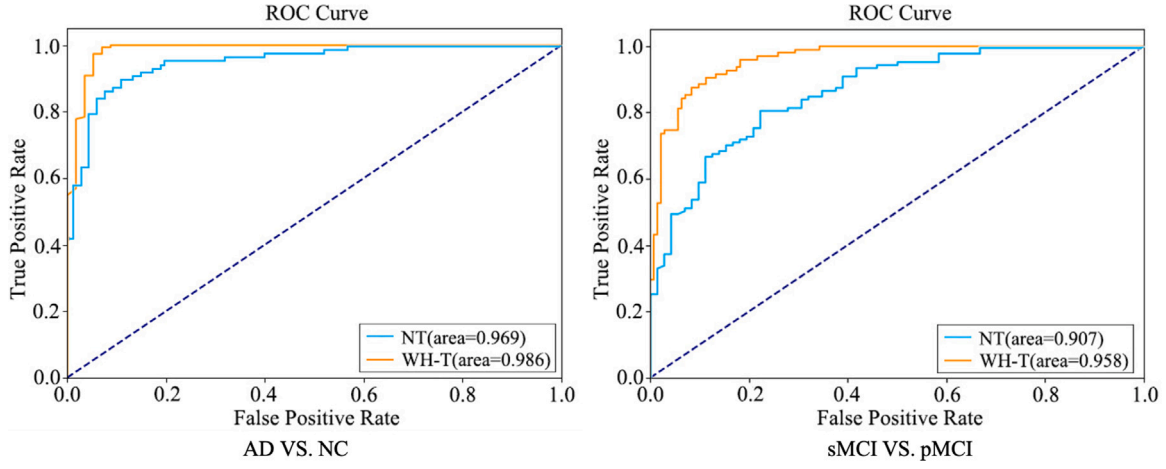

Fig. 10. ROC curves of NT and WH-T methods in the two diagnostic methods. The orange lines represent our proposed model.

samples. From these figures, the WH-T classification accuracy is higher than others. That is, as many positive samples as possible are judged as positive ones, and negative samples are judged as unfavorable ones. Therefore, the multimodal feature fusion diagnostic model proposed in this paper improves the ability of the AD diagnostic model.

In this paper, we propose a feature transformation method, which extends the features of patients' clinical information data and biological information data, and mines the deeper features significant for AD diagnosis through the proposed feature filtration algorithm. Through the ablation experiments of the feature transformation module above, we can see that our proposed WH-T method has obvious advantages. In addition, another consideration of feature transformation is to reduce the dimension of the patients' neuroimaging features, make the dimension consistent with the patients' clinical information data and biological information data, and carry out fusion diagnosis. Therefore, we reduce the dimension of the neuroimaging features extracted by the CNN feature extraction model proposed in this paper through the ANN layers, and fuse them with the unprocessed patients' clinical information data and biological information data to obtain the diagnosis result (named NT). Fig. 9 shows the performance of the two methods under four evaluation indexes. Comparing the NT method with the WH-T method, the performance of the WH-T method is better than that of the NT method in two diagnostic tasks. In addition, we plotted the ROC curves of the two methods, as shown in Fig. 10. The performance of the WH-T method is apparently better than that of the NT method. Objectively, although the NT method reduces the computing time to a

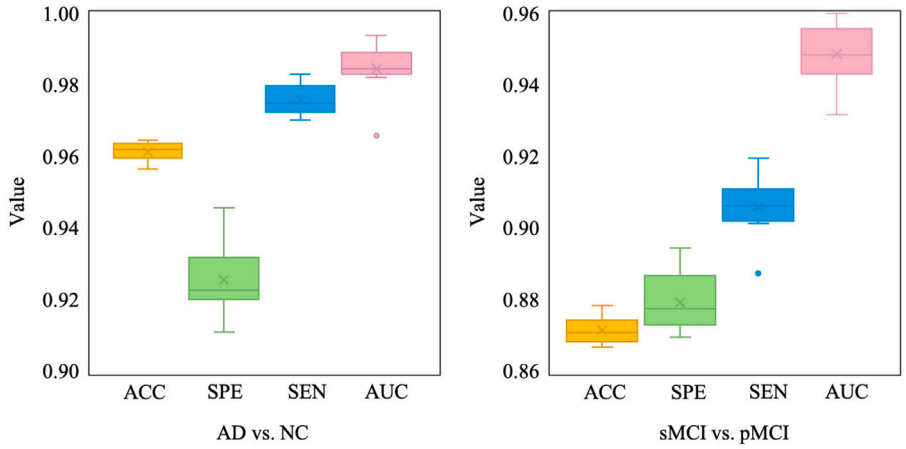

Fig. 11. The boxplots of the cross-validation results in the two diagnostic tasks. The line in the middle of each box represents the median. The cross indicates the mean of the current evaluation index. Independent points outside the box are considered outliers.

Table 5
Diagnostic performance of our proposed method compared with other techniques applied to diagnose AD.

Reference	Subject	AD vs. NC				sMCI vs. pMCI			
		ACC	SPE	SEN	AUC	ACC	SPE	SEN	AUC
Jan et al. [48]	231 NC + 63 sMCI + 168 pMCI +198 AD	0.87	0.81	0.91	–	0.71	0.68	0.74	–
Suk et al. [49]	101 NC + 128 sMCI +76 pMCI + 93 AD	0.92	0.92	0.95	0.97	0.72	0.37	0.91	0.73
Liu et al. [50]	204 NC + 180 AD	0.79	0.83	0.87	0.78	–	–	–	–
Liu et al. [51]	128 NC + 117 sMCI + 117 pMCI + 97 AD	0.93	**0.95**	0.90	0.96	0.79	0.88	0.76	0.83
Korolev et al. [52]	61 NC + 77 sMCI + 43 pMCI + 50 AD	0.80	–	–	0.87	0.52	–	–	0.52
Karasawa et al. [11]	574 NC + 346 AD	0.94	–	–	–	–	–	–	–
Shi et al. [53]	52 NC + 56 sMCI + 43 pMCI + 51 AD	0.95	0.94	0.96	0.96	0.75	0.63	0.85	0.72
Khv et al. [54]	58 NC + 48 AD	0.85	0.88	0.90	–	–	–	–	–
Lin et al. [55]	229 NC + 193 sMCI + 169 pMCI + 188 AD	0.80	0.84	0.75	0.86	0.80	0.86	0.69	0.84
Xu et al. [56]	165 NC + 95 sMCI + 126 pMCI + 142 AD	0.90	0.92	0.89	0.95	0.64	0.79	0.45	0.68
Cui et al. [22]	223NC + 231 sMCI + 165 pMCI + 192 AD	0.91	0.95	0.87	0.93	0.72	0.76	0.65	0.73
Zhu et al. [57]	419 NC + 345 AD	0.92	0.87	0.95	0.96	–	–	–	–
Lian et al. [18]	429 NC + 465 sMCI + 205 pMCI + 358 AD	0.90	0.82	0.97	0.95	0.80	0.53	0.85	0.78
Duc et al. [13]	198 NC + 133 AD	0.85	0.67	–	–	–	–	–	–
Poloni et al. [58]	302 NC + 209 AD	0.83	–	–	0.90	–	–	–	–
Alinsaif et al. [8]	50 NC + 50 sMCI + 50 pMCI + 50 AD	0.90	0.85	0.95	–	0.70	0.60	0.80	–
Zhang et al. [59]	40 NC + 38 AD	0.95	–	–	–	–	–	–	–
Proposed	100 NC + 117 sMCI + 53 pMCI + 78 AD	**0.96****	0.93**	**0.97***	**0.99***	**0.87****	**0.88*****	**0.91****	**0.96***

Note:
*Indicates extremely significant statistical difference ($P < 0.01$).
**Indicates significant statistical difference ($P < 0.05$).
***Indicates no significant statistical difference ($P > 0.05$).

certain extent, it has a poor ability to mine the information contained in patients' clinical information data and biological information data, and ignores some potential features that have a significant impact on the diagnosis of AD. Therefore, the diagnostic ability of the NT method is lower than that of the WH-T method.

4.2. Performance of fusion diagnostic model

In this subsection, the overall performance of the fusion diagnostic model is given. Because the dataset of medical images is not large enough, it is easy to produce overfitting phenomena and significant generalization errors in model training. Hence, a 10-fold cross-validation method is used during training our proposed fusion diagnostic model to ensure that the trained model has better stability and fidelity. Fig. 11 shows the experimental results in different diagnostic tasks. The figure shows that our proposed diagnostic model always performs well in two diagnostic tasks when we transform different data into training and testing sets. Meanwhile, the experimental results also demonstrate that the influence of input data on the model is limited. That is, the method we proposed has specific stability and robustness.

The results of our proposed model are compared with the recent techniques applied to diagnose AD. All these methods are tested on MRI data of the ADNI dataset. The experimental results are shown in Table 5. These methods include traditional feature engineering-based methods [48,51,56,58,59] and deep learning-based methods. According to the different models used, the methods based on deep learning are divided into deep Boltzmann machine-based methods [49], stacked autoencoder-based methods [50], recurrent neural network-based methods [22], deep polynomial network-based methods [53], and CNN-based methods [8,11,18,52,54,55,57] widely used in computer vision. It is worth noting that although the datasets used in all the literatures are from ADNI open-access datasets, the number of subjects used in different literatures are various. However, we can still obtain some observations with a rough comparison of our proposed method (the last row in Table 5) with these SOTA methods. First of all, although we use a different number of subjects dataset from the other literatures in the table for performance evaluation, our method achieves the best performance on both two diagnostic tasks regardless of whether the baseline literatures select more (e.g. [18]) or less (e.g. [8]) data. This comparison is obviously relatively fair. Second, we use a more challenging evaluation method, i.e., we set up an independent training set and test set, which makes our proposed method more credible. Third, our proposed method shows the best performance compared with several CNN-based diagnostic methods, which indicates

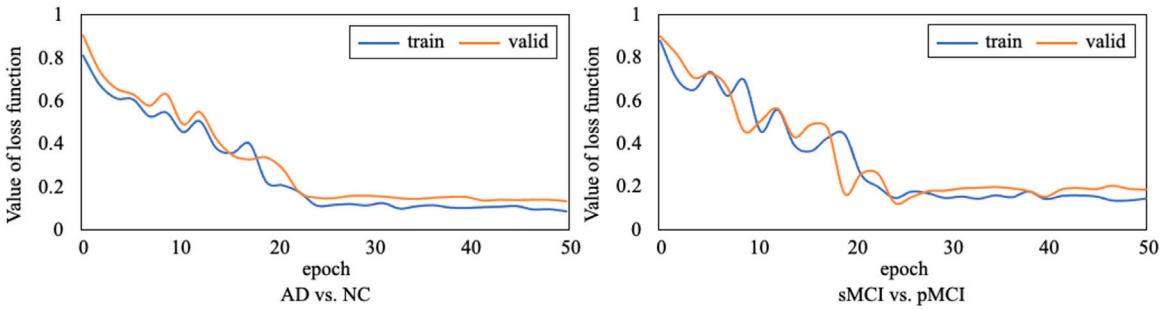

Fig. 12. The training and validation loss function curves of two diagnostic tasks. The blue curve represents the change curve of the loss function on the training set. The orange curve represents the change curve of the loss function on the validation set.

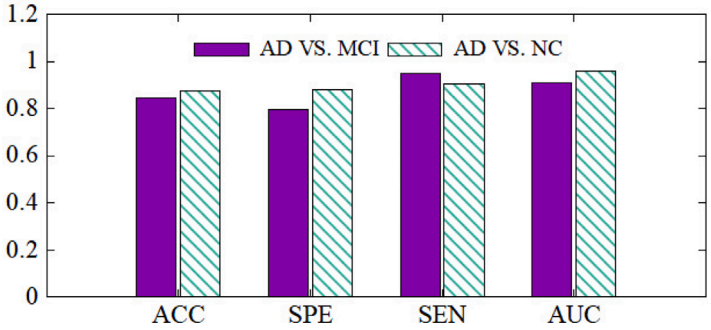

Fig. 13. The comparison results in AD vs. MCI task and AD vs. NC task. Four evaluation indexes (includes ACC, SPE, SEN, AUC) are used to measure diagnostic performance. AD vs. MCI task achieves comparable performance with AD vs. NC task in almost all evaluation indexes.

that the performance of AD diagnosis can be improved by introducing multimodal transformation module. Furthermore, our method achieves better results than the method using untransformed patient clinical information data [13]. It means that the feature transformation method proposed in this paper can mine the latent features of patients at a deeper level and provide more comprehensive patient information for the diagnostic model through multimodal information data feature transformation and fusion. It makes the features more influential for AD diagnosis through feature extension and filtration. Therefore, our method can achieve excellent results.

To demonstrate the convergence of our proposed method, we document the training loss and validation loss values at each epoch in both AD vs. NC task and sMCI vs. pMCI task, as shown in Fig. 12. The loss function values gradually decrease in the first 40 epochs and converge within 50 epochs. It is worth noting that the training loss and validation loss decrease synchronously with the increase of epoch, and finally tend to be stable. This shows that our method can achieve the optimal state and avoid the occurrence of the overfitting phenomenon.

AD vs. MCI task is more challenging to diagnose for AD vs. NC tasks. MCI patients usually have similar neuroimaging and clinical features as AD patients so that MCI patients are hard to be distinguished from AD patients in the diagnosis. Therefore, we mix sMCI class and pMCI class to form MCI class, and use our proposed method to classify AD and MCI. The performance is shown in Fig. 13. As can be seen from the figure, AD vs. MCI task achieves comparable performance with AD vs. NC task in almost all evaluation indexes. Although AD vs. MCI task is more difficult to diagnose than AD vs. NC task, our proposed method can still perform well. This shows that our proposed method has excellent performance and broad applicability.

5. Conclusion

In this study, a novel AD diagnostic model with multimodal feature transformation and fusion is developed. We propose a multimodal feature transformation method to mine more information from patients' profiles, gene sequences, and mental state examination data. A geometric algebra-based multimodal feature extension method is proposed to obtain higher and deeper features. An influence degree-based feature filtration algorithm is proposed to cut out those features, which have no apparent guiding significance for AD diagnosis. Then, a CNN-based neuroimaging feature extracting model is designed to extract features from MRI data. Also, an ANN-based fusion model is designed to fuse the transformed features with the neuroimaging features extracted by CNN. The proposed model has been applied for automated diagnosis of AD in two classification tasks and obtained excellent performance. The proposed model improves the diagnostic accuracy and accelerates the convergence of the diagnostic model. It has been observed that our proposed model can achieve 96.2% accuracy in the AD vs. NC diagnostic task and 87.4% accuracy in the sMCI vs. pMCI diagnostic task when the model has been trained to the optimal situation. Our proposed feature fusion diagnostic model can obtain the optimal diagnostic performance on both diagnostic tasks compared with SOTA methods.

Considering that multimodal features can better guide the model for AD diagnosis, in the future, we will collect more patients' feature information clinically and fuse more multimodal information data for diagnosis to further improve the effectiveness of the model. In addition, if the patients' clinical or biological information data are incomplete, it will be challenging to use the method proposed in this paper for diagnosis. Therefore, the completion of missing data will be regarded as an extension of this work.

Funding

This research was funded by Science and Technology Development Project of Liaoning Province of China under Grant number 2021JH6/10500127 and Science Research Project of Liaoning Department of Education of China under Grant number LJKZ0008.

Declaration of competing interest

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

References

[1] M. Goedert, M.G. Spillantini, A century of Alzheimer's disease, Science 314 (5800) (2006) 777–781.

[2] G.L. Wenk, et al., Neuropathologic changes in Alzheimer's disease, J. Clin. Psychiatry 64 (2003) 7–10.

[3] A.I. Bush, The metallobiology of Alzheimer's disease, Trends Neurosci. 26 (4) (2003) 207–214.

[4] C.G. Lyketsos, M.C. Carrillo, J.M. Ryan, A.S. Khachaturian, P. Trzepacz, J. Amatniek, J. Cedarbaum, R. Brashear, D.S. Miller, Neuropsychiatric symptoms in Alzheimer's disease, 2011.

[5] S. Gauthier, B. Reisberg, M. Zaudig, R.C. Petersen, K. Ritchie, K. Broich, S. Belleville, H. Brodaty, D. Bennett, H. Chertkow, et al., Mild cognitive impairment, Lancet 367 (9518) (2006) 1262–1270.

[6] A. Brugnolo, F. Nobili, M. Barbieri, B. Dessi, A. Ferro, N. Girtler, E. Palummeri, D. Partinico, U. Raiteri, G. Regesta, et al., The factorial structure of the mini mental state examination (MMSE) in Alzheimer's disease, Arch. Gerontol. Geriatr. 49 (1) (2009) 180–185.

[7] J.C. Morris, The clinical dementia rating (cdr): Current version and scoring rules, Young 41 (1991) 1588–1592.

[8] S. Alinsaif, J. Lang, A.D.N. Initiative, et al., 3D shearlet-based descriptors combined with deep features for the classification of Alzheimer's disease based on MRI data, Comput. Biol. Med. (2021) 104879.

[9] A. Puente-Castro, E. Fernandez-Blanco, A. Pazos, C.R. Munteanu, Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques, Comput. Biol. Med. 120 (2020) 103764.

[10] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, M. Filippi, A.D.N. Initiative, et al., Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks, NeuroImage: Clin. 21 (2019) 101645.

[11] H. Karasawa, C.-L. Liu, H. Ohwada, Deep 3d convolutional neural network architectures for alzheimer's disease diagnosis, in: Asian Conference on Intelligent Information and Database Systems, Springer, 2018, pp. 287–296.

[12] M. Liu, J. Zhang, E. Adeli, D. Shen, Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 66 (5) (2018) 1195–1206.

[13] N.T. Duc, S. Ryu, M.N.I. Qureshi, M. Choi, K.H. Lee, B. Lee, 3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI, Neuroinformatics 18 (1) (2020) 71–86.

[14] S. Li, X. Yuan, F. Pu, D. Li, Y. Fan, L. Wu, W. Chao, N. Chen, Y. He, Y. Han, Abnormal changes of multidimensional surface features using multivariate pattern classification in amnestic mild cognitive impairment patients, J. Neurosci. 34 (32) (2014) 10541–10553.

[15] J. Baron, G. Chetelat, B. Desgranges, G. Perchey, B. Landeau, V. De La Sayette, F. Eustache, In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease, Neuroimage 14 (2) (2001) 298–309.

[16] B. Richhariya, M. Tanveer, A. Rashid, A.D.N. Initiative, et al., Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE), Biomed. Signal Process. Control 59 (2020) 101903.

[17] M. Liu, D. Cheng, K. Wang, Y. Wang, Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis, Neuroinformatics 16 (3) (2018) 295–308.

[18] C. Lian, M. Liu, J. Zhang, D. Shen, Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE Trans. Pattern Anal. Mach. Intell. 42 (4) (2020) 880–893.

[19] J. Zhang, Y. Gao, Y. Gao, B.C. Munsell, D. Shen, Detecting anatomical landmarks for fast Alzheimer's disease diagnosis, IEEE Trans. Med. Imaging 35 (12) (2016) 2524–2533.

[20] S.-H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, H. Cheng, Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling, J. Med. Syst. 42 (5) (2018) 1–11.

[21] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, Y. Li, A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease, Neurocomputing 320 (2018) 195–202.

[22] R. Cui, M. Liu, A.D.N. Initiative, et al., RNN-based longitudinal analysis for diagnosis of Alzheimer's disease, Comput. Med. Imaging Graph. 73 (2019) 1–10.

[23] J. Islam, Y. Zhang, Early diagnosis of Alzheimer's disease: A neuroimaging study with deep learning architectures, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 1881–1883.

[24] Y. Pan, M. Liu, C. Lian, T. Zhou, Y. Xia, D. Shen, Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 455–463.

[25] H.-C. Shin, A. Ihsani, Z. Xu, S. Mandava, S.T. Sreenivas, C. Forster, J. Cha, A.D.N. Initiative, et al., GANDALF: Generative adversarial networks with discriminator-adaptive loss fine-tuning for Alzheimer's disease diagnosis from MRI, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 688–697.

[26] H. Wang, Y. Shen, S. Wang, T. Xiao, L. Deng, X. Wang, X. Zhao, Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and alzheimer's disease, Neurocomputing 333 (2019) 145–156.

[27] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, B. Lei, Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM, IEEE Access 7 (2019) 63605–63618.

[28] S. Basu, K. Wagstyl, A. Zandifar, L. Collins, A. Romero, D. Precup, Early prediction of alzheimer's disease progression using variational autoencoders, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 205–213.

[29] C. Lian, M. Liu, L. Wang, D. Shen, End-to-end dementia status prediction from brain mri using multi-task weakly-supervised attention network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 158–167.

[30] T. Zhou, K.-H. Thung, X. Zhu, D. Shen, Feature learning and fusion of multimodality neuroimaging and genetic data for multi-status dementia diagnosis, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2017, pp. 132–140.

[31] X. Hao, Y. Bao, Y. Guo, M. Yu, D. Zhang, S.L. Risacher, A.J. Saykin, X. Yao, L. Shen, A.D.N. Initiative, et al., Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease, Med. Image Anal. 60 (2020) 101625.

[32] A. Abrol, Z. Fu, Y. Du, V.D. Calhoun, Multimodal data fusion of deep learning and dynamic functional connectivity features to predict Alzheimer's disease progression, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 4409–4413.

[33] F. Ramzan, M.U.G. Khan, A. Rehmat, S. Iqbal, T. Saba, A. Rehman, Z. Mehmood, A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks, J. Med. Syst. 44 (2) (2020) 1–16.

[34] E. Ryzhikova, N.M. Ralbovsky, V. Sikirzhytski, O. Kazakov, L. Halamkova, J. Quinn, E.A. Zimmerman, I.K. Lednev, Raman spectroscopy and machine learning for biomedical applications: Alzheimer's disease diagnosis based on the analysis of cerebrospinal fluid, Spectrochim. Acta A 248 (2021) 119188.

[35] E.N. Marzban, A.M. Eldeib, I.A. Yassine, Y.M. Kadah, A.D.N. Initiative, Alzheimer's disease diagnosis from diffusion tensor images using convolutional neural networks, PLoS One 15 (3) (2020) e0230409.

[36] G. Fiscon, E. Weitschek, M.C. De Cola, G. Felici, P. Bertolazzi, An integrated approach based on EEG signals processing combined with supervised methods to classify Alzheimer's disease patients, in: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2018, pp. 2750–2752.

[37] C.R. Jack Jr., M.A. Bernstein, N.C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P.J. Britson, J. L. Whitwell, C. Ward, et al., The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods, J. Magn. Reson. Imaging: Off. J. Int. Soc. Magn. Reson. Med. 27 (4) (2008) 685–691.

[38] S. Lovestone, P. Francis, I. Kloszewska, P. Mecocci, A. Simmons, H. Soininen, C. Spenger, M. Tsolaki, B. Vellas, L.-O. Wahlund, et al., AddNeuroMed—the European collaboration for the discovery of novel biomarkers for Alzheimer's disease, Ann. New York Acad. Sci. 1180 (1) (2009) 36–46.

[39] C.-C. Liu, T. Kanekiyo, H. Xu, G. Bu, Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy, Nat. Rev. Neurol. 9 (2) (2013) 106–118.

[40] L. Jia, H. Xu, S. Chen, X. Wang, J. Yang, M. Gong, C. Wei, Y. Tang, Q. Qu, L. Chu, et al., The APOE ε4 exerts differential effects on familial and other subtypes of Alzheimer's disease, Alzheimer's Dement. 16 (12) (2020) 1613–1623.

[41] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[42] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines Vinod Nair, in: Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel, 2010, pp. 807–814.

[43] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, PMLR, 2015, pp. 448–456.

[44] Y.T. Zhou, R. Chellappa, Computation of optical flow using a neural network, in: Neural Networks, 1988., IEEE International Conference on, 2002, pp. 71–78.

[45] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.

[46] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[47] A. Mehmood, S. Yang, Z. Feng, M. Wang, A.S. Ahmad, R. Khan, M. Maqsood, M. Yaqub, A transfer learning approach for early diagnosis of alzheimer's disease on MRI images, Neuroscience 460 (2021) 43–52.

[48] E. Janoušová, M. Vounou, R. Wolz, K.R. Gray, D. Rueckert, G. Montana, et al., Biomarker discovery for sparse classification of brain images in Alzheimer's disease, Ann. BMVA (2) (2012).
[49] H.-I. Suk, S.-W. Lee, D. Shen, A.D.N. Initiative, et al., Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis, NeuroImage 101 (2014) 569–582.
[50] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M.J. Fulham, et al., Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease, IEEE Trans. Biomed. Eng. 62 (4) (2015) 1132–1140.
[51] M. Liu, D. Zhang, D. Shen, Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment, IEEE Trans. Med. Imaging 35 (6) (2016) 1463–1474.
[52] S. Korolev, A. Safiullin, M. Belyaev, Y. Dodonova, Residual and plain convolutional neural networks for 3D brain MRI classification, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), IEEE, 2017, pp. 835–838.
[53] J. Shi, X. Zheng, Y. Li, Q. Zhang, S. Ying, Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease, IEEE J. Biomed. Health Inf. 22 (1) (2018) 173–183.
[54] A. Khvostikov, K. Aderghal, J. Benois-Pineau, A. Krylov, G. Catheline, 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies, 2018, arXiv preprint arXiv:1801.05968.
[55] W. Lin, T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, X. Qu, et al., Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment, Front. Neurosci. 12 (2018) 777.
[56] L. Xu, Z. Yao, J. Li, C. Lv, H. Zhang, B. Hu, Sparse feature learning with label information for Alzheimer's disease classification based on magnetic resonance imaging, IEEE Access 7 (2019) 26157–26167.
[57] T. Zhu, C. Cao, Z. Wang, G. Xu, J. Qiao, Anatomical landmarks and DAG network learning for Alzheimer's disease diagnosis, IEEE Access 8 (2020) 206063–206073.
[58] K.M. Poloni, I.A.D. de Oliveira, R. Tam, R.J. Ferrari, A.D.N. Initiative, et al., Brain MR image classification for Alzheimer's disease diagnosis using structural hippocampal asymmetrical attributes from directional 3-D log-Gabor filter responses, Neurocomputing 419 (2021) 126–135.
[59] Y. Zhang, S. Wang, K. Xia, Y. Jiang, P. Qian, A.D.N. Initiative, et al., Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion, Inf. Fusion 66 (2021) 170–183.

Yue Tu was born in Liaoning Province, China, in 1994, received the B.S. degree and M.S. degrees in computer science and technology from Northeastern University in 2016 and 2018. He is currently pursuing the Ph.D. degree in computer science and technology at Northeastern University, China. His current research areas include medical image analysis and computer vision.

Shukuan Lin was born in Jilin Province, China, in 1966, graduated from the Department of Computer Science of Jilin University in 1988. Currently, she is a professor of Northeastern University. Her current research areas include medical image analysis and computer vision.

Jianzhong Qiao was born in Liaoning Province, China, in 1964, graduated from the Department of Computer Science of Xi'an Jiaotong University in 1986. Currently, he is a professor of Northeastern University. His current research areas include pattern recognition and computer vision.

Yilin Zhuang was born in Jilin Province, China, in 1998, received the B.S. degree in computer science and technology from Changchun University of Science and Technology in 2020. He is currently pursuing the M.S. degree in computer science and technology at Northeastern University, China. His current research areas include medical image analysis and computer vision.

Peng Zhang was born in Liaoning Province, China, in 1996, received the B.S. degree and M.S. degrees in computer science and technology from Northeastern University in 2018 and 2021. His current research areas include medical image analysis and computer vision.