Pharmacological Research 199 (2024) 107038

Contents lists available at ScienceDirect

Pharmacological Research

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

Disease2Vec: Encoding Alzheimer's progression via disease embedding tree

Lu Zhang [a], Li Wang [b], Tianming Liu [c], Dajiang Zhu [a,*]

[a] Department of Computer Science and Engineering, The University of Texas at Arlington, Arlington, TX, USA
[b] Department of Mathematics, The University of Texas at Arlington, Arlington, TX, USA
[c] Department of Computer Science, The University of Georgia, Athens, GA, USA

ARTICLE INFO

Keywords:
AD Progression
Disease Embedding
Disease Embedding Tree

ABSTRACT

For decades, a variety of predictive approaches have been proposed and evaluated in terms of their prediction capability for Alzheimer's Disease (AD) and its precursor – mild cognitive impairment (MCI). Most of them focused on prediction or identification of statistical differences among different clinical groups or phases, especially in the context of binary or multi-class classification. The continuous nature of AD development and transition states between successive AD related stages have been typically overlooked. Though a few progression models of AD have been studied recently, they were mainly designed to determine and compare the order of specific biomarkers. How to effectively predict the individual patient's status within a wide spectrum of continuous AD progression has been largely understudied. In this work, we developed a novel learning-based embedding framework to encode the intrinsic relations among AD related clinical stages by a set of meaningful embedding vectors in the latent space (*Disease2Vec*). We named this process as disease embedding. By Disease2Vec, our framework generates a disease embedding tree (DETree) which effectively represents different clinical stages as a tree trajectory reflecting AD progression and thus can be used to predict clinical status by projecting individuals onto this continuous trajectory. Through this model, DETree can not only perform efficient and accurate prediction for patients at any stages of AD development (across five fine-grained clinical groups instead of typical two groups), but also provide richer status information by examining the projecting locations within a wide and continuous AD progression process. (Code will be available: https://github.com/qidianzl/Disease2Vec.)

1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia that cannot be prevented, cured, or even slowed. Earlier studies have shown that AD pathogenesis involve widespread alterations in brain structure and/or function, such as hippocampi [1], gray matter atrophy [2], white matter disruption [3] and abnormal functional connectivity in default mode network (DMN) [4]. Based on these brain alterations, many approaches have been developed for early diagnosis of AD and its prodromal stage – mild cognitive impairment (MCI), such as voxel-based analysis [5], tract-based spatial statistics [6], and recently developed machine learning/deep learning-based models [7–9]. However, as a neurodegenerative disorder with a long pre-clinical period, the spectrum of AD spans from clinically asymptomatic to severely impaired [10]. For example, heterogeneity in clinical presentation, rate of atrophy and cognitive decline [11] may occur in the prodromal stage of AD [12]. Furthermore, individual variations may also contribute to the heterogeneity of AD: earlier studies suggested that the gap between cognitive function and brain pathology (i.e., cognitive reserve) is typically larger in highly educated individuals [13]. In general, traditional predictive approaches (e.g., classification-based models) may be limited in describing the continuum of AD development and individual variations in clinical prediction. To address this potential limitation, hypothetical models [14] for AD progression have been proposed and followed by various progression studies using cross-sectional or short-term follow-up dataset. These attempts include regression-based models [15], event-based models [16] and other computational models [17]. Nevertheless, most of them were designed to determine the order of biomarkers. Because these models consider different measures or biomarkers separately, they create a different trajectory for each biomarker. Consequently, different models and assumptions may lead to inconsistent results and interpretations [16]. More importantly, previous AD progression models are based on population analysis, they cannot be directly used for individualized diagnosis and prediction.

* Correspondence to: Department of Computer Science and Engineering, The University of Texas at Arlington, Arlington, TX 76019, USA.
E-mail address: dajiang.zhu@uta.edu (D. Zhu).

https://doi.org/10.1016/j.phrs.2023.107038
Received 22 August 2023; Received in revised form 6 November 2023; Accepted 7 December 2023
Available online 10 December 2023

Recent advancements in deep learning have brought about a paradigm shift in representation learning, leading to the emergence of powerful embedding techniques across various domains. Notably, word embedding methods have made remarkable contributions in the field of natural language processing (NLP) [18]. These methods aim to generate dense, continuous vector representations of words, effectively representing and capturing the complex semantic relationships among them. That is, the words sharing similar semantic meanings can be projected as the vector representations that are closely situated in a high-dimensional space. These embedding methods have demonstrated exceptional performance across a range of downstream tasks [18]. Similarly, a groundbreaking approach known as cortex2vector [19] has recently been introduced. This method focuses on encoding the cortical folding patterns into a group of anatomically meaningful embedding vectors. By leveraging this technique, the anatomical similarity of different brain landmarks can be effectively represented by the similarity of their corresponding embedding vectors. The core idea underlying these embedding methods lies in the direct shaping of the latent space (*embedding space*), aligning it with the semantic space. By mapping complex data into the latent space, different positions within this space correspond to distinct semantic or anatomical information. This alignment enables efficient representation and analysis of complex relationships and therefore enhances our ability to explore and understand intricate data structures.

Inspired by the abovementioned remarkable successes of embedding methods, in this work, we designed a new learning-based embedding framework to encode the entire AD progression by a set of meaningful embedding vectors in the latent space (*Disease2Vec*). By employing this approach, the latent space is aligned with the AD progression trajectory, allowing for the representation of intrinsic relationships between different clinical stages through the learned embedding vectors. Fig. 1 provides a visual depiction of the core idea behind our framework. During the training process, we introduce a novel ordered embedding method that shapes the latent space in alignment with the disease development process. This results in a set of learnable embeddings for different clinical groups. These group embedding vectors are orderly arranged in the latent space, directly corresponding to different stages of disease progression. To effectively capture individuality, we transform the input features of individuals into the latent space to obtain individual embeddings. Through the learning process, the distribution of these individual embeddings in the space reflects the corresponding individual's clinical status within the entire disease development process. The proposed framework jointly learns the clinical group embeddings and individual embeddings to better shape the latent space. As a result, we obtain a tree-based trajectory within the latent space, known as the Disease Embedding Tree (*DETree*). This trajectory effectively integrates AD progression modeling and individual prediction. During the prediction process, new individuals can be projected onto the continuous trajectory of the learned DETree. This allows our model not only to assign clinical groups to new patients but also to indicate their clinical status throughout the entire development process, spanning from normal cognition (NC) to AD. With the learned DETree, our model achieves a relatively high classification accuracy – 77.8% for multi-class classification (NC vs. SMC vs. EMCI vs. LMCI vs. AD), compared to other established machine learning/deep learning methods [7,13,20–28]. Furthermore, the proposed DETree framework is versatile and can be adapted to a wide range of neurodevelopmental, neurodegenerative, and psychiatric disorders that exhibit multiple clinical stages during the development.

2. Materials and methods

2.1. Data

2.1.1. Datasets description and data pre-processing

In this work, we used 266 subjects (60 NC, 34 SMC, 51 EMCI, 62 LMCI and 59 AD) from the ADNI dataset (http://adni.loni.usc.edu/). Each subject has both structure MRI (T1-weighted) and resting state fMRI (rs-fMRI) data. For T1-weighted MRI, FOV $= 240 \times 256 \times 208$ mm^3, voxel size $= 1.0$ mm isotropic, and TR $= 2.3$ s. The rs-fMRI data has 197 volumes, FOV $= 220 \times 220 \times 163$ mm^3, voxel size $= 3.3$ mm isotropic, TR $= 3$ s, TE $= 30$ ms and flip angle $= 90^\circ$. We followed the standardized preprocessing procedures adopted in [29,30] for imaging data. Specifically, we applied skull removal for both T1 and rs-fMRI modalities. And for rs-fMRI images, the first 6 volumes were discarded during preprocessing procedures to ensure magnetization equilibrium. Then we applied spatial smoothing, slice time correction, temporal pre-whitening, global drift removal and band pass filtering (0.01–0.1 Hz). All these preprocessing steps are implemented using FMRIB Software Library (FSL) (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) FEAT. For T1 images, we conducted segmentation by FreeSurfer package (https://surfer.nmr.mgh.harvard.edu/). After the segmentation, we adopted the Destrieux Atlas for ROI labeling, and the brain cortex is partitioned into 148 regions.

2.1.2. Generation of functional connectivity

We calculated averaged fMRI signal for each brain region. Previous studies [31,32] suggested that for rs-fMRI 14 time points (when TR=2 s) are sufficient to capture functional dynamic patterns. To enlarge the dataset, we divided the signal into four non-overlapping segments and each segment has 45 time points. We used Pearson Correlation

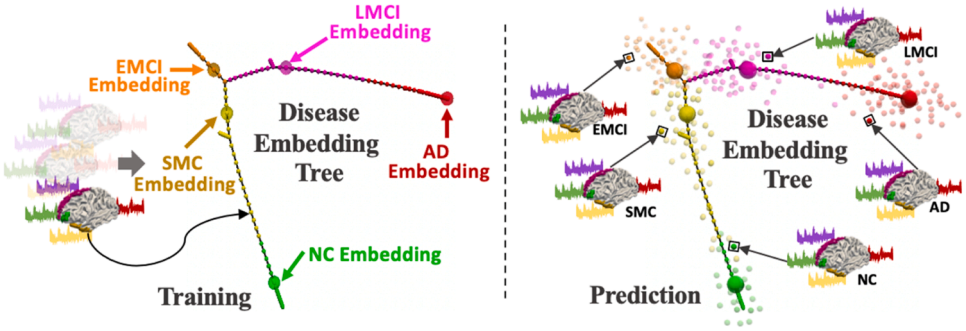


Fig. 1. Training: we used functional connectivity as input and learned a Disease Embedding Tree (*DETree*) to model the entire progression of AD in the latent space. In the tree structure, each small bubble represents a single subject, and the colors indicate different clinical groups, including normal control – NC (*green*), significant memory concern – SMC (*yellow*), early MCI – EMCI (*orange*), late MCI – LMCI (*pink*) and AD (*red*). Each edge in the DETree indicates that the connecting two nodes have higher similarity in the latent space. The five larger bubbles represent the learned group embeddings. **Prediction:** During the prediction, new patients will be projected into the latent space which are represented as scattered bubbles. The color of the bubble indicates the true label, the location of the bubble shows its state in the entire development process from NC to AD, and the prediction of the bubble is based on the nearest group embedding.

Coefficient to calculate functional connectivity for each of the four groups of the signal segments and obtained four functional connectivity matrices for each subject. These functional connectivity matrices were vectorized and used as input of our model.

2.2. Method overview

We proposed a DETree framework to represent the continuum of AD development process as a tree structure embedded in a latent space. Here, an embedding is an abstract representation defined in latent space that is associated with a specific clinical stage (*group embedding*) or a specific individual (*individual embedding*). We parameterized a set of group embeddings as hidden variables in latent space (Section 2.3) and used the order information of clinical groups (NC → SMC → EMCI → LMCI → AD) to guide the embedding process (Section 2.4). In general, the proposed model aims to learn a deep representation of the input signals in a latent space that is specially optimized for both tasks simultaneously: the individual prediction and the AD progression learning. As a result, on the learned tree structure, the patients with similar clinical status are close and distant otherwise. Moreover, DETree can predict the clinical stage for a new patient by projecting it to the appropriate location on the learned tree structure (Section 2.4). Next, we will present the details of DETree and its predictive capability for new patients.

2.3. Disease embedding learning

Let $\{(x_i, y_i)\}_{i=1}^{n}$ be the training data consisting of n labeled data with the i^{th} input $x_i \in \mathscr{R}^{d'}$ and class label $y_i \in \{1, \ldots, C\}$ with C disease stages. To maintain representative instances for different disease stages, we parameterized and learned a set of embeddings in the latent space and used embedding matching for classification.

First, we learned a non-linear function $h(x, \theta) : \mathscr{R}^d \rightarrow \mathscr{R}^k$ to transform any given input $x \in \mathscr{R}^d$ to a latent space $\mathscr{R}^k$ with learnable model parameter θ to obtain individual embedding. And then, we defined a set of clinical group embeddings as $\mathscr{E} = \{e_{ij} \in \mathscr{R}^k | i = 1, 2, \ldots, C; j = 1, 2, \ldots, K\}$ where K is the number of the embeddings in each class. With the help of the non-linear transformation function $h(x, \theta)$ and the set of embeddings $\mathscr{E}$ in the latent space, we can make prediction for any given data. Specifically, given an input data $x \in \mathscr{R}^d$, we first generated its representation (individual embedding) $h(x, \theta)$ in latent space, then we compared the individual embedding with all clinical group embeddings and classified it to the category y, which is the nearest clinical group that the embedding belongs to:

$$y = \underset{i \in \{1,2,\cdots C\}}{\operatorname{argmin}} \min_{j \in \{1,2,\cdots K\}} \left\| h(x, \theta) - e_{ij} \right\|_2^2 \tag{1}$$

The network parameters θ and clinical group embeddings $\mathscr{E}$ can be trained jointly in an end-to-end manner, which can make the model $h(x, \theta)$ and clinical group embeddings interact with each other for better performance. To train the model, we need to define a proper loss function such that 1) it is differentiable with respect to θ and $\mathscr{E}$, and 2) it should be closely related to the classification accuracy.

Embedding Learning Based Cross Entropy Loss. In our DETree model, we used distance to measure the similarity between the individual embeddings and the clinical group embeddings. The class label of the clinical group embedding e_{ij} can be denoted by y_{ij}, to indicate the j^{th} embedding of class y_i. Thus, the probability of an input x belongs to class y_i (i.e., e_{ij} is the nearest embedding of x) is formulated as:

$$P(y_{ij}|x) = \frac{\exp\left\{ -\alpha \left\| h(x, \theta) - e_{ij} \right\|_2^2 \right\}}{\sum_{l=1}^{C} \sum_{m=1}^{K} \exp\left\{ -\alpha \left\| h(x, \theta) - e_{lm} \right\|_2^2 \right\}} \tag{2}$$

where α is a hyper-parameter that controls the hardness of distance in probability assignment. Given the definition of $P(y_{ij}|x)$, we can further define the probability of an input x belonging to the category $c \in \{1, 2, \ldots, C\}$ as:

$$P(c|x) = \sum_{j=1}^{K} P(y_{cj}|x) \tag{3}$$

Then, we defined a classification loss function based on the probability $P(c|x)$ and named it as embedding learning based cross entropy loss given by:

$$\mathscr{L}_{\mathscr{E}}((x, y); \theta_{\mathscr{E}}) = -\frac{1}{C} \sum_{c=1}^{C} \mathbb{I}(c = y) \log P(c|x) \tag{4}$$

where indicator function $\mathbb{I}(c = y)$ is 1 if predicator $c = y$ is true and 0 otherwise.

From (2), (3) and (4), we can see that optimizing the embedding based cross entropy loss essentially corresponds to decreasing the distance between the individual embedding $h(x, \theta)$ of input sample x and the clinical group embedding vector, which comes from the true category of x. By this way, the distance of two input samples at the same disease stage will be small in the latent space, and the disease related representative clinical group embeddings can be automatically learned from data.

To improve the generalization performance and prevent over-fitting, we also proposed a new embedding-based regularization term:

$$\mathscr{L}_{\mathscr{ER}}((x, y); \theta_{\mathscr{E}}) = \left\| h(x, \theta) - e_{y,*} \right\|_2^2 \tag{5}$$

where $e_{y,*}$ is the closest group embedding of $h(x, \theta)$ with class label y. The regularization term pulls the individual embedding $h(x, \theta)$ of input sample x close to its corresponding clinical group embedding, making the individual embeddings within the same class more compact, so it is beneficial for classification.

2.4. Ordered Embedding Constraint

The class labels y provides not only the separability of their inputs, but also the underlying relationship of the clinical groups, which corresponds to different disease stages during the progression of AD. It is generally assumed that the ordering of the clinical groups is NC → SMC → EMCI → LMCI → AD. Even though the ordering of each input sample is unknown, the ordering of the classes can still provide valuable information to guide the embedding learning. To take advantage of this prior knowledge, we constructed an affinity matrix $\mathscr{A} = [a_{(ij),(i'j')}] \in \mathscr{R}^{N \times N}$ for the similarity among embedding class labels. $N = C \times K$ is the total number of embeddings, where C is the number of clinical stages and K is the number of the embeddings in each class. $a_{(ij),(i'j')} = 1$ if the $(i, j)^{th}$ embedding and the $(i', j')^{th}$ embedding are from the same class, that is $y_i = y_{i'}$, $a_{(ij),(i'j')} = 0.5$ if y_i is the neighbor of $y_{i'}$ in the ordering of class labels, and 0 otherwise.

To leverage this prior information for learning the path of AD progression, we added an additional neural network layer with softmax function onto the embeddings to link the clinical group embeddings and the different classes (stages) of AD. As a result, the output probability of clinical group embedding e_{ij} belonging to the class c is formulated as:

$$O_c(e_{ij}; W, b) = \frac{\exp\{(W_c^T e_{ij} + b_c)\}}{\sum_{l=1}^{C} \exp\{(W_l^T e_{ij} + b_l)\}} \tag{6}$$

where $\{W_l, b_l\}$ are the parameters of the neural network layer. The final prediction is:

$$y_{ij} = \underset{c \in \{1,2,\cdots C\}}{\operatorname{argmax}} O_c(e_{ij}; W, b) \tag{7}$$

According to (6) and (7), the classification loss of clinical group embeddings is defined as:

$$\mathscr{L}_{\mathscr{O}}\left(e_{i,j};W,b\right)=-\frac{1}{C}\sum_{c=1}^{C}\mathbb{I}(c=i)\quad\log\quad O_c\left(e_{i,j};W,b\right)\tag{8}$$

Then, we proposed the following regularization term to incorporate the ordering information of the class labels in terms of the affinity matrix $\mathscr{A}$ based on the manifold assumption: if two labels are similar, their probabilities of predictions should be close. The regularization term is then formulated as:

$$\mathscr{L}_{\mathscr{O}\mathscr{R}}(\mathscr{E}W,b)=trace\left(OL_yO^T\right)\tag{9}$$

where $O=[O_1;O_2;\cdots;O_C]\in\mathscr{R}^{C\times(C\times K)}$ and $O_c=\left[O_c\left(e_{ij};W,b\right)\right]_{\{ij\}}\in\mathscr{R}^{1\times(C\times K)},\forall c$, $L_y=\mathscr{D}-\mathscr{A}$ is the graph Laplacian matrix of $\mathscr{A}$ and $\mathscr{D}=\mathrm{diag}\left(\sum_{j=1}^{n}\mathscr{A}_{ij}\right)$.

Together with (4), (5), (8), (9) in hand, we are now ready to formulate our DETree model with the loss function defined as:

$$\mathscr{L}=\sum_{i=1}^{n}[\mathscr{L}_{\mathscr{E}}((x_i,y_i);\theta\mathscr{E})+\beta\mathscr{L}_{\mathscr{E}\mathscr{R}}((x_i,y_i);\theta\mathscr{E})]+\gamma\sum_{i=1}^{C}\times\sum_{j=1}^{K}\mathscr{L}_{\mathscr{O}}\left(e_{i,j};W,b\right)+\delta\mathscr{L}_{\mathscr{O}\mathscr{R}}(\mathscr{E}W,b)\tag{10}$$

This loss function (10) is derivable with respect to θ, $\mathscr{E}$, W and b. The whole model can be trained in an end-to-end manner. Once the model is trained, the latent space will demonstrate a clear alignment with the different stages of AD progression. The clinical group embeddings will be organized in a meaningful sequence within the latent space, and the distances between individual embeddings will accurately capture the relatedness of disease states among corresponding individuals throughout the entire course of the disease. To visually represent the intricate relationships among individuals and offer a clear depiction of each individual's position throughout the disease progression, we attempt to create a structured tree-like representation based on these learned embeddings. In this tree structure, individuals that share similar disease states are connected by edges, providing a visual reflection of the disease's developmental process. To achieve this, we initiated the process by computing pairwise distances between individual embeddings, resulting in an $n\times n$ distance matrix, where n is the number of embeddings. Subsequently, upon the obtained distance matrix we generated a minimum spanning tree using Kruskal's algorithm. Within the tree structure, the connecting individuals have shortest latent distances and share similar disease status. This approach effectively provides a visual and structural representation of the disease progression, facilitating a deeper understanding of the relationships among individuals at different stages of the disease and their positions within the progression. We named it Disease Embedding Tree (DETree).

For a new patient x, the DETree can provide two sets of predictions. Firstly, we can obtain the probabilities of assigning the new patient x to each of the given clinical groups using (3). Based on these probabilities, we can make the best prediction regarding the clinical group that patient x belongs to. Secondly, DETree enables us to determine the location of the individual patient within the learned tree using the function $h(x,\theta)$. This location reflects the specific stage in the progression of AD where the patient is situated. By utilizing these predictions, we can gain valuable insights into both the patient's clinical group assignment and their disease progression stage.

3. Results

3.1. Experimental Setting

3.1.1. Data Setting

In this work, we used 266 subjects (60 NC, 34 SMC, 51 EMCI, 62 LMCI, 59 AD) in our experiment. Based on Section 2.1 each subject has four functional matrices and we obtained 1064 data samples in total. In our experiments, the training, validation, and testing datasets were split according to subjects, that is, four matrices of the same subject will be divided into the same dataset. As the functional matrix is symmetric, to reduce the redundant data, we used the vectorized upper triangle of each matrix as input features.

3.1.2. Model setting

In this work, the non-linear function $h(x,\theta)$ was implemented by 6-layer fully connected network. The dimensions of the fully connected network are 1024–512–256–64–16-k, where k is the dimension of the latent space (Section 2.3). We tested $k=5$, 10, 15, 20, 25. We showed the results of $k=25$ which gives the best classification performance in Section 3.2 and Section 3.3, and compared the results of $k=5$, 10, 15, 20 and 25 in Section 3.4. Activation function Relu and Batchnorm were used at each layer. C = 5 is the number of classes (NC/SMC/EMCI/LMCI/AD). For each class, we initialized one learnable group embedding ($K=1$). We conducted hyper-parameter tuning for α, β, γ, and δ by searching a grid of powers of 10 within the range of 10^{-4} to 10^{1}. The selection of the best hyper-parameter values was based on the performance of models using the training and validation datasets, resulting in $\alpha=1.0$, $\beta=0.001$, $\gamma=1.0$ and $\delta=1.0$. The entire model was trained in an end-to-end manner. Adam optimizer was used to train the whole model with standard learning rate 0.001, weight decay 0.01, and momentum rates (0.9, 0.999).

3.2. Classification performance

In this section, we showed the classification performance of the proposed DETree. For fair comparisons, we used two strategies to compare the proposed method with other widely used methods. Firstly, we repeated experiments 5 times with random seeds to compare the results with other four broadly used machine learning methods including support vector machine (SVM), k-nearest neighbors (KNN), logistic regression and random forest. We conducted grid searching using training and validation datasets to select the best model parameters for the four models. As a result, the SVM with linear kernel, regularization strength = 0.8, the KNN with $k=9$, the logistic regression with lbfgs solver, l2 regularization, regularization strength = 1.2, and random forest with 100 trees obtained the optimal performance. The classification performance was measured by F_1 scores: $F_1=2\times\frac{precision\times recall}{precision+recall}$ and accuracy (Acc). The results are showed in Table 1. We can see that the F_1 score of DETree model is over 0.75 which is more than 10% higher than the second-best results. And for some classes it can reach 0.80, which is outstanding in multi-class classification of AD and significantly outperforms the other four methods.

Secondly, we compared the multi-class classification performance with latest deep learning methods on AD and reported the results in Table 2. As shown in Table 2, [20] obtains a very high F_1 score for AD group, however the F_1 scores for other groups are considerably lower. Although the total accuracy of [24] is slightly higher than our results (0.780 >0.778), it is important to note that they only considered three classes, whereas our approach encompasses five classes in this study. In comparison with these methods, our proposed approach not only achieves a high overall accuracy but also maintains high accuracies for each class without significant disparities.

3.3. The learned disease embedding tree

In addition to its outstanding classification performance, our DETree's most significant contribution lies in the introduction of a novel ordered embedding method to direct shape the latent space (*embedding space*). Through this method, the latent space is effectively aligned with the AD progression trajectory and a tree structure is learned to model the entire spectrum of AD progression. To evaluate the effectiveness of this method, we analyzed the results from two different perspectives.

Table 1
Classification Performance of DETree and Four Traditional Machine Learning Methods. The orange shade and blue shade highlight the best and the second-best results, respectively.

Method	F1						Acc (All)
	All	AD	LMCI	EMCI	SMC	CN	
SVM	0.658±0.05	0.684±0.06	0.641±0.05	0.630±0.05	0.659±0.05	0.675±0.04	0.672±0.03
KNN	0.526±0.05	0.427±0.09	0.576±0.05	0.546±0.07	0.547±0.08	0.535±0.04	0.532±0.04
Logistic Regression	0.642±0.03	0.651±0.06	0.634±0.04	0.627±0.05	0.649±0.05	0.651±0.03	0.649±0.02
Random Forest	0.422±0.03	0.377±0.05	0.516±0.07	0.426±0.04	0.436±0.06	0.354±0.04	0.428±0.03
DETree (segmented)	**0.777**±0.01	**0.785**±0.03	**0.762**±0.03	**0.801**±0.06	**0.765**±0.02	**0.773**±0.03	**0.778**±0.02

Table 2
Classification Performance of DETree and Other Deep Learning Methods. cMCI/pMCI: MCI patient who converted to AD within 36 months; sMCI: MCI patients who didn't convert to AD within 36 months. EMCI/LMCI: early/late MCI. The orange shade and blue shade highlight the best and the second-best results, respectively.[7,13,20–26,26,27,28].

Study	Modality	Participants (n)	Method	Task / Performance
Amorosoa et al. (2018) [20]	Predefined Features	60AD/60NC/ 60cMCI/60MCI	Deep Random Forest	AD vs. cMCI vs. sMCI vs. NC F_1: 0.805/0.518/0.305/0.525
Zhou et al. (2018) [21]	MRI, PET, SNP	190AD/226NC/ 157pMCI/205sMCI	Multi-modal Fusion	AD vs. pMCI vs. sMCI vs. NC *Acc*: 0.574/0.622/0.342/0.625
Brand et al. (2019) [22]	MRI, SNP	412(AD+MCI+NC)	Joint Regression-Classification	AD vs. MCI vs. NC F_1: 0.566/0.513/0.683
Lei et al. (2020) [13]	MRI	192AD/402MCI/ 220NC	Multiple Templates, Adaptive Feature Selection	AD vs. MCI vs. NC *Acc*: 0.775
Wang et al. (2020) [7]	rs-fMRI	253NC/45EMCI/ 88LMCI	Deep Autoencoder	LMCI vs. EMCI vs. NC *Acc*: 0.730
Puente-Castro et al. (2020) [29]	MRI	297AD/921MCI/ 525NC	Multi-Plane Features, Transfer Learning	AD vs. MCI vs. NC F_1: 0.433/0.760/0.616
Liu et al. (2021) [30]	MRI	90AD/136MCI/ 266NC	Depthwise Separable Convolution (kernel size: 3 × 3; maximum pooling size: 2 × 2)	AD vs. MCI vs. NC *Acc*: 0.780
Xu et al. (2021) [31]	MRI	85AD/244MCI/ 133NC	Tresnet, Selective Kernel module (kernel size: 3 × 3 and 1 × 1)	AD vs. MCI vs. NC *Acc*: 0.632
Lin et al. (2021) [32]	MRI, PET, CSF, genetic data	105AD/441MCI/ 200NC	Linear Discriminant Analysis, PCA, Multimodal Fusion	AD vs. MCI vs. NC *Acc*: 0.667
Lin et al. (2021) [32]	MRI, PET, CSF, genetic data	105AD/200NC/ 110pMCI/208sMCI	Linear Discriminant Analysis, PCA, Multimodal Fusion	AD vs. pMCI vs. sMCI vs. NC *Acc*: 0.573
Mulyadi et al. (2023) [33]	MRI	1540 (AD+MCI+NC)	Clinically Guided Prototype Learning	AD vs. MCI vs. NC *Acc*: 0.632
Sudharsan et al. (2023) [34]	MRI	80AD/84MCI/80NC	Regularized Extreme Learning Machine	AD vs. MCI vs. NC *Acc*: 0.628
Proposed	rs-fMRI	59AD/62LMCI/ 51EMCL/34SMC/ 60NC	Disease Embedding, DETree	AD vs. LMCI vs. EMCI vs. SMC vs. NC F_1: 0.785/0.762/0.801/0.765/ 0.773 *Acc*: 0.778

Firstly, we conducted analyses from a group-level perspective. The well-trained model mapped the samples at different disease stages into the embedding space. We attempted to evaluate the alignment between embedding space with the AD progression by visualizing the feature distribution of different clinical groups in the embedding space. Since the learned DETree is in a high dimensional embedding space ($k = 5, 10, 15, 20, 25$), we adopted Principal Component Analysis (PCA) to project high-dimensional features into two-dimensional space. PCA is a dimensionality reduction technique widely used in data analysis and machine learning. Its primary objective is to simplify complex datasets by transforming them into a lower-dimensional form while preserving the most critical information. The results are shown in Fig. 2(a), where five subfigures (a1 to a5) correspond to the results of five runs of our experiments. From the visualization, it is evident that the feature distribution corresponding to the five different clinical stages exhibit a clear order (*highlighted by dashed arrows*), ranging from NC (*green*) to SMC (*yellow*), EMCI (*orange*), LMCI (*pink*), and eventually ends with AD (*red*), which is consistent with the AD progression trajectory. This demonstrates that the embedding space aligns well with the AD progression. This result indicates that the proposed ordered embedding method is effective in capturing the progression of AD in the embedding space.

Secondly, we conducted analyses at the individual level. For each subject, the well-trained model mapped the input individual feature to the embedding space, obtaining the corresponding embedding vector – $h(x, \theta)$. To analyze the relationship between subjects, we employed Kruskal's algorithm to create a minimum spanning tree over the embedding vectors (*DETree*). The results are shown in Fig. 2(b). In this tree structure, each small bubble represents a single subject, with its color indicating the clinical group to which the subject belongs. Each edge in the tree structure indicates higher similarity between the connected nodes. The five larger bubbles represent the learned group embeddings. From the results we can see that, the learned DETree structure precisely displays a trajectory of AD progression. It starts with the NC, goes through SMC, EMCI, LMCI and eventually ends with AD. The DETree captures the underlying progression pattern, demonstrating its ability to represent the continuous spectrum of AD progression in the embedding space. Moreover, a significant advantage of DETree over traditional classification methods is its ability to represent not only clinical group information but also the specific states of individuals along the entire disease development process. To further illustrate this, we mapped the Mini-Mental State Examination (MMSE) score [33] and the Alzheimer's Disease Assessment Scale - Cognitive subscale (ADAS--cog) score [34], two commonly used tools for assessing dementia, to DETree and presented the results in Fig. 2(c) and (d), respectively. It is evident from the two score trees that the cognitive impairment severity shows an increasing trend from normal control (NC) to Alzheimer's disease (AD) in both score trees, in alignment with the disease progression. Furthermore, considerable variabilities exist in the impairment severity among individuals within the same clinical group. DETree's such capability to preserve substantial individuality in AD progression sets DETree apart as a significant advantage in disease modeling and enhances its potential utility in clinical applications and personalized medicine.

In this work, we used functional connectivity to learn the DETree. To further explore which functional connections contribute most to the learned tree structure, we sorted them with Laplacian score (LS) [35]. LS is a robust feature selection technique widely employed in machine learning and data analysis. It proves particularly useful when dealing with high-dimensional datasets and tasks that require a deep understanding of underlying data structures. LS is grounded in Laplacian Eigenmaps and Locality Preserving Projection, focusing on the

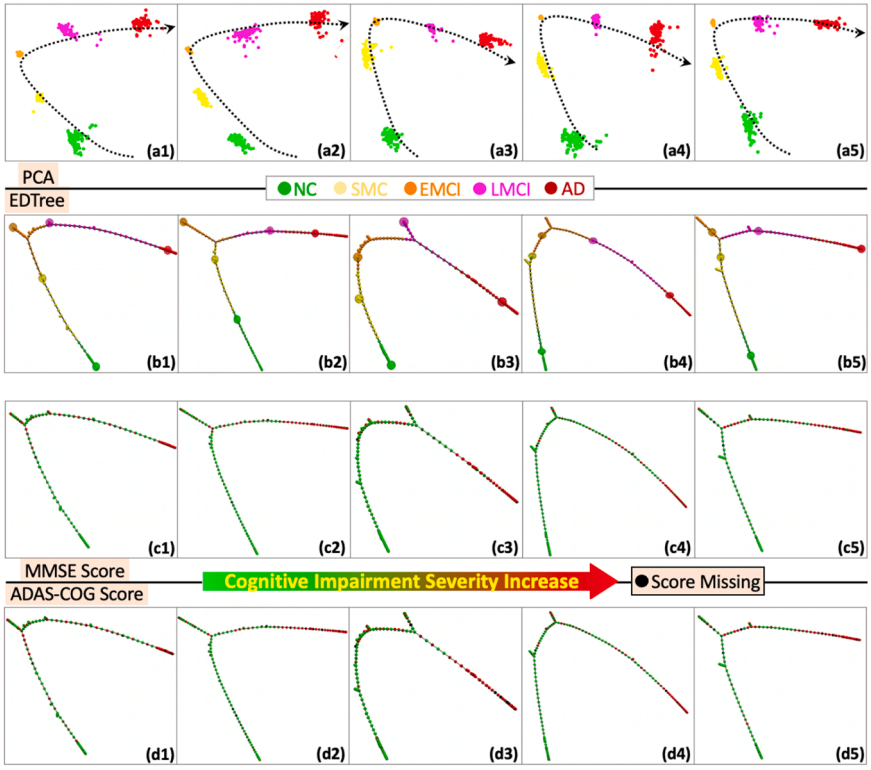


Fig. 2. (a) Visualization of the feature distribution of different clinical groups in the embedding space. (b): DETree learned from multiple clinical groups, including NC, SMC, EMCI, LMCI, and AD. Each small bubble in the tree represents a single subject color-coded according to their clinical group. Each edge indicates higher similarity between connected nodes. The five larger bubbles represent the group level embeddings. (c): the Mini-Mental State Examination (MMSE) score [33] mapped to DETree. (d): the Alzheimer's Disease Assessment Scale - Cognitive subscale (ADAS-cog) score [34] mapped to DETree. The small bubbles in (c) and (d) correspond to the small bubbles in (b) at the same location.

evaluation of features based on their locality preserving power. The concept behind LS is straightforward: data points that belong to the same class are typically close to one another. LS uses the nearest neighbor graph to obtain the local structure of the data and obtains the LS value of each feature. Features that respect this graph structure will obtain higher LS values and will be selected. Fig. 3 shows the top 5, 10 and 15 connectivity selected by LS that have the most contributions during the learning of DETree structure. In each subfigure, the first row shows the brain regions involved in the connectivity. The second row shows the connectivity, and the corresponding regions are represented by bubbles with the same color. Most of the regions in Fig. 3 are reported in previous studies for the close relationship to AD, such as the regions in frontal lobe and temporal lobe [7–9].

3.4. Ablation Study

In our DETree model, the hyper-parameter that has the most significant influence on the DETree structure is k, which represents the dimension of the embedding space. We conducted experiments with different values of k, specifically $k = 5$, 10, 15, 20, and 25, and analyzed the results from three perspectives: the classification performance, feature distribution in the latent space, and the learned DETree structure.

Firstly, we evaluated the influence of k on the classification performance. To augment the training dataset and improve model training, the fMRI signals of each individual in the training and validation datasets were divided into four non-overlapping segments, effectively quadrupling the dataset size. In order to assess whether the segmented methods impact the performance estimation on the testing dataset, we conducted experiments in both segmented and unsegmented settings on the testing datasets and compared the results. These comparisons are presented in Fig. 4. As depicted in Fig. 4(A) and (B), increasing the value of k leads to enhanced classification performance in both segmented and unsegmented settings on the testing dataset. This improvement is attributed to the fact that low-dimensional embedding spaces may not fully capture the intricate relationships within the brain network data, while higher-dimensional spaces provide a more comprehensive representation of these relationships, resulting in improved classification performance. To highlight the distinctions between the segmented and unsegmented testing dataset settings, we calculated the differences in F_1 and Acc measures and presented the results in Fig. 4(C). The findings indicate that segmented and unsegmented settings exhibit similar classification performance, with variations within the range of $[-0.05, 0.06]$.

Secondly, we assessed the impact of k on the feature distribution in the embedding space, particularly its alignment with the AD progression process. The results are displayed in Fig. 5. Notably, as we varied the value of k from 5 to 25, the feature distribution of the five clinical groups consistently maintained the order from NC to SMC, EMCI, LMCI, and eventually AD. This alignment of embedding space with the AD progression trajectory demonstrates that the choice of k does not affect the feature distribution in the embedding space. This indicates that our proposed ordered embedding method exhibits excellent robustness across different dimensions of embedding space. The flexibility to choose different dimensions of the embedding space enhances the potential of the proposed model in adapting to diverse data characteristics and complexities, making it applicable to a wide range of applications.

Thirdly, we explored the impact of k on the learned DETree structure and presented the results in Fig. 6. From Fig. 6 we can see that if k is too small (*corresponding to lower dimensional embedding space*), the distances among different embeddings tend to be small (*highlighted by blue circle*). As a result, the corresponding embeddings exhibit high similarities. This leads to insufficient dissimilarity between embeddings, which can limit the capability of DETree in representing multiple clinical stages in AD progression and compromise its prediction performance when estimating new samples. These findings are consistent with the results in Fig. 4, where lower-dimensional spaces exhibited inferior classification performance compared to higher-dimensional embedding spaces.

3.5. Reproducibility and Generalizability

To evaluate the reproducibility of our proposed model across different datasets and its generalizability to different tasks, we collected the most recent released subset of ADNI, including 145 subjects (40 NC: 22 females, 18 males, 73.64±7.03 yrs.; 28 progressive MCI (pMCI): 13 females, 15 males, 72.44±7.43 yrs.; 42 stable MCI (sMCI): 18 females, 14 males, 71.32±6.94 yrs.; and 35 AD: 15 females, 20 males, 72.93±8.56 yrs.). The pMCI group includes patients who progressed to AD within 36 months [20,21], while sMCI consists of individuals who did not progress. We divided the 145 subjects (dataset-2) into training, validation, and testing datasets. For the training and validation datasets, we employed the same four-segment approach as in dataset-1. However, in the testing dataset, we used an unsegmented setting. Using a similar experimental setup as in dataset-1, we conducted a series of experiments, repeating each experiment 5 times with different random seeds. Our objective was to evaluate the performance of the proposed DETree model with varying embedding dimensions (k) for the new classification task: NC vs. sMCI vs. pMCI vs. AD. We evaluated the model's performance from three perspectives: classification performance (Fig. 7), feature distributions in the latent space (Fig. 8), and the learned DETree structures (Fig. 9).

As shown in the results, increasing the value of k leads to improved classification performance, consistent feature distributions in the latent space, and larger distances among different embeddings within the learned DETree structure. These findings are consistent with the conclusions drawn from dataset-1, indicating that the DETree model exhibits excellent reproducibility and generalizability. However, it's worth noting that in the new classification task, our model achieved a maximum accuracy of 0.708, slightly lower than the 0.777 accuracy obtained in task-1 based on dataset-1. This drop in accuracy can be attributed to the introduction of new categories, sMCI and pMCI, which replaced EMCI and LMCI from task-1. This observation is consistent with the results in Table 2, where tasks related to MCI progression, as discussed in references [20,21,26], typically exhibit slightly lower classification accuracy compared to other works. This discrepancy may be due to the increased challenge of distinguishing between sMCI and pMCI compared to the previous EMCI and LMCI categories. Nevertheless, even in this context, our model consistently outperforms other models in the MCI progression task [20,21,26].

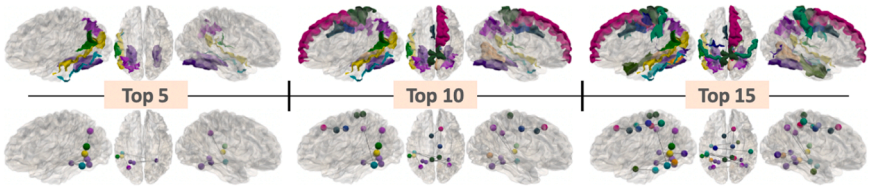
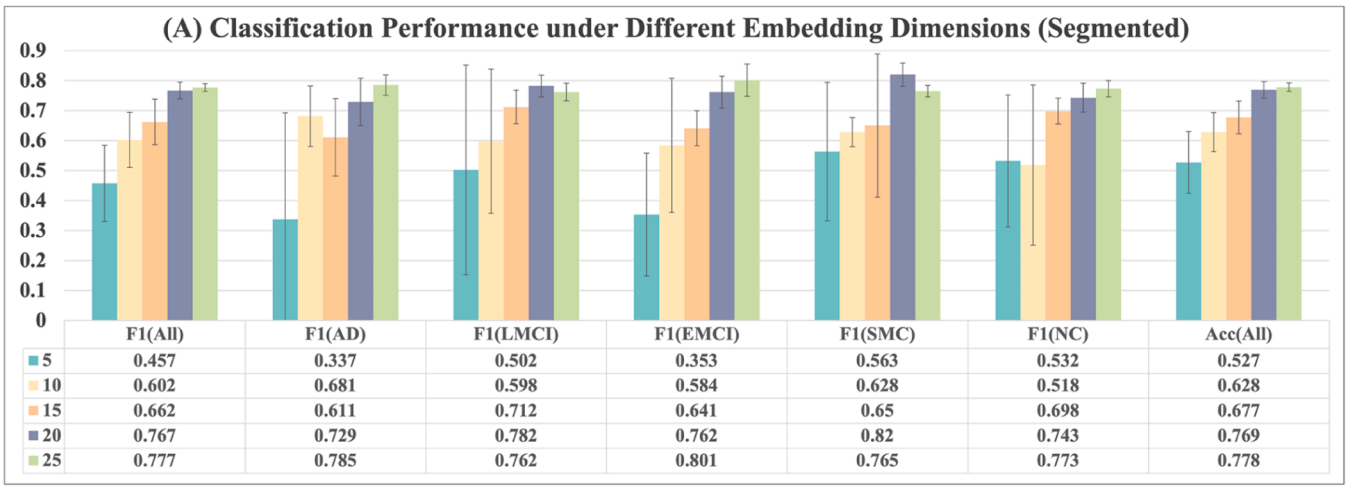
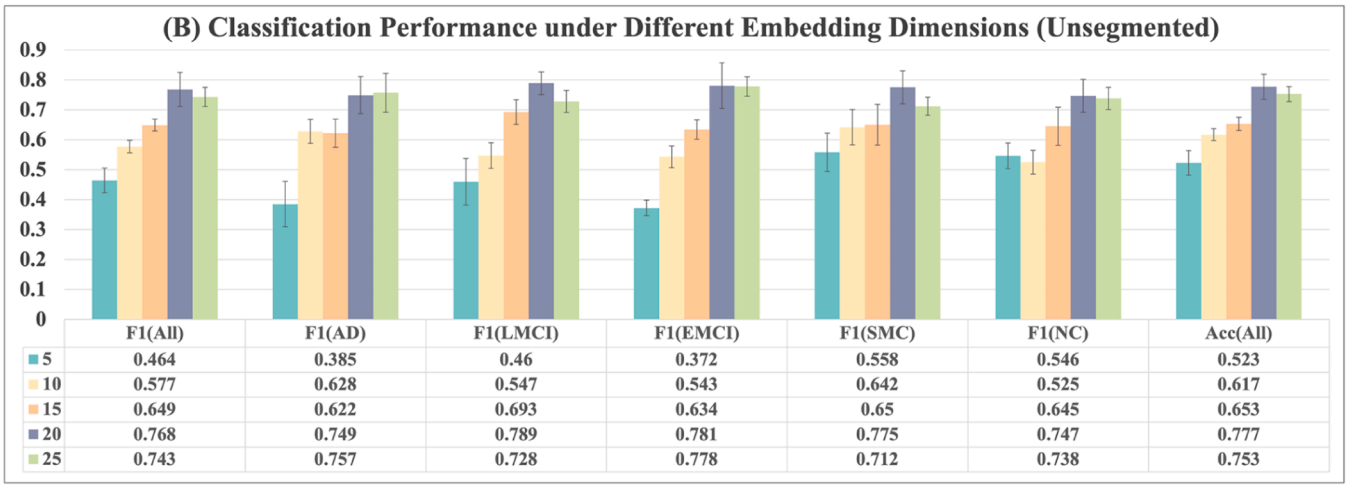


Fig. 3. Top connectivity that contributes most to the learned DETree structure. In each block, the top and bottom rows display the involved brain regions and connectivity, respectively.



	F1(All)	F1(AD)	F1(LMCI)	F1(EMCI)	F1(SMC)	F1(NC)	Acc(All)
5	0.457	0.337	0.502	0.353	0.563	0.532	0.527
10	0.602	0.681	0.598	0.584	0.628	0.518	0.628
15	0.662	0.611	0.712	0.641	0.65	0.698	0.677
20	0.767	0.729	0.782	0.762	0.82	0.743	0.769
25	0.777	0.785	0.762	0.801	0.765	0.773	0.778



	F1(All)	F1(AD)	F1(LMCI)	F1(EMCI)	F1(SMC)	F1(NC)	Acc(All)
5	0.464	0.385	0.46	0.372	0.558	0.546	0.523
10	0.577	0.628	0.547	0.543	0.642	0.525	0.617
15	0.649	0.622	0.693	0.634	0.65	0.645	0.653
20	0.768	0.749	0.789	0.781	0.775	0.747	0.777
25	0.743	0.757	0.728	0.778	0.712	0.738	0.753

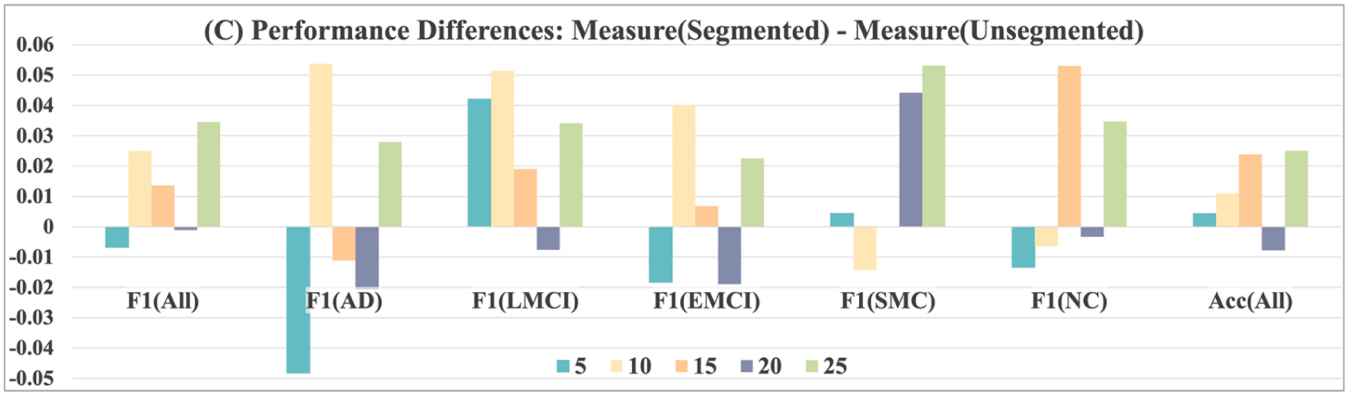


Fig. 4. (A) and (B): Classification performance of various dimensions of the embedding space in two testing dataset settings. (C): The differences in F_1 and Acc measures of two settings.

4. Conclusion and discussion

In this study, we introduced a novel DETree framework that seamlessly integrates individual prediction with AD progression modeling. The learned DETree structure effectively represents the trajectory of AD progression and achieves an impressive prediction performance of over 77.8% for multiple AD-related stages. One of the key strengths of our approach is its ability to not only predict the clinical status of individual patients but also provide valuable information about their specific state within the entire spectrum of AD progression. We summarize the advantages and limitations of our current work and provide some insights for future research.

4.1. Advantages

4.1.1. DETree is a general framework for modeling continuous diseases development

In this work, we only applied DETree to Alzheimer's disease, but it is a versatile framework that can be extended to a wide range of diseases. The proposed DETree framework allows for flexibility in implementing

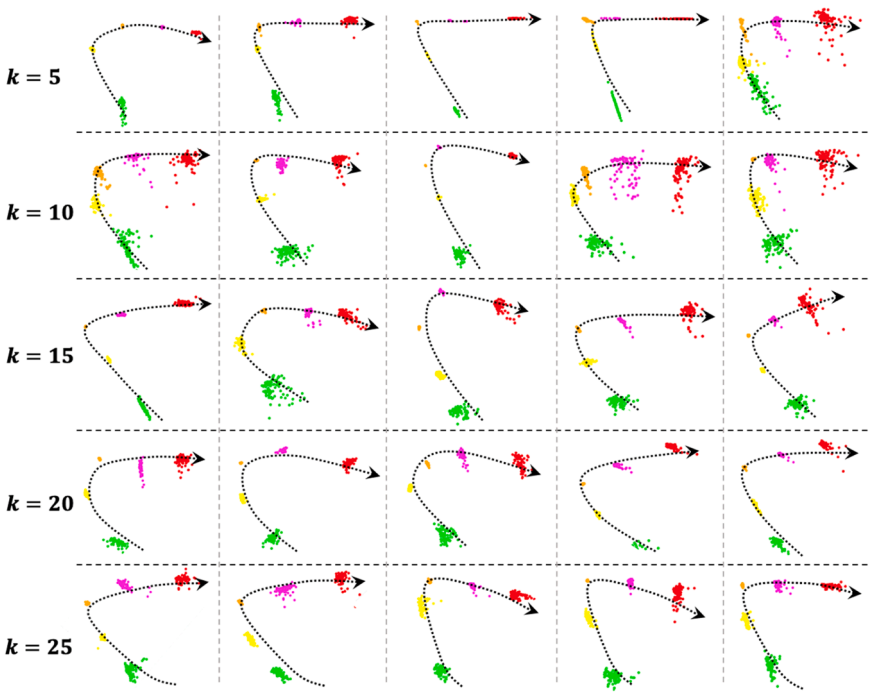


Fig. 5. Visualization of the feature distribution of different clinical groups in the embedding space with varying dimensions: NC – *green*, SMC – *yellow*, EMCI – *orange*, LMCI – *pink*, and AD – *red*.

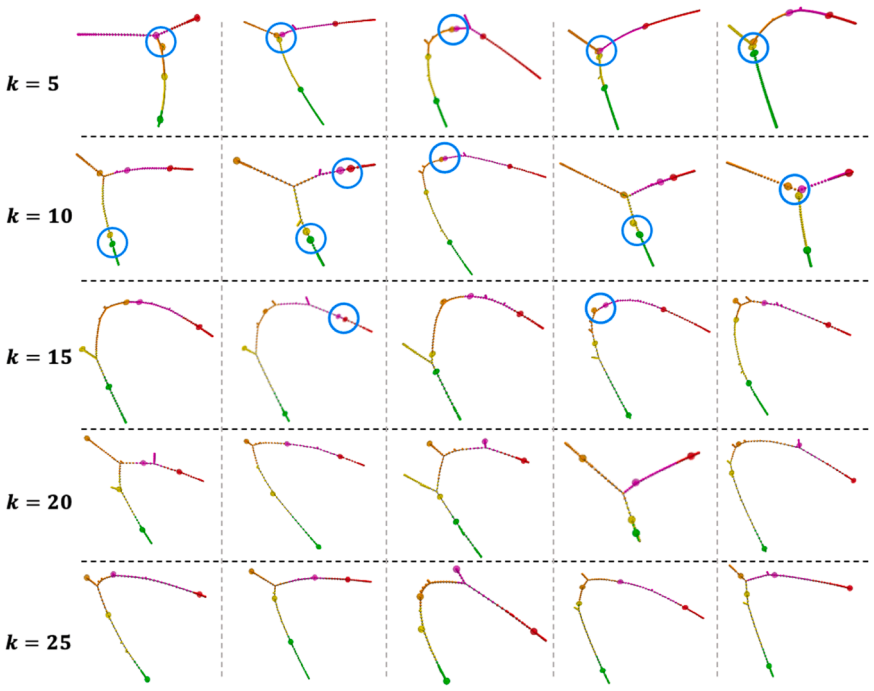
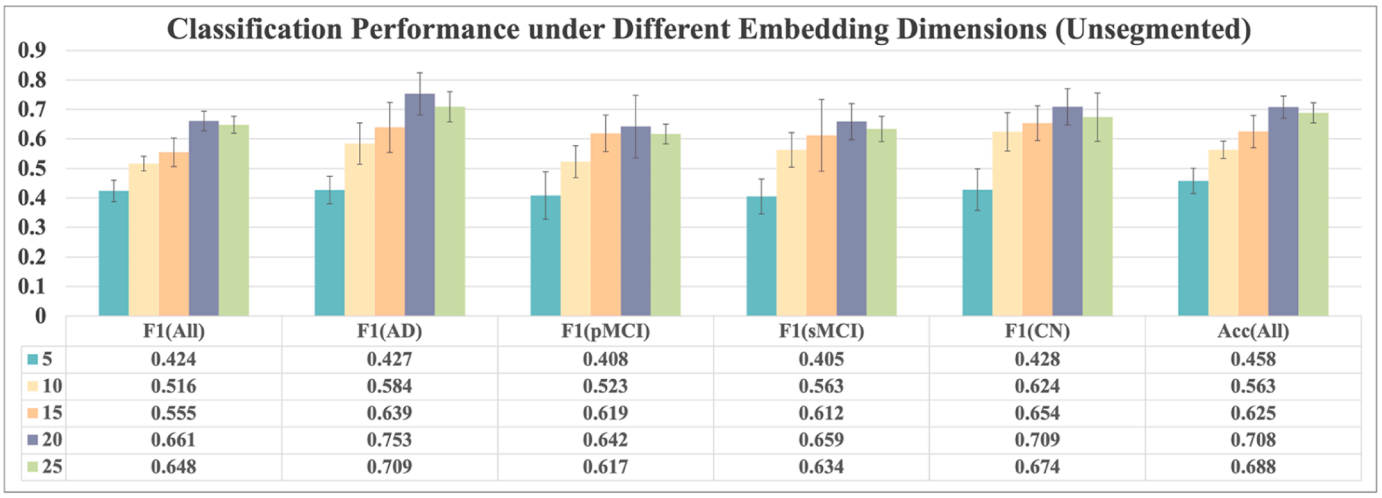


Fig. 6. Different DETree structures learned in the embedding space with varying dimensions. The blue circles are used to highlight the embeddings from different clinical groups with small distance.

non-linear function $h(x,\theta)$, making it adaptable to any disease that exhibits multiple clinical stages during its development. Researchers can choose a suitable model architecture for $h(x,\theta)$ implementation and input relevant features into the model based on the specific disease and tasks at hand. More important, by modifying the affinity matrix $\mathscr{A}$, the prior knowledge about the disease can be easily introduced into the DETree model.

4.1.2. *DETree exhibits versatility beyond the realm of classification tasks*

With minor adjustments to the additional neural network layer in Section 2.4, DETree can seamlessly extend its applicability to regression problems. For instance, by substituting the discrete clinical labels with continuous clinical scores, such as the MMSE score [33], the existing classification framework can be readily transformed into a robust regression-based model. A straightforward approach to achieve this is as follows: We take the feature and score pairs $\{(x_i, s_i)\}_{i=1}^{n}$ as input to train a



	F1(All)	F1(AD)	F1(pMCI)	F1(sMCI)	F1(CN)	Acc(All)
5	0.424	0.427	0.408	0.405	0.428	0.458
10	0.516	0.584	0.523	0.563	0.624	0.563
15	0.555	0.639	0.619	0.612	0.654	0.625
20	0.661	0.753	0.642	0.659	0.709	0.708
25	0.648	0.709	0.617	0.634	0.674	0.688

Fig. 7. Classification performance of various dimensions of the embedding space based on dataset2.

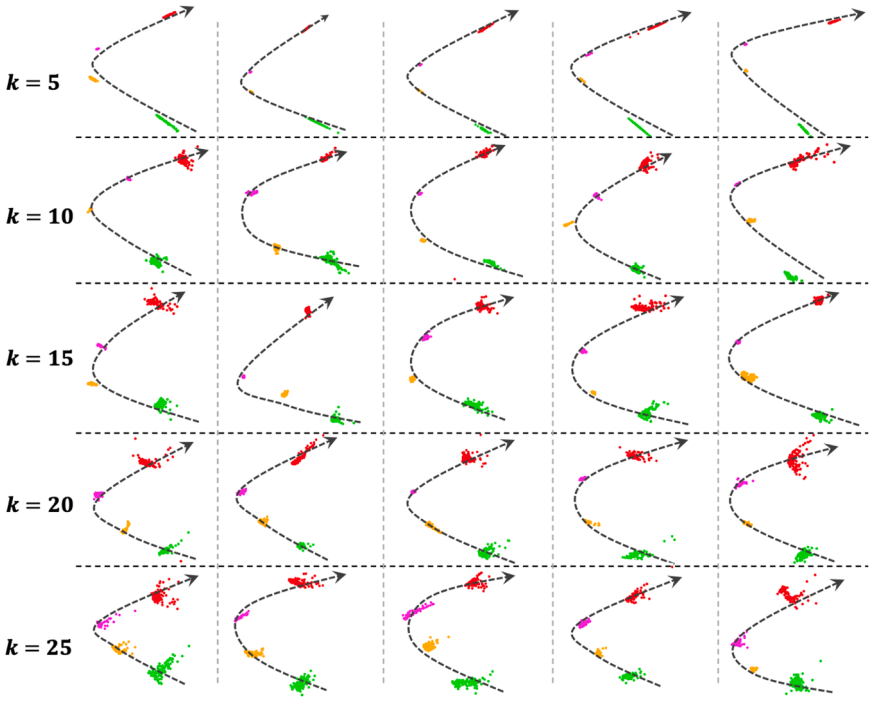


Fig. 8. Visualization of the feature distribution of different clinical groups in the embedding space with varying dimensions based on dataset2.

non-linear function $h(x,\theta): \mathscr{R}^d \rightarrow \mathscr{R}^k$, with learnable model parameter θ. This transforms any given input $x \in \mathscr{R}^d$ to a latent space $\mathscr{R}^k$. Based on the latent feature $h(x,\theta)$, we will integrate two tasks: clinical score prediction and disease progression representation. This means $h(x,\theta)$ is influenced by both tasks, allowing it to capture key information related to both scores and disease progression simultaneously. To implement the prediction task, we can train a submodule $g(x, W_r, b_r)$ to predict the clinical score as follows: $s_i^{'} = g(h(x_i,\theta), W_r, b_r)$, and a regression loss $\mathscr{L}_r = \frac{1}{n}\sum_{i=1}^{n}(s_i - s_i^{'})^2$ can be used to control the regression learning. Regarding the disease progression task, we can utilize a ranking loss function [36]. In contrast to other loss functions like Cross-Entropy Loss or Mean Square Error Loss, which aim to directly predict a label, a value, or a set of values based on an input, Ranking Losses focus on predicting relative distances between inputs, commonly referred to as metric learning. Specifically, let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be the objects that need to be ranked, associated with multi-level ratings denoted as $L = \{l(1), \ldots, l(n)\}$, where each $l(i) \in \{r_1, \ldots, r_k\}$ and signifies the label of x_i [36]. In our work, the rating $l(i)$ is the clinical score of x_i. In the ranking context, when $l(i) > l(j)$, it implies that object x_i should take precedence over x_j in the ranking order. Let $\mathscr{F}$ represent the class of functions, and $f \in \mathscr{F}$ denote a specific ranking function. The objective is to learn the optimal ranking function from training data by minimizing a specific loss function. This loss function is defined based on the objects, their associated labels, and the ranking function itself. Several methodologies have been proposed to facilitate the learning of this optimal ranking function, including pointwise approaches [37,38], pairwise approaches [39,40], and e listwise approaches [41,42]. Moreover, it's worth delving deeper into exploring the integration of disease-related prior knowledge to tailor a ranking function that better aligns with the demands of our specific task. This is an avenue that deserves further investigation in our future work.

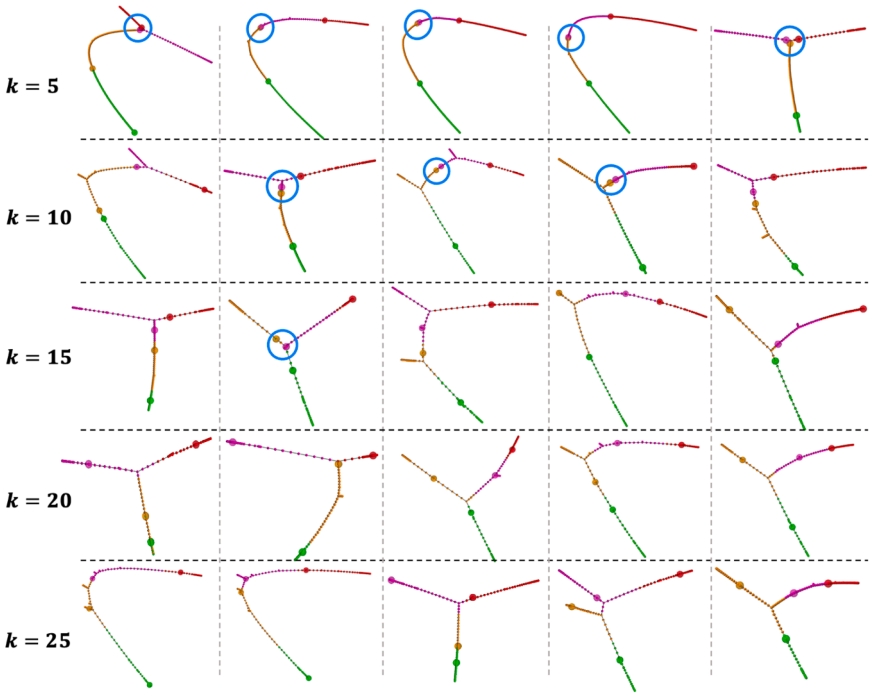


Fig. 9. Different DETree structures learned in the embedding space with varying dimensions based on dataset2.

5. Limitations and future work

The current work has focused on a specific choice of the embedding number – *K*, for each class. In Section 2.3, we provide a general form of group embeddings as $\mathscr{E} = \{e_{ij} \in \mathscr{P}^k | i = 1, 2, \ldots, C; j = 1, 2, \ldots, K\}$ where K is the number of the embeddings in each class and could be equal to or greater than 1. However, for the particular application addressed in this study, K is intentionally set to 1. It's important to note that in many scenarios, the choice of K can be a subject for more in-depth exploration. For example, in the case of specific diseases where individuals in the same disease stage exhibit diverse clinical symptoms, the choice of K becomes crucial. In such cases, the objective is that each of the K learned embeddings will capture the specific features associated with the varying clinical symptoms observed at the same disease stage. It's essential to emphasize that, given the potential subtle differences among patients in the same disease stage, a significant amount of data is required to adequately train the model for each clinical response.

Current work only covers single modality.

Due to the scarcity of multi-modal data, where each subject is required to have multiple modalities of data, such as both structural and functional image data, the sample size drastically decreases. Consequently, we have only focused on utilizing a single modality for this study. Previous studies [43,44] suggest that functional abnormalities may precede structural alterations, making functional data more suitable for a comprehensive modeling of the entire AD progression process. Therefore, we chose to use functional connectivity data in this study. In the future, as we acquire sufficient multi-modal data, extending our current model to accommodate multiple modalities becomes feasible by treating each modality as a distinct view. Specifically, we can tailor a modality-specific model for each modality, mapping the original data into a unified embedding space from its unique perspective. This flexible strategy not only effectively conceals heterogeneities and inconsistencies between modalities, such as variations in numerical values, dimensions, and representations, but also allows us to optimize the model architecture for each modality to capture its specific characteristics most effectively. For instance, recent advancements in Large Language Models (LLMs) have demonstrated impressive achievements in various domains [45–49], including NLP and Computer Vision (CV). We can leverage these pre-trained LLMs on large datasets as feature extractors and fine-tune them to adapt to healthcare data. This enables us to fully exploit the remarkable generalization and feature extraction capabilities of these large models. Through this approach, we can harness cutting-edge models and technologies from the NLP and CV domains to study brain diseases effectively. By integrating NLP and CV advancements into our research on brain diseases, we have the opportunity to gain new insights and breakthroughs that were previously not attainable. This interdisciplinary approach opens up exciting possibilities for the advancement of neuroscience and healthcare research.

Funding

This work was supported by National Institutes of Health (R01AG075582, RF1NS128534).

CRediT authorship contribution statement

Lu Zhang: Conceptualized and designed the method presented in the paper, conducted data collection and experiments, and drafted the initial manuscript. **Li Wang:** Provided expert insights on the mathematical methods discussed in the article and reviewed and synthesized the key findings. **Tianming Liu and Dajiang Zhu:** Reviewed and synthesized the central findings and contributed to enhancing the overall coherence of the manuscript.

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.

Data Availability

The data we used is from the publicly available dataset ADNI, and the link has been provided in the manuscript.

References

[1] W. Henneman, J. Sluimer, J. Barnes, W. Van Der Flier, I. Sluimer, N. Fox, P. Scheltens, H. Vrenken, F. Barkhof, Hippocampal atrophy rates in Alzheimer disease: added value over whole brain volume measures, Neurology 72 (11) (2009) 999–1007.

[2] G. Karas, P. Scheltens, S.A. Rombouts, P.J. Visser, R.A. van Schijndel, N.C. Fox, F. Barkhof, Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease, Neuroimage 23 (2) (2004) 708–716.

[3] S. Li, F. Pu, F. Shi, et al., Regional white matter decreases in Alzheimer's disease using optimized voxel-based morphometry, Acta Radiol. 49 (1) (2008) 84–90.

[4] M.D. Greicius, G. Srivastava, A.L. Reiss, V. Menon, Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional mri, Proc. Natl. Acad. Sci. 101 (13) (2004) 4637–4642.

[5] J. Ashburner, K.J. Friston, Voxel-based morphometry–the methods, Neuroimage 11 (6) (2000) 805–821.

[6] S.M. Smith, M. Jenkinson, Johansen-Berg, et al., Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data, Neuroimage 31 (4) (2006) 1487–1505.

[7] L. Wang, L. Zhang, D. Zhu, Learning latent structure over deep fusion model of mild cognitive impairment. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 1039–1043.

[8] L. Zhang, L. Wang, J. Gao, S.L. Risacher, J. Yan, G. Li, T. Liu, D. Zhu, et al., Deep fusion of brain structure-function in mild cognitive impairment, Med. Image Anal. 72 (2021), 102082.

[9] Zhang, L., Wang, L., Zhu, D.: Jointly analyzing Alzheimer's disease related structure-function using deep cross-model attention network pp. 563–567 (2020).

[10] P.S. Aisen, J. Cummings, C.R. Jack, J.C. Morris, et al., On the path to 2025: understanding the Alzheimer's disease continuum, Alzheimer's Res. Ther. 9 (1) (2017) 1–10.

[11] M. Ten Kate, E. Dicks, P.J. Visser, W.M. van der Flier, C.E. Teunissen, F. Barkhof, P. Scheltens, B.M. Tijms, A.D.N. Initiative, Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline, Brain 141 (12) (2018) 3443–3456.

[12] S.J. Vos, F. Verhey, L. Frölich, J. Kornhuber, et al., Prevalence and prognosis of Alzheimer's disease at the mild cognitive impairment stage, Brain 138 (5) (2015) 1327–1338.

[13] B. Lei, Y. Zhao, Z. Huang, X. Hao, F. Zhou, A. Elazab, J. Qin, H. Lei, Adaptive sparse learning using multi-template for neurodegenerative disease diagnosis, Med. Image Anal. 61 (2020), 101632.

[14] G.B. Frisoni, N.C. Fox, C.R. Jack, P. Scheltens, P.M. Thompson, The clinical use of structural mri in Alzheimer disease, Nat. Rev. Neurol. 6 (2) (2010) 67–77.

[15] A. Mouiha, S. Duchesne, A.D.N. Initiative, et al., Toward a dynamic biomarker model in Alzheimer's disease, J. Alzheimer's Dis. 30 (1) (2012) 91–100.

[16] A.L. Young, N.P. Oxtoby, P. Daga, D.M. Cash, N.C. Fox, S. Ourselin, J.M. Schott, D. C. Alexander, A data-driven model of biomarker changes in sporadic Alzheimer's disease, Brain 137 (9) (2014) 2564–2577.

[17] D. Li, S. Iddi, W.K. Thompson, M.C. Donohue, A.D.N. Initiative, Bayesian latent time joint mixed effect models for multicohort longitudinal data, Stat. Methods Med. Res. 28 (3) (2019) 835–845.

[18] R. Patil, S. Boit, V. Gudivada, J. Nandigam, A survey of text representation and embedding techniques in NLP, IEEE Access (2023).

[19] L. Zhang, L. Zhao, D. Liu, Z. Wu, X. Wang, T. Liu, D. Zhu, Cortex2vector: anatomical embedding of cortical folding patterns, Cereb. Cortex, 33(10) (2023) 5851–5862.

[20] N. Amoroso, D. Diacono, A. Fanizzi, M. La Rocca, A. Monaco, A. Lombardi, C. Guaragnella, R. Bellotti, S. Tangaro, A.D.N. Initiative, et al., Deep learning reveals Alzheimer's disease onset in mci subjects: results from an international challenge, J. Neurosci. Methods 302 (2018) 3–9.

[21] T. Zhou, K.H. Thung, X. Zhu, D. Shen, Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis, Hum. brain Mapp. 40 (3) (2019) 1001–1016.

[22] L. Brand, K. Nichols, H. Wang, L. Shen, H. Huang, Joint multi-modal longitudinal regression and classification for Alzheimer's disease prediction, IEEE Trans. Med. Imaging 39 (6) (2019) 1845–1855.

[23] 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.

[24] J. Liu, M. Li, Y. Luo, S. Yang, W. Li, Y. Bi, Alzheimer's disease detection using depthwise separable convolutional neural networks, Comput. Methods Prog. Biomed. 203 (2021), 106032.

[25] Z. Xu, H. Deng, J. Liu, Y. Yang, Diagnosis of Alzheimer's disease Based on the Modified Tresnet, Electronics 10 (16) (2021) 1908.

[26] W. Lin, Q. Gao, M. Du, W. Chen, T. Tong, Multiclass diagnosis of stages of Alzheimer's disease using linear discriminant analysis scoring for multimodal data, Comput. Biol. Med. 134 (2021), 104478.

[27] A.W. Mulyadi, W. Jung, K. Oh, J.S. Yoon, K.H. Lee, H.I. Suk, Estimating explainable Alzheimer's disease likelihood map via clinically-guided prototype learning, NeuroImage 273 (2023), 120073.

[28] M. Sudharsan, G. Thailambal, Alzheimer's disease prediction using machine learning techniques and principal component analysis (PCA), Mater. Today.: Proc. (2021).

[29] Zhang, L., Wang, L. and Zhu, D.: Recovering brain structural connectivity from functional connectivity via multi-gcn based generative adversarial network. In Medical Image Computing and Computer Assisted Intervention, Proceedings, Part VII 23, pp. 53–61 (2020).

[30] L. Zhang, L. Wang, D. Zhu, Alzheimer's Disease Neuroimaging Initiative, Predicting brain structural network using functional connectivity, Med. Image Anal. 79 (2022), 102463.

[31] X. Zhang, L. Guo, X. Li, T. Zhang, D. Zhu, K. Li, H. Chen, J. Lv, C. Jin, Q. Zhao, et al., Characterization of task-free and task-performance brain states via functional connectome patterns, Med. Image Anal. 17 (8) (2013) 1106–1122.

[32] Zhang, L., Zaman, A., Wang, L., Yan, J. and Zhu, D., 2019. A cascaded multi-modality analysis in mild cognitive impairment. In Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI, Proceedings 10, pp. 557–565 (2019).

[33] T.N. Tombaugh, N.J. McIntyre, The mini-mental state examination: a comprehensive review, J. Am. Geriatr. Soc. 40 (9) (1992) 922–935.

[34] J.K. Kueper, M. Speechley, M. Montero-Odasso, The Alzheimer's disease assessment scale–cognitive subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. a narrative review, J. Alzheimer'S. Dis. 63 (2) (2018) 423–444.

[35] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, Adv. Neural Inf. Process. Syst. (18) (2005).

[36] W. Chen, T.Y. Liu, Y. Lan, Z.M. Ma, H. Li, Ranking measures and loss functions in learning to rank, Adv. Neural Inf. Process. Syst. (2009) 22.

[37] D. Cossock, T. Zhang, Statistical analysis of Bayes optimal subset ranking, IEEE Trans. Inf. Theory 54 (11) (2008) 5140–5154.

[38] P. Li, Q. Wu, C. Burges, Mcrank: learning to rank using multiple classification and gradient boosting, Adv. Neural Inf. Process. Syst. (2007) 20.

[39] R. Herbrich, Large margin rank boundaries for ordinal regression, Adv. Large Margin Classif. (2000) 115–132.

[40] Y. Freund, R. Iyer, R.E. Schapire, Y. Singer, An efficient boosting algorithm for combining preferences, J. Mach. Learn. Res. 4 (Nov) (2003) 933–969.

[41] Z. Cao, T. Qin, T.Y. Liu, M.F. Tsai, H. Li, Learning to rank: from pairwise approach to listwise approach (June), Proc. 24th Int. Conf. Mach. Learn. (2007) 129–136.

[42] F. Xia, T.Y. Liu, J. Wang, W. Zhang, H. Li, Listwise approach to learning to rank: theory and algorithm (July), Proc. 25th Int. Conf. Mach. Learn. (2008) 1192–1199.

[43] L. Wang, L. Zhang, D. Zhu, Accessing latent connectome of mild cognitive impairment via discriminant structure learning. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, 2019, pp. 164–168.

[44] F. Palesi, G. Castellazzi, L. Casiraghi, E. Sinforiani, P. Vitali, C.A. Gandini Wheeler-Kingshott, E. D'Angelo, Exploring patterns of alteration in Alzheimer's disease brain networks: a combined structural and functional connectomics analysis, Front. Neurosci. 10 (2016) 380.

[45] W.X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, A survey of large language models, arXiv Prepr. arXiv 2303 (2023) 18223.

[46] L. Zhao, L. Zhang, Z. Wu, Y. Chen, H. Dai, X. Yu, Z. Liu, T. Zhang, X. Hu, X. Jiang, X. Li, When brain-inspired ai meets agi, Meta-Radiol. (2023), 100005.

[47] X. Li, L. Zhang, Z. Wu, Z. Liu, L. Zhao, Y. Yuan, J. Liu, G. Li, D. Zhu, P. Yan, Q. Li, Artificial general intelligence for medical imaging, arXiv Prepr. arXiv (2023), 2306.05480.

[48] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, Learning transferable visual models from natural language supervision (July), Int. Conf. Mach. Learn. (2021) 8748–8763.

[49] Z. Xiao, Y. Chen, L. Zhang, J. Yao, Z. Wu, X. Yu, Y. Pan, L. Zhao, C. Ma, X. Liu, W. Liu, Instruction-vit: multi-modal prompts for instruction learning in vit, arXiv Prepr. arXiv (2023), 2305.00201.