NeuroImage 297 (2024) 120737

Contents lists available at ScienceDirect

NeuroImage

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

Disentangling brain atrophy heterogeneity in Alzheimer's disease: A deep self-supervised approach with interpretable latent space

Sohyun Kang[a], Sung-Woo Kim[b,c,d], Joon-Kyung Seong[a,b,e,*], for the Alzheimer's Disease Neuroimaging Initiative

[a] *Department of Artificial Intelligence, College of Informatics, Korea University, Seoul, 02841, South Korea*
[b] *School of Biomedical Engineering, College of Health Science, Korea University, Seoul, 02841, South Korea*
[c] *Department of Neurology, Wonju Severance Christian Hospital, Yonsei University Wonju College of Medicine, Wonju, 26426, South Korea*
[d] *Research Institute of Metabolism and Inflammation, Yonsei University Wonju College of Medicine, Wonju, 26426, South Korea*
[e] *Interdisciplinary Program in Precision Public Health, College of Health Science, Korea University, Seoul, 02841, South Korea*

ARTICLE INFO

Keywords:
Alzheimer's disease
Atrophy
Heterogeneity
Longitudinal trajectory
Deep latent representation
Self-supervised learning

ABSTRACT

Alzheimer's disease (AD) is heterogeneous, but existing methods for capturing this heterogeneity through dimensionality reduction and unsupervised clustering have limitations when it comes to extracting intricate atrophy patterns. In this study, we propose a deep learning based self-supervised framework that characterizes complex atrophy features using latent space representation. It integrates feature engineering, classification, and clustering to synergistically disentangle heterogeneity in Alzheimer's disease. Through this representation learning, we trained a clustered latent space with distinct atrophy patterns and clinical characteristics in AD, and replicated the findings in prodromal Alzheimer's disease. Moreover, we discovered that these clusters are not solely attributed to subtypes but also reflect disease progression in the latent space, representing the core dimensions of heterogeneity, namely progression and subtypes. Furthermore, longitudinal latent space analysis revealed two distinct disease progression pathways: medial temporal and parietotemporal pathways. The proposed approach enables effective latent representations that can be integrated with individual-level cognitive profiles, thereby facilitating a comprehensive understanding of AD heterogeneity.

1. Introduction

Alzheimer's disease (AD) is a complex disorder with various clinical and pathological manifestations (Scheltens et al., 2017; Armstrong et al., 2000; Murray et al., 2011; Ritchie and Touchon, 1992; Lam et al., 2013; Whitwell et al., 2012) that render its diagnosis and treatment challenging. Over the last decade, to explain the diversity of these manifestations or heterogeneity, researchers have categorized the disease into various subtypes. These classifications have often been based on factors such as atrophy patterns (Noh et al., 2014; Park et al., 2017; Ferreira et al., 2017) or the presence of neurofibrillary tangles (Lowe et al., 2018; Whitwell et al., 2018). Multiple studies have identified four subtypes, namely typical, limbic predominant, hippocampal predominant, and minimal atrophy (Byun et al., 2015; Ferreira et al., 2020; Poulakis et al., 2018). However, it remains uncertain whether these subtypes represent systematic variation or merely reflect differences due to the different stages of disease progression. Poulakis et al. (2022) demonstrated that the minimal atrophy, limbic predominant, and typical atrophy subgroups occur at different stages of the same disease trajectory, supporting that the disease stage is crucial in understanding heterogeneity (Young et al., 2018; Vogel et al., 2021; Young et al., 2022; Mohanty et al., 2023; Ossenkoppele et al., 2019; Habes et al., 2020). Severity and typicality have been identified as two critical dimensions that can contribute to this variation (Ferreira et al., 2020). Therefore, it is essential to disentangle disease progression and the subtypes to unravel the complexities of disease heterogeneity, which will aid in facilitating diagnosis and personalized treatments.

Previous studies on heterogeneity of AD have commonly employed clustering, an unsupervised technique, to group individuals based on their similarities (Noh et al., 2014; Park et al., 2017; Poulakis et al., 2018; Kim et al., 2019). Recently, researchers have been increasingly focusing on dimensionality reduction techniques to extract meaningful

---

[*] Corresponding author at: Department of Biomedical Engineering and Department of Artificial Intelligence, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, South Korea.
E-mail address: jkseong@korea.ac.kr (J.-K. Seong).

https://doi.org/10.1016/j.neuroimage.2024.120737
Received 6 March 2024; Received in revised form 3 July 2024; Accepted 11 July 2024

features from the data to uncover heterogeneity (Zhang et al., 2016; Ten Kate et al., 2018; Kwak et al., 2021; Wen et al., 2022). These techniques operate in a latent space, a compressed and low-dimensional representation that retains essential information, while discarding irrelevant details from the original data (Hinton and Salakhutdinov, 2006; Bishop and Nasrabadi, 2006; Murphy, 2012; Bengio et al., 2013; Goodfellow et al., 2016). Particularly with deep learning methods, many studies have demonstrated the power of latent representations in identifying significant disease features from neuroimaging dataset (Yang et al., 2021; Yang et al., 2024; Martí-Juan et al., 2023; Yu et al., 2019; Li et al., 2021). By establishing an easily interpretable and separable latent space, these methods offer a valuable approach to disentangle disease heterogeneity along arbitrary axes. Cho et al. (2020) attempted to disentangle the underlying axes of heterogeneity in a data-driven manner, while its use of a linear dimensionality reduction method limited its ability to capture complex nonlinear relationships between the original data and the latent features.

Self-supervised learning has emerged as a promising technique to capture complex features and generate biologically interpretable latent representations (Wang et al., 2021; Ericsson et al., 2022; Jing and Tian, 2020; Leeb et al., 2022). However, its application in unraveling disease heterogeneity has been limited because of its reliance on large unlabeled datasets, such as web-scale corpora (Brown et al., 2020), whereas the neuroimaging datasets are limited (Esteva et al., 2019; Pan et al., 2023; Krishnan et al., 2022). Nevertheless, self-supervised learning has the ability to generate its own labels, denoted as pseudo-labels, guiding the creation of latent representation through clustering-based models (Zhan et al., 2020; Caron et al., 2018; Caron et al., 2020) despite the limited availability of datasets, making it applicable in this domain-specific context. As such, clustering-based self-supervised learning is particularly suitable for extracting intrinsic heterogeneous features from brain atrophy. To the best of our knowledge, this is the first attempt to introduce self-supervision to address heterogeneity of AD.

In this study, we used a deep self-supervised framework to disentangle intrinsic heterogeneous features from brain atrophy in AD patients. Our approach learned a clustered latent space that revealed distinct atrophy patterns and clinical characteristics. It did not directly align with disease subtypes but rather emerged as a result of a mixing effect with disease progression. Notably, the latent space learned in AD was disentangled along two dimensions: disease progression and subtypes. We were able to replicate this disentangled pattern in prodromal AD, indicating sustained heterogeneity in the early stage of the disease. Furthermore, we identified two distinct longitudinal progression patterns: medial-temporal and parietotemporal pathways, which exhibited a directed bottom-to-top flow in the latent space, without crossing subtypes. This observation is significant as it demonstrates the consistent disentangled dimensions of progression and subtypes across various stages of AD, and individual longitudinal observations. Our compelling results would provide valuable insights into heterogeneity of the disease.

2. Materials and methods

2.1. Participants

We utilized 1724 T1-weighted structural magnetic resonance imaging (MRI) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (http://adni.loni.usc.edu) and 1080 from Open Access Series of Imaging Studies (OASIS). The ADNI dataset were categorized into three, namely discovery, prodromal AD validation, and longitudinal validation. The discovery dataset, comprising 343 individuals with AD, was used to develop our model and identify subgroups. The prodromal AD validation dataset comprised 722 individuals with mild cognitive impairment (MCI) allowing us to replicate the discovery experiment and assess consistency in the early continuum. The longitudinal validation datasets included 110 individuals with AD at baseline and two-year follow-up visits, enabling the analysis of longitudinal progression. Notably, the prodromal AD and longitudinal validation datasets were not included in our model training. Additionally, 439 cognitively normal (CN) individuals were included as a control group. For external validation, we used the OASIS dataset, comprising 218 individuals with AD and 862 CN individuals to demonstrate the generalizability of our model across different data.

2.2. MRI preprocessing and cortical atrophy

In the ADNI dataset without overlapping images, the following MRI scan types were used: 1.5 T non-accelerated magnetization-prepared rapid acquisition gradient echo (MP-RAGE) scans for the ADNI-1 group, 3 T non-accelerated MP-RAGE or inversion recovery spoiled gradient echo (IR-SPGR) scans for the ADNI2 and ADNI-GO cohorts, and 3 T non-accelerated MP-RAGE scans for ADNI3. Detailed MRI scanner protocols and quality control are available online at https://adni.loni.usc.edu/methods/documents/mri-protocols/. In the OASIS dataset, we selected the latest version, OASIS3, which includes MRI scans with two different field strengths: 1.5 T and 3 T (LaMontagne et al., 2019). Details of MR quality control are provided by OASIS (Marcus et al., 2007).

We utilized cortical thickness obtained by processing raw MRI using the standard FreeSurfer v.6.0 pipeline (http://surfer.nmr.mgh.harvard.edu/). The individual cortical thickness values were transformed into fsaverage space and resampled to ic3 vertex-level by barycentric interpolation (Winkler et al., 2012). The ic3 level has 1284 vertices, including 642 vertices in each hemisphere. We selected this resolution as it is known to offer a balance between feature dimension and resolution for learning. This resolution is high enough to interpolate the brain surface into a sphere and low enough to maintain a manageable feature dimension considering data size (Winkler et al., 2018).

We calculated cortical atrophy of each patient, as described in previous research (Park et al., 2017; Kim et al., 2019; Lee et al., 2022). W-score quantifies the extent of cortical thinning in the disease group as compared to the CN group. To achieve this, we used multiple linear regression (MLR) models to fit each vertex's thickness of the normative reference group to the covariates age, sex, and years of education. The W-score was calculated as the difference between the real thickness of the disease groups and the predicted thickness from the MLR model (residuals in the disease group), divided by the standard deviation of residuals in the normative reference group. Finally, atrophy is a vector consisting of positive values created by converting the positive values of W-score to 0 and multiplying the remaining negative values by -1. This allowed us to provide a quantitative measure of neurodegeneration in our analysis.

2.3. Clinical characteristics

We used demographics and clinical characteristics, such as global cognition, genetic biomarker, and cerebrospinal fluid (CSF) biomarker provided by ADNI (Table 1). Individuals with one or two copies of the *APOE4* gene were considered to be *APOE4* carriers. Individuals with "AV45$>$1.1″ and "ABETA$<$880″ were categorized into amyloid-positive (Landau et al., 2013; Hansson et al., 2018). In particular, longitudinal diagnostic information was used to determine dementia conversion for the prodromal AD validation dataset. The categories for dementia conversion were converter, non-converter, and missing. We classified individuals who were converted to AD within 36 months from the first visit without being censored as converter (Supplementary Note). Furthermore, we used cognitive scores from four domains: memory, executive, language, and visuospatial. Among the numerous cognitive variables provided by ADNI, we used only scores that all subjects have. For prodromal AD dataset, the scores of three domains except for visuospatial were used due to missing data. For the OASIS3 dataset, we also included demographics and clinical characteristics such as global cognition, genetic biomarker, and CSF biomarker. Individuals with one

Table 1
Demographic and clinical characteristics of the overall study population.

ADNI	Discovery	Prodromal AD validation	Longitudinal validation		Cognitively Normal
			Baseline	Follow-up	
n	343	722	110	110	439
Diagnosis	AD	MCI	AD	AD	CN
Age, years	75.2 ± 7.9	73.0 ± 7.5	75.7 ± 7.4	77.7 ± 7.4	74.0 ± 5.8
Sex, female (%)	162 (47)	301 (41)	50 (45)	50 (45)	228 (51)
Education, years	15.1 ± 3.0	15.9 ± 2.8	14.7 ± 3.0	14.7 ± 3.0	16.3 ± 2.7
MMSE	23.1 ± 2.0	27.5 ± 1.8	23.0 ± 2.1	18.3 ± 6.1	29.0 ± 1.1
CDRSOB	4.4 ± 1.6	1.5 ± 0.8	4.6 ± 1.7	7.7 ± 3.6	0.0 ± 0.1
APOE4 carrier[a] (%)	221 / 330 (67)	370 / 720 (51)	74 / 106 (69)	74 / 106 (69)	126 / 438 (28)
Amyloid positivity[b] (%)	124 / 141 (87)	209 / 468 (56)	57 / 65 (87)	34 / 37 (91)	80 / 232 (33)
Converter[c] (%)	–	217 / 516 (42)	–	–	–

OASIS	External validation	Cognitively Normal
n	218	862
Diagnosis	AD	CN
Age, years	76.2 ± 7.2	68.7 ± 8.7
Sex, female (%)	100 (45)	498 (57)
Education, years	14.8 ± 2.9	16.1 ± 2.5
MMSE	24.8 ± 3.7	29.0 ± 1.2
CDRSOB	3.5 ± 1.8	0.0 ± 0.1
APOE4 carrier[a] (%)	136 / 215 (63)	299 / 858 (34)
Amyloid positivity[b] (%)	88 / 101 (87)	151 / 643 (23)

Data are presented as n (%) or mean ± standard deviation.

[a] The number of *APOE4* carriers among non-missing patients / the total number of *APOE4* non-missing patients (%).

[b] The number of amyloid-positive patients among non-missing patients / the total number of amyloid non-missing patients (%).

[c] The number of converters among non-missing patients / sum of the number of converters and non-converters (%).

or two copies of the *APOE4* gene were classified as *APOE4* carriers. Amyloid positivity was determined using the cut-off provided by OASIS (https://theunitedconsortium.com/wp-content/uploads/2021/07/OASIS-3_Imaging_Data_Dictionary_v1.8.pdf).

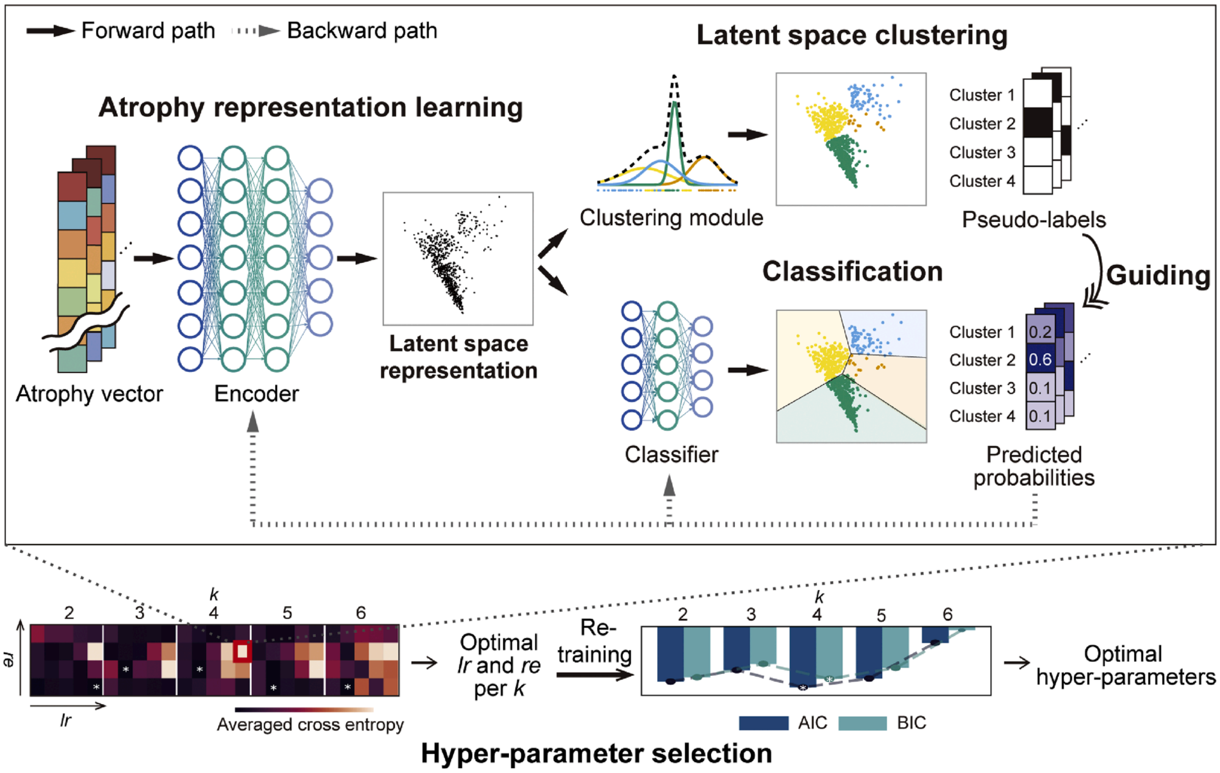

Fig. 1. Methodological overview. Our model comprises three components: atrophy representation learning, latent space clustering, and classification. Atrophy vectors are transformed into latent space representations through an encoder. These representations are inputted into a clustering module and a classifier simultaneously. The clustering module produces pseudo-labels that guide the classifier's predicted probabilities in a self-supervised manner. The model is trained to minimize cross-entropy between the pseudo-labels and the predicted probabilities. Hyper-parameters, such as learning rate (*lr*) and the number of classification iterations between clustering (*re*), are selected through 10-fold cross-validation and grid search. Optimal number of clusters (*k*) is determined by re-training the model with the complete dataset.

2.4. Clustering-based deep self-supervised learning

We used the clustering-based deep self-supervised learning framework (Caron et al., 2018) to obtain deep latent representation unraveling AD heterogeneity (Fig. 1). It facilitates simultaneous atrophy feature extraction, clustering, and classification through an end-to-end learning with an integrated loss function. Our framework consisted of three steps: atrophy representation learning, latent space clustering, and classification. An encoder compressed original atrophy vectors and yielded latent representations of atrophy. A clustering module clustered the latent representations of atrophy and assigned a pseudo-label to the individual atrophy. Simultaneously, a classifier predicted the label (forward path). By minimizing cross-entropy loss between the pseudo-labels and predicted probabilities, both the classifier and encoder were refined. It contributes to the latent representations of atrophy (backward path). The latent representations help to disentangle heterogeneity of the disease into key dimensions in the latent space.

The encoder (E) is a five-layer multi-layer perceptron (MLP) with hidden neurons (1284, 256, 128, 64, and 16), transforming original atrophy vectors into a 16-dimensional latent space denoted as $\mathscr{L}$. The encoder function $\mathrm{E}: \mathbb{R}^{1284} \rightarrow \mathscr{L}$ is represented as $\mathbf{h}_i = E_\theta(\mathbf{x}_i)$, where $\mathbf{x}_i$ is the atrophy vector of the i th subject, $\mathbf{h}_i$ is its latent atrophy feature vector, and θ represents the encoder parameters. The encoder employed batch normalization and Relu activation functions between layers.

The output $\mathbf{h}_i$ is then fed into two components: the clustering module and the classifier. In the clustering module, the set $\{\mathbf{h}_i\}_{i=1}^n$ is clustered into c subgroups using a Gaussian mixture model (GMM) (Bishop and Nasrabadi, 2006) to address the issue associated with cluster vanishing in k-means clustering[47]. The GMM function is expressed as[29]:

$$p(\mathbf{h}_i|\eta) = \sum_{k=1}^{c} \pi_k \mathscr{N}(\mathbf{h}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{1}$$

Here, η represents the parameters of the clustering module, $\pi_k \in [0, 1]$, and $\sum \pi_k = 1$, with a multivariate Gaussian distribution $\mathscr{N}$ characterized by mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$. The Mahalanobis distance serves as the clustering metric to consider the covariance of each variable in the latent space. The clustering module assigns a pseudo-label $\mathbf{y}_i$ to each individual through $G: \mathrm{L} \rightarrow \mathscr{Y}$ as $\mathbf{y}_i = G_\eta(\mathbf{h}_i)$, where $\mathbf{y}_i \in \{0, 1\}^c$, $1^{\mathrm{T}}\mathbf{y}_i = 1$, and $\mathscr{Y}$ is the pseudo-label space.

The classifier, $C: \mathrm{L} \rightarrow \widehat{\mathscr{Y}}$, consists of a fully-connected layer and a soft-max function. It predicts pseudo-labels from the clustering module, generating the predicted probability $\widehat{\mathbf{y}}_i = C_\lambda(\mathbf{h}_i)$, where $\widehat{\mathbf{y}}_i \in [0, 1]^c$, $1^{\mathrm{T}}\widehat{\mathbf{y}}_i = 1$, $\widehat{\mathscr{Y}}$ is the predicted probability space, and λ represents the parameters of the classifier.

2.5. Training

Our model's training alternates between optimizing the clustering module and the neural backbone (encoder and classifier), with multiple neural backbone updates for each clustering module update. Specifically, the atrophy vectors first enter the MLP-based encoder and are mapped into the latent space. These latent features are then fed into both the clustering module and the classifier. In the clustering module, the latent features are clustered by GMM and assigned pseudo-labels (first learning: clustering module). In the classifier, the latent features are converted into class probabilities. The encoder and classifier weights are updated to minimize the cross-entropy between the pseudo-labels and the class probabilities (second learning: neural backbone). During training, the encoder and classifier undergo multiple updates between clustering sessions. The GMM is optimized using the expectation-maximization algorithm to maximize the likelihood estimate represented in Eq. (1). The loss function for the clustering module L_c is defined as:

$$L_c = \sum_{i=1}^{n} -\log p(\mathbf{h}_i|\eta) \tag{2}$$

The neural-backbone's parameters are updated to minimize the cross-entropy between the pseudo-labels (target value of the classifier) and the predicted probabilities. The loss function for the neural-backbone L_n is expressed as:

$$L_n = \sum_{i=1}^{n} -\mathbf{y}_i^{\mathrm{T}} \log \widehat{\mathbf{y}}_i \tag{3}$$

The overall loss function is given by:

$$(\theta^*, \eta^*, \lambda^*) = \underset{(\theta,\ \eta,\ \lambda)}{\operatorname{argmax}} \sum_{i=1}^{n} \left[\log p(E_\theta(\mathbf{x}_i)|\eta) + G_\eta(E_\theta(\mathbf{x}_i))^{\mathrm{T}} \log q(E_\theta(\mathbf{x}_i)|\lambda)\right]. \tag{4}$$

2.6. Hyper-parameter selection

For hyper-parameter selection, we employed grid search coupled with 10-fold cross-validation. Notably, the choice of the number of clusters (denoted as "k") held substantial importance, as it impacted both disease heterogeneity identification and latent representations. Hyper-parameters such as learning rate (*lr*) and iteration counts between cluster assignments (*re*) were selected for each k based on averaged cross-entropy loss. The optimal value for k was chosen using information criteria, specifically AIC (Akaike information criterion) and BIC (Bayes information criterion).

2.7. Model evaluation

We implemented nested cross-validation to evaluate our model's generalizability irrespective of hyperparameters. This process involved 10 outer folds and 2 inner folds. Each training split within the outer folds was divided into 2 inner folds, where a grid search was conducted to tune hyperparameters based on the average loss. The model's performance was then assessed on the test split of each outer fold, and the results were averaged across all outer folds.

To assess the effectiveness of deep latent representations for feature extraction, we conducted comparisons with two alternative models, both equipped with a clustering module and a classifier but treated separately. One model abstained from utilizing any dimensionality reduction procedure, denoted as "GM+MLP", while the other employed the linear dimensionality reduction method principle component analysis (PCA), denoted as "PCA+GM+MLP". Evaluation of these models was based on classification accuracy and F1 score. Furthermore, we calculated the adjusted mutual information (AMI) for two sets of pseudo-labels generated by the clustering module in consecutive epochs. This provided insights into the stability of clustering during the training process in our model.

2.8. Latent representation

The latent space was structured as a metric space, using an averaged Mahalanobis distance defined as follows:

$$\mathrm{D}((\mathbf{x}_1, \mathbf{x}_2)) = \frac{1}{k} \sum_{j=1}^{k} \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^T \Sigma_j^{-1} (\mathbf{x}_1 - \mathbf{x}_2)} \tag{5}$$

Here, $\sum_j$ represents the covariance matrix of the j-th cluster in the GMM. The k Mahalanobis distances between two points, considering the distribution of each cluster, were computed for all clusters and then averaged to encompass the entire cluster distribution. To visualize the latent space, we applied multi-dimensional scaling (MDS) (Kruskal and Wish, 1978), a method that preserves distances between points in the latent space, which is not achievable through other methods like PCA.

2.9. *Trajectory visualization*

Individual trajectories were delineated in the latent space using the longitudinal validation dataset. Progression vectors, depicted as arrows from baseline to follow-up points, were formulated. Employing a quadtree approach (Gargantini, 1982), the latent space was partitioned into squared boxes to ensure a uniform subject distribution. Within each box, progression vectors were averaged, forming tensors indicating the direction of progression. In cases where a box lacked a progression vector, these tensors were utilized for interpolation. Individual trajectories were established from each person's baseline point. By connecting or disconnecting the tensors among neighboring boxes, each trajectory represents the path of the individuals.

2.10. *Atrophy reconstruction*

We reconstructed atrophy patterns within the latent space for two objectives: unveiling the significance of the latent space's axis and comprehending longitudinal trajectory. This reconstruction employed a decoder mirroring the encoder's structure, forming an encoder-decoder. We fixed the encoder's parameters and trained the decoder using the same hyper-parameters selected for the encoder. To explore the axes, 10,000 atrophy vectors were reconstructed, accompanied by two reference line segments. These vectors were grouped into four, and their atrophy values were averaged. Additionally, atrophy vectors were reconstructed along with two longitudinal trajectories at intervals of 0.1 distance in the latent space.

2.11. *Statistical analysis*

To compare atrophy patterns between each cluster and CN group, a general linear model (GLM) was used. False discovery rate (FDR) correction was applied to mitigate the impact of multiple comparisons. The clinical characteristics of each cluster were compared using GLM for continuous variables and logistic regression for categorical variables. These comparisons were adjusted for age, sex, and year of education.

To estimate the change in cognitive function over time, a linear mixed model (LMM) was employed. The response variable comprised individuals' mini-mental status examination (MMSE) scores. The explanatory variables included a dummy-coded cluster indicator, time, and their interaction. With random effects attributed to individuals and the inclusion of covariates like as age, sex, and years of education, the model accounted for individual variability and potential confounders. The analysis entailed statistical tests for cognitive decline rates in each cluster.

For estimating dementia conversion in prodromal AD, a Kaplan-Meier survival analysis was conducted (Peto et al., 1976). In this analysis, individuals reverting to CN status were treated as censored at the reference year. A log-rank test was implemented to compare survival rates across clusters.

3. Results

3.1. *Clustering-based self-supervised learning to understand AD heterogeneity: model selection and evaluation*

Our model learned latent representations of cortical atrophy in 343 patients with AD, grouping individuals with common latent atrophy features. For training, we utilized the Adam optimizer with L2 regularization at 10^{-5} and a momentum of 0.9, using a mini-batch size of 64. Optimal hyper-parameters per *k* were selected through cross-validation (Fig. S1). The k selection and model evaluation process revealed consistent clustering of latent features into four clusters (Fig. 2-a). Our model demonstrated stability with a convergence AMI of 0.941 ± 0.052 (Fig. 2-b). Comparing our model with others using both clustering and classification, GM+MLP lacks feature extraction, while PCA+GM+MLP incorporates linear extraction; optimal *k* values were two and four, respectively (Fig. 2-c). Accuracy, F1 score, and cross-entropy for the optimal *k* values per comparison model were presented alongside our model's, highlighting our model's superior performance across all evaluation metrics with the discovery and external validation dataset (Table 2). Additionally, our model showed better performance across prodromal AD and longitudinal datasets except for a higher cross-entropy loss observed in the longitudinal baseline dataset (Table 3).

3.2. *Disentangled latent representation in patient with AD*

The encoded atrophy features were distinctly separated along two axes (axis 1 and axis 2) within the latent space, representing the four

Table 2
Model comparison results using nested cross validation to assess performance with discovery and external validation dataset. Our proposed model consistently outperformed other models. Specifically, among the two comparison models, the one employing a linear extraction procedure demonstrated superior performance.

Model	*k*	ACC (%)	F1 score	Cross-entropy
GM+MLP	2	63.58 ± 6.36	0.42 ± 0.07	2.00 ± 1.45
PCA+GM+MLP	4	79.00 ± 0.07	0.72 ± 0.10	0.53 ± 0.14
Ours (discovery)	4	**91.53 ± 3.66**	**0.89 ± 0.06**	**0.26 ± 0.09**
Ours (external)	4	91.72 ± 7.29	0.92 ± 0.05	0.45 ± 0.51

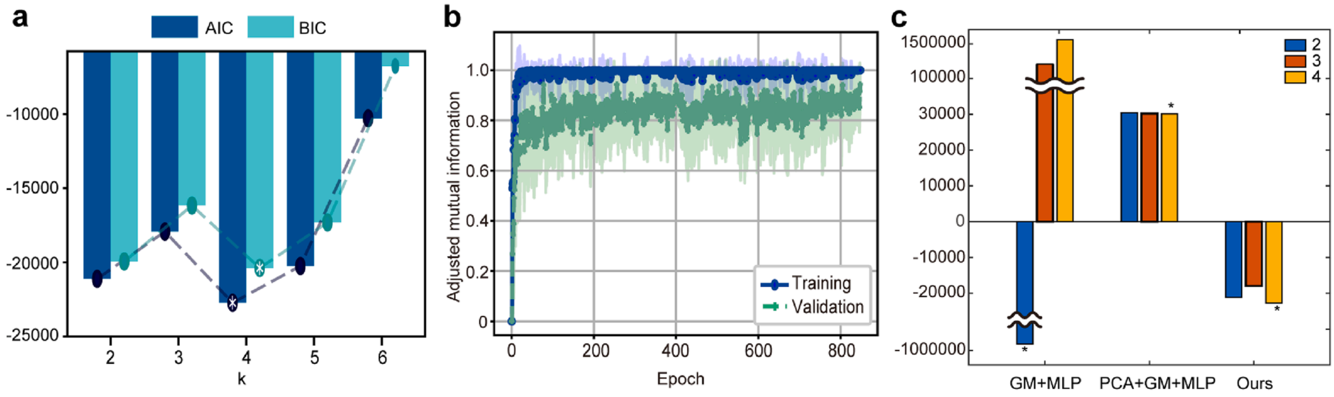

Fig. 2. k selection and model evaluation process. (a) A k value for our model was selected based on Akaike information criterion and Bayesian information criterion, with a consistent choice of four. (b) The stability evaluation of our model was assessed using adjusted mutual information (mean ± standard deviation), showing convergence. (c) The selection of k values using the Akaike information criterion for comparison models. For the GM+MLP model, a value of two was chosen, while for PCA+GM+MLP and our model, four was selected.

Table 3

Test results for prodromal AD and longitudinal (baseline and follow-up) datasets. Our model showed better performance than others across all datasets except for a higher cross-entropy loss observed in the longitudinal baseline dataset.

Data	Model	ACC (%)	F1 score	Cross-entropy
Prodromal AD	GM+MLP	64.27	0.39	5.33
	PCA+GM+MLP	76.04	0.58	0.85
	Ours	**88.64**	**0.82**	**0.83**
Longitudinal (Baseline)	GM+MLP	86.79	0.83	0.80
	PCA+GM+MLP	80.19	0.73	**0.52**
	Ours	**100**	**1.0**	0.70
Longitudinal (Follow-up)	GM+MLP	84.91	0.81	1.31
	PCA+GM+MLP	73.58	0.63	0.80
	Ours	**91.50**	**0.870**	**0.79**

cardinal directions (Fig. 3a). Each of the four clusters exhibited a distinct pattern of atrophy (Fig. 3b). Out of 343 individuals, 78 (22.7 %) individuals displayed the typical atrophy pattern of AD with balanced atrophy in the medial-temporal and parieto-temporal cortices and greater atrophy throughout the cortex compared to that of CN individuals (typical atrophy, Typ). The second cluster consisted of 121 (35.3 %) individuals with medial-temporal predominant atrophy, where substantial atrophy occurred in the entorhinal and parahippocampal areas (medial-temporal predominant, MT). Additionally, 38 (11.1 %) individuals exhibited predominant parieto-temporal atrophy, where parietal areas, including the precuneus and superior and inferior parietal lobules, were affected (parieto-temporal predominant, PT). The remaining 106 (30.9 %) patients showed minimal atrophy with no significant cortical thinning, except in the entorhinal cortex (minimal atrophy, Min).

We further compared the demographics and clinical characteristics of each cluster, including genetic traits (*APOE4* carrier status), CSF

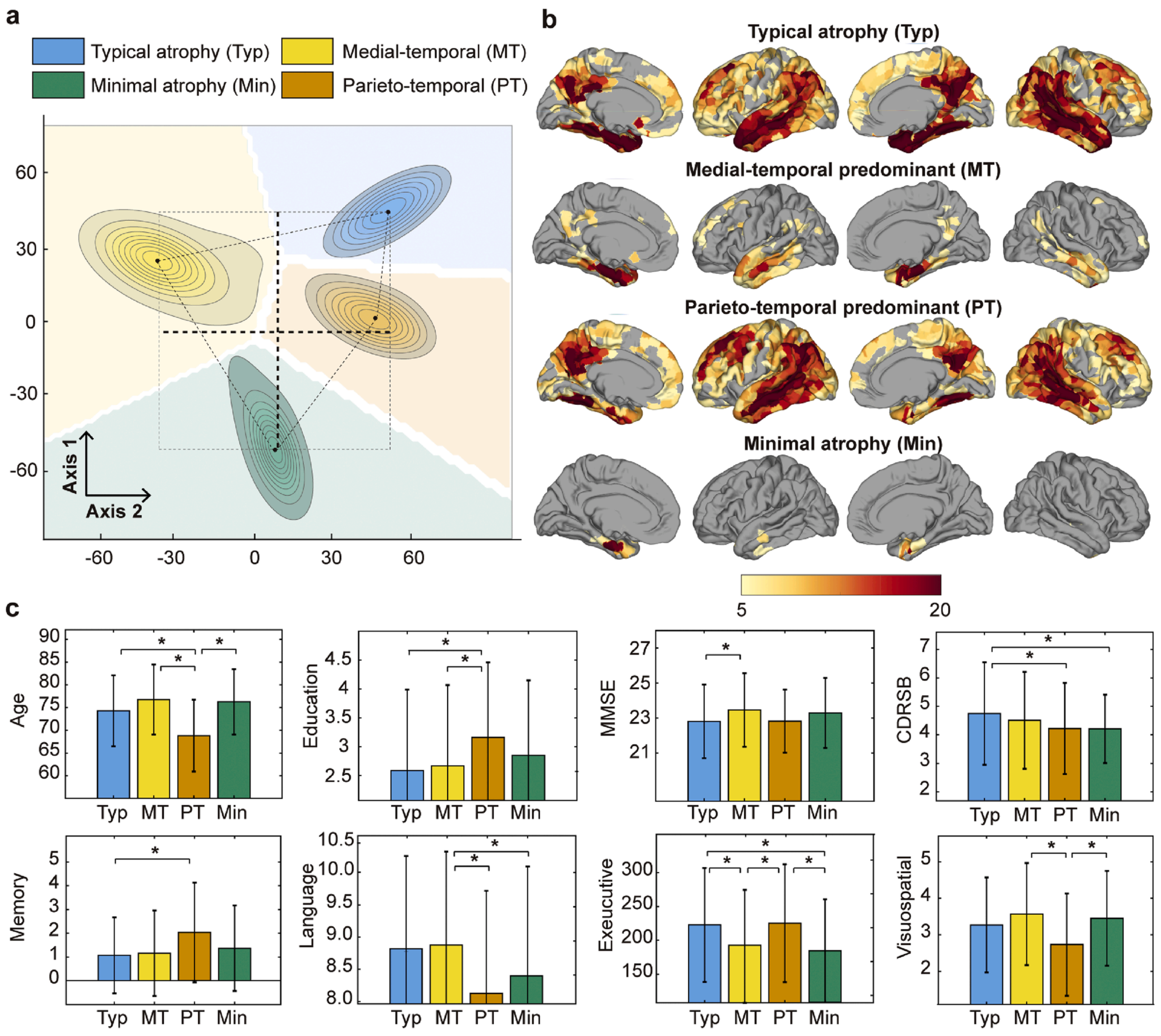

Fig. 3. Four distinct clusters in patients with AD. a. The latent space is effectively disentangled by vertical (axis 1) and horizontal (axis 2) axes. Level curves are overlaid on decision regions, illustrating each cluster's density distribution in the latent space. b. Each row represents atrophy patterns for identified clusters. Color signifies a negative logarithm p-value, denoting statistically significant increases in vertex level atrophy compared to CN individuals. P-values were derived from fitted GLMs with subsequent FDR correction. c. Clinical characteristics of each cluster were compared, adjusting age (except in the case of age), sex, and years of education (except in the case of education). An asterisk (*) indicates significant differences in characteristics between two clusters.

biomarkers of amyloid-beta, and cognitive scores (Fig. 2c and Table S2). In the MMSE, the Typ cluster exhibited worse scores than the MT cluster ($p = 0.0200$, $t = 2.3462$). Similarly, in the clinical dementia rating sum of boxes (CDRSB), the Typ cluster performed worse than the PT and Min clusters ($p_{PT}= 0.0495$, $t_{PT}= 1.9038$; $p_{Min}= 0.0423$, $t_{Min}= 1.9538$). The PT cluster included significantly younger patients than the other clusters ($p_{Typ}< 0.001$, $t_{Typ}= 3.7976$; $p_{MT}< 0.001$, $t_{MT}= 5.3491$; $p_{Min}< 0.001$, $t_{Typ}= 5.0913$). They were also significantly more educated than the Typ and MT clusters ($p_{Typ}= 0.0442$, $t_{Typ}= 2.0120$; $p_{MT}= 0.0419$, $t_{MT}= 1.7996$). In addition, individuals with PT displayed significantly worse executive ($p_{MT}= 0.0129$, $t_{MT}= 2.5171$; $p_{Min}= 0.0046$, $t_{Min}= 2.8810$) and visuospatial ($p_{MT}= 0.0413$, $t_{MT}= 2.0603$; $p_{Min}= 0.0048$, $t_{Min}= 2.8767$) function than those in the MT and Min clusters, but better memory function than those in the Typ cluster ($p = 0.0426$, $t = 1.19611$). On the other hand, the MT cluster showed worse language ($p_{PT}= 0.0361$, $t_{PT}= 2.1158$; $p_{Min}= 0.0442$, $t_{Min}= 1.9363$) and memory scores (not significant) than the PT and Min clusters.

3.3. Exploration of heterogeneity in prodromal AD

Our model was applied to 722 prodromal AD patients to explore heterogeneity in the early continuum. The distribution pattern of atrophy features in the latent space was consistent with the AD case (Fig. 4a), having similar atrophy patterns per cluster, albeit with a lower overall atrophy level than AD (Fig. 4b). Furthermore, we observed variations in the distribution ratios of each cluster between AD and prodromal AD (Fig. 4c). The proportion of Typ individuals in the prodromal AD group was approximately half lower than that in the AD group, while the proportion of individuals assigned to the Min cluster was doubled in the prodromal AD group. Moreover, the PT cluster in the prodromal AD group had fewer individuals than that in the AD group, whereas the MT cluster in the prodromal AD accounted for a comparable

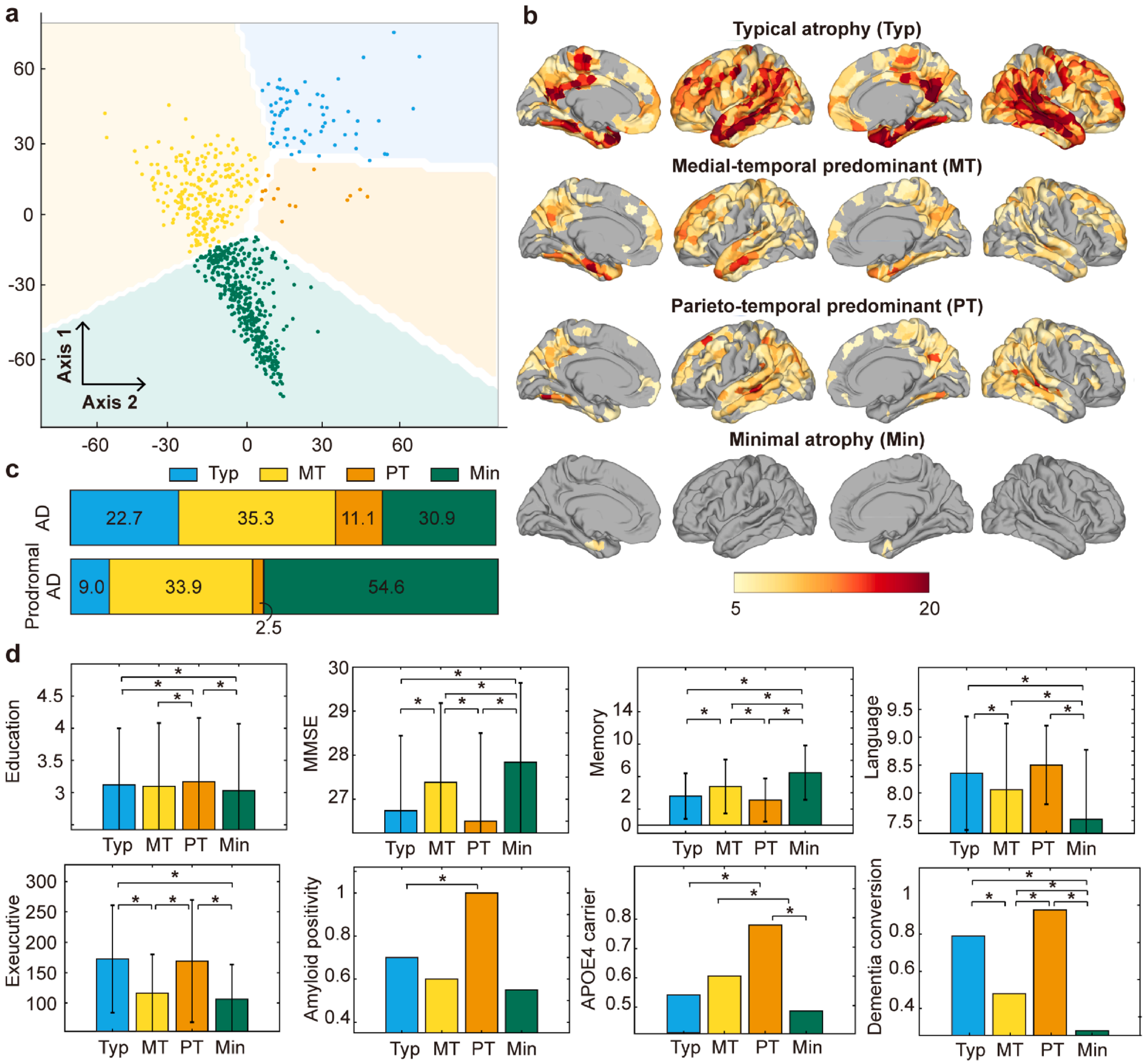

Fig. 4. Validation in prodromal AD. a. Prodromal AD individuals exhibit latent distributions akin to AD. Cluster-specific data points are marked by colored dots. b. Atrophy maps highlight significant levels of atrophy compared to CN individuals. P-values were determined through GLMs and subjected to FDR correction. c. Disparities in cluster proportions between AD and prodromal AD. d. Clinical characteristics of each cluster were compared, with adjustments for age (except in the case of age), sex, and years of education (except in the case of education). An asterisk (*) means that two clusters show significantly different characteristics.

percentage to that in the AD group.

Similar to the clinical characteristics of clusters in AD, the Typ cluster exhibited significantly worse MMSE scores than the MT and Min clusters ($p_{MT}= 0.0111$, $t_{MT}= 2.5558$; $p_{Min}< 0.001$, $t_{Min}= 4.4734$), while the Min cluster showed significantly better MMSE scores across all the clusters ($p_{Typ}< 0.001$, $t_{Typ}= 4.4734$; $p_{MT}= 0.0047$, $t_{MT}= 2.8396$; $p_{PT}< 0.001$, $t_{PT}= 3.431$) in prodromal AD (Fig. 3d and Table S2). The PT cluster included significantly more highly educated individuals than the Typ and Min clusters ($p_{Typ}= 0.0246$, $t_{Typ}= 1.8811$; $p_{Min}= 0.0022$, $t_{Min}= 1.2615$). However, there were several distinct characteristics compared to the AD cases. Individuals in the PT cluster were more likely to be *APOE4* carriers than those in the Typ and Min clusters ($p_{Typ}= 0.0246$, $t_{Typ}= 2.248$; $p_{Min}= 0.0022$, $t_{Min}= 3.0578$) and more likely to be amyloid-positive than those in the Typ ($p < 0.001$, $t = 26.99$). Furthermore, this cluster scored significantly worse on memory ($p_{MT}= 0.0141$, $t_{MT}= 2.4725$; $p_{Min}< 0.001$, $t_{Min}= 4.7462$) and execution ($p_{MT}< 0.001$, $t_{MT}= 3.5143$; $p_{Min}< 0.001$, $t_{Min}= 4.7464$) function than the MT and Min clusters and on language functions than the Min cluster ($p < 0.001$, $t = 3.4611$).

The dementia conversion ratios in the PT cluster were observed to be significantly higher than those in the MT and Min clusters ($p_{MT}= 0.0097$, $t_{MT}= 2.5873$; $p_{Min}< 0.001$, $t_{Min}= 3.3991$). We estimated survival distributions for all the clusters and compared the distributions using the log-rank tests (Fig. 5). The median survival time of the Typ cluster was 19.2 months and that of the PT cluster was 21.2 months. We found that all the clusters showed a significant difference from one another, except for the Typ and PT clusters (Table S3).

3.4. Disentangled key dimensions: progression and subtype

To understand two key axes in the latent space (Figs. 3a and 4a), we reconstruct atrophy vectors using a decoder that takes the latent representations as inputs. We utilized 110 atrophy of AD patients at baseline in the validation dataset. The atrophy was reconstructed along the two dashed black lines (Fig. 3a). Along axis 1, progressive one-sided (from left to right) atrophy was observed in various regions, including the medial and lateral temporal, cingulate, and frontal cortices (Fig. 6a). Conversely, for axis 2, moving in the negative direction, atrophy primarily affected the medial-temporal region, while becoming less pronounced in other regions. In contrast, moving in the positive direction, atrophy became more prominent in the parieto-temporal and prefrontal cortices.

We further examined the significance of the two axes by assessing the association between each axis's coordinate values and potential key determinants of heterogeneity, such as age, cognition, and atrophy progression. Notably, we found that axis 1 was strongly associated with global atrophy levels ($r = 0.7364$, $p < 0.001$) and cognition ($r = -0.4597$, $p = 0.0037$), rather than age (Fig. 6b). Moreover, axis 1 showed a significantly positive correlation with vertex-level atrophy throughout the brain (Fig. S2a). The MMSE scores were negatively correlated with axis 1 across all clusters (Fig. 6c). Additionally, the slope of the PT cluster was significantly steeper than those of the other clusters (Fig. 6d and Table S4). These observations suggest that as axis 1 increases, the rate of cognitive decline in the PT cluster would be faster than those in the other clusters. Conversely, axis 2 exhibited significant but not very strong correlations with age, global cognition, and global atrophy ($r_{age}= -0.2547$, $p_{age}< 0.001$; $r_{cog}= -0.1997$, $p_{cog}= 0.0237$; $r_{atrophy}= 0.3510$, $p_{atrophy}< 0.001$). In the vertex analysis, atrophy demonstrated a positive association with axis 2, primarily in the parieto-temporal area, while a negative association was observed in the frontal and entorhinal cortices (Fig. 6e). Overall, these analyses suggest that the latent space of brain atrophy in AD, or its heterogeneity, can be disentangled into two bases: progression (axis 1) and subtypes (axis 2).

3.5. Two distinct trajectories in latent space

Using 220 atrophy vectors of patients at baseline and follow-up in the longitudinal validation datasets, the progression vectors of individuals were examined in the latent space as arrows (Fig. 7a). This analysis

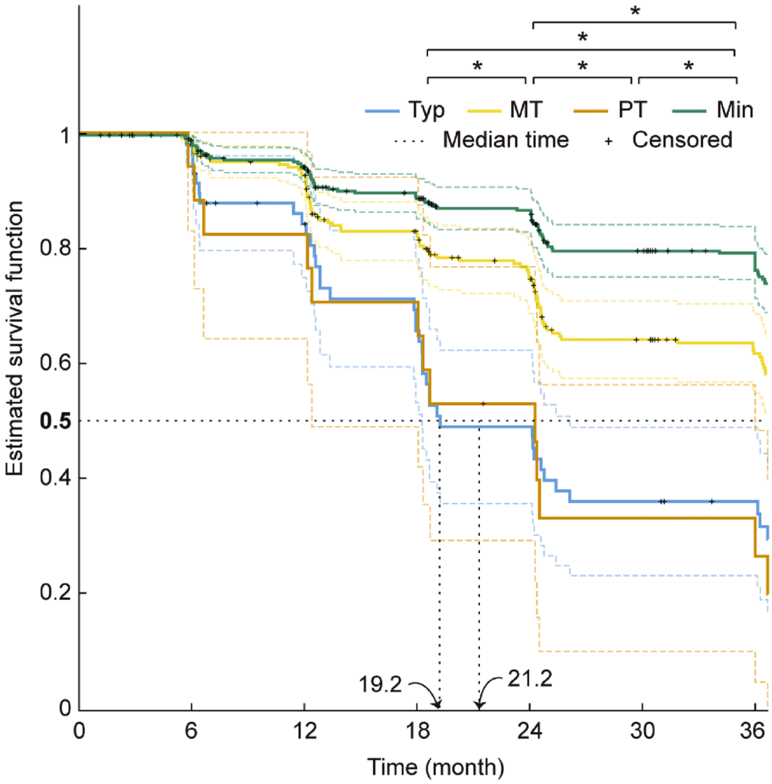

Fig. 5. Kaplan-Meier analysis for dementia conversion in prodromal AD. Survival distributions for each cluster are depicted in this figure. Log-rank tests were applied to compare distributions between two clusters. An asterisk (*) indicates significant differences between the two distributions.

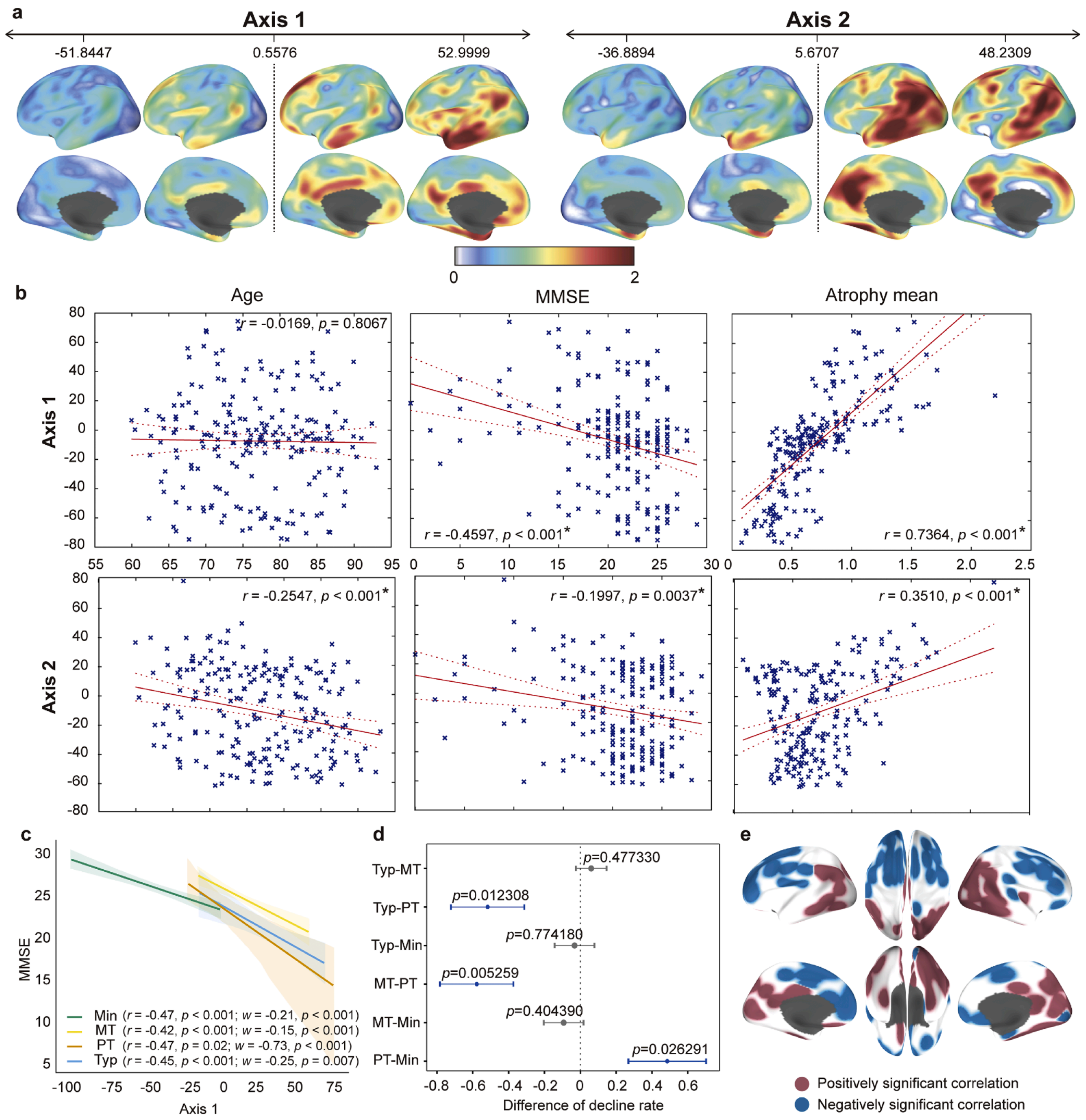

Fig. 6. Interpreting latent space axes. a. Patterns of reconstructed atrophy along each axis exhibit distinctive characteristics. b. Correlations were computed between each axis and determinants of heterogeneity: age, cognition, and atrophy. An asterisk (*) denotes significant correlations, with "*r*" indicating the correlation coefficient. c. Linear lines were fitted between axis 1 and MMSE for each cluster. "*w*" represents the corrected estimated slope between axis 1 and MMSE, adjusting for age, sex, and year of education. d. Comparison of MMSE decline rates across clusters while increasing axis 1. Significant differences are highlighted in blue. e. Colors indicate significant positive or negative correlations between vertex-level atrophy and axis 2.

revealed two distinct pathways in this progression: one pathway led from Min to MT and then to the Typ cluster, and the other involved progression from Min to PT and then to the Typ cluster. There was no transition between the MT and PT clusters (Fig. S2b). By estimating individual trajectories, as expected, they exhibited a diamond shape with a Y-shaped bifurcation in the bottom half and an inverted Y-shaped convergence in the upper half (Fig. 7b and Fig. S3). Two branches originating from the Min cluster progressed towards the MT and PT clusters and merged into a single large trunk within the Typ cluster. This trunk then stretched in the top-right direction, indicating the convergence of neurodegeneration. Notably, these pathways were clearly separated with no overlapping between the PT and MT clusters, before converging into the Typ cluster. These findings not only provide additional evidence that the disease heterogeneity can be disentangled by stage and subtype but also suggest two distinct progression pathways.

To understand the atrophy patterns along these pathways, we used the decoder, which reconstructs atrophy from latent representations. Atrophy was reconstructed along the two trajectories, T1 and T2. Several points on each trajectory were sampled with even lengths, interpolated by the corresponding latent representations. We revealed

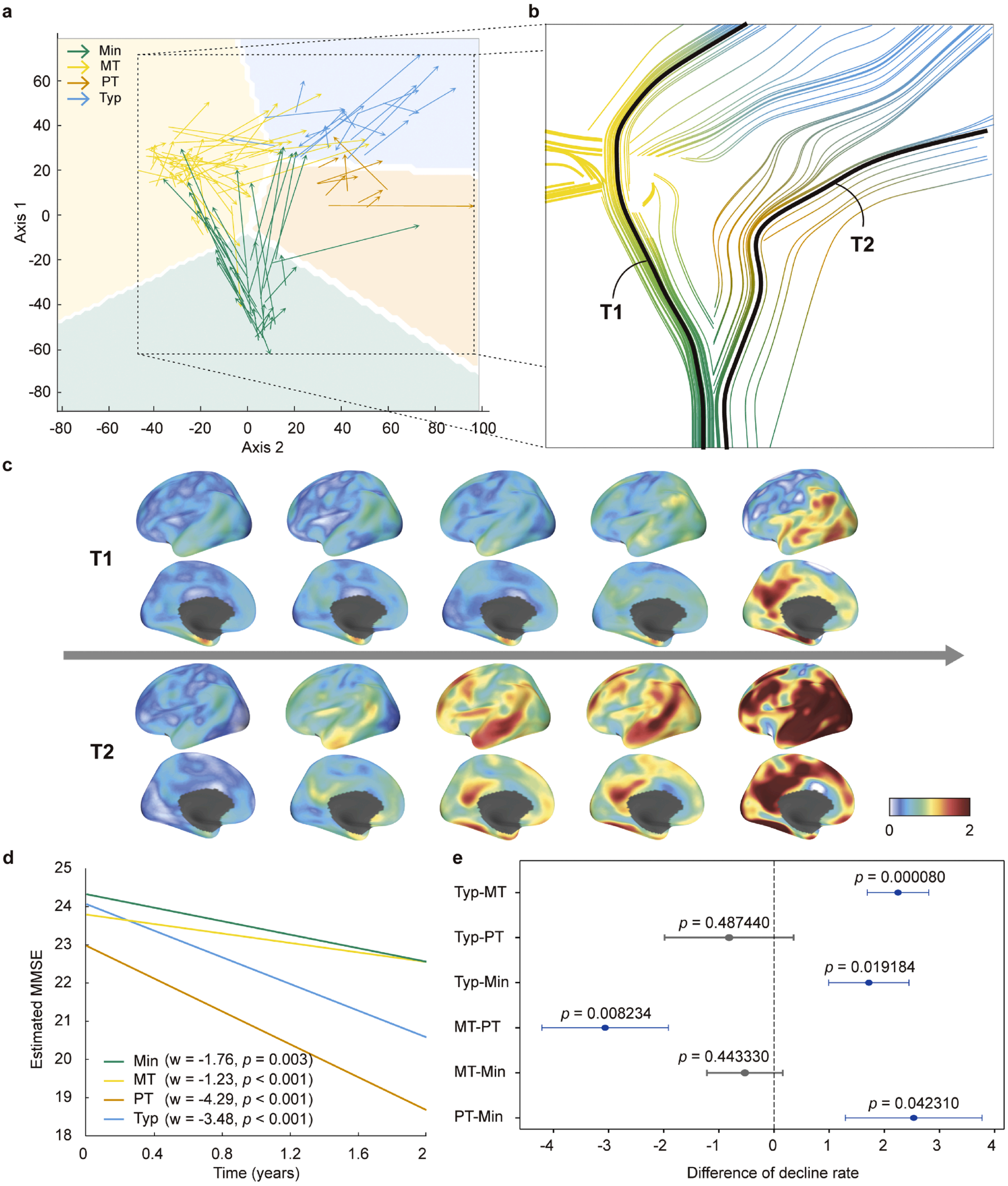

Fig. 7. Longitudinal progression in latent space. a. Individual progressions from baseline to follow-up are indicated by arrows within latent space. Baseline cluster assignments are color-coded. b. Individual trajectories are estimated. Each trajectory is depicted by a color-varying line, denoting the corresponding cluster at that location. c. Atrophy progression patterns of two trajectories (T1 and T2) are reconstructed. These trajectories commence at Min and culminate at Typ, traversing MT and PT, respectively. d. Longitudinal MMSE scores are estimated per cluster via linear mixed effect models, adjusted for age, sex, and education. e. Comparison of MMSE decline slopes across clusters. Significant differences are highlighted in blue.

two progressive patterns of neurodegeneration (Fig. 7c). Specifically, T1 showed atrophy that focused on the entorhinal cortex in the early stage, which spread to the hippocampal gyrus, followed by severe engagement of the temporal and parietal areas (medial-temporal pathway). This pathway had a slow progression rate. On the other hand, T2 showed similar atrophy in all the areas in the early stage, which spread to the fusiform, precuneus, and superior temporal gyri, and then displayed deeper atrophy in the parieto-temporal and frontal cortices (parieto-temporal pathway). This pathway exhibited a faster deterioration rate. To further investigate the rate of progression in each cluster, the longitudinal MMSE scores were estimated using LMM (Fig. 7d). These results show that except for the Typ cluster, the PT cluster experienced a significantly faster decline in MMSE scores than all the other clusters (p_{MT}= 0.0082, t_{MT}= 2.6691; p_{Min}= 0.0423, t_{Min}= 2.0436) (Fig. 7e). In contrast, the MT cluster exhibited a significantly slower decline than all the other clusters, except for the Min cluster (p_{Typ}< 0.001, t_{Typ}= 4.0274) (Table S5).

4. Discussion

To understand AD heterogeneity, many studies have focused on systematic subtypes in the past decade (Whitwell et al., 2012; Noh et al., 2014; Park et al., 2017; Ferreira et al., 2017; Lowe et al., 2018; Whitwell et al., 2018; Poulakis et al., 2018; Habes et al., 2020). However, there has been a disagreement regarding the number of these subtypes (Ferreira et al., 2020; Habes et al., 2020) and the underlying factors contributing to this heterogeneity (Ritchie and Touchon, 1992; Lambert and Amouyel, 2007; Lo et al., 2019; Smirnov et al., 2022). To address this, we disentangled AD heterogeneity into two factors, revealed as progression and subtype, using the deep latent representation with a self-supervised framework.

Our model excelled compared to clustering alone or clustering with linear dimensionality reduction. It showcased superior feature extraction using deep latent representations over linear methods (Cho et al., 2020). Self-supervised learning with pseudo-labels improved classification performance compared to other semi-supervised methods (Varol et al., 2017). Our external validation demonstrated our model's generalizability across different datasets, given the same number of parameters and possible choices of architecture in the framework. With the highest F1-score, surpassing even accuracy, it demonstrated robustness in handling class-imbalance, consistent with previous studies (Caron et al., 2018). Predictive ability for prodromal AD patients and longitudinal scans closely followed that for AD cases, affirming the model's efficacy across the entire AD cycle.

In line with previous studies (Ferreira et al., 2017; Byun et al., 2015; Ferreira et al., 2020; Poulakis et al., 2018), our analysis revealed four clusters each separated into four cardinal directions within the latent space. In particular, the Typ and Min clusters, which represent disease severity, were located at the top and bottom of the latent space, respectively. On the other hand, the MT and PT clusters, which are well-known representatives of variation in AD (Habes et al., 2020), were located at the left and right sides of the latent space, respectively. This clear separation in the latent space indicates the effective disentanglement of progression and subtype. Based on the global cognitive scores, the Typ cluster demonstrated the worst scores, while the Min cluster displayed the best score across all the clusters. Additionally, prodromal AD, which precedes AD, showed higher proportions of the Min cluster and lower proportions of the Typ cluster than AD, supporting the hypothesis that the Min and Typ clusters imply a different disease continuum rather than subtypes. Furthermore, the atrophy patterns decoded along the vertical axis showed progressive deterioration, strongly associated with cognitive scores. In contrast, the reconstructed atrophy along the horizontal axis showed two distinct patterns, reflecting an atypical subtype.

We identified clinical differences between the PT and MT types in patients with AD and prodromal AD. In AD, the individuals in the PT type were likely to be younger than those in the MT type, which is consistent with studies investigating the relationship between early age at onset and extensive parietal involvement (Ossenkoppele et al., 2012). The PT type also demonstrated higher education levels, consistent with previous studies (Ferreira et al., 2020; Poulakis et al., 2022). Executive and visuospatial scores were worse in the PT group, whereas the memory and language functions were more impaired in the MT group, consistent with previous studies on cognitive domain differences according to subtype (Byun et al., 2015; Zhang et al., 2016; Smits et al., 2012). Additionally, the PT type exhibited a faster cognitive decline rate, whereas the MT type showed a slower decline rate, which is consistent with the previous studies (Byun et al., 2015; Stanley et al., 2019). In prodromal AD, the PT type had the highest proportion of *APOE4* carriers and amyloid-beta positive, which is consistent with the results of previous studies (Kim et al., 2019; Ten Kate et al., 2018). Additionally, the PT type was more likely to convert to dementia with a faster conversion time than that of the MT type. This is consistent with the results of previous studies on the characteristics of PT type dementia conversion (Kim et al., 2019; Na et al., 2016). As mentioned above, there was a smaller PT type population in prodromal AD than in AD, which is attributed to we thought the early conversion to dementia of the PT type, whereas a moderate incidence of dementia conversion for the MT type in prodromal AD could lead to an equivalent MT type population in AD. Despite the early and faster declines of atrophy and cognition in the PT type, there was no difference in the global cognitive scores at the disease stage between MT and PT types. We postulated that it is likely attributed to modulatory factors such as the cognitive reserve since the education level known as a proxy of cognitive reserve was higher in the PT type than the MT type.

Our longitudinal analysis of latent space provided valuable insights into the progression of AD. The diamond-shaped trajectories and the reconstructed atrophy along the trajectories demonstrated two distinct pathways of progression. The first, represented by the T1 branch that sequentially proceeded through three clusters, Min, MT, and Typ, is analogous to the Braak stages, which describe the stereotypical tau pathology and neurodegeneration, beginning with the entorhinal region and spreading to the hippocampal cortex and temporal area (Braak et al., 2006). These branches were referred to as "medial-temporal pathways". In contrast, the other, represented by the T2 branch from Min through PT to Typ, showed that atrophy occurs in the perirhinal cortex and spreads to the fusiform, precuneus, and temporal regions. These branches were referred to as "parieto-temporal pathways." The worse visuospatial function of the PT type than that of the MT type is likely attributed to the deterioration in the perirhinal cortex and fusiform gyrus involving the visual function (Ferko et al., 2022), even more than the entorhinal cortex (Weiner and Zilles, 2016). The progressive patterns are closely aligned with those reported in previous findings (Young et al., 2018) and the faster atrophy deterioration in the parieto-temporal pathways than that in the medial-temporal pathways is in line with previous studies (Vogel et al., 2021). Notably, the disease progression eventually converged into the typical pattern. Additionally, this latent space analysis provides a more specific investigation of heterogeneity over time, unlike previous studies that mainly demonstrated the longitudinal change of subtype via the cluster transition result (Fig. S2b). In our case, for example, a large proportion of individuals in MT (36 / 50 = 72 %) and PT (5 / 9 = 55 %) appeared to remain within their respective clusters after 2 years according to the cluster transition result. Not quite the same as this result, we observed that they exhibited movement in the latent space (Fig. S2c).

5. Limitation

Our study had some limitations. Only atrophy data was used to identify disease heterogeneity and other imaging modalities were not considered. To ensure the fundamental assumption of deep learning, the central limit theorem, a sufficiently large dataset is required.

Unfortunately, owing to the shortage of data, we were unable to utilize other imaging modalities, such as tau PET molecular imaging. In future studies, we plan to utilize cumulated tau PET imaging to reveal the relationships between tau accumulation patterns and neurodegeneration subtype by adopting causality-based explainable artificial intelligence techniques to explore heterogeneity of AD. Secondly, longitudinal analysis requires data with more time points to capture disease progression over an extended period. Nevertheless, our study consistently demonstrated reliable results, effectively validating our findings and uncovering significant insights in the latent space. Future studies would greatly benefit from expanding the dataset with additional time points to enhance the depth of analysis and capture the longitudinal dynamics of AD.

6. Conclusions

This study disentangles the key factors of heterogeneity: progression and subtypes in a fully data-driven manner and presents two progressive pathways in AD using a self-supervised deep learning mechanism. We established deep latent representations of brain atrophy, which could be integrated with other deep learning model enabling to be fused with other modalities or to explain the results, which lead to comprehensively understanding the disease. Furthermore, we demonstrated the consistency of our model for prodromal AD, thereby establishing a foothold for the early diagnosis of AD. We believe that this framework could help in understanding disease stages and subtypes, unraveling the complexity of the disease heterogeneity and paving the way for more effective treatments and patient-tailored brain rehabilitation (Yu et al., 2019; Yu et al., 2023; Yu et al., 2018) based on various deep learning model.

Abbreviations

AD = Alzheimer's disease; ADNI = Alzheimer's Disease Neuroimaging Initiative; AIC = Akaike information criterion; BIC = Bayes information criterion; CDRCOB = clinical dementia rating sum of boxes; CN = cognitively normal; FDR = false discovery rate; GLM = general linear model; GMM = Gaussian mixture model; LMM = linear mixed model; MT = medial-temporal predominant cluster; MCI = mild cognitive impairment; Min = minimal atrophy cluster; MDS = multidimensional scaling; MLP = multi-layer perceptron; MLR = multiple linear regression; PT = parieto-temporal predominant cluster; Typ = typical atrophy cluster

Ethics statement

Ethical approval for ADNI was granted by the review boards of the involved institutions, and all participants provided informed consent before enrolling in the study.

Consent for publication

Not applicable.

CRediT authorship contribution statement

Sohyun Kang: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Sung-Woo Kim:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Joon-Kyung Seong:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no competing interests.

Data availability

The data from this study were sourced from the ADNI database (http://adni.loni.usc.edu). The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. The code is available on GitHub (https://github.com/SohyunKang/DecoNet).

Funding

This research was supported by the Basic Research Program through the National Research Foundation of Korea (NRF) funded by the MSIT (No. 2022R1A4A1033856), and by the grant of the Research of Korea Disease Control and Prevention Agency (grant No. 2024-ER1003-00), funded by the Ministry of Health & Welfare, Republic of Korea. Additionally, This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2023R1A2C2006201, Development of simulation-based Digital-Brain editing technology).

Acknowledgments

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2024.120737.

References

Armstrong, R.A., Nochlin, D., Bird, T.D., 2000. Neuropathological heterogeneity in Alzheimer's disease: a study of 80 cases using principal components analysis. Neuropathology. 20 (1), 31–37.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. 35 (8), 1798–1828.

Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning, 4. Springer.

Braak, H., et al., 2006. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta Neuropathol. 112 (4), 389–404.

Brown, T., et al., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.

Byun, M.S., et al., 2015. Heterogeneity of regional brain atrophy patterns associated with distinct progression rates in Alzheimer's disease. PLoS ONE 10 (11), e0142756.

Caron, M., et al., 2018. Deep clustering for unsupervised learning of visual features. In: Proceedings of the European conference on computer vision (ECCV).

Caron, M., et al., 2020. Unsupervised learning of visual features by contrasting cluster assignments. Adv. Neural Inf. Process. Syst. 33, 9912–9924.

Cho, H., et al., 2020. Principal components of tau positron emission tomography and longitudinal tau accumulation in Alzheimer's disease. Alzheimers Res. Ther. 12 (1), 1–11.

Ericsson, L., et al., 2022. Self-supervised representation learning: introduction, advances, and challenges. IEEE Signal Process. Mag. 39 (3), 42–62.

Esteva, A., et al., 2019. A guide to deep learning in healthcare. Nat. Med. 25 (1), 24–29.

Ferko, K.M., et al., 2022. Activity in perirhinal and entorhinal cortex predicts perceived visual similarities among category exemplars with highest precision. Elife 11, e66884.

Ferreira, D., et al., 2017. Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. Sci. Rep. 7 (1), 46263.

Ferreira, D., Nordberg, A., Westman, E., 2020. Biological subtypes of Alzheimer disease: a systematic review and meta-analysis. Neurology 94 (10), 436–448.

Gargantini, I., 1982. An effective way to represent quadtrees. Commun. ACM 25 (12), 905–910.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT press.

Habes, M., et al., 2020. Disentangling heterogeneity in Alzheimer's disease and related dementias using data-driven methods. Biol. Psychiatry 88 (1), 70–82.

Hansson, O., et al., 2018. CSF biomarkers of Alzheimer's disease concord with amyloid-β PET and predict clinical progression: a study of fully automated immunoassays in BioFINDER and ADNI cohorts. Alzheimer's Dementia 14 (11), 1470–1481.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507.

Jing, L., Tian, Y., 2020. Self-supervised visual feature learning with deep neural networks: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 43 (11), 4037–4058.

Kim, H.J., et al., 2019. Cortical atrophy pattern–based subtyping predicts prognosis of amnestic MCI: an individual-level analysis. Neurobiol. Aging 74, 38–45.

Krishnan, R., Rajpurkar, P., Topol, E.J., 2022. Self-supervised learning in medicine and healthcare. Nat. Biomed. Eng. 6 (12), 1346–1352.

Kruskal, J.B., Wish, M., 1978. Multidimensional Scaling. Sage.

Kwak, K., et al., 2021. Subtyping of mild cognitive impairment using a deep learning model based on brain atrophy patterns. Cell Rep. Med. 2 (12).

Lam, B., et al., 2013. Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. Alzheimers Res. Ther. 5 (1), 1–14.

Lambert, J.-C., Amouyel, P., 2007. Genetic heterogeneity of Alzheimer's disease: complexity and advances. Psychoneuroendocrinology 32, S62–S70.

LaMontagne, P.J., et al., 2019. OASIS-3: Longitudinal neuroimaging, clinical, and Cognitive Dataset For Normal Aging and Alzheimer disease. MedRxiv, p. 2019.12. 13.19014902.

Landau, S.M., et al., 2013. Amyloid-β imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. J. Nucl. Med. 54 (1), 70–77.

Lee, W.J., et al., 2022. Regional Aβ-tau interactions promote onset and acceleration of Alzheimer's disease tau spreading. Neuron 110 (12), 1932–1943 e5.

Leeb, F., et al., 2022. Exploring the latent space of autoencoders with interventional assays. Adv. Neural Inf. Process. Syst. 35, 21562–21574.

Li, K., et al., 2021. Feature extraction and identification of Alzheimer's disease based on latent factor of multi-channel EEG. IEEE Trans. Neural Syst. Rehab. Eng. 29, 1557–1567.

Lo, M.-T., et al., 2019. Identification of genetic heterogeneity of Alzheimer's disease across age. Neurobiol. Aging 84, 243 e1-243. e9.

Lowe, V.J., et al., 2018. Widespread brain tau and its association with ageing, Braak stage and Alzheimer's dementia. Brain 141 (1), 271–287.

Marcus, D.S., et al., 2007. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. 19 (9), 1498–1507.

Martí-Juan, G., et al., 2023. MC-RVAE: multi-channel recurrent variational autoencoder for multimodal Alzheimer's disease progression modelling. Neuroimage 268, 119892.

Mohanty, R., et al., 2023. Neuropathological features of antemortem atrophy-based subtypes of Alzheimer's disease. Alzheimer's Dementia 19, e067339.

Murphy, K.P., 2012. Machine learning: a Probabilistic Perspective. MIT press.

Murray, M.E., et al., 2011. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. Lancet Neurol. 10 (9), 785–796.

Na, H.K., et al., 2016. Malignant progression in parietal-dominant atrophy subtype of Alzheimer's disease occurs independent of onset age. Neurobiol. Aging 47, 149–156.

Noh, Y., et al., 2014. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. Neurology 83 (21), 1936–1944.

Ossenkoppele, R., et al., 2012. Amyloid burden and metabolic function in early-onset Alzheimer's disease: parietal lobe involvement. Brain 135 (7), 2115–2125.

Ossenkoppele, R., et al., 2019. Distinct Tau PET Patterns in Atrophy-Defined Subtypes of Alzheimer's disease. Alzheimer's & dementia.

Pan, D., et al., 2023. Deep learning for brain MRI confirms patterned pathological progression in Alzheimer's disease. Adv. Sci. 10 (6), 2204717.

Park, J.Y., et al., 2017. Robust Identification of Alzheimer's Disease subtypes based on cortical atrophy patterns. Sci. Rep. 7, 43270.

Peto, R., et al., 1976. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br. J. Cancer 34 (6), 585–612.

Poulakis, K., et al., 2018. Heterogeneous patterns of brain atrophy in Alzheimer's disease. Neurobiol. Aging 65, 98–108.

Poulakis, K., et al., 2022. Multi-cohort and longitudinal Bayesian clustering study of stage and subtype in Alzheimer's disease. Nat. Commun. 13 (1), 4566.

Ritchie, K., Touchon, J., 1992. Heterogeneity in senile dementia of the Alzheimer type: individual differences, progressive deterioration or clinical sub-types? J. Clin. Epidemiol. 45 (12), 1391–1398.

Scheltens, N.M., et al., 2017. Cognitive subtypes of probable Alzheimer's disease robustly identified in four cohorts. Alzheimer's Dementia 13 (11), 1226–1236.

Smirnov, D.S., et al., 2022. Association of neurofibrillary tangle distribution with age at onset–related clinical heterogeneity in Alzheimer disease: an autopsy study. Neurology 98 (5), e506–e517.

Smits, L.L., et al., 2012. Early onset Alzheimer's disease is associated with a distinct neuropsychological profile. J. Alzheimer's Dis. 30 (1), 101–108.

Stanley, K., et al., 2019. Rate of cognitive decline in Alzheimer's disease stratified by age. J. Alzheimer's Dis. 69 (4), 1153–1160.

Ten Kate, M., et al., 2018. Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. Brain 141 (12), 3443–3456.

Varol, E., et al., 2017. HYDRA: revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. Neuroimage 145 (Pt B), 346–364.

Vogel, J.W., et al., 2021. Four distinct trajectories of tau deposition identified in Alzheimer's disease. Nat. Med. 27 (5), 871–881.

Wang, T., et al., 2021. Self-supervised learning disentangled group representation as feature. Adv. Neural Inf. Process. Syst. 34, 18225–18240.

Weiner, K.S., Zilles, K., 2016. The anatomical and functional specialization of the fusiform gyrus. Neuropsychologia 83, 48–62.

Wen, J., et al., 2022. Multi-scale semi-supervised clustering of brain images: deriving disease subtypes. Med. Image Anal. 75, 102304.

Whitwell, J.L., et al., 2012. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: a case-control study. Lancet Neurol. 11 (10), 868–877.

Whitwell, J.L., et al., 2018. [18F] AV-1451 clustering of entorhinal and cortical uptake in Alzheimer's disease. Ann. Neurol. 83 (2), 248–257.

Winkler, A.M., et al., 2012. Measuring and comparing brain cortical surface area and other areal quantities. Neuroimage 61 (4), 1428–1443.

Winkler, A.M., et al., 2018. Joint analysis of cortical area and thickness as a replacement for the analysis of the volume of the cerebral cortex. Cereb. Cortex 28 (2), 738–749.

Yang, Z., et al., 2021. A deep learning framework identifies dimensional representations of Alzheimer's Disease from brain structure. Nat. Commun. 12 (1), 7065.

Yang, Z., et al., 2024. Gene-SGAN: discovering disease subtypes with imaging and genetic signatures via multi-view weakly-supervised deep clustering. Nat. Commun. 15 (1), 354.

Young, A.L., et al., 2018. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. Nat. Commun. 9 (1), 4273.

Young, C.B., et al., 2022. Divergent cortical tau positron emission tomography patterns among patients with preclinical Alzheimer disease. JAMA Neurol. 79 (6), 592–603.

Yu, H., et al., 2018. Modulation of spectral power and functional connectivity in human brain by acupuncture stimulation. IEEE Trans. Neural Syst. Rehab. Eng. 26 (5), 977–986.

Yu, H., et al., 2019a. Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification. IEEE Trans. Fuzzy Syst. 28 (1), 60–71.

Yu, H., et al., 2019b. Modulation effect of acupuncture on functional brain networks and classification of its manipulation with EEG signals. IEEE Trans. Neural Syst. Rehab. Eng. 27 (10), 1973–1984.

Yu, H., et al., 2023. Probing the flexible internal state transition and low-dimensional manifold dynamics of human brain with acupuncture. Biomed. Signal Process. Control 82, 104494.

Zhan, X., et al., 2020. Online deep clustering for unsupervised representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Zhang, X., et al., 2016. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. Proc Natl. Acad. Sci. U S A, 113 (42), E6535–E6544.