Big Data and Information Analytics

https://www.aimspress.com/journal/BDIA

BDIA, 10: 308–337.
DOI:10.3934/bdia.2026016
Received: 26 August 2026
Revised: 22 September 2026
Accepted: 28 September 2026
Published: 27 October 2026

Research article

Subtype identification of Alzheimer's disease based on multimodal fusion and deep clustering

Yueli Chen[1*], **Taihe Yu**[1] **and Hui Chang**[2]

[1] School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China

[2] Department of Critical Care Medicine, Luoyang Orthopedic-Traumatological Hospital of Henan Province (Henan Provincial Orthopedic Hospital), Zhengzhou 450016, China

* **Correspondence:** Email: chenyueli@zzu.edu.cn.

Abstract: With growing global population aging, Alzheimer's disease (AD) poses a severe public health threat. Mild cognitive impairment (MCI), a subtle transitional state between normal cognition and AD, exhibits substantial heterogeneity and hinders accurate early diagnosis. This study aimed to resolve MCI heterogeneity via a multimodal data fusion approach for reliable subtype identification. Integrating clinical scores, genetic profiles, biomarkers, and structural magnetic resonance imaging (sMRI) data from the Alzheimer's disease neuroimaging initiative (ADNI) database, we constructed a standardized multimodal dataset for model development and validation. On the basis of the original Ordinary graph-based deep temporal clustering (OG-DTC) model, we incorporated an attention mechanism and optimized the loss function to establish an improved Attention-OG-DTC model. Five-fold cross-validation verified that Attention-OG-DTC outperformed the original model in silhouette coefficient and Davies-Bouldin index (DBI). Two distinct MCI subtypes were identified. Subtype 1 showed severer baseline cognitive impairment and broader cerebral atrophy, with significantly faster AD progression and poorer longitudinal prognosis ($p \leq 0.001$, log-rank test). shapley additive explanations (SHAP) analysis further determined key covariates governing subtype differentiation. This data-driven stratification reveals prominent MCI heterogeneity, providing credible evidence for the continuous and heterogeneous nature of the AD spectrum. The proposed method facilitates early refined screening and stratified intervention for AD.

Keywords: Alzheimer's disease; multimodal data fusion; deep clustering; subtype identification

1. Introduction

Alzheimer's disease (AD) is a prevalent neurodegenerative disorder in the elderly, characterized by progressive cognitive deterioration. As the leading cause of late-life dementia, AD accounts

for 60%–70% of all dementia cases. Affected individuals typically experience memory loss, language impairment, executive dysfunction, and behavioral and personality changes. Over the disease course, patients gradually lose the capacity for independent daily living, leading to a severely diminished quality of life. Moreover, AD imposes substantial caregiving burdens and significant economic strain on families and society at large.

Existing AD studies commonly partition participants into three diagnostic categories: cognitively normal (CN), mild cognitive impairment (MCI) [1], and AD. As a transitional stage between normal aging and dementia [2], MCI is notoriously difficult to identify accurately in conventional classification frameworks [3]. Individuals with MCI exhibit cognitive decline beyond age-related expectations while preserving most daily functional abilities. Beyond the binary clinical question of whether a patient will develop dementia, determining disease subtypes and long-term progression trajectories holds greater clinical relevance. The MCI population is notably heterogeneous: some patients remain cognitively stable over years, whereas others convert rapidly to AD. Traditional classification models, however, fail to capture such interindividual variability. To address this, the present study adopts an unsupervised deep clustering approach to identify intrinsic MCI subtypes. This data-driven strategy avoids subjective biases inherent in prior-knowledge-based stratification and offers objective evidence for interpreting the continuity and heterogeneity of the AD spectrum. When combined with longitudinal survival analyses that quantify subtype-specific conversion risks and progression rates, our framework facilitates early identification of high-risk individuals and supports personalized follow-up and intervention planning.

Identifying MCI—the prodromal stage of AD—is crucial for elucidating disease mechanisms and improving early diagnosis. Unsupervised clustering has become the predominant tool for MCI subtyping, with research shifting from static cross-sectional analyses toward dynamic modeling that integrates multimodal clinical data and longitudinal temporal features. Early subtyping studies relied on traditional algorithms such as K-means, BIRCH, and hierarchical clustering applied to single-time-point static features. Pourzinal et al. [4] employed K-means clustering in Parkinson's disease patients with comorbid MCI and identified four distinct cognitive phenotypes. Using FDG-PET images, Ryoo et al. [5] trained a conditional variational autoencoder and performed K-means clustering on latent representations, uncovering four metabolic AD subtypes; further application to MCI cohorts revealed significant differences in conversion rates across subtypes. Diverging from K-means-based methods, Chen et al. [6] utilized MCI subjects from the ADNI-2 database and implemented incremental clustering via BIRCH, yielding six subtypes with distinct cognitive and laboratory profiles. Edmonds et al. [7] conducted Ward's hierarchical clustering on neuropsychological data from ADNI-2, demonstrating that their redefined MCI subgroups exhibited steeper cognitive decline curves and more pronounced biomarker abnormalities. Nonetheless, these conventional clustering approaches generally neglect multidimensional clinical information and longitudinal disease dynamics, resulting in weak associations between derived subtypes and clinical outcomes and limiting their translational utility.

To overcome the limitations of static clustering, recent work has incorporated clinical priors and temporal information into subtype discovery. Sindhu et al. [8] constructed a hybrid optimized deep fuzzy clustering network to stratify cohorts into CN, early MCI (EMCI), MCI, late MCI (LMCI), and AD. Using structural MRI (sMRI), Manouvriez et al. [9] computed the brain age gap (BrainAGE) and applied K-means to distinguish early-onset AD patients with divergent progression speeds, identifying two optimal subtypes. The high-BrainAGE subtype showed accelerated cognitive deterioration and

cerebral atrophy, with more aggressive disease courses, supporting high-risk screening and clinical trial design. Chang et al. [10] built a multi-head deep learning architecture to classify participants into fast, moderate, slow, and non-converters; further stratification by race and gender produced 16 subgroups with distinct progression trajectories, and they screened core predictive features, validating heterogeneous risk factors across subpopulations.

Although these studies integrate temporal information, most still rely on conventional clustering or shallow networks. Recent advances in deep representation learning have spurred the development of deep clustering models that capture complex temporal patterns. Ribino et al. [11] extended K-means for multivariate time-series clustering to capture the coordinated dynamics of clinical features and identify subgroups with divergent progression patterns. Jiao et al. [12] proposed diagnosis-guided deep subspace clustering association (DDSCA), which uses diagnostic labels as constraints to jointly model high-level associations between brain connectomes and genetic variants, screening disease-relevant imaging and genetic markers. Wang et al. [13] established a two-stage Slice-to-Voxel Feature Representation (SVFR) framework to automatically extract salient sMRI features for AD diagnosis, achieving competitive performance against state-of-the-art methods. In the realm of temporal clustering integrated with clinical objectives, the outcome-guided deep temporal clustering (OG-DTC) framework proposed by Wang et al. [14] stands out. It employs gated recurrent unit (GRU) networks to model temporal dynamics, optimizes clustering via a K-means loss, and incorporates clinical outcome regression. By unifying representation learning, temporal modeling, and clinical priors within a single optimization objective, OG-DTC produces reproducible AD subtypes with favorable clinical interpretability. Following this outcome-guided paradigm, Bayta et al. [15] introduced adversarial perturbation into MCI conversion prediction, quantifying the perturbation distance required to shift patient features toward the AD diagnostic boundary and generating personalized risk metrics to characterize vulnerability to AD conversion. Hou et al. [16] developed an interpretable survival model combining discriminative and generative mechanisms; by learning weighted mixtures of parametric survival distributions, the model stratifies MCI conversion risks and uncovers latent temporal patterns underlying subtype transitions, enhancing biological interpretability.

In summary, MCI subtyping research has evolved from static grouping toward deep modeling integrating clinical guidance, temporal awareness, and prognostic prediction. However, existing deep clustering frameworks cannot adaptively weight distinct time points and clinical variables during feature learning, limiting discriminative temporal pattern extraction and reducing subtyping accuracy. To address this challenge, this study incorporates an attention mechanism into the OG-DTC framework [14]. This attention module adaptively focuses on critical time steps and informative features, enhances discriminative representation learning, and improves the plausibility of subtype partitions. This improvement refines learned clustering representations and supports subsequent disease progression modeling and individualized prognostic prediction.

2. Materials and method

2.1. Data sources

All data used in this study were obtained from the publicly available Alzheimer's disease neuroimaging initiative (ADNI) database [17]. Data were downloaded from the official ADNI website

following approval of a formal access request, in compliance with the ADNI Data Use Agreement.

ADNI is a pioneering, large-scale, multicenter international collaborative research project. Launched in 2004, it was initiated and funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the U.S. Food and Drug Administration, over a dozen pharmaceutical companies, and more than 20 leading research institutions. Its primary objective is to establish a standardized biomarker database that characterizes the clinical progression of AD.

The ADNI project has progressed through five phases:

(1) ADNI-1 (2004–2010) [18]: established preliminary standards for AD biomarkers, enrolling approximately 800 participants;

(2) ADNI-GO (2009–2011): expanded the research scope to investigate early-stage mild cognitive impairment (MCI);

(3) ADNI-2 (2011–2016): enlarged the sample size to 1500 participants and incorporated novel biomarkers and imaging sequences;

(4) ADNI-3 (2016–2022) [19]: introduced advanced tau-PET imaging, high-resolution MRI, and blood-based biomarkers;

(5) ADNI-4 (2023–2026): further validates biomarkers and optimizes MRI techniques, including compressed sensing and arterial spin labeling (ASL). This phase aims to improve sample diversity by recruiting participants from underrepresented populations, who are expected to constitute 50%–60% of the total cohort.

As of 2025, the ADNI database has enrolled more than 2500 participants, covering the full disease spectrum, including cognitively normal (CN) individuals, patients with MCI, and those with AD dementia. Longitudinal follow-up data spanning up to 15 years are available for these cohorts. ADNI employs rigorous participant screening criteria to ensure scientific validity and data comparability. The age range for inclusion was 55–90 years in ADNI-1 and was extended to 55–100 years in subsequent phases. Participants were classified into three diagnostic groups (AD, CN, MCI) based on the criteria of the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association [20] and Petersen's criteria [21]. Standardized cognitive assessment scales, including the mini-mental state examination (MMSE) [22], clinical dementia rating (CDR) [23], and Alzheimer's disease assessment scale–cognitive subscale (ADAS-Cog) [24], were administered. Subjects with other major neurological disorders, severe psychiatric history, or unstable systemic diseases were excluded to minimize confounding effects on AD-related analyses.

In this study, the structured data primarily include participant age, sex, genetic information, biomarker measurements, and region-of-interest (ROI) features derived via computational software. The imaging data comprise MRI scans from 956 participants in the ADNI database. To facilitate the MCI subtype classification analyses, we additionally leveraged the ADNIMERGE dataset, a curated derivative of the ADNI database. The dataset encompasses the following categories of variables:

(1) **Identifiers and temporal variables**. These variables enable robust participant tracking and facilitate seamless integration of cross-sectional and longitudinal data. Specifically, RID and PTID serve as primary participant identifiers, ensuring precise linkage across diverse data

modules; VISCODE denotes distinct follow-up visits; and temporal metrics such as Month_bl and Years_bl are essential for constructing longitudinal models that quantify cognitive trajectory changes over time. A detailed summary is provided in Table 1.

Table 1. Identifiers and temporal variables.

Variable name	Explanation
RID	Unique participant identifier
PTID	Internal project participant identifier
VISCODE	Visit code (e.g., baseline, month 6)
Month/Month_bl/Years_bl	Temporal metrics (months/years since baseline)
EXAMDATE_bl	Baseline assessment date
update_stamp	Data update timestamp

(2) **Demographic and genetic characteristics**. These include age, sex, years of education, and apolipoprotein E (APOE) $\varepsilon 4$ allele carrier status. The APOE4 genotype is a well-established genetic risk factor for late-onset AD, associated with earlier age of onset, altered lipid metabolism, and accelerated cognitive decline. Detailed variable descriptions are shown in Table 2.

Table 2. Demographic and genetic variables.

Variable name	Description
AGE	Age at baseline (years)
PTGENDER	Biological sex
PTEDUCAT	Years of education
APOE4	APOE $\varepsilon 4$ allele carrier status (count of $\varepsilon 4$ alleles)

(3) **Neuropsychological and clinical assessment scales**. To comprehensively evaluate participants' cognitive status, we incorporated multiple standardized neuropsychological instruments and clinical rating scales. These variables cover diverse cognitive and functional domains, including global dementia severity (CDRSB), core cognitive impairments (ADAS series: ADAS11, ADAS13, ADASQ4), verbal episodic memory (RAVLT series: immediate recall, learning, forgetting), executive function (DIGITSCOR, TRABSCOR), functional dependence (FAQ), and subjective cognitive decline (EcoG series). Additionally, the modified preclinical Alzheimer's cognitive composite scores (mPACCdigit and mPACCtrailsB) were included to capture subtle cognitive changes in preclinical and early-stage AD. A comprehensive inventory of these neuropsychological variables is presented in Table 3.

(4) **Brain regional volume variables**. To characterize structural brain morphometry, T1-weighted structural images were processed using FreeSurfer [25] for automated segmentation and volumetric quantification, yielding volume estimates for multiple key brain regions. These metrics (summarized in Table 4) encompass the core structures of the medial temporal lobe memory system (e.g., hippocampus and entorhinal cortex), the global parenchymal integrity (whole-brain volume), and the regions implicated in higher-order cognitive functions (e.g., fusiform and middle temporal gyri). Intracranial volume (ICV) was included to correct for inter-individual differences in head size. Additionally, we accounted for the technical heterogeneity across the multi-site data by incorporating scanner field strength ("FLDSTRENG") and FreeSurfer software version ("FSVERSION") as nuisance covariates in the subsequent analyses.

Table 3. Neuropsychological and clinical assessment scale variables.

Variable name	Description
CDRSB	Clinical dementia rating scale–sum of boxes
ADAS11	Alzheimer's disease assessment scale–11-item cognitive subscale
ADAS13	Alzheimer's disease assessment scale–13-item cognitive subscale
ADASQ4	Alzheimer's disease assessment scale–delayed word recall (q4)
MMSE	Mini-mental state examination
RAVLT_immediate	Rey auditory verbal learning test–immediate recall (sum of trials 1–5)
RAVLT_learning	Rey auditory verbal learning test–learning slope (trial 5–trial 1)
RAVLT_forgetting	Rey auditory verbal learning test–absolute forgetting (trial 5–delayed recall)
RAVLT_perc_forgetting	Rey auditory verbal learning test–percentage forgetting
LDELTOTAL	Logical memory delayed recall total (wechsler memory scale–revised)
DIGITSCOR	Digit symbol substitution test (dsst) score
TRABSCOR	Trail making test–part b time (seconds)
FAQ	Functional activities questionnaire
MOCA	Montreal cognitive assessment
EcoGpTMem	Everyday cognition scale–participant-rated memory
EcoGpTLang	Everyday cognition scale–participant-rated language
EcoGpTVisspat	Everyday cognition scale–participant-rated visuospatial function
EcoGpTPlan	Everyday cognition scale–participant-rated planning
EcoGpTOrgan	Everyday cognition scale–participant-rated organization
EcoGpTDivatt	Everyday cognition scale–participant-rated divided attention
EcoGpTTotal	Everyday cognition scale–participant-rated total score
EcoGSPMem	Everyday cognition scale–study-partner-rated memory
EcoGSPLang	Everyday cognition scale–study-partner-rated language
EcoGSPVisspat	Everyday cognition scale–study-partner-rated visuospatial function
EcoGSPPlan	Everyday cognition scale–study-partner-rated planning
EcoGSPOrgan	Everyday cognition scale–study-partner-rated organization
EcoGSPDivatt	Everyday cognition scale–study-partner-rated divided attention
EcoGSPTotal	Everyday cognition scale–study-partner-rated total score
mPACCdigit	Modified preclinical alzheimer's cognitive composite–digit symbol substitution test component
mPACCtrailsB	Modified preclinical alzheimer's cognitive composite–trail making test part b component (log-transformed)

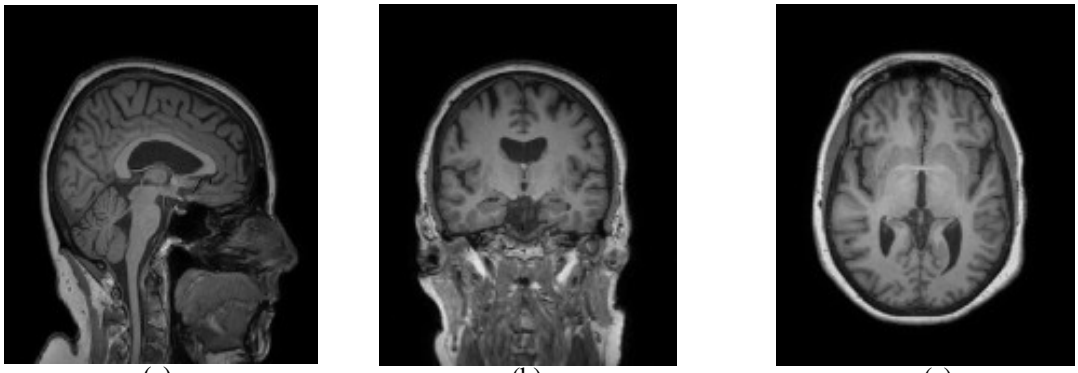
Table 4. Brain regional volume variables.

Variable Name	Description
Ventricles	Total ventricular volume (sum of lateral and third ventricles)
Hippocampus	Total hippocampal volume (left + right)
WholeBrain	Total parenchymal volume (excluding ventricles and CSF)
Entorhinal	Entorhinal cortex volume (left + right)
Fusiform	Fusiform gyrus volume (left + right)
MidTemp	Middle temporal cortex volume (left + right)
ICV	Intracranial volume (total cranial capacity)
FLDSTRENG	MRI scanner field strength (e.g., 1.5T or 3.0T)
FSVERSION	FreeSurfer software version (e.g., v5.1, v6.0)

2.2. Image data processing

Representative raw T1-weighted structural MRI scans from the ADNI dataset are displayed in Figure 1. To ensure high image quality and cross-subject comparability, all raw images were subjected to a standardized preprocessing pipeline designed to minimize non-biological variance and

enhance the reliability of subsequent quantitative analyses. The preprocessing was conducted in MATLAB (The MathWorks, Inc.) using Statistical Parametric Mapping (SPM12) [26] and the Computational Anatomy Toolbox (CAT12) [27]. The pipeline encompassed the following sequential steps: bias-field correction to remove intensity inhomogeneities; skull-stripping and tissue segmentation into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF); spatial normalization to the Montreal Neurological Institute (MNI) standard space; modulation (i.e., scaling by Jacobian determinants) to preserve regional volumetric information; and finally, atlas-based extraction of brain regional volume measures.



(a) (b) (c)

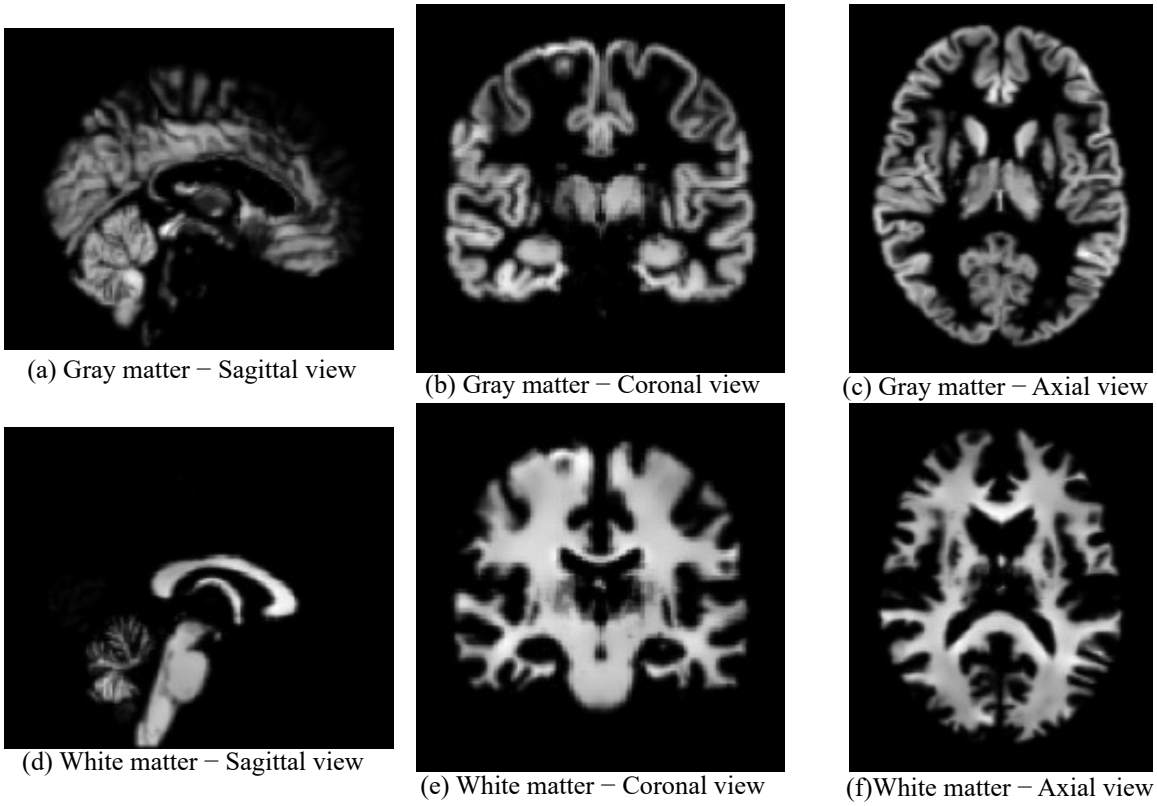
Figure 1. Raw MRI images. (a) Sagittal view; (b) Coronal view; (c) Axial view.

First, raw T1-weighted structural images were processed through skull-stripping [28] and tissue segmentation, which automatically removed non-brain tissues such as scalp, skull, and dura mater. The resulting brain parenchyma was precisely extracted and further partitioned into three tissue classes: GM, WM, and CSF. This initial step not only establishes well-defined anatomical boundaries for subsequent tissue-specific analyses but also mitigates potential interference from non-brain signals in later registration and statistical modeling.

Second, to account for inherent inter-individual variability in brain size, shape, and spatial orientation, all images underwent spatial normalization [29] to the Montreal Neurological Institute (MNI) standard space [30] via non-linear registration. This procedure aligns homologous anatomical structures across subjects within a unified coordinate framework, thereby providing a robust spatial basis for group-level comparisons and voxel-wise statistical inference.

Third, modulation was applied to the normalized tissue images to compensate for volume changes (i.e., local stretching or compression) induced by the deformation fields of spatial normalization. In this process, the intensity of each voxel is multiplied by the Jacobian determinant of the corresponding deformation, effectively correcting for scaling effects while preserving the absolute regional volumes from each subject's native space. Consequently, the modulated voxel values directly reflect true volumetric differences between individuals, enabling valid inter-subject comparisons within the standardized MNI space.

To derive the quantitative volumetric indices for individual brain regions, we adopted the Automated Anatomical Labeling 90 (AAL90) atlas [31] to parcellate the original images into 90 major cortical and subcortical structures. Derived from the widely used AAL116 atlas, the AAL90 atlas simplifies cerebellar subdivisions and merges several subregions, thereby improving parcellation efficiency and



(a) Gray matter − Sagittal view (b) Gray matter − Coronal view (c) Gray matter − Axial view

(d) White matter − Sagittal view (e) White matter − Coronal view (f)White matter − Axial view

Figure 2. Gray-matter and white-matter images obtained via computational segmentation. (a)(d) Sagittal view; (b)(e) Coronal view; (c)(f) Axial view.

stability while maintaining comprehensive anatomical coverage. Subsequently, each brain region was segmented into GM and WM; the resulting tissue-class images are presented in Figure 2. Volumetric measures for GM, WM, and CSF were then computed separately for each of the 90 regions. This process yielded a total of 270 volumetric features per subject (90 regions $\times$ 3 tissue types), providing a robust and reproducible quantitative basis for subsequent region-of-interest (ROI)-based morphometric analyses, inter-group comparisons, and clinical correlation studies.

In summary, the preprocessing pipeline employed in this study is systematic and rigorously standardized. It effectively removes non-brain tissue artifacts, achieves precise cross-subject spatial normalization, and—through modulation—preserves true volumetric information of individual brain structures. Collectively, these procedures deliver a high-quality, fully comparable imaging dataset, which lays a solid foundation for mechanistic explorations and quantitative evaluations of cognitive impairment using multimodal MRI.

2.3. Structured data preprocessing

To ensure the robustness and comparability of subsequent model-based analyses, we applied a systematic preprocessing protocol to the structured clinical and demographic features. This pipeline comprises three key steps: (i) imputation of missing values to avoid selection bias and maximize sample retention; and (ii) z-score standardization of continuous variables to eliminate scale

discrepancies, ensuring that all features contribute equally to the distance-based metrics underlying our unsupervised clustering framework.

2.3.1. Missing-value handling

For AD classification, we incorporated seven key biomarkers derived from cerebrospinal fluid and blood, each of which serves a distinct role in the underlying pathological cascade:

- **pT217_F** (phosphorylated tau 217, p-tau217): currently one of the most blood-based biomarkers with high specificity for AD, p-tau217 reflects the extent of cerebral tau pathology. Its concentration is typically low (<0.2 pg/mL) in cognitively normal individuals but markedly elevated in AD patients.
- **Aβ42_F** (amyloid-β42): as the primary component of senile plaques, Aβ42 levels in cerebrospinal fluid (CSF) decline early in AD. Healthy subjects exhibit CSF Aβ42 concentrations of 800–1200 pg/mL, whereas values below 600 pg/mL are frequently observed in AD patients.
- **Aβ40_F** (amyloid-β40): the most abundant Aβ isoform in CSF, Aβ40 alone has limited diagnostic utility; however, its ratio to Aβ42 is clinically informative. Normal CSF Aβ40 levels typically range from 5000 to 12000 pg/mL.
- **Aβ42_Aβ40_F** (Aβ42/Aβ40 ratio): this ratio provides greater stability than standalone Aβ42 and partially corrects for inter-individual baseline variations. In healthy individuals, it ranges from 0.06 to 0.12, and decreases significantly in AD.
- **pT217_Aβ42_F** (p-tau217/Aβ42 ratio): recent evidence indicates that this ratio improves diagnostic accuracy for AD. Although no universal reference range exists, a value below 0.02 is commonly adopted as the threshold for the negative classification.
- **NfL_Q** (neurofilament light chain, NfL): NfL reflects neuronal axonal injury and serves as a non-specific marker of neurodegeneration. In healthy older adults, blood NfL concentrations are generally below 15–20 pg/mL, but are substantially elevated in patients with neurodegenerative disorders.
- **GFAP_Q** (glial fibrillary acidic protein, GFAP): GFAP indicates astrocyte activation and reactive gliosis. Circulating GFAP levels in healthy individuals are usually below 150–200 pg/mL, with marked elevations observed in AD patients, particularly among those positive for Aβ.

For the seven AD-related biomarkers described above, the physiological values are expected to be non-negative. During the data cleaning process, any raw value coded as -4 (indicating invalid or anomalous measurements) was replaced with missing (NaN). Additionally, any negative readings—which may arise from acquisition, transmission, or recording errors—were also treated as invalid and converted to missing to prevent distortion of downstream analyses.

For the classification task, which relies on cross-sectional data, we devised a systematic missing-data imputation strategy. First, we applied the K-nearest neighbor (KNN) algorithm [32] as the primary imputation method. For a sample with missing entries, KNN identifies the k complete samples most similar to it in the feature space and estimates the missing value by averaging (or weighted averaging) the corresponding feature values from these neighbors. The imputation estimate is given by

$$\hat{x}_{ij} = \frac{1}{k} \sum_{t \in N_k(i)} x_{tj}, \tag{1}$$

where $\hat{x}_{ij}$ is the imputed value for the i-th sample and j-th feature, $N_k(i)$ is the index set of the k nearest complete samples, and x_{tj} is the observed value of the t-th neighbor on feature j. For samples that lacked sufficient valid neighbors, we sequentially applied diagnostic-group–specific mean imputation, followed by global mean imputation, to ensure complete feature vectors for all subjects. All imputed values were then checked for non-negativity, ensuring that the final dataset preserves the physiological plausibility of the biomarkers.

For the MCI subtype clustering analysis, which incorporates a broader set of multi-modal features—including demographics (AGE, PTEDUCAT, APOE4), cognitive scales (MMSE, ADAS13, RAVLT series, CDRSB), self-reported functional assessments (FAQ, Ecog series), and neuroimaging measures (ventricular, hippocampal, and whole-brain volumes)—we first defined theoretical validity ranges for each variable based on clinical knowledge (e.g., MMSE scores range from 0 to 30). Out-of-range values were recoded as missing to avoid distorting the clustering structure.

To handle missingness in this multi-modal longitudinal dataset, we developed a hierarchical imputation strategy that respects both temporal and cross-sectional dependencies. For each patient, we preferentially applied forward-fill and backward-fill along the visit timeline to preserve individual disease trajectories. For features that remained missing after this step, we employed KNN imputation using complete cases from the same time point. When local matching was insufficient, we fell back to diagnostic-group–specific or global mean imputation. Finally, all imputed values were clipped to their theoretical bounds to maintain clinical validity. This protocol integrates temporal continuity with feature correlation, yielding a robust multi-modal input for subsequent MCI subtype clustering and ensuring the stability of the resulting models.

2.3.2. Data standardization

After the imputation and outlier removal, we standardized all structured features using Z-score normalization to eliminate the scale-induced biases arising from different units and value ranges. The transformation is given by

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}, \tag{2}$$

where x_{ij} is the original value of the j-th feature for the i-th sample, μ_j and σ_j are the sample mean and standard deviation of that feature, respectively, and z_{ij} is the standardized value. This procedure centers each feature at zero and scales it to unit variance, ensuring that all features contribute equally to distance-based metrics in subsequent modeling.

2.4. Architecture of the attention-enhanced OG-DTC model

This subsection introduces Attention-OG-DTC, an enhanced deep clustering approach for MCI subtyping that integrates an attention module into the original OG-DTC framework [14]. The model distinguishes between two types of input features: (i) time-invariant covariates, including demographic and baseline clinical measures that remain constant throughout the study; and (ii) time-variant longitudinal indicators that evolve over follow-up. A schematic overview of the proposed architecture is provided in Figure 3.

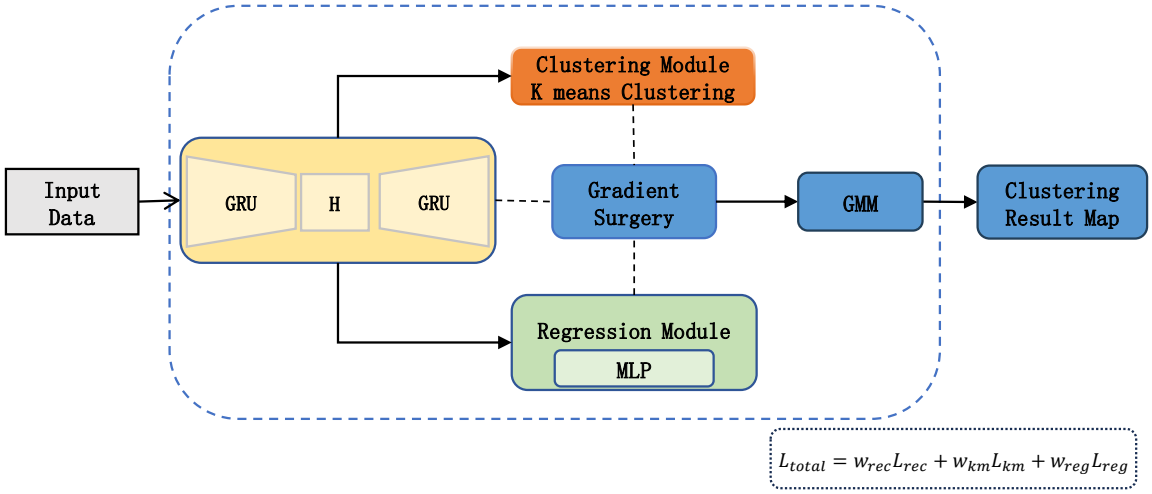


Figure 3. OG-DTC architecture diagram.

The proposed attention-OG-DTC model consists of three synergistic modules: (i) a sequence-to-sequence reconstruction module based on a GRU autoencoder; (ii) a K-means clustering loss module to enforce compact cluster structures; and (iii) an outcome-guided module that aligns latent representations with clinical endpoints.

The reconstruction module uses a GRU-based encoder-decoder with a sequence-to-sequence architecture. The encoder stacks two GRU layers with batch normalization to capture temporal dynamics and multi-scale patterns from both time-invariant features (e.g., sex, genetic markers) and time-varying clinical indicators (e.g., MMSE, neuropsychological scale scores), while a single-layer GRU decoder reconstructs original input sequences. This reconstruction objective ensures learned latent embeddings preserve comprehensive temporal information.

The clustering loss module addresses the limitation that representations derived from standard sequence-to-sequence models lack inherent cluster structures. By incorporating a K-means loss into the training objective, we explicitly encourage compact and well-separated clusters in the embedding space, thereby grouping patients with similar progression trajectories into distinct subtypes.

The outcome-guided module leverages clinical endpoints—quantitative measures of disease severity and treatment response—to further shape the latent space. By predicting downstream outcomes such as ADAS-Cog and CDR-SB scores, the model identifies subgroups with distinct progression profiles (e.g., rapid versus slow cognitive decline, varying dementia severity) that are clinically meaningful. This guidance ensures that the learned representations effectively discriminate across clinically relevant endpoints.

Although the GRU is capable of handling variable-length sequences, the original OG-DTC framework tends to lose early critical information (e.g., baseline cognitive impairment) when compressing long sequences and lacks sensitivity to abrupt deterioration points. To overcome this limitation, we augment the architecture with a multi-head self-attention layer positioned after the GRU encoder. This attention mechanism dynamically weights each time step, enabling the model to adaptively emphasize the most discriminative temporal features during representation learning. The

derived latent representations are thus optimized for both subtype discovery and accurate progression prediction. An overview of the proposed architecture is shown in Figure 4.

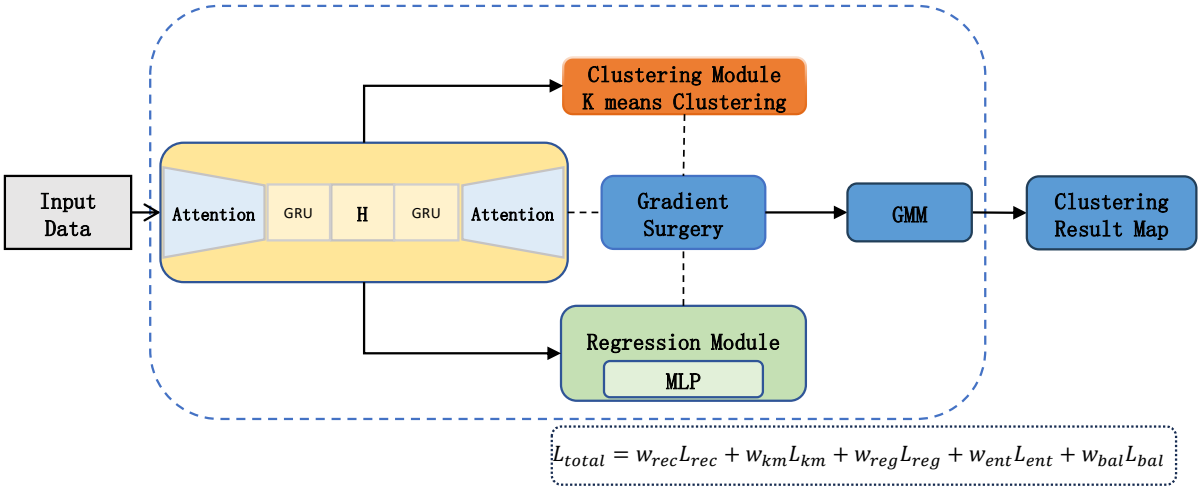


Figure 4. Attention-OG-DTC architecture diagram.

The baseline model (OG-DTC) consists of a GRU encoder, a GRU decoder, and two prediction heads (a clustering head and a regression head). For the input sequence $X_i \in \mathbb{R}^{T_i \times D}$ of the i-th patient, the encoder computes the hidden state via recurrent calculation:

$$h_t = \mathrm{GRU}(x_t, h_{t-1}). \tag{3}$$

The baseline model only adopts the final-time-step hidden state h_{T_i} as the fixed feature vector z_i for each patient. Nevertheless, this approach exhibits obvious limitations: for long-sequence data, early subtle pathological changes (e.g., mild cognitive decline at baseline) tend to be forgotten, resulting in information loss. To address the above limitation, our improved model (Attention-OG-DTC) introduces a temporal attention mechanism. To capture the importance of different time points throughout the disease course, a multi-head self-attention layer is appended after the GRU layers. Instead of solely relying on the final hidden state h_{T_i}, the model computes attention weights using the complete hidden-state sequence $H = \{h_1, \ldots, h_{T_i}\}$:

$$Q = HW_Q, \quad K = HW_K, \quad V = HW_V, \tag{4}$$

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V. \tag{5}$$

The final latent representation z_i is obtained by weighted aggregation of features across all time steps:

$$z_i = \mathrm{Pool}(\mathrm{Attention}(H)). \tag{6}$$

This enables the model to adaptively focus on time windows most predictive of disease progression and enhances feature representation capability.

The total loss function L_{total} of OG-DTC is composed of temporal reconstruction loss, deep spectral clustering loss, and clinical outcome regression loss. Based on the original loss terms, the proposed Attention-OG-DTC further introduces two additional clustering regularization terms. The total loss function is defined as shown in Eq (7):

$$L_{\text{total}} = w_{\text{rec}}L_{\text{rec}} + w_{km}L_{km} + w_{\text{reg}}L_{\text{reg}} + w_{\text{ent}}L_{\text{ent}} + w_{\text{bal}}L_{\text{bal}}. \tag{7}$$

where w_{rec}, w_{km}, w_{reg}, w_{ent}, and w_{bal} denote balancing hyper-parameters for each task. The temporal reconstruction loss L_{rec}, given in Eq (8), ensures that the attention-weighted latent representation h_i still preserves critical temporal evolution information from the original sequence. Considering variable follow-up lengths among patients, masked mean squared error is adopted to ignore the contribution of padded entries.

$$L_{\text{rec}} = \frac{1}{\sum M_{i,t}} \sum_{i=1}^{B} \sum_{t=1}^{T} M_{i,t} \left\| x_{i,t} - \hat{x}_{i,t} \right\|_2^2 . \tag{8}$$

Here, B denotes the batch size and T denotes the maximum time-step length. $x_{i,t}$ and $\hat{x}_{i,t}$ represent the raw input and reconstructed output of the i-th sample at time step t, respectively. $M_{i,t} \in \{0, 1\}$ is the mask indicator variable.

The deep spectral clustering loss L_{km} (Eq (9)) is designed to discover nonlinear disease subtype structures within the attention-enhanced latent space. This loss relaxes the spectral-clustering objective to maximize intra-cluster affinity while minimizing orthogonal error:

$$L_{km} = L_{\text{trace}} + \lambda_{\text{orth}}L_{\text{orth}}, \tag{9}$$

where

$$L_{\text{trace}} = \frac{\text{Tr}(HH^{\text{T}}) - \text{Tr}(Y^{\text{T}}HH^{\text{T}}Y)}{B^2}. \tag{10}$$

Based on the relaxed form of the min-cut criterion, the model maximizes the similarity among samples within the same cluster; $\boldsymbol{H} \in \mathbb{R}^{B\times d}$ denotes the latent representation matrix within a batch, and $\boldsymbol{Y} \in \mathbb{R}^{B\times K}$ is the soft-assignment matrix. The orthogonal regularization term $L_{orth} = \left\| \boldsymbol{Y}^{\text{T}}\boldsymbol{Y} - I_K \right\|_F^2$ enforces column-orthogonality for the clustering assignment matrix to avoid inter-cluster overlap.

The clinical regression loss L_{reg} endows the discovered subtypes with clinical interpretability. The regressor $f_{reg}(\cdot)$ predicts clinical scores at the last follow-up (e.g., ADAS13 and CDRSB), as formulated in Eq (11):

$$L_{reg} = \frac{1}{B} \sum_{i=1}^{B} \left\| y_i - f_{reg}(h_i) \right\|_2^2 , \tag{11}$$

where y_i is the ground-truth clinical-score vector, and h_i is the latent vector output from the attention mechanism. This loss ensures that the latent space encodes disease severity and progression trajectories.

To improve the clustering stability and quality, two auxiliary loss terms are introduced in this work. The entropy-minimization loss is given in Eq (12)):

$$L_{ent} = -\frac{1}{B} \sum_{i=1}^{B} \sum_{k=1}^{K} Y_{i,k} \log(Y_{i,k} + \delta), \tag{12}$$

where K denotes the pre-defined number of clusters, B is the batch size, $Y_{i,k}$ represents the soft-assignment probability that the i-th sample belongs to the k-th cluster, and δ is a tiny constant to avoid numerical instability caused by log(0) in logarithmic computation. This term encourages the model to produce high-confidence clustering assignments.

The class-balance loss is shown in Eq (13):

$$L_{bal} = \frac{1}{K} \sum_{k=1}^{K} \left(p_k - \frac{1}{K} \right)^2, \tag{13}$$

which prevents trivial solutions, where $p_k = \frac{1}{B} \sum_{i=1}^{B} Y_{i,k}$, B denotes the batch size, and K is the pre-specified number of clusters.

Considering potential competing relationships among the above multiple tasks, simple weighted summation tends to yield suboptimal solutions. This study adopts the PCGrad algorithm. Before each parameter update, gradient vectors for all tasks are computed. If the angle between gradient vectors of two tasks exceeds 90° (i.e., negative cosine similarity), the gradient of one task is projected onto the normal plane of the other gradient to eliminate conflicting components, so as to guarantee stable model convergence across all tasks.

The model was trained using the Adam optimizer with a learning rate of 4.9×10^{-3}, a batch size of 64, and 30 training epochs. The encoder GRU hidden layer dimension was 256, the latent representation dimension was 64, the number of GRU layers was 3, and the number of attention heads was 2. The number of clusters k was set to 2. The weights of the five losses were as follows: reconstruction loss 1.0, clustering loss 1.4, regression loss 0.47, orthogonality regularization 1.1, and class balance 0.1; the entropy minimization loss was set to 0.1 for the first 9 epochs and to 0 thereafter. Pretraining was performed by optimizing only the reconstruction and regression losses, after which the clustering loss and PCGrad were introduced for multi-task joint optimization. During training, the clustering temperature was set to 2.0 for the first 9 epochs and then restored to 1.0; the clustering loss weight was reduced to 0.7 for the first 9 epochs and then restored to 1.4. All hyperparameters were kept consistent across five-fold cross-validation, and the random seed was fixed at 42.

2.5. Feature selection

To mitigate the interference of high-dimensional data on clustering performance and screen out the most discriminative features, feature importance screening is performed prior to model training in this section. Considering that deep clustering adopts unsupervised data without external labels, this study employs a comprehensive unsupervised feature-selection framework. Two complementary metrics, namely the Laplacian score and PCA communality, are combined to identify the most discriminative features from high-dimensional datasets.

Laplacian score is a manifold-preserving feature selection method. Its core idea is to evaluate how well each feature contributes to preserving the nearest-neighbor graph structure of data. For a dataset containing n samples, we first construct a graph G that reflects the adjacency relationships among samples. The adjacency matrix $\boldsymbol{S}$ is generally defined via K-nearest neighbors or the thermal kernel

function, as shown in Eq (14):

$$S_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & \text{if } x_i \text{ and } x_j \text{ are neighbors,} \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Define the degree matrix $\boldsymbol{D}$ as a diagonal matrix whose diagonal entries satisfy $D_{ii} = \sum_j S_{ij}$. Then the graph Laplacian matrix is $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{S}$. For the r-th feature $\boldsymbol{f}_r = [f_{r1}, f_{r2}, \ldots, f_{rm}]^{\mathrm{T}}$, its Laplacian Score is defined in Eq (15):

$$L_r = \frac{\sum_{ij}(f_{ri} - f_{rj})^2 S_{ij}}{\mathrm{Var}(\boldsymbol{f}_r)} = \frac{\tilde{\boldsymbol{f}}_r^{\mathrm{T}} \boldsymbol{L} \tilde{\boldsymbol{f}}_r}{\tilde{\boldsymbol{f}}_r^{\mathrm{T}} \boldsymbol{D} \tilde{\boldsymbol{f}}_r}, \tag{15}$$

where $\tilde{\boldsymbol{f}}_r$ denotes the feature vector after mean centering. This score measures the smoothness of a feature within local neighborhoods. If neighboring samples have similar values on feature $\boldsymbol{f}_r$, the term $(f_{ri} - f_{rj})^2$ will be small, leading to a low L_r. This indicates that the feature well preserves the local geometric structure of data. Accordingly, features with lower Laplacian Scores are preferred during feature screening.

PCA communality evaluates feature importance from the perspective of global variance. Derived from factor analysis theory, communality within the principal component analysis framework represents the proportion of variance of an original variable that can be explained by the first m principal components. We perform principal component decomposition on the raw data matrix $\boldsymbol{X} \in \mathbb{R}^{n \times d}$ and obtain the principal component loading matrix $\boldsymbol{A} \in \mathbb{R}^{d \times m}$ (where the first m principal components are typically selected such that the cumulative explained variance ratio exceeds a predefined threshold). For the r-th feature, its communality h_r^2 is calculated as Eq (16):

$$h_r^2 = \sum_{j=1}^{m} a_{rj}^2, \tag{16}$$

where a_{rj} is the loading coefficient of the r-th feature on the j-th principal component. Communality ranges within $[0, 1]$. A value of h_r^2 closer to 1 implies that the variance of this feature is better captured by the global principal component structure, which corresponds to higher importance in terms of global variance.

The Laplacian score focuses on the local manifold preservation capability of features, while PCA communality emphasizes the global variance contribution of features. By integrating these two complementary metrics, this study aims to screen key features that can not only preserve the intrinsic geometric structure of data but also explain most of the variance, so as to provide an optimized feature subset for the subsequent clustering analysis.

The feature selection procedure adopted in this work is described as follows. First, the feature importance metrics of all variables on the complete dataset are calculated. Next, variables whose composite metric values are lower than the predefined threshold (0.6) are eliminated. It should be noted that the threshold of 0.6 is an empirical threshold determined after the normalized fusion of the two indicators. Its rationality is indirectly verified by the clustering stability from subsequent 5-fold cross-validation: the 20 features retained under this threshold yield a silhouette coefficient greater than 0.89 and a Davies-Bouldin (DB) index less than 0.15 across all folds, indicating a good match between the feature subset and the clustering structure. Finally, 20 variables are retained from the original 42

variables. In addition, since the K-means clustering module needs to be trained within the model, the cluster number k should be specified prior to model training. To this end, the elbow method curve is plotted to determine the optimal k, as illustrated in Figure 5. It can be observed from Figure 5 that the decreasing rate of the sum of squared errors (SSE) tends to flatten out when $k = 2$. Accordingly, we set $k = 2$ for the K-means module in the proposed deep clustering framework.

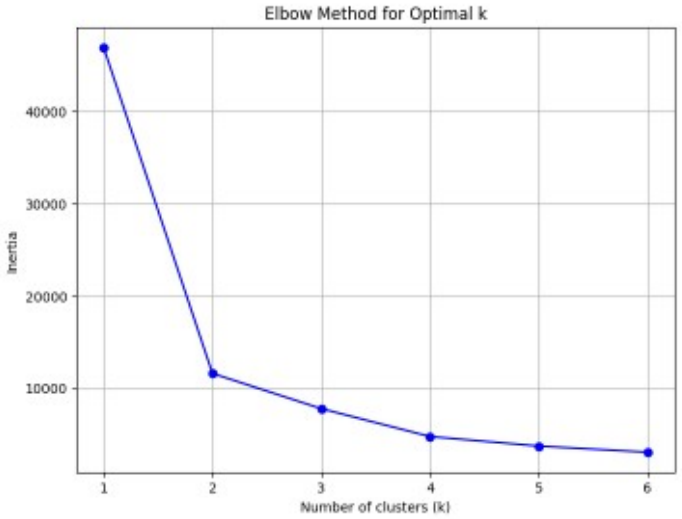


Figure 5. Elbow plot for K-means clustering.

3. Results

To enhance the capability of the OG-DTC model to capture temporal information, this work integrates an attention mechanism into the GRU module of the original architecture, enabling the model to effectively focus on critical time steps. To comprehensively evaluate the effectiveness of the proposed improvement, we conduct comparative analyses between the improved model and the original model from two perspectives: model stability and performance gain.

First, five-fold cross-validation is utilized to assess the stability of the Attention-OG-DTC model, as listed in Table 5. The five-fold cross-validation results demonstrate that the optimized model achieves favorable clustering performance across all folds. The values of all evaluation metrics are closely distributed, revealing satisfactory consistency and reliability. Specifically, the silhouette coefficient exceeds 0.89 on every fold and reaches a maximum of 0.96, which indicates satisfactory intra-cluster compactness and inter-cluster separation of clustering outputs. DBI remains below 0.15 across all folds, with a minimum of 0.075, suggesting prominent discriminability between different clusters. The Calinski-Harabasz index (CHI) is higher than 6700 for all folds and peaks near 28,500, further verifying the compactness and stability of the clustering results. The concentrated distribution and small standard deviation of the three metrics across different folds demonstrate that the proposed model possesses strong stability and generalization ability.

Meanwhile, to quantify the improvement brought by the attention module, we compare the average five-fold cross-validation performance of the Attention-OG-DTC model with that of the original OG-

Table 5. Results of five-fold cross-validation.

Fold	Silhouette coefficient	DBI	CHI
1st	0.930	0.101	17,918.5
2nd	0.960	0.075	28,408.9
3rd	0.951	0.077	27,155.5
4th	0.898	0.143	6728.3
5th	0.947	0.078	24,808.2

DTC model. As presented in Table 6, the model equipped with the attention mechanism achieves consistent improvements on all clustering metrics.

Table 6. Evaluation metrics of clustering models.

Model	Silhouette coefficient	DBI	CHI
OG-DTC	0.760	0.325	1964.76
Attention-OG-DTC	0.937	0.095	21,003.9

As observed from Table 6, the average silhouette coefficient, CHI and DBI of the original OG-DTC without the attention mechanism are 0.760, 1964.76 and 0.325, respectively. After incorporating the attention mechanism, the corresponding metrics are updated to 0.937, 21,003.9 and 0.095. The silhouette coefficient increases by 0.177, implying stronger intra-cluster homogeneity and clearer subtype partition. The DBI decreases by 0.23, which reflects higher inter-subtype separation and reduced cluster overlap. These results confirm that the improved model outperforms the original counterpart on all evaluation metrics and can effectively capture information from critical time steps.

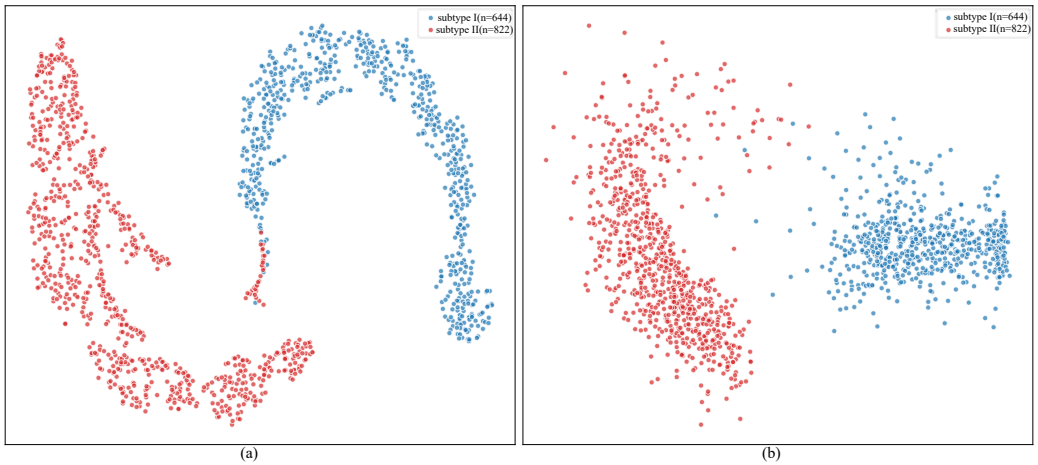


Figure 6. Visualization of OG-DTC clustering results. (a) t-SNE visualization; (b) PCA visualization.

To intuitively demonstrate the clustering performance, two dimensionality reduction methods, namely principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), are adopted to conduct visual analysis on the features extracted by the model. An ideal clustering outcome should exhibit a distribution pattern with high intra-cluster compactness and distinct inter-cluster separation in the low-dimensional projection space.

Figure 6 reveals that the sample points of each cluster obtained by the original OG-DTC model are relatively scattered within the PCA projection space. In particular, obvious dispersion exists inside Subtype I and Subtype II, and partial overlapping regions appear at inter-cluster boundaries. This phenomenon indicates that the original model fails to achieve favorable class separability in feature learning.

As illustrated in Figure 7, the improved model equipped with the attention mechanism achieves prominent advantages in both intra-class aggregation and inter-class separation. Samples belonging to the same cluster gather more tightly in the low-dimensional space, and the gaps between different subtypes become clearer, demonstrating a stronger capability to learn discriminative features. This result visually verifies the effectiveness of the attention mechanism in enhancing the discriminative power of the model for clustering.

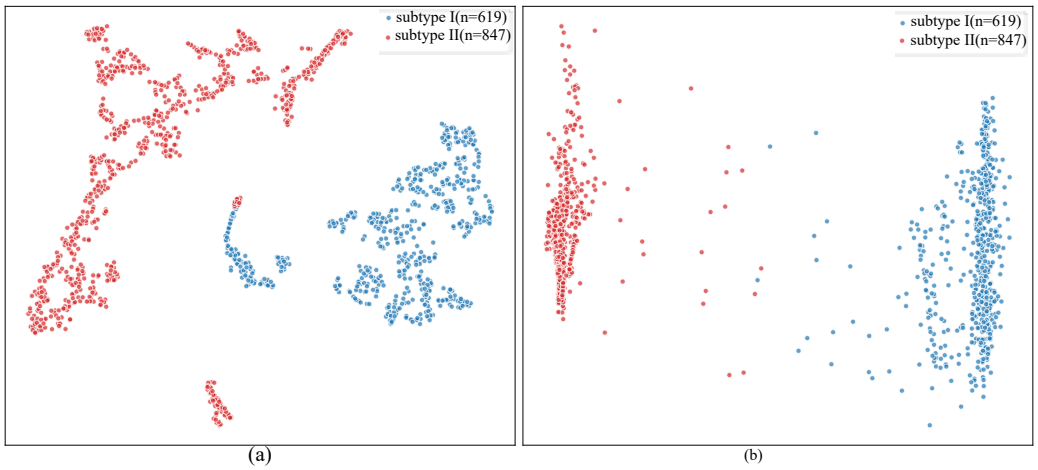


Figure 7. Visualization of clustering results from the improved OG-DTC. (a) t-SNE visualization; (b) PCA visualization.

The above visualization results demonstrate that the improved model equipped with the attention mechanism can effectively distinguish the two subgroups of patients. Superior clustering performance is reflected not only by sample separability in the mathematical feature space but also by whether the identified patient subgroups exhibit statistically significant differences from a clinical perspective. Therefore, we further conduct statistical comparisons of the mean values of various clinical indicators between the two patient subtypes and plot their longitudinal trajectories during follow-up, so as to verify the biological rationality and interpretability of the clustering results from a clinical dimension. Meanwhile, independent two-sample t-tests are performed to identify significant inter-subtype differences, and the false discovery rate (FDR) correction is adopted to control the false positive rate.

As summarized in Table 7, Subtype I exhibits worse performance on all RAVLT indicators. In terms of the forgetting score, Subtype I achieves a value of 4.96, which is significantly higher than the value of 3.93 for Subtype II (adjusted $p < 0.01$), indicating poorer memory retention. For the learning ability index, Subtype I obtains a score of 2.70, significantly lower than 5.65 for Subtype II (adjusted $p < 0.01$), suggesting more severe impairment in information encoding. Regarding the percentage of forgetting, the value reaches 82.87% for Subtype I, substantially higher than 40.36% for Subtype II (adjusted $p < 0.01$), further confirming severe deficits in memory consolidation.

Table 7. Overall levels of memory function-related indicators.

Variable	Subtype I	Subtype II	P-Value
RAVLT_forgetting	4.96	3.93	< 0.01
RAVLT_learning	2.70	5.65	< 0.01
RAVLT_perc_forgetting	82.87	40.36	< 0.01

Figure 8 illustrates the longitudinal trajectories of memory function indicators during follow-up. For memory retention, the RAVLT_forgetting score of Subtype I remains persistently high within the range of 4.5–5.5, whereas Subtype II maintains a relatively intact level around 4.0. In terms of learning ability, the RAVLT_learning score of Subtype I stays at a low level of 2.5–3.0, while Subtype II steadily maintains a high level of 5.5–6.0. The discrepancy between the two patient groups persists throughout the follow-up period. With respect to memory consolidation, the RAVLT_perc_forgetting of Subtype I remains above 80% and presents a fluctuating upward trend over follow-up time. By contrast, Subtype II maintains a relatively stable level around 40%.

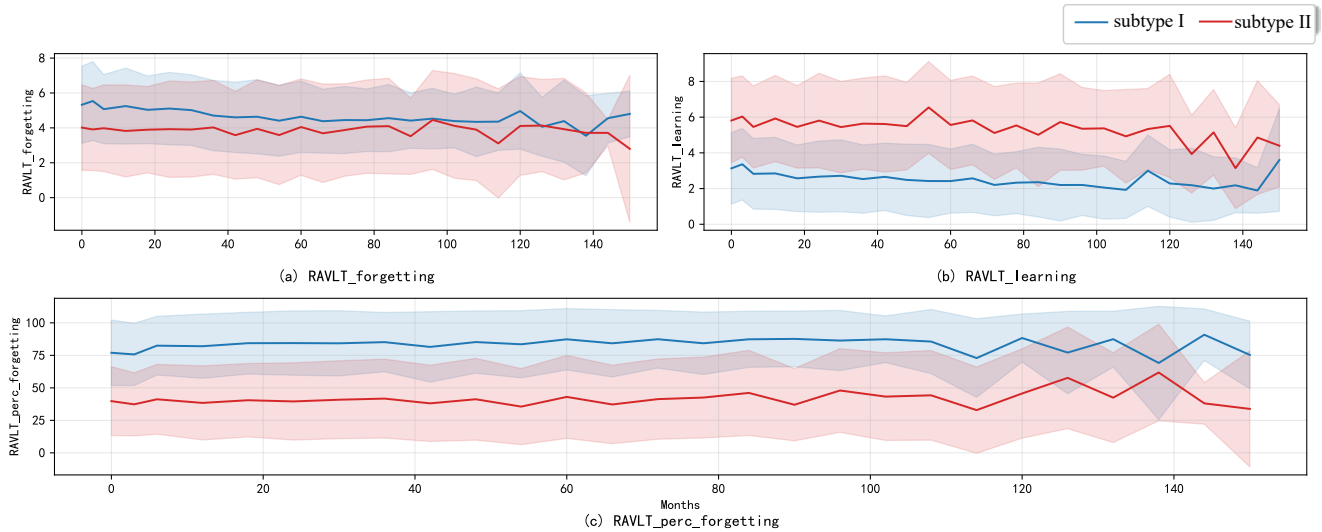


Figure 8. Line charts illustrating longitudinal changes in memory function-related indicators during follow-up.

The above results reveal that patients in Subtype I exhibit a pattern characterized by persistently impaired learning ability, rapid memory decay, and poor functional stability, accompanied by more severe progressive deterioration in learning capacity and memory consolidation.

4. Discussion

The identified MCI subtypes exhibited distinct clinical functional profiles in daily-living capacity and cognitive performance. As presented in Table 8, Subtype I demonstrates worse performance in activities of daily living and executive function. In terms of daily function, the FAQ score of Subtype I reaches 8.12, which is significantly higher than the value of 1.45 in Subtype II (adjusted $p < 0.01$), indicating more severe impairment in activities of daily living. Regarding executive function, for

the sustained attention task, Subtype I achieves an mPACCdigit score of −11.13, significantly lower than −1.40 in Subtype II (adjusted $p < 0.01$). In the task-switching test, the mPACCtrailsB score of Subtype I is −10.36, markedly lower than −1.20 observed in Subtype II (adjusted $p < 0.01$). For the completion time of Trail Making Test Part B, Subtype I requires 145.76 seconds, substantially longer than the 85.44 seconds recorded for Subtype II (adjusted $p < 0.01$).

Table 8. Overall levels of indicators related to daily function and executive ability.

Variable	Subtype I	Subtype II	P-Value
FAQ	8.12	1.45	< 0.01
mPACCdigit	−11.13	−1.40	< 0.01
mPACCtrailsB	−10.36	−1.20	< 0.01
TRABSCOR	145.76	85.44	< 0.01

As illustrated in Figure 9, for daily function, the FAQ score of Subtype I increases from approximately 4.8 to above 10, while Subtype II remains stable around 2.5. In terms of sustained attention, the mPACCdigit score of Subtype I ranges from −11 to −15 with a continuous declining trend, whereas Subtype II maintains a stable level between −1.5 and −2.5. For task switching, the mPACCtrailsB score of Subtype I fluctuates between −10 and −15 and keeps decreasing, and Subtype II stays steadily within −1 to −2. With respect to complex task processing, the TRABSCOR value of Subtype I rises from 120 seconds to 170 seconds, in contrast to Subtype II, which remains stable at roughly 85 seconds.

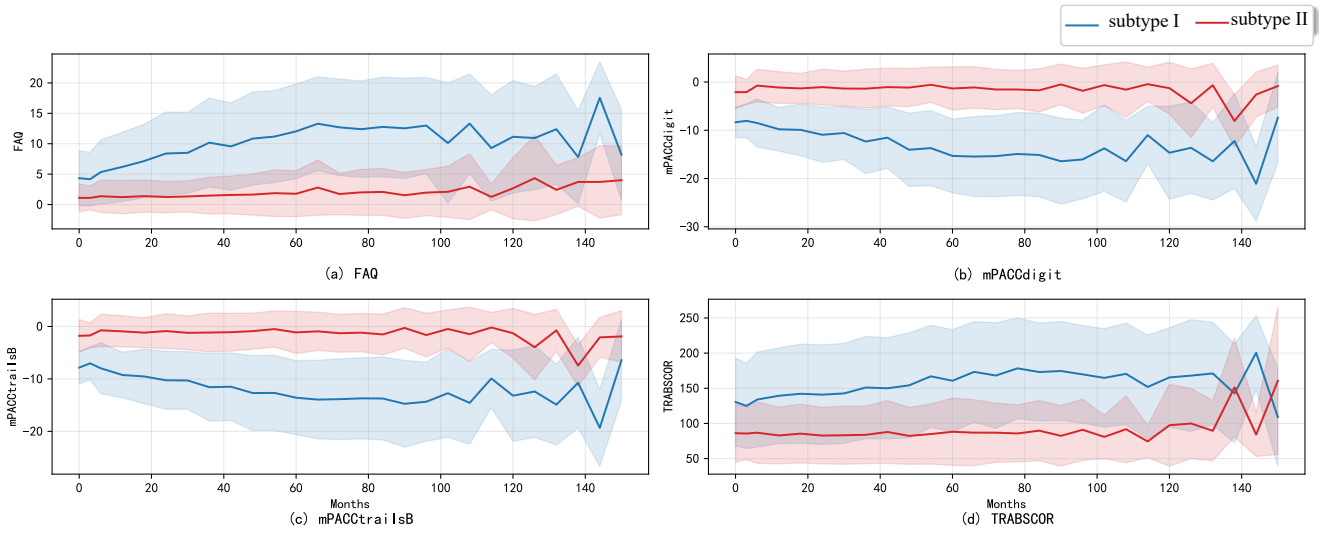


Figure 9. Line chart of indicators related to daily function and executive ability.

As shown in Table 9, Subtype I exhibited worse performance across all global cognitive measures. The MMSE score of Subtype I was 25.22, which was significantly lower than the value of 28.72 observed in Subtype II (adjusted $p < 0.01$). Higher scores on ADAS11, ADAS13, ADASQ4, and CDRSB indicate greater cognitive impairment. Subtype I achieved scores of 14.96, 23.93, 7.40, and 3.06 for these scales, all significantly higher than the corresponding values of 6.40, 10.01, 3.13, and

0.85 in Subtype II (all adjusted $p < 0.01$).

Table 9. Global levels of global cognitive assessment indicators.

Variable	Subtype I	Subtype II	*P*-Value
MMSE	25.22	28.72	< 0.01
ADAS11	14.96	6.40	< 0.01
ADAS13	23.93	10.01	< 0.01
ADASQ4	7.40	3.13	< 0.01
CDRSB	3.06	0.85	< 0.01

Figure 10 illustrates global cognitive status trajectories. The MMSE score of Subtype I declined from approximately 27 at baseline to around 25 at the final follow-up, whereas Subtype II remained stably preserved at a high level of 28–30 points. With regard to core cognitive domains, Subtype I maintained consistently elevated scores on ADAS11, ADAS13, ADASQ4, and CDRSB, ranging from 10–20, 20–30, 6–9, and 2–5 points, respectively. In contrast, Subtype II remained stably low on these measurements, with ranges of 5–6, 8–10, 2–3, and 0.5–1 points, indicating relatively intact cognitive function.

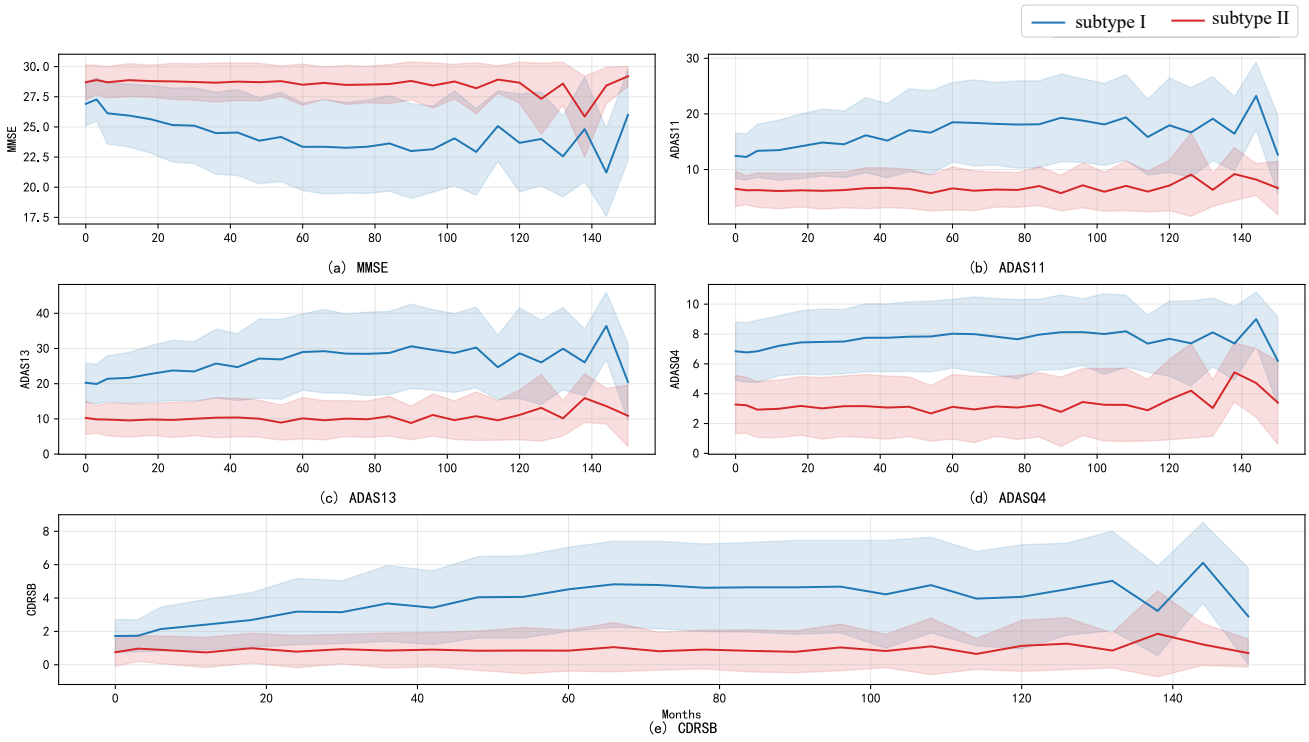


Figure 10. Line chart showing trajectories of comprehensive cognitive assessment indicators.

As presented in Table 10, Subtype I exhibited higher scores on all ECoG measures relative to Subtype II. Regarding informant-rated assessments, Subtype I obtained scores of 1.91, 1.80 and 1.47 for divided attention, global cognition and visuospatial ability, respectively, which were higher than the corresponding values of 1.88, 1.67 and 1.33 in Subtype II. The between-group difference in divided attention reached statistical significance (adjusted $p = 0.023$), and the adjusted p values for the remaining two indices were both less than 0.01. For patient self-rated assessments, Subtype I demonstrated significantly higher scores in organization, planning and global cognition (2.35, 2.13 and 2.35) compared with Subtype II (1.46, 1.39 and 1.52; all adjusted $p < 0.01$).

These findings indicate that Subtype I experienced greater difficulties across multiple daily cognitive domains including divided attention, visuospatial ability, organizational ability and planning ability, based on both informant ratings and patient self-reports.

Table 10. Global levels of self-rated and informant-rated cognitive function.

Variable	Subtype I	Subtype II	*P*-Value
EcogPtDivatt	1.91	1.88	0.023
EcogPtTotal	1.80	1.67	< 0.01
EcogPtVisspat	1.47	1.33	< 0.01
EcogSPOrgan	2.35	1.46	< 0.01
EcogSPPlan	2.13	1.39	< 0.01
EcogSPTotal	2.35	1.52	< 0.01

Table 11 shows that Subtype I had smaller volumes in key brain regions than Subtype II. The MidTemp volume of Subtype I was $18,512.90\ mm^3$, lower than the value of $20,452.93\ mm^3$ in Subtype II. The Fusiform volume of Subtype I was $16,609.06\ mm^3$, compared with $18,214.41\ mm^3$ in Subtype II. The between-group differences in both regions were statistically significant (all adjusted $p < 0.01$). These results suggest that patients in Subtype I exhibited more pronounced neurodegenerative alterations at the brain structural level, further supporting the pathological basis of their cognitive impairment.

Table 11. Global levels of brain structural indices.

Variable	Subtype I	Subtype II	*P*-Value
Fusiform	16609.06	18214.41	< 0.01
MidTemp	18512.90	20452.93	< 0.01

Consistent differences across the above multi-dimensional indicators systematically reveal more severe cognitive and functional impairment in Subtype I from multiple perspectives, including learning and memory, executive function, global cognitive function, activities of daily living, and brain structural changes. These results verify the biological plausibility and clinical interpretability of the clustering outputs derived from the present model.

To further visualize longitudinal trajectories of disease status among patients in distinct clusters, a disease state transition trajectory plot was generated in Figure 11.

As illustrated in Figure 11, which depicts disease-state transition trajectories for the two subgroups, Subtype I and Subtype II exhibited markedly divergent patterns of clinical progression. Patients within Subtype I showed evident progressive cognitive decline: the large majority of participants transitioned gradually from MCI to dementia over follow-up, with no observed reverse transitions from dementia back to MCI or cognitively normal (CN) status. This suggests that patients

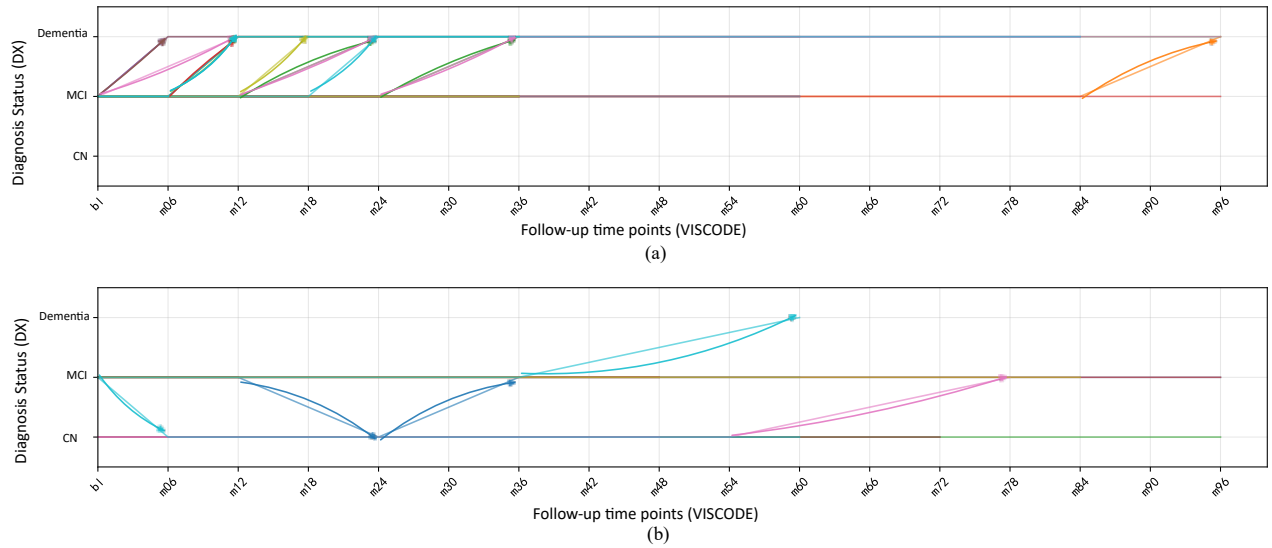


Figure 11. Longitudinal disease state transition trajectories of patients in different clusters. (a) Temporal changes in Subtype I; (b) Temporal changes in Subtype II.

in this subtype not only presented severe cognitive impairment at baseline, but their disease course was characterized by unidirectional and irreversible neurodegeneration.

In contrast, disease trajectories of Subtype II patients revealed greater heterogeneity and a relatively favorable prognosis. Most patients remained in the MCI state, and some even reverted from MCI to CN, suggesting that their cognitive impairment possessed a certain degree of stability or reversibility; only a small proportion progressed to AD. Overall, patients in Subtype II experienced slower disease progression, with more trajectories tending toward stabilization or improvement.

In summary, Subtype I and Subtype II identified by the present model differed not only in cross-sectional characteristics but also demonstrated substantially divergent longitudinal disease trajectories. Subtype I represented a high-risk neurodegenerative group with rapid progression, whereas Subtype II consisted of stable or reversible individuals with slower progression and relatively favorable outcomes. These findings further validate the capacity of the model to identify subtypes with distinct disease courses and provide important evidence for early intervention and prognostic evaluation. To further evaluate disparities in disease progression and clinical prognosis between the two subtypes, survival analysis was conducted based on clustering results. Figure 12 displays survival curves over time for patients in the two subgroups, which were constructed using the Kaplan-Meier method (product-limit estimator). This method estimates the survival function $S(t)$ to describe the probability that an individual remains event-free beyond time t. Assume there are n samples with k distinct event time points $t_1 < t_2 < \cdots < t_k$. At each time point t_i, d_i events (disease progression) occur, and n_i denotes the number of individuals at risk prior to t_i. The conditional survival probability at time t_i is estimated as $(n_i - d_i)/n_i$. Accordingly, the Kaplan–Meier estimator of the survival function $S(t)$ can be expressed as:

$$\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right) = \prod_{t_i \le t} \frac{n_i - d_i}{n_i}. \tag{4.27}$$

By multiplying conditional survival probabilities at each time point, this method recursively calculates cumulative survival probabilities from baseline to any time point t. It fully accommodates censored data and dynamically illustrates changes in survival status for the two subgroups across distinct follow-up time points. The blue curve represents patients in Subtype I, and the orange curve corresponds to patients in Subtype II.

Marked intergroup differences can be observed in the survival curves presented in Figure 12. The survival curve of Subtype I exhibits a steeper declining trend: the survival probability decreased sharply to approximately 0.25 at around 100 months of follow-up and continued declining to nearly 0.15, indicating faster disease progression and poorer prognosis in this subgroup. In contrast, the survival curve of Subtype II declined moderately throughout the follow-up period. The survival probability remained at approximately 0.8 until about 150 months, suggesting that patients in this subgroup maintained relatively stable survival status over an extended timeframe with a lower risk of disease progression.

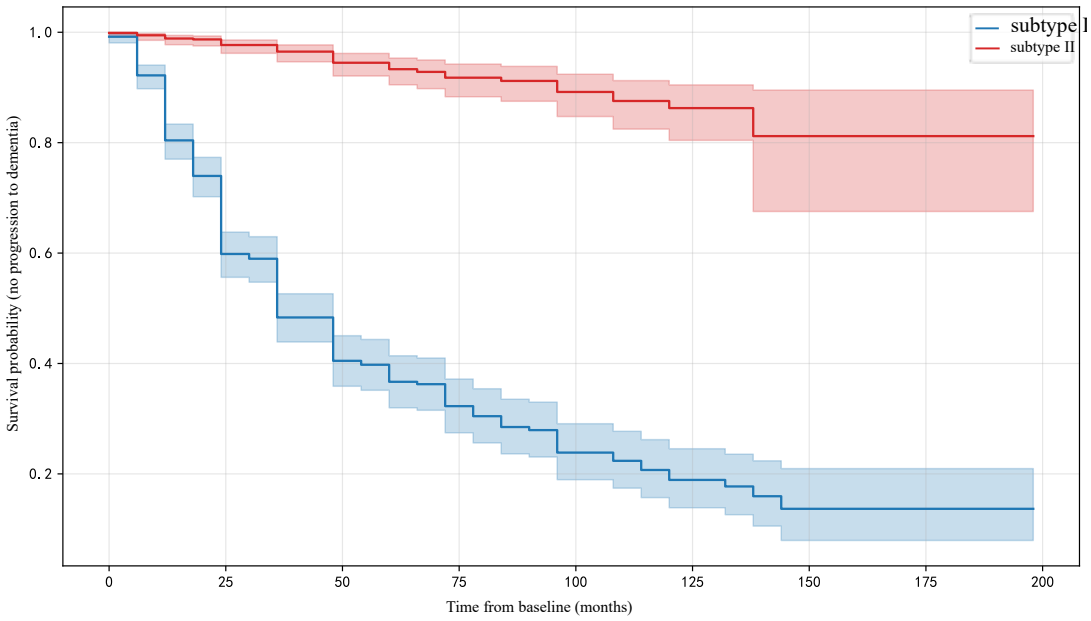


Figure 12. Kaplan-Meier survival curves for distinct clusters (participants with baseline MCI).

The survival analysis results indicated that Subtype I and Subtype II identified via the deep clustering approach not only exhibited distinct cross-sectional clinical and neuroimaging features, but also displayed significant differences in longitudinal prognosis. Subtype II represented the subgroup with slow disease progression and low survival risk, whereas Subtype I corresponded to the subgroup characterized by rapid disease progression and poor survival outcomes. These findings further verify the effectiveness and biological significance of the proposed model in identifying subtypes with divergent clinical prognoses.

To examine whether statistically significant differences in survival outcomes existed between the two subgroups, the Log-rank test was performed to compare the two survival curves.

Table 12. Results of the log-rank test.

Index	Value
Log-rank p-value	$1.6267184319960783 \times 10^{-171}$

As presented in Table 12, the Log-rank p-value was less than 0.01, suggesting that the null hypothesis of identical survival distributions between the two subgroups could be rejected statistically. Accordingly, the two subgroups derived from deep clustering had significantly different survival prognoses. This statistical evidence supports the heterogeneity between Subtype I and Subtype II with respect to disease progression risk and clinical outcomes, and further confirms the effectiveness and clinical value of the proposed model for identifying prognosis-related subtypes.

Meanwhile, to explore the practical value of the model in interpreting clinical phenotypes and uncovering feature-driven mechanisms, model-driven feature importance analysis was conducted based on clustering outputs. Specifically, the cluster labels obtained from deep clustering were treated as the dependent variable, and all multimodal features included in clustering were used as independent variables to train a random forest classification model. The contribution of each feature to discriminating the two patient subgroups was quantified using this model, and the results are illustrated in Figure 13.

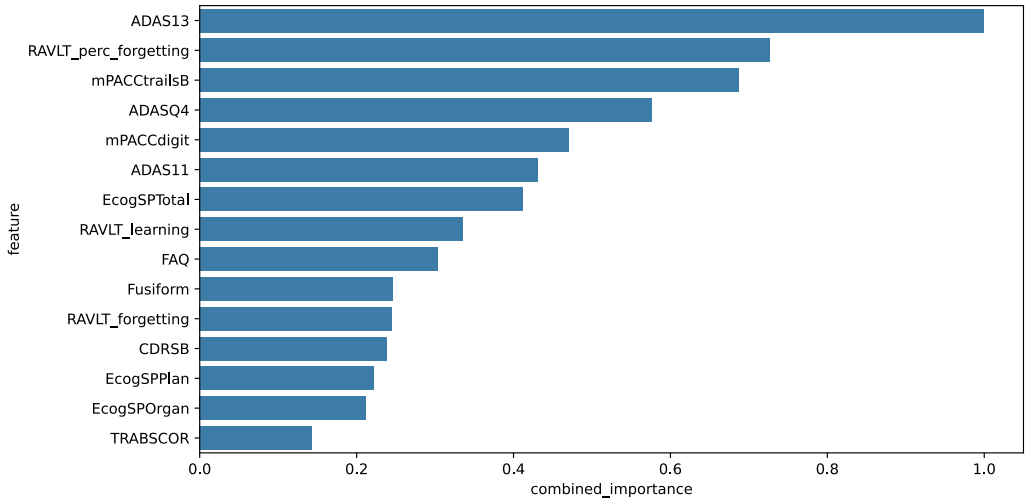


Figure 13. Ranking results of feature importance based on clustering results.

As shown in Figure 13, the top five features exerting the most prominent influence on clustering results were, in sequence: Alzheimer's disease assessment scale-cognitive subscale (ADAS13), RAVLT_perc_forgetting (Percent forgetting on the Rey Auditory Verbal Learning Test), mPACCtrailsB (executive function and processing speed), ADASQ4 (orientation item of the ADAS scale), and mPACCdigit (episodic memory). In the feature importance ranking, ADAS13, the core scale for evaluating the overall severity of cognitive impairment, ranked first. This indicates that the model primarily relies on the overall severity of cognitive decline across language, memory, and praxis domains when stratifying patients into subtypes. The second feature, RAVLT_perc_forgetting, directly reflects memory retention capacity. Its high importance weight suggests that the rate of

forgetting serves as a primary cognitive dimension for differentiating patients in terms of cognitive trajectories and memory stability. The third feature, mPACCtrailsB, mainly measures executive function and information processing speed, highlighting the critical contribution of individual differences in complex task switching and mental flexibility to subtype classification. The fourth feature ADASQ4 further focuses on orientation ability, implying that the integrity of basic cognitive function also constitutes an important basis for subtyping. mPACCdigit captures sustained attention and processing speed; its relatively high importance reveals systematic disparities in information processing efficiency between the two patient subgroups.

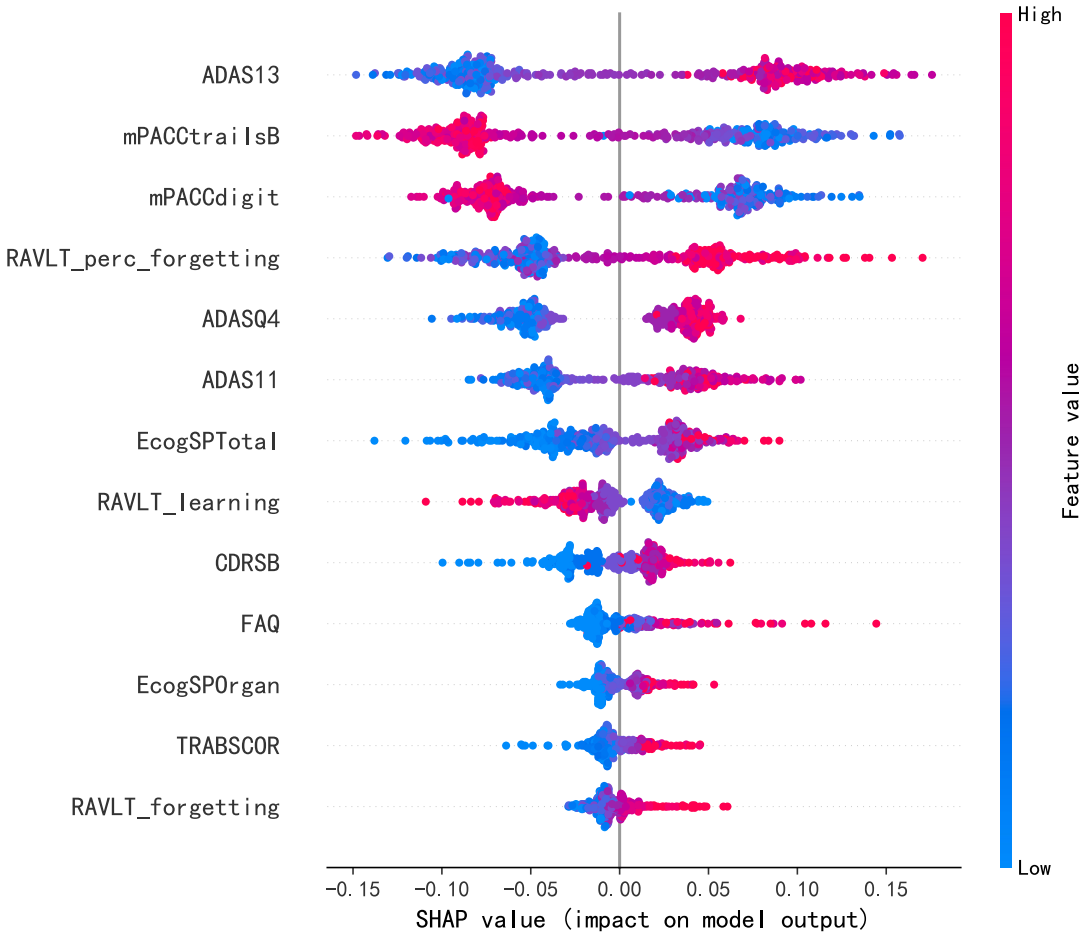


Figure 14. SHAP summary plot of input features. Features on the vertical axis are ranked in descending order of mean absolute SHAP value. The horizontal axis shows SHAP values, and the color gradient encodes feature values (red: high values; blue: low values).

Furthermore, features immediately following the top five, including ADAS11, EcogSPTotal and RAVLT_learning, also exhibited considerable importance. These findings further validate the clinical plausibility of the clustering results from multiple dimensions, encompassing core cognitive domains, self-perceived cognitive function, and learning ability. To deeply interpret the prediction mechanism of the model and enhance its clinical interpretability, this study further quantifies and visualizes feature importance based on the SHAP (Shapley additive explanations) method [33]. The SHAP

values can reflect both the direction and magnitude of each feature's contribution to the model output, and reveal the nonlinear relationships between feature values and prediction outcomes. Figure 14 illustrates the SHAP-value distribution of each feature, where each row corresponds to one feature sorted in descending order by the mean absolute SHAP value. A higher rank indicates stronger influence of the feature on model prediction.

As shown in Figure 14, the most influential features for model prediction are ADAS13, mPACCtrailsB, mPACCdigit, RAVLT_perc_forgetting, and ADASQ4. In addition, clinical and cognitive assessment metrics including ADAS11, EcogSPTotal, RAVLT_learning, and CDRSB also yield substantial contributions; these features are mainly derived from neuropsychological assessments and clinical cognitive scales.

Among them, ADAS13 and mPACCtrailsB exhibit the widest range of SHAP values, indicating that they serve as the strongest predictors for discriminating patient prognosis. Specifically, higher scores of ADAS13 and ADASQ4 are significantly associated with elevated risk of progression from MCI to dementia, suggesting that overall cognitive decline and impaired orientation constitute critical warning signals for disease progression. Higher values of RAVLT_perc_forgetting also correspond to increased progression risk, reflecting the adverse prognostic effect of rapid memory forgetting.

In contrast, higher scores of mPACCtrailsB are correlated with lower progression risk, demonstrating the protective effect of preserved executive function and information-processing speed on disease stability. As an indicator of sustained attention and processing speed, higher scores of mPACCdigit are likewise associated with low progression risk, implying that preserved information-processing efficiency is an essential factor for cognitive stability. Better performance on RAVLT_learning is also linked to reduced risk, which highlights the vital role of episodic-memory encoding and new-information learning in maintaining cognitive stability. Furthermore, better informant-reported everyday cognitive function (EcogSPTotal, EcogSPOrgan) and lower CDRSB scores act as reliable indicators for predicting stable disease status.

5. Conclusions

This study proposed an improved unsupervised deep clustering framework, namely Attention-OG-DTC, to explore intrinsic disease heterogeneity among patients with mild cognitive impairment (MCI). Compared with the original OG-DTC model, the introduced attention mechanism and optimized loss function effectively enhanced feature integration and discriminative capability, as validated by improved silhouette coefficient, DB index, and CH index, together with favorable PCA and t-SNE visualization results.

Based on the optimized clustering model, two clinically distinct MCI subtypes were identified. Cross-sectional analyses demonstrated that Subtype I exhibited more severe cognitive dysfunction, impaired daily activities, and reduced brain regional volumes, whereas Subtype II showed relatively preserved clinical and structural profiles. Longitudinal survival analysis further verified significant prognostic disparities between the two subgroups. Specifically, Subtype I corresponded to a high-risk subtype with rapid neurodegeneration and unfavorable outcomes, while Subtype II represented a stable subtype with slow disease progression, which was statistically confirmed by the Log-rank test.

Model-driven feature importance and SHAP interpretability analyses further validated the clinical rationality of the proposed subtyping scheme. Key discriminative biomarkers, including ADAS13,

mPACCtrailsB, and MMSE, were quantitatively proven to dominate subtype differentiation and prognostic stratification. Overall, this study provides a novel and interpretable subtyping paradigm for MCI patients, which reveals heterogeneous disease progression patterns and offers a reliable foundation for future personalized prognosis assessment and early intervention.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the Key Scientific Research Projects of Colleges and Universities in Henan Province, China, under Grant No.22A520010.

Conflict of interest

The authors declare that they have no conflict of interest with respect to the research, authorship, and publication of the article.

References

1. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, et al. (2006) Mild cognitive impairment. *Lancet* 367: 1262–1270. https://doi.org/10.1016/S0140-6736(06)68542-5
2. Mitchell AJ, Shiri-Feshki M (2009) Rate of progression of mild cognitive impairment to dementia—meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand* 119: 252–265. https://doi.org/10.1111/j.1600-0447.2008.01326.x
3. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dementia* 7: 270–279. https://doi.org/10.1016/j.jalz.2011.03.008
4. Pourzinal D, Yang JHJ, Byrne GJ, O'Sullivan JD, Mitchell L, McMahon KL, et al. (2020) Identifying subtypes of mild cognitive impairment in Parkinson's disease using cluster analysis. *J Neurol* 267: 3213–3222. https://doi.org/10.1007/s00415-020-09977-z
5. Ryoo HG, Choi H, Shi K, Rominger A, Lee DY, Lee DS, et al. (2024) Distinct subtypes of spatial brain metabolism patterns in Alzheimer's disease identified by deep learning-based FDG PET clusters. *Eur J Nucl Med Mol Imaging* 51: 443–454. doilinkhttps://doi.org/10.1007/s00259-023-06440-9
6. Chen R, Herskovits EH, for the Alzheimer's Disease Neuroimaging Initiative (2022) Machine learning detects distinct subtypes of minimal cognitive impairment. *J Sign Process Syst* 94, 437–443. https://doi.org/10.1007/s11265-021-01667-x
7. Edmonds EC, McDonald CR, Marshall A, Thomas KR, Eppig J, Eppig AJ, et al. (2019) Early versus late MCI: Improved MCI staging using a neuropsychological approach. *Alzheimer's Dementia* 15: 699–708. https://doi.org/10.1016/j.jalz.2018.12.009

8. Sindhu TS, Kumaratharan N, Anandan P (2024) Hybrid optimized deep fuzzy clustering-based segmentation and Deep Maxout Network for Alzheimer's disease classification. *Biomed Signal Process Control* 93: 106118. https://doi.org/10.1016/j.bspc.2024.106118

9. Manouvriez D, Kuchcinski G, Roca V, Sillaire AR, Bertoux M, Delbeuck X, et al. (2025) BrainAGE latent representation clustering is associated with longitudinal disease progression in early-onset Alzheimer's disease. *J Neuroradiology* 52: 101365. https://doi.org/10.1016/j.neurad.2025.101365

10. Chang CY, Slowiejko D, Win N (2024) Prediction and clustering of Alzheimer's disease by race and sex: A multi-head deep-learning approach to analyze irregular and heterogeneous data. *Sci Rep* 14: 26668. https://doi.org/10.1038/s41598-024-77829-1

11. Ribino P, Di Napoli C, Paragliola G, Chicco D, Gasparini F (2025) Multivariate longitudinal clustering reveals neuropsychological factors as dementia predictors in an Alzheimer's disease progression study. *BioData Min* 18: 26. https://doi.org/10.1186/s13040-025-00441-0

12. Jiao CN, Shang J, Li F, Cui X, Wang YL, Gao YL, et al. (2024) Diagnosis-guided deep subspace clustering association study for pathogenetic markers identification of Alzheimer's disease based on comparative atlases. *IEEE J Biomed Health Inf* 28: 3029–3041. https://doi.org/10.1109/JBHI.2024.3372294

13. Wang R, Gao L, Zhang X, Han J (2024) SVFR: A novel slice-to-volume feature representation framework using deep neural networks and a clustering model for the diagnosis of Alzheimer's disease. *Heliyon* 10: e23008. https://doi.org/10.1016/j.heliyon.2023.e23008

14. Wang D, Ma X, Schulz PE, Jiang X, Kim Y, (2024) Clinical outcome-guided deep temporal clustering for disease progression subtyping. *J Biomed Inf* 158: 104732. https://doi.org/10.1016/j.jbi.2024.104732

15. Baytaş İM (2024) Predicting progression from mild cognitive impairment to Alzheimer's dementia with adversarial attacks. *IEEE J Biomed Health Inf* 28: 3750–3761. https://doi.org/10.1109/JBHI.2024.3373703

16. Hou B, Wen Z, Bao J, Zhang R, Tong B, Yang S, et al. (2024) Interpretable deep clustering survival machines for Alzheimer's disease subtype discovery. *Med Image Anal* 97: 103231. https://doi.org/10.1016/j.media.2024.103231

17. Jack Jr CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. (2008) The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn. Reson Imaging* 27: 685–691. https://doi.org/10.1002/jmri.21049

18. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack C, Jagust W, et al. (2005) The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin* 15: 869–877. https://doi.org/10.1016/j.nic.2005.09.008

19. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, et al. (2017) The Alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's Dementia* 13: 561–571. https://doi.org/10.1016/j.jalz.2016.10.006

20. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack Jr CR, Kawas CH, et al. (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dementia* 7: 263–269. https://doi.org/10.1016/j.jalz.2011.03.005
21. Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *J Intern Med* 256: 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x
22. Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12: 189–198. https://doi.org/10.1016/0022-3956(75)90026-6
23. Morris JC (1993) The clinical dementia rating (CDR) current version and scoring rules. *Neurology* 43: 2412–2412. https://doi.org/10.1212/WNL.43.11.2412-a
24. Rosen WG, Mohs RC, Davis KL (1984) A new rating scale for Alzheimer's disease. *Am J Psychiatry* 141: 1356–1364. https://doi.org/10.1176/ajp.141.11.1356
25. Fischl B (2012) FreeSurfer. *Neuroimage* 62: 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021
26. Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ (1994) Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp* 2: 189–210. https://doi.org/10.1002/hbm.460020402
27. Gaser C, Dahnke R, Thompson PM, Kurth F, Luders E, the Alzheimer's Disease Neuroimaging Initiative (2024) CAT: A computational anatomy toolbox for the analysis of structural MRI data. *Gigascience* 3: giae049. https://doi.org/10.1093/gigascience/giae049
28. Ashburner J, Friston KJ (2005) Unified segmentation. *Neuroimage* 26: 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018
29. Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* 38: 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007
30. Mazziotta JC (1995) A probabilistic atlas of the human brain: Theory and rationale for its development: The international consortium for brain mapping (ICBM). *Neuroimage* 2: 89–101. https://doi.org/10.1006/nimg.1995.1012
31. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15: 273–289. https://doi.org/10.1006/nimg.2001.0978
32. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323: 533–536. https://doi.org/10.1038/323533a0
33. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 30.



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)