

# MRI Embeddings Complement Clinical Predictors for Cognitive Decline Modeling in Alzheimer’s Disease Cohorts

Nathaniel Putera, Daniel Vilet Rodríguez, Noah Videcrantz,  
Julia Machnio, and Mostafa Mehdipour Ghazi

Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

## ABSTRACT

Accurate modeling of cognitive decline in Alzheimer’s disease is essential for early stratification and personalized management. While tabular predictors provide robust markers of global risk, their ability to capture subtle brain changes remains limited. In this study, we evaluate the predictive contributions of tabular and imaging-based representations, with a focus on transformer-derived Magnetic Resonance Imaging (MRI) embeddings. We introduce a trajectory-aware labeling strategy based on Dynamic Time Warping clustering to capture heterogeneous patterns of cognitive change, and train a 3D Vision Transformer (ViT) via unsupervised reconstruction on harmonized and augmented MRI data to obtain anatomy-preserving embeddings without progression labels. The pretrained encoder embeddings are subsequently assessed using both traditional machine learning classifiers and deep learning heads, and compared against tabular representations and convolutional network baselines. Results highlight complementary strengths across modalities. Clinical and volumetric features achieved the highest AUCs of around 0.70 for predicting mild and severe progression, underscoring their utility in capturing global decline trajectories. In contrast, MRI embeddings from the ViT model were most effective in distinguishing cognitively stable individuals with an AUC of 0.71. However, all approaches struggled in the heterogeneous moderate group. These findings indicate that clinical features excel in identifying high-risk extremes, whereas transformer-based MRI embeddings are more sensitive to subtle markers of stability, motivating multimodal fusion strategies for AD progression modeling.

**Keywords:** Cognitive Decline, Alzheimer’s Disease, MRI, Deep Learning, Vision Transformer.

## 1. INTRODUCTION

Cognitive decline is an indicator of Alzheimer’s disease (AD), a progressive neurodegenerative disorder affecting memory, language, and executive function [1]. As the most common cause of dementia worldwide [2], AD imposes not only individual suffering but also substantial societal and economic burdens through increased care needs and loss of independence. Yet, its clinical course is highly variable; some individuals decline gradually over many years, while others progress rapidly to severe impairment. This heterogeneity complicates prognosis and care planning. Identifying individuals at risk of faster decline is thus essential for enabling timely intervention, optimizing clinical trial design, and advancing personalized treatment strategies [3].

Conventional diagnosis and staging of AD rely on neuropsychological tests, structural magnetic resonance imaging (MRI), and fluid or molecular biomarkers [4]. Structural MRI is widely available and non-invasive, and it captures atrophy patterns associated with disease progression [5]. However, cross-sectional measures such as hippocampal volume provide only limited information on the temporal dynamics of decline. This limitation has motivated the development of longitudinal methods that characterize trajectories of cognitive changes.

Recent advances in deep learning have enabled data-driven modeling of disease trajectories from neuroimaging [3, 6]. Nevertheless, most studies remain focused on static classification of AD stages rather than predicting future decline [7]. Directly training large models on MRI data also risks overfitting in modest cohort sizes [8]. To mitigate this, unsupervised and self-supervised learning strategies are increasingly employed to extract robust, transferable representations of brain structure without requiring large labeled datasets [9].

In this work, we aim to predict cognitive decline trajectories in AD by utilizing baseline MRIs and tabular descriptors. We derive four progression groups of stable, mild, moderate, and severe decline, based on clustering of temporal Clinical Dementia Rating Sum of Boxes (CDR-SB) scores. Besides, we train a 3D vision transformer

(ViT) inspired by [9] using MRI reconstruction as a pretext task, enabling the extraction of meaningful embeddings independent of clinical labels. These representations are subsequently used for downstream classification and compared against traditional classifiers, tabular predictors, and pretrained convolutional networks.

Our work provides three main contributions. First, we derive progression labels using trajectory clustering with Dynamic Time Warping (DTW) to capture heterogeneous temporal patterns despite irregular follow-up schedules. Second, we train a 3D ViT model through unsupervised MRI reconstruction to obtain anatomy-aware embeddings, thereby eliminating reliance on clinical labels. Third, we perform a systematic cross-modal comparison between tabular predictors, transformer-derived imaging features, and convolutional network baselines to characterize modality-specific strengths in predicting distinct progression subtypes.

## 2. METHODS

### 2.1 Study Data

We used two complementary subsets from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) [10]. The first subset from ADNI1 consists of screening-visit data with 503 T1 MRI scans. These scans were employed to train the vision encoder and support image-based predictive modeling. All MRIs underwent preprocessing with the FreeSurfer pipeline [11], including motion and bias-field correction, affine registration to the MNI305 template space, resampling to 1 mm isotropic resolution, and intensity normalization to  $[0, 255]$  [12]. To ensure fair representation across patient groups, the MRI subset was first stratified by age, sex, and clinical diagnosis. The stratified data were then divided into training and testing sets using an 80:20 split. We applied a series of data augmentation techniques to increase variability and mitigate overfitting, following the protocol of [13]. These transformations simulate common clinical MRI artifacts as well as natural anatomical variability, thereby encouraging the model to learn robust features that generalize beyond the training data.

The second subset, ADNI-Merge [14], provides longitudinal tabular information, including cerebrospinal fluid (CSF) biomarkers (amyloid-beta, tau, p-tau), positron emission tomography (PET) measurements (FDG, AV45, PIB, FBB), risk factors (age, sex, education, APOE4), neuropsychological test scores (CDR, ADAS, MMSE, RAVLT, LDEL, DIGIT, TRAB, FAQ, MOCA), and brain volumetrics (hippocampus, entorhinal, middle temporal, fusiform, ventricles). From this subset, we derived cognitive progression labels using the CDR-SB. Specifically, we applied trajectory clustering to assign individuals to distinct progression groups that capture the heterogeneity of cognitive decline. The resulting labels served as ground truth for downstream prediction, while the remaining imaging and tabular data were used as predictors in classification experiments.

### 2.2 Cognitive Decline Groups

Cognitive decline groups were defined using the CDR-SB scores from ADNI-Merge participants with at least two visits. It was chosen because it is a well-established measure of functional and cognitive status in AD, widely used in clinical trials and observational studies [15]. To capture temporal similarities across individual trajectories, we applied  $k$ -means clustering using DTW distances [16]. DTW enables alignment of trajectories of varying length and timing, allowing us to identify homogeneous patterns of change even when visits occur at different intervals. We set  $k = 4$  to define four clinically interpretable subgroups. These were labeled as stable, mild, moderate, and severe, corresponding to increasing levels of cognitive decline over time. These clusters served as target labels for downstream predictive modeling. An overview of the clustered trajectories is shown in Figure 1, indicating the distinct CDR-SB progression profiles across groups.

### 2.3 3D ViT Pretraining

To obtain robust MRI-based embeddings while reducing the risk of overfitting associated with direct supervised training, we adopted an unsupervised pretraining strategy. Specifically, we implemented a 3D ViT-V-Net model inspired by [9], designed to learn representations through image reconstruction. The model was trained for 200 epochs with a mean squared error (MSE) loss, using both original and artifact-simulated MRI scans. Reconstruction quality was monitored with the Structural Similarity Index Measure (SSIM) to ensure that the learned representations preserved fine-grained anatomical structure.

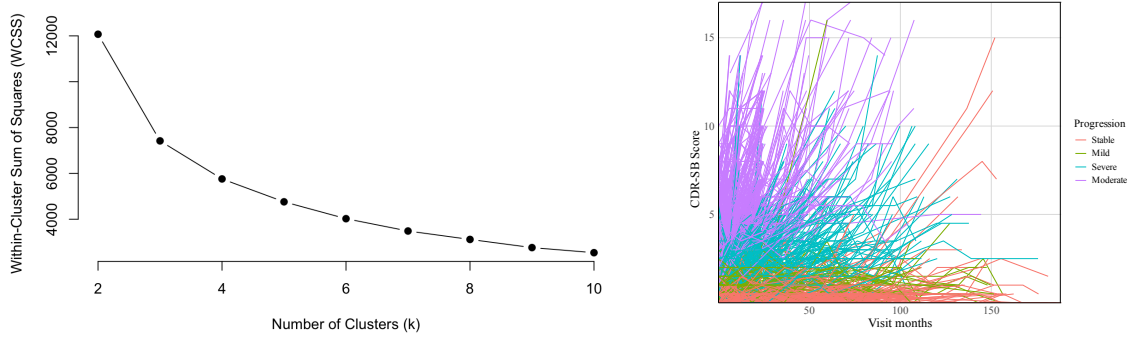


Figure 1: Elbow plot for selecting the optimal number of clusters (left), and the resulting 4 clusters (right) obtained from  $k$ -means clustering with DTW distances, grouped by CDR-SB progression.

Following pretraining, the decoder was discarded, and the encoder was retained to extract latent feature embeddings from MRI scans. These embeddings served as input representations for downstream classification tasks, where we evaluated both traditional machine learning classifiers and fully connected (FC) classifier heads. To provide a comparative baseline, we fine-tuned ResNet-3D models [17, 18] on the same MRI subset, enabling a direct assessment of performance differences between transformer-based and convolutional architectures.

## 2.4 Cognitive Decline Prediction

To evaluate predictive strategies for AD progression, we considered models based on tabular predictors, transformer embeddings, and pretrained convolutional neural networks (CNNs). All approaches were trained to classify individuals into the four cognitive decline groups defined by clustered CDR-SB trajectories.

**Tabular Predictors.** Tabular features included CSF markers, PET measures, brain volumetrics, neuropsychological test scores (excluding CDR-SB), and risk factors. To obtain compact and noise-robust representations, each feature set was processed through an autoencoder before classification. The resulting representations were then used to train an Extreme Gradient Boosting (XGBoost) model [19] for the four-class prediction task.

**3D ViT Predictors.** MRI-based representations were extracted from the encoder of the pretrained ViT-V-Net model. To reduce dimensionality and mitigate overfitting, we applied principal component analysis (PCA), retaining 15 principal components that together explained approximately 95% of the variance. These reduced features were classified with XGBoost. In a complementary setup, we attached a lightweight FC head to the frozen ViT-V-Net encoder, consisting of layers with ReLU activations and dropout for regularization. Only the classifier head was trained, using a learning rate of  $1 \times 10^{-4}$ .

**3D CNN Predictors.** For comparison with established CNN architectures, we fine-tuned pretrained 3D models. In the ResNet-Med3D framework [17], the original segmentation head was replaced by a custom classifier. The backbone was updated with a learning rate of  $1 \times 10^{-4}$ , while the new classifier head was trained with  $1 \times 10^{-2}$ . Intensity normalization was applied to match the input distribution of the pretraining model. Similarly, in the ResNet-MONAI implementation [18], the final output layer of the backbone model was adapted to predict the four cognitive decline categories.

## 3. RESULTS

We evaluated four-way classification of cognitive decline trajectories using one-vs-rest AUC as the primary metric. Summary AUCs for tabular predictors are reported in Table 1, and a per-class breakdown across all other methods is provided in Table 2. As can be seen, using clinical and volumetric features compressed via autoencoders, with XGBoost yielded the strongest performance for the mild and severe groups, indicating that structured clinical information and brain volumetrics are particularly informative for progressive groups.

Table 1: Test AUCs (mean  $\pm$  std) of different feature sets for predicting AD progression stages over 5 repetitions.

| Feature Set       | Stable                 | Mild                   | Moderate               | Severe                 |
|-------------------|------------------------|------------------------|------------------------|------------------------|
| Cognitive Scores  | <b>0.59</b> $\pm$ 0.03 | <b>0.75</b> $\pm$ 0.01 | <u>0.56</u> $\pm$ 0.03 | <b>0.77</b> $\pm$ 0.02 |
| CSF Markers       | 0.55 $\pm$ 0.02        | 0.64 $\pm$ 0.02        | <b>0.59</b> $\pm$ 0.02 | 0.59 $\pm$ 0.01        |
| PET Measures      | 0.49 $\pm$ 0.02        | 0.69 $\pm$ 0.02        | <b>0.59</b> $\pm$ 0.02 | 0.69 $\pm$ 0.02        |
| Risk Factors      | 0.53 $\pm$ 0.01        | 0.63 $\pm$ 0.02        | 0.51 $\pm$ 0.02        | 0.62 $\pm$ 0.02        |
| Brain Volumetrics | <u>0.58</u> $\pm$ 0.03 | <u>0.70</u> $\pm$ 0.02 | 0.53 $\pm$ 0.03        | <u>0.70</u> $\pm$ 0.02 |

Table 2: Test AUCs (mean  $\pm$  std) of different methods for predicting AD progression stages over 5 repetitions.

| Method            | Stable                 | Mild                   | Moderate               | Severe                 |
|-------------------|------------------------|------------------------|------------------------|------------------------|
| Brain Volumetrics | 0.58 $\pm$ 0.03        | <b>0.70</b> $\pm$ 0.02 | <b>0.53</b> $\pm$ 0.03 | <b>0.70</b> $\pm$ 0.02 |
| ViT + XGBoost     | <u>0.65</u> $\pm$ 0.02 | <u>0.59</u> $\pm$ 0.02 | <u>0.51</u> $\pm$ 0.03 | 0.46 $\pm$ 0.02        |
| ViT + FC          | <b>0.71</b> $\pm$ 0.01 | 0.58 $\pm$ 0.03        | 0.37 $\pm$ 0.02        | 0.41 $\pm$ 0.02        |
| ResNet-Med3D      | 0.62 $\pm$ 0.02        | 0.53 $\pm$ 0.02        | 0.43 $\pm$ 0.02        | <u>0.54</u> $\pm$ 0.01 |
| ResNet-MONAI      | 0.55 $\pm$ 0.03        | 0.52 $\pm$ 0.02        | 0.47 $\pm$ 0.02        | 0.48 $\pm$ 0.02        |

We next assessed MRI-derived embeddings from the pretrained ViT-V-Net encoder combined with PCA and XGBoost. Relative to tabular models, this approach improved discrimination of the stable group, suggesting that the learned anatomical representations capture features associated with minimal near-term change. However, performance in the severe and mild groups did not surpass the tabular baseline, potentially reflecting information loss from dimensionality reduction or the need for features more tightly coupled to disease severity.

Attaching a lightweight FC head to the frozen ViT-V-Net encoder produced the highest AUC for the stable class among all imaging models. Despite these gains, the moderate class remained the most challenging across methods, reinforcing that borderline progression states are difficult to separate from baseline MRI alone. Moreover, among CNN baselines, ResNet-Med3D outperformed the ResNet-MONAI, but both CNN variants underperformed the ViT-based models on the stable class and showed limited improvement for moderate and severe trajectories. Overall, tabular predictors excel at identifying mild and severe decline, whereas transformer-based MRI representations are most effective for detecting stable courses. This complementary pattern motivates multimodal fusion and class-aware learning in future work to leverage clinical severity cues alongside anatomy-driven stability signatures for more robust AD progression modeling.

#### 4. CONCLUSION

This study investigated how different data modalities and representation learning strategies contribute to the prediction of cognitive decline in AD progression. We first applied stratified preprocessing and realistic data augmentation to ensure balanced group representation and robustness to common MRI artifacts. To obtain meaningful imaging features, we pretrained a 3D ViT model for unsupervised reconstruction, retaining the encoder as a feature extractor for downstream tasks. The resulting embeddings were compared against clinical predictors, volumetric measures, and CNN architectures.

Our results demonstrated complementary strengths of different modalities. Tabular predictors were most effective in identifying mild and severe progression groups, achieving average AUC values of approximately 0.70. In contrast, MRI-derived embeddings from ViT were most informative for the stable progression group, achieving the highest AUC of 0.71. The difficulty in classifying the moderate group highlights the clinical and biological heterogeneity of this stage, which may require multimodal integration for more reliable discrimination.

Overall, clinical features capture global risk patterns, while deep learning-based imaging embeddings characterize subtle structural signatures of stability and transition. The complementary performance profiles indicate that future work should move toward multimodal fusion strategies, combining structured tabular data with transformer-derived imaging features. Such integrative approaches hold promise for more robust AD progression modeling, ultimately supporting early stratification and personalized disease management.

## ACKNOWLEDGMENTS

This project is supported by the Pioneer Centre for AI, funded by the Danish National Research Foundation (grant number P1).

## REFERENCES

- [1] Better, M. A., “Alzheimer’s disease facts and figures,” *Alzheimer’s & Dementia* **19**(4), 1598–1695 (2023).
- [2] Gustavsson, A., Norton, N., Fast, T., Frölich, L., Georges, J., Holzapfel, D., Kirabali, T., Krolak-Salmon, P., Rossini, P. M., Ferretti, M. T., et al., “Global estimates on the number of persons across the Alzheimer’s disease continuum,” *Alzheimer’s & Dementia* **19**(2), 658–670 (2023).
- [3] Yi, F., Zhang, Y., Yuan, J., Liu, Z., Zhai, F., Hao, A., Wu, F., Somekh, J., Peleg, M., Zhu, Y.-C., et al., “Identifying underlying patterns in Alzheimer’s disease trajectory: A deep learning approach and Mendelian randomization analysis,” *EClinicalMedicine* **64** (2023).
- [4] Jack Jr, C. R., Andrews, J. S., Beach, T. G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Ho, C., Jagust, W., McDade, E., et al., “Revised criteria for diagnosis and staging of Alzheimer’s disease: Alzheimer’s association workgroup,” *Alzheimer’s & Dementia* **20**(8), 5143–5169 (2024).
- [5] Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack Jr, C. R., Kaye, J., Montine, T. J., et al., “Toward defining the preclinical stages of 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**(3), 280–292 (2011).
- [6] Helaly, H. A., Badawy, M., and Haikal, A. Y., “Deep learning approach for early detection of Alzheimer’s disease,” *Cognitive Computation* **14**(5), 1711–1727 (2022).
- [7] Chamakuri, R. and Janapana, H., “A systematic review on recent methods on deep learning for automatic detection of Alzheimer’s disease,” *Medicine in Novel Technology and Devices* **25**, 100343 (2025).
- [8] Jo, T., Nho, K., and Saykin, A. J., “Deep learning in Alzheimer’s disease: diagnostic classification and prognostic prediction using neuroimaging data,” *Frontiers in Aging Neuroscience* **11**, 220 (2019).
- [9] Chen, J., He, Y., Frey, E. C., Li, Y., and Du, Y., “ViT-V-Net: Vision transformer for unsupervised volumetric medical image registration,” *CoRR* (2021).
- [10] Wyman, B. T., Harvey, D. J., Crawford, K., Bernstein, M. A., Carmichael, O., Cole, P. E., Crane, P. K., DeCarli, C., Fox, N. C., Gunter, J. L., et al., “Standardization of analysis sets for reporting results from ADNI MRI data,” *Alzheimer’s & Dementia* **9**(3), 332–337 (2013).
- [11] Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al., “Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain,” *Neuron* **33**(3), 341–355 (2002).
- [12] Nielsen, M. E., Nielsen, M., and Mehdipour Ghazi, M., “Assessing the efficacy of classical and deep neuroimaging biomarkers in early Alzheimer’s disease diagnosis,” in *[Medical Imaging 2025: Computer-Aided Diagnosis]*, **13407**, 504–509, SPIE (2025).
- [13] Mehdipour Ghazi, M. and Nielsen, M., “FAST-AID Brain: Fast and accurate segmentation tool using artificial intelligence developed for brain,” in *[Scandinavian Conference on Image Analysis]*, 161–176, Springer (2025).
- [14] Mehdipour Ghazi, M., Selnes, P., Timón-Reina, S., Tecelão, S., Ingala, S., Bjørnerud, A., Kirsebom, B.-E., Fladby, T., and Nielsen, M., “Comparative analysis of multimodal biomarkers for amyloid-beta positivity detection in Alzheimer’s disease cohorts,” *Frontiers in Aging Neuroscience* **16**, 1345417 (2024).
- [15] Cedarbaum, J. M., Jaros, M., Hernandez, C., Coley, N., Andrieu, S., Grundman, M., Vellas, B., et al., “Rationale for use of the clinical dementia rating sum of boxes as a primary outcome measure for Alzheimer’s disease clinical trials,” *Alzheimer’s & Dementia* **9**(1), S45–S55 (2013).
- [16] Mehdipour Ghazi, M., Urdanibia-Centelles, O., Bakhtiari, A., Fagerlund, B., Vestergaard, M. B., Larsson, H. B. W., Mortensen, E. L., Osler, M., Nielsen, M., Benedek, K., et al., “Cognitive aging and reserve factors in the Metropolit 1953 Danish male cohort,” *GeroScience* **47**(2), 2475 (2024).
- [17] Chen, S., Ma, K., and Zheng, Y., “Med3D: Transfer learning for 3D medical image analysis,” *arXiv preprint arXiv:1904.00625* (2019).

- [18] Cardoso, M. J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al., “MONAI: An open-source framework for deep learning in healthcare,” *arXiv preprint arXiv:2211.02701* (2022).
- [19] Ali, Z. A., Abduljabbar, Z. H., Tahir, H. A., Sallow, A. B., and Almufti, S. M., “eXtreme gradient boosting algorithm with machine learning: A review,” *Academic Journal of Nawroz University* **12**(2), 320–334 (2023).