Article:
Huang, Y., Zhang, H., Ma, B. et al. (2026) Converse or reverse? Machine-learning modeling for disease progression: A study based on Alzheimer's disease continuum cohort. NeuroImage, 327. 121754. ISSN: 1053-8119

https://doi.org/10.1016/j.neuroimage.2026.121754

Converse or Reverse? Machine-learning Modeling for Disease Progression: A Study Based on Alzheimer's Disease Continuum Cohort

Abstract

INTRODUCTION: Longitudinal trajectories from healthy aging to Mild Cognitive Impairment and Alzheimer's Disease involve complex mechanisms.

METHODS: We evaluated five machine learning approaches (Random Forest, Support Vector Machines, Radial Basis Function Networks, Backpropagation Networks, Convolutional Neural Network) to assess the importance of potential predictive markers across the health-to-dementia continuum. Using the ADNI cohort across four phases (ADNI1, ADNIGO, ADNI2, ADNI3), we analyzed participants with distinct trajectories: stable, convertible, and reverse progression.

RESULTS: Random Forest outperformed other models across key effectiveness metrics and achieved a macro-averaged sensitivity of 70.28% and specificity of 96.28% across all participant groups. Random Forest identified visuospatial and memory-related cognitive dysfunction as key predictive clinical features and several amyloid-related neuroimaging biomarkers — including temporal variations of amyloid uptake within inferior lateral ventricles, para-hippocampus—for classifying participant groups. Additionally, plasma APOE4 and long neurofilament light chain levels emerged as promising predictors for tracking progression.

Conclusion: These findings highlight the potential of machine learning in classifying disease trajectories.

Key words

Machine learning; Random Forest; Healthy-MCI-AD continuum; ADNI

1.Introduction

The transition from normal cognitive function to mild cognitive impairment (MCI) or Alzheimer's disease (AD) involves a complex pathological process. Patients with MCI may progress to dementia, remain stable for years, or even revert to normal cognition [1-3]. Studies indicates that over 20% of MCI exhibit a reverse to normal cognition [4-6]. Thus, the heterogeneity and variability of MCI and AD present significant challenges for accurate prediction and prevention in patients.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal, multi-center study spanning the United States and Canada. The original goal of ADNI was to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Based on the ADNI cohort, a systems biology framework combining neuroimaging, biospecimen biomarkers, and clinical metrics is critical for deciphering the disease continuum from normal cognition to MCI and AD [7-9]. By synthesizing existing literature, machine learning enables the integration of multimodal biomarkers to assess how exposures modulate MCI and AD trajectories [10,11]. A deep learning algorithm at the individual level was developed to generate visualizations of AD risk stratification within the ADNI cohort [11]. However, due to the predominance of opaque 'black-box' algorithms [12], it is still unresolved about the reliable computerized predictors for disease progression across the normal-to-dementia continuum.

We aim to assess the comparative effectiveness of five machine learning algorithms (Random Forest, Support Vector Machines, Radial Basis Function Networks, Backpropagation Networks, and Convolutional Neural Network) in predicting progression across the cognitive spectrum from normal aging to dementia in the ADNI cohort. These machine learning algorithms have been proven to be powerful data-driven methods for classification and prediction applications [13-16]. Machine learning is widely used for tasks such as regression, classification, dimensionality reduction. In this study, we selected five representative classification methods: Backpropagation Networks (a type of artificial neural networks), Support Vector Machines and Radial Basis Function Networks (both classical Kernel Methods), Random Forest (an Ensemble Method based on Decision Tree Learning), and Convolutional Neural Network (a machine-learning algorithm to extract features from grid-like matrix datasets). We screened participants across four phases of the ADNI cohort (ADNI1, ADNIGO, ADNI2, and ADNI3). We selected participants with varying longitudinal trajectories, including those with stable status (stable health, stable MCI, stable AD), convertible status (Health-to-MCI, MCI-to-AD, Health-to-MCI-to-AD), and reversible status (MCI-to-Health, AD-to-MCI). There are features in each machine learning model, encompassing the demographic characteristics, clinical assessments, neuroimaging biomarkers (magnetic resonance imaging and positron emission tomography), biospecimen biomarkers (e.g. plasma and cerebrospinal fluid). We predicted that amyloid-related biomarkers and clinical decline (e.g. episodic memory or visual-spatial dysfunction) may play a pivotal role in predicting the conversion or reversion of AD or MCI patients from the normal state.

2.Methods

2.1 Participants

We screened all data summaries in the ADNI cohort across four phases (ADNI1, ADNIGO, ADNI2, ADNI3 with participant Rostered ID ranging from 1 to 7123) and initially identified 1492 participants with complete records for diagnosis outcome, APOE4 and key demographic variables (age, gender, education, ethnicity, race, marital status, and handedness). After excluding 293 participants due to missing family history data (from mothers, fathers or siblings), we retained multi-modal data summaries from 1199 participants for subsequent analyses (see Figure 1). Since tracking the rate of conversion from normal to MCI and MCI to AD is a primary outcome measure in the ADNI cohort, site physicians reviewed visit-related clinical measures and completed Diagnostic Summary and Diagnostic Summary Baseline Changes forms. When conversion was triggered, the conversion committee reviewed the Clinical Dementia Rating (CDR) for the visit and complete the Consensus Diagnosis and Conversion forms. We screened all subjects across four ADNI phases (ADNI 1, ADNI GO, ADNI 2, ADNI 3) and then selected the subjects with different longitudinal trajectories based on diagnostic summaries in accordance with several criteria:

(1) All subjects across all ADNI phases (with Participant Rostered IDs ranging from 1 to 7123) were reviewed for diagnostic results at each visit time point;

(2) For each subject, the duration from baseline to the final visit and corresponding diagnostic results was recorded. In the ADNI cohort, the diagnostic results for each subject were typically reported every 6 months. Subjects who only completed the baseline visit were excluded.

(3) Subjects with a baseline diagnosis of AD, and who maintained a consistent AD diagnosis for 12 months or longer, were defined as *StableAD* patients.

(4) Subjects with a baseline diagnosis of MCI, and who maintained a consistent MCI diagnosis 48 months or longer, were defined as *StableMCI* patients.

(5) Subjects with a baseline diagnosis of HC, and who maintained a consistent HC for 48 months or longer, were defined as *StableHC* participants.

(6) Subjects with a baseline diagnosis of HC, and who converted to MCI or AD during follow-up, and maintained the MCI or dementia diagnosis respectively after conversion, were defined as *UnstableConvertibleHC* participants.

(7) Subjects with a baseline diagnosis of MCI, and who converted to AD during follow-up, and maintained the AD diagnosis after conversion, were defined *as UnstableConvertibleMCI* participants.

(8) Subjects with a baseline diagnosis of MCI or AD who reverted to normal cognition (MCI-to-HC) or a less severe dementia state (AD-to-MCI) respectively during follow-up, and maintained the improved cognitive/dementia severity state in subsequent visits after reversion, were defined as *UnstableReverse* participants. Subjects who experienced both conversion and reversion events were classified as UnstableReverse participants.

(9) Subjects with a baseline diagnosis of HC, whose condition gradually progressed to MCI and subsequently to AD during follow-up were classified as *UnstableProgress* participants.

The study was approved by the local research ethics committee in the ADNI cohort and all participants provided written informed consent prior to the study before their participation according to the Declaration of Helsinki.

2.2 Data Preprocessing

We merged key variables using MATLAB (version R2021a, Mathworks, Natick, MA, USA), integrating data from case report forms, neuroimaging lab summaries, and biospecimen biomarker lab summaries across four ADNI phases (ADNI1, ADNIGO, ADNI2, and ADNI3). The resulting dataset was structured with rows representing participant visits and columns containing key variables, including diagnostic outcome, demographic features, cognitive assessments, neuroimaging measures (brain atrophy, FDG PET, 18F-AV45 PET, AV1451 PET), fluid biomarkers (plasma and CSF pTau, amyloid), and laboratory. After merging, the dataset comprised a total of 6240 raw variables per participant visit.

For the neuroimaging lab summaries, brain atrophy was preprocessed by using FreeSurfer image analysis suite. Briefly, this preprocessing includes: (1) motion correction and averaging of multiple volumetric T1 weighted images; (2) removal of non-brain tissue using a hybrid watershed/surface deformation procedure; (3) automated Talairach transformation; (4) segmentation of the subcortical white matter and deep gray matter volumetric structures (including hippocampus, amygdala, caudate, putamen, ventricles) intensity normalization; (5) tessellation of the gray matter white matter boundary; (6) automated topology correction; (7) surface deformation following intensity gradients to optimally place the gray/white and gray/cerebrospinal fluid borders at the location where the greatest shift in intensity defines the transition to the other tissue class [17]. PET image analysis pipeline includes: (1) using Freesurfer to define regions (frontal, anterior/posterior cingulate, lateral parietal, lateral temporal) and make up a summary of cortical ROI. (2) defining five reference regions (cerebellar grey matter, whole cerebellum, brainstem/pons, eroded subcortical white matter, and a composite reference region); (3) coregistering each PET image scan to the corresponding MRI to calculate

the mean standardized uptake value ratio (SUVR) within the cortical and reference regions [18-20].

To assess longitudinal trajectories for each raw variable, we calculated three intra-individual metrics across participant visits:

(1) mean value ('Global level'), representing overall measure during follow-up. The 'Global Level', which was computed from multiple follow-ups, refers to the averaging process of the values of the same clinical symptom at multiple time points. Global Level can reduce the random error of a single measurement and more objectively reflect the long-term baseline level or overall trend of the clinical feature.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

In the equation, '$\bar{x}$' denotes the 'Global level' of each raw variable per subject, 'n' represents the total number of measurements(from baseline to the final visit) for each raw variable per subject, and 'x_i' indicates the value of the raw variable, where 'i' denotes the raw variable ID (ranging from 1 to 6240).

(2) standard error ('Temporal Variation'), indicating visit-to-visit longitudinal fluctuations. The 'Temporal Variation' refers to the dynamic fluctuations of physiological, pathological, or symptomatic indicators in patients at different follow-up time points.

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

In the equation, 'S' denotes the 'Temporal Variation' of each raw variable per subject, 'n' represents the total number of measurements (from baseline to the final visit) for each raw variable per subject, $\bar{x}$ indicates the global level of each raw variable per subject. 'x_i' indicates the value of the raw variable, where 'i' denotes the raw variable ID (ranging from 1 to 6240).

(3) linear regression beta coefficients ('Causal Inference'), quantifying the association between follow-up time and variable changes. The regression coefficient is a core parameter to quantify the magnitude and direction of the impact of time on clinical characteristics. When time is taken as the independent variable and clinical characteristics as the dependent variable, a positive regression coefficient indicates that the clinical feature shows an upward trend over time, while a negative coefficient indicates a downward trend. The larger the absolute value of the coefficient, the stronger the impact of time on the clinical characteristics.

$$\beta = regress(x_i, t_i)$$

In the equation, 'β' denotes the 'Causal Inference' of each raw variable per subject, which was derived from the slope of the linear model $x_i = \beta * t_i$. Here, 'x_i' represents the value of the raw variable,where 'i' denotes the raw variable ID (ranging from 1 to 6240). 't_i' indicates the visit time corresponding to the value of raw variable x_i.

For variables measured only once or remaining constant, only mean values $\bar{x}$ were computed. The preprocessing procedure yielded 11761 intra-individual variables per participant, comprising Global level, Temporal Variation, Causal Inference measures for each raw variable. Following quality control, we excluded intra-individual variables with missing values exceeding 31.76%, retaining 2324 intra-individual variables as features for machine learning. For the selected features, we imputed remaining missing values using group-wise mean imputation (calculated within each participant group).

2.3 Machine Learning Analyses

By using the same preprocessed data, we evaluated and compared five machine learning models — Random Forest, Support Vector Machines, Radial Basis Function Networks, Backpropagation Networks, and Convolutional Neural Network — using the same preprocessed data to predict

participant groups. All machine learning models were executed using MATLAB (version R2021a, Mathworks, Natick, MA, USA). The dataset was randomly divided into training (70%) and testing (30%) sets for each model. Prior to modeling, features were normalized.

The optimal parameter pairs for each machine learning model were obtained via a grid searching method by taking into account the leave-one-out (LOO) cross-validation error. For each model, the parameter values that yielded the minimum LOO cross-validation error were selected as the optimal parameters. Model-specific parameters were configured as follows:

Random Forest: 50 decision trees with the minimum leaf of 1 for modeling. The importance value of each feature was calculated. The out-of-bag error was performed to estimate the accuracy.

Support Vector Machines: cost=10.0, and gamma=0.01.

Radial Basis Function Networks: spread = 100.

Backpropagation Networks: the train parameters of epochs, goal, and learning ratio were set to 1000, 1e-6, and 0.01 respectively.

Convolutional Neural Network: Adam algorithm, MaxEpochs=500, InitialLearnRate=1e-3, L2Regularization=1e-4, LearnRateDropFactor=0.1, LearnRateDropPeriod=400.

Model performance was assessed using sensitive, specificity, accuracy, precision, and F1 score per participant group. The confusion matrices summarized the results of modeling classification (please see Figure 2).

2.4 Validation Analyses

We used Lasso regression to assess the relationships between the selected features and the outcome variable (seven participant groups), with 10-fold cross-validation. Potential predictors

of longitudinal trajectories in participants were defined as those meeting two criteria: (1) ranking in the top 1.2% of absolute beta coefficients (≥0.1) in Lasso regression; (2) ranking in the top 17.2% of importance values (≥0.15) in Random Forest. To further validate these predictors, we performed one-way ANOVA to compare differences among seven participant groups with different longitudinal trajectories at each visit time. The F test p values (uncorrected) were applied to validate those features with high Lasso β and importance values. Post-hoc analysis were based on LSD algorithm when equal variances assumed, and on Tamhane's T2 algorithm when equal variances not assumed. Cross-tabulation analysis was performed to examine associations between categorical variables. Chi-square tests were applied to assess statistical significance in contingency tables.

To validate the potential predictors — particularly voxel-level alterations from multimodal imaging datasets — we further integrated a convolutional neural network with SHAP (Shapley Additive exPlanations) interpretability values. This combined strategy ensures strong predictive power for the promising predictors, which is critical for validating the identified features. Furthermore, for those potential predictors with higher importance and higher SHAP values, we compared the change rates of raw variables across different groups to explore the dynamic progression between visits:

$$\text{Change Rate}=(x_i - x_j)/(t_i - t_j)$$

Here, t_i, t_j indicate the last and the first visit time for a raw variable respectively; x_i *means the raw value of the variable at the visit time* t_i *and* x_j *denotes the raw value of the variable at the visit time* t_j.

3.Results

3.1 The prediction effectiveness of machine learning models

Macro-averaged method was used to evaluate sensitivity, specificity, precision, accuracy, and F1 score for each group. We observed that Random Forest consistently outperformed other machine learning models across most metrics (Figure 2). Notably, sensitivity was particularly high for three participant groups — StableHC, StableAD, and UnstableConvertibleMCI — across five models. For instance, Random Forest achieved sensitivities of 97.84% (StableHC), 88.31% (StableAD) and 85.55% (UnstableConvertibleMCI). In contrast, specificity showed minimal variation, ranging narrowly from 86.39% to 100% across all groups and models.

Machine learning performance of other metrics — precision, accuracy, F1 score— ranges across participant groups as follows:
(1) Random Forest: precision 65.95% - 100%, accuracy 91.38%-98.61%, and F1 score 54.54%-91.91%; *(2) Support Vector Machines*: precision 37.5%-98.59%, accuracy 88.05%-97.77%, and F1 score 40%-94.59%. (3) *Radial Basis Function Networks*: precision 50%-96.96%, accuracy 86.11%-98.05%, and F1 score 46.15%-89.51%. (4) *Backpropagation Networks*: precision 44.11%-100%, accuracy 84.44%-98.61%, and F1 score 42.85%-92.10%. (5) *Convolutional Neural Network*: precision 54.83%-100%, accuracy 90% - 99.16% and F1 score 48.48% - 93.42%.

The detailed confusion matrix analysis confirms that while all machine learning models achieved accuracies exceeding 70%, Random Forest provided the most reliable overall classification across participant groups. This model showed particularly robust performance in identifying StableHC cases (Figure 3).

To further validate the prediction capability of Random Forest, we calculated the out-of-bag error, which remained below 0.2 when using 50 decision trees (Figure 4). Given these findings, we selected Random Forest for subsequent feature importance analyses in the ADNI cohort.

3.2 Demographics

Participants with seven longitudinal trajectory types showed no significant differences in certain demographics, including handedness, race, ethnicity, and primary language (Supplementary Table 1). However, the StableAD group had a lower BMI and fewer years of education compared to StableHC and StableMCI groups (Figure 5). Additionally, age varied across groups: the StableAD group was older than the StableHC and StableMCI groups, while the UnstableReverse group was significantly younger than other unstable groups. Gender and marital status also differed among groups. For instance, the proportion of females was higher in the StableHC, UnstableConvertibleHC and UnstableProgress groups compared to others, while married participants were more prevalent in the StableAD group.

Analysis of family history and APOE4 allele revealed that APOE4 significantly correlated with the etiology of AD, which carrying one or two alleles increased the likelihood of StableAD or UnstableConvertibleMCI development (Supplementary Table 2). Additionally, family history — particularly having a father ($\chi2=27.20$, $p<0.01$) or sibling with AD ($\chi2=27.67$, $p<0.01$) (rather than other forms of dementia) — appeared more strongly associated with AD risk than a maternal history ($\chi2=19.11$, $p>0.05$). The population ratios of APOE4 with 0, 1 and 2 alleles across groups stratified by longitudinal trajectories were significantly different ($\chi2=170.97$, $p<0.001$):

StableHC: 71% with 0 allele, 27% with 1 allele, 2% with 2 alleles

StableMCI: 62% with 0 allele, 31% with 1 allele, 7% with 2 alleles

StableAD: 30% with 0 allele, 49% with 1 allele, 21% with 2 alleles

UnstableReverse: 64% with 0 allele, 31% with 1 allele, 5% with 2 alleles

UnstableConvertibleHC: 65% with 0 allele, 30% with 1 allele, 5% with 2 alleles

UnstableConvertibleMCI: 35% with 0 allele, 50% with 1 allele, 15% with 2 alleles

UnstableProgress: 48% with 0 allele, 52% with 1 allele

3.3 Importance Values and Potential Predictors in Random Forest

Random Forest importance values were used to identify the potential predictors of participants' longitudinal trajectories. Potential predictors of longitudinal trajectories in participants were defined as those meeting two criteria: (1) ranking in the top 1.2% of absolute beta coefficients (≥0.1) in Lasso regression; (2) ranking in the top 17.2% of absolute importance values (≥0.15) in Random Forest. As shown in Figure 6-Figure 8 (for details, see supplementary 3), amyloid uptake emerged as a key predictor across all analyses (global level, temporal variation or causal inference). As shown in Figure 7, temporal variation in Clinical Dementia Rating global scores (CDGLOBAL_sem) and Logical Memory delayed recall scores (LDELTOTAL_sem) were strongly associated with participant groups and disease progression. The visuospatial dysfunction also appeared predictive, as indicated by the importance values of causal inference measures (Figure 8): clock copying task performance (COPYSCORE_beta) and visual-spatial dysfunction in everyday cognition (EcogPtVisspat_beta). Finally, causal inference analysis confirmed that plasma APOE4 levels may help predict participant groups. For those potential predictors, subgroup analyses for three intra-individual metrics were conducted (as shown in Table 1).

3.4 Cognition Dysfunction Predictors

Using Random Forest, we quantified the predictive importance of cognitive measures for AD trajectories (Table 2). Global scores from the Clinical Dementia Rating Scale (Temporal Variation Importance=0.25, Lasso β=1.76), Mini-Mental State Examination Scale (Causal Inference Importance=0.13, Lasso β=0.15), and Everyday Cognition Scale (Causal Inference Importance=0.20, Lasso β=-0.49) showed stronger predictive value for overall cognitive decline than ADAS13 (all Lasso β=0) or MOCA (|Lasso β|$\leq$0.03) global scores.

Regarding specific cognitive domains, visuospatial function and memory function emerged as particularly important predictors of AD progression (see Figure 9). These included performance on: clock drawing task, ADAS13 sub-items (object naming, orientation, word recognition), and logical memory II delayed recall.

3.5 Neuroimaging and Biospecimen Predictors

Our analyses revealed that the 18F-AV45 PET demonstrated superior predictive value for MCI/AD progression compared to T1 MRI and FDG PET (Table 3). Amyloid standard uptake value ratios (SUVRs) showed particular predictive significance in the parahippocampus, ventricles, and caudal cingulate regions.

Global amyloid deposition patterns were predictive across multiple regions:

Bilateral anterior cingulate: Global Level Importance=0.14, Lasso β =0.12;

Left amygdala: Global Level Importance=-0.14, Lasso β =0.58;

Left thalamus: Global Level Importance=0.10, Lasso β =0.28;

Right Lingual: Global Level Importance=0.15, Lasso β =-0.10;

Right temporal pole: Global Level Importance=0.16, Lasso β =-0.12;

Right caudate: Global Level Importance=0.14, Lasso β =0.24.

Temporal variation in amyloid uptake implicated a distributed neural network:
Bilateral inferior lateral ventricles: Temporal Variation Importance=0.14, Lasso β =0.24;
Bilateral parahippocampus: Temporal Variation Importance=0.20, Lasso β =-2.10;
Left choroid plexux: Temporal Variation Importance=0.13, Lasso β =0.37;
Left caudal middle frontal: Temporal Variation Importance=0.14, Lasso β =-0.60.

Notably, blood-based biomarkers also showed classification utility for the etiology of MCI/AD:
Long Neurofilament light chain (NFL): Causal Inference Importance=0.28, Lasso β =-0.17;
plasma APOE4: Causal Inference Importance=0.29, Lasso β =1.35 (Figure 10).

3.6 *Validation by Integrating Convolutional Neural Network with SHAP Interpretability Values*

Figure 11 reveals that high-SHAP-value predictors encompass three major categories: (1) Clinical Dementia Rating metrics (*global level*: CDGLOBAL for StableHC, StableMCI, and UnstableConvertibleHC cohorts; *causal inference*: CDSOB for StableAD and UnstableConvertibleMCI cohorts) and Logical Memory scores; (2) plasma APOE4 status and dementia family history (*global level*: FHQMOM for StableMCI, and FHQDAD for StableAD and UnstableReverse); and (3) neuroimaging and fluid biomarkers (*temporal variation*: ventricular amyloid uptake for StableAD group, lingual/temporal/hippocampus volume changes for UnstableConvertibleHC and StableAD, UnstableProgress groups, and *causal inference*: Cerebral Spinal Fluid Aβ40/Aβ42 ratio for ConvertibleMCI and UnstableProgress groups). These predictors identified by convolutional neural networks and validated via interpretability

tools (e.g., SHAP values) were consistent with the feature importance scores derived from the Random Forest algorithm (for more details, see supplementary 4).

We compared the rates of change for the potential predictors with higher importance and SHAP values — Clinical Dementia Rating Sum of Boxes Scores, Logical Memory Scores, amyloid uptake from AV45 Positron Emission Computed Tomography, brain volumes (ventricles, hippocampus) — to further validate the dynamic progression between visits (Figure 12).

4.Discussion

We evaluated predictors to complex MCI/AD progression patterns, encompassing stable, convertible, and reversible disease trajectories. Our results demonstrate that machine learning approaches — particularly the Random Forest algorithm —effectively classify these progression patterns among healthy, MCI, and AD states. The Random Forest model achieved an average sensitivity of 70.28% and specificity of 96.28% across all participant groups. Importantly, this method generates feature importance weights that identified the most predictive biomarkers across cognitive assessments, neuroimaging measures, and biospecimen analyses.

This study evaluated four machine learning models for classifying progression along the healthy-MCI-AD continuum. We systematically integrated multi-modal inputs — demographic data, cognitive measures, MRI/PET neuroimaging, and plasma/CSF biomarkers— into distinct model architectures. Our approach facilitates intuitive visualization of individual healthy-MCI-AD continuum risk profiles while advancing understanding of neurodegenerative pathophysiology. Among the tested algorithms, Random Forest model (50 decision trees) demonstrated superior diagnostic accuracy in the ADNI cohort. This finding aligns with growing cross-disciplinary

efforts to apply data-driven computational approaches in AD research [21-22]. For instance, Qiu et al developed an individual-level interpretable deep learning framework for extracting AD signatures from heterogeneous data sources [11]. Looking ahead, machine learning will become increasingly vital for AD diagnosis, particularly in modeling the complex interactions among diverse risk factors. Our methodology not only improves disease classification but also establishes a framework for linking computational predictions with underlying biological mechanisms.

Our classification of participants along the health-to-MCI-to-AD continuum aligns with existing literature demonstrating heterogeneous disease trajectories, where some MCI patients progress to dementia while others may revert to cognitively normal states [19-22]. Growing evidence suggests that MCI/AD progression can be modified through lifestyle intervention and targeted therapies [23]. For instance, sodium oligomannate has been shown to modulate gut microbiota and attenuate neuroinflammation, thereby potentially slowing AD progression [24]. Longitudinal studies further indicate that physical activities like fieldwork or gardening may enhance the likelihood of MCI reversion [25]. A recent meta-analysis of 25 studies reported an aggregated reversion rates of approximately 24% [26]. Leveraging the ADNI cohort with >4 years of follow-up data, our study provides unique insights into factors associated with dynamic transitions along the healthy-MCI-AD continuum.

Our demographic analyses revealed significant gender, BMI, and marital status differences across the healthy-MCI-AD continuum. For example, stable healthy group showed a higher proportion of females compared to disease-progressing groups. Sex differences across

participants indicate that there are complex biological mechanisms in the etiology of MCI/AD, involving multiple risk factors such as longevity, neuroanatomical variations, hormonal influences, genetic predisposition, and immune system modulation [27-28]. Our study also identified significant lower BMI in the stable AD group compared to other groups, potentially implicating gut microbiota composition in AD progression risk [29].

A defining feature of AD is the presence of extracellular amyloid-beta plaques and intraneuronal neurofibrillary tangles. In the current study, several insightful predictors — including APOE4 genotype, plasma NFL, visuospatial cognition, episodic memory, parahippocampal structural alterations, and amyloid uptake — extracted by machine learning models are highly consistent with existing literature on AD. The progression of AD involves dysfunction across multiple cellular and molecular processes. Accumulating evidence suggests that key contributing factors such as APOE4, aβ, and tau disrupt a myriad of biological pathways, leading to neuroinflammation, neuropathological lesions, and neural network deficits, which ultimately manifest as clinically observable cognitive decline [30]. The heterogenous pathological hallmarks of AD include [31]: (1) aggregation of aβ peptides and formation of amyloid plaques; (2) hyperphosphorylation and aggregation of tau, followed by its propagation to connected neurons; (3) neuroinflammation mediated by microglia and astrocytes.

According to our findings, key distinguishing factors among MCI/AD progression include education level, age, and family history. Specifically, APOE4 carriers with a self-report family history of AD demonstrated elevated risk for disease progression. Interestingly, paternal or sibling history of AD showed stronger association with pathological development than maternal

history or other dementia types in family members. This aligns with previous findings that sibling-reported family history improves AD diagnostic sensitivity to >90% [32]. Furthermore, growing evidence supports the association between paternal dementia and biomarkers profiles (e.g. CSF Aβ, tau/Aβ ratio, and global PiB uptake) [33], suggesting family history patterns may help elucidate genetic mechanisms in AD pathogenesis.

The predictors of MCI/AD pathology form a complex network encompassing cognitive dysfunction, neuroimaging biomarkers, and biospecimen biomarkers. For cognitive assessment, global measures like Clinical Dementia Rating and Mini-Mental State Examination scores provide meaningful evaluations of overall impairment. The Random Forest quantified the outperformance of these scales compared to MOCA and ADAS13 scales. Furthermore, visuospatial and memory-related dysfunction offer particular valuable insights for differential predicting MCI/AD longitudinal trajectories. Our results suggest that disorientation and visuospatial deficits may serve as particularly robust diagnostic markers for AD, consistent with existing literature [33-35]. For example, Yew et al. demonstrated that disorientation appears specific to AD, likely stemming from posterior hippocampus deficits, while frontal-temporal dementia patients typically maintain relatively preserved orientation abilities [33]. Additionally, our findings align with previous reports of object naming and word recognition impairments in AD [36-38], which may reflect degeneration in conceptual/lexical retrieval systems and semantic networks characteristic of AD pathology.

Our results identified several amyloid-related neuroimaging biomarkers as significant predictors of disease progression. Notably, global amyloid deposition in key neural networks —

particularly the caudal anterior cingulate, amygdala, thalamus, lingual cortex, temporal pole and caudate— demonstrated strong predictive value for healthy-MCI-AD continuum. These regions are critically involved in executive attention, emotional processing, and related cognitive functions [39]. Longitudinal analyses further revealed the importance of temporal changes in amyloid accumulation within specific areas, including the inferior lateral ventricles, parahippocampus, and caudal middle frontal cortex. These findings align with established literature demonstrating impaired parahippocampal connectivity in MCI and AD, which correlates with characteristic episodic memory deficits [40].

Additionally, we identified plasma long NFL as a promising predictor for tracking progression along the healthy-MCI-AD continuum. While plasma biomarkers (including amyloid-beta 42/40 ratio, pTau181, and NFL, glial fibrillary acidic protein) remain controversial due to varying clinical utility [41]. NFL shows particular promise as a marker of axonal damage. This neuronal cytoplasmic protein, expressed in large-caliber myelinated axons, has been consistently associated with various neurological disorders including neurodegenerative, inflammatory, and cerebrovascular diseases [42-43].

The core value of leveraging machine learning algorithms to identify potential disease predictive indicators lies in promoting the transformation of the medical model from "passive diagnosis and treatment" to "proactive prevention and precision stratified management". Machine learning can integrate multi-dimensional data (including clinical signs, laboratory indicators, imaging features, lifestyle factors, genomics/proteomics data, and electronic health records) to identify hidden predictive markers that are difficult to detect via traditional statistical methods. For high-

risk populations, personalized and intensified intervention plans can be formulated; moreover, disease risk prediction indicators based on machine learning can assist in developing regional prevention and control strategies. By screening biomarkers with both high sensitivity and specificity (e.g. APOE4, Clinical Dementia Rating, Amyloid Uptake and Ventrical/Parahippocampus Atrophy), machine learning models integrated with AD convertible risk or MCI reverse prediction indicators can be embedded into clinical decision-making systems.

In the present study, three core metrics—global-level characteristics, temporal variability, and causal inference—were employed to quantify the longitudinal trajectories of the Healthy-to-MCI-to-AD continuum. However, capturing the dynamic, time-dependent nature of disease progression is inherently complex, and this challenge will likely necessitate the development of more sophisticated analytical frameworks in future research. For instance, a novel integrated algorithm that incorporates multiple temporal parameters (e.g., coefficient of variation, intra-timepoint standard deviation, rate of clinical/biomarker change, and longitudinal phenotypic patterns) would offer greater interpretive power for delineating the heterogeneous dynamic progression profiles of participants with stable, progressive, or even reversible cognitive statuses.

5.Conclusion

In conclusion, machine learning approaches provide valuable tools for identifying key predictors of cognitive decline and neural deterioration along the healthy-MCI-AD continuum. Our comprehensive comparison of machine learning models demonstrates the potential of Random Forest to enhance diagnostic accuracy in clinical practice. These data-driven methods offer clinicians objective evidence to support diagnostic decision-making, while also advancing our

understanding of disease mechanisms. Future research should focus on translating these computational insights into targeted prevention strategies and personalized therapeutic interventions.

Contributors:

Yujing Huang, Gaoyi Yang, Qingguo Ma: Conceptualization, methodology, Writing-Original Draft Preparation; **Hao Zhang**: Visualization, Investigation; **Buqing Ma, Zhe Yu, Li Su**: Data Curation; **Shenyi Dai, Lu Cheng**: Validation, Manuscript Revision; **ADNI**: Data Source.

Declaration of interests

The authors declare no competing interests.

Acknowledgements

We appreciate the funding from Zhejiang Provincial Key Laboratory Construction Project (2024ZY01026), National Key R&D Program of China (2024YFA1306600), National Key R&D Program of China (2024YFA1306603). We appreciated the comments from Ju-Sheng Zheng for polishing the manuscript.

Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health

(FNIH) including generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:
http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

References

[1] J.E. Ahlskog, Y.E. Geda, N.R. Graff-Radford, et al, Physical exercise as a preventive or disease-modifying treatment of dementia and brain aging, Mayo Clin. Proc. 86 (2011), pp. 876-884.

[2] E.D Huey, Course and etiology of dysexecutive MCI in a community sample, Alzheimers Dement. 9 (2013), pp. 632-639.

[3] C. Hu, D. Yu, X. Sun, et al, The prevalence and progression of mild cognitive impairment among clinic and community populations: a systematic review and meta-analysis, Int. Psychogeriatr. 29(10) (2017) , pp. 1595-1608.

[4] M Malek-Ahmadi, Reversion from mild cognitive impairment normal cognition: a meta-analysis, Alzheimer Dis. Assoc. Disord. 30 (2016), pp. 324-330.

[5] K. Makino, S. Lee, S. Bae, et al, Diabetes and Prediabetes inhibit reversion from mild cognitive impairment to normal cognition, J. Am. Med. Dir. Assoc. 22(9) (2021), pp. 1912-1918.

[6] M. Tsujimoto, K. Suzuki, N. Saji, et al, Organized registration for the assessment of dementia by the nationwide general consortium toward effective treatment (ORANGE) registry: current status and perspectives of mild cognitive impairment, J. Alzheimers Dis. 88(4) (2022), pp. 1423-1433.

[7] I. Gueorguieva, L. Chua, B.A. Willis, et al, Disease progression model using the integrated Alzheimer's Disease rating scale, Alzheimers Dement. 19(6) (2023), pp. 2253-2264.

[8] S. Jamalian, M. Dolton, P. Chanu, et al, Alzheimer's Disease Neuroimaging Initiative, CPT Pharmacometrics Syst. Pharmacol. 12(7) (2023), pp. 1029-1042.

[9] J. Chen, J. Yang, D. Shen, et al, A predictive model of the progression to Alzheimer's Disease in patients with mild cognitive impairment based on the MRI enlarged perivascular spaces, J. Alzheimers Dis. 101(1) (2024), pp. 159-173.

[10] X. Gao, J. Wang, Y. Wang, et al, Combining multiple factors to predict Alzheimer's Disease, Neurosci. Bull. 38 (2022), pp. 969-972.

[11] S. Qiu, P.S. Joshi, M.I. Miller, et al, Development and validation of an interpretable deep learning framework for Alzheimer's disease classification, Brain 143(6) (2020), pp. 1920-1933.

[12] D. Castelvecchi, Can we open the black box of AI? Nature 538 (2016), pp. 20-3.

[13] A. Loe, S. Murray, Z. Wu, Random forest for dynamic risk prediction of recurrent events: a pseudo-observation approach, Biostatistics 26(1) (2024), pp. kxaf007.

[14] V. Franc, V. Hlavac, Multi-class support vector machine. In: Proceedings of the IEEE 16th International Conference on Pattern Recognition. 2002. 2. pp. 236-239.

[15] D. Rad, G.C.Jr. Magulod, E. Balas, et al, A Radial Basis Function neural betwork approach to predict preschool teachers' technology acceptance behavior, Front. Psychol. 13 (2022), pp. 880753.

[16] M. Rapp, Y. Yarom, I. Segev, Modeling back propagating action potential in weakly excitable denrites of neocortical pyramidal cells. Proc. Natl. Acad. Sci. U. S. A. 93(21) (1996), pp. 11985-11990.

[17] M. Hartig, D. Truran-Sacrey, S. Raptentsetsang, et al, UCSF Freesurfer Methods. Alzheimer's Disease Neuroimaging Initiative (2012).

[18] S. Landau, R. Koeppe, W. Jagust, et al, Florbetaben processing and positivity threshold derivation. Alzheimer's Disease Neuroimaging Initiative (2011).

[19] S. Landau, W Jagust, Flortaucipir (AV-1451) processing methods. Alzheimer's Disease Neuroimaging Initiative (2016).

[20] S. Landau, W. Jagust, UC Berkeley FDG MetaROI methods. Alzheimer's Disease Neuroimaging Initiative (2022).

[21] H. Geerts, M. Hofmann-Apitius, T.J. Anastasio, Brain Health Modeling Initiative. Knowledge-driven computational modeling in Alzheimer's disease research: current state and future trends, Alzheimers Dement. 13(11) (2017), pp. 1292-1302.

[22] B. Yew, S. Alladi, M. Shailaja, et al, Lost and forgotten? Orientation versus memory in Alzheimer's disease and frontotemporal dementia, J. Alzheimers Dis. 33(2) (2013), pp. 473-481.

[23] R.V. Rao, K.G. Subramaniam, J. Gregory, et al, Rationale for a multi-factorial approach for the reversal of cognitive decline in Alzheimer's Disease and MCI: a review, Int. J. Mol. Sci. 24(2) (2023), pp. 1659.

[24] X. Wang, G. Sun, T. Feng, et al, Sodium oligomannate therapeutically remodels gut microbiota and supresses gut bacterial amino acids-shaped neuroinflammation to inhibit Alzheimer's disease progression, Cell Res. 29(10) (2019), pp. 787-803.

[25] H. Shimada, T. Doi, S. Lee, et al, Reversible predictors of reversion from mild cognitive impairment to normal cognition: a 4-year longitudinal study, Alzheimers Res. Ther. 11(1) (2019), pp. 24.

590 [26] M. Malek-Ahmadi, Reversion from mild cognitive impairment to normal cognition: a meta-
591 analysis, Alzheimer Dis. Assoc. Disord. 30(4) (2016), pp. 324-330.
592 [27] D. Zhu, A. Montagne, Z Zhao, Alzheimer's pathogenic mechanisms and underlying sex
593 difference, Cell Mol. Life Sci. 78(11) (2021), pp. 4907-4920.
594 [28] E.M. Arenaza-Urquijo, R. Boyle, K. Casaletto, et al, Sex and genderr differences in
595 cognitive resilience to ageing and Alzheimer's disease, Alzheimers Dement. 20(8) (2024), pp.
596 5695-5719.
597 [29] B.J.H. Verhaar, H.M.A. Hendriksen, F.A. de Leeuw, et al, Gut microbiota composition is
598 related to AD pathology, Front. Immunol. 12 (2022), pp. 794519.
599 [30] N. Koutsodendris, M.R. Nelson, A. Rao, Y. Huang. Apolipoproein E and Alzheimer's
600 Disease: Findings, Hypotheses, and Potential Mechanisms. Annu. Rev. Pathol. Mech. Dis.
601 17(2022), pp. 73-99
602 [31] C.R.J. Jr, D.A. Bennett, K. Blennow, et al. NIA-AA Research Framework: Toward a
603 biologica definition of Alzheimer's disease. Alzheimers Dement. 14(4) (2018), pp. 535-562.
604 [32] G. Devi, K. Marder, P.W. Schofield, Validity of family history for the diagnosis of dementia
605 among siblings of patients with late-onset Alzheimer's disease, Genet. Epidemiol. 15(3) (1998),
606 pp. 215-223.
607 [33] R.A. Honea, E.D. Vidoni, R.H. Swerdlow, J.M. Burns, et al, Maternal family history is
608 associated with Alzheimer's disease biomarkers, J. Alzheimers. Dis. 31(3) (2012), pp. 659-668.
609 [34] D. Bazadona, I. Fabek, M.B. Leko, et al, A non-invasive hidden-goal test for spatial
610 orientation deficit detection in subjects with suspected mild cognitive impairment, J. Neurosci.
611 Methods 332 (2020), pp. 108547.

[35] S. Tu, S. Wong, J.R. Hodges, et al, Lost in spatial translation-A novel tool to objectively assess spatial disorientation in Alzheimer's disease and frontotemporal dementia, Cortex 67(2015), pp. 83-94.

[36] D.J. Williamson, J.C. Adair, A.M. Raymer, et al, Object and action naming in Alzheimer's disease, Cortex 34(4) (1998), pp. 601-610.

[37] F. Cuetos, N. Arce, C. Martinez, et al, Word recognition in Alzheimer's disease: Effects of semantic degeneration, J. Neuropsychol. 11(1) (2017), pp. 26-39.

[38] M. Alenius, L. Hokkanen, S. Koskinen, et al, Cognitive performance at time of AD diagnosis: A clinically augmented register-based study, Front. Psychol. 13(2022), pp. 901945.

[39] A. Etkin, T. Egner, R. Kalisch, Emotional processing in anterior cingulate and medial prefrontal cortex, Trends Cogn. Sci. 15(2) (2011), pp. 85-93.

[40] J. Liu, X Zhang., C. Yu, et al, Impairment parahippocampus connectivity in mild cognitive impairment and Alzheimer's Disease, J. Alzheimers Dis. 49(4) (2016), pp. 1051-1064.

[41] Y. Yakoub, N.J. Ashton, C. Strikwerda-Brown, et al, Longitudinal blood biomarker trajectories in preclinical Alzheimer's disease, Alzheimers Dement. 19(12) (2023), pp. 5620-5631.

[42] L. Gaetani, K. Blennow, P. Calabresi, et al, Neurofilament light chain as a biomarker in neurological disorders, J. Neurol. Neurosurg. Psychiatry 90(8) (2019), pp. 870-881.

[43] L. Chen, L. Sun, J. Sun, J. Wang, D. Zhang, M. Xia et al. Brain-clinical signatures of basal ganglia-related dysfunctional reorganization in Parkinson's disease. eBioMedicine, 120(2025), pp. 105917.

Figure Legends

Figure 1. Participant Classification Based on Different Longitudinal Trajectories. *StableHC* group: participants consistently diagnosed as healthy from baseline to ≥48 months. *StableMCI* group: participants consistently diagnosed as MCI from baseline to ≥48 months. *StableAD* group: participants consistently diagnosed as AD from baseline to ≥12 months. *UnstableReverse* group: participants diagnosed with MCI or AD at baseline. However, during follow-up, these MCI or AD patients reverted to normal cognition or a reduced dementia severity state respectively. *UnstableConvertibleHC* group: participants are healthy at baseline, but converted to MCI during follow-up. *UnstableConvertibleMCI* group: patients diagnosed as MCI at baseline but converted to AD during follow-up. *UnstableProgress* group: participants are healthy at baseline. During follow-up, their conditions gradually worsened, converting to MCI, and then AD. The x-axis represents the follow-up time, with the unit of "months". Different colors represent the diagnostic outcomes. Blue=Healthy, yellow=MCI, red= AD. The Patient RID is the patient registration number in the ADNI cohort. Each row indicates a participant's diagnostic trajectory over follow-up visits.

Figure 2. Effectiveness of Machine Learning Models: the performance of four machine learning models —Random Forest (RF), Support Vector Machines (SV), Radial Basis Function Networks (RBF), Backpropagation Networks (BP) , and Convolutional Neural Network (CNN) — was assessed by using the following metrics: the sensitivity, specificity, precision, accuracy and F1 score.

Figure 3. Confusion Matrices of Machine Learning Models: For each machine learning algorithm, the dataset was split into training set (70% of participants) and testing set (30% of participants). The classification accuracy was evaluated on the testing set for all seven participant groups: StableHC, StableMCI, StableAD, UnstableReverse, UnstableConvertibleHC, UnstableConvertibleMCI, and UnstableProgress. The actual class vs. predicted class labels were visualized using confusion matrices to illustrate classification performance.

Figure 4. Random Forest-Specific Analysis: the out-of-bag error rate was computed across 0 to 50 decision trees (X axis: number of decision tress; Y axis: error values). Importance values quantified the contribution of each feature as predictive biomarkers in classifying participant groups. A positive importance value indicates a positive association between the feature and the outcome (participant group classification). A negative importance value suggests an inverse relationship.

Figure 5. Demographic Differences. (1) For the continuous variables (Education, BMI, Age) in the dual-axis plots, left Y axis is 'Education' (years) and 'BMI' (body mass index) while right axis is 'Age' (years). Mean values with standard error bars were shown for all participant groups. For post-hoc comparisons, 'Age' used least significant difference (LSD) while 'Education' and 'BMI' used Tamhane's T2. (2) For the categorical variables (Gender, Marriage Status), Y axis and X axis indicates the participant number and participant groups respectively. 'Non-Married' combines the status of 'widowed', 'divorced', 'never married' and 'unknown'. Chi-square tests for gender and marriage status comparisons. Significance levels (2-sided asymptotic) are illustrated as follows: '*', '**', '***' for $p<0.05$, $p<0.01$, $p<0.001$ respectively.

Figure 6. Random Forest Importance Values for Feature Global Level. 'Global level' represents overall measure during follow-up. The green dots marked the potential predictors which achieved absolute beta coefficients (≥0.1) in Lasso regression and importance values (≥0.15) in Random Forest while the grey dots failed to meet the criteria. A positive importance value indicates a positive association between the feature and the outcome (participant group classification) while a negative importance value suggests an inverse relationship.
Abbreviations: EMBIC_R3_avg= Probability of retrieving from the DURABLY LEARNED State after a 5-minute delay with distraction in the task EMBICDCB (Embic Corporation Digital Cognitive Biomarkers); **Eightmm_CTX_RH_LINGUAL_SUVR_avg**= right-lingual cortex SUVR value from UC Berkeley - AV45 analysis;
Eightmm_CTX_RH_TEMPORALPOLE_SUVR_avg=right-temporalpole cortex SUVR value from UC Berkeley - AV45 analysis; **PETsum**=Summaries of Positron Emission Tomography Results; **FDG PET**=Fluorodeoxyglucose Positron Emission Tomography Results; **CSF**= Cerebral Spinal Fluid

Figure 7. Random Forest Importance Values for Feature Temporal Variation. 'Temporal Variation' indicates visit-to-visit longitudinal fluctuations. The green dots marked the potential predictors which achieved absolute beta coefficients (≥0.1) in Lasso regression and importance values (≥0.15) in Random Forest while the grey dots failed to meet the criteria. A positive importance value indicates a positive association between the feature and the outcome (participant group classification) while a negative importance value suggests an inverse relationship**. Abbreviations**: **FAQ_sem**=Functional Assessment Questionnaire; **CDGLOBAL_sem**=Global Clinical Dementia Rating; **LDELTOTAL_sem**=Logical Memory -

Delayed Recall; **Eightmm_CTXPARAHIPPOCAMPAL_SUVR_sem**=parahippocampal-cortex SUVR value from UC Berkeley - AV45 analysis; **Eightmm_Right_VESSEL_SUVR_sem**=right-vessel SUVR value from UC Berkeley - AV45 analysis; **Eightmm_CTX_LH_PARAHIPPOCAMPAL_SUVR_sem**=left-parahippocampal-cortex SUVR value from UC Berkeley - AV45 analysis; **Eightmm_CC_CENTRAL_SUVR_sem**=central SUVR value from UC Berkeley - AV45 analysis'; **RIGHT_CEREBELLUM_WHITE_MATTER_SUVR_sem**=right-cerebellum-white-matter SUVR from UC Berkeley - AV45 analysis;**UCB_FDGAngRight_MIN_sem**=Min of FDG-PET within Right Angular Gyrus from UC Berkeley -FDG analysis; **PETsum**=Summaries of Positron Emission Tomography Results; **FDG PET**= Fluorodeoxyglucose Positron Emission Tomography Results; **CSF**=Cerebral Spinal Fluid.

Figure 8. Random Forest Importance Values for Feature Causal Inference. 'Causal Inference' quantifies the association between follow-up time and variable changes. The green dots marked the potential predictors which achieved absolute beta coefficients (≥0.1) in Lasso regression and importance values (≥0.15) in Random Forest while the grey dots failed to meet the criteria. A positive importance value indicates a positive association between the feature and the outcome (participant group classification) while a negative importance value suggests an inverse relationship. Abbreviations: **CDRSB_beta**=Clinical Dementia Rating Sum of Boxes; **Other_Clinical_Q7SCORE_beta**=Question 7 Score from ADAS-Cognitive Behavior; **EcogPtVisspat_beta**= Participant ECog - Visual/Spatial Score from Everyday Cognition - Participant Self-Report; **EcogPtTotal_beta**=Participant ECog - Total Score from Everyday Cognition - Participant Self-Report; **EcogPtDivatt_beta**=Participant ECog - Division attention

726 Score from Everyday Cognition - Participant Self-Report; **EcogPtPlan_beta**=Participant ECog -
727 Plan Score from Everyday Cognition - Participant Self-Report; **COPYSCOR_beta**=Copy Clock
728 Total Score; **CLOCKSCOR_beta**=Draw Clock Total Score;
729 **BAI_AV45PET_MCSUVRWM_beta**=Mean Cortical SUVR Corpus Callosum and Centrum
730 Semiovale Combined from Banner Alzheimer's Institute PET NMRC AV45 Summaries;
731 **UCB_FDGCingBilat_STDEV_beta**= Standard Deviation of FDG-PET within Bilateral
732 Cingulate Cortex from UC Berkeley -FDG analysis;
733 **PLASMA_APOEQ1_APOEGenotype_beta**=PLASMA ApoE Genotyping - Results;
734 **PLASMA_APOEQ1_APOE4_beta**=APOE4 Results;
735 **BLENNOW_PLASMA_NFLLONG_PLASMA_NFL_beta**=Plasma neurofilament light
736 longitudinal results from Blennow Lab ADNI1-2 Plasma neurofilament light longitudinal
737 Analysis; **CSF_MSD_MSD_STREM2_CV_beta**=Haass CSF sTREM2 Intraplate Coefficient
738 of Variation (%);**CSF_MSD_MSD_PGRN_CV_beta**=Haass CSF PGRN Intraplate Coefficient
739 of Variation (%); **PETsum**=Summaries of Positron Emission Tomography Results; **FDG**
740 **PET**=Fluorodeoxyglucose Positron Emission Tomography Results; **CSF**=Cerebral Spinal Fluid.
741
742 **Figure 9. Cognitive Dysfunction Predictors**. The score distributions of specific cognitive
743 domains across follow-up visits are illustrated for each participant group. (1) *Object Naming*.
744 The scores (behavioral errors in the object naming task in ADAS13 scale, Question 5) range
745 from 0 to 5. (2) *Orientation*. The scores (behavioral errors in the temporal/spatial orientation task
746 in ADAS13 scale, Question 7) range from 0 to 8. (3) *Word Recognition*. The scores (behavioral
747 errors in the word recognition task in ADAS13 scale, Question 9) range from 0 to 5. (4) *Delayed*
748 *Recall*. The scores (accuracy in the delayed recall task from Logical Memory II scale) range

749 from 0-25. (5) *Clock Drawing*. The scores (behavioral dysfunction in the clock drawing task
750 from Clock Drawing Test) range from 0 to 5. (6) *Clock Copying*. The scores (behavioral
751 dysfunction in the clock copying task from Clock Drawing Test) range from 0 to 5. Y axis
752 indicates the follow-up time points [Baseline (BL) to month 120 (M120)]: bl, m06, m12, m24,
753 m36, m48, m60, m72, m84, m96, m108, m120. However, for the StableAD group, the follow-up
754 time points range from BL to month 48 (M48): bl, m06, m12, m24, m36, m48. X axis indicates
755 the participant number. The participant groups include StableHC, StableMCI, UnstableReverse,
756 UnstableConvertibleHC, UnstableConvertibleMCI, and UnstableProgress.
757
758 **Figure 10. Plasma Long NFL Predictor**. Plasma long Neurofilament light chain protein (NFL),
759 a blood-based biomarker, also showed classification utility for the etiology of MCI/AD. X axis
760 indicates the follow-up time points from baseline (bl) to month 60 (m60): bl, m12, m24, m48,
761 m60. However, for the StableAD group, the follow-up time points range from BL to month 24
762 (M24): bl, m06, m12, m24. Y axis means the NFL concentration (pg/ml).
763
764 **Figure 11. Features with Top Mean Absolute SHAP Values**. Top 11 features with high
765 absolute mean SHAP values were illustrated for different groups. **Abbreviations:**
766 **LIMMTOTAL**= Logical Memory - Immediate Recall Score; **CDGLOBAL**= Global Clinical
767 Dementia Rating Score; **CDRSB** = Clinical Dementia Rating Sum of Boxes Score;
768 **EMBIC_N2**= Probability of encoding into the TRANSIENTLY LEARNED State from Embic
769 Corporation Digital Cognitive Biomarkers; **EMBIC_M3** = Probability of delayed recall of a
770 durably stored episodic memory from Embic Corporation Digital Cognitive Biomarkers;
771 **EcogPtLang** = Participant Language Score from Everyday Cognition - Participant Self-Report

Scale; **EcogPtMem** = Participant Memory Score from Everyday Cognition - Participant Self-Report Scale; **HMSCORE**= Total Score of Modified Hachinski Scale; **UCC_FDGPNS_AVEREF**=Average value of reference region from UC Berkeley - Fluorodeoxyglucose analysis; **UCC_FDGPNS_ AVEASSOC** = Average frontal, parietal and temporal cortices, norm to reference region from UC Berkeley - Fluorodeoxyglucose analysis; **UCC_FDGPNS_AVEFRONT**= Average frontal cortex, norm to reference region from UC Berkeley - Fluorodeoxyglucose analysis; **UCC_AV45WM_AVEFRONT** = Average frontal cortex (norm to reference region) from UC Berkeley Positron Emission Tomography analysis; **UCC_AV45WM_X3SDSIGPXL**= Number of pixels with Zscores in the range [3, inf) SD from UC Berkeley PET analysis; **UCC_AV45WM_X2SDSIGPXL** = Number of pixels with Zscores in the range [2, 3) SD from UC Berkeley Positron Emission Tomography analysis; **UCC_AV45WM_AVEASSOC**= Average frontal, parietal and temporal cortices, norm to reference region from UC Berkeley Positron Emission Tomography analysis; **UCC_AV45CBL_X2SDSIGPXL** = Number of pixels with Zscores in the range [2, 3) SD from UC Berkeley Positron Emission Tomography analysis; **UCC_AV45CBL_ X3SDSIGPXL**= Number of pixels with Zscores in the range [3, inf) SD from UC Berkeley PET analysis; **BAI_AV45_ACI**=amyloid convergence index for AV-45 subjects from Banner Alzheimer's Institute PET NMRC Summaries; **BAI_FDG_SROI**= statistical region of interest developed and validated for optimal longitudinal Fluorodeoxyglucose - Positron Emission Tomography CMRgl changes from Banner Alzheimer's Institute Positron Emission Tomography Summaries; **BAI_AV45_MCSUVRWM3** = Mean Cortical SUVR with Corpus Callosum and Centrum Semiovale Combined as reference region from Banner Alzheimer's Institute Positron Emission Tomography Summaries; **ADAS Q8 Score**= Question 8 Word Recognition from Alzheimer's

Disease Assessment Scale; **ADAS Q7 Score**=Question 7 Orientation from Alzheimer's Disease Assessment Scale; **VENTRICLE_3RD_SUVR**=3rd-ventricle Amyloid uptake from UC Berkeley - AV45 analysis; **TRANSVERSETEMPORAL_VOLUME** = transversetemporal Region of Interest size in mm^3 from UC Berkeley - AV45 analysis; **MOCA**= Montreal Cognitive Assessment Score; **SUMMARYSUVR_COMPOSITE**= Summary florbetapir cortical SUVR normalized by composite ref region from UC Berkeley - AV45 analysis; **FHQDAD**=Dad Dementia history from Family History Questionnaire; **FHQDADAD**=Dad Alzheimer's Dementia history from Family History Questionnaire; **FHQMOM**= Mother Dementia history from Family History Questionnaire; **FHQSIB**=Sibling Status from Family History Questionnaire; **TRABSCOR**= Trail Making Test Part B - Time to complete; **LH_LINGUAL_VOLUME**=Left Hemisphere Lingual Volume from UC Berkeley - AV45 analysis; **CTX_PARACENTRAL_VOLUME**=Paracentral ROI size in mm^3 from UC Berkeley - AV45 analysis; **UCB_FDGTempLeft_MODE**= Mode of Fluorodeoxyglucose - Positron Emission Tomography Results within Left Temporal Areas from UC Berkeley - Fluorodeoxyglucose analysis; **LH_INFERIORTEMPORAL_VOLUME**=Left Hemisphere Inferior Temporal Volume from UC Berkeley - AV45 analysis; **Left_Hippo**=Left Hippocampus Volume from UC Berkeley - AV45 analysis.

Figure 12. Change Rate of Potential Predictors. Change rates were compared across different groups. **Abbreviations: CDRSB**=Clinical Dementia Rating Sum of Boxes Scores; **PET**= Fluorodeoxyglucose Positron Emission Tomography Results.