



Machine learning for missing data imputation in Alzheimer's research: predicting medial temporal lobe dynamic flexibility

Soodeh Moallemian¹ · Abolfazl Saghafi² · Rutvik Deshpande¹ · Jose M. Perez¹ · Miray Budak¹ · Bernadette A. Fausto¹ · Fanny M. Elahi³ · Mark A. Gluck¹

Received: 2 December 2025 / Revised: 15 January 2026 / Accepted: 24 January 2026
© The Author(s), under exclusive licence to Springer Nature B.V. 2026

Abstract

Alzheimer's disease (AD) pathology begins years before symptoms appear, and dynamic flexibility of the medial temporal lobe (MTL) may serve as an early functional biomarker. Using data from 656 older adults in the Rutgers Aging and Brain Health Alliance study, we evaluated whether cognitive, genetic, biochemical, and demographic predictors could estimate MTL dynamic flexibility, despite substantial missingness (1,866 missing values; 25.86%). Only 42 participants (6.40%) had complete data; therefore, we compared case deletion with five imputation strategies (MICE, GAIN, MissForest, MIWAE, ReMasker) and eight regression models, assessing prediction accuracy using repeated 5-fold cross-validation. Complete-case analysis yielded limited performance (average $MAE = 0.220$, $CCC = 0.253$). After imputation, all methods improved accuracy, with MissForest paired with Bagging Trees or Random Forest achieving the lowest prediction error ($MAE = 0.186$). The greatest improvement in concordance occurred when GAIN was combined with Bagging Trees/Random Forest ($CCC = 0.464$), representing a 57% gain over the best complete-case model. A Scheirer–Ray–Hare ANOVA confirmed significant differences across imputation strategies ($p < 0.001$). Runtime analyses showed GAIN and MissForest to be both accurate and computationally efficient, while deep generative imputers were slower. These findings demonstrate that robust imputation is essential for maximizing data utility and predictive reliability in high-missingness neuroimaging studies and highlight the potential of ensemble tree models combined with advanced imputation techniques for estimating MTL dynamic flexibility in aging populations.

Keywords Alzheimer's disease · Missing value imputation · Missing completely at random · Machine learning

✉ Abolfazl Saghafi
asaghafi@sju.edu

Soodeh Moallemian
s.moallemian@rutgers.edu

Rutvik Deshpande
rvd39@rutgers.edu

Jose M. Perez
jose.mojicaperez@rutgers.edu

Miray Budak
miray.budak@rutgers.edu

Bernadette A. Fausto
bernadette.fausto@rutgers.edu

Fanny M. Elahi
fanny.elahi@mssm.edu

Mark A. Gluck
gluck@newark.rutgers.edu

- ¹ Center for Molecular & Behavioral Neuroscience, Rutgers, The State University of New Jersey, Newark, United States
- ² Department of Mathematics, Saint Joseph's University, Philadelphia, United States
- ³ Departments of Neurology, Neuroscience, Pathology, Molecular and Cell-based Medicine, Icahn School of Medicine at Mount Sinai, New York, United States

Introduction

Alzheimer's disease (AD), the most common form of dementia, is characterized by progressive cognitive decline and widespread neurodegeneration (DeTure and Dickson 2019). Pathological changes associated with AD can begin decades before the onset of clinical symptoms (Chen-Chen et al. 2014), underscoring the importance of early detection for the development of effective diagnostic and therapeutic strategies, particularly in the absence of a definitive cure. Integrating multimodal predictors provides a more comprehensive and potentially more accurate means of detecting early signs of AD by leveraging complementary information from cognitive, genetic, and biochemical domains.

MTL dynamic flexibility

The medial temporal lobe (MTL), particularly subregions such as the transentorhinal cortex and hippocampus, is among the earliest brain regions to undergo neurodegeneration in AD (Braak and Braak 1991; Moallemian et al. 2023). Functional connectivity, defined as the temporal correlation of neural activity between brain regions during resting-state functional magnetic resonance imaging (fMRI), has been shown to increase within the MTL in association with mild cognitive impairment (Das et al. 2013), age-related cognitive decline (Salami et al. 2014), and AD (Pasquini et al. 2015). Furthermore, a study involving healthy older individuals of African ancestry found that carriers of the ABCA7 rs115550680 G allele exhibited elevated MTL connectivity, which was linked to an increased risk of developing AD (Neha et al. 2019). These findings suggest that MTL functional connectivity may serve as a promising biomarker for early detection and tracking the progression of AD.

However, static functional connectivity measures fall short in capturing the dynamic nature of brain network interactions. Time-varying functional connectivity may offer increased sensitivity to early neural alterations in AD (Sizemore and Bassett 2018). One such dynamic measure is network flexibility, which quantifies how frequently a brain region reconfigures its connections over time (Bassett et al. 2011). Higher dynamic flexibility has been associated with improved memory performance in humans (Sinha et al. 2021; Budak et al. 2024). In preclinical stages of AD, alterations in MTL network dynamic flexibility may serve as early markers of pathology, correlate with cognitive decline, and reflect underlying neurodegeneration (Budak et al. 2024; Berron et al. 2020). Therefore, this study aims to predict MTL dynamic flexibility, a key marker for cognitive adaptability and effective information processing.

Cognitive

Older adults often exhibit reduced MTL dynamic flexibility in applying previously learned knowledge to new contexts, which contributes to difficulties in problem-solving

(Uddin 2021). Similarly, individuals in the preclinical stages of AD frequently show impairments in generalizing past experiences to novel situations (Petok et al. 2018). These deficits are believed to reflect age-related changes in the MTL and prefrontal cortex—regions essential for integrating and transferring information across different contexts (Azam et al. 2021). In this study, generalization, as measured by the Fish Task (Myers et al. 2003), is used as a cognitive predictor of MTL dynamic flexibility. Prior work has shown that performance on this task is associated with individual differences in MTL network dynamics, including flexibility, among older adults (Sinha et al. 2021).

In addition, acquisition performance, which reflects the ability to learn initial stimulus-outcome associations, is also modeled as a predictor. Acquisition is strongly dependent on hippocampal and broader MTL integrity, particularly in tasks requiring the formation of arbitrary or relational associations (Frank and Claus 2006). Prior studies have shown that better acquisition performance is associated with greater hippocampal activation and connectivity (Zeithamova et al. 2012). Given that MTL dynamic flexibility reflects the brain's capacity to dynamically reconfigure in response to new cognitive demands (Sinha et al. 2021), individuals with stronger acquisition performance may exhibit more adaptive and resilient MTL network dynamics. Therefore, both acquisition and generalization are included as complementary predictors of individual variability in MTL dynamic flexibility, shedding light on how learning processes relate to neural adaptability.

Genetic

Genetic variation plays a critical role in Alzheimer's disease, influencing both susceptibility and disease progression. In this study, two genetic biomarkers of AD—Apolipoprotein E (APOE) and ATP-binding cassette subfamily A member 7 (ABCA7)—measured via saliva samples, are examined as predictors of MTL dynamic flexibility. The $\epsilon 4$ isoform of APOE (APOE $\epsilon 4$) is the strongest known genetic risk factor for late-onset AD, associated with increased amyloid-beta ($A\beta$) accumulation, neuroinflammation, and synaptic dysfunction (Mecca et al. 2018; Butterfield and Johnson 2018). Conversely, the $\epsilon 2$ isoform (APOE $\epsilon 2$) appears to be protective, facilitating $A\beta$ clearance and reducing tau pathology.

ABCA7, another lipid transporter gene linked to AD, has been particularly associated with elevated risk in individuals of African ancestry (Reitz et al. 2013; Neha et al. 2019). Research suggests that ABCA7 modulates the processing of amyloid precursor protein (APP), affecting the generation and clearance of $A\beta$ peptides (Sakae et al. 2016). Loss-of-function mutations in ABCA7 can impair $A\beta$ phagocytosis and promote amyloid plaque formation—one of the pathological hallmarks of AD.

$A\beta$ accumulation begins years before clinical symptoms and is especially damaging to memory-related brain regions such as the hippocampus and surrounding MTL areas. $A\beta$ disrupts synaptic signaling, promotes oxidative stress, and impairs both structural and functional connectivity (Busche and Konnerth 2015; Selkoe and Hardy 2016). These disruptions can interfere with the brain's capacity for network dynamic flexibility, especially in the MTL. Given that dynamic flexibility reflects the ability to shift functional connections in response to changing demands, genetic variants that drive $A\beta$ pathology may lead to reduced neural adaptability. Thus, APOE and ABCA7 genotypes are included as predictors to evaluate their impact on dynamic MTL network properties in preclinical AD.

Biochemical

The intracellular accumulation of hyperphosphorylated tau, forming neurofibrillary tangles, is another central hallmark of AD (Jack et al. 2018). Elevated tau levels in cerebrospinal fluid (CSF) likely result from increased phosphorylation and neuronal secretion of tau in response to $A\beta$, marking early neurodegenerative changes (Maia et al. 2013). Blood-based tau biomarkers such as plasma P-tau181 have been shown to correlate with both amyloid and tau PET imaging (Mielke et al. 2018). Recent findings indicate that P-tau217 outperforms P-tau181 in sensitivity and specificity, offering a more robust marker of AD pathology due to its stronger correlation with tau-PET (Janelidze et al. 2020; Budak et al. 2024). Therefore, plasma P-tau217, measured through blood samples, is included as a biochemical predictor in this study.

Missing Values

A significant challenge in AD research is the prevalence of missing data, which is common in clinical and neuroimaging studies (Coley et al. 2011). Missing data in AD studies can arise from several factors. Logistical constraints, such as requiring participants to make multiple visits to a research or clinical facility, can lead to incomplete datasets. Additionally, some participants may be unable to undergo MRI scans due to non-MRI-compatible implants, claustrophobia, or other medical contraindications. Cognitive assessments may also be incomplete due to participant fatigue, lack of motivation, or cognitive decline preventing task completion. Furthermore, external factors such as human errors in data collection, budget limitations, evolving study protocols, and changes in funding sources can contribute to missing data, particularly in longitudinal studies.

In the present study, this challenge is amplified by the multimodal design: predictors span cognitive, genetic, and biochemical domains, whereas the response variable (MTL dynamic flexibility) is derived from MRI and is therefore only available for participants who completed imaging. As a result, missingness is not uniformly distributed across

variables and can be tied to domain-specific constraints (e.g., MRI contraindications or incomplete task completion), making complete-case analysis particularly inefficient and potentially biased.

Missing data mechanisms can be classified into three categories (Rubin 1976): Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). In MCAR, data are missing for reasons unrelated to both observed and unobserved variables, meaning the likelihood of missingness is entirely random. For example, in a survey, some respondents may accidentally skip the question about their age due to a formatting issue in the survey software. In MAR, the probability of a missing value depends only on observed data, allowing for statistical estimation of the missing values. For instance, individuals with higher education may be more likely to withhold their income information in a study on earnings, not because of their actual income level but due to their educational background, an observed variable. Lastly, in MNAR, missingness depends on both observed and unobserved values. For example, employees who are dissatisfied with their jobs may be less likely to disclose their salary, as their unwillingness to report income is influenced not just by observable characteristics but also by the unobserved variable job satisfaction.

Addressing missing data is essential for maintaining the reliability and validity of research findings. Two common approaches for handling missing data are case deletion and imputation. Case deletion can be listwise, where an entire case is removed if it has any missing values, or pairwise, which minimizes data loss by excluding cases only from specific analyses requiring the missing data. Alternatively, imputation involves replacing missing values with predicted estimates based on observed data, helping to retain more information in the dataset. Choosing the appropriate method depends on the nature and extent of the missing data (Jamshidian and Mata 2007; Emmanuel et al. 2021).

Importantly, the motivation for imputation in this work is not merely to “fill in” missing entries, but to preserve the information contained in partially observed records when estimating MTL dynamic flexibility from heterogeneous predictors. In multimodal AD research, complete-case deletion can drastically reduce the effective sample size and may distort the predictor distribution when missingness is related to observed covariates (e.g., health or task performance), thereby limiting generalizability of predictive models. Conversely, principled imputation can leverage cross-variable associations to retain informative cases and stabilize downstream model fitting and validation.

The high level of missingness, which is commonly encountered in medical and clinical datasets (Weiskopf and Weng 2013), highlights the importance of implementing appropriate missing data imputation strategies. First, using

only the few complete cases for analysis would severely limit the reliability and generalizability of regression models due to the drastically reduced sample size. Second, although many cases are incomplete, they still contain substantial partial information that can contribute to the predictive power and robustness of statistical models. By leveraging imputation techniques, this incomplete yet valuable data can be retained and utilized, thereby improving the overall performance and validity of analyses.

Given the growing availability of modern imputation algorithms and their differing assumptions (e.g., parametric versus nonparametric, discriminative versus generative), a comparative evaluation is needed to determine how imputation choices influence downstream prediction of MTL dynamic flexibility in this specific setting. Accordingly, we benchmark multiple established and recent imputation approaches and assess how each impacts predictive performance under a consistent cross-validation framework.

To date, various techniques have been developed to impute missing values. Multiple Imputation by Chained Equations (MICE)(Van Buuren and Oudshoorn 2011), imputation using random forests (MissForest)(Stekhoven and Buehlmann 2012), and Generative Adversarial Imputation Networks (GAIN)(Yoon et al. 2018) are among the top-tier missing value imputation methods (Zhou et al. 2024). Each of these techniques has been effectively utilized in various medical and clinical studies to address high rates of missing data. For instance, in a study analyzing data from stroke patients, MICE missing value imputation enabled the development of a regression model to assess factors influencing the time from patient arrival to computed tomography (CT) imaging, a critical quality indicator in stroke care (Chang et al. 2022). In another study MissForest and MICE are used in two clinical datasets: a cirrhosis cohort (446 patients) and an inflammatory bowel disease cohort (395 patients) (Waljee et al. 2013). MissForest demonstrated the least imputation error for both continuous and categorical variables across varying frequencies of missingness. Additionally, predictive models utilizing MissForest imputed data exhibited the smallest prediction differences, underscoring its efficacy in maintaining predictive accuracy in clinical settings. Some studies show GAIN outperforming both MissForest and MICE in imputing missing data, especially at high missingness rates (50%) and for skewed continuous variables (Dong et al. 2021).

These techniques are explained in Sect. 2.1. Section 2.2 describes the utilized regression models to estimate MTL dynamic flexibility using available predictors.

Methods

This section describes procedures for handling missing values, fitting regression models to estimate MTL dynamic flexibility, and assessing predictive performance.

Handling Missing Values

Missing data were addressed using pairwise case deletion and five imputation approaches: MICE, MissForest, GAIN, MIWAE, and ReMasker. To support fair comparison and reproducibility, imputation-related hyperparameters were selected using restricted grids or method-recommended defaults applied within training folds only i.e., no information from validation folds was used for tuning.

Case Deletion

Pairwise deletion was used, where each analysis uses all available observations for the variables involved (Little and Rubin 2002). This strategy retains partial information but reduces the effective sample size for downstream modeling when predictors or outcomes are missing. Consequently, pairwise deletion provides an informative baseline that avoids imputation assumptions while illustrating the cost of incomplete data in downstream prediction.

MICE

Multiple Imputation by Chained Equations (MICE) imputes missing values by iteratively fitting conditional models for each partially observed variable given all other variables (Van Buuren and Oudshoorn 2011). The procedure cycles through variables multiple times, updating imputations at each step; repeating this process yields multiple completed datasets that reflect imputation uncertainty. In this study, MICE was implemented using `MiceForest` (White 2021) via the `miceforest` package (v6.0.5), which preserves the MICE framework while employing gradient boosting trees as conditional learners. These boosting trees are grown sequentially and dependently, with each tree focusing on residual errors from previous trees. The number of iterations was set to a maximum of 50, and $m = 5$ completed datasets were generated and subsequently aggregated.

MissForest

MissForest imputes missing values using an iterative random-forest framework (Stekhoven and Buehlmann 2012). At each iteration, each variable with missingness is treated as an outcome and predicted from the remaining variables using a random forest trained on observed cases; predicted values replace missing entries and the cycle repeats until a maximum iteration count is reached. The random forests used by MissForest consist of independently grown trees trained on bootstrap samples with random feature selection at each split. Unlike MICE, MissForest is a

single-imputation algorithm and produces one completed dataset without explicitly modeling or propagating imputation uncertainty. In this study, MissForest was implemented using the `missForest` package (v4.2.3) with default settings for moderate-sized datasets.

GAIN

Generative Adversarial Imputation Nets (GAIN) uses an adversarial learning formulation to impute missing values in tabular data (Yoon et al. 2018). A generator network proposes imputations for missing entries conditional on observed values and a noise component; a discriminator network then attempts to identify which entries were originally observed versus generated. A hint mechanism provides the discriminator with partial information about the missingness mask, stabilizing training and encouraging the generator to produce imputations that are difficult to distinguish from true observations. Default batch size of 64, hint rate of 0.90, alpha of 60, and 1000 iterations are used as hyperparameters. In addition, prior to training, missing entries were initialized using zero imputation to provide a feasible starting point for optimization.

MIWAE

The Missing-data Importance Weighted Autoencoder (MIWAE) is a deep latent variable model that learns from incomplete data by maximizing an importance-weighted lower bound on the observed-data likelihood (Mattei and Frellsen 2019). Unlike methods that require an initial imputation to create a complete training matrix, MIWAE directly optimizes the likelihood of observed entries under a generative model, using latent variables to represent shared structure across features. After training, imputations are obtained by approximating the conditional distribution of missing entries given observed entries via Monte Carlo sampling, enabling both single and multiple imputation. In this study, MIWAE was implemented using general default hyperparameter settings recommended for moderate datasets (including latent dimensionality, learning rate, and number of importance samples) to support stable convergence and a consistent comparison across imputation approaches.

ReMasker

ReMasker extends masked autoencoding to tabular missing-value imputation by combining naturally missing entries with additional stochastic “re-masking” of observed values during training (Du et al. 2024). An autoencoder is trained to reconstruct the re-masked entries from the remaining observed context, which encourages learning representations that are robust to missingness patterns and helps prevent leakage of trivial identity mappings. After training, the learned model is applied to predict true missing entries from the observed subset. ReMasker was implemented using the default configuration recommended for small datasets, including standard layer sizes, learning rate, and masking

ratio settings, thereby prioritizing stable training and limiting overfitting risk.

Regression Models

Five regression approaches were evaluated: Ridge regression, k-Nearest Neighbors (k-NN), Support Vector Regression (SVR), tree-based ensembles, and an Artificial Neural Network (ANN). All hyperparameters were selected via grid search using an inner 3-fold cross-validation on training folds only. Standard scaling (zero mean and unit variance) was applied to Ridge, k-NN, and SVR to ensure comparability of feature magnitudes in distance- and margin-based learning. Min-max scaling to $[0, 1]$ was applied to ANN inputs to stabilize gradient-based optimization. Tree-based models used no feature scaling because split criteria depend on within-feature ordering rather than absolute magnitude. All regression models were implemented using scikit-learn v1.6 (Pedregosa et al. 2011).

Baseline Regression

A baseline model was included using `DummyRegressor(strategy="mean")`. This baseline predicts the training-set mean of the response for all validation cases and provides an interpretable reference point for quantifying the added value of covariate-informed models.

Ridge Regression

Ridge regression was used as a penalized linear model to mitigate instability from collinearity among predictors (Hastie et al. 2009). The regularization hyperparameter α (shrinkage strength) was selected by cross-validation to balance bias and variance.

k-NN

k-NN regression predicted outcomes by averaging the responses of the k nearest neighbors under Euclidean distance (Hastie et al. 2009). The neighborhood size k was selected by cross-validation, reflecting the bias-variance trade-off between local smoothing and robustness to noise.

SVR

SVR was fit using the radial basis function (RBF) kernel as the primary specification to capture potential nonlinear associations (Hastie et al. 2009). Hyperparameters C (regularization), ϵ (insensitive loss margin), and γ (RBF kernel scale) were selected via cross-validation. Linear and polynomial kernels were also examined; however, the RBF kernel yielded superior validation performance and is therefore the focus of the reported results.

Regression Trees

Tree-based models were evaluated using three ensembles: bagging, random forests, and boosting (Hastie et al. 2009). Bagging reduces variance via averaging across trees trained on bootstrap samples. Random forests further decorrelate trees by selecting a random subset of features at each

split. Boosting constructs trees sequentially, where each new tree targets residual errors from the current ensemble, improving bias at the risk of overfitting if overly aggressive. The number of trees was set to 1000; for bagging/random forests, the maximum samples parameter was set to 10 after case deletion and 50 after imputation, reflecting the effective sample sizes in those settings. For boosting trees, the learning rate was set to 0.90.

ANN

A compact feed-forward Artificial Neural Network (ANN) was used to reduce overfitting risk in a modest-sample setting (Hastie et al. 2009). The network consisted of a single hidden layer with 5 neurons and an output layer for continuous prediction. Training used the Adam optimizer with batch size 8 for 500 epochs. No dropout regularization was applied because exploratory checks did not indicate overfitting under this architecture, and additional regularization did not improve validation performance. The architecture was selected to balance expressive capacity with generalizability, consistent with common heuristics for small-sample regression (e.g., hidden width on the order of the input dimension).

Performance Assessment

Model performance was evaluated using 5-fold cross-validation. Data were randomly partitioned into five folds of approximately equal size; in each fold iteration, four folds were used for training (including all preprocessing, imputation tuning, and model selection steps) and the remaining fold was held out for validation. The full 5-fold procedure was repeated 25 times using independent random fold assignments to obtain stable estimates of predictive performance and runtime.

Predictive error was quantified using Mean Absolute Error (MAE),

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (1)$$

and Root Mean Squared Error (RMSE),

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2)$$

where n denotes the number of observations in the validation fold, y_i the observed response values, and \hat{y}_i corresponding predictions. MAE provides an interpretable measure of typical absolute error, whereas RMSE emphasizes larger deviations and is therefore more sensitive to outliers.

To complement these error-based metrics, Concordance Correlation Coefficient (CCC) is employed to quantify

overall predictive agreement. Unlike correlation-based or variance-explained measures such as R^2 , the CCC jointly evaluates precision (the correlation between observed and predicted values) and accuracy (systematic deviation from the 45° line of perfect concordance). This makes CCC particularly well suited for comparing models with diverse functional forms, including linear, kernel-based, ensemble, and neural network approaches, applied to a mixture of binary and continuous predictors. The CCC is defined as:

$$\rho_c = \frac{2\rho s_y s_{\hat{y}}}{s_y^2 + s_{\hat{y}}^2 + (\mu_y - \mu_{\hat{y}})^2}, \quad (3)$$

where ρ denotes the Pearson correlation coefficient and μ and s represent the means and variances of the observed (y) and predicted (\hat{y}) values.

To obtain stable performance estimates, the full 5-fold cross-validation procedure was repeated 25 independent times, each with a new random fold assignment. Mean RMSE, MAE, CCC, and computational runtime across these repetitions are reported for each missing-data imputation method and regression model. Consistent with the observed scale of MTL dynamic flexibility, MAE values below 0.1 were interpreted as indicating high predictive accuracy, corresponding to less than 10% absolute error relative to the observed range. Experiments were conducted on Windows 10 with Intel Core i7-10750 H CPU @ 2.60GHz, 6 cores with 32 GB RAM.

Materials

A total of 656 individuals of African ancestry (average age \pm SD: 69.81 \pm 7.38 years; 80.6% female; average education \pm SD: 13.93 \pm 2.29 years) were included in this study, recruited through the ongoing longitudinal Pathways to Healthy Aging and Brain Health Alliance study. Although the parent study is longitudinal, the present analysis included only one observation per participant. When multiple visits were available for an individual, the most recent visit was selected. All cases were de-identified and indexed by unique participant IDs, ensuring that observations used in model training and validation were independent at the participant level. This initiative investigates the relationships among genetics, cognition, health, and lifestyle factors in older individuals of African ancestry residing in the Greater Newark, NJ area. The study was conducted at Rutgers University–Newark and funded by the National Institute on Aging (NIA) under grant #1R01AG053961 (NIH/NIA) (Neha et al. 2019). Further details on community engagement and recruitment strategies are available in Gluck et al. (2025) (Gluck et al. 2025). The study received ethical approval

from the Rutgers University Institutional Review Board (IRB) and was conducted in accordance with the principles of the 1975 Helsinki Declaration.

All participants were fluent English speakers and provided written informed consent prior to participation. The initial screening was conducted over the phone. At this stage, individuals who self-reported a diagnosis of mild cognitive impairment (MCI) or dementia were excluded. Additional exclusion criteria included the use of medications typically prescribed for cognitive impairments, a history of learning disabilities, colorblindness, self-reported excessive alcohol and/or drug use, or having undergone a medical procedure requiring general anesthesia within the past three months.

Blood samples were collected from all participants to measure biomarkers, including P-tau217. Additionally, most participants provided a saliva sample, which was used to assess APOE and ABCA7 genetic variants. All participants completed either the Montreal Cognitive Assessment (MoCA) (Nasreddine et al. 2005) or the Mini-Mental State Examination (MMSE) (Folstein et al. 1975). These two tests are widely used screening tools for detecting and monitoring cognitive impairments, such as those associated with Alzheimer's disease, mild cognitive impairment (MCI), and other forms of dementia. To ensure consistency across assessments and enable direct comparisons, MMSE scores were converted to MoCA-equivalent scores using the method described by Fasnacht et al. (2023).

Participants' generalization and acquisition abilities were assessed using the Fish Task (Myers et al. 2003), which is freely available at <https://osf.io/mky3n/>. This cognitive task is designed to evaluate learning, memory, and generalization, particularly in relation to associative learning and the integrity of the MTL dynamic flexibility, including structures such as the hippocampus.

MRI acquisition

Participants were asked whether they consented to undergo an MRI session at the Rutgers University Brain Imaging Center (RUBIC). A total of 224 participants were eligible and agreed to participate in MRI scanning, representing 34.15% of the total eligible sample. Magnetic resonance imaging data were acquired at the Rutgers University Brain Imaging Center (RUBIC) on a 3T Siemens TRIO scanner using a 32-channel multiband parallel-encoding head coil (Sinha et al. 2021). A high-resolution T1-weighted 3D MP-RAGE structural scan was collected in the sagittal plane with the following parameters: TR = 1900 ms, TE = 2.52 ms, flip angle = 9°, 176 slices (no gap), voxel size = 1.0 × 1.0 × 1.2 mm, FOV = 270 × 254 × 212 mm, total acquisition time ≈ 9 min. High-resolution multiband echo-planar images (EPI) data were collected with FOV = 208 × 208 × 125 mm, TR = 664 ms, TE = 30 ms, flip angle = 30°, isotropic resolution = 1.8 mm, multiband acceleration factor =

5, and 45 axial slices covering the whole brain (Sinha et al. 2021). Multiband acquisition enabled rapid whole-brain sampling while reducing susceptibility-related distortions.

fMRI preprocessing and quality control

Neuroimaging data were preprocessed and analyzed using Analysis of Functional NeuroImages (AFNI) on Linux and Mac OSX platforms, largely following standardized `afni_proc.py` procedures (Reynolds et al. 2024). Preprocessing included de-spiking (`3dDespike`), slice-timing correction (`3dTshift`), alignment of functional data to the T1-weighted anatomical scan (`align_epi_anat.py`), motion correction (`3dvolreg`), spatial smoothing to 2 mm isotropic using a Gaussian FWHM kernel (`3dmerge`), and automated brain masking (`3dautomask`). The time series were band-pass filtered to retain the 0.06-0.12 Hz frequency band. Volumes with motion exceeding 0.3 mm were censored from the time series (scrubbing). To reduce motion- and hardware-related artifacts, nuisance regression included local white-matter and ventricular signals using ANATICOR (Jo et al. 2010), alongside six rigid-body motion parameters (pitch, roll, yaw; and x, y, z displacement) and linear scanner drift. Final voxel time courses were estimated via univariate regression (`3dDeconvolve`).

Spatial normalization

Advanced Normalization Tools (ANTs) was used to warp each participant's T1-weighted image to an in-house 0.65 mm isotropic template using diffeomorphic nonlinear registration (SyN), and the resulting transforms were applied to the denoised functional time series to align all data to the same template space (Sinha et al. 2021).

MTL network definition and dynamic connectivity

Dynamic functional connectivity was examined within an a priori medial temporal lobe (MTL) network comprising seven regions of interest (ROIs): perirhinal cortex, parahippocampal cortex, anterolateral entorhinal cortex, posteromedial entorhinal cortex, and hippocampal subfields (subiculum, CA1, and DG/CA3). ROIs were defined using a structural ROI approach based on manual delineations on the custom template following published segmentation protocols. For each ROI, the mean time series was extracted (`3dmaskave`), yielding 812 time points. Time series were subdivided into 16 non-overlapping windows of 50 time points (approximately 33 s), after discarding the first 6 and last 6 time points to reduce boundary effects. Within each window, connectivity between ROI pairs was quantified as the magnitude-squared spectral coherence, producing subject-specific $7 \times 7 \times 16$ connectivity matrices with coherence values in $[0,1]$ (Sinha et al. 2021).

MTL dynamic flexibility

Each time window was treated as a layer in a multilayer network, with inter-layer links connecting each node to itself in adjacent time windows. Community structure was

Table 1 Summary of the variables in this study (N = 656)

Variable	N valid cases (%)	N missing (%)	Range	Ave. (SD) / N (%)	Skewness
MTL dynamic flexibility	224 (34.15%)	432 (65.85%)	0.09-0.99	0.42 (0.23)	0.471
Generalization	284 (43.29%)	372 (56.71%)	0-1	0.60 (0.27)	-0.009
Acquisition	284 (43.29%)	372 (56.71%)	0.19-1	0.74 (0.14)	-0.523
MoCA	624 (95.12%)	32 (4.88%)	11-29	23.07 (3.46)	-0.548
APOE [1]	571 (87.04%)	85 (12.96%)	0,1	high:190 (33.3%)	0.712
ABCA7 50 [1]	608 (92.70%)	48 (7.30%)	0,1	high:306 (50.3%)	-0.013
ABCA7 80 [1]	608 (92.70%)	48 (7.30%)	0,1	high:80 (13.2%)	2.185
Age (yrs)	635 (96.80%)	21 (3.20%)	54-92	69.81 (7.38)	0.344
Sex [1]	649 (98.93%)	7 (1.07%)	0,1	females:529 (80.6%)	1.627
Education (yrs)	646 (98.48%)	10 (1.52%)	6-22	13.93 (2.29)	0.382
P-tau217	217 (33.08%)	439 (66.92%)	0.06-1.87	0.34 (0.27)	2.824

¹ Binary categorical variable

estimated using a Louvain-like locally greedy algorithm to optimize multilayer modularity (Mucha et al. 2010). Consistent with prior multilayer modularity applications, the intra-layer resolution and inter-layer coupling parameters were set to $\gamma = 1$ and $\omega = 1$, respectively. MTL flexibility

was quantified for each ROI as the number of changes in community assignment across consecutive windows, normalized by the total possible number of changes; the MTL dynamic flexibility measure was computed as the mean flexibility across the seven ROIs (Sinha et al. 2021).

Summary of demographic information along with summary statistics of the variables used in this study is presented in Table 1.

Results

A total of 1,866 values were missing, representing approximately 25.86% of the dataset. When cases (i.e., individual participants) with any missing values were removed, only 42 complete cases remained, indicating that 93.60% of participants had at least one missing value. Using the available predictors, several state-of-the-art regression techniques were employed to model medial temporal lobe (MTL) dynamic flexibility after handling missing values. Table 2 and Table 3 present the average validation MAE and CCC, respectively, computed over 25 independent 5-fold cross-validation runs for each combination of missing-data handling method and regression model.

In the first phase of the analysis, only the 42 complete cases (6.40% of the total sample) were used. Across the evaluated models, Ridge, Bagging Trees, and Random Forest yielded the best performance, with average MAE values of approximately 0.19, indicating that predictions deviated

Table 2 Average (SD) validation MAE from 25 independent runs for predicting MTL dynamic flexibility

Handling Missing	Case Deletion	MICE	GAIN	MissForest	MIWAE	ReMasker	Average
Baseline	0.279 (0.0028)	0.209 (0.0007)	0.209 (0.0007)	0.209 (0.0007)	0.209 (0.0007)	0.209 (0.0007)	0.221
Ridge	0.191 (0.0039)	0.190 (0.0022)	0.190 (0.0056)	0.191 (0.0021)	0.192 (0.0022)	0.192 (0.0022)	0.191
k-NN	0.198 (0.0048)	0.192 (0.0037)	0.189 (0.0060)	0.190 (0.0028)	0.195 (0.0030)	0.195 (0.0029)	0.193
SVR	0.198 (0.0055)	0.191 (0.0020)	0.191 (0.0035)	0.189 (0.0022)	0.190 (0.0021)	0.190 (0.0022)	0.191
Bagging Trees	0.192 (0.0061)	0.195 (0.0047)	0.189 (0.0103)	0.186 (0.0026)	0.200 (0.0037)	0.200 (0.0034)	0.194
Random Forest	0.192 (0.0062)	0.195 (0.0049)	0.189 (0.0102)	0.186 (0.0026)	0.201 (0.0038)	0.200 (0.0033)	0.194
Boosting Trees	0.210 (0.0098)	0.192 (0.0038)	0.188 (0.0079)	0.187 (0.0028)	0.195 (0.0021)	0.195 (0.0023)	0.194
ANN	0.299 (0.0096)	0.208 (0.0103)	0.204 (0.0113)	0.210 (0.0137)	0.210 (0.0116)	0.210 (0.0111)	0.224
Average	0.220 (0.0124)	0.197 (0.0040)	0.194 (0.0069)	0.193 (0.0037)	0.199 (0.0037)	0.199 (0.0035)	0.200

Table 3 Average (SD) validation CCC from 25 independent runs for predicting MTL dynamic flexibility

Handling Missing	Case Deletion	MICE	GAIN	MissForest	MIWAE	ReMasker	Average
Baseline	0.000 (0.0000)	0.000 (0.0000)	0.000 (0.0000)	0.000 (0.0000)	0.000 (0.0000)	0.000 (0.0000)	0.000
Ridge	0.210 (0.0197)	0.310 (0.0209)	0.350 (0.0549)	0.298 (0.0176)	0.309 (0.0185)	0.309 (0.0179)	0.298
k-NN	0.233 (0.0309)	0.378 (0.0396)	0.425 (0.0646)	0.388 (0.0212)	0.373 (0.0228)	0.369 (0.0193)	0.361
SVR	0.269 (0.0045)	0.392 (0.0181)	0.431 (0.0358)	0.415 (0.0163)	0.411 (0.0137)	0.410 (0.0151)	0.388
Bagging Trees	0.284 (0.0571)	0.437 (0.0468)	0.464 (0.1000)	0.449 (0.0220)	0.423 (0.0317)	0.425 (0.0287)	0.414
Random Forest	0.285 (0.0578)	0.437 (0.0473)	0.464 (0.0999)	0.449 (0.0227)	0.422 (0.0321)	0.425 (0.0286)	0.414
Boosting Trees	0.295 (0.0908)	0.340 (0.0343)	0.401 (0.0806)	0.392 (0.0215)	0.325 (0.0155)	0.325 (0.0194)	0.346
ANN	0.196 (0.1276)	0.294 (0.0560)	0.332 (0.0720)	0.282 (0.0610)	0.284 (0.0506)	0.274 (0.0363)	0.277
Average*	0.253	0.370	0.410	0.382	0.364	0.362	0.357

*Average is calculated excluding Baseline regression model

from the true MTL flexibility values by about 0.19 units on average. Given that the target variable ranges from 0.09 to 0.99, this level of error remains relatively large and reflects limited predictive accuracy. Consistent with this, the corresponding *CCC* values for Ridge, Bagging Trees, and Random Forest (0.210, 0.284, and 0.285, respectively) were only modestly higher than would be expected from a naïve baseline model, indicating poor concordance between observed and predicted values. By contrast, the ANN model exhibited the highest *MAE* (0.299), consistent with the well-established requirement for larger training datasets to achieve stable neural-network performance. When aggregated across all regression models, the mean (SD) *MAE* was 0.220 (0.0124), underscoring the substantial information loss associated with case deletion and highlighting the need for robust imputation to support more reliable downstream prediction.

In the second phase, the five imputation techniques described in Sect. 2.1 were applied to the full dataset of 656 cases to estimate missing values. Using the complete dataset for imputation allowed each method to exploit associations among all variables, thereby improving the accuracy of estimated values for partially observed cases. After imputation, only the 224 cases with available MTL dynamic flexibility measurements (the response variable) were retained for predictive modeling. These 224 cases were then used to fit and evaluate regression models under a 5-fold cross-validation framework, repeated 25 times with different random splits. The resulting average (SD) validation *MAE* and *CCC* values are summarized in Table 2 and Table 3.

Across all regression models, the results in Table 2 show that handling missing data through imputation substantially improved predictive accuracy compared with case deletion. The average validation *MAE* for case deletion (0.220) was consistently higher than that of all imputation strategies, with GAIN and MissForest achieving the lowest overall *MAE* values (0.194 and 0.193, respectively). As expected, the baseline regression model performed poorly under case deletion ($MAE = 0.279$) but its error decreased markedly and uniformly across all imputation methods ($MAE = 0.209$). Among traditional linear and kernel-based models, Ridge, k-NN, and SVR all showed meaningful reductions in error once imputation was applied, with *MAE* values converging around 0.189–0.192 for the best-performing configurations. These findings demonstrate that the choice of missing-data strategy has a pronounced effect on downstream predictive performance.

Tree-based ensemble models, particularly Bagging Trees and Random Forest, consistently ranked among the top performers in terms of *MAE*. When combined with MissForest or GAIN imputation, these models achieved the lowest *MAE* observed in the study (0.186–0.189), indicating strong

robustness under conditions of high missingness. Boosting Trees also performed competitively after imputation, with *MAE* values approaching those of Bagging Trees and Random Forest. In contrast, the ANN model exhibited the highest *MAE* across all methods, especially under case deletion, reflecting its sensitivity to small effective sample sizes and incomplete data. Overall, imputation not only reduced prediction error across all regression models but also allowed ensemble tree methods to emerge as the most reliable predictors of MTL dynamic flexibility (Fig. 1).

The *CCC* results in Table 3 further demonstrate that imputation markedly improved concordance between observed and predicted MTL dynamic flexibility values compared with case deletion. Under case deletion, *CCC* values were uniformly low, ranging from 0.196 (ANN) to 0.295 (Boosting Trees), reflecting poor predictive agreement when analysis was restricted to complete cases. In contrast, every imputation method substantially increased *CCC*, with model-specific gains ranging from moderate (e.g., Ridge: 0.210 \rightarrow 0.310–0.350) to pronounced (e.g., Bagging Trees: 0.284 \rightarrow 0.423–0.464). GAIN consistently produced the highest *CCC* values across nearly all regression models, yielding the strongest concordance within each model class.

Across the full set of imputation methods, tree-based ensemble models again emerged as the top performers in terms of concordance. Bagging Trees and Random Forest achieved the highest overall *CCC* values, with GAIN-imputed datasets producing a peak *CCC* of 0.464—the strongest concordance observed in the study. MissForest and MICE also provided substantial improvements, although typically to a slightly lesser degree. In contrast, ANNs showed the lowest *CCC* across all missing-data strategies, reinforcing their sensitivity to limited sample sizes. When averaged across models (excluding the baseline regression), GAIN achieved the highest overall *CCC* (0.410), followed by MissForest (0.382) and MICE (0.370), highlighting the clear advantage of advanced imputation methods in enhancing predictive agreement under high-missingness conditions. Notably, the combination of GAIN with Bagging Trees/Random Forest yielded a 54.7% improvement in predictive concordance relative to the best-performing case-deletion configuration, emphasizing the substantial benefit of leveraging all available partially observed data.

The robust Scheirer–Ray–Hare ANOVA was used to evaluate differences in prediction performance across missing-data handling methods. The analysis demonstrated a significant effect of imputation strategy on predictive performance ($p < 0.001$), as indexed by both average validation *MAE* and average validation *CCC*. All regression models substantially outperformed the baseline regression model, and all imputation methods markedly improved

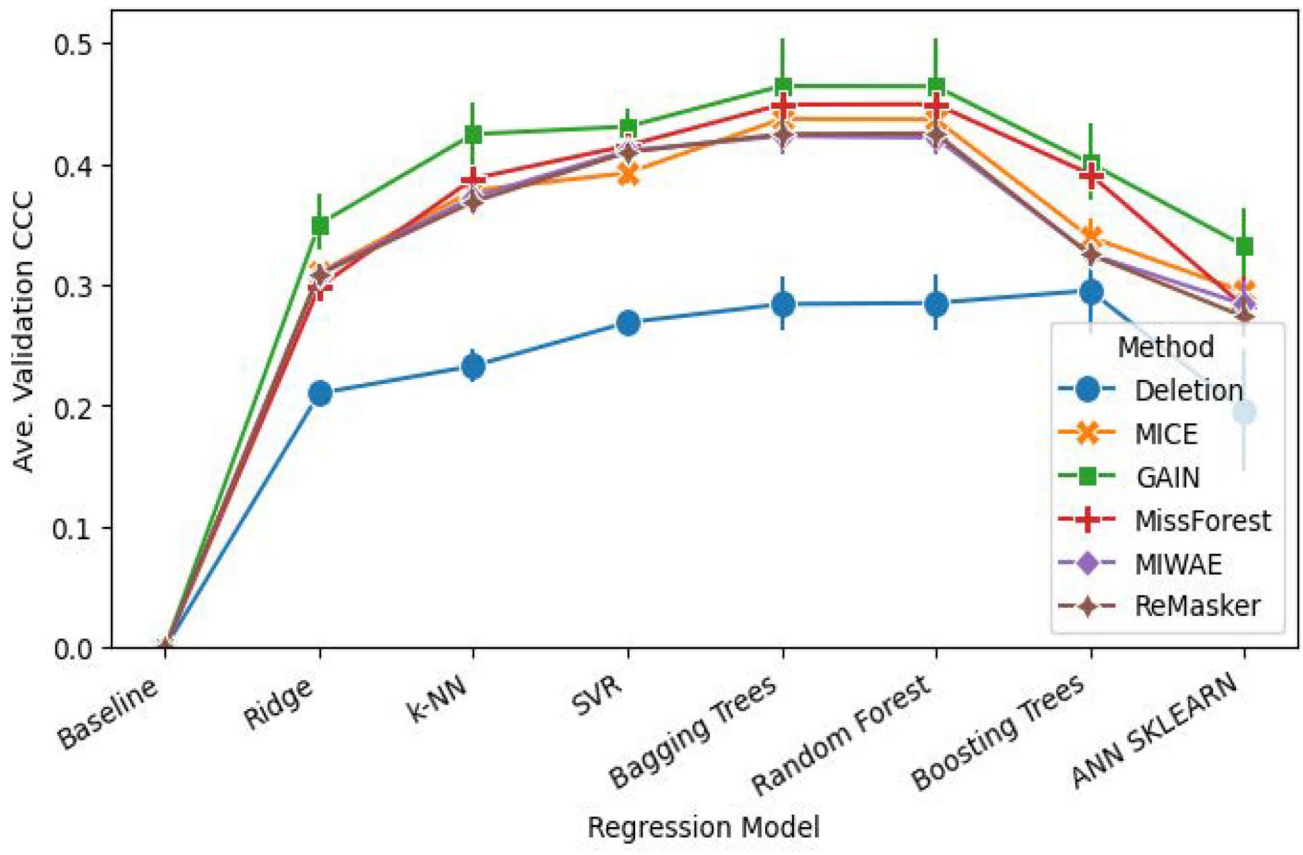


Fig. 1 Average validation CCC and 95% error bars of regression models across different missing value handling methods

Table 4 Average (SD) imputation time (seconds) from 25 independent runs for predicting MTL dynamic flexibility

Handling Missing	Case Deletion	MICE	GAIN	MissForest	MIWAE	ReMasker
Average	0.000 (0.00)	7.54 (0.99)	2.46 (0.06)	3.56 (0.17)	259.59 (1.72)	78.30 (2.81)

MTL dynamic flexibility prediction relative to case deletion. Among the imputation strategies, GAIN produced the largest improvement, increasing average CCC by 61.81%. This was followed by MissForest (50.87%), MICE (46.07%), MIWAE (43.73%), and ReMasker (43.10%). Despite these gains, the resulting CCC values remained in the poor-to-moderate range, indicating that even the best-performing models only partially capture the complexity of dynamic MTL flexibility.

Furthermore, the imputation-time results in Table 4 show substantial variability in computational cost across methods, highlighting important trade-offs between efficiency and methodological complexity. As expected, case deletion required no computation time. Among the imputation approaches, GAIN was the fastest, completing imputation of all 1,866 missing values in an average (SD) of 2.46 (0.06) seconds. MissForest also demonstrated relatively efficient performance with an average of 3.56 (0.17) seconds, whereas MICE required considerably more time at 7.54 (0.99) seconds. In contrast, deep generative models such as

ReMasker and especially MIWAE were substantially more computationally expensive, with average times of 78.30 (2.81) and 259.59 (1.72) seconds, respectively. These differences underscore that while advanced neural-network-based imputers may offer theoretical advantages, they do so at a significant computational cost compared with faster, more lightweight methods such as GAIN and MissForest.

Finally, Table 5 presents the summary statistics of the variables following missing value imputation by MissForest. The distributions of the variables remained consistent with those of the original data, preserving both the range and the central tendencies. The means and standard deviations of the imputed datasets closely matched the original values, indicating that the imputation did not introduce substantial bias or distortion. This preservation of key statistical properties is essential for maintaining the interpretability and comparability of the variables, particularly when they are used as predictors in regression models. It also suggests that the imputation technique effectively filled in missing

Table 5 Summary of the variables after missing value imputation using MissForest (N = 224)

Variable	N valid cases (%)	N imputed cases (%)	Range	Ave. (SD) / N (%)	Skewness
MTL dynamic flexibility	224 (100%)	0 (0%)	0.09-0.98	0.43 (0.23)	0.471
Generalization	224 (100%)	109 (48.66%)	0–1	0.57 (0.20)	– 0.138
Acquisition	224 (100%)	109 (48.66%)	0.42–1	0.75 (0.10)	– 0.632
MoCA	224 (100%)	6 (2.68%)	12–29	23.42 (3.48)	– 0.465
APOE [1]	224 (100%)	20 (8.93%)	0,1	high:77 (34.4%)	0.662
ABCA7 50 [1]	224 (100%)	11 (4.91%)	0,1	high:129 (57.6%)	– 0.309
ABCA7 80 [1]	224 (100%)	11 (4.91%)	0,1	high:36 (16.1%)	1.860
Age (yrs)	224 (100%)	4 (1.79%)	55–90	69.75 (7.04)	0.500
Sex [1]	224 (100%)	0 (0%)	0,1	females:178 (79.5%)	1.469
Education (yrs)	224 (100%)	0 (0%)	7–20	14.05 (2.16)	0.515
P-tau217	224 (100%)	148 (66.07%)	0.06-1.24	0.32 (0.23)	2.263

¹ Binary categorical variable

values without significantly altering the overall structure of the dataset.

Conclusion

Missing-data handling is a prerequisite for reliable modeling in multimodal clinical and neuroimaging studies, where incomplete acquisition across cognitive, genetic, biochemical, and MRI-derived variables is common. In the present dataset, 1,866 values were missing (25.86% of all entries), 93.60% of participants had at least one missing value, and only 42 complete cases (6.40%) were available for a complete-case analysis. Consistent with expectations for high-missingness settings, complete-case modeling yielded relatively large prediction errors (average $MAE = 0.220$) and low agreement (average $CCC = 0.253$), with particularly poor performance for neural network models which typically require larger datasets.

Beyond reiterating the general principle that imputation can increase data utility, the contribution of this work is to quantify how *specific* imputation choices propagate to downstream prediction of an MRI-derived network biomarker (MTL dynamic flexibility) in an underrepresented cohort of older adults of African ancestry. The evaluation emphasized (i) a consistent modeling framework across

heterogeneous regressors, (ii) repeated cross-validation to stabilize estimates, (iii) agreement-based evaluation via CCC in addition to error metrics, and (iv) computational runtime as a practical constraint for clinical research workflows. In addition, the imputation suite included both established approaches (MICE, MissForest, GAIN) and modern deep generative imputers (MIWAE and ReMasker) that have been reported as competitive in recent benchmarking studies (Mattei and Frellsen 2019; Du et al. 2024). These design choices enable a more actionable comparison than a purely conceptual argument for imputation.

Across all regression models, every imputation method reduced error relative to case deletion, with the best combinations achieving an MAE of 0.186 (MissForest with Bagging Trees/Random Forest), corresponding to approximately a 15–16% reduction in prediction error relative to the complete-case average ($MAE = 0.220$). Concordance showed even larger gains: while the best complete-case model attained a CCC of 0.295, the highest concordance was observed for GAIN combined with Bagging Trees/Random Forest ($CCC = 0.464$), representing an improvement of approximately 57% in predictive agreement. These findings align with prior evidence that nonparametric imputers and ensemble tree models can be robust in tabular biomedical settings with nonlinear effects (Waljee et al. 2013; Stekhoven and Buehlmann 2012) and with broader evaluations suggesting that conventional imputers often remain strong baselines in limited-sample tabular data (Sun et al. 2023). At the same time, absolute CCC values remained in the poor-to-moderate range, indicating that, although performance improved substantially with imputation, current models still only partially explain variability in MTL dynamic flexibility.

Computational trade-offs further differentiate methods in practice. MissForest and GAIN provided strong accuracy–efficiency profiles (average imputation times of 3.56 and 2.46 s, respectively), whereas MIWAE and ReMasker were considerably more computationally demanding (259.59 and 78.30 s on average). This pattern is consistent with the broader literature showing that deep generative imputers can be effective but may impose nontrivial compute overhead and tuning complexity relative to established approaches (Mattei and Frellsen 2019; Du et al. 2024). In downstream regression, tree-based ensembles (bagging and random forests, and to a lesser extent boosting) generally outperformed linear, kernel-based, and neural models, underscoring the importance of matching model complexity to effective sample size and noise conditions (Jerez et al. 2010).

Several limitations temper generalizability and motivate future work. First, the reported imputation–model combinations were evaluated on a single cohort; external validity should be established through replication on independent datasets with differing recruitment, imaging protocols,

Table 6 Average (SD) validation RMSE from 25 independent runs for predicting MTL dynamic flexibility

Handling Missing	Case Deletion	MICE	GAIN	MissForest	MIWAE	ReMasker	Average
Baseline	0.305 (0.0004)	0.257 (0.0001)	0.257 (0.0001)	0.257 (0.0001)	0.257 (0.0001)	0.257 (0.0001)	0.265
Ridge	0.208 (0.0006)	0.230 (0.0002)	0.234 (0.0005)	0.241 (0.0002)	0.230 (0.0003)	0.230 (0.0002)	0.229
k-NN	0.213 (0.0005)	0.234 (0.0004)	0.231 (0.0005)	0.238 (0.0003)	0.236 (0.0004)	0.236 (0.0003)	0.231
SVR	0.210 (0.0006)	0.231 (0.0002)	0.230 (0.0004)	0.237 (0.0002)	0.227 (0.0002)	0.227 (0.0002)	0.227
Bagging Trees	0.227 (0.0007)	0.240 (0.0005)	0.233 (0.0009)	0.235 (0.0003)	0.246 (0.0005)	0.246 (0.0004)	0.238
Random Forest	0.227 (0.0007)	0.240 (0.0005)	0.233 (0.0009)	0.235 (0.0003)	0.246 (0.0005)	0.246 (0.0004)	0.238
Boosting Trees	0.247 (0.0010)	0.233 (0.0004)	0.229 (0.0007)	0.232 (0.0003)	0.233 (0.0002)	0.233 (0.0003)	0.234
ANN	0.368 (0.0080)	0.257 (0.0013)	0.251 (0.0014)	0.261 (0.0020)	0.258 (0.0017)	0.258 (0.0016)	0.276
Average	0.251	0.240	0.237	0.193	0.242	0.241	0.242

and missingness mechanisms. Second, while MIWAE and ReMasker provide a meaningful step toward state-of-the-art comparison, the rapidly evolving landscape of imputation models (including additional deep generative and masked-modeling variants) warrants broader benchmarking under harmonized evaluation criteria and identical downstream tasks (Mattei and Frellsen 2019; Du et al. 2024). Third, potential coupling between algorithmically related imputers and predictors (e.g., MissForest with random-forest regression) may influence relative rankings; future evaluations could isolate imputation fidelity by masking observed entries and reporting imputation error in parallel with downstream prediction outcomes.

Overall, the results support the use of robust, nonparametric imputation strategies, particularly MissForest and GAIN, paired with ensemble tree regression as a practical default for modeling outcomes in high-missingness settings. Although the empirical focus here is on predicting MTL dynamic flexibility from multimodal neuroimaging and clinical data, the methodological insights are expected to generalize to other fields that rely on heterogeneous, partially observed tabular data, including electronic health records, epidemiology, behavioral and social sciences, and environmental studies. In such contexts, similar challenges arise from limited complete cases, nonlinear relationships, and mixed variable types, making principled imputation and model–data alignment equally critical. At the same time, absolute performance levels and optimal imputation–model pairings are likely to be dataset-dependent, underscoring the need for validation on independent cohorts and application-specific benchmarking to establish external validity beyond the present neuroimaging setting.

Appendix A RMSE Table

See Table 6

Acknowledgements The authors thank the reviewers for their thoughtful and constructive comments, which substantially improved the clarity, rigor, and presentation of this manuscript.

Author Contributions All authors contributed equally to the study and approved the final version.

Funding The study is carried out at Rutgers, the State University of New Jersey-Newark, funded by the National Institutes of Aging (NIA), under Grant No. 1R01AG053961 (NIH / NIA).

Availability of data and material The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

- Azam S, Haque ME, Balakrishnan R, Kim IS, Choi DK (2021) The ageing brain: Molecular and cellular basis of neurodegeneration. *Front Cell Develop Biol* 9:683459. <https://doi.org/10.3389/fcell.2021.683459>
- Budak M, Fausto BA, Osiecka Z, Sheikh M, Perna R, Ashton N, Blennow K, Zetterberg H, Fitzgerald-Bocarsly P, Gluck MA (2024) Elevated plasma p-tau231 is associated with reduced generalization and medial temporal lobe dynamic network flexibility among healthy older african americans. *Alzheimer's Res Ther* 16(1):253. <https://doi.org/10.1186/s13195-024-01619-0>
- Butterfield DA, Johnson LA (2018) Apoe in alzheimer's disease and neurodegeneration. *Neurobiol Dis* 139:104847. <https://doi.org/10.1016/j.nbd.2020.104847>
- Busche MA, Konnerth A (2015) Neuronal hyperactivity—a key defect in alzheimer's disease? *BioEssays: news and reviews in molecular, cellular and developmental biology* 37(6), 624–632 <https://doi.org/10.1002/bies.201500004>
- Berron D, Van Westen D, Ossenkoppelle R, Strandberg O, Hansson O (2020) Medial temporal lobe connectivity and its associations with cognition in early alzheimer's disease. *Brain* 143(4):1233–1248. <https://doi.org/10.1093/brain/awaa068>
- Bassett DS, Wymbs NF, Porter MA, P.J., M., J.M., C., S.T., G. (2011) Dynamic reconfiguration of human brain networks during learning. *Proc Natl Acad Sci* 108(18):7641–7646. <https://doi.org/10.1073/pnas.1018985108>
- Chen-Chen T, Jin-Tai Y, Lan T (2014) Biomarkers for preclinical Alzheimer's disease. *J Alzheimer's Dis* 42(4):1051–1069. <https://doi.org/10.3233/JAD-140843>
- Chang C, Deng Y, Jiang X, Long Q (2022) Multiple imputation for analysis of incomplete data in distributed health data networks. *Nat Commun* 11(1):5467. <https://doi.org/10.1038/s41467-020-19270-2>

- Coley N, Gardette V, Cantet C, Gillette-Guyonnet S, Nourhashemi F, Vellas B, Andrieu S (2011) How should we deal with missing data in clinical trials involving alzheimers disease patients? *Curr Alzheimer Res* 8(4):421–433. <https://doi.org/10.2174/156720511795745339>
- DeTure MA, Dickson DW (2019) The neuropathological diagnosis of alzheimer's disease. *Mol Neurodegener* 14(1):32. <https://doi.org/10.1186/s13024-019-0333-5>
- Dong W, Fong D, Yoon J, Wan E, Bedford LE, Tang E, Lam C (2021) Generative adversarial networks for imputing missing data for big data clinical research. *BMC Med Res Methodol* 21(1) <https://doi.org/10.1186/s12874-021-01272-3>
- Du T, Melis L, Wang T (2024) Remasker: Imputing tabular data with masked autoencoding. In: The Twelfth International Conference on Learning Representations. ICLR. International Conference on Learning Representations, <https://openreview.net/forum?id=K19NqjLVDT>
- Das SR, Pluta J, Mancuso L, Kliot D, Orozco S, Dickerson BC, Yushkevich PA, Wolk DA (2013) Increased functional connectivity within medial temporal lobe in mild cognitive impairment. *Hippocampus* 23(1):1–6. <https://doi.org/10.1002/hipo.22051>
- Braak H, Braak E. (1991) Neuropathological staging of alzheimer-related changes. *Acta Neuropathol* 82(4):239–259. <https://doi.org/10.1007/BF00308809>
- Emmanuel T, Maupong T, Mpoeleng D, Semong T, Mphago B, Tabona O (2021) A survey on missing data in machine learning. *J Big Data* 8(1):140. <https://doi.org/10.1186/s40537-021-00516-9>
- Frank MJ, Claus ED (2006) Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev* 113(2):300–326. <https://doi.org/10.1037/0033-295X.113.2.300>
- 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(3):189–198. <https://doi.org/10.1037/t07757-000>
- Fasnacht JS, Wuest AS, Berres M, Thomann AE, Krumm S, Gutbrod K, Steiner LA, Goettel N, Monsch AU (2023) Conversion between the montreal cognitive assessment and the mini-mental status examination. *J Am Geriatr Soc* 71(3):869–879. <https://doi.org/10.1111/jgs.18124>
- Gluck MA, Fausto BA, Thornton-Hammonds, D., Wright, G., Wilson, G.B. (2025) There is no quick, easy, or cheap way to recruit older african americansto aging and brain health research: Ten evidence-based strategies for investing in community engagement. *Alzheimer's & Dementia* 20:091673. <https://doi.org/10.1002/alz.091673>
- Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo JL, Montine T, Phelps C, Rankin KP, Rowe CC, Scheltens P, Siemers E, Snyder HM, Sperling R (2018) NIA-AA research framework: Toward a biological definition of alzheimer's disease. *Alzheimer's & dementia* 14(4):535–562. <https://doi.org/10.1016/j.jalz.2018.02.018>
- Jamshidian M, Mata M (2007) Advances in analysis of mean and covariance structure when data are incomplete. In: *Handbook of Latent Variable and Related Models*, pp. 21–44. North-Holland, Amsterdam <https://doi.org/10.1016/B978-044452044-9/50005-7>
- Jerez JM, Molina I, Garcia-Laencina PJ, Alba E, Ribelles N, Martin M, Franco L (2010) Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artif Intell Med* 50(2):105–115. <https://doi.org/10.1016/j.artmed.2010.05.002>
- Jo H, Saad Z, Simmons W, Milbury L, Cox R (2010) Mapping sources of correlation in resting state fmri, with artifact detection and removal. *Neuroimage* 52(2):571–582. <https://doi.org/10.1016/j.neuroimage.2010.04.246>
- Janelidze S, Stomrud E, Smith R, Palmqvist S, Mattsson N, Airey DC, Proctor NK, Chai X, Shcherbinin S, Sims JR, Triana-Baltzer G, Theunis C, Slemmon R, Mercken M, Kolb H, Dage JL, Hansson O (2020) Cerebrospinal fluid p-tau217 performs better than p-tau181 as a biomarker of alzheimer's disease. *Nat Commun* 11(1):1683. <https://doi.org/10.1038/s41467-020-15436-0>
- Little RA, Rubin DB (2002) *Statistical Analysis with Missing Data*. Wiley, Hoboken, NJ <https://doi.org/10.1002/9781119013563>
- Mattei PI, Frellsen J (2019) Miwae: Deep generative modelling and imputation of incomplete data sets. In: Chaudhuri, K., Salakhutdinov, R. (eds.) *Proceedings of the 36th International Conference on Machine Learning*. PMLR, vol. 97, pp. 4413–4423. *Proceedings of Machine Learning Research*, Long Beach, CA, USA. <https://proceedings.mlr.press/v97/mattei19a.html>
- Mielke MM, Hagen CE, Xu J, Chai X, Vemuri P, Lowe VJ, Airey DC, Knopman DS, Roberts RO, Machulda MM, Jack CR Jr, Petersen RC, Dage JL (2018) Plasma phospho-tau181 increases with alzheimer's disease clinical severity and is associated with tau- and amyloid-positron emission tomography. *Alzheimer's Dementia* 14(8):989–997. <https://doi.org/10.1016/j.jalz.2018.02.013>
- Maia LF, Kaeser SA, Reichwald J, Hruscha M, Martus P, Staufenbiel M, Jucker M (2013) Changes in amyloid- β and tau in the cerebrospinal fluid of transgenic mice overexpressing amyloid precursor protein. *Sci Translat Med*. <https://doi.org/10.1126/scitranslmed.3006446>
- Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328:876–878. <https://doi.org/10.1126/science.1184819>
- Moallemian S, Salmon E, Bahri MA, Belyi N, Delhaye E, Baletau E, Degueldre C, Phillips C, Bastin C (2023) Multimodal imaging of microstructural cerebral alterations and loss of synaptic density in alzheimer's disease. *Neurobiol Aging* 132:24–35. <https://doi.org/10.1016/j.neurobiolaging.2023.08.001>
- Myers CE, Shohamy D, Gluck MA, Grossman S, Kluger A, Ferris S, Golomb J, Schnirman G, Schwartz R (2003) Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *J Cogn Neurosci* 15(2):185–193. <https://doi.org/10.1162/089892903321208123>
- Mecca AP, Barcelos NM, Wang S, Brück A, Nabulsi N, Planeta-Wilson B, Nadelmann J, Benincasa AL, Ropchan J, Huang Y, Gelernter J (2018) Cortical β -amyloid burden, gray matter, and memory in adults at varying apoe ϵ 4 risk for alzheimer's disease. *Neurobiol Aging* 61:207–214. <https://doi.org/10.1016/j.neurobiolaging.2017.09.027>
- Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H (2005) The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 53(4):695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- Neha N, Reagh ZM, Tustison NJ, Berg CN, Shaw A, Myers CE, Hill D, Yassa MA, Gluck MA (2019) Abca7 risk variant in healthy older african americans is associated with a functionally isolated entorhinal cortex mediating deficient generalization of prior discrimination training. *Hippocampus* 29(6):527–538. <https://doi.org/10.1002/hipo.23042>
- Petok JR, Myers CE, Pa J, Hobel Z, Wharton DM, Medina LD, Casado M, Coppola G, Gluck MA, Ringman JM (2018) Impairment of memory generalization in preclinical autosomal dominant alzheimer's disease mutation carriers. *Neurobiol Aging* 65:149–157. <https://doi.org/10.1016/j.neurobiolaging.2018.01.022>

- Pasquini L, Scherr M, Tahmasian M, Meng C, Myers NE, Ortner M, Muhlau M, Kurz A, Forstl H, Zimmer C, Grimmer T, Wohlschlagel AM, Riedl V, Sorg C (2015) Link between hippocampus' raised local and eased global intrinsic connectivity in ad. Alzheimer's Dementia 11(5):475–484. <https://doi.org/10.1016/j.alz.2014.02.007>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12:2825–2830
- Reynolds RC, Glen DR, Chen G, Saad ZS, Cox RW, Taylor PA (2024) Processing, evaluating, and understanding fmri data with `afni_proc.py`. *Imaging Neuroscience* 2:2–00347. https://doi.org/10.1162/imag_a_00347
- Reitz C, Jun G, Naj A, Rajbhandary R, Vardarajan BN, Wang LS, Valadares O, Lin CF, Larson EB, Graff-Radford NR, Evans D, De Jager PL, Crane PK, Buxbaum JD, Murrell JR, Raj T, Ertekin-Taner N, Logue M, Baldwin CT, Green R, et al. (2013) Variants in the atp-binding cassette transporter (*abca7*), apolipoprotein *e* $\epsilon 4$, and the risk of late-onset alzheimer disease in african americans. *JAMA* 309(14):1483–1492. <https://doi.org/10.1001/jama.2013.2973>
- Rubin DB (1976) Inference and missing data. *Biometrika* 63(3):581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Stekhoven DJ, Buehlmann P (2012) Missforest - non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28(1):112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Sizemore AE, Bassett DS (2018) Dynamic graph metrics: Tutorial, toolbox, and tale. *Neuroimage* 180:417–427. <https://doi.org/10.1016/j.neuroimage.2017.06.081>
- Sinha N, Berg CN, Yassa MA, Gluck MA (2021) Increased dynamic flexibility in the medial temporal lobe network following an exercise intervention mediates generalization of prior learning. *Neurobiol Learn Mem* 177:107340. <https://doi.org/10.1016/j.nlm.2020.107340>
- Selkoe DJ, Hardy J (2016) The amyloid hypothesis of alzheimer's disease at 25 years. *EMBO Mol Med* 8(6):595–608. <https://doi.org/10.15252/emmm.201606210>
- Sakae N, Liu CC, Shinohara M, Frisch-Daiello J, Ma L, Yamazaki Y, Tachibana LM, amd Younkin, Kurti, A., Carrasquillo, M.M., Zou, F., Sevlever, D., Bisceglia, G., Gan, M., Fol, R., Knight, P., Wang, M., Han, X., Fryer, J.D., Fitzgerald, M.L., Ohyagi, Y., Younkin, S.G., Bu, G., Kanekiyo, T. (2016) *Abca7* deficiency accelerates amyloid- β generation and alzheimer's neuronal pathology. *J Neurosci* 36(13):3848–3859. <https://doi.org/10.1523/JNEUROSCI.3757-15.2016>
- Sun Y, Li J, Xu Y, Zhang T, Wang X (2023) Deep learning versus conventional methods for missing data imputation: A review and comparative study. *Expert Syst Appl* 227:120201. <https://doi.org/10.1016/j.eswa.2023.120201>
- Salami A, Pudas S, Nyberg L (2014) Elevated hippocampal resting-state connectivity underlies deficient neurocognitive function in aging. *Proceed Natl Acad Sci United States* 111(49):17654–17659. <https://doi.org/10.1073/pnas.1410233111>
- Uddin LQ (2021) Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nat Rev Neurosci* 22:167–179. <https://doi.org/10.1038/s41583-021-00428-w>
- Van Buuren S, Oudshoorn KG (2011) MICE: Multivariate imputation by chained equations in r. *J Stat Softw* 45(3):1–67. <https://doi.org/10.18637/jss.v045.i03>
- White C (2021) Miceforest: Fast, Memory-Efficient Multiple Imputation with LightGBM. <https://github.com/AnotherSamWilson/miceforest>. Accessed: 2025-05-07
- Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, Marrero J, Zhu J, Higgins PD (2013) Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open* 3(8):002847. <https://doi.org/10.1136/bmjopen-2013-002847>
- Weiskopf NG, Weng C (2013) Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 20(1):144–151. <https://doi.org/10.1136/amiajnl-2011-000681>
- Yoon J, Jordan J, Van Der Schaar M (2018) GAIN: Missing data imputation using generative adversarial nets. In: Proceedings of the 35th International Conference on Machine Learning, vol. 80, pp. 5689–5698. PMLR, <https://doi.org/10.48550/arXiv.1806.02920>
- Zhou Y, Aryal S, Bouadjenek M (2024) Review for handling missing data with special missing mechanism. arXiv <https://doi.org/10.48550/arXiv.2404.04905>
- Zeithamova D, Dominick AL, Preston AR (2012) Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron* 75(1):168–179. <https://doi.org/10.1016/j.neuron.2012.05.010>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.