

RESEARCH ARTICLE

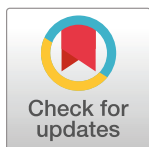
# Machine learning based multi-modal prediction of future decline toward Alzheimer's disease: An empirical study

Batuhan K. Karaman<sup>1,2</sup>, Elizabeth C. Mormino<sup>3</sup>, Mert R. Sabuncu<sup>1,2\*</sup>, for the Alzheimer's Disease Neuroimaging Initiative<sup>¶</sup>

**1** School of Electrical and Computer Engineering, Cornell University and Cornell Tech, New York, NY, United States of America, **2** Department of Radiology, Weill Cornell Medicine, New York, NY, United States of America, **3** Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA, United States of America

<sup>¶</sup> Membership of the Alzheimer's Disease Neuroimaging Initiative is provided in the Acknowledgments.

\* [msabuncu@cornell.edu](mailto:msabuncu@cornell.edu)



## OPEN ACCESS

**Citation:** Karaman BK, Mormino EC, Sabuncu MR, for the Alzheimer's Disease Neuroimaging Initiative (2022) Machine learning based multi-modal prediction of future decline toward Alzheimer's disease: An empirical study. PLoS ONE 17(11): e0277322. <https://doi.org/10.1371/journal.pone.0277322>

**Editor:** Kim Han Thung, University of North Carolina at Chapel Hill, UNITED STATES

**Received:** July 4, 2022

**Accepted:** October 24, 2022

**Published:** November 16, 2022

**Copyright:** © 2022 Karaman et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data used in preparation of this article were obtained from the Alzheimer's Disease neuroimaging Initiative (ADNI) database <http://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](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf) Others can use the

## Abstract

Alzheimer's disease (AD) is a neurodegenerative condition that progresses over decades. Early detection of individuals at high risk of future progression toward AD is likely to be of critical significance for the successful treatment and/or prevention of this devastating disease. In this paper, we present an empirical study to characterize how predictable an individual subjects' future AD trajectory is, several years in advance, based on rich multi-modal data, and using modern deep learning methods. Crucially, the machine learning strategy we propose can handle different future time horizons and can be trained with heterogeneous data that exhibit missingness and non-uniform follow-up visit times. Our experiments demonstrate that our strategy yields predictions that are more accurate than a model trained on a single time horizon (e.g. 3 years), which is common practice in prior literature. We also provide a comparison between linear and nonlinear models, verifying the well-established insight that the latter can offer a boost in performance. Our results also confirm that predicting future decline for cognitively normal (CN) individuals is more challenging than for individuals with mild cognitive impairment (MCI). Intriguingly, however, we discover that prediction accuracy decreases with increasing time horizon for CN subjects, but the trend is in the opposite direction for MCI subjects. Additionally, we quantify the contribution of different data types in prediction, which yields novel insights into the utility of different biomarkers. We find that molecular biomarkers are not as helpful for CN individuals as they are for MCI individuals, whereas magnetic resonance imaging biomarkers (hippocampus volume, specifically) offer a significant boost in prediction accuracy for CN individuals. Finally, we show how our model's prediction reveals the evolution of individual-level progression risk over a five-year time horizon. Our code is available at <https://github.com/batuhankmkaraman/mlbasedad>.

ADNI website to access and download the data by submitting a data access request to ADNI. The authors had no special privileges and any researcher should be able to obtain the data the same way as them.

**Funding:** This study was funded by the NIH National Library of Medicine (grant number R01LM012719), the NIH National Institute on Aging (grant number R01AG053949), the National Science Foundation NeuroNex (grant number 1707312), and the National Science Foundation CAREER (grant number 1748377), all awarded to MRS.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Alzheimer's Disease (AD) is the most common type of dementia among the elder population, accounting for nearly 70% of all dementia patients [1] and ranking as the seventh leading cause of death in the United States [2]. Many mechanisms in the development of AD have been uncovered by decades of experimental and clinical studies [3–5], but the jigsaw is still unsolved. In the realm of AD, public and private databases [6–10] serve as important resources for the application of machine learning algorithms that help characterize disease heterogeneity [11], guide therapy [12], and develop and evaluate potential treatments [13, 14]. An area where machine learning can play a crucial role is the prediction of future clinical decline at the individual level, which can inform clinical and other life decisions.

Clinically, in the context of Alzheimer's disease, individuals are classically grouped into one of the following three stages: cognitively normal (CN), mild cognitive impairment (MCI), and AD dementia. MCI is considered a high-risk, transitional stage between healthy aging and dementia. Future clinical decline toward dementia is considered to be more predictable in those with MCI than in CN individuals [15, 16]. As our results further corroborate, CN-to-MCI conversion is inherently a more difficult prediction problem than MCI-to-AD conversion. The vast majority of the published studies dealing with individual-level future decline predictions with machine learning focus on early detection of MCI-to-AD conversion [17].

There is growing literature showing factors, including certain biomarkers, that predict progression from CN to MCI [18, 19]. This work has converged to suggest that the risk associated with specific factors is relatively small, and requires a long follow-up to observe the effect. Furthermore, papers tackling the CN-to-MCI conversion prediction problem with machine learning have been relatively limited [20, 21]. Thus, there is a need to understand what combination of variables can yield accurate individual-level predictions; and whether the significance of the variables changes as a function of disease stage (e.g., CN or MCI at baseline).

Many prior studies have primarily focused on building models that predict future conversion within a specific time horizon, e.g., three years [22, 23]. Although some of these studies test their modeling strategy for variable follow-up years, they do this by training new models for each time horizon. Studies that utilize survival (event-time) models can, in theory, make predictions for any future time-point [24], yet, they need to make strong assumptions about the evolution of the underlying hazard function, which might potentially limit performance.

In this work, we rely on the neural network (deep learning) framework, which gives us the flexibility to explore the effect of various modeling choices, namely nonlinearity and predicting arbitrary time horizons, while holding other design parameters constant. We implement three different classifiers, two of which are trained to predict the clinical status (CN, MCI, or AD) at a single time point in the future. The first of these two models is linear and called the “Linear Single-year Model (LSM)”. The other one is nonlinear and referred to as the “Nonlinear Single-year Model (NSM)”. In the third model, we employ a nonlinear architecture and modify it to make it capable of predicting the clinical status at any time point in the future. We refer to this as the “Nonlinear Multi-year Model (NMM).” Comparing these models allow us to deduce the predictive importance of nonlinear models and accounting for different time horizons. For instance, our results verify that a linear model (LSM) can yield high-quality predictions for the relatively easy MCI-to-AD conversion prediction problem, whereas higher capacity nonlinear models are needed for making more accurate predictions about the more challenging task of predicting CN-to-MCI conversion.

Our analyses further offer some novel insights. In CN individuals, predicting who will convert to MCI within a year is easier than predicting for a 5-year time window. For individuals with MCI, however, the situation is reversed. It is harder to predict who will progress to AD in

the shorter term. We train our models to handle arbitrary missingness patterns in the input data, which in turn, allows us to interrogate the predictive value of different data types. For example, our results suggest that the molecular biomarkers we consider in our study are very helpful in the MCI stage, but not as much in CN individuals. On the other hand, structural magnetic resonance imaging (MRI) biomarkers (hippocampus volume, specifically) offer a significant performance boost for CN-to-MCI conversion but not MCI-to-AD conversion. Finally, we present our model's prediction of future conversion risk as a continuous function of time, which reveals different trajectories.

## Materials and methods

In this section; we first present the dataset, which was derived from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [25]. We then discuss important implementation details that were critical in handling missingness, and unbalanced classes at different time points.

### Dataset

All participants used in this work are from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. ADNI aims to evaluate the structure and function of the brain across different disease states and uses clinical measures and biomarkers to monitor disease progression. We select the participants who did not have clinical AD at the baseline (screening) visit and had at least one follow-up diagnosis within the next five years ( $n = 1404$ ). We excluded the CN baseline participants who converted to AD in the next five years because they are very few ( $n = 6$ ), the CN baseline participants who converted to a later stage but reverted back to an earlier stage ( $n = 23$ ), and the MCI baseline participants who either were diagnosed as CN in a later follow-up year ( $n = 87$ ) or converted to AD and reverted back ( $n = 18$ ) since these subjects might have been diagnosed incorrectly at some point. Including these individuals did not alter our main conclusions. After the exclusions, we are left with 1404 participants. We note that all individuals were either CN or MCI at baseline. Table 1 lists summary statistics for the participants; including sex, age, number of years of education completed, count of Apolipoprotein E4 (APOE4) allele, Clinical Dementia Rating (CDR), and Mini Mental State Examination (MMSE) scores at baseline.

A critical aspect of the data, as is common in many real-world longitudinal studies, is that there are missing follow-up visits, with imperfect timings, and many subjects drop out of the study before the planned completion. Table 2 shows the number of available subjects in each diagnostic group for annual follow-up visits. We note that Table 2 can be used to infer the number of subjects progressing from one stage to another during follow-up years of interest.

**Table 1. Summary statistics of the participants at baseline.**

	CN baseline ( $n = 615$ )	MCI baseline ( $n = 789$ )
Female/Male	335/280	324/465
Age (yr)	$73.19 \pm 6.18$	$73.46 \pm 7.39$
Education (yr)	$16.51 \pm 2.57$	$15.93 \pm 2.81$
APOE4 (0/1/2)	430/169/14	371/313/98
CDR	$0.04 \pm 0.13$	$1.55 \pm 0.89$
MMSE	$29.11 \pm 1.11$	$27.52 \pm 1.82$

Mean  $\pm$  standard deviations are listed. APOE4 row represents the number of alleles.

<https://doi.org/10.1371/journal.pone.0277322.t001>

**Table 2. The number of available subjects in each diagnostic group for annual follow-up visits.**

Follow-up year	CN baseline		MCI baseline	
	CN	MCI	MCI	AD
1	427	14	674	110
2	527	32	431	218
3	181	41	317	261
4	230	49	202	286
5	123	54	127	292

The follow-up diagnoses are not actually exactly 12 months apart. They have been rounded to the nearest time horizon in years.

<https://doi.org/10.1371/journal.pone.0277322.t002>

In Table 2, and in all subsequent analyses, any subject who progressed (from CN to MCI, or from MCI to AD) before dropping out of the study was considered to remain MCI or AD until year 5. Non-converter subjects without a visit in a certain follow-up year are not used for either training or testing in that follow-up year. Some subject groups have different follow-up visit schedules. For example, in ADNI-2 and ADNI-3, CN baseline participants are only clinically evaluated every other follow-up year. Therefore years 1 and 3 have fewer CN diagnoses than years 2 and 4, respectively. ADNI includes a limited number of subjects who have been monitored for more than five years. However, due to the very limited number of visits beyond 5 years, we excluded these timepoints from our analyses.

## Input features

We use the clinical data and biomarkers collected at baseline as our input features. Clinical data includes subject demographics (age, gender, number of years of education completed, ethnicity, race, marital status), genotype (number of APOE4 alleles), clinical assessments (clinical dementia rating, or CDR; Activities of Daily Living, or FAQ; Everyday Cognition, or ECog), cognitive assessments (Mini-Mental State Exam, or MMSE; Alzheimer's Disease Assessment Scale, or ADAS-Cog; Montreal Cognitive Assessment, or MoCA; Rey Auditory Verbal Learning Test Trials 1–6; Logical Memory Delayed Recall; Trail Making Test Part B; Digit Symbol Substitution, Digit and Trails B versions of Preclinical Alzheimer's Cognitive Composite score [26, 27]), and the baseline diagnosis (CN or MCI). The biomarkers are Cerebrospinal Fluid (CSF) measurements [28, 29] (Amyloid-Beta 1–42; Total Tau, or T-Tau-; Phosphorylated Tau, or P-Tau), Magnetic Resonance Imaging (MRI) volume measurements [30, 31] (Ventricles; Hippocampus; WholeBrain; Entorhinal; Fusiform; MidTemp; Intracranial Volume, or ICV; all computed using the FreeSurfer software [32–51]), and Positron Emission Tomography (PET) standardized uptake value ratio (SUVR) scores (for following tracers: Fluorine-18-Fluorodeoxyglucose, or FDG; Florbetapir, or AV45; Pittsburgh Compound, or PIB) [52, 53]. We note that CSF, FDG, and PIB biomarkers are referred to as molecular biomarkers. Furthermore, we employ single, global PET SUVR measurements instead of regional values.

The regional volume measurements derived from the MRI scans were computed, quality controlled and publicly made available by UCSF researchers. In this pipeline, the images are processed to implement the following steps: Talairach transform computation, intensity normalization, skull stripping, creation of the white-matter and pial surfaces, segmentation of the gray and white matter volumetric regions of interest and creation of the cortical parcellation as

described in [54]. Note that we input ICV as a separate feature, instead of normalizing other volumetric MRI measurements with it.

The ADNI study consists multiple phases (1, GO, 2, and 3). Each phase implemented a slightly different data acquisition protocol. Additionally, as mentioned above, follow-up data collection was also heterogeneous, with missing visits and non-uniform visit intervals. The degree of missingness for the different baseline data modalities is shown in Table 3 for the 1404 participants we use. Rather than dropping out the subjects with incomplete baseline data, we substitute placeholder values for their missing features and keep them in our dataset. Our substitution procedure consists of two parts. First, we record the binary missingness mask for the feature set. Each participant has their own binary missingness mask indicating what variable is observed or not for that particular individual. Then, following prior work [55], we perform mode substitution for missing categorical variables (sex, ethnicity, race, marital status, APOE4), and mean substitution for missing numerical variables. It is true that when the class labels are unbalanced, these substituted values will be biased by the values in the majority class. However, we would like to emphasize that we consider these substituted values as dummy placeholders. In other words, we use mode/mean values solely to make sure that the substituted values are within an appropriate range and thus the numerical optimization is not compromised. By concatenating the missingness mask to the feature vector, we expect the model to learn to treat these placeholders appropriately. To prevent any information leakage, we substitute the mode and mean values computed in the non-missing portion of the training set as placeholders for the missing values in training, validation, and test sets. We compute a single mode/mean value for each feature, weighing all non-missing values in the training set equally.

The categorical variables except the baseline diagnosis are one-hot encoded (i.e., are represented with dummy variables encoding presence) and numerical variables are z-score normalized in the last step of feature processing. We note that the z-score normalization is first performed on training data and mean and variance values are saved to be used for validation and test data later on, which is similar to what we do in the second step of the imputation procedure.

We have six discrete (categorical) variables, which are sex (encoded as a one-hot vector for either female or male), ethnicity (encoded as a one-hot vector for either Hispanic/Latino or not Hispanic/Latino), race (encoded as a one-hot vector for one of the following: Asian, Black, Hawaiian/Other PI, Indian/Alaskan, More than one, White), marital status (encoded as a one-hot vector for one of the following: divorced, married, never married, widowed), number of

**Table 3. The degree of missingness (%) in different baseline data modalities for two patient groups.**

Data type	CN baseline	MCI baseline
Clinical assessments	27.87	37.51
Cognitive assessments	7.38	7.34
CSF	48.29	37.14
MRI	33.36	21.22
FDG	45.37	24.08
AV45	46.99	50.95
PIB	100	98.48

There is significant missingness in the biomarker data. Crucially, there is no CN baseline patient with an AV45 scan at baseline. Moreover, a substantially limited amount of MCI subjects have AV45 scans at baseline.

<https://doi.org/10.1371/journal.pone.0277322.t003>

APOE4 alleles (encoded as 0, 1 or 2 copies of the E4 allele), and baseline diagnosis (a scalar that is 0 for CN and 1 for MCI). Real-valued variables are the number of years of education completed, clinical test scores, cognitive assessments, and biomarker values. All real-valued (numerical) features are scalars except for the clinical assessment of ECog, and the cognitive assessments of ADAS-Cog, and Rey Auditory Verbal Learning Test Trials 1–6, which are vectors with dimensionalities 14, 3, and 4, respectively. In total, we have 45 real-valued features. We note that we also compute a binary missingness mask for the input feature set. In the mask, there is no entry for the baseline diagnosis, since that variable has no missingness. Therefore, the binary missingness mask has a dimensionality of 50. Concatenating the categorical features, numerical features, and binary missingness mask yields a feature vector of length 113.

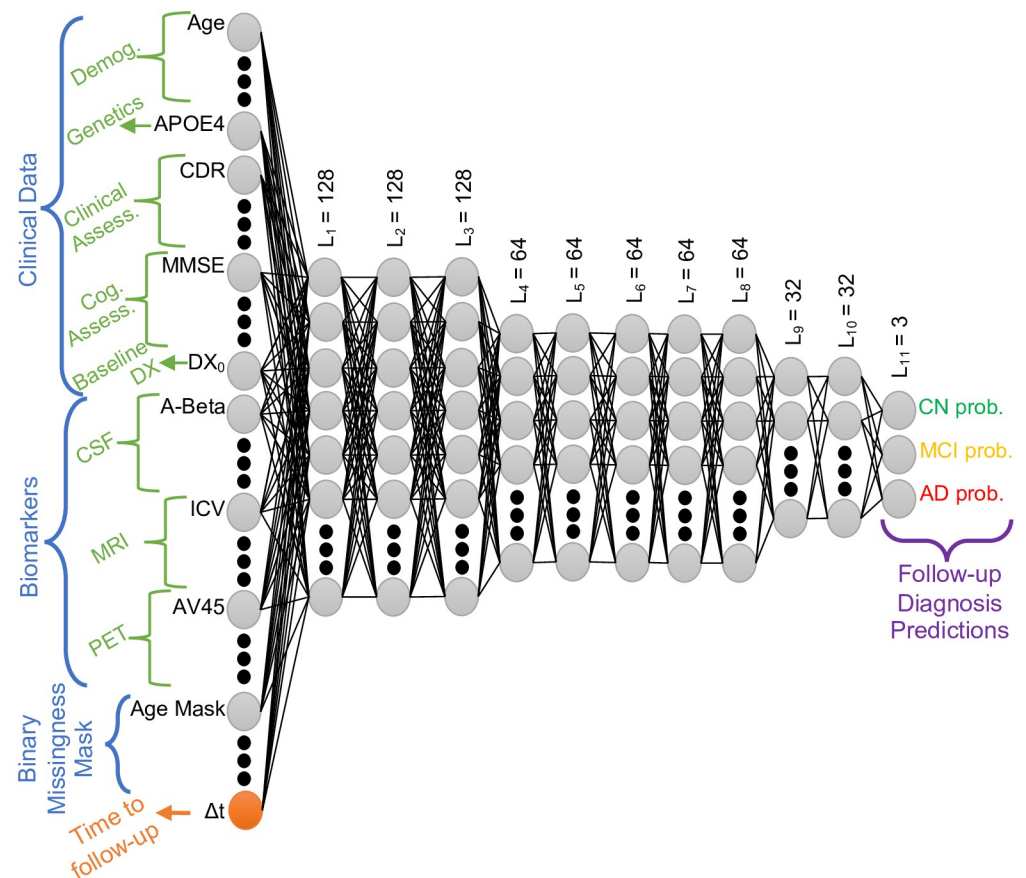
## Models

We are interested in the prediction of an individual's future diagnostic status (CN, MCI, or AD) based on the input features at a baseline visit. A large number of studies have looked at this question as a classification problem, often at a single follow-up time, e.g., three years after the baseline. However, this formulation has two drawbacks. First, many subjects might drop out of the study before that follow-up time, which means these subjects cannot be used for training. Secondly, this approach groups together subjects who convert after the intended follow-up time, with those who are stable through the study. To distinguish these two subject groups, one would need to train a new model from scratch corresponding to a different follow-up time. An alternative approach that addresses these issues involves survival modeling [56]. However, these methods require strong assumptions about the underlying hazard function that can constrain the model's performance.

All our models follow the neural network architecture template depicted in Fig 1, which was designed based on optimizing empirical performance on validation data in a single split. The output is a length-3 probability vector, computed by a soft-max layer, that corresponds to the individual's CN, MCI, and AD probabilities at the future time point. In the single-year models, the prediction is for a fixed follow-up time and thus time-to-follow-up is not an input feature. Therefore, the single-year models have an input layer width of 113, and the input layer width of the multi-year model is 114. We train a separate single-year model for each time horizon. The linear single-year model, LSM, is made up of linear (fully connected) layers, whereas its nonlinear counterpart, NSM, contains nonlinear activation functions, which are element-wise rectified linear units, or ReLUs, between linear layers. The most flexible model, the nonlinear multi-year model, or NMM, accepts the time-to-follow-up (in months) as an input feature and computes the output corresponding to that input value. Thus, we train a single NMM for all follow-up times. The NMM can be viewed as a family of models, parameterized by the follow-up time. We note that all three models have roughly the same number of learnable parameters.

We also experimented with two alternate models. The alternative single-year linear model was a standard linear regression model, implemented as a single fully connected layer neural network, with L2 penalty (weight decay) on the weights (coefficients). This is equivalent to a ridge regression approach. As the results presented in the Supplementary Material demonstrate this model performs no better than the LSM model described above. The second alternative is a slight modification of the NMM model, where the input time-to-follow-up feature is encoded as the closest annual visit time. This model (results in S1 and S2 Tables) performed very close to the NMM model we present here.





**Fig 1. Feed-forward, fully-connected neural network architecture.** Nonlinear models have rectified linear units (ReLU) between layers. Final layer implements a soft-max.  $L_i$ : Number of neurons in layer  $i$ . Input features include following. Demog.: Deomographics. Clinical Assess.: Clinical Assessments. Cog. Assess.: Cognitive Assessments. Baseline DX: Baseline Diagnosis. prob.: probability.

<https://doi.org/10.1371/journal.pone.0277322.g001>

## Loss function

We use categorical cross-entropy loss to train our neural networks:

$$-\frac{1}{N} \sum_{i=1}^N \langle y_i, \log(\hat{y}_i) \rangle \quad (1)$$

where  $N$  is the number of training datapoints,  $\langle \cdot, \cdot \rangle$  is the inner product operator,  $y_i$  is the one-hot encoded vector for the ground truth label of sample  $i$  (i.e., is a dummy variable vector encoding presence), and  $\hat{y}_i$  is the probability vector of the sample point  $i$  that is computed by the classifier. The expression in Eq (1) represents the average of the losses across the entire training dataset, which implies that each sample has the same weight. This is undesirable in unbalanced problems due to the fact that the contribution of the majority class to the loss function will be higher compared to the minority class.

As can be seen in Table 2, there are two types of imbalance we need to consider.

1. The distribution of class labels varies significantly over the years. The number of CN baseline participants who convert to MCI is smaller than the number of non-converters over the five-year period, but the relative difference shrinks in later years. For participants who

are MCI at baseline the situation is more drastic. Those who convert to AD represent a small minority in the early years, yet MCI-to-AD converters are the majority at the 5-year mark.

2. The number of available clinical labels decreases with each follow-up year, since individuals drop out of the longitudinal study.

Naively using the loss term in Eq (1) would encourage certain types of errors. For example, the model would care less about accurately classifying CN-to-MCI converters, particularly in earlier years. This would affect all three of our models. The second imbalance factor, on the other hand, would exclusively impact the performance of NMM. With the loss in Eq (1), NMM would pay less attention to later follow-up years than earlier ones which means the performance of NMM would suffer in longer time horizons. We note that the LSM and NSM approaches are not affected by this because a separate model is trained for each follow-up year.

There are well-established ways of addressing imbalance issues. Under-sampling the majority class, over-sampling the minority class, and using re-weighted loss functions are the most popular options. In this work, we employ a loss re-weighting scheme. In this approach, the model is penalized more for an error in the minority class than an error in the majority class, using sample-level weights. The loss function we use in this work is

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^N \mathbf{w}_i \langle y_i, \log(\hat{y}_i) \rangle \quad (2)$$

where  $\mathbf{w}_i$  is the weight of the sample point  $i$ . We propose a weighting scheme that accounts for both sources of imbalance we discussed above. Although converter CN baseline participants and non-converter MCI baseline participants belong to the same ground-truth class, we do not weigh those samples equally as they represent different prediction scenarios. Therefore, for a given follow-up year, we consider four possible categories of participants: CN baseline non-converters, CN baseline converters, MCI baseline non-converters, and MCI baseline converters. In the first step, we compute the weights for each category as one over the size of each group. In step 2, we scale these weights so that the total weight of each follow-up year is equal. By doing so, we are addressing the second imbalance.

In the single-year models, the ground-truth clinical status, e.g. for one year after the baseline, was the diagnosis made at the visit corresponding to that follow-time. We note that the timing of this visit is typically not exact and can deviate by several months. For example, a planned 1-year follow-up could have occurred around 15 months after the baseline visit. Thus, the ground-truth labels can be viewed as noisy. On the other hand, for the NMM model, since the follow-time is not fixed and is treated as an input variable (coded in months), the ground-truth label can be viewed as more accurate. That said, as we described above, we implemented a version of the NMM model that accepts the rounded annualized visit times as input (see Supplementary Material) and we observed no meaningful difference in results.

## Experimental details

We implemented a randomized, diagnosis stratified 80–20 split of the data into train-test sets. We repeated this 80–20 split 200 times and all presented results are averaged over these repeats. For each 80–20 split, we also conducted a 5-fold cross-validation on the train sets, where the validation loss was used for early stopping. For each cross-validation fold and modeling choice, we trained 5 different models with different random initializations. Thus, for each test case, the final prediction is computed as the average of the 25 model predictions (5 cross-validation and 5 random initializations each). For the NMM model, we ensure that all



longitudinal follow-up data for a participant is in the same partition. All three of our models use the same data splits.

The model architecture and hyperparameter values are fixed for NMM and single-year models, with the only difference being the missing input neuron of  $\Delta t$  in LSM and NSM. These choices were manually determined based on inspecting the validation loss in a single split. The architecture, illustrated in Fig 1, has 3 hidden layers with a width of 128, followed by 5 hidden layers of width 64 and 2 hidden layers of width 32. We perform early stopping during training based on the validation loss. We employ an L2 penalty loss on the weights and biases, with a weight of  $10^{-6}$  in each hidden layer. We implement dropout after each hidden layer with a rate of 0.2. Our optimizer is Adam with a learning rate of  $10^{-5}$ . We use the softmax activation function at the output layer.

In order to explore the influence of the network architecture on results, we conduct two more experiments where we use the same training strategy, activation functions and hyperparameters as NMM, however we modify the architecture that is illustrated in Fig 1. The first experiment uses an elementary 3-layer nonlinear architecture with fewer parameters than NMM. The second experiment is a computationally expensive practice where we optimize the nonlinear architecture via a grid search over depth and width hyperparameters in each one of the 200 train/test splits. In this experiment, each test set has its optimal architecture, identified by the hyperparameter values that yield the best performance on a validation set. Results of both experiments are presented in S1 Fig. We note that the overall trend and pattern of the NMM's results hold consistent across these architectural choices. The 3-layer model's performance is slightly worse and the optimized architecture results are the best, as expected. The 11-layer model we present in Fig 1 under-performs slightly compared to the optimized architectures. We note that the 11-layer model was manually designed to optimize performance in a single train/validation split.

We analyze prediction accuracy in two different patient groups: CN at baseline and MCI at baseline. Therefore, each result we show has two parts, one corresponding to the CN-to-MCI conversion task and the other to the MCI-to-AD conversion task. In each task, clinical conversion is considered a positive event. For example, for the CN-to-MCI conversion task, a true positive sample refers to the subject progressing to MCI and the model predicting this correctly. Accordingly, the true positive rate is defined as the ratio of true positive samples against all converter subjects, and the false positive rate is the ratio of false positives against all stable subjects.

Due to the heavily unbalanced nature of the data, which can be seen in Table 2, we use the receiver operating characteristic (ROC) curve to inspect the performance of our models. Although there are no subjects who convert from CN to AD or from MCI to CN in our dataset, we do not implement any mechanism to prevent our models from making such predictions. Therefore, both CN-to-MCI conversion and MCI-to-AD conversion tasks are multi-class problems for our models. There are two different types of ROC analyses for multi-class problems: one-versus-one analysis and one-versus-rest analysis. As we demonstrate with our results, our models capture the disease progression dynamics sufficiently that both analyses give nearly identical ROC curves. In other words, the predicted AD probability for CN baseline subjects and the predicted CN probability for MCI baseline subjects are close to 0. Thus, we only share one-vs-rest results where the positive class for CN-to-MCI conversion is MCI and for MCI-to-AD conversion it is AD.

The area under the ROC curve (ROC AUC) is a scalar that summarizes the overall performance of a classifier. Our data has a time horizon of five follow-up years and we evaluate against annual diagnoses. For each of our three models, we compute an ROC AUC value corresponding to each follow-up year and each baseline group (CN baseline and MCI baseline).

Therefore, each model has five ROC AUC values associated with each baseline group. To statistically compare the ROC AUC values achieved by two different models, we implement a pairwise permutation testing strategy, yielding a p-value for the null hypothesis that the two models' predictions are indistinguishable. Our test statistic is the difference between the mean ROC AUC values (averaged over the annual follow-ups) for the two models in a given 80–20 train-test split. We then average this over all 200 random splits of our data. To create the null distribution of the test statistic, we randomly permute ( $10^5$  times) the two models when computing the ROC AUC difference for each split. Finally, the normalized rank of the observed (unpermuted) test statistic value among all sorted (permuted) test statistic values yields the p-value, which we denote with  $\rho$ .

## Results

### Impact of modeling choices

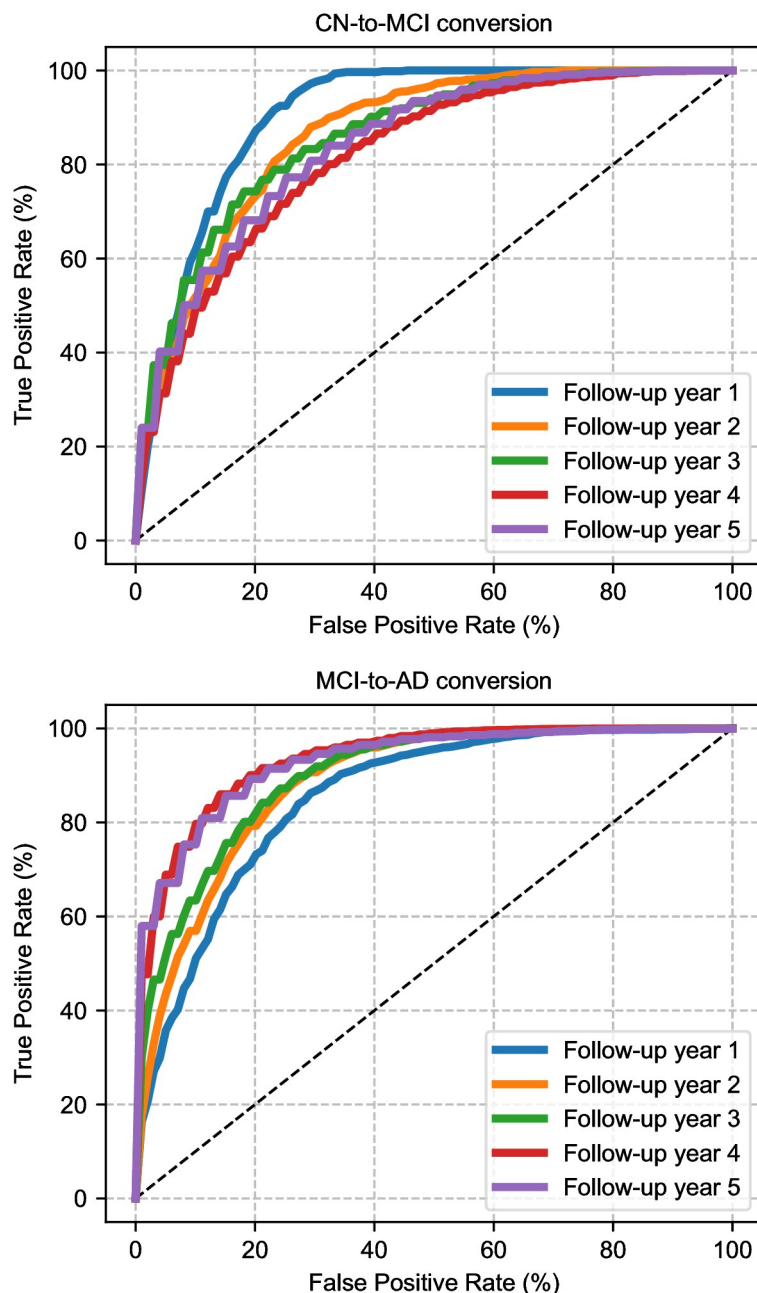
In CN-to-MCI conversion, we observe that there is a substantial difference between the linear and nonlinear models. For example, for the 1-year follow-up, LSM yields 83.88% ROC AUC, whereas its nonlinear counterpart, NSM, achieves 88.73%. This difference remains stable over all follow-up years and is statistically significant ( $\rho < 0.0001$ ). The multi-year training strategy, on the other hand, further boosts prediction accuracy. For instance, for the 1-year follow-up, NMM achieves an ROC AUC of 90.40%. The difference with the NSM model is consistent over the follow-up years and statistically significant ( $\rho = 0.0001$ ). Finally, we note that for CN-to-MCI conversion, all models tend to achieve worse performance as the time-horizon increases. For instance, the best-performing NMM model suffers more than a 6% drop in ROC AUC between 1- and 5-year follow-up predictions. This result suggests that it is easier to predict who will convert from CN to MCI in the relatively short term, say within a year, than in the longer term, say within 5 years.

We notice that the performance of LSM fluctuates as a function of the time horizon. There are two local minima, one at 2- and another at 4-year follow-up. This is likely because those two years include a higher percentage of CN subjects, due to the study design of ADNI 2 and 3, as can be seen in Table 2. We see that this affects the performance of the nonlinear single-year model, NSM, too. However, for the NMM the issue is mitigated, which is likely because the multi-year model can leverage the data from the other follow-up years to “smooth out” its predictions.

In MCI-to-AD conversion, there is an overall diminished difference between the performance of the three models. For the single-year models, the linear and nonlinear counterparts are statistically indistinguishable ( $\rho = 0.3735$ ). The multi-year model, on the other hand, offers a statistically significant ( $\rho = 0.0004$  against LSM,  $\rho = 0.0014$  against NSM), yet subtle boost in ROC AUC, specifically for 1- and 2-year follow-ups. In the remaining follow-up years, all three models achieve essentially the same performance level. The most striking observation from the MCI-to-AD conversion results is that prediction accuracy improves for later years, and there is a very consistent increase in ROC AUC values across all modeling choices. This indicates that it is relatively easier to predict who will convert from MCI to AD in the 4–5 year horizon compared to the 1–2 year horizon. Fig 2 shows corresponding ROC curves of NMM for each follow-up year and each patient group.

### Contribution of different biomarkers

As mentioned above, our models are capable of handling missing values in the input. This allows us to inspect the contribution of different data types to prediction accuracy. We perform this analysis on our best performing model, NMM, where we focus only on test



**Fig 2. ROC curves of NMM for CN-to-MCI and MCI-to-AD conversion in five-year time horizon.** Displayed are averages of 200 train-test splits.

<https://doi.org/10.1371/journal.pone.0277322.g002>

participants with complete baseline data and systematically mask each input feature, treating it as missing.

Our baseline scenario is where only clinical data (CD) is available. Fig 4 shows the difference in AUC ROC ( $\Delta$  AUC ROC) achieved with the utilization of additional biomarkers: FDG PET (a single global marker of sugar metabolism), CSF (global markers of tau and amyloid burden), AV45 PET (a single global marker of brain amyloid load), MRI volumetric measurements (markers of brain atrophy). For MRI, we consider two scenarios. First, we only use the

value of hippocampus volume, normalized by the intracranial volume (ICV) (CD+ICV normalized Hippocampus size in Fig 4). In the second scenario, we use seven MRI-derived AD-associated biomarkers (CD+MRI in Fig 4). As a reference, we also show the results for including all available biomarkers in these test subjects that have complete baseline data (CD+All Biomarkers in Fig 4). We were not able to quantify the contribution of PIB PET, as only a very limited number of participants have PIB PET scans.

In CN-to-MCI conversion, molecular biomarkers (FDG, CSF, and AV45), by themselves, do not significantly improve performance over the baseline CD-only scenario, particularly beyond the 1-year follow-up ( $\rho = 0.1898$  for CD+FDG,  $\rho = 0.2082$  for CD+CSF,  $\rho = 0.3001$  for CD+AV45). However, we observe a substantial accuracy boost when MRI data are available ( $\rho < 0.0001$ ), much of which can be attributed to the hippocampus volume ( $\rho < 0.0001$ ). All biomarkers combined achieve the highest ROC AUC values ( $\rho < 0.0001$ ). The performance gain grows over the years, suggesting that additional biomarkers are more useful for making longer-term predictions.

Overall, the performance gain offered by additional biomarkers is relatively smaller for the easier MCI-to-AD conversion problem. Here, MRI markers add around 1% ROC AUC to the CD-only baseline. FDG consistently yields a greater boost than the MRI biomarkers in each follow-up year, which is in contrast to what we observe in CN-to-MCI conversion. Crucially, we find that hippocampus volume does not provide a statistically significant performance boost ( $\rho = 0.2412$ ), while FDG and MRI markers improve the model performance subtly but significantly ( $\rho < 0.0001$  for CD+FDG,  $\rho = 0.0024$  for CD+MRI). Beyond year 1, CSF consistently outperforms AV45 ( $\rho < 0.0010$  for CD+CSF,  $\rho = 0.0253$  for CD+AV45), where the latter yields a boost on par with MRI. This highlights the potential importance of tau markers, particularly in the MCI stage. Overall, however, a striking observation is that the model that has access to all the biomarkers is substantially more accurate than a model with a single biomarker type.

## Disease progression risk predictions

Even though we consider the problem as a three-label classification task for a given follow-up time, the underlying process can be viewed as a continuous evolution of MCI and AD dementia risk [57]. Using our NMM model, we can compute a prediction for arbitrary time horizons for the test subjects and interpret the output probabilities as a longitudinal estimation of risk. The softmax outputs of the MCI channel for CN baseline participants and AD channel for the MCI baseline participants are shown in Figs 5 and 6, respectively. We average these values over test subjects who have the same conversion time profile.

For individuals who remain stable CN throughout the 5-year follow-up period, we observe that NMM's MCI prediction is consistently less than 50%. Intriguingly, for those stable subjects who were last observed earlier, the predicted MCI probabilities tend to be higher. In fact, for stable CN subjects last seen before the end of year 2, average predicted MCI probabilities exceed 50% around the year 4 mark. We emphasize that the model has no access to follow-up information, as the only input is baseline data. For subjects who convert to MCI at year 1, average predicted MCI probabilities exceed 0.5 before the first annual follow-up visit. Similarly, for those who convert around the second year, the average predicted MCI probabilities exceed 0.5, between years one and two. One notable exception is the group of individuals who progress to MCI at the third-year visit. In this group, the NMM prediction is that MCI conversion will happen, on average, at around the 5-year mark.

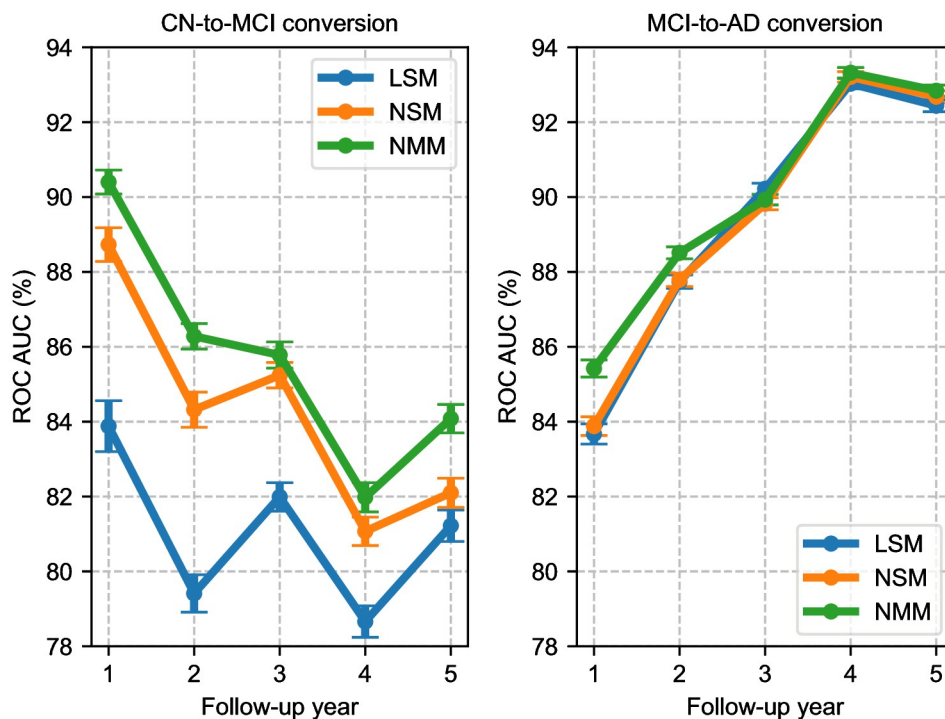
For the MCI baseline subjects, we observe similar patterns. For the stable subjects, the predicted AD probabilities remain under 0.5 until the last follow-up visit. For MCI-to-AD

converters, the average predicted AD probability exceeds 0.5 before the AD diagnosis, except for the subjects who convert at the 5-year follow-up, where the average predicted AD probability is slightly below 0.5 at the 5-year mark. On the other hand, when we examine the timing of the average predicted conversion, it seems to be less accurate than with CN baselines. In most scenarios (e.g. conversion at 2, 3, and 4 years), the average predicted AD probability exceeds 0.5 before the corresponding time interval. This suggests that the NMM's predictions tend to estimate an earlier MCI-to-AD prediction than observed.

## Discussion

In this work, we present an empirical study to characterize how predictable an individual subject's future AD-associated clinical trajectory is, several years in advance, based on rich multi-modal data, and using modern deep learning methods. We present a novel machine learning strategy that can handle variable follow-up time queries, missingness patterns, and unbalanced class labels in the data, to make accurate predictions about the future decline in CN and MCI baseline participants.

Comparing the prediction accuracy for CN-to-MCI and MCI-to-AD conversions in Fig 3, our results verify that the CN-to-MCI conversion prediction is a harder task than the MCI-to-AD conversion prediction. On the other hand, we also confirm that more sophisticated modeling, such as a nonlinear multi-year (NMM) architecture, offers a larger boost for the harder CN-to-MCI conversion prediction task. This verifies that there is a bigger gap in performance between what a relatively simple model can achieve and the upper bound of what



**Fig 3. Predictive performance of different models for different follow-up years.** ROC AUC values are averaged across 200 80–20 data splits. Error bars indicate the standard error across these splits. LSM, Linear Single-year Model; NSM, Nonlinear Single-year Model; NMM, Nonlinear Multi-year Model.

<https://doi.org/10.1371/journal.pone.0277322.g003>

is achievable (also known as the Bayes-optimal performance) in the harder problem of CN-to-MCI conversion.

Five years is a relatively short time window for studying CN-to-MCI conversion. On the other hand, in many real-world clinical scenarios, 5 years is a useful horizon to consider. Moreover, we note that at year 5, around 30% of the baseline CN subjects who remained in the study had converted to MCI, as we show in [Table 2](#). Our analysis demonstrates that the prediction of CN-to-MCI conversion gets harder for distant time horizons, and we achieve higher accuracy for shorter time frames. This insight might be useful in detecting those CN subjects who might be on the cusp of developing MCI.

Despite the missingness in the data, [Fig 4](#) suggests that NMM does not rely solely on a single modality. Additional biomarkers, in general, do not make the prediction performance worse. This finding is in parallel with the fact that multi-modal data, such as different MRI sequences and various PET tracers are often combined in the literature for predicting MCI-to-AD conversion [58–60]. However, our results also demonstrate that the predictive value of each additional biomarker can vary. For example, for CN baseline participants, although there is a substantial accuracy increase with the use of MRI; molecular biomarkers (CSF, FDG, and AV45) do not offer a significant boost beyond the first year horizon. For the prediction of MCI-to-AD conversion, however, the situation is different—molecular biomarkers offer a significant boost. Furthermore, using the different MRI biomarkers together seems to be much more helpful for predicting MCI-to-AD conversion, rather than relying on a single MRI biomarker, namely the ICV-normalized hippocampal volume. These results highlight the importance of characterizing the diagnostic and predictive utility of different data types, at different stages of the disease process.

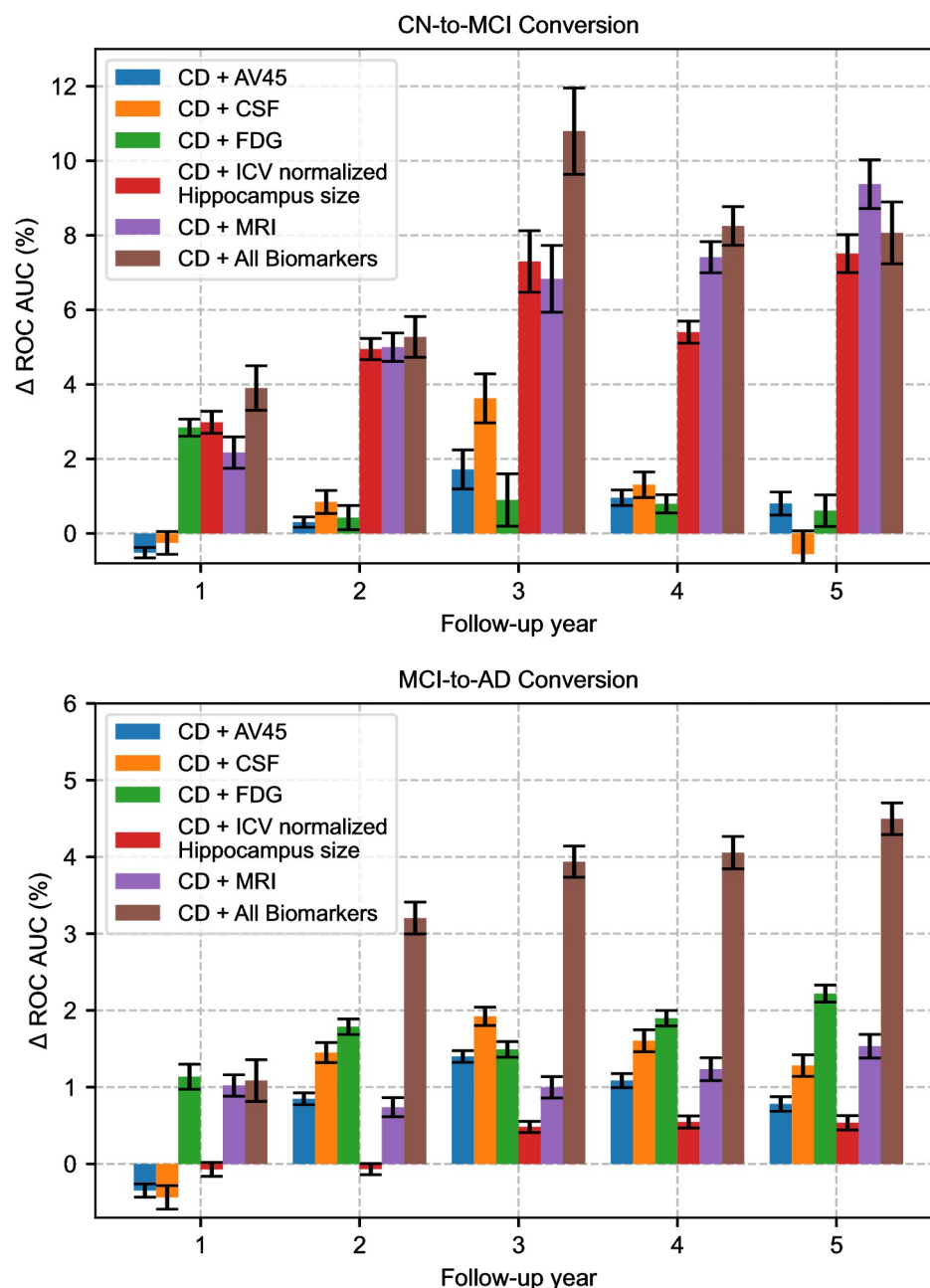
One interpretation of the patterns of results we present in this study might be that amyloid or tau-associated biomarker changes have a relatively longer timecourse than MRI derived measurements, such as hippocampal volume. Furthermore, MRI markers may be less specific and reflect a multitude of effects that result in atrophy, particularly at later ages. Thus, MRI might predict more proximal decline from CN to MCI, but its utility will be less during the MCI stage, where tau/amyloid markers might offer some specific insights into the Alzheimer's pathology dynamics that will play out over the next several years.

The conversion risk predictions that we show in [Figs 5 and 6](#) suggest that NMM captures the continuous disease dynamics. However, NMM's predictions are not always exactly aligned with the timing of events. This issue can be related to various biases in subject recruitment and follow-up in the ADNI [61]. For example, the data suffer from a “temporal bias” [62] that is caused by the fact that baseline visits are not distributed uniformly over latent disease stages. These shortcomings require further investigations, likely demanding novel methodological approaches that can address the selection and temporal biases in the data and possibly exploiting other cohorts, as in [63].

## Conclusion

We have presented a machine learning approach that uses participants' multimodal baseline data with arbitrary missingness, to predict their future diagnostic status at any time point. We have demonstrated that our model can capture disease progression dynamics and produce future conversion predictions that are highly accurate. Our analyses allow us to dissect the impact of modeling choices and input data types. We found that molecular biomarkers are more useful for predicting MCI-to-AD conversion than CN-to-MCI conversion. Our results show that MRI features are essential for both types of predictions, yet different types of MRI-derived measurements can be useful in different stages.





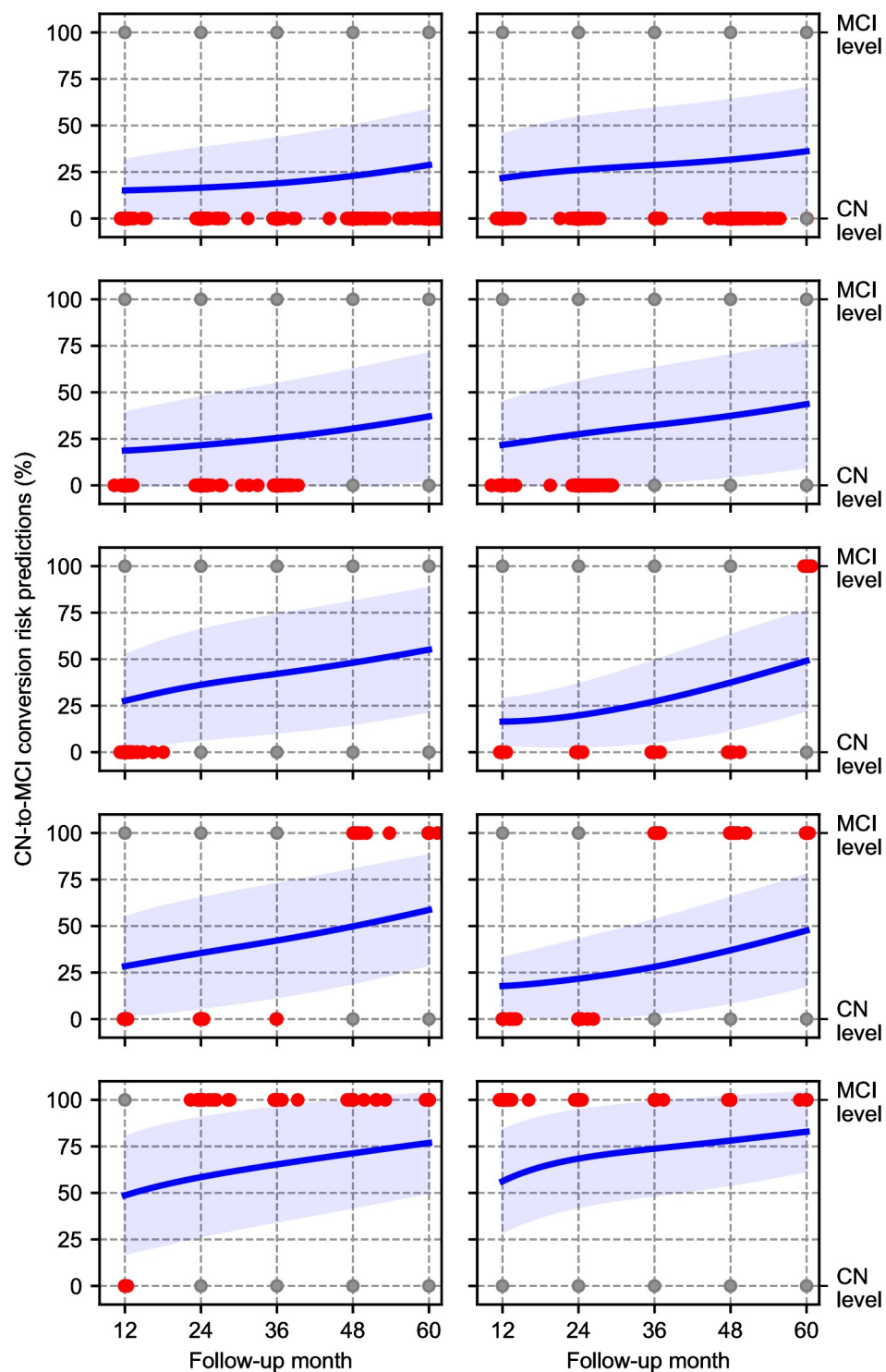
**Fig 4.**  $\Delta$  ROC AUC values obtained with NMM by the addition of various biomarker combinations to the clinical data (participant demographics, clinical assessments, and cognitive assessments). ROC AUC values are averaged across 200 80–20 data splits. Error bars indicate the standard error across these splits. +: Used together. CD, Clinical data; AV45, Florbetapir PET; CSF, Cerebrospinal Fluid; FDG, Fluorine-18-Fluorodeoxyglucose PET; MRI, Magnetic Resonance Imaging; ICV, Intracranial Volume.

<https://doi.org/10.1371/journal.pone.0277322.g004>

## Supporting information

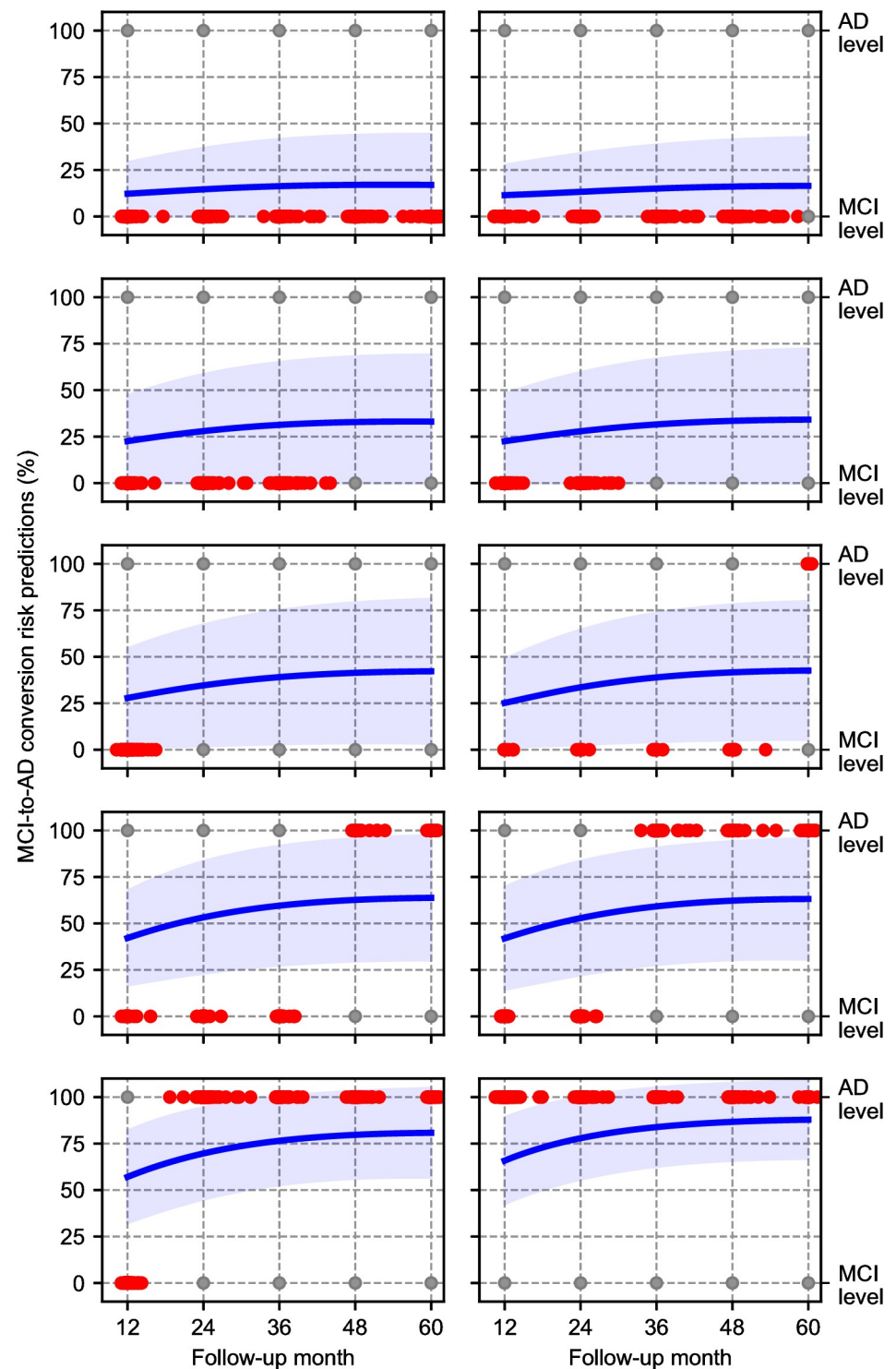
### S1 Table. Performance of each model in terms of AUC ROC for CN baseline participants.

Data format is mean  $\pm$  standard error. LSM<sup>†</sup> is a standard linear ridge regression model that is an alternative implementation of LSM (Linear Single-year Model). NMM<sup>†</sup> is a slight



**Fig 5. Conversion risk predictions of NMM for CN baseline participants with different ground truth disease trajectories.** Blue line is the average MCI conversion risk with 68% confidence. Red dots represent the observed diagnosis time (x-coordinate) and the observed diagnosis (y-coordinate) of the participants with the corresponding trajectory. Grey dots are for reference.

<https://doi.org/10.1371/journal.pone.0277322.g005>



**Fig 6. Conversion risk predictions of NMM for MCI baseline participants with different ground truth disease trajectories.** Blue line is the average AD conversion risk with 68% confidence. Red dots represent the observed diagnosis time (x-coordinate) and the observed diagnosis (y-coordinate) of the participants with the corresponding trajectory. Grey dots are for reference.

<https://doi.org/10.1371/journal.pone.0277322.g006>

modification of NMM (Nonlinear Multi-year Model), where the input time-to-follow-up ( $\Delta t$ ) feature is encoded as the closest annual visit time. NSM: Nonlinear Single-year Model. (PDF)

**S2 Table. Performance of each model in terms of AUC ROC for MCI baseline participants.**

See caption of [S1 Table](#).

(PDF)

**S1 Fig. Predictive performance of NMMs with different architectures.** ROC AUC values are averaged across 200 80–20 data splits. Error bars indicate the standard error across these splits. NMM, Nonlinear Multi-year Model with the architecture shown in [Fig 1](#); NMM (3-layer), Nonlinear Multi-year Model with a three-layer architecture; NMM (Optimized), Nonlinear Multi-year Model with optimized architectures for each test set. Details of NMM (3-layer) and NMM(optimized) can be found in [S1 Text](#).

(PDF)

**S1 Text. Details of NMM (3-layer) and NMM (optimized).** We use the same hyperparameters and activation functions for NMM (3-layer) and NMM (optimized) as NMM. NMM (3-layer) has an architecture consisting of 2 hidden layers with a width of 128 and an output layer of width 3. NMM (optimized) architectures for each test split are searched over a  $3 \times 3$  grid, characterized by two parameters: width (W) and depth (D). W represents the number of neurons in the first hidden layer, and it can be either 64, 128, or 256. D represents the depth of the architecture in terms of equally wide blocks in [Fig 1](#), i.e., a D of 1 means the architecture has 3 hidden layers of width W; a D of 2 means the architecture has 3 hidden layers of width W, followed by 5 hidden layers of width W/2; and a depth of 3 means the architecture has 3 hidden layers of width W, followed by 5 hidden layers of width W/2, followed by 2 hidden layers of width W/4. All architectures have an output layer with a width of 3. The best architecture is chosen by monitoring the validation loss in one of the train/validation splits.

(PDF)

## Acknowledgments

The authors would like to acknowledge the anonymous reviewers for their valuable revision of the manuscript.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database <http://adni.loni.usc.edu>. Applications for ADNI data use can be submitted through the ADNI website at <https://adni.loni.usc.edu/data-samples/access-data/>. Others would be able to access the data in the same manner as the authors. The authors did not have any special access privileges that others would not have. 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. Michael Weiner (E-mail: [Michael.Weiner@ucsf.edu](mailto:Michael.Weiner@ucsf.edu)) serves as the principal investigator for ADNI. A complete listing of ADNI investigators and their affiliations can be found below.

Michael Weiner<sup>4</sup>, Paul Aisen<sup>5</sup>, Ronald Petersen<sup>6</sup>, Clifford R. Jack Jr.<sup>6</sup>, William Jagust<sup>7</sup>, John Q. Trojanowki<sup>8</sup>, Arthur W. Toga<sup>9</sup>, Laurel Beckett<sup>10</sup>, Robert C. Green<sup>11</sup>, Andrew J. Saykin<sup>12</sup>, John Morris<sup>13</sup>, Leslie M. Shaw<sup>14</sup>, Enchi Liu<sup>15</sup>, Tom Montine<sup>16</sup>, Ronald G. Thomas<sup>5</sup>, Michael Donohue<sup>5</sup>, Sarah Walter<sup>5</sup>, Devon Gessert<sup>5</sup>, Tamie Sather<sup>5</sup>, Gus Jimenez<sup>5</sup>, Danielle Harvey<sup>10</sup>, Michael Donohue<sup>5</sup>, Matthew Bernstein<sup>6</sup>, Nick Fox<sup>17</sup>, Paul Thompson<sup>18</sup>, Norbert Schuff<sup>19</sup>, Charles DeCarli<sup>10</sup>, Bret Borowski<sup>6</sup>, Jeff Gunter<sup>6</sup>, Matt Senjem<sup>6</sup>, Prashanthi Vemuri<sup>6</sup>, David Jones<sup>6</sup>, Kejal Kantarci<sup>6</sup>, Chad Ward<sup>6</sup>, Robert A. Koeppe<sup>20</sup>, Norm Foster<sup>21</sup>, Eric M. Reiman<sup>22</sup>,

Kewei Chen<sup>22</sup>, Chet Mathis<sup>23</sup>, Susan Landau<sup>7</sup>, Nigel J. Cairns<sup>13</sup>, Erin Householder<sup>13</sup>, Lisa Taylor Reinwald<sup>13</sup>, Virginia Lee<sup>24</sup>, Magdalena Korecka<sup>24</sup>, Michal Figurski<sup>24</sup>, Karen Crawford<sup>9</sup>, Scott Neu<sup>9</sup>, Tatiana M. Foroud<sup>12</sup>, Steven Potkin<sup>25</sup>, Li Shen<sup>12</sup>, Faber Kelley<sup>12</sup>, Sungeun Kim<sup>12</sup>, Kwangsik Nho<sup>12</sup>, Zaven Kachaturian<sup>26</sup>, Richard Frank<sup>27</sup>, Peter J. Snyder<sup>28</sup>, Susan Molchan<sup>29</sup>, Jeffrey Kaye<sup>30</sup>, Joseph Quinn<sup>30</sup>, Betty Lind<sup>30</sup>, Raina Carter<sup>30</sup>, Sara Dolen<sup>30</sup>, Lon S. Schneider<sup>31</sup>, Sonia Pawluczyk<sup>31</sup>, Mauricio Beccera<sup>31</sup>, Liberty Teodoro<sup>31</sup>, Bryan M. Spann<sup>31</sup>, James Brewer<sup>32</sup>, Helen Vanderswag<sup>32</sup>, Adam Fleisher<sup>22</sup>, Judith L. Heidebrink<sup>20</sup>, Joanne L. Lord<sup>20</sup>, Ronald Petersen<sup>6</sup>, Sara S. Mason<sup>6</sup>, Colleen S. Albers<sup>6</sup>, David Knopman<sup>6</sup>, Kris Johnson<sup>6</sup>, Rachelle S. Doody<sup>33</sup>, Javier Villanueva Meyer<sup>33</sup>, Munir Chowdhury<sup>33</sup>, Susan Rountree<sup>33</sup>, Mimi Dang<sup>33</sup>, Yaakov Stern<sup>34</sup>, Lawrence S. Honig<sup>34</sup>, Karen L. Bell<sup>34</sup>, Beau Ances<sup>35</sup>, John C. Morris<sup>35</sup>, Maria Carroll<sup>35</sup>, Sue Leon<sup>35</sup>, Erin Householder<sup>13</sup>, Mark A. Mintun<sup>35</sup>, Stacy Schneider<sup>35</sup>, Angela OliverNG<sup>36</sup>, Randall Griffith<sup>36</sup>, David Clark<sup>36</sup>, David Geldmacher<sup>36</sup>, John Brockington<sup>36</sup>, Erik Roberson<sup>36</sup>, Hillel Grossman<sup>37</sup>, Effie Mitsis<sup>37</sup>, Leyla de Toledo-Morrell<sup>38</sup>, Raj C. Shah<sup>38</sup>, Rannjan Duara<sup>39</sup>, Daniel Varon<sup>39</sup>, Maria T. Greig<sup>39</sup>, Peggy Roberts<sup>39</sup>, Marilyn Albert<sup>40</sup>, Chiadi Onyike<sup>40</sup>, Daniel D'Agostino II<sup>40</sup>, Stephanie Kielb<sup>40</sup>, James E. Galvin<sup>41</sup>, Dana M. Pogorelec<sup>41</sup>, Brittany Cerbone<sup>41</sup>, Christina A. Michel<sup>41</sup>, Henry Rusinek<sup>41</sup>, Mony J. de Leon<sup>41</sup>, Lidia Glodzik<sup>41</sup>, Susan De Santi<sup>41</sup>, P. Murali Doraiswamy<sup>42</sup>, Jeffrey R. Petrella<sup>42</sup>, Terence Z. Wong<sup>42</sup>, Steven E. Arnold<sup>14</sup>, Jason H. Karlawish<sup>14</sup>, David Wolk<sup>14</sup>, Charles D. Smith<sup>43</sup>, Greg Jicha<sup>43</sup>, Peter Hardy<sup>43</sup>, Partha Sinha<sup>43</sup>, Elizabeth Oates<sup>43</sup>, Gary Conrad<sup>43</sup>, Oscar L. Lopez<sup>23</sup>, MaryAnn Oakley<sup>23</sup>, Donna M. Simpson<sup>23</sup>, Anton P. Porsteinsson<sup>44</sup>, Bonnie S. Goldstein<sup>44</sup>, Kim Martin<sup>44</sup>, Kelly M. Makino<sup>44</sup>, M. Saleem Ismail<sup>44</sup>, Connie Brand<sup>44</sup>, Ruth A. Mulnard<sup>45</sup>, Gaby Thai<sup>45</sup>, Catherine Mc Adams Ortiz<sup>45</sup>, Kyle Womack<sup>46</sup>, Dana Mathews<sup>46</sup>, Mary Quiceno<sup>46</sup>, Ramon Diaz Arrastia<sup>46</sup>, Richard King<sup>46</sup>, Myron Weiner<sup>46</sup>, Kristen Martin Cook<sup>46</sup>, Michael DeVous<sup>46</sup>, Allan I. Levey<sup>47</sup>, James J. Lah<sup>47</sup>, Janet S. Cellar<sup>47</sup>, Jeffrey M. Burns<sup>48</sup>, Heather S. Anderson<sup>48</sup>, Russell H. Swerdlow<sup>48</sup>, Liana Apostolova<sup>49</sup>, Kathleen Tingus<sup>49</sup>, Ellen Woo<sup>49</sup>, Daniel H. S. Silverman<sup>49</sup>, Po H. Lu<sup>49</sup>, George Bartzokis<sup>49</sup>, Neill R. Graff Radford<sup>50</sup>, Francine Parfitt<sup>50</sup>, Tracy Kendall<sup>50</sup>, Heather Johnson<sup>50</sup>, Martin R. Farlow<sup>12</sup>, Ann Marie Hake<sup>12</sup>, Brandy R. Matthews<sup>12</sup>, Scott Herring<sup>12</sup>, Cynthia Hunt<sup>12</sup>, Christopher H. van Dyck<sup>51</sup>, Richard E. Carson<sup>51</sup>, Martha G. MacAvoy<sup>51</sup>, Howard Chertkow<sup>52</sup>, Howard Bergman<sup>52</sup>, Chris Hosein<sup>52</sup>, Sandra Black<sup>53</sup>, Bojana Stefanovic<sup>53</sup>, Curtis Caldwell<sup>53</sup>, Ging Yuek Robin Hsiung<sup>54</sup>, Howard Feldman<sup>54</sup>, Benita Mudge<sup>54</sup>, Michele Assaly Past<sup>54</sup>, Andrew Kertesz<sup>55</sup>, John Rogers<sup>55</sup>, Dick Trost<sup>55</sup>, Charles Bernick<sup>56</sup>, Donna Munic<sup>56</sup>, Diana Kerwin<sup>57</sup>, Marek Marsel Mesulam<sup>57</sup>, Kristine Lipowski<sup>57</sup>, Chuang Kuo Wu<sup>57</sup>, Nancy Johnson<sup>57</sup>, Carl Sadowsky<sup>58</sup>, Walter Martinez<sup>58</sup>, Teresa Villena<sup>58</sup>, Raymond Scott Turner<sup>59</sup>, Kathleen Johnson<sup>59</sup>, Brigid Reynolds<sup>59</sup>, Reisa A. Sperling<sup>60</sup>, Keith A. Johnson<sup>60</sup>, Gad Marshall<sup>60</sup>, Meghan Frey<sup>60</sup>, Jerome Yesavage<sup>61</sup>, Joy L. Taylor<sup>61</sup>, Barton Lane<sup>61</sup>, Allyson Rosen<sup>61</sup>, Jared Tinklenberg<sup>61</sup>, Marwan N. Sabbagh<sup>62</sup>, Christine M. Belden<sup>62</sup>, Sandra A. Jacobson<sup>62</sup>, Sherye A. Sirrel<sup>62</sup>, Neil Kowall<sup>63</sup>, Ronald Killiany<sup>63</sup>, Andrew E. Budson<sup>63</sup>, Alexander Norbash<sup>63</sup>, Patricia Lynn Johnson<sup>63</sup>, Thomas O. Obisesan<sup>64</sup>, Saba Wolday<sup>64</sup>, Joanne Allard<sup>64</sup>, Alan Lerner<sup>65</sup>, Paula Ogrocki<sup>65</sup>, Leon Hudson<sup>65</sup>, Evan Fletcher<sup>66</sup>, Owen Carmichael<sup>66</sup>, John Olichney<sup>66</sup>, Charles DeCarli<sup>66</sup>, Smita Kittur<sup>67</sup>, Michael Borrie<sup>68</sup>, T. Y. Lee<sup>68</sup>, Rob Bartha<sup>68</sup>, Sterling Johnson<sup>69</sup>, Sanjay Asthana<sup>69</sup>, Cynthia M. Carlsson<sup>69</sup>, Steven G. Potkin<sup>70</sup>, Adrian Preda<sup>70</sup>, Dana Nguyen<sup>70</sup>, Pierre Tariot<sup>22</sup>, Adam Fleisher<sup>22</sup>, Stephanie Reeder<sup>22</sup>, Vernice Bates<sup>71</sup>, Horacio Capote<sup>71</sup>, Michelle Rainka<sup>71</sup>, Douglas W. Scharre<sup>72</sup>, Maria Kataki<sup>72</sup>, Anahita Adeli<sup>72</sup>, Earl A. Zimmerman<sup>73</sup>, Dzintra Celmins<sup>73</sup>, Alice D. Brown<sup>73</sup>, Godfrey D. Pearlson<sup>74</sup>, Karen Blank<sup>74</sup>, Karen Anderson<sup>74</sup>, Robert B. Santulli<sup>75</sup>, Tamar J. Kitzmiller<sup>75</sup>, Eben S. Schwartz<sup>75</sup>, Kaycee M. SinkS<sup>76</sup>, Jeff D. Williamson<sup>76</sup>, Pradeep Garg<sup>76</sup>, Franklin Watkins<sup>76</sup>, Brian R. Ott<sup>77</sup>, Henry Querfurth<sup>77</sup>, Geoffrey Tremont<sup>77</sup>, Stephen Salloway<sup>78</sup>, Paul Malloy<sup>78</sup>, Stephen Correia<sup>78</sup>, Howard J. Rosen<sup>4</sup>, Bruce L. Miller<sup>4</sup>, Jacobo Mintzer<sup>79</sup>, Kenneth Spicer<sup>79</sup>, David Bachman<sup>79</sup>, Elizabeth Finger<sup>80</sup>, Stephen Pasternak<sup>80</sup>, Irina Rachinsky<sup>80</sup>, John



Rogers<sup>55</sup>, Andrew Kertesz<sup>55</sup>, Dick Drost<sup>80</sup>, Nunzio Pomara<sup>81</sup>, Raymundo Hernando<sup>81</sup>, Antero Sarrael<sup>81</sup>, Susan K. Schultz<sup>82</sup>, Laura L. Boles Ponto<sup>82</sup>, Hyungsub Shim<sup>82</sup>, Karen Elizabeth Smith<sup>82</sup>, Norman Relkin<sup>83</sup>, Gloria Chaing<sup>83</sup>, Lisa Raudin<sup>83</sup>, Amanda Smith<sup>84</sup>, Kristin Faragher<sup>84</sup>, Balebail Ashok Raj<sup>84</sup>

**4** UC San Francisco, San Francisco, CA, USA. **5** UC San Diego, San Diego, CA, USA. **6** Mayo Clinic, Rochester, NY, USA. **7** UC Berkeley, Berkeley, CA, USA. **8** U Pennsylvania, Pennsylvania, CA, USA. **9** USC, Los Angeles, CA, USA. **10** UC Davis, Davis, CA, USA. **11** Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. **12** Indiana University, Bloomington, IN, USA. **13** Washington University St. Louis, St. Louis, MO, USA. **14** University of Pennsylvania, Philadelphia, PA, USA. **15** Janssen Alzheimer Immunotherapy, South San Francisco, CA, USA. **16** University of Washington, Seattle, WA, USA. **17** University of London, London, UK. **18** USC School of Medicine, Los Angeles, CA, USA. **19** UCSF MRI, San Francisco, CA, USA. **20** University of Michigan, Ann Arbor, MI, USA. **21** University of Utah, Salt Lake City, UT, USA. **22** Banner Alzheimer's Institute, Phoenix, AZ, USA. **23** University of Pittsburgh, Pittsburgh, PA, USA. **24** UPenn School of Medicine, Philadelphia, PA, USA. **25** UC Irvine, Newport Beach, CA, USA. **26** Khachaturian, Radebaugh & Associates, Inc and Alzheimer's Association's Ronald and Nancy Reagan's Research Institute, Chicago, IL, USA. **27** General Electric, Boston, MA, USA. **28** Brown University, Providence, RI, USA. **29** National Institute on Aging/National Institutes of Health, Bethesda, MD, USA. **30** Oregon Health and Science University, Portland, OR, USA. **31** University of Southern California, Los Angeles, CA, USA. **32** University of California San Diego, San Diego, CA, USA. **33** Baylor College of Medicine, Houston, TX, USA. **34** Columbia University Medical Center, New York, NY, USA. **35** Washington University, St. Louis, MO, USA. **36** University of Alabama Birmingham, Birmingham, MO, USA. **37** Mount Sinai School of Medicine, New York, NY, USA. **38** Rush University Medical Center, Chicago, IL, USA. **39** Wien Center, Vienna, Austria. **40** Johns Hopkins University, Baltimore, MD, USA. **41** New York University, New York, NY, USA. **42** Duke University Medical Center, Durham, NC, USA. **43** University of Kentucky, city of Lexington, NC, USA. **44** University of Rochester Medical Center, Rochester, NY, USA. **45** University of California, Irvine, CA, USA. **46** University of Texas Southwestern Medical School, Dallas, TX, USA. **47** Emory University, Atlanta, GA, USA. **48** University of Kansas, Medical Center, Lawrence, KS, USA. **49** University of California, Los Angeles, CA, USA. **50** Mayo Clinic, Jacksonville, FL, USA. **51** Yale University School of Medicine, New Haven, CT, USA. **52** McGill Univ., Montreal Jewish General Hospital, Montreal, WI, USA. **53** Sunnybrook Health Sciences, Toronto, ON, Canada. **54** U.B.C. Clinic for AD & Related Disorders, British Columbia, BC, Canada. **55** Cognitive Neurology St. Joseph's, Toronto, ON, Canada. **56** Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, NV, USA. **57** Northwestern University, Evanston, IL, USA. **58** Premiere Research Inst Palm Beach Neurology, West Palm Beach, FL, USA. **59** Georgetown University Medical Center, Washington, DC, USA. **60** Brigham and Women's Hospital, Boston, MA, USA. **61** Stanford University, Santa Clara County, CA, USA. **62** Banner Sun Health Research Institute, Sun City, AZ, USA. **63** Boston University, Boston, MA, USA. **64** Howard University, Washington, DC, USA. **65** Case Western Reserve University, Cleveland, OH, USA. **66** University of California, Davis Sacramento, CA, USA. **67** Neurological Care of CNY, New York, NY, USA. **68** Parkwood Hospital, Parkwood, CA, USA. **69** University of Wisconsin, Madison, WI, USA. **70** University of California, Irvine BIC, Irvine, CA, USA. **71** Dent Neurologic Institute, Amherst, MA, USA. **72** Ohio State University, Columbus, OH, USA. **73** Albany Medical College, Albany, NY, USA. **74** Hartford Hosp, Olin Neuropsychiatry Research Center, Hartford, CT, USA. **75** Dartmouth Hitchcock Medical Center, Albany, NY, USA. **76** Wake Forest University Health Sciences, Winston-Salem, NC, USA. **77** Rhode Island Hospital, Rhode Island, USA. **78** Butler Hospital, Providence, RI, USA. **79**



Medical University South Carolina, Charleston, SC, USA. **80** St. Joseph's Health Care, Toronto, Canada. **81** Nathan Kline Institute, Orangeburg, SC, USA. **82** University of Iowa College of Medicine, Iowa City, IA, USA. **83** Cornell University, Ithaca, NY, USA. **84** University of South Florida, USF Health Byrd Alzheimer's Institute, Tampa, FL, USA.

## Author Contributions

**Conceptualization:** Batuhan K. Karaman, Elizabeth C. Mormino, Mert R. Sabuncu.

**Data curation:** Batuhan K. Karaman.

**Funding acquisition:** Mert R. Sabuncu.

**Investigation:** Batuhan K. Karaman.

**Methodology:** Batuhan K. Karaman, Mert R. Sabuncu.

**Project administration:** Mert R. Sabuncu.

**Resources:** Mert R. Sabuncu.

**Software:** Batuhan K. Karaman.

**Validation:** Batuhan K. Karaman.

**Visualization:** Batuhan K. Karaman.

**Writing – original draft:** Batuhan K. Karaman, Mert R. Sabuncu.

**Writing – review & editing:** Elizabeth C. Mormino.

## References

1. Organization WH. Dementia; 2021. Available from: <https://www.who.int/news-room/fact-sheets/detail/dementia>.
2. for Disease Control C, Prevention. Leading causes of death; 2022. Available from: <https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>.
3. James BD, Bennett DA. Causes and Patterns of Dementia: An Update in the Era of Redefining Alzheimer's Disease. *Annual Review of Public Health*. 2019; 40(1):65–84. <https://doi.org/10.1146/annurev-publhealth-040218-043758> PMID: 30642228
4. Breijyeh Z, Karaman R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules*. 2020; 25(24). <https://doi.org/10.3390/molecules25245789> PMID: 33302541
5. Munoz DG, Feldman H. Causes of Alzheimer's disease. *CMAJ*. 2000; 162(1):65–72. PMID: 11216203
6. LaMontagne PJ, Benzinger TL, Morris JC, Keefe S, Hornbeck R, Xiong C, et al. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease. *medRxiv*. 2019.
7. Malone IB, Cash D, Ridgway GR, MacManus DG, Ourselin S, Fox NC, et al. MIRIAD—Public release of a multiple time point Alzheimer's MR imaging dataset. *NeuroImage*. 2013; 70:33–36. <https://doi.org/10.1016/j.neuroimage.2012.12.044> PMID: 23274184
8. Birkenbihl C, Westwood S, Shi L, Nevado-Holgado A, Westman E, Lovestone S, et al. ANMerge: A Comprehensive and Accessible Alzheimer's Disease Patient-Level Dataset. *Journal of Alzheimer's Disease*. 2021; 79:423–431. <https://doi.org/10.3233/JAD-200948> PMID: 33285634
9. Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International Psychogeriatrics*. 2009; 21(4):672–687. <https://doi.org/10.1017/S1041610209009405> PMID: 19470201
10. Beekly DL, Ramos EM, van Belle G, Deitrich W, Clark AD, Jacka ME, et al. The National Alzheimer's Coordinating Center (NACC) Database: An Alzheimer Disease Database. *Alzheimer Disease & Associated Disorders*. 2004; 18:270–277. PMID: 15592144
11. Zhang X, Mormino EC, Sun N, Sperling RA, Sabuncu MR, Yeo BT, et al. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proceedings of the*

- National Academy of Sciences. 2016; 113(42):E6535–E6544. <https://doi.org/10.1073/pnas.1611073113> PMID: 27702899
12. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nature Reviews Neurology*. 2018; 14:653–666. <https://doi.org/10.1038/s41582-018-0070-3> PMID: 30291317
  13. Mangialasche F, Solomon A, Winblad B, Mecocci P, Kivipelto M. Alzheimer's disease: clinical trials and drug development. *The Lancet Neurology*. 2010; 9(7):702–716. [https://doi.org/10.1016/S1474-4422\(10\)70119-8](https://doi.org/10.1016/S1474-4422(10)70119-8) PMID: 20610346
  14. Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drug development pipeline: 2019. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*. 2019; 5:272–293. <https://doi.org/10.1016/j.trci.2019.05.008> PMID: 31334330
  15. Rosenberg PB, Mielke MM, Appleby BS, Oh ES, Geda YE, Lyketsos CG. The Association of Neuropsychiatric Symptoms in MCI with Incident Dementia and Alzheimer Disease. *The American Journal of Geriatric Psychiatry*. 2013; 21(7):685–695. <https://doi.org/10.1016/j.jagp.2013.01.006> PMID: 23567400
  16. Feldman H, Scheltens P, Scarpini E, Hermann N, Mesenbrink P, Mancione L, et al. Behavioral symptoms in mild cognitive impairment. *Neurology*. 2004; 62(7):1199–1201. <https://doi.org/10.1212/01.WNL.0000118301.92105.EE> PMID: 15079026
  17. Grueso S, Viejo-Sobera R. Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review. *Alzheimer's Research & Therapy*. 2021; 13. <https://doi.org/10.1186/s13195-021-00900-w> PMID: 34583745
  18. Chen Y, Denny KG, Harvey D, Farias ST, Mungas D, DeCarli C, et al. Progression from normal cognition to mild cognitive impairment in a diverse clinic-based and community-based elderly cohort. *Alzheimer's & Dementia*. 2017; 13:399–405. <https://doi.org/10.1016/j.jalz.2016.07.151>
  19. Peavy GM, Jacobson MW, Salmon DP, Gamst AC, Patterson TL, Goldman S, et al. The Influence of Chronic Stress on Dementia-related Diagnostic Change in Older Adults. *Alzheimer Disease & Associated Disorders*. 2012; 26:260–266. <https://doi.org/10.1097/WAD.0b013e3182389a9c> PMID: 22037597
  20. Popuri K, Balachandar R, Alpert K, Lu D, Bhalla M, Mackenzie IR, et al. Development and validation of a novel dementia of Alzheimer's type (DAT) score based on metabolism FDG-PET imaging. *NeuroImage: Clinical*. 2018; 18:802–813. <https://doi.org/10.1016/j.nicl.2018.03.007> PMID: 29876266
  21. Yee E, Popuri K, Beg MF. Quantifying brain metabolism from FDG–PET images into a probability of Alzheimer's dementia score. *Human Brain Mapping*. 2019. PMID: 31507022
  22. Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage*. 2017; 155:530–548. <https://doi.org/10.1016/j.neuroimage.2017.03.057> PMID: 28414186
  23. Ocasio E, Duong TQ. Deep learning prediction of mild cognitive impairment conversion to Alzheimer's disease at 3 years after diagnosis using longitudinal and whole-brain 3D MRI. *PeerJ Computer Science*. 2021; 7:e560. <https://doi.org/10.7717/peerj-cs.560> PMID: 34141888
  24. Pavisic IM, Nicholas JM, O'Connor A, Rice H, Lu K, Fox NC, et al. Disease duration in autosomal dominant familial Alzheimer disease. *Neurology Genetics*. 2020; 6(5). <https://doi.org/10.1212/NXG.0000000000000507> PMID: 33225064
  25. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, et al. Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's & Dementia*. 2005; 1:55–66. <https://doi.org/10.1016/j.jalz.2005.06.003>
  26. Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, et al. The Preclinical Alzheimer Cognitive Composite. *JAMA Neurology*. 2014; 71:961. <https://doi.org/10.1001/jamaneurol.2014.803> PMID: 24886908
  27. Donohue MC, Sperling RA, Petersen R, Sun CK, Weiner MW, Aisen PS, et al. Association Between Elevated Brain Amyloid and Subsequent Cognitive Decline Among Cognitively Normal Persons. *JAMA*. 2017; 317:2305–2316. <https://doi.org/10.1001/jama.2017.6669> PMID: 28609533
  28. Olsson A, Vanderstichele H, Andreasen N, De Meyer G, Wallin A, Holmberg B, et al. Simultaneous measurement of beta-amyloid(1–42), total tau, and phosphorylated tau (Thr181) in cerebrospinal fluid by the xMAP technology. *Clinical Chemistry*. 2005; 51:336–345. <https://doi.org/10.1373/clinchem.2004.039347> PMID: 15563479
  29. Jellinger KA, Janetzky B, Attems J, Kienzl E. Biomarkers for early diagnosis of Alzheimer disease: 'Alzheimer ASSociated gene'- a new blood biomarker? *Journal of Cellular and Molecular Medicine*. 2008; 12:1094–1117. <https://doi.org/10.1111/j.1582-4934.2008.00313.x> PMID: 18363842
  30. Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*. 2008; 27:685–691. <https://doi.org/10.1002/jmri.21049> PMID: 18302232

31. Jack CR, Barnes J, Bernstein MA, Borowski BJ, Brewer J, Clegg S, et al. Magnetic resonance imaging in Alzheimer's Disease Neuroimaging Initiative 2. *Alzheimer's & Dementia*. 2015; 11:740–756. <https://doi.org/10.1016/j.jalz.2015.05.002> PMID: 26194310
32. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*. 2006; 31(3):968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021> PMID: 16530430
33. Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, et al. Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*. 2004; 14(1):11–22. <https://doi.org/10.1093/cercor/bhg087> PMID: 14654453
34. Fischl B, Salat DH, van der Kouwe AJW, Makris N, Ségonne F, Quinn BT, et al. Sequence-independent segmentation of magnetic resonance images. *NeuroImage*. 2004; 23(Supplement 1):S69–S84. <https://doi.org/10.1016/j.neuroimage.2004.07.016> PMID: 15501102
35. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97(20):11050–11055. <https://doi.org/10.1073/pnas.200033797> PMID: 10984517
36. Fischl B, Liu A, Dale AM. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Medical Imaging*. 2001; 20(1):70–80. <https://doi.org/10.1109/42.906426> PMID: 11293693
37. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002; 33:341–355. [https://doi.org/10.1016/S0896-6273\(02\)00569-X](https://doi.org/10.1016/S0896-6273(02)00569-X) PMID: 11832223
38. Fischl B, Sereno MI, Tootell RBH, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*. 1999; 8(4):272–284. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)8:4<272::AID-HBM10>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4) PMID: 10619420
39. Jovicich J, Czanner S, Greve D, Haley E, van der Kouwe A, Gollub R, et al. Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *NeuroImage*. 2006; 30(2):436–443. <https://doi.org/10.1016/j.neuroimage.2005.09.046> PMID: 16300968
40. Kuperberg GR, Broome M, McGuire PK, David AS, Eddy M, Ozawa F, et al. Regionally localized thinning of the cerebral cortex in Schizophrenia. *Archives of General Psychiatry*. 2003; 60:878–888. <https://doi.org/10.1001/archpsyc.60.9.878> PMID: 12963669
41. Rosas HD, Liu AK, Hersch S, Glessner M, Ferrante RJ, Salat DH, et al. Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology*. 2002; 58(5):695–701. <https://doi.org/10.1212/WNL.58.5.695> PMID: 11889230
42. Salat D, Buckner RL, Snyder AZ, Greve DN, Desikan RS, Busa E, et al. Thinning of the cerebral cortex in aging. *Cerebral Cortex*. 2004; 14:721–730. <https://doi.org/10.1093/cercor/bbh032> PMID: 15054051
43. Segonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, et al. A hybrid approach to the skull stripping problem in MRI. *NeuroImage*. 2004; 22(3):1060–1075. <https://doi.org/10.1016/j.neuroimage.2004.03.032> PMID: 15219578
44. Dale A, Fischl B, Sereno MI. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*. 1999; 9(2):179–194. <https://doi.org/10.1006/nimg.1998.0395> PMID: 9931268
45. Fischl B, Sereno MI, Dale A. Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System. *NeuroImage*. 1999; 9(2):195–207. <https://doi.org/10.1006/nimg.1998.0396> PMID: 9931269
46. Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage*. 2006; 32(1):180–194. <https://doi.org/10.1016/j.neuroimage.2006.02.051> PMID: 16651008
47. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*. 1998; 17:87–97. <https://doi.org/10.1109/42.668698> PMID: 9617910
48. Segonne F, Pacheco J, Fischl B. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Trans Med Imaging*. 2007; 26:518–529. <https://doi.org/10.1109/TMI.2006.887364> PMID: 17427739
49. Reuter M, Rosas HD, Fischl B. Highly Accurate Inverse Consistent Registration: A Robust Approach. *NeuroImage*. 2010; 53(4):1181–1196. <https://doi.org/10.1016/j.neuroimage.2010.07.020> PMID: 20637289
50. Reuter M, Fischl B. Avoiding Asymmetry-Induced Bias in Longitudinal Image Processing. *NeuroImage*. 2011; 57(1):19–21. <https://doi.org/10.1016/j.neuroimage.2011.02.076> PMID: 21376812

51. Reuter M, Schmansky NJ, Rosas HD, Fischl B. Within-Subject Template Estimation for Unbiased Longitudinal Image Analysis. *NeuroImage*. 2012; 61(4):1402–1418. <https://doi.org/10.1016/j.neuroimage.2012.02.084> PMID: 22430496
52. Jagust WJ, Bandy D, Chen K, Foster NL, Landau SM, Mathis CA, et al. The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. *Alzheimer's & Dementia*. 2010; 6:221–229. <https://doi.org/10.1016/j.jalz.2010.03.003>
53. Jagust WJ, Landau SM, Koeppe RA, Reiman EM, Chen K, Mathis CA, et al. The Alzheimer's Disease Neuroimaging Initiative 2 PET Core: 2015. *Alzheimer's & Dementia*. 2015; 11:757–771. <https://doi.org/10.1016/j.jalz.2015.05.001>
54. Hartig M, Truran-Sacrey D, Raptentsetsang S, Simonson A, Mezher A, Schuff N, et al. UCSF FreeSurfer Methods; 2014. Available from: [https://adni.bitbucket.io/reference/docs/UCSFFSX51/UCSF%20FreeSurfer%20Methods%20and%20QC\\_OFFICIAL.pdf](https://adni.bitbucket.io/reference/docs/UCSFFSX51/UCSF%20FreeSurfer%20Methods%20and%20QC_OFFICIAL.pdf).
55. Campos S, Pizarro L, Valle C, Gray KR, Rueckert D, Allende H. Evaluating Imputation Techniques for Missing Data in ADNI: A Patient Classification Study. In: Pardo A, Kittler J, editors. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Cham: Springer International Publishing; 2015. p. 3–10.
56. Wu Y, Zhang X, He Y, Cui J, Ge X, Han H, et al. Predicting Alzheimer's disease based on survival data and longitudinally measured performance on cognitive and functional scales. *Psychiatry Research*. 2020; 291:113201. <https://doi.org/10.1016/j.psychres.2020.113201> PMID: 32559670
57. Li D, Iddi S, Aisen PS, Thompson WK, Donohue MC. The relative efficiency of time-to-progression and continuous measures of cognition in presymptomatic Alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*. 2019; 5:308–318. <https://doi.org/10.1016/j.trci.2019.04.004> PMID: 31367671
58. Lin W, Tong T, Gao Q, Guo D, Du X, Yang Y, et al. Convolutional Neural Networks-Based MRI Image Analysis for the Alzheimer's Disease Prediction From Mild Cognitive Impairment. *Frontiers in Neuroscience*. 2018; 12. <https://doi.org/10.3389/fnins.2018.00777> PMID: 30455622
59. Pagani M, Nobili F, Morbelli S, Arnaldi D, Giuliani A, Öberg J, et al. Early identification of MCI converting to AD: a FDG PET study. *European Journal of Nuclear Medicine and Molecular Imaging*. 2017; 44:2042–2052. <https://doi.org/10.1007/s00259-017-3761-x> PMID: 28664464
60. Nozadi SH, Kadoury S. Classification of Alzheimer's and MCI Patients from Semantically Parcelled PET Images: A Comparison between AV45 and FDG-PET. *International Journal of Biomedical Imaging*. 2018; 2018:1–13. <https://doi.org/10.1155/2018/1247430> PMID: 29736165
61. Mendelson AF, Zuluaga MA, Lorenzi M, Hutton BF, Ourselin S. Selection bias in the reported performances of AD classification pipelines. *NeuroImage: Clinical*. 2017; 14:400–416. <https://doi.org/10.1016/j.nicl.2016.12.018> PMID: 28271040
62. Yuan W, Beaulieu-Jones BK, Yu KH, Lipnick SL, Palmer N, Loscalzo J, et al. Temporal bias in case-control design: preventing reliable predictions of the future. *Nature Communications*. 2021; 12. <https://doi.org/10.1038/s41467-021-21390-2>
63. Shishegar R, Cox T, Rolls D, Bourgeat P, Doré V, Lamb F, et al. Using imputation to provide harmonized longitudinal measures of cognition across AIBL and ADNI. *Scientific Reports*. 2021; 11:23788. <https://doi.org/10.1038/s41598-021-02827-6> PMID: 34893624