Predicting Future Brain Atrophy Based on Longitudinal MRI

Maryam Hadji*†, Elaheh Moradi*, Jussi Tohka*†

[1] A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio 70150, Finland.

*Equal contribution.

†Corresponding authors: Jussi Tohka Email: jussi.tohka@uef.fi, Maryam Hadji Email: maryam.hadji@uef.fi .

Abstract

Neuron loss is a key feature of neurodegenerative diseases often leading to brain atrophy detectable through magnetic resonance imaging (MRI). Various brain atrophy measures are essential in research of Alzheimer' disease (AD) and related dementias. This study aims to forecast future annual percentage changes in hippocampal, ventricular, and total gray matter (TGM) volumes in individuals with varying cognitive statuses, from healthy to dementia. We developed a machine learning model using elastic net linear regression and tested two approaches: (1) a baseline model using predictors from a single-time-point and (2) a longitudinal model using predictors derived from longitudinal MRI. Both approaches were evaluated with MRI-only models and models that combined MRI with additional risk factors (age, sex, APOE4, and baseline diagnosis). Cross-validated Pearson correlation scores between predicted and actual annual percentage changes were 0.62 for the hippocampus, 0.51 for the ventricles, and 0.41 for TGM, using the longitudinal MRI + risk factor model. Longitudinal models consistently outperformed baseline models, and models including risk factors outperformed the MRI only model. Validation using an external dataset confirmed these findings, highlighting the value of predictors derived based on longitudinal data. We

further studied the value of the predicted atrophy/enlargement rates for clinical status progression prediction across three different datasets. Predicted atrophy was a consistently better indicator of progression to mild cognitive impairment and dementia than present-day regional volumes, with the longitudinal atrophy prediction model typically outperforming the baseline model in terms of clinical status prediction. Future atrophy prediction has significant potential for assessing the risk of cognitive decline, even in cognitively unimpaired individuals, and can aid in selecting participants for clinical trials of disease-modifying drugs for AD.

Keywords: Machine learning, MRI, Progression prediction, Dementia, Mild cognitive impairment, Hippocampal atrophy, ventricle enlargement, Total gray matter atrophy

1 Introduction

Dementia impacts approximately 57 million people globally [1], with Alzheimer's disease (AD) being its most common cause [2]. Machine learning approaches for predicting future diagnoses have been extensively studied in the context of dementia [3], with numerous studies focusing on the progression from mild cognitive impairment (MCI) to dementia [4]. However, as the clinical focus shifts from prognosis to selecting the most effective treatments, there is a growing need to predict more nuanced outcomes [5, 6]. This shift in focus could also enhance the selection of study participants for clinical trials, ensuring that the most appropriate candidates are chosen based on their predicted responses to treatment.

Predicting the change in cognitive tests scores, such as mini mental state examination (MMSE) and Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog), is perhaps the most natural way towards these more nuanced outcomes. However, predicting changes in cognitive scores has proven to be challenging, partly due to the validity challenges in the longitudinal measurement of cognitive changes in less impaired populations with the well-established tests [7, 8] and partly due to challenges in assessing cognition in general [9]. These challenges include learning effects, which can skew results

due to repeated testing, and high variability, where outcomes depend significantly on the individual's condition on the test day. For example, in the recent TADPOLE challenge, none of the competing teams was able to perform better than a simple baseline method for the ADAS13 score prediction while the predictions for the future diagnosis and future ventricular volume were significantly better than with the baseline methods [10]. Moreover, while the future cognitive scores may be predictable, this predictive power is many times driven by the current baseline scores, yielding change estimates of limited usability [11]. As a result, there is a pressing need for serial biomarkers that reflect the underlying biology of AD and its connection to dementia.

This work, instead, explores the prediction of future atrophy based on serial magnetic resonance imaging (MRI). The MRI-based brain atrophy has long been identified as an important measure to track dementia progression [12]. It possesses several important characteristics for a prediction target: 1) atrophy measures are quantitative, 2) MRI is widely available and non-invasive, 3) various atrophy measures have been adopted as secondary endpoints in randomized controlled trials of dementia/AD interventions [13, 14] and 4) hippocampal atrophy has been qualified for enrichment in pre-dementia trials by the European Medicines Agency [15]. By focusing on atrophy prediction, this study aims to provide a more reliable and clinically useful tool for both treatment selection and trial participant selection.

We focus on three distinct atrophy measures: hippocampal atrophy, enlarged ventricular volume, and whole-brain atrophy which all have been established as key structural MRI markers for AD and dementia progression, with slightly different time windows of the best predictive utility [16–19]. In clinical trials on dementia and AD, the atrophy of these three key brain regions are often considered as secondary endpoints (e.g. [20]). These regions provide valuable insights into short-term dementia risks and changes in their volumes has been linked with the cognition [21–23].

Surprisingly, while the association of brain atrophy measures to dementia and AD progression is well documented, the prediction of future changes in these measures has not been studied widely. Prediction of the future ventricle volume was one of the tasks in a recent TADPOLE challenge [10], where the best performing model was based on the concept of rate of AD progression [24]. Liedes et al

[25] focused on prediction of hippocampal atrophy with promising results. However, in [25] they studied only prediction of hippocampal atrophy and used data from single time-points only for predictions. In this work, we analyze three regions (hippocampus, total gray matter, and ventricles) together within the same dataset and workflow and develop a modeling framework able to use predictors derived from longitudinal data. Finally, we study the value of forecasted atrophy rates to predict the progression of the level of the memory concern from healthy to MCI/dementia and from MCI to dementia. We demonstrate that the models that use predictors derived from longitudinal data consistently outperform corresponding models based on single time-point predictors. Also, we show that predicted atrophy is a better indicator of progression to mild cognitive impairment and dementia than the volumes of the regions at current day.

2 Materials and Methods

2.1 Data for atrophy/enlargement prediction

Data used in this work were obtained from the ADNI (http://adni.loni.usc.edu) and the AIBL study group (https://aibl.org.au). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see (www.adni-info.org). AIBL study methodology has been reported previously [26].

We included all ADNI participants who had a T1-weighted MRI acquired at the time of entry, 24-month visit, and 48-month visit. Moreover, we required participants to have information on demographics (age, sex), APOE4 status, and diagnosis status obtained from from the ADNIMERGE.csv table. 2,365 participants had MRI data at the time of entry. After 24 months of follow-up, MRI information was available for 1,373 participants, and for 586 participants after 48 months. 565 individuals had MRI data at entry, 24 months, and 48 months. After excluding participants without APOE4 information, we finalized our dataset with 550 individuals who had all the considered information. We refer to this dataset as ADNI-main dataset.

We used the Australian Imaging Biomarkers and Lifestyle Flagship Study of Aging (AIBL, ADNI-subcohort) as a external test data in our study. At the study entry, MRI scans were available for 605 participants. After 18 months of follow-up, 232 individuals had MRI scans, and 149 individuals had scans at the 36-month follow-up. Among them, 95 participants had MRI scans at all three time points. All these 95 individuals have complete demographic data, including age, sex, diagnosis, and APOE4 status. The characteristics of study cohorts are presented in Table 1 and participant's RIDs are available as supplementary material.

Table 1 Characteristics of the ADNI and AIBL cohorts, including sample size, age, sex distribution, APOE4 status, and clinical diagnosis. Age is presented as mean (standard deviation). Diagnosis and age are from the time of entry. CN refers to cognitively normal. APOE4 status is displayed as the count of individuals with 0, 1, or 2 alleles.

Baseline characteristics	**ADNI**	**AIBL**
Sample size, N	550	95
Age, years	73.04 (6.07)	73.31 (6.93)
Sex, M/F	297 / 253	48 / 47
APOE4, 0/1/2	321 / 186 / 43	60 / 29 / 6
Diagnosis, CN/MCI/AD	246 / 236 / 68	76 / 12 / 7

2.2 MRI preprocessing

The MRI data were preprocessed following the framework described in [27]. We used the volumetric and thickness measures computed with CAT12 toolbox (https://neuro-jena.github.io/cat/, version 8.1) using the default settings [28]: For 136 volume measures, the T1-weighted MRI was segmented into gray matter (GM) and white matter (WM) and non-linearly normalized to a stereotactic space using the shooting approach [29]. Based on spatially normalized GM and WM segments, the neuromorphometrics atlas was used to extract the regional volumes[1]. We denote these volume measures by $V_{ij}(t)$ for participant i, region $j = 1, \ldots, 136$, and time point t. The cortical thickness measures [30] were computed and registered to the FSaverage surface template [31], where DK40 atlas [32] was used

[1]The atlas is derived from the maximum probability tissue labels derived from the "MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling" http://www.neuromorphometrics.com/2012_MICCAI_Challenge_Data.htm with the MRIs from OASIS project and the labeled data provided by Neuromorphometrics, Inc. (Neuromorphometrics.com) under academic subscription.

to define 69 regionally averaged cortical thickness measurements. These cortical thickness measures are denoted by $V_{ij}(t)$ for participant i, region $j = 137 \ldots, 206$, and time point t. The MRI data from the participant i at time point t is collected into a vector $V_i(t)$.

The annual percentage change between the time points t_1 and t_2 (expressed in years) for a region i is defined as

$$\Delta_{ij}(t_1, t_2) = \frac{V_{ij}(t_1) - V_{ij}(t_2)}{V_{ij}(t_1)(t_2 - t_1)}. \tag{1}$$

If $\Delta_{ij}(t_1, t_2)$ is positive, it refers to atrophy, and if $\Delta_{ij}(t_1, t_2)$ is negative, it refers to enlargement.

2.3 Prediction framework

Our purpose is to predict future $\Delta_{ik}(t_2, t_3)$ based on the data available during time points t_1 and t_2 where $t_3 > t_2 > t_1$. For simplifying the terminology, we name the time-point t_1 as past, t_2 as current day, and t_3 as future. We formulate this prediction as supervised learning problem. Denote future change $\Delta_{ik}(t_2, t_3)$ by y_{ik} and the number of participants by N. Provided the data $\mathbf{x}_i, i = 1, \ldots, N$ (see Table 2) from the past and current day, we apply the Elastic Net linear regression, a regularization technique that combines both Lasso (L1) and Ridge penalties (L2) [33]. Elastic Net is advantageous in high-dimensional data environments, especially when multicollinearity (correlation among predictors) is present. This method allows us to improve model generalization by reducing over-fitting while selecting a sparse subset of informative features. Dropping the index k, elastic net solves the coefficients $\hat{\beta}$ of the linear model as

$$\hat{\beta} = \min \arg_{\beta} \sum_{i=1}^{n} (y_i - \mathbf{x}_i \beta)^2 + \lambda \left(\alpha \sum_{j=1}^{p} |\beta_j| + (1 - \alpha) \sum_{j=1}^{p} \beta_j^2 \right)$$

where $\beta = [\beta_1, \ldots, \beta_p]^T$, λ controls the overall regularization strength, and α adjusts the balance between L1 and L2 penalties. When $\alpha = 1$, Elastic Net behaves like Lasso regression, driving some coefficients (β_j) to exactly zero, effectively selecting relevant features. Conversely, when $\alpha = 0$, it acts as Ridge regression, reducing the effect of multicollinearity without eliminating predictors. By setting

$0 < \alpha < 1$, Elastic Net captures the advantages of both Lasso and Ridge regularization, yielding a stable and interpretable model even in the presence of correlated predictors.

We used (nested) cross-validation to optimize the parameter λ, ensuring that the selected model provided the best trade-off between model complexity and predictive accuracy. The parameter α was fixed at 0.5 following our previous works [34]. This approach allowed us to obtain a model that is both interpretable and well-suited for generalization to new data.

We performed the prediction for three different regions of interest: 1) hippocampus, defined as the union of left and right hippocampus; 2) ventricles, defined as the union of left and right lateral ventricles and inferior lateral ventricles, and 3) total gray matter (TGM) volume. Hippocampus and ventricles were defined according to the neuromorphometrics atlas (https://neuromorphometrics.com/Seg/html/segmentation/) and the TGM volume was defined as in the CAT12 toolbox [28].

We designed four models based on two approaches with different feature combinations, presented in Table 2. In the baseline approach, in which we used current day data (ADNI 24 month visits) to predict atrophy in the future (ADNI 48 month visits), we defined two models: the MRI only model was trained using only MRI information and the MRI and risk factors model used in addition demographic/risk factor information containing age, sex, the number of APOE4 alleles, and the current day diagnosis (normal cognition, MCI, dementia). We represented this demographic/risk-factor information by $\mathbf{r}_i$ for participant i. The longitudinal approach used past and current day data (ADNI 0 and 24 month visits) to predict atrophy in the future (ADNI 48 month visits), and we again defined MRI only and MRI+ risk factors models. Since the MRIs were acquired both with 1.5T and 3T scanners, we provided this information to the learning algorithm as additional predictors (see [35]), encoding the field-strength combination for participant i as a binary vector $\mathbf{f}_i$.

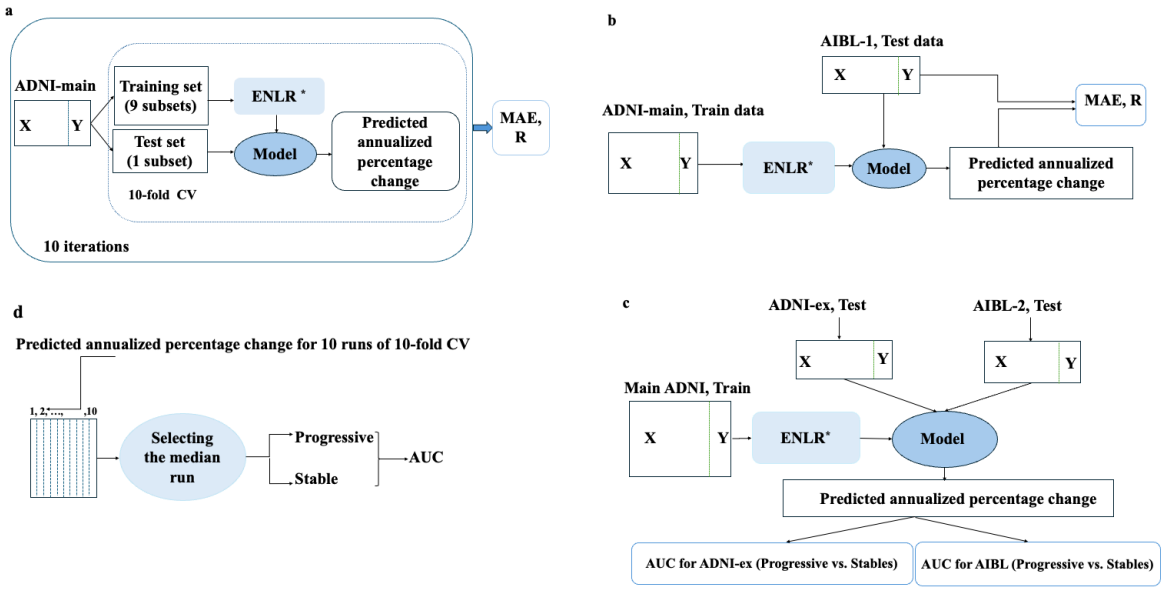

Fig. 1 Schematic representation of the prediction framework. **(a)** Atrophy prediction framework: The model is trained and validated using a 10-fold cross-validation approach on the ADNI-main dataset. Elastic Net Linear Regression (ELNR) is used to predict the annualized percentage change. The performance is evaluated using Mean Absolute Error (MAE) and Pearson correlation (R). **(b)** External validation: The trained model is applied to the AIBL-1 test dataset (Table 1) predict the annualized percentage change, and the performance is assessed using MAE and Pearson correlation (R). **(c)** Progression prediction: The model is trained on the ADNI-main dataset and applied to external test datasets (ADNI-ex and AIBL-2, Table 3) to predict clinical progression. The Area Under the Curve (AUC) is used to evaluate classification performance for progressive vs. stable cases. **(d)** Selection of the representative run: The median run is selected from 10 iterations of 10-fold cross-validation, and participants are classified into progressive or stable groups based on predicted values. AUC is then calculated for the classification task. We use AIBL-1 and AIBL-2 to refer to two different versions of the AIBL test data, described in Tables 1 and 3, respectively.

Table 2 Definitions of the predictive models used in the study. The models vary by the inclusion of MRI data, MRI changes over time, and additional risk factors such as age, sex, diagnosis, and APOE4 status.

Model	**Predictors**	$\mathbf{x}_i$
Baseline, MRI only	Current day MRI	$V_i(t_2), \mathbf{f}_i$
Baseline, MRI+Risk factors	Current day MRI, age, sex, diagnosis, APOE4	$V_i(t_2), \Delta_i(t_1, t_2), \mathbf{f}_i, \mathbf{r}_i$
Longitudinal, MRI only	Current day MRI, MRI change between the past and the current day	$V_i(t_2), \Delta_i(t_1, t_2), \mathbf{f}_i$
Longitudinal, MRI+Risk factors	Current day MRI, MRI change between the past and the current day, age, sex, diagnosis, APOE4	$V_i(t_2), \Delta_i(t_1, t_2), \mathbf{f}_i \mathbf{r}_i$

2.4 Implementation and performance evaluation

Following the framework described in [36], we implemented a 10-fold cross-validation (CV) approach to prevent overfitting and evaluate model performance, see Figure 1. The data was divided into training and test sets using a nested and stratified 10-fold CV framework [37]. This involved two levels of CV: an outer loop for model evaluation and an inner loop for selecting hyperparameters (λ). Participants were randomly divided into 10 subsets. In each fold of the outer loop, one subset was used for testing, and the remaining nine were designated for training. During the inner loop, optimal λ were selected by minimizing the mean square error (MSE). This nested approach ensures robust model evaluation and minimizes overfitting by effectively utilizing the entire dataset for training and testing.

We assessed the models' performance using the cross-validated Pearson correlation coefficient (PCC) and the mean absolute error (MAE) between the predicted and actual rates of cognitive decline expressed in percentage points. To ensure robustness and minimize the influence of random variations, the results were averaged over 10 runs of nested 10-fold CV. To compare the correlation coefficients between the two approaches, we tested the results of a computation run with median PCC (6th largest PCC among 10 runs) using Hittner-modification of the Dunn-Clark test [38, 39] implemented in the CoCor package [40]. We computed the confidence intervals using a bootstrap procedure specifically designed for repeated CV [41]. We considered all the comparisons planned and therefore report the results of all comparisons without adjustments for multiple testing.

Data in the ADNI originated from different phases of ADNI (ADNI1, ADNI2, ADNI3, ADNIGO) with different MRI protocols and scanner field strengths (1.5T or 3.0T). Our baseline approach to correct for differences in MRI markers due to scanner field strength was to include this information in the prediction model as explained in Section 2.3. Additionally, we tested ComBat to harmonize MRI data. ComBat, an empirical Bayes method for batch correction in microarray expression data [42], has been successfully utilized for harmonization in MRI [43]. However, the addition of a harmonization step using the ComBat approach did not improve the prediction performance compared to our baseline approach (Supplementary Table 1).

For ADNI and hippocampal atrophy prediction, we excluded data from two individuals with biologically implausible hippocampal change rate, indicating annual volume increase above 2%. We performed visual quality control on the data from these individuals that indicated poor quality registration or segmentation.

We further utilized the AIBL-cohort as an independent test dataset. The ADNI data served as the training set, while AIBL data was used for testing. The visit schedule of AIBL is different from the ADNI visit schedule: the current day are 18 (AIBL) and 24 (ADNI) month visit, and future are 36 (AIBL) and 48 (ADNI) month visit. However, as the prediction task was formulated as the prediction of annualized change, this causes no differences to the prediction framework presented above. All the analyses were done using R (version 4.1.1), with the following packages: glmnet [44], caret [45], cocor [46], pROC [47], ggplot2 [48], and complexheatmap [49]. The codes to reproduce the work are available at https://github.com/MaryamHadji/Future-prediction

2.5 MCI/dementia progression prediction

Table 3 Demographic and clinical characteristics of cohort used for validation of clinical progression prediction. The cohorts include ADNI-Main, AIBL, and ADNI-External datasets, with subgroups categorized as CN Progressive, CN Stable, MCI Progressive, and MCI Stable. Age is presented as mean (standard deviation). Sex is shown as M/F, and APOE4 status is displayed as the count of individuals with 0, 1, or 2 alleles.

datasets	Group	N	Age, years	Sex, M/F	APOE4 (0/1/2)
ADNI-Main	CN Progressive	60	73.74 (6.63)	32 /28	51 / 8 / 1
	CN Stable	133	74.71 (5.34)	64 /69	85 /46 /2
	MCI Progressive	79	71.90 (7.24)	44 /35	53 /24 /2
	MCI Stable	119	72.64 (7.15)	71 /48	56 /46 /17
AIBL	CN Progressive	13	74.38 (6.56)	10 / 3	4 / 7 / 2
	CN Stable	45	73.51 (7.07)	16 / 29	36 / 8 / 1
ADNI-External	CN Progressive	25	75.42 (5.27)	18 / 7	13 / 12 / 0
	CN Stable	56	69.69 (5.74)	21 / 35	42 / 13 / 1
	MCI Progressive	46	73.81 (6.78)	29 /17	23 / 17 /6
	MCI Stable	43	71.08 (7.56)	24 /19	30 /13 /0

We evaluated the effectiveness of our estimated atrophy/enlargement rates in predicting dementia progression (Figure 1 c&d). This analysis included participants from the CN and MCI groups. We predicted their future atrophy/enlargement rates and assessed how well these estimates predicted

later changes in cognitive status, specifically CN-to-MCI/dementia or MCI-to-dementia progressions. Model's performance was measured using the area under the receiver operating characteristic curve (AUC), and we considered each region of interest (hippocampus, ventricles, and TGM) separately. We divided participants within the CN and MCI groups into two categories based on their diagnosis at the current visit and follow-up data: 1) Stable group: CN/MCI individuals who have remained with the same diagnosis status for five years or more following the current day visit; 2) Progressive group: a) Individuals with CN diagnosis at the current day visit who later develop MCI or dementia (the last two diagnoses must be MCI or dementia), and b) MCI individuals who later develop to dementia (the last two diagnoses must be dementia). We then compared the ability of predicted future atrophy/enlargement rates to MCI/dementia progression prediction against a baseline using present-day brain region volumes. We hypothesized that predicted atrophy/enlargement rates better reflect future MCI/dementia progression than present-day brain region volume.

We used three different datasets for MCI/Dementia progression prediction: First, we used the main ADNI dataset, which we had previously used for atrophy prediction, comprising data from 550 individuals. Among the 246 CN individuals, 60 individuals progressed MCI/dementia during the follow-up period, and 133 remained stable with a follow-up duration of at least five years. Similarly, in the MCI group, 79 individuals converted to dementia, whereas 119 remained with MCI status, see Table 3. In this dataset, atrophy/enlargement rate predictions were generated within a cross-validation framework. For MCI/dementia progression prediction, we selected results from the cross-validation run with the median PCC and used the predicted atrophy/enlargement rates from that run to predict MCI/dementia progression.

Second, we trained our models to predict atrophy/enlargement rates $\hat{\Delta}_k(t_2, t_3)$ using the main ADNI dataset, and applied them to the AIBL cohort. The predicted rates in AIBL were then used to assess MCI/Dementia progression prediction. In the CN group, 13 individuals progressed from CN to MCI/dementia, while 45 remained stable. Due to the limited number of participants in the MCI group ($n = 8$) with available follow-up data extending five years beyond the current visit (i.e., the 18-month assessment), prediction of MCI-to-dementia progression was not conducted in the AIBL cohort. Lastly,

we selected a subset of ADNI participants who were excluded from the main experiments due to missing 48-month data. However, since these individuals had both 0- and 24-month follow-up visits along with sufficient clinical follow-up information, they were eligible for inclusion in the progression prediction analysis. We refer to this subset as the ADNI-external test set. Within this subset, 25 individuals in the CN group progressed to MCI/dementia, while 56 remained stable, but the mean age of these two groups was highly different. In the MCI group, 46 individuals progressed to dementia, whereas 43 remained unchanged in their clinical status, see Table 3.

3 Results

3.1 Prediction of atrophy/enlargement: Cross-validation with ADNI

The validation results of all models are presented in Figure 2. Overall, our findings highlight that the longitudinal approach, particularly when incorporating additional risk factors, yielded the best performance for predicting annual atrophy/enlargement.

Hippocampal atrophy

For the baseline approach utilizing only MRI data, the PCC for predicting the hippocampal atrophy rate was 0.55 (95% CI: 0.49, 0.61), with MAE of 1.36 (95% CI: 1.26, 1.46). Incorporating risk factors into the baseline MRI model, the PCC improved significantly to 0.59 (95% CI: 0.53, 0.65, $p = 0.02$), and the MAE decreased to 1.32 (95% CI: 1.22, 1.41). For the longitudinal approach using the MRI-only model, the PCC was 0.60 (95% CI: 0.54, 0.65), with an MAE of 1.31 (95% CI: 1.22, 1.41). When the longitudinal MRI model was combined with risk factors, the PCC significantly improved to 0.62 (95% CI: 0.56, 0.67, $p = 0.04$), with an estimated MAE of 1.29 (95% CI: 1.20, 1.38). In the MRI-only model, the PCC for the baseline approach significantly improved from 0.55 (95% CI: 0.49, 0.61) to 0.60 (95% CI: 0.54, 0.65) for the longitudinal approach ($p = 0.02$). Similarly, in the model incorporating both MRI and risk factors, the PCC increased from 0.59 (95% CI: 0.53, 0.65) in the baseline approach to 0.62 (95% CI: 0.56, 0.67) in the longitudinal approach, representing a significant improvement ($p = 0.03$).

Ventricle enlargement

For predicting ventricle enlargement rate using the baseline approach relying solely on MRI data, the PCC was 0.40 (95% CI: 0.32, 0.48), with an MAE of 2.67 (95% CI: 2.48, 2.88). Incorporating risk factors into the baseline MRI model improved the PCC significantly to 0.46 (95% CI: 0.37, 0.54, p = 0.004), while reducing the MAE to 2.57 (95% CI: 2.38, 2.79). For the longitudinal approach using the MRI-only model, PCC increased to 0.49 (95% CI: 0.39, 0.58), and the MAE decreased to 2.52 (95% CI: 2.33, 2.72). The best performance was achieved with the longitudinal approach combining the MRI model and additional risk factors, yielding the PCC 0.51 (95% CI: 0.41, 0.59) and the MAE of 2.48 (95% CI: 2.29, 2.69), however, this improvement was not significant (p = 0.4) over the longitudinal MRI-only model. In order to compare the models using only MRI information for ventricle enlargement rate, the PCC for the baseline approach was 0.40 (95% CI: 0.32, 0.48). For the longitudinal approach, the PCC increased to 0.49 (95% CI: 0.39, 0.58), representing a significant improvement (p = 0.004). For the model incorporating both MRI and risk factor information, there was no significant improvement, with the PCC increasing from 0.46 (95% CI: 0.37, 0.54) in the baseline approach to 0.51 (95% CI: 0.41, 0.59) in the longitudinal approach (p = 0.1).

Gray matter atrophy

To predict the TGM atrophy, based on baseline approach utilizing only MRI data, the PCC was 0.28 (95% CI: 0.20, 0.35), with a MAE of 0.90 (95% CI: 0.83, 0.97). Incorporating risk factors into the MRI model improved the PCC to 0.32 (95% CI: 0.24, 0.39) and slightly reduced the MAE to 0.89 (95% CI: 0.82, 0.95), the difference between the two models was not significant (p = 0.1). For the longitudinal approach using only the MRI model, the PCC increased further to 0.40 (95% CI: 0.33, 0.47), with the MAE of 0.86 (95% CI: 0.80, 0.93). The longitudinal MRI+Risk factors model achieved a PCC of 0.41 (95% CI: 0.34, 0.48) and an MAE of 0.86 (95% CI: 0.80, 0.93). The improvement over longitudinal MRI-only model was not significant (p = 0.6). In the MRI-only model, the PCC was 0.28 (95% CI: 0.20, 0.35) for the baseline approach and increased to 0.40 (95% CI: 0.33, 0.47) for the longitudinal approach, demonstrating a significant improvement (p= 0.0009). Furthermore, after incorporating risk

factors alongside MRI information, we observed a significant improvement from 0.32 (95% CI: 0.24, 0.39) in the baseline approach to 0.41 (95% CI: 0.34, 0.48) in the longitudinal approach ($p = 0.006$).

Predictor importance

The most important predictors for atrophy/enlargement rates in the longitudinal MRI+risk factor model are displayed in Fig. 3. For predicting hippocampal atrophy, the five most important predictors were dementia diagnosis, right hippocampal volume, and changes in volumes of left middle temporal gyrus, right hippocampus and left cerebral white matter. Both present day volumes and changes in volumes of various brain structures contributed to models. However, only present day cortical thickness values (and not the rate of change in cortical thickness) were included as top predictors. From risk factors, only dementia diagnosis and the number of APOE4 alleles were included in the top predictors. Unsurprisingly, only the left or right (but not both) structure was included in the top predictors, with the exception of hippocampus volume. For predicting ventricle enlargement, the five most important predictors were changes in volumes of left lateral ventricle, left middle temporal gyrus, and left parahippocampal gyrus, dementia diagnosis, and the number of APOE4 alleles. The change rates of volumes of various structures appeared more often in the top predictors than the present day volumes of structures. Few cortical thickness measures were included as well as change rate of left rostral middle frontal cortex. Again, from risk factors, only dementia diagnosis and the number of APOE4 alleles were included in the top predictors. For TGM atrophy, the five most important predictors were dementia diagnosis, change rates of right parahippocampal gyrus volume, left transverse temporal cortical thickness, right cerebellum white matter volume, and left isthmus cortical thickness. Again, the change rates of volumes of various structures appeared more often in the top predictors than the present day volumes of structures. Both changes of and present day cortical thicknesses were included as top 20 predictors. The only risk factor in the top predictor list was the dementia diagnosis. For TGM, one of variables coding the MRI field strength entered to the model (strength3, indicating that all the measurements were done with 3.0T scanner).

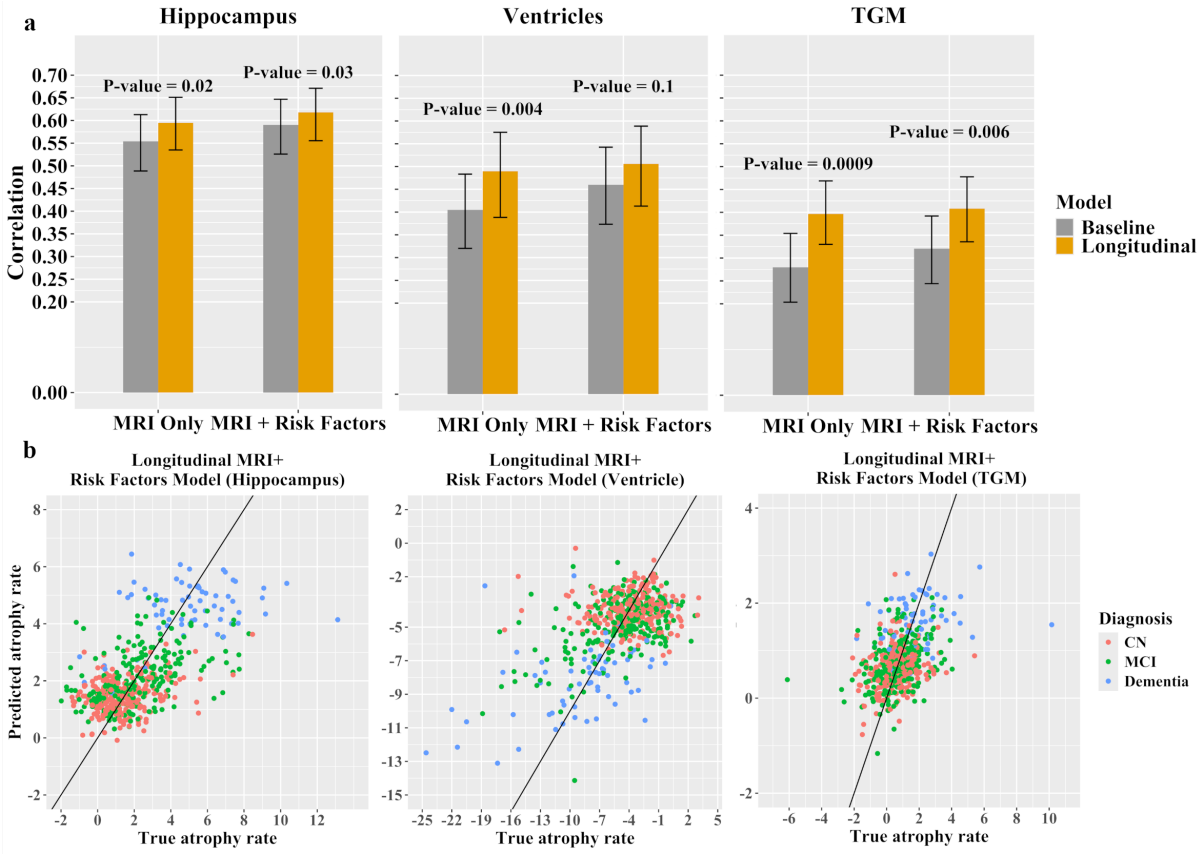

Fig. 2 Prediction of annual percentage change of volumes of different brain regions using 10-fold cross-validation on the main ADNI dataset. **(a)** Bar plots showing the correlation of different models based on baseline and longitudinal approaches. **(b)** Scatter plot illustrating the prediction of atrophy rates for the longitudinal approach of MRI + Risk Factors model, derived from a single computation run with median performance.

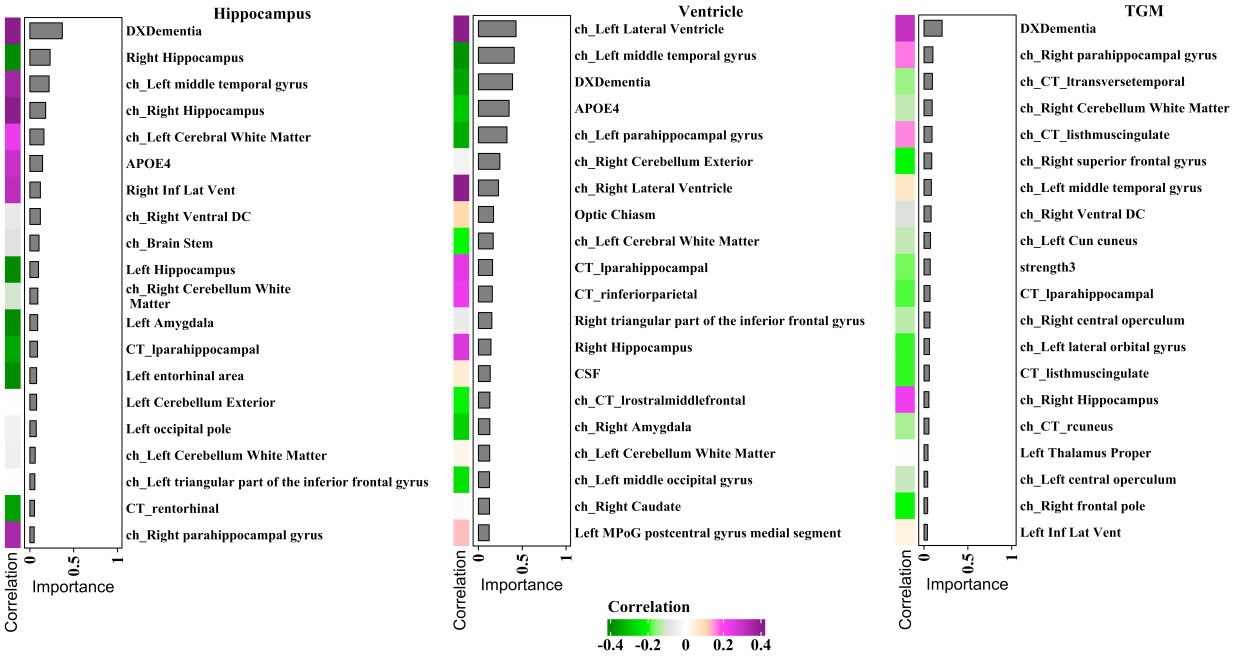

Fig. 3 Importance of predictors across 10 runs of 10-fold CV for the longitudinal, MRI+Risk factors model. Each heatmap consists of a separate single-column heatmap showing the correlation scores between each variable and the target label. Additionally, a bar graph illustrates the importance of each predictor, calculated as the mean of absolute values of regression coefficients across the cross-validation folds. Note that the regression coefficients are computed for standardized models. For brain imaging predictors, "ch" indicates change and "CT" indicates cortical thickness. Predictor names without "CT" refer to volumes. "DX" refers to the diagnosis at 24 months, defined across three columns: "DX-CN", "DX-Dementia", and "DX-MCI". Strength refers to variable coding of the field strengths of the scanners used to acquire the scans.

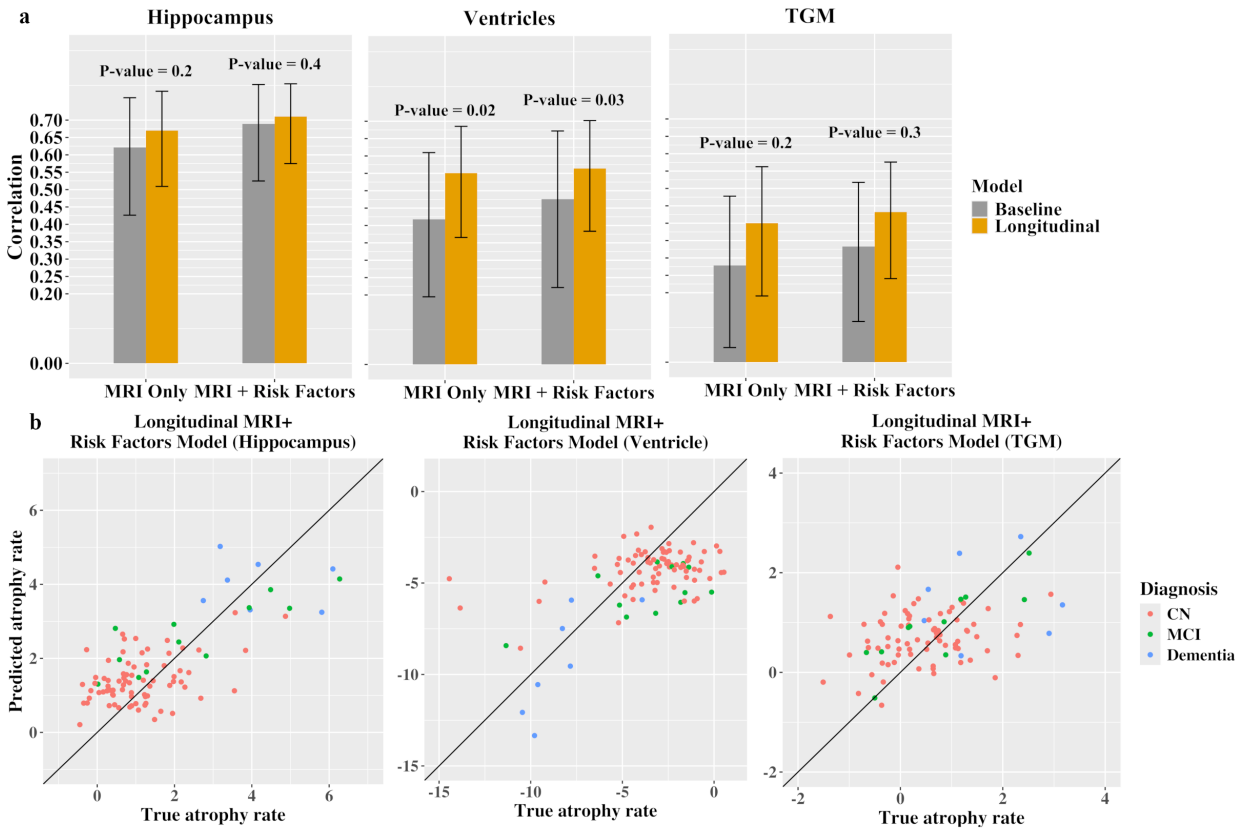

Fig. 4 Prediction of annual percentage change of volumes in different brain regions on the AIBL (test) dataset. **(a)** Bar plots showing the correlation scores of different models based on baseline and longitudinal approaches. **(b)** Scatter plots illustrating the prediction of atrophy rates for the longitudinal approach of MRI + Risk Factors model, derived from a single computation run with median performance.

3.2 Prediction of atrophy/enlargement: External evaluation with AIBL

Figure 4 presents the external evaluation results for atrophy/enlargement prediction using the AIBL dataset. These results demonstrate that the longitudinal approach outperforms the baseline approach when tested on an external dataset.

The external evaluation for hippocampal atrophy prediction using the baseline approach with the MRI only model resulted in a PCC of 0.62 (95% CI: 0.43, 0.76) and a MAE of 1.00 (95% CI: 0.87,

1.15). When applying the longitudinal approach and incorporating risk factors alongside MRI data, the PCC improved to 0.71 (95% CI: 0.57, 0.80), while the MAE decreased to 0.89 (95% CI: 0.78,1.01).

The external evaluation for ventricle enlargement prediction using the baseline approach, which relied solely on MRI data, resulted in the PCC of 0.42 (95% CI: 0.19, 0.61) and the MAE of 2.40 (95% CI: 2.05, 2.79). In comparison, the longitudinal approach that combined MRI data with risk factor information achieved a higher PCC of 0.56 (95% CI: 0.38, 0.70) and a lower MAE of 2.14 (95% CI: 1.83, 2.46).

The external evaluation for TGM atrophy prediction using the baseline approach, which relied solely on MRI data, resulted in the PCC of 0.28 (95% CI: 0.04, 0.48) and the MAE of 0.82 (95% CI: 0.69,0.96). In comparison, the longitudinal approach that combined MRI data with risk factor information achieved a higher PCC of 0.43 (95% CI: 0.24, 0.58) and a slightly lower MAE of 0.78(95% CI: 0.66,0.90).

3.3 Prediction of MCI/dementia progression

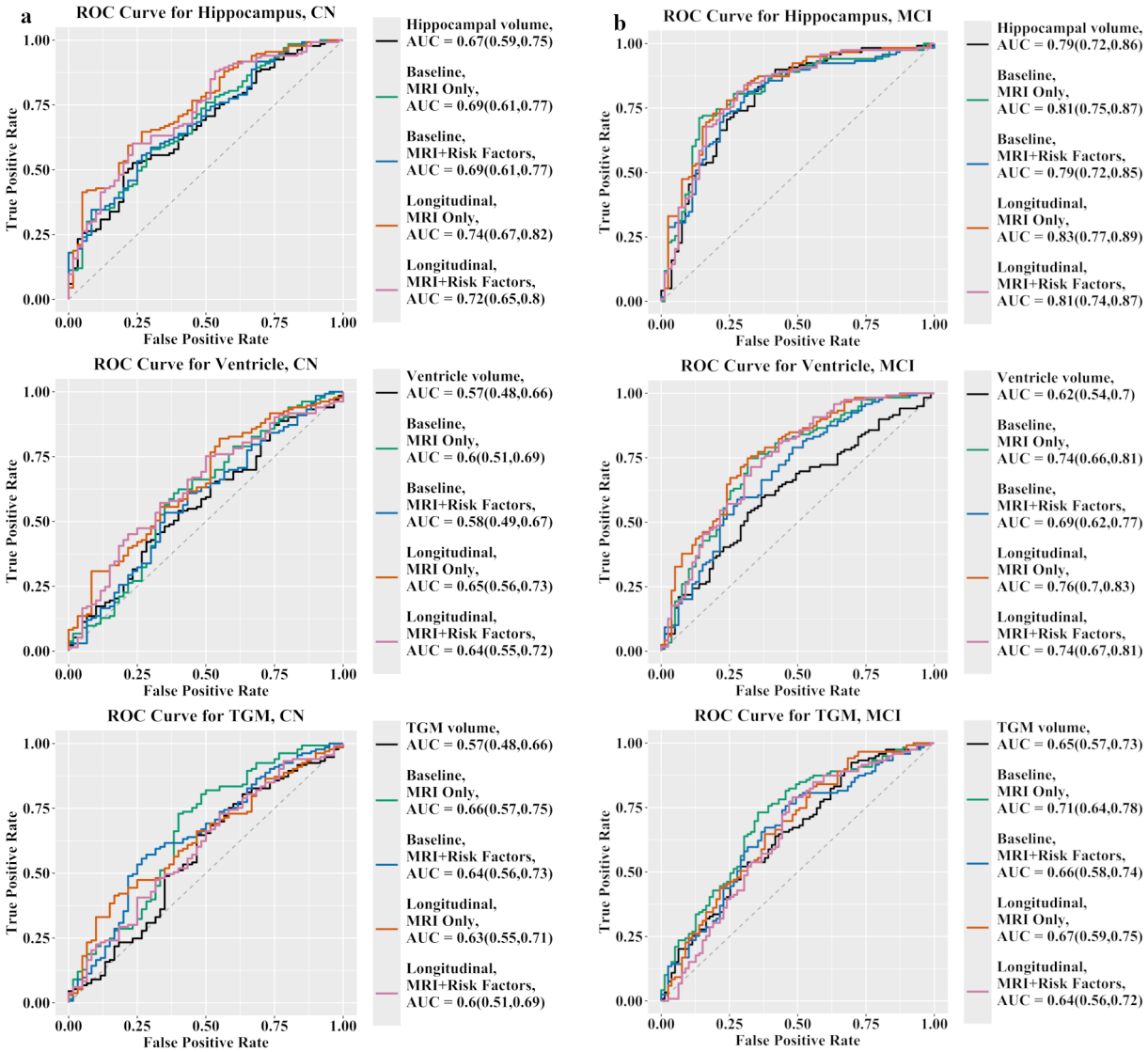

Fig. 5 ROC curves of clinical status progression prediction in **a** CN and **b** MCI groups with different models and brain regions using ADNI-Main dataset.

The ROC curves assessing performance of the predicted atrophy/enlargement rates in clinical status progression prediction are presented in Figure 5 for ADNI main dataset and in Supplementary Figures

1 and 2 for AIBL and ADNI external datasets. All AUC statistics are collected in Table 4. Overall, the clinical status progression prediction results highlight that a) predicted atrophy yielded better than chance progression predictions and the predicted atrophy scores typically reached higher AUCs than the corresponding present day volume information.

ADNI-Main Cohort

For the ADNI-Main cohort, using present day hippocampus volume achieved an AUC of 0.67 (95% CI: 0.59, 0.75) for the CN-to-MCI progression prediction. The predicted hippocampal atrophy (with longitudinal MRI + risk factors model) resulted in the improved progression prediction performance, yielding an AUC of 0.72 (95% CI: 0.65, 0.80). However, this improvement was not statistically significant ($p = 0.10$, DeLong test [50]). Similarly, the present day ventricle volume produced an AUC of 0.57 for CN-to-MCI progression prediction (95% CI:0.48, 0.66), which increased to 0.64 (95% CI:0.55,0.72) with the predicted ventricle enlargement (longitudinal MRI + risk factors model). This improvement was not statistically significant ($p = 0.30$, DeLong test). The present day TGM volume predicted CN-to-MCI progression with an AUC of 0.57 (95% CI: 0.48, 0.66), which increased to 0.60 (95% CI: 0.51, 0.69) with the predicted TGM atrophy (longitudinal MRI +risk factors). This increment was not statistically significant ($p = 0.60$).

Within MCI group, present day hippocampus volume predicted progression to dementia with an AUC of 0.79 (95% CI: 0.72, 0.86). Predicted atrophy (longitudinal MRI + risk factors model) did not significantly improve the performance ($p = 0.50$), yielding an AUC of 0.81 (95% CI: 0.74, 0.87). For ventricle volume, the AUC was 0.62 (95% CI: 0.54, 0.70), which significantly improved to 0.74 (95% CI: 0.67, 0.81) when predicting progression to dementia using predicted atrophy (longitudinal MRI + risk factors model, $p = 0.04$). For the TGM volume, the AUC for the TGM volume was 0.65 (95% CI: 0.57, 0.73) and predicted atrophy for longitudinal MRI + risk factors model yielded a similar AUC of 0.64 (95% CI:0.56, 0.72) ($p=0.9$).

Table 4 AUCs for different models for clinical status progression prediction across three databases. The values in parentheses refer to 95% confidence intervals of AUC.

Brain Regions	**Models**	**Datasets**				
		ADNI-main		**AIBL**	**ADNI-external**	
		CN-to-MCI	**MCI-to-Dementia**	**CN-to-MCI**	**CN-to-MCI**	**MCI-to-Dementia**
Hippocampus	Hippocampal volume	0.67 (0.59, 0.75)	0.79 (0.72, 0.86)	0.62 (0.42, 0.82)	0.69 (0.56, 0.82)	0.78 (0.69, 0.88)
	Baseline, MRI only	0.69 (0.61, 0.77)	0.81 (0.75, 0.87)	0.77 (0.62, 0.92)	0.73 (0.61, 0.84)	0.73 (0.62, 0.83)
	Baseline, MRI + Risk factors	0.69 (0.61, 0.77)	0.79 (0.72, 0.85)	0.84 (0.72, 0.96)	0.81 (0.71, 0.90)	0.83 (0.75, 0.91)
	Longitudinal, MRI only	0.74 (0.67, 0.82)	0.83 (0.77, 0.89)	0.84 (0.72, 0.95)	0.83 (0.72, 0.93)	0.85 (0.77, 0.93)
	Longitudinal, MRI + Risk factors	0.72 (0.65, 0.80)	0.81 (0.74, 0.87)	0.87 (0.76, 0.98)	0.83 (0.72, 0.93)	0.86 (0.79, 0.94)
Ventricles	Ventricle volume	0.57 (0.48, 0.66)	0.62 (0.54, 0.70)	0.65 (0.49, 0.81)	0.77 (0.66, 0.88)	0.61 (0.49, 0.73)
	Baseline, MRI only	0.60 (0.51, 0.69)	0.74 (0.66, 0.81)	0.66 (0.46, 0.87)	0.59 (0.45, 0.73)	0.70 (0.59, 0.81)
	Baseline, MRI + Risk factors	0.58 (0.49, 0.67)	0.69 (0.62, 0.77)	0.70 (0.49, 0.92)	0.56 (0.42, 0.70)	0.65 (0.54, 0.77)
	Longitudinal, MRI only	0.65 (0.56, 0.73)	0.76 (0.70, 0.83)	0.82 (0.66, 0.97)	0.69 (0.56, 0.81)	0.76 (0.66, 0.86)
	Longitudinal, MRI + Risk factors	0.64 (0.55, 0.72)	0.74 (0.67, 0.83)	0.83 (0.67, 0.99)	0.67 (0.54, 0.80)	0.75 (0.64, 0.85)
TGM	TGM volume	0.57 (0.48, 0.66)	0.65 (0.57, 0.73)	0.56 (0.34, 0.77)	0.49 (0.35, 0.62)	0.58 (0.46, 0.70)
	Baseline, MRI only	0.66 (0.57, 0.75)	0.71 (0.64, 0.78)	0.74 (0.57, 0.91)	0.57 (0.43, 0.71)	0.55 (0.42, 0.67)
	Baseline, MRI + Risk factors	0.64 (0.56, 0.73)	0.66 (0.58, 0.74)	0.74 (0.58, 0.90)	0.48 (0.34, 0.62)	0.56 (0.44, 0.69)
	Longitudinal, MRI only	0.63 (0.55, 0.71)	0.67 (0.59, 0.75)	0.74 (0.57, 0.90)	0.62 (0.50, 0.75)	0.55 (0.43, 0.67)
	Longitudinal, MRI + Risk factors	0.60 (0.51, 0.69)	0.64 (0.56, 0.72)	0.69 (0.51, 0.88)	0.60 (0.47, 0.73)	0.59 (0.46, 0.71)

The comparison between different atrophy prediction models (baseline vs. longitudinal, MRI only vs. MRI + risk factors) did not yield any clearly best model, but all the models performed similarly in the progression prediction.

ADNI-External Cohort

For the ADNI-external cohort among the CN group, present day hippocampus volume achieved an AUC of 0.69 (95% CI: 0.56, 0.82) for the CN-to-MCI progression prediction. The predicted hippocampal atrophy (with longitudinal MRI + risk factors model) resulted in the increased progression prediction performance, yielding an AUC of 0.83 (95% CI: 0.72,0.93). This improvement was statistically significant ($p = 0.04$, DeLong test). Similarly, the present day ventricle volume produced an AUC of 0.77 for CN-to-MCI progression prediction (95% CI:0.66, 0.88), which reduced to 0.67 (95% CI:0.54,0.80) with the predicted ventricle enlargement (longitudinal MRI + risk factors model). This reduction was not statistically significant ($p = 0.2$, DeLong test) and we hypothesize that the high AUC for ventricle volume based prediction was driven by a large age difference between the stable and progressive groups (see Table 3). The present day TGM volume predicted CN-to-MCI progression with an AUC of 0.49 (95% CI: 0.35, 0.62), which increased to 0.60 (95% CI: 0.47, 0.73) with the predicted TGM atrophy (longitudinal MRI +risk factors). This increment was not statistically significant ($p = 0.2$).

For the MCI group in the ADNI-external, the model using hippocampal volume alone achieved an AUC of 0.78 (95% CI: 0.69, 0.88). With the predicted hippocampal atrophy (longitudinal MRI + risk factors model) the progression prediction performance improved significantly, yielding an AUC of 0.86 (95% CI: 0.79, 0.94) ($p=0.03$). The ventricle volume model achieved an AUC of 0.61 (95% CI: 0.49,0.73), which increased to 0.75 (95% CI: 0.64,0.85) with the predicted ventricle enlargement (longitudinal MRI + risk factors mode), but the increase was not significant ($p = 0.1$). The model using TGM volume resulted in an AUC of 0.58 (95% CI: 0.46,0.70), while the predicted TGM atrophy (longitudinal MRI + risk factors model) resulted to a similar AUC to 0.59 (95% CI: 0.46, 0.71) ($p = 0.9$).

Notably, the AUCs for clinical status progression prediction were consistently higher for atrophy scores predicted based on longitudinal model than for the atrophy scores based on baseline model. Addition of risk factors to the atrophy score prediction did not have much influence on the progression prediction, with the exception of the hippocampus and the baseline model, where atrophy scores predicted based on MRI + risk factors model provided better progression predictions than those based on MRI only model.

AIBL Cohort

Using hippocampus volume, we could predict the progression from CN to MCI with an AUC of 0.62 (95% CI: 0.42, 0.82). Using predicted atrophy scores significantly improved the prediction performance, resulting in an AUC of 0.87 (95% CI: 0.76, 0.98) (p=0.02), highlighting the enhanced predictive value of this approach. The ventricle volume achieved an AUC of 0.65 (95% CI: 0.49, 0.81) while the AUC with predicted ventricle enlargement improved to 0.83 (95% CI: 0.67, 0.99), however, the improvement was not significant (p = 0.1). Using TGM volume, the AUC was 0.56 (95% CI: 0.34, 0.77) while the AUC with predicted TGM was increased to 0.69 (95% CI: 0.51, 0.88) (p = 0.3).

Here, the longitudinal MRI + risk factors model was clearly superior to the baseline MRI only model for progression prediction using predicted hippocampal atrophy and ventricle enlargement. With TGM, all the four models performed similarly in terms of progression prediction.

4 Discussion

Hippocampal atrophy, ventricle enlargement, and whole-brain atrophy are three key MRI markers used to predict AD/dementia [16, 51, 52]. This study aimed to predict the annual change in the hippocampus, ventricles, and total gray matter volumes (used as a proxy for whole-brain atrophy), in individuals with varying levels of cognitive impairment, ranging from healthy to dementia. To achieve this, we developed an ML model based on ENLR. We explored two approaches: (1) a baseline method relying on single-time-point data and (2) a longitudinal method incorporating information from two years prior to baseline. Both approaches were tested using models that utilized only MRI data and models

that combined MRI data with additional risk factors, including age, sex, and baseline diagnosis. Our analysis also explored how incorporating longitudinal data affected the performance of these models.

The predictions for all three MRI markers were clearly better than the baseline prediction of no chance that would have the PCC of zero. Based on the PCC values, the hippocampal atrophy could be best predicted, followed by the ventricle enlargement, while the TGM atrophy was the most challenging to predict. Indeed, with the external test data (AIBL) hippocampal atrophy prediction had PCC of 0.71 while TGM atrophy prediction only reached PCC of 0.43.

Our findings showed that incorporating additional information into the baseline MRI model significantly enhanced its ability to predict hippocampal atrophy and ventricular enlargement. In our previous works, we have demonstrated the advantages of combining multiple data modalities—such as MRI and cognitive assessments — for prediction tasks relevant to dementia management[27, 36, 53]. Also, other studies have emphasized the importance of integrating diverse data sources for enhanced predictive accuracy[25, 54, 55]. Our current results align with these findings, but with a notable difference in the methodology. Instead of using baseline cognitive assessments as predictors, we utilized baseline diagnoses as predictors. Baseline diagnosis serves as a summary of cognitive assessments, encapsulating an individual's cognitive status. Using a single summary variable reduced data dimensionality and redundancy, as many cognitive measures convey overlapping information.

Incorporating data from two years prior to the current visit into a longitudinal model consistently enhanced predictive performance even if the performance improvement was in relatively modest especially in hippocampal atrophy prediction. With ADNI, the improvement was statistically significant when comparing baseline versus longitudinal using MRI-only models across all cases, including hippocampal atrophy, TGM atrophy, and ventricular enlargement. For MRI + risk factor models, adding historical data also improved performance, but the improvement was smaller compared to the MRI-only models. The improvement remained statistically significant for hippocampal and TGM atrophy but not for ventricular enlargement. Overall, the best predictions were achieved using the longitudinal MRI + risk factor model across all three brain regions. External validation using the AIBL dataset confirmed these findings, demonstrating that longitudinal models consistently provided more accurate

predictions for all three types of annual change. Our findings extend those of earlier studies that have shown the importance of considering longitudinal data in prediction tasks based on brain imaging data [56, 57]. We speculate that the greater improvement observed in the MRI-only model compared to the MRI + risk factor model may be due to the absence of cognitive status information in MRI-only models. Adding historical data to baseline MRI data provides additional insights into individual structural changes over time, as the differences between past and baseline MRI scans vary across diagnostic groups (CN, MCI, AD). This temporal information helps the model differentiate individuals more effectively. In contrast, the MRI + risk factor model already includes diagnostic group information, which partially accounts for cognitive status, thereby reducing the relative contribution of historical data.

In longitudinal models, we investigated the importance of various predictors, including risk factors, MRI features and their annualized percentage changes between baseline and two years prior. A closer examination of the coefficient values in the longitudinal model revealed that baseline diagnosis is one of the most important features for predicting the annual percentage change in all three studied markers. Among the top selected features, APOE4 emerged as a significant predictor for hippocampal atrophy and ventricular enlargement, but not for TGM atrophy. Interestingly, many longitudinal MRI features were identified as important within the top selected features, emphasizing the value of incorporating longitudinal information into the models. Another notable observation is that, for predicting the annual percentage change in each biomarker, multiple brain regions contribute to the prediction task, rather than the respective region alone holding the highest importance. Among the top selected MRI features for predicting hippocampal atrophy and ventricular enlargement were the hippocampus, middle temporal gyrus, inferior lateral ventricles, and amygdala, all of which are highly relevant in AD progression [51]. However, the top selected features for predicting TGM atrophy differed slightly from those for hippocampal and ventricular predictions and were not as strongly associated with AD. These results align with our expectations, as the hippocampus and ventricles are the primary brain regions affected by AD, even in the early stages of disease. In contrast, TGM represents the whole brain and

exhibits a weaker association with AD compared to the hippocampus and ventricles. Overall, our findings indicate that multiple brain regions, along with their longitudinal information and associated risk factors, play a crucial role in predicting the annual percentage change in all these biomarkers.

The MRI scans in the ADNI dataset were drawn from different ADNI phases, and MRI data was acquired using different field strengths. This variability posed a challenge [58, 59], particularly as our analysis considered multiple time points for each participant, two time points (current day and future) for baseline models, and three time points (past, current day, and future) for longitudinal models. For many participants, the field strength was not consistent across all time points. To address this issue, we introduced a new categorical variable representing the field strength at each time point and included it as a predictor in our models (see [35]). By doing so, we accounted for the variability in field strength and ensured it was considered in the analysis, reducing its potential impact on the model's performance. To streamline our analysis, we opted for a cross-sectional pipeline for MRI preprocessing using CAT12, prioritizing its simplicity and efficiency in reducing processing time. Our preliminary analyses revealed no significant performance differences between the cross-sectional and longitudinal preprocessing approaches for atrophy prediction in CAT12. Consequently, the cross-sectional pipeline was selected as a practical and effective choice for our study and the finding that longitudinal predictors outperformed cross-sectional ones supports this choice. However, the optimal approach for addressing the construction of the longitudinal predictors should be addressed systematically in future studies.

Although hippocampal and TGM atrophy, as well as ventricular enlargement, are critical factors in dementia progression, predicting volume changes in these regions has not been extensively studied. In particular, limited research has evaluated all three regions within the same dataset and pipeline to assess their roles in dementia progression. While various studies have investigated the associations of these biomarkers with progression to dementia [52, 60, 61], the prediction of changes over time and their relationship with progression to MCI or dementia remains underexplored. The prediction of total ventricular volume has been previously examined in the TADPOLE challenge [10], where the tasks were predicting clinical diagnosis, ADAS-Cog13 scores, and total ventricular volume. Another study

by Liedes et al. [25] focused on predicting the annualized rate of hippocampal atrophy using multivariate ML algorithms on the ADNI dataset. Our study shares similarities with Liedes et al [25] in predicting annualized hippocampal atrophy but differs in three key ways: (1) we analyzed all three regions—hippocampus, TGM, and ventricles—together within the same dataset and workflow, rather than focusing solely on hippocampal atrophy; (2) we developed a longitudinal model that incorporates data from prior time points, enabling us to investigate the impact of longitudinal predictors on the performance of our predictive models; and (3) we evaluated the ability of our models to predict progression to MCI or dementia in both healthy and MCI groups.

We assessed the effectiveness of our predictive models in estimating the annual percentage changes in hippocampal, TGM, and ventricular volumes to predict dementia progression in both healthy and MCI groups. Results from the main ADNI dataset showed that predicted hippocampal atrophy and ventricular enlargement outperformed actual biomarker volumes in predicting progression to MCI or dementia in both CN and MCI groups. The most notable improvements came from predicted hippocampal atrophy, particularly using longitudinal approaches. For CN individuals, the predictive accuracy for progression to MCI/dementia improved from an AUC of 0.67 (using actual hippocampal volume) to 0.73 (using predicted atrophy with the longitudinal method). Similarly, in MCI individuals, the AUC for predicting progression to dementia increased from 0.79 to 0.83. However, no such improvement was observed for TGM atrophy, as predicted TGM atrophy did not provide better predictions compared to actual TGM volume and AUCs remained mostly below 0.70. Results from the external validation datasets, (AIBL and ADNI-external) showed similar trends. While our models were not specifically designed to predict progression to MCI or dementia, the predicted annual percentage changes—especially for hippocampal atrophy—performed well and the performance was comparable to previous MRI-based approaches for predicting dementia progression in CN individuals [62, 63]. This analysis demonstrates that predicted future atrophy scores can function as single valued and easily interpretable indexes of potential future cognitive decline. Also, they might be applicable when selecting/stratifying population for clinical trials, particularly when combined with disease-specific biomarkers.

In summary, our study highlights the importance of incorporating longitudinal data to improve predictions of annual changes in key biomarkers linked to dementia. By integrating MRI data and risk factors into machine learning models, we significantly enhanced predictive accuracy, particularly for hippocampal atrophy and ventricular enlargement. The inclusion of longitudinal data allowed us to capture individual structural changes over time, improving the model's ability to differentiate between diagnostic groups. These findings emphasize the advantages of combining diverse data sources and tracks brain changes over time to better predict dementia progression. Future research should further explore the integration of multiple data modalities and assess the long-term impact of longitudinal data to refine predictive models for dementia.

Availability of data and materials

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu/) and and the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) (https://aibl.csiro.au/adni/index.html). Details about data access are detailed there. The participant's RIDs are provided in Supplementary Materials. The codes to reproduce the work are available at https://github.com/MaryamHadji/Future-prediction. CAT12 toolbox is available at https://neuro-jena.github.io/cat.

Competing interests

The authors have no actual or potential conflicts of interest.

Funding. This research has been supported by The Academy of Finland, grant 351849 under the frame of ERA PerMed ("Pattern-Cog"), grants 346934 (PRIMAL), and 358944 (Flagship of Advanced Mathematics for Sensing Imaging and Modeling) from the Research Council of Finland.

Author Contributions

M.H: Methodology, formal analysis, investigation, software, visualization, writing—original draft, and writing—review and editing; E.M: Methodology, formal analysis, software, writing—original draft,

and writing—review and editing; J.T: Conceptualization, methodology, resources, funding acquisition, writing—original draft, and writing—review and editing.

Acknowledgments

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.adni.loni.usc.edu) and the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), which was made available at the ADNI database. The investigators within ADNI and AIBL contributed to the design and implementation of ADNI/AIBL 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. AIBL researchers are listed at www.aibl.csiro.au.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is

coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

During the preparation of this work, the authors used GPT-4 from OpenAI and Microsoft Copilot to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed. The authors take full responsibility for the content of the publication.

The computational analyses were partly performed on servers provided by the UEF Bioinformatics Center and Biocenter Kuopio, Biocenter Finland, University of Eastern Finland, Finland.

References

[1] Nichols, E., Vos, T.: The estimation of the global prevalence of dementia from 1990-2019 and forecasted prevalence through 2050: an analysis for the global burden of disease (gbd) study 2019. Alzheimer's & Dementia **17**, 051496 (2021)

[2] Lanctôt, K.L., Hahn-Pedersen, J.H., Eichinger, C., Freeman, C., Clark, A., Tarazona, L., Cummings, J.: Burden of illness in people with alzheimer's disease: a systematic review of epidemiology, comorbidities and mortality. The Journal of Prevention of Alzheimer's Disease **11**(1), 97–107 (2024)

[3] Colliot, O.: Machine Learning for Brain Disorders (2023)

[4] Ansart, M., Epelbaum, S., Bassignana, G., *et al.*: Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review. Medical Image Analysis **67**, 101848 (2021)

[5] Frisoni, G.B., Altomare, D., et al.: Dementia prevention in memory clinics: recommendations from the European task force for brain health services. The Lancet Regional Health–Europe **26** (2023)

[6] Reuben, D.B., Kremen, S., Maust, D.T.: Dementia prevention and treatment: A narrative review. JAMA Internal Medicine **184**(5), 563–572 (2024)

https://jamanetwork.com/journals/jamainternalmedicine/articlepdf/2815813/jamainternal_reuben_2024_rv_230006_17

[7] Kim, H., Lee, S., Levine, A., Huber, B., Andrews, H., Kerner, N.A., Cohen, D., Carlson, S., Bell, S.A., Rivera, A.M., *et al.*: Novel measures of cognition and function for the ad spectrum in the novel measures for alzheimer's disease prevention trials (nomad) project: Psychometric properties, convergent validation, and contrasts with established measures. Alzheimer's & dementia **20**(8), 5089–5101 (2024)

[8] Grochowalski, J.H., Liu, Y., Siedlecki, K.L.: Examining the reliability of adas-cog change scores. Aging, Neuropsychology, and Cognition **23**(5), 513–529 (2016)

[9] Salthouse, T.A., Nesselroade, J.R., Berish, D.E.: Short-term variability in cognitive performance and the calibration of longitudinal change. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences **61**(3), 144–151 (2006)

[10] Marinescu, R.V., Oxtoby, N.P., Young, A.L., Bron, E.E., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Eshaghi, A., Toni, T., et al.: The alzheimer's disease prediction of longitudinal evolution (tadpole) challenge: Results after 1 year follow-up. arXiv preprint arXiv:2002.03419 (2020)

[11] Prakash, M., Abdelaziz, M., Zhang, L., Strange, B.A., Tohka, J., Initiative, A.D.N., *et al.*: Quantitative longitudinal predictions of alzheimer's disease by multi-modal predictive learning. Journal of Alzheimer's Disease **79**(4), 1533–1546 (2021)

[12] Jack Jr, C., Shiung, M., Gunter, J., O'brien, P., Weigand, S., Knopman, D.S., Boeve, B.F., Ivnik, R.J., Smith, G.E., Cha, R., *et al.*: Comparison of different mri brain atrophy rate measures with clinical disease progression in ad. Neurology **62**(4), 591–600 (2004)

[13] Cash, D.M., Rohrer, J.D., Ryan, N.S., Ourselin, S., Fox, N.C.: Imaging endpoints for clinical trials in alzheimer's disease. Alzheimer's research & therapy **6**, 1–10 (2014)

[14] Hunter, T.R., Santos, L.E., Tovar-Moll, F., De Felice, F.G.: Alzheimer's disease biomarkers and their current use in clinical research and practice. Molecular Psychiatry, 1–13 (2024)

[15] Hill, D.L., Schwarz, A.J., Isaac, M., Pani, L., Vamvakas, S., Hemmings, R., Carrillo, M.C., Yu, P., Sun, J., Beckett, L., *et al.*: Coalition against major diseases/european medicines agency biomarker qualification of hippocampal volume for enrichment of clinical trials in predementia stages of alzheimer's disease. Alzheimer's & Dementia **10**(4), 421–429 (2014)

[16] Drago, V., Babiloni, C., Bartrés-Faz, D., Caroli, A., Bosch, B., Hensch, T., Didic, M., Klafki, H.-W., Pievani, M., Jovicich, J., *et al.*: Disease tracking markers for alzheimer's disease at the prodromal (mci) stage. Journal of Alzheimer's disease **26**(s3), 159–199 (2011)

[17] Younes, L., Albert, M., Moghekar, A., Soldan, A., Pettigrew, C., Miller, M.I.: Identifying change-points in biomarkers during the preclinical phase of alzheimer's disease. Frontiers in aging neuroscience **11**, 74 (2019)

[18] Bertens, D., Knol, D.L., Scheltens, P., Visser, P.J., Initiative, A.D.N., *et al.*: Temporal evolution of biomarkers and cognitive markers in the asymptomatic, mci, and dementia stage of alzheimer's disease. Alzheimer's & Dementia **11**(5), 511–522 (2015)

[19] Nestor, S.M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J.L., Fogarty, J., Bartha, R., Initiative, A.D.N.: Ventricular enlargement as a possible measure of alzheimer's disease progression validated using the alzheimer's disease neuroimaging initiative database. Brain **131**(9), 2443–2454 (2008)

[20] Soininen, H., Solomon, A., Visser, P.J., Hendrix, S.B., Blennow, K., Kivipelto, M., Hartmann, T., Group, L.C.S.: 36-month lipididiet multinutrient clinical trial in prodromal alzheimer's disease. Alzheimer's & dementia **17**(1), 29–40 (2021)

[21] Dawe, R.J., Yu, L., Arfanakis, K., Schneider, J.A., Bennett, D.A., Boyle, P.A.: Late-life cognitive decline is associated with hippocampal volume, above and beyond its associations with traditional

neuropathologic indices. Alzheimer's & Dementia **16**(1), 209–218 (2020)

[22] Apostolova, L.G., Green, A.E., Babakchanian, S., Hwang, K.S., Chou, Y.-Y., Toga, A.W., Thompson, P.M.: Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment (mci), and alzheimer disease. Alzheimer Disease & Associated Disorders **26**(1), 17–27 (2012)

[23] Goukasian, N., Porat, S., Blanken, A., Avila, D., Zlatev, D., Hurtz, S., Hwang, K.S., Pierce, J., Joshi, S.H., Woo, E., *et al.*: Cognitive correlates of hippocampal atrophy and ventricular enlargement in adults with or without mild cognitive impairment. Dementia and geriatric cognitive disorders extra **9**(2), 281–293 (2019)

[24] Venkatraghavan, V., Bron, E.E., Niessen, W.J., Klein, S., Initiative, A.D.N., *et al.*: Disease progression timeline estimation for alzheimer's disease using discriminative event based modeling. NeuroImage **186**, 518–532 (2019)

[25] Liedes, H., Lötjönen, J., Kortelainen, J.M., Novak, G., Van Gils, M., Gordon, M.F., Initiative, A.D.N., Biomarkers, A.I., Ageing, L.F.S.: Multivariate prediction of hippocampal atrophy in alzheimer's disease. Journal of Alzheimer's Disease **68**(4), 1453–1468 (2019)

[26] Ellis, K.A., Bush, A.I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, 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 **21**(4), 672–687 (2009)

[27] Moradi, E., Dahnke, R., Imani, V., Gaser, C., Solomon, A., Tohka, J.: Integrating plasma, mri, and cognitive biomarkers for personalized prediction of decline across cognitive domains. medRxiv, 2024–12 (2025)

[28] Gaser, C., Dahnke, R., Thompson, P.M., Kurth, F., Luders, E., Alzheimer's Disease Neuroimaging Initiative: CAT: a computational anatomy toolbox for

the analysis of structural MRI data. GigaScience **13**, 049 (2024) https://doi.org/10.1093/gigascience/giae049 https://academic.oup.com/gigascience/article-pdf/doi/10.1093/gigascience/giae049/58736633/giae049.pdf

[29] Ashburner, J., Friston, K.J.: Diffeomorphic registration using geodesic shooting and gauss–newton optimisation. NeuroImage **55**(3), 954–967 (2011)

[30] Dahnke, R., Yotter, R.A., Gaser, C.: Cortical thickness and central surface estimation. NeuroImage **65**, 336–348 (2013) https://doi.org/10.1016/j.neuroimage.2012.09.050

[31] Yotter, R.A., Thompson, P.M., Gaser, C.: Algorithms to improve the reparameterization of spherical mappings of brain surface meshes. Journal of Neuroimaging **21**(2), 134–147 (2011) https://doi.org/10.1111/j.1552-6569.2010.00484.x https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1552-6569.2010.00484.x

[32] Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J.: An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. NeuroImage **31**(3), 968–980 (2006) https://doi.org/10.1016/j.neuroimage.2006.01.021

[33] Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology **67**(2), 301–320 (2005)

[34] Tohka, J., Moradi, E., Huttunen, H., Initiative, A.D.N.: Comparison of feature selection techniques in machine learning for anatomical brain mri in dementia. Neuroinformatics **14**, 279–296 (2016)

[35] Linn, K.A., Gaonkar, B., Doshi, J., Davatzikos, C., Shinohara, R.T.: Addressing confounding in predictive models with an application to neuroimaging. The international journal of biostatistics **12**(1), 31–44 (2016)

[36] Moradi, E., Prakash, M., Hall, A., Solomon, A., Strange, B., Tohka, J., Initiative, A.D.N.: Machine

learning prediction of future amyloid beta positivity in amyloid-negative individuals. Alzheimer's Research & Therapy **16**(1), 46 (2024)

[37] Tohka, J., Van Gils, M.: Evaluation of machine learning algorithms for health and wellness applications: A tutorial. Computers in Biology and Medicine **132**, 104324 (2021)

[38] Hittner, J.B., May, K., Silver, N.C.: A monte carlo evaluation of tests for comparing dependent correlations. The Journal of general psychology **130**(2), 149–168 (2003)

[39] Dunn, O.J., Clark, V.: Correlation coefficients measured on the same individuals. Journal of the American Statistical Association **64**(325), 366–377 (1969)

[40] Diedenhofen, B., Musch, J.: cocor: A comprehensive solution for the statistical comparison of correlations. PloS one **10**(4), 0121945 (2015)

[41] Lewis, J.D., Evans, A.C., Tohka, J., Group, B.D.C., *et al.*: T1 white/gray contrast as a predictor of chronological age, and an index of cognitive performance. Neuroimage **173**, 341–350 (2018)

[42] Johnson, W.E., Li, C., Rabinovic, A.: Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics **8**(1), 118–127 (2007)

[43] Fortin, J.-P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooper, C., Fava, M., McGrath, P.J., *et al.*: Harmonization of cortical thickness measurements across scanners and sites. Neuroimage **167**, 104–120 (2018)

[44] Hastie, T., Qian, J., Tay, K.: An introduction to glmnet. CRAN R Repositary (2021)

[45] Kuhn, M.: Building predictive models in r using the caret package. Journal of statistical software **28**, 1–26 (2008)

[46] Diedenhofen, B., Diedenhofen, M.B.: Package 'cocor'. Comprehensive R Archive Network. Available online at: https://cran. r-project. org/web/packages (accessed March 24, 2022) (2016)

[47] Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Müller, M., Siegert, S., Doering, M., Billings, Z., et al.: Package 'pROC'. Technical Report. Available online: https://cran. r-project. org/web ... (2017)

[48] Wickham, H.: ggplot2. Wiley interdisciplinary reviews: computational statistics **3**(2), 180–185 (2011)

[49] Gu, Z., Eils, R., Schlesner, M.: Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics **32**(18), 2847–2849 (2016)

[50] DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics, 837–845 (1988)

[51] Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M., Frisoni, G.B.: Brain atrophy in alzheimer's disease and aging. Ageing research reviews **30**, 25–48 (2016)

[52] Lee, J.-s., Heo, D.-y., Choi, K.-H., Kim, H.-J.: Impact of the ventricle size on alzheimer's disease progression: A retrospective longitudinal study. Dementia and Neurocognitive Disorders **23**(2), 95 (2024)

[53] Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., Initiative, A.D.N., *et al.*: Machine learning framework for early mri-based alzheimer's conversion prediction in mci subjects. Neuroimage **104**, 398–412 (2015)

[54] Rahim, N., El-Sappagh, S., Ali, S., Muhammad, K., Del Ser, J., Abuhmed, T.: Prediction of alzheimer's progression based on multimodal deep-learning-based fusion and visual explainability of time-series data. Information Fusion **92**, 363–388 (2023)

[55] Reas, E.T., Shadrin, A., Frei, O., Motazedi, E., McEvoy, L., Bahrami, S., Meer, D., Makowski,

C., Loughnan, R., Wang, X., *et al.*: Improved multimodal prediction of progression from mci to alzheimer's disease combining genetics with quantitative brain mri and cognitive measures. Alzheimer's & Dementia **19**(11), 5151–5158 (2023)

[56] Khundrakpam, B., Booij, L., Jeon, S., Karama, S., Tohka, J., Evans, A.C.: Individualized prediction of future cognition based on developmental changes in cortical anatomy. Neuroimage: Reports **2**(4), 100127 (2022)

[57] Bhagwat, N., Viviano, J.D., Voineskos, A.N., Chakravarty, M.M., Initiative, A.D.N., *et al.*: Modeling and prediction of clinical symptom trajectories in alzheimer's disease using longitudinal data. PLoS computational biology **14**(9), 1006376 (2018)

[58] Turrisi, R., Pati, S., Pioggia, G., Tartarisco, G., Initiative, A.D.N., et al.: Adapting to evolving mri data: A transfer learning approach for alzheimer's disease prediction. NeuroImage, 121016 (2025)

[59] Imani, V., Prakash, M., Zare, M., Tohka, J.: Comparison of single and multitask learning for predicting cognitive decline based on mri data. IEEE Access **9**, 154275–154291 (2021)

[60] Zhang, B., Xu, M., Wu, Q., Ye, S., Zhang, Y., Li, Z., Initiative, A.D.N.: Definition and analysis of gray matter atrophy subtypes in mild cognitive impairment based on data-driven methods. Frontiers in Aging Neuroscience **16**, 1328301 (2024)

[61] Sapkota, S., McFall, G.P., Masellis, M., Dixon, R.A., Black, S.E.: Differential cognitive decline in alzheimer's disease is predicted by changes in ventricular size but moderated by apolipoprotein e and pulse pressure. Journal of Alzheimer's Disease **85**(2), 545–560 (2022)

[62] Strange, B.A., Zhang, L., Sierra-Marcos, A., Alfayate, E., Tohka, J., Medina, M.: Predicting the future development of mild cognitive impairment in the cognitively healthy elderly. bioRxiv, 2020–07 (2020)

[63] Achterberg, H.C., Lijn, F., Heijer, T., Vernooij, M.W., Ikram, M.A., Niessen, W.J., Bruijne, M.:

Hippocampal shape is predictive for the development of dementia in a normal, elderly population. Human brain mapping **35**(5), 2359–2371 (2014)