Received: 2 February 2023 | Revised: 21 August 2023 | Accepted: 26 September 2023

DOI: 10.1002/dad2.12493

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

NeuropsychBrainAge: A biomarker for conversion from mild cognitive impairment to Alzheimer's disease

Jorge Garcia Condado[1,2] | **Jesus M. Cortes**[1,3,4] | **for the Alzheimer's Disease Neuroimaging Initiative**

[1]Computational Neuroimaging Laboratory, Biobizkaia Health Research Institute, Barakaldo, Bizkaia, Spain

[2]Biomedical Research Doctorate Program, University of the Basque Country, Leioa, Bizkaia, Spain

[3]Department of Cell Biology and Histology, University of the Basque Country, Leioa, Bizkaia, Spain

[4]IKERBASQUE Basque Foundation for Science, Bilbao, Spain

Correspondence
Jorge Garcia Condado, Computational Neuroimaging Laboratory, Biobizkaia Health Research Institute, Plaza Cruces S/N, 48903 Barakaldo, Bizkaia, Spain.
Email: jorge.garciacondado@bio-bizkaia.eus

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or the 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.

Funding information
'la Caixa' Foundation Fellowship: LCF/BQ/DI21/11860030; Ikerbasque: The Basque Foundation for Science; Department of Economic Development and Infrastructure of the Basque Country, Grant/Award Number: KK-2021-00009; ADNI, Grant/Award Number: U01 AG024904; DOD ADNI, Grant/Award Number: W81XWH-12-2-0012; National Institute on Aging; National Institute of Biomedical Imaging and Bioengineering

Abstract

INTRODUCTION: BrainAge models based on neuroimaging data have diagnostic classification power but have replicability issues due to site and patient variability. BrainAge models trained on neuropsychological tests could help distinguish stable mild cognitive impairment (sMCI) from progressive MCI (pMCI) to Alzheimer's disease (AD).

METHODS: A linear regressor BrainAge model was trained on healthy controls using neuropsychological tests and neuroimaging features separately. The BrainAge delta, predicted age minus chronological age, was used to distinguish between sMCI and pMCI.

RESULTS: The cross-validated area under the receiver-operating characteristic (ROC) curve for sMCI versus pMCI was 0.91 for neuropsychological features in contrast to 0.68 for neuroimaging features. The BrainAge delta was correlated with the time to conversion, the time taken for a pMCI subject to convert to AD.

DISCUSSION: The BrainAge delta from neuropsychological tests is a good biomarker to distinguish between sMCI and pMCI. Other neurological and psychiatric disorders could be studied using this strategy.

KEYWORDS
Alzheimer's disease, biomarker, BrainAge, mild cognitive impairment, neuropsychological test

Highlights

- BrainAge models based on neuropsychological tests outperform models based on neuroimaging features when distinguishing between stable mild cognitive impairment (sMCI) from progressive MCI (pMCI) to Alzheimer's disease (AD).
- The combination of neuropsychological tests with neuroimaging features does not lead to an improvement in sMCI versus pMCI classification compared to using neuropsychological tests on their own.
- BrainAge delta of both neuroimaging and neuropsychological models was correlated with the time to conversion, the time taken for a pMCI subject to convert to AD.

1 | INTRODUCTION

Recent advances in aging modeling have been aided by the use of machine learning and deep learning to create BrainAge models[1–4] based on neuroimaging data. BrainAge models have been applied rapidly to the medical fields to identify neurological disorders, such as mild cognitive impairment (MCI) and Alzheimer's disease (AD),[5–11] traumatic brain injury,[12,13] and multiple sclerosis,[14,15] and also psychiatric disorders, such as schizophrenia[16–19] and bipolar disorder.[19,20] Two types of models are used in classification tasks: those that use the difference between subject's predicted age trained on healthy controls (cognitively normal [CN]) and subject's chronological age, BrainAge delta, as a biomarker for classification,[1] and those that modify a deep learning model originally trained on BrainAge prediction and retrain the network on a classification task to distinguish CN subjects from patients.[4] The first type of model suffers from a lack of specificity for a given disease and the BrainAge delta seems to vary considerably between studies and models.[3] The second type of model does not have a specificity problem, but may suffer from the lack of enough training data for the patient subjects. One of the critical limitations of using neuroimaging is the variability intrinsic to this type of imaging across sites. If analyses are not carried out appropriately, site effects can dominate and make the models unusable. This poses a great challenge when thinking about bringing these methods to a clinical setting. There have been recent advances on using deep learning BrainAge models to minimize these effects and ensure replicability.[2] An alternative is to train these types of models with features better suited to the disease under study, such as neuropsychological tests.[21,22]

We aimed to develop a BrainAge model trained on neuropsychological features that can be used to identify a biomarker of MCI to AD conversion. We used the data consisting only of CN subjects for training. The output of the model, defined as NeuropsychBrainAge delta in this study, represents the difference between the subject's predicted age by the model and the subject's chronological age. To show the applicability of this model to a neurological disorder in the clinical setting, we applied the model to a cohort of subjects with MCI, of whom some remained stable and others progressed to AD. The proposed biomarker is capable of distinguishing with good accuracy between stable MCI (sMCI), those who remain MCI, and progressive MCI (pMCI), those who converted to AD. There have been BrainAge models based on neuropsychological features to study cognitive age[23] and to predict age directly from behavioral tests.[24] However, to the authors' knowledge, there are no previous studies using these tests to predict age and then use the difference between predicted age and chronological age as a biomarker.

RESEARCH IN CONTEXT

1. **Systematic review**: We reviewed the literature extensively. The use of machine-learning based BrainAge models has been adopted widely; however, they have focused on neuroimaging-based features. The use of neuropsychological features to create a BrainAge model with diagnostic classification power is unexplored.
2. **Interpretation**: We present a neuropsychological BrainAge model that has discriminative power to differentiate stable mild cognitive impairment (sMCI) from progressive MCI (pMCI) to Alzheimer's disease (AD). The biomarker has high accuracy and is correlated with the time to conversion, the time taken for a pMCI subject to convert to AD.
3. **Future directions**: This article poses a framework for the creation of BrainAge models based on neuropsychological tests that can be applied to other neurological disorders and psychiatric orders. BrainAge models based on similar types of features have the possibility of being more discriminant due to the specificity of the tests to each pathology.

2 | METHODS

2.1 | Subjects

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).[25] The ADNI was launched in 2003 as a public–private partnership, led by principal investigator, Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. (For up-to-date information, see www.adni-info.org.)

All subjects in the ADNI2 and ADNI3 phases who had an initial visit T1-weighted imaging and neuropsychological evaluation were extracted from the ADNI database. This included healthy controls (CN), and MCI and AD subjects. The data set statistics can be summarized as follows: 629 CN with mean age of 72.1 ± 6.8 of which 265 are male, 635 MCI with mean age of 72.5 ± 7.8 of which 356 are male, and 208 AD with mean age of 74.7 ± 8.2 of which 121 are male. Using longitudinal data, we identified conversors: those who converted from CN to MCI ($N = 64$), those who converted from MCI to AD ($N = 152$), and those who converted from CN to AD ($N = 7$). The main sociodemographic, cognitive, and brain indices for training and test sets and subjects separated by clinical label and prognosis can be found in Table S1.

A second data set was created to study the effect of homogenization on the results. For this data set only, sites with more than 10 CN subjects were used, and which had at least 1 MCI and 1 AD subject. The CN and MCI cohorts were also resampled to ensure that they had the same age distribution as the AD cohort.

The model was tested in the task of distinguishing between sMCI subjects, those who remained MCI, and pMCI subjects, those who

converted to AD. As it is impossible to know whether or an sMCI subject will convert to AD at a future date, sMCI patients were considered stable if they remained with a diagnosis of MCI after 3 years from the initial visit. Similarly, only subjects with pMCI who converted to AD within 3 years from the initial visit were used. As there are more sMCI subjects, these were randomly sampled to obtain a balanced data set containing equal amounts of sMCI ($N = 98$) and pMCI ($N = 98$) subjects.

2.2 | Features

Models were built using two different types of features: structural brain features extracted from T1-weighted images and neuropsychological features. The first set of features were used to build a reference model so that we can compare our second model.

T1-weighted images from each subject's first baseline visit were processed to obtain volumes of different brain structures. Structural Image Evaluation with Normalisation of Atrophy Cross-sectional (SIENAX),[26] part of FMRIB Software Library (FSL)[27] was used to obtain gray matter volume, white matter volume, cerebrospinal fluid volume, and peripheral grey matter volume as well as a volume scale value. FMRIB's Integrated Registration and Segmenration Tool (FIRST)[28] was used to segment and calculate the volumes of the thalamus, caudate, putamen, pallidum, brainstem, hippocampus, amygdala, and accumbens. The volume scale value was used to control for differences in brain size. For the data set used for more powerful homogenization, PyCombat[29] was used to homogenize, taking into account the volume scale value, gender, and phase in which the patient was recruited. Twelve features were used in total.

Neuropsychological assessments from each subject's first baseline visit were used as features for the second model. This consisted of scores from standard neuropsychological tests: Mini-Mental State Examination (MMSE),[30] Alzheimer's Disease Assessment Scale (ADAS),[31] Functional Assessment Questionaire (FAQ),[32] and Montreal Cognitive Assessment (MoCA),[33] as well as two metrics generated in the ADNI study: ADNI Memory score[34] and ADNI Executive Function.[35] Thus six features were used in total. The MMSE is a widely used screening test that assesses several cognitive domains, including orientation, attention, memory, language, and visuospatial abilities. It is commonly employed as a brief initial assessment tool for cognitive impairment and is often used to detect potential signs of dementia or other cognitive disorders. The ADAS is a comprehensive test battery designed to evaluate cognitive functions related primarily to AD measuring a range of cognitive abilities, including memory, language, attention, and problem-solving. The ADAS is used frequently used in research settings and clinical trials to track the progression of AD and assess the effects of interventions or treatments. The FAQ is a questionnaire-based assessment that focuses on a person's ability to perform activities of daily living, measuring the level of impairment in various functional domains, such as finance management, medication management, meal preparation, housekeeping, and transportation. The FAQ is commonly used to evaluate functional decline and independence in individuals with cognitive impairment or dementia. The MoCA assesses multiple cognitive domains, including attention and concentration, executive function, memory, language, visuospatial ability, and orientation. The MoCA is often utilized to detect MCI and to screen for potential cognitive impairments associated with various conditions, including AD and other dementias. ADNI Memory score is a composite memory score that measures the memory cognitive domain by combining memory domain scores of the Rey Auditory Verbal Learning Test, ADAS, MMSE, and Logical memory. ADNI Executive Function score is a composite score that measures the executive function cognitive domain by combining executive function domain scores of the WAIS-R Digit Symbol Substitution, Digit Span Backwards, Trails A and B, Category Fluency, and Clock Drawing. Both ADNI Memory and ADNI Executive Function have better psychometric characteristics than the widely used individual components.[34,35] The sensitivity of all of these tests can vary depending on various factors, such as the population being tested, the scoring procedures used, and the specific version or form of the test.[36] Generally these tests demonstrate good to excellent sensitivity in terms of internal consistency and test–retest reliability, particularly when administered by trained professionals according to standardized protocols.[37] The ADNI study ensured standardization during the administration of the tests to ensure consistency and reliability. However, it is important to note that reliability measures may vary across different studies or versions of the tests, and specific reliability coefficients may be available in the respective test manuals or research literature.

2.3 | Model

A linear regressor was trained in a supervised fashion (code can be found in https://github.com/JGarciaCondado/ADNIBrainAge). Python 3.9 and the package scikit-learn 1.0 were used for the analysis. Other models such as Ridge with polynomial feature expansion, K-Nearest Neighbors, and XGBoost were also tested. The task at hand is to use the features extracted from each subject at the initial visit to predict the age of the subject at the initial visit (Figure 1A). The model is trained using only CN subjects. For training, 85% of all CN subjects are used and 15% are used for testing. The features are normalized by subtracting the training mean and dividing by the training SD. The model can be summarized below:

$$BrainAge = \sum_i w_i f_i + b \quad (1)$$

where f_i is the value of each normalized feature, w_i the weight of each feature, and b the intercept. The weights and the intercept are chosen to minimize the mean square error between the predicted age (Brain Age) and the subject's age at the initial visit (Chronological Age, Ω):

$$\min_{w,\,b} \|BrainAge - \Omega\|_2^2 \quad (2)$$

There is a bias in the model, as younger controls tend to be given higher ages and older controls are given lower ages than they are as it

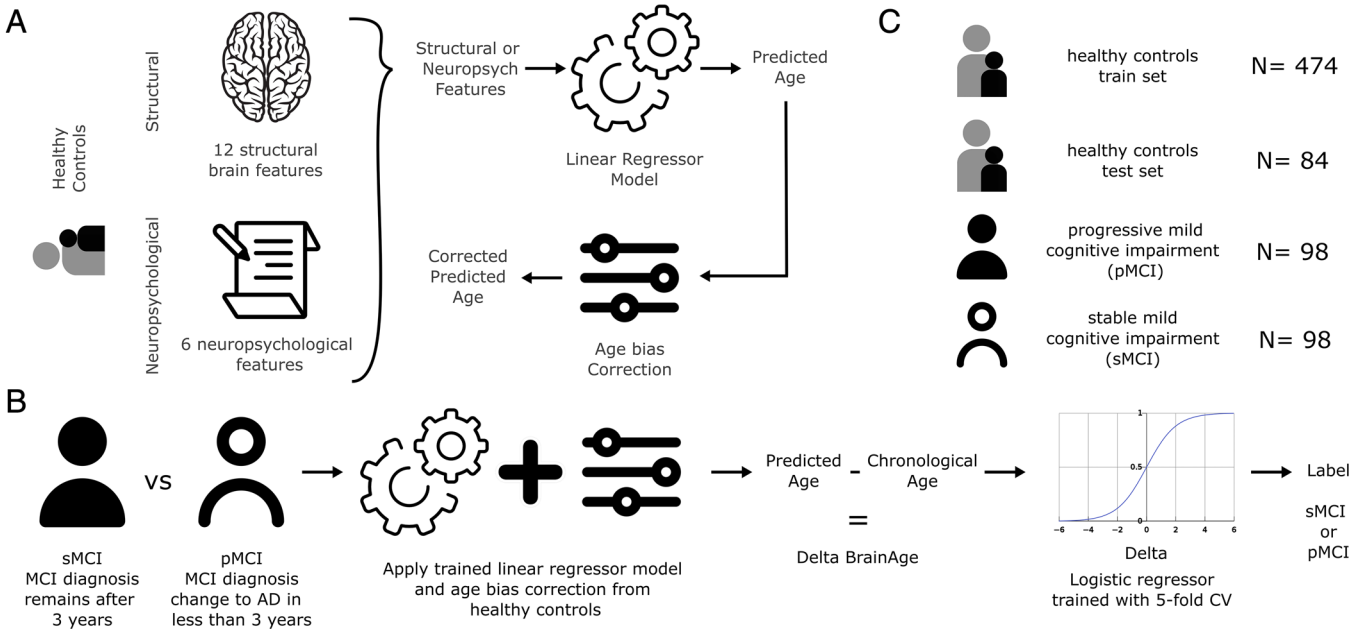

FIGURE 1 Overview of the BrainAge model and classification task. (A) Training of the BrainAge model on healthy controls with input data consisting of structural features or neuropsychological features. A total of 12 structural brain features were used, consisting of volume measured in mm^3 for: white matter, gray matter, peripheral gray matter, cerebrospinal fluid, thalamus, caudate, putamen, pallidum, hippocampus, amygdala, accumbens, and brainstem. A total of six neuropsychological features were used: Mini-Mental Status Examination (MMSE), Alzheimer's Disease Assessment Scale (ADAS), Functional Assessment Questionaire (FAQ), Montreal Cognitive Assessment (MoCA), ADNI Memory, and ADNI Executive Function. After training the linear regressor on the healthy controls age estimation task, an age bias correction was applied to deal with the inherent bias of regression to the mean problem. (B) Description of the classification task between stable mild cognitive impairment (sMCI) and progressive (pMCI). First, features were extracted for each subject as with healthy controls. Then, using either neuropsychological features or structural features, the trained model and bias correction were applied to obtain a predicted age. The BrainAge delta was calculated by subtracting the chronological age from the predicted age. This delta was then used as an input to a logistic regressor to determine a threshold for labeling using a five-fold cross validation (CV) scheme. (C) Number of subjects used for training with healthy controls, the number of subjects used to test the performance of BrainAge models on unseen healthy controls, and number of sMCI and pMCI used in the classification task. pMCI, progressive mild cognitive impairment.

is a regression to the mean problem.[38,39] It can be thought of in terms of our best estimate for a subject that we know nothing about, being the mean, so younger patients tend to be given higher ages and older patients lower ages. The predicted BrainAge can be adjusted by taking into account the actual age of subjects by fitting the following model[38]:

$$BrainAge = \alpha \times \Omega + \beta \quad (3)$$

The coefficients α and β represent the slope and intercept. The training set is used to fit α and β. Careful analysis and considerations have to be taken according to which method is used to do the correction, such as wether the predicted age or the delta, the difference between predicted and true age, will be used for further analysis, so as not to introduce new biases such as inflating correlation effects by choosing an appropriate method such as the one used; an alternative method to that used in this study is to use age as a covariate when studying group differences.[38,40,41] The fitted α and β are then used to correct the predictions in a test set using:

$$Corrected\ BrainAge = BrainAge + [\Omega - \alpha \times \Omega + \beta] \quad (4)$$

2.4 | Biomarker

The model is trained on CN subjects only, so it is then applied to our test CN subjects, MCI subjects, and AD subjects to predict their BrainAge. Our biomarker to differentiate our subjects is the BrainAge delta, defined as:

$$BrainAge\ delta = Corrected\ BrainAge - \Omega \quad (5)$$

Notice that by using the corrected BrainAge, the BrainAge delta will not correlate with age but will be related to our variable of interest, the cohort each subject belongs to.

Finally, to assess the potential of BrainAge delta to differentiate between sMCI and pMCI subjects, a logistic regressor is trained using a five-fold, cross-validation strategy on our task data set (Figure 1B). Four logistic regressors are trained with the following inputs: one using StructBrainAge delta (delta found using the model trained using structural imaging features), one using NeuropsychBrainAge delta (delta found using the model trained using neuropsychological features), one using the sum of both deltas, and finally one using both deltas

TABLE 1 Summary metrics of accuracy of BrainAge prediction in healthy controls for training and test sets of a linear regressor.

Set	Input Features	MAE	RMSE	r	R^2
Training	Neuroimaging	4.02	5.19	0.65	0.42
	Neuropsychological	4.91	6.31	0.38	0.15
Test	Neuroimaging	4.14	5.17	0.49	0.22
	Neuropsychological	4.51	5.59	0.32	0.08

Abbreviations: MAE, mean absolute error; r, Pearson's correlation; R^2, R-squared; RMSE, root mean squared error.

independently. Moreover, we also follow a different strategy of training logistic regressors directly on the features (using only structural features, using neuropsychological features, and using all features) adjusted for age, sex, and education to compare to the logistic regressors trained on BrainAge deltas. In total, seven different logistic regressors were trained. Principal component analysis (PCA) decomposition was also carried out on each set of features (structural, neuropsychological, and all features) and logistic regressors trained with the components that explained 90% of the variance.

3 | RESULTS

3.1 | Correlation between features and age

As a first initial assessment of the suitability of the proposed features for predicting age, Pearson correlation coefficients were calculated between each feature and age of the CN subjects. Correlations between structural features and age are shown in Figure 2A, ranging from $c_{min} = 0.12$ (brainstem, $p < 0.02$) to $c_{max} = 0.55$ (gray matter, $p < 3 \times 10^{-39}$). As noted in previous research,[42] there is a strong linear correlation between certain brain structures and population aging above 50 years. Correlations between neuropsychological features and age are shown in Figure 2B, ranging from $c_{min} = 0.08$ (FAQ, $p < 0.10$) to $c_{max} = 0.35$ (ADNI Executive Function, $p < 7 \times 10^{-15}$). These have a weaker correlation yet there is still one.

3.2 | A model trained only on healthy controls

A BrainAge model was trained on CN subjects on two set of features (neuroimaging and neuropsychological) and tested on a different set of CN subjects. Summary metrics of the age-prediction accuracy for both train and test sets can be found in Table 1. The more complex models—Ridge with polynomial features expansion, K-Nearest Neighbors, and XGBoost—all performed similarly to the linear regression models. Hence, given the linear relationship between age and features as shown in Figure 2 and following Occam's razor, the linear model was used for the rest of the study.

3.3 | Differences in BrainAge deltas between cohorts

We applied both BrainAge models, one trained on neuroimaging features and another trained on neuropsychological features, to all cohorts: the test set CN, MCI, and AD subjects. We also applied age-bias correction with parameters fitted on the training set as outlined in Equation (4) on all cohorts. Then we calculated the BrainAge delta for each subject. The BrainAge delta was significantly higher between CN and MCI for both neuroimaging (-0.32 ± 3.28 vs 1.92 ± 3.97, $p < 1 \times 10^{-7}$) and neuropsychology (-0.21 ± 2.19 vs 4.42 ± 3.95, $p < 1 \times 10^{-26}$), as well as between MCI and AD subjects, for neuroimaging (1.92 ± 3.97 vs 4.98 ± 4.50, $p < 1 \times 10^{-17}$) and neuropsychology (4.42 ± 3.95 vs 16.07 ± 6.38, $p < 1 \times 10^{-83}$). The differences are shown in Figure 3A.

We next addressed the BrainAge delta of the pMCI cohort, those labeled as MCI at the baseline visit but who would progress to AD during follow-ups. There was a correlation between the BrainAge delta and the conversion time, the number of years until the subject was labeled AD. Pearson's correlation between conversion time and BrainAge delta was stronger when neuropsychological features were used (-0.48, $p < 1 \times 10^{-9}$) as compared to neuroimaging features (-0.23, $p < 0.005$), illustrated in Figure 3B. The absolute value of Pearson's correlation between time to conversion and PCA decomposition of structural features ranged from $c_{min} = 0.002$ ($p < 0.99$) to $c_{max} = 0.19$ ($p < 0.02$) and for neuropsychological features ranged from $c_{min} = 0.15$ ($p < 0.07$) to $c_{max} = 0.40$ ($p < 1 \times 10^{-8}$). The absolute value of Pearson's correlation between time to conversion and individual features corrected by age, education, and sex for structural features ranged from $c_{min} = 0.004$ (palidum, $p < 0.96$) to $c_{max} = 0.26$ (peripheral gray matter, $p < 0.002$) and for neuropsychological features ranged from $c_{min} = 0.23$ (MoCA, $p < 0.006$) to $c_{max} = 0.37$ (ADNI Executive Function, $p < 1 \times 10^{-5}$). The neuropsychological based BrainAge delta has the highest absolute Pearson's correlation with time to conversion.

3.4 | Feasibility of application to identify sMCI versus pMCI subjects using BrainAge deltas

Next we asked whether the BrainAge Delta could discriminate between sMCI subjects (those who remain MCI for at least 3 years after the initial visit) from pMCI subjects (those who convert to AD within 3 years of the initial visit). A logistic regressor was trained on these BrainAge deltas using the neuroimaging model (StructBrainAge delta), the neuropsychological model (NeuropsychBrainAge delta), the sum of both deltas, or each delta individually. The ROC curve for each can be seen in Figure 4. The best results were obtained when using the NeuropsychBrainAge deltas only as seen in Table 2.

Further investigations were carried out to compare whether it would be better to use the features directly in a logistic regressor. For this purpose, logistic regressors to classify sMCI versus pMCI were

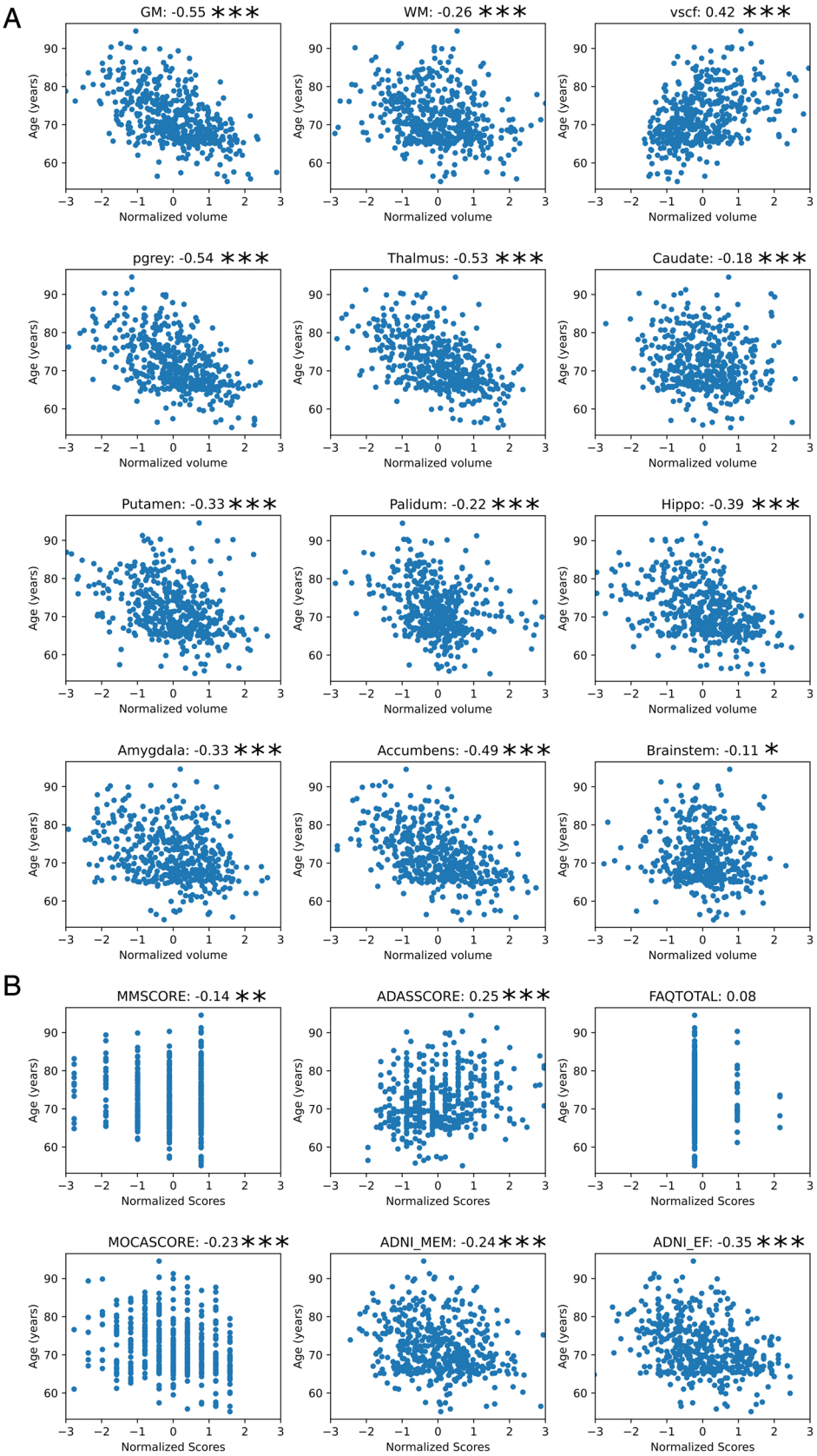

FIGURE 2 Correlation between BrainAge model features and age in healthy controls (cognitively normal [CN]) subjects resulting from (A) structural imaging features and (B) neuropsychological test features. The title of each graph describes the precise feature and its value of the Pearson's correlation coefficient with age. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

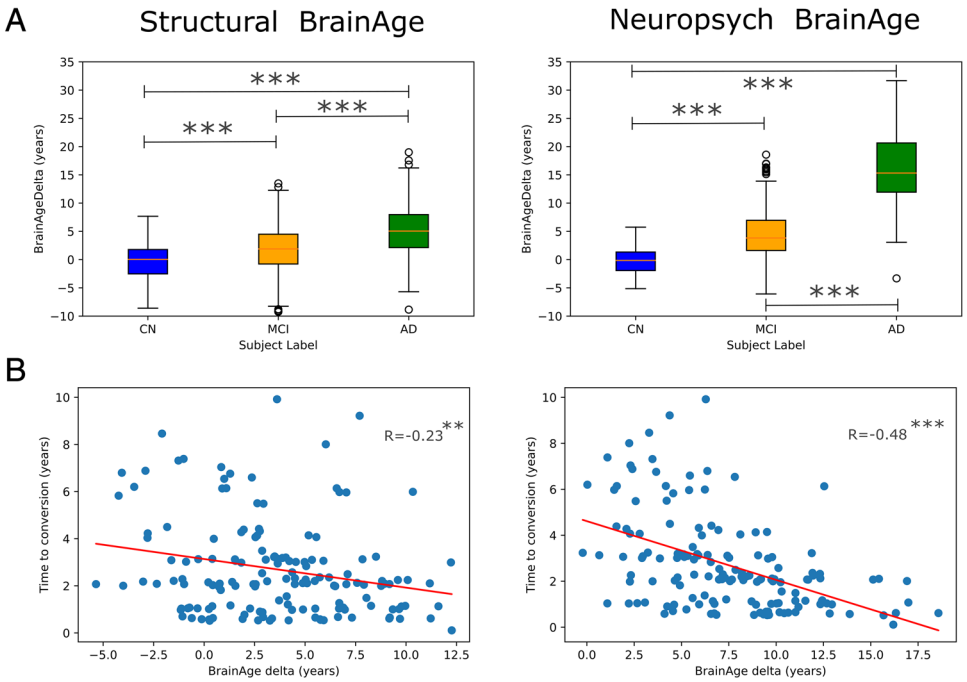

FIGURE 3 BrainAge delta for each subject group and its relation to conversion time for subjects with progressive mild cognitive impairment. Structural BrainAge refers to the model trained on neuroimaging features and NeuropsychBrainAge refers to the model trained on neuropsychological features. (A) BrainAge delta for each cohort, healthy control (CN), MCI, and Alzheimer's disease (AD). (B) Time to conversion for subjects with pMCI as a function of BrainAge delta, where R indicates Pearson's correlation coefficient. Time to conversion is defined as the time between the subject's first baseline visit and the visit at which the subject is labeled with AD. $^{**}p < 0.01$, $^{***}p < 0.001$. MCI, mild cognitive impairment; pMCI, progressive mild cognitive impairment.

TABLE 2 Results of different logistic regressors trained in a 5-fold CV scheme to distinguish between stable mild cognitive impairment and progressive mild cognitive impairment subjects using different BrainAge deltas as inputs and possible combinations as well as using neuroimaging and neuropsychological features corrected by age, sex and education directly as inputs.

Model	N° Inputs	AUC	Accuracy	Sensitivity	Specificity
StructBrainAge delta (SBAD)	1	0.68 [0.53–0.83]	0.61 [0.51–0.72]	0.63 [0.49–0.76]	0.63 [0.47–0.80]
NeuropsychBrainAge delta (NBAD)	1	0.91 [0.84–0.98]	**0.85 [0.76–0.94]**	**0.87 [0.78–0.97]**	**0.87 [0.76–0.99]**
Addition (SBAD+NBAD)	1	0.86 [0.79–0.94]	0.79 [0.69–0.89]	0.81 [0.70–0.93]	0.81 [0.66–0.96]
Multidomain (SBAD, NBAD)	2	0.90 [0.85–0.96]	**0.85 [0.77–0.93]**	0.86 [0.77–0.95]	0.85 [0.72–0.95]
Neuroimaging features	12	0.63 [0.59–0.67]	0.58 [0.55–0.61]	0.58 [0.52–0.64]	0.58 [0.50–0.67]
Neuropsychological features	6	**0.92 [0.86–0.98]**	0.83 [0.73–0.94]	0.86 [0.75–0.97]	0.86 [0.73–1.00]
All features	18	0.89 [0.81–0.97]	0.84 [0.77–0.90]	0.86 [0.78–0.94]	0.85 [0.71–0.99]

Abbreviation: AUC, area under the curve.
Bold indicates the model with highest value for that specific metric.

trained using a five-fold CV scheme directly on neuroimaging and neuropsychological features corrected for age, sex, and education instead of BrainAge deltas as the input variables. The NeuropsychBrainAge delta–trained logistic regressor outperformed all other logistic regressors trained directly on neuroimaging or neuropsychological features on accuracy, sensitivity, and specificity, as shown in Table 2. However, the model trained on neuropsychological features corrected for age, sex, and education had better AUC and very similar accuracy, sensitivity, and specificity. PCA decomposition showed no increases in performance compared to using the features directly; performance metrics can be seen in Table S2.

3.5 | Effects of homogenization of neuroimaging features

Finally, we also tested whether the use of homogenization techniques on neuroimaging features improved the performance of logistic

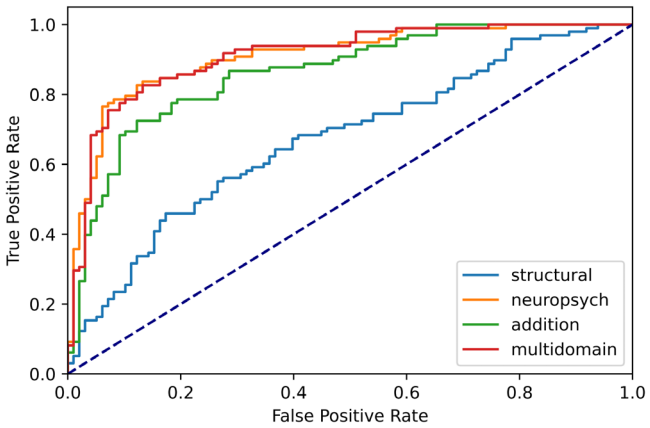

FIGURE 4 Receiver-operating characteristic (ROC) curve for different logistic regressors models trained in a five-fold CV scheme to discriminate between sMCI and pMCI subjects. (light blue) Structural refers to using the Structural BrainAge delta as the input to the logistic regressor. (orange) Neuropsych refers to using the Neuropsych BrainAge delta as the input to the logistic regressor. (green) Addition refers to using the sum of the Structural BrainAge delta and the Neuropsych BrainAge delta for each subject as a single input to the logistic regressor. (red) Multidomain refers to using Neuropsych BrainAge delta and Structural BrainAge delta as two separate inputs to the logistic regressor, hence in this logistic regressor there are two input features instead of one. (dashed) Represents the performance of a truly random classifier. sMCI, stable mild cognitive impairment; pMCI, progressive mild cognitive impairment.

regressors in the sMCI versus pMCI discrimination task. Neuroimaging homogenization did improve performance (Table S3), but it was still lower than the performance achieved by NeuropsychBrainAge delta.

4 | DISCUSSION

One of the key issues in the clinical application of BrainAge models is developing models that are reliable and reproducible. To achieve this, one strategy is to accurately track the trajectory of aging in a normal healthy population. Deviations from this normal aging trajectory can then indicate risks of developing certain conditions. The advantage of this type of modeling is two-fold. The first is that by training the model on a healthy population and then tracking deviations from this, no assumptions are made about the condition under study that could bias the results. Second, it is much easier to collect data for healthy individuals and hence get larger data sets to train with. These models are then tested on the cohort with an underlying condition that we aim to identify to ensure that the biomarker can correctly distinguish subjects in each cohort. BrainAge models were trained on healthy controls ($N = 474$) and then the logistic regressors only had to find a threshold to divide the BrainAge deltas (a single feature) of sMCI ($N = 95$) versus pMCI ($N = 95$). On the other hand, logistic regressors trained directly on features had to combine information from multiple features (12 in the case of neuroimaging, 6 in the case of neuropsychological, or 18 in total) directly only on our task set of sMCI ($N = 95$) versus pMCI ($N = 95$). This could be leading to worse generalization due to a lower number of datapoints because the confidence intervals (Cis) of the cross-validation set are wider for the logistic regressor trained directly on features. More research is needed to validate these findings with diverse data sets and different pathologies.

The logistic regressor trained on NeuropsychBrainAge deltas was able to outperform all other models based on BrainAge deltas. It performs very similarly to using each BrainAge delta as separate inputs to the logistic regressor (multidomain approach), showing that all the information captured by StructBrainAge is already captured by NeuropsychBrainAge. In comparison to other state-of-the-art models such as the BrainAge model original developed by Gaser et al.,[1] based on structural imaging features and therefore similar to our study's StructBrainAge. Gaser et al's model achieved accuracies of 0.81, lower than the 0.85 achieved by ours. It should be noted that the BrainAge model based on neuropsychological features performs worse in the task of predicting age, since its mean absolute error (MAE) in the test set for healthy controls is worse before bias correction than the model trained on neuroimaging data.

Age-prediction models have utility in capturing atypical aging, but it is crucial to minimize methodological variance by building accurate models that capture biological variance. Overfitted models can result in smaller differences between AD and CN individuals, whereas models with lower accuracy might capture better biological variance as shown by previous studies.[4,43] Although more research is needed, since other studies argue that "loose fitting" violates the conceptual foundation of BrainAge modeling, can be highly problematic from a methodological point of view, and might lead to results that are uninterpretable.[44] Some moderately accurate models demonstrate a high delta for AD and a strong correlation with AD scales, whereas the model with the highest delta shows a weaker correlation with behavior.[43] Nonlinear relationships exist between model accuracy, deltas, and behavioral correlations. Therefore, having higher model accuracy in the age-prediction task does not result in better differentiation of sMCI versus pMCI subjects as shown by our NeuropsychBrainAge model being worse at age prediction but better at differentiation compared to the StructBrainAge. Regularized models in the patient population, despite lower accuracy, may be beneficial, as they focus on specific features related to atypical aging. Comparing models based on their performance and delta-behavior correlations in patient data is a promising but open area, requiring further studies to define appropriate model-selection procedures.

The large difference in accuracy between models trained in neuroimaging and neuropsychological tests, and after careful inspection of other models in the literature, raises the question of whether the concept of BrainAge as a biomarker is robust and not strongly model dependent. Further studies are required to better understand whether there is a correlation between MAE before age bias correction and better performance on different classification tasks. There is a wide variety in reported BrainAge deltas between studies for similar cohorts. For example, AD subjects may have mean BrainAge deltas ranging from +5.35 years in one study to +10.70 in another.[3] This great variability indicates the need for further research on how to ensure

that BrainAge-derived biomarkers are robust for clinical application across sites and subjects.

There might be concern in using neuropsychological tests to assess conversion, since they are used to assign the healthy control, MCI, and AD labels in the first place. In ADNI, a selection of these tests are used in combination with cutoff points to assign those labels as well as other clinical assessments.[25] Various tests, such as clinical dementia rating,[45] are used to assign labels that we did not use in our study. However, the most important remark is that we trained our BrainAge model on healthy controls and then assessed the MCI conversion from baseline scores. There is no data leakage in terms of biases in the BrainAge deltas, as the model is not trained on MCI subjects but on healthy controls. These tests have been used to assign subjects who belong to MCI at baseline, but we can predict future conversion to AD with those same scores without requiring future scores beyond the AD cutoff points. This is evidence to show that hard cutoff points are not the optimal tool for assigning labels, as with hard cutoffs we label subjects into the same category (MCI) that have different pathological trajectories (sMCI vs pMCI). It is important to note that pMCIs are assigned as MCIs with the baseline scores because of the test cutoff scores, and we can distinguish them from sMCI in the future with the baseline scores, but never using future assessment scores, which make their label switch from MCI to AD. This shows that more information can be extracted from neuropsychological tests than is obvious by using cutoff points.

The logistic regressor trained directly on the neuropsychological features performs similarly to the logistic regressor trained on the NeuropsychBrainAge deltas. This is to be expected because the NeuropsychBrainAge delta is just a linear combination of the neuropsychological features, hence they both contain similar information. Therefore, the logistic regressor trained on NeuropsychBrainAge delta is just two linear transformations of the neuropsychological features instead of a single linear transformation when training directly on the neuropsychological features. However, the correlation between NeuropsychBrainAge delta and time to conversion is stronger than for any individual neuropsychological feature after correcting for age, education, and sex, and also shows a higher correlation than PCA decomposition.

The advantage of using the NeuropsychBrainAge delta over using the neuropsychological features is two-fold. First, it is a good summary metric for clinicians because in one single value it is both a good predictor of conversion and has easy interpretation due to its high correlation with time to conversion. The higher the NeuropsychBrainAge delta the less time to conversion from MCI to AD. Second, it is trained on a healthy population with a higher number of subjects and no assumptions except that the subject are cognitively healthy, this could potentially lead to a reduction in bias and less overfitting. The CIs of the NeuropsychBrainAge models in cross-validation accuracy, sensitivity, and specificity are smaller than the model directly trained on the neuropsychological features, thus showing some robustness to changes in the input training data. Similarly, a positive effect of first training on the age task in healthy subjects has been shown with larger cohorts.[2,4] Further research is needed to investigate to what extent this is the case with neuropsychological tests. Moreover, further explorations are also required combining different cohorts and data sets to show robust generalizability, although this is also challenging as different studies use different neuropsychological tests.

A major limitation to our study is lack of generalizability to different types of populations; this study is intended more as a case study. The model will not be able to perform well if the population under study varies significantly in comparison to the data used to train the model—for example, if the subjects under study come from sociocultural backgrounds different from those from the ADNI cohort. Similarly, it would require the neuropsychological batteries to be the same as those set out in ADNI. However, it is a valuable study to showcase the potential of this type of approach when using neuropsychological tests.

In this study we have shown that neuropsychological features combined with BrainAge modeling can yield a valuable biomarker to distinguish between sMCI and pMCI subjects. This biomarker also shows a strong correlation with conversion time, which is a sign of a robust biomarker, as good biomarkers should have higher values with higher levels of severity, in this case, less time to conversion. Neuropsychological tests are better suited than neuroimaging-derived features for clinical application due to their larger effect size and lower cost to administer.[22] Even when accounting for site effects in neuroimaging-derived features by homogenization techniques, the neuropsychological-based models still outperformed the neuroimaging ones. This shows that the use of BrainAge models combined with specific neuropsychological tests for a specific condition can provide valuable and accurate biomarkers.

5 | CONCLUSION

We present a BrainAge model trained on neuropsychological tests able to discriminate sMCI subjects from pMCI subjects. BrainAge models have the advantage of training on a larger cohort of healthy controls (CN) to measure deviations from the norm. By using features tightly linked to a specific condition, such as the neuropsychological tests for AD determination, we were able to achieve good performance on the classification tasks. Furthermore, we have shown that this is a robust biomarker because it was correlated with the conversion time from MCI patients to AD. We expect that our approach can be extended to other neurological and psychological disorders by applying the same models but with different neuropsychological tests specific to each condition.

AUTHOR CONTRIBUTIONS

Jorge Garcia Condado and Jesus M. Cortes designed the study. Jorge Garcia Condado analyzed the data. Jorge Garcia Condado. and Jesus M. Cortes interpreted results and supported the analysis. All authors wrote and edited the article. All authors approved the article.

ACKNOWLEDGMENTS

We are grateful to Alzheimer's Disease Neuroimaging Initiative (ADNI) study investigators and participants for providing the data used in this

article. The project that gave rise to these results received the support of a fellowship from "la Caixa" Foundation (ID 100010434). The fellowship code is LCF/BQ/DI21/11860030. J.M.C is funded by Ikerbasque: The Basque Foundation for Science and by the Department of Economic Development and Infrastructure of the Basque Country (Elkartek Program Grant KK-2021-00009).

Data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (U.S. 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.;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 Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

CONFLICT OF INTEREST STATEMENT

All authors declare no competing interests. Author disclosures are available in the supporting information.

CONSENT STATEMENT

All Alzheimer's Disease Neuroimaging Initiative (ADNI) participants consented to ADNI's data policy more information can be found in: https://adni.loni.usc.edu/data-samples/access-data/.

ORCID

Jorge Garcia Condado https://orcid.org/0000-0002-9463-1924
Jesus M. Cortes https://orcid.org/0000-0002-9059-8194

REFERENCES

1. Gaser C, Franke K, Klöppel S, Koutsouleris N, Sauer H. Alzheimer's disease neuroimaging initiative. BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease. *PLoS ONE*. 2013;8:e67346. doi:10.1371/journal.pone.0067346
2. Leonardsen EH, Peng H, Kaufmann T, et al. Deep neural networks learn general and clinically relevant representations of the ageing brain. *NeuroImage*. 2022;256:119210. doi:10.1016/j.neuroimage.2022.119210
3. Baecker L, Garcia-Dias R, Vieira S, Scarpazza C, Mechelli A. Machine learning for brain age prediction: introduction to methods and clinical applications. *EBioMedicine*. 2021;72:103600. doi:10.1016/j.ebiom.2021.103600
4. Bashyam VM, Erus G, Doshi J, et al. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain*. 2020;143:2312-2324. doi:10.1093/brain/awaa160
5. Franke K, Ziegler G, Klöppel S, Gaser C. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. *NeuroImage*. 2010;50:883-892. doi:10.1016/j.neuroimage.2010.01.005
6. Varikuti DP, Genon S, Sotiras A, et al. Evaluation of non-negative matrix factorization of grey matter in age prediction. *NeuroImage*. 2018;173:394-410. doi:10.1016/j.neuroimage.2018.03.007
7. Mohajer B, Abbasi N, Mohammadi E, et al. Gray matter volume and estimated brain age gap are not linked with SLEEP-DISORDERED breathing. *Hum Brain Mapp*. 2020;41:3034-3044. doi:10.1002/hbm.24995
8. Beheshti I, Maikusa N, Matsuda H. The association between "Brain-Age Score" (BAS) and traditional neuropsychological screening tools in Alzheimer's disease. *Brain Behav*. 2018;8:e01020. doi:10.1002/brb3.1020
9. Franke K, Gaser C. Longitudinal changes in individual *BrainAGE* in healthy aging, mild cognitive impairment, and Alzheimer's disease. *GeroPsych*. 2012;25:235-245. doi:10.1024/1662-9647/a000074
10. Ly M, Yu GZ, Karim HT, et al. Improving brain age prediction models: incorporation of amyloid status in Alzheimer's disease. *Neurobiol Aging*. 2020;87:44-48. doi:10.1016/j.neurobiolaging.2019.11.005
11. Löwe LC, Gaser C, Franke K, for the Alzheimer's Disease Neuroimaging Initiative. The effect of the APOE genotype on individual BrainAGE in normal aging, mild cognitive impairment, and Alzheimer's disease. *PLoS ONE*. 2016;11:e0157514. doi:10.1371/journal.pone.0157514
12. Savjani RR, Taylor BA, Acion L, Wilde EA, Jorge RE. Accelerated changes in cortical thickness measurements with age in military service members with traumatic brain injury. *J Neurotrauma*. 2017;34:3107-3116. doi:10.1089/neu.2017.5022
13. Cole JH, Leech R, Sharp DJ, for the Alzheimer's Disease Neuroimaging Initiative. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Ann Neurol*. 2015;77:571-581. doi:10.1002/ana.24367
14. Høgestøl EA, Kaufmann T, Nygaard GO, et al. Cross-Sectional and longitudinal MRI brain scans reveal accelerated brain aging in multiple sclerosis. *Front Neurol*. 2019;10:450. doi:10.3389/fneur.2019.00450
15. Cole JH, Raffel J, Friede T, et al. Longitudinal assessment of multiple sclerosis with the brain-age paradigm. *Ann Neurol*. 2020;88:93-105. doi:10.1002/ana.25746
16. Schnack HG, van Haren NEM, Nieuwenhuis M, Hulshoff Pol HE, Cahn W, Kahn RS. Accelerated brain aging in schizophrenia: a longitudinal pattern recognition study. *Am J Psychiatry*. 2016;173:607-616. doi:10.1176/appi.ajp.2015.15070922
17. Koutsouleris N, Davatzikos C, Borgwardt S, et al. Accelerated brain aging in schizophrenia and beyond: a neuroanatomical marker of psychiatric disorders. *Schizophr Bull*. 2014;40:1140-1153. doi:10.1093/schbul/sbt142
18. Lee WH, Antoniades M, Schnack HG, Kahn RS, Frangou S. Brain age prediction in schizophrenia: does the choice of machine learning algorithm matter? *Psychiatry Res Neuroimaging*. 2021;310:111270. doi:10.1016/j.pscychresns.2021.111270
19. Nenadić I, Dietzek M, Langbein K, Sauer H, Gaser C. BrainAGE score indicates accelerated brain aging in schizophrenia, but not bipolar disorder. *Psychiatry Res. Neuroimaging*. 2017;266:86-89. doi:10.1016/j.pscychresns.2017.05.006

20. Van Gestel H, Franke K, Petite J, et al. Brain age in bipolar disorders: effects of lithium treatment. *Aust N Z J Psychiatry*. 2019;53:1179-1188. doi:10.1177/0004867419857814
21. for the Alzheimer's Disease Neuroimaging Initiative, LQ Park, Gross AL, McLaren DG, et al, for the Alzheimer's Disease Neuroimaging Initiative. Confirmatory factor analysis of the ADNI neuropsychological battery. *Brain Imaging Behav*. 2012;6:528-539. doi:10.1007/s11682-012-9190-3
22. for the NEURONORMA Study Team, Sánchez-Benavides G, Peña-Casanova J, Casals-Coll M, et al, for the NEURONORMA Study Team. Cognitive and neuroimaging profiles in mild cognitive impairment and Alzheimer's disease: data from the Spanish multicenter normative studies (NEURONORMA Project). *J Alzheimers Dis*. 2014;41:887-901. doi:10.3233/JAD-132186
23. Anatürk M, Kaufmann T, Cole JH, et al. Prediction of brain age and cognitive age: quantifying brain and cognitive maintenance in aging. *Hum Brain Mapp*. 2021;42:1626-1640. doi:10.1002/hbm.25316
24. Statsenko Y, Habuza T, Charykova I, et al. Predicting age from behavioral test performance for screening early onset of cognitive decline. *Front Aging Neurosci*. 2021;13:661514. doi:10.3389/fnagi.2021.661514
25. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010;74:201-209. doi:10.1212/WNL.0b013e3181cb3e25
26. Smith SM, Zhang Y, Jenkinson M, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage*. 2002;17:479-489. doi:10.1006/nimg.2002.1040
27. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*. 2004;23:S208-S219. doi:10.1016/j.neuroimage.2004.07.051
28. Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*. 2011;56:907-922. doi:10.1016/j.neuroimage.2011.02.046
29. Behdenna A, Haziza J, Azencott C-A, Nordor A. pyComBat, a python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods. *Bioinformatics*. 2020. doi:10.1101/2020.03.17.995431. bioRxiv.
30. Folstein MF, Folstein SE, McHugh PR. Mini-mental state. *J Psychiatr Res*. 1975;12:189-198. doi:10.1016/0022-3956(75)90026-6
31. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *J Prev Alzheimers Dis*. 1984;141:1356-1364. doi:10.1176/ajp.141.11.1356
32. Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S. Measurement of functional activities in older adults in the community. *Journal of Gerontology*. 1982;37:323-329. doi:10.1093/geronj/37.3.323
33. Nasreddine ZS, Phillips NA, Bédirian V, et al. The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment: mOCA: a Brief screening tool for MCI. *J Am Geriatr Soc*. 2005;53:695-699. doi:10.1111/j.1532-5415.2005.53221.x
34. for the Alzheimer's Disease Neuroimaging Initiative, PK Crane, Carle A, Gibbons LE, et al, for the Alzheimer's Disease Neuroimaging Initiative. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav*. 2012;6:502-516. doi:10.1007/s11682-012-9186-z
35. for the Alzheimer's Disease Neuroimaging Initiative, Gibbons LE, Carle AC, Mackin RS, et al, for the Alzheimer's Disease Neuroimaging Initiative. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav*. 2012;6:517-527. doi:10.1007/s11682-012-9176-1
36. Franzen S, Van Den Berg E, Goudsmit M, et al. A systematic review of neuropsychological tests for the assessment of dementia in non-western, low-educated or illiterate populations. *J Int Neuropsychol Soc*. 2020;26:331-351. doi:10.1017/S1355617719000894
37. Sherman EMS, Tan JE, Hrabok M. *A compendium of neuropsychological tests: administration, tests, norms, and commentary*. Fourth ed. Oxford University Press; 2020.
38. de Lange A-MG, Cole JH. Commentary: correction procedures in brain-age prediction. *NeuroImage Clin*. 2020;26:102229. doi:10.1016/j.nicl.2020.102229
39. Zhang B, Zhang S, Feng J, Zhang S. Age-level bias correction in brain age prediction. *NeuroImage Clin*. 2023;37:103319. doi:10.1016/j.nicl.2023.103319
40. de Lange AG, Anatürk M, Rokicki J, et al. Mind the gap: performance metric evaluation in brain-age prediction. *Hum Brain Mapp*. 2022;43:3113-3129. doi:10.1002/hbm.25837
41. Butler ER, Chen A, Ramadan R, et al. Pitfalls in brain age analyses. *Hum Brain Mapp*. 2021;42:4092-4101. doi:10.1002/hbm.25533
42. Dima D, Modabbernia A, Papachristou E, et al. Subcortical volumes across the lifespan: data from 18,605 healthy individuals aged 3-90 years. *Hum Brain Mapp*. 2022;43:452-469. doi:10.1002/hbm.25320
43. More S, Antonopoulos G, Hoffstaedter F, Caspers J, Eickhoff SB, Patil KR. Brain-age prediction: a systematic comparison of machine learning workflows. *NeuroImage*. 2023;270:119947. doi:10.1016/j.neuroimage.2023.119947
44. Hahn T, Fisch L, Ernsting J, et al. From 'loose fitting' to high-performance, uncertainty-aware brain-age modelling. *Brain*. 2021;144:e31-e31. doi:10.1093/brain/awaa454
45. Hughes CP, Berg L, Danziger W, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry*. 1982;140:566-572. doi:10.1192/bjp.140.6.566

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Garcia Condado J, Cortes JM. NeuropsychBrainAge: A biomarker for conversion from mild cognitive impairment to Alzheimer's disease. *Alzheimer's Dement*. 2023;15:e12493. https://doi.org/10.1002/dad2.12493

APPENDIX

COLLABORATORS

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.