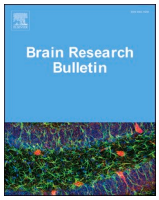
Brain Research Bulletin 205 (2023) 110811

Contents lists available at ScienceDirect

Brain Research Bulletin

journal homepage: www.elsevier.com/locate/brainresbull



Research report

Feasibility of brain age predictions from clinical T1-weighted MRIs

Pedro A. Valdes-Hernandez [a,b,c,1], Chavier Laffitte Nodarse [a,b,c,1], James H. Cole [d,e], Yenisel Cruz-Almeida [a,b,c,f,*]

[a] Department of Community Dentistry and Behavioral Science, University of Florida, USA
[b] Pain Research and Intervention Center of Excellence, University of Florida, USA
[c] Center for Cognitive Aging and Memory, McKnight Brain Institute, University of Florida, USA
[d] Centre for Medical Image Computing, Department of Computer Science, University College London, UK
[e] Dementia Research Centre, Queen Square Institute of Neurology, University College London, UK
[f] Department of Neuroscience, College of Medicine, University of Florida, USA

ARTICLE INFO

Keywords:
Patients
Brain-PAD
DeepBrainNet
Brain age bias

ABSTRACT

An individual's brain predicted age minus chronological age (brain-PAD) obtained from MRIs could become a biomarker of disease in research studies. However, brain age reports from clinical MRIs are scant despite the rich clinical information hospitals provide. Since clinical MRI protocols are meant for specific clinical purposes, performance of brain age predictions on clinical data need to be tested. We explored the feasibility of using DeepBrainNet, a deep network previously trained on research-oriented MRIs, to predict the brain ages of 840 patients who visited 15 facilities of a health system in Florida. Anticipating a strong prediction bias in our clinical sample, we characterized it to propose a covariate model in group-level regressions of brain-PAD (recommended to avoid Type I, II errors), and tested its generalizability, a requirement for meaningful brain age predictions in new single clinical cases. The best bias-related covariate model was scanner-independent and linear in age, while the best method to estimate bias-free brain ages was the inverse of a scanner-independent and quadratic in brain age function. We demonstrated the feasibility to detect sex-related differences in brain-PAD using group-level regression accounting for the selected covariate model. These differences were preserved after bias correction. The Mean-Average Error (MAE) of the predictions in independent data was ~8 years, 2–3 years greater than reports for research-oriented MRIs using DeepBrainNet, whereas an R^2 (assuming no bias) was 0.33 and 0.76 for the uncorrected and corrected brain ages, respectively. DeepBrainNet on clinical populations seems feasible, but more accurate algorithms or transfer-learning retraining is needed.

1. Introduction

In recent years, there has been an explosion of machine learning methods for the prediction of brain age based on brain structural Magnetic Resonance Images (MRIs) (Bashyam et al., 2020; Cole and Franke, 2017; Yin et al., 2023). This methodology has already proven its potential as a biomarker of brain aging and disease in several research studies (Chen et al., 2022; Christman et al., 2020; Cole et al., 2018a; Elliott et al., 2021; Franke and Gaser, 2019; Jawinski et al., 2022; Koutsouleris et al., 2014; Millar et al., 2023; Montesino-Goicolea et al., 2022b; Wei et al., 2022). The difference between the predicted brain age and the chronological age [namely, the 'brain-PAD' or 'brain age gap' (BAG)] can be affected by pathologies and/or lifestyle factors. Specifically, the higher the positive brain-PAD or BAG values, the poorer the health of the person, and the higher the risk of health deterioration; whereas the lower the negative brain-PAD or BAG, the healthier the person.

Most brain aging applications utilize three dimensional (3D) T1-weighted (T1w) brain MRIs obtained from databases primarily devised for research purposes. This is the natural choice given most of the brain age prediction methods were trained using data from research projects. However, this leaves out the highly rich and diverse clinical information that a massive amount of daily brain MRIs acquired at a myriad of hospital across the world can offer. Clinical brain MRI protocols are usually configured for specific clinical purposes and often lack the quality of the research MRIs that were used for training most of the

* Correspondence to: 1329 SW 16th Street, Ste. 5180 (zip 32608), Gainesville, FL 32610, USA.
E-mail address: cryeni@ufl.edu (Y. Cruz-Almeida).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.brainresbull.2023.110811
Received 31 May 2023; Received in revised form 7 November 2023; Accepted 8 November 2023
Available online 10 November 2023

above-cited brain age prediction methods.

The recent study by Wood et al. (2022) stands out among the paucity of brain age prediction methods applied to clinical MRIs. They used clinical-grade T2 weighted (T2w) and diffusion-weighted (DWI) axial MRIs to train a convolutional neural network (CNN) for predicting brain age. However, considering the increasing use of T1w MRIs with relatively high resolution and nearly isotropic voxel sizes in clinical settings, as suggested by the clinical data obtained from our medical system, we propose that existing T1w-based brain age prediction methods can now be applied to clinical cohorts.

In this investigation, we explored the feasibility of using an already trained CNN, DeepBrainNet (Bashyam et al., 2020), to predict brain age from clinical brain MRIs. DeepBrainNet is based on the Visual Geometry Group 16 (VGG16) (Simonyan et al., 2014), and it was trained on a large and heterogeneous T1w MRI dataset of 11,729 participants from 18 studies spanning different scanners, ages, and locations. Due to its moderately "loose" training, which avoids an exclusive prioritization of maximum prediction accuracy, DeepBrainNet promises to be sensible to clinical conditions. Our sample was a set of clinical brain T1w MRIs, distributed among 8 different scanners, from 840 patients that were scanned in 15 medical facilities within our university healthcare system (UFHealth) in Florida, USA.

When predicting brain age, the "regression toward the mean" phenomenon (Galton, 1886) yields a distribution of brain ages where younger ages tend to be overestimated and older ages underestimated. If not accounted for, this age-related bias can affect group-level analyses using brain-PAD as the dependent variable via two possible mechanisms: 1) a spurious correlation between brain-PAD and the independent variable of interest driven by correlations between the latter and chronological age (Le et al., 2018) (Type I Error) and 2) an excess in age-related variance in brain-PAD that attenuates the estimated effect size of the independent variable of interest (Beheshti et al., 2019) (Type II Error). That is why, under the assumption that the bias is linear, the literature recommends to add chronological age as a covariate in group analyses to minimize the residuals (Beheshti et al., 2019; de Lange and Cole, 2020). In our clinical data, the bias resulting from the use of DeepBrainNet needs to be characterized so we can determine the appropriate covariate terms (the covariate model) needed in future group-level regressions of the brain-PAD. For example, we do not know whether other factors like the type of scanner may play a role in such bias.

The above-described statistically-based approach to account for the bias is not possible in single-case or smaller samples, and an approach to directly correct the brain ages is needed. This requires testing whether a model of brain age correction is generalizable and can thus be used to obtain unbiased brain ages in new samples. Additionally, correcting the bias enables report of unbiased summary statistics (e.g., the mean) of brain ages in certain subsamples of interest. At the expense of a potential loss in prediction accuracy, we propose methods to obtain unbiased brain ages that do not inflate this accuracy, a risk posed by some correction efforts in the literature (Butler et al., 2021).

The aims of the present study are to: 1) determine the feasibility of employing clinically collected T1-weighted MRIs using a research-derived and already trained algorithm for brain age prediction (i.e., DeepBrainNet) and characterize its age-related bias; 2) propose a generalizable parsimonious covariate model for group-level regression; 3) propose a method to obtain individual unbiased brain ages in new clinical samples; 4) determine whether it is possible to detect, using group-level regression accounting for the bias, the sex-related differences in brain aging that has been previously reported in the literature (Beheshti et al., 2021; Goyal et al., 2019; Király et al., 2016; Sanford et al., 2022) and 5) whether these sex-related differences are preserved after the correction.

We hypothesized that 1) DeepBrainNet can estimate brain ages from clinical T1-weighted MRIs with an accuracy comparable to that reported in research studies, 2) the most generalizable model of brain age bias is linear in age and scanner-independent, 3) it is possible to obtain unbiased brain ages in newer samples with an accuracy at least similar to that of the uncorrected brain ages, 4) after accounting for the bias, brain-PAD is higher in males compared to females, and 5) this difference is preserved after the bias correction. Our sex-related hypothesis is based on recent findings in the brain age literature (Beheshti et al., 2021; Goyal et al., 2019; Király et al., 2016; Sanford et al., 2022) supporting the hypothesis that female brain neoteny is present in young adults and persists throughout the adult life span (Goyal et al., 2019).

2. Materials and methods

2.1. Participants and MRI data

All participants, or their legal guardians, gave informed consent for their clinical data to be used for research purposes. MRI acquisition was carried out after all participants completed a screening form to determine MRI eligibility. To test the feasibility of applying machine learning algorithms to derive brain aging biomarkers from MRIs obtained from electronic medical records from clinical visits, we requested and obtained IRB approval to receive 30,000 unique clinical images. The Institutional Review Board of the University of Florida IRB01 approved the study as non-Human exempt since the study team received coded and de-identified data from the University of Florida Integrated Data Repository with a confidentiality agreement put in place. The study was approved under IRB202101469 protocol on 8/20/2021.

2.2. Brain age prediction

DeepBrainNet is a convolutional neural network recently developed to predict brain age (Bashyam et al., 2020). It was trained using the slices of the T1w MRI images from 11,729 individuals (ages 3–95 years) from a diverse range of geographic locations, scanners, acquisition protocols, and studies, and tested in an independent sample of 2739 individuals. In this study, we used the version of DeepBrainNet that is available online, which is based on the VGG16 (Simonyan et al., 2014) to predict and evaluate prediction accuracy in our clinical MRI data (objective 1 of the Introduction).

Features for the DeepBrainNet are calculated as follows. First, the T1w needs to be skull-stripped. Second, the skull-stripped image has to be spatially normalized to the 1-mm isotropic voxel FSL skull-stripped T1w template using a 12-parameter linear affine transformation. For training, each of the skull-stripped MRIs was divided into 80 2D slices (centered on the $z = 0$ plane in MNI coordinates) and considered as an independent sample. To obtain a final age prediction for a test sample, each of 80 slices of the test scan is input to the trained model independently and the median prediction is calculated as the subject's predicted brain age. To obtain skull-stripped images in our sample, we used *smriprep*.[2] Briefly, the T1w image was corrected for intensity non-uniformity using *N4BiasFieldCorrection* (Tustison et al., 2010) distributed with ANTs 2.2.0 and skull-stripped with a Nipype implementation of the *antsBrainExtraction.sh* workflow from ANTs (Avants et al., 2009), using OASIS30ANTs as target template. Finally, images values were scaled from 0 to 255.

2.3. MRI/preprocessing quality control

Preprocessed MRI images were submitted to a careful quality control (QC) procedure. Only a subset of the preprocessed MRIs that were likely to have bad quality were visually inspected. The selection of this subset of MRIs was carried out as follows. We calculated the normalized mutual information (NMI) (Studholme et al., 1998) between the preprocessed MRIs and the 1-mm isotropic voxel FSL skull-stripped T1w template. We

[2] https://www.nipreps.org/smriprep/usage.html

then plotted the histogram of the NMIs and visually defined a threshold based on those values appearing to be significantly below the main unimodal distribution. We inspected all images below this threshold and those above it until they had no visible preprocessing errors. Since the goal is to demonstrate feasibility of the brain age estimation for clinical images, which have generally less quality than those intended for research purposes, we were lenient regarding the consideration of what a "processing error" was. We only removed preprocessed MRIs that were indisputably unrecognizable due to motion, a brain extraction that did not remove significant non-brain tissues, a normalization that performed poorly, or the presence of structural abnormalities, e.g., tumors, stroke lesions, tissue loss, but not atrophy or ventricular enlargement unless extremely pronounced.

2.4. Modeling the bias

In response to part of our objective 1, we considered a bias model with generic form $brain\ age = f_{\beta}(age) + \varepsilon$, where *age* denotes the chronological age, ε is a Gaussian noise and $f_{\beta}(age)$ is a function of *age* that depends on a set of regression coefficients contained in the vector $\boldsymbol{\beta}$. We proposed the following forms, written in Wilkinson notation for the sake of simplicity:

1. $f_{\beta}(age) = age$, where $\boldsymbol{\beta}$ contains 2 elements;
2. $f_{\beta}(age) = age^2$, where $\boldsymbol{\beta}$ contains 3 elements;
3. $f_{\beta}(age) = age * scanner$, where $\boldsymbol{\beta}$ contains 16 elements; and
4. $f_{\beta}(age) = age^2 * scanner$, where $\boldsymbol{\beta}$ contains 24 elements;

We shall denote $\widehat{\boldsymbol{\beta}}$ as the estimated values of the regression coefficients and $f_{\widehat{\beta}}(age)$ the estimated function characterizing the bias. For example, for model 1, $\widehat{\boldsymbol{\beta}} = [\widehat{\beta}_{intercept}, \widehat{\beta}_{linear}]$ contains the estimated intercept and slope, and $f_{\widehat{\beta}}(age) = \widehat{\beta}_{intercept} + \widehat{\beta}_{linear} age$ is its estimated algebraic form.

2.5. Methods for bias-correcting the brain ages

The use of age as a covariate in group-level analyses commonly seen in the literature invites the following correction of the brain age to remove the bias:

$$corrected\ \ brain\ \ age = age + brain\ \ age - f_{\widehat{\beta}}(age) \quad (1)$$

However, even though using $f_{\widehat{\beta}}(age)$ as a covariate in group regressions of brain-PAD is appropriate to account for unwanted age-related variance, the presence of chronological age-dependent terms in Eq. (1) misleads the accuracy of the unbiased brain ages obtained using this equation. For example, for model 1 above, i.e., $f_{\beta}(age) = age$ in Wilkinson notation, Butler et al. (2021) theoretically and empirically demonstrated that the correlation between the chronological ages and the corrected predicted brain ages obtained via Eq. (1), $r_{corrected\ brain\ age,age}$, is inflated and never below ~0.87, even if there is no relationship between the MRIs and age at all (Butler et al., 2021). And this lower bound is when the sample used to estimate the coefficients is not the same as the sample for which brain age is corrected. If the same sample is used for both, the situation worsens, as $r_{corrected\ brain\ age,age} \geq$ ~0.9177 (Butler et al., 2021). Consequently, the mean absolute error (MAE) of the corrected brain ages is also spuriously lower than the MAE of the uncorrected ones.

We thus propose the following chronological age-independent correction:

$$corrected\ \ brain\ \ age = f_{\widehat{\beta}^{-1}(brain\ \ age)} \quad (2)$$

Note that Eq. (2) is a generalization of the correction proposed by Cole et al. (2018b) for model 1, i.e., $corrected\ brain\ age = \frac{brain\ age - \widehat{\beta}_{intercept}}{\widehat{\beta}_{linear}}$. In this particular case, there is no risk of overestimation of the accuracy since $r_{corrected\ brain\ age,age}$ is identical to $r_{brain\ age,age}$, whereas the standard deviation of the corrected brain ages is $1/\widehat{\beta}_{linear}$ times higher than the standard deviation of the uncorrected brain ages. Note that, for model 4, a quadratic equation should yield two real solutions from which one has to be discarded based on the range of realistic age values—in very rare cases two complex conjugated are produced and those data points have to be discarded. Corrections based on models 3 and 4 are equivalent to models 1 and 2, respectively, if done for each scanner independently.

2.6. Model selection

Models including nonlinear age-related and/or scanner-related dependencies may not necessarily generalize well in newer samples due to the risk of overfitting related to over-parameterization. Thus, with the objectives outlined as follows, we selected the model that maximized the trade-off between parsimony and the similarity between:

a) *age* and $f_{\widehat{\beta}}(age)$, to recommend the covariate model to account for the bias in group analysis (objective 2); and
b) *age* and $corrected\ brain\ age = f_{\widehat{\beta}^{-1}(brain\ age)}$, to recommend the formula to correct the bias (objective 3).

Parsimony is important to minimize the information needed (e.g., scanner used) in clinical applications. For a) parsimony is also important to maximize power when the sample size is small, since it minimizes the loss in degrees of freedom; while for b), parsimony warrants generalizability. Thus, our goodness-of-fit measure was the Akaike Information Criterion (AIC), which favors maximum likelihood while penalizing the complexity of the model (given by the number of parameters in $\widehat{\boldsymbol{\beta}}$)—the lower the AIC the better the fit. We shall denote $AIC_a(\widehat{\boldsymbol{\beta}})$ and $AIC_b(\widehat{\boldsymbol{\beta}})$ the estimated AIC values for a) and b) above.

We divided the subjects into a training and a test sample, containing 80 % and 20 % of the subjects, respectively. The former was used for model selection, and the latter to evaluate generalization of the correction, i.e., the accuracy of the model in predicting the unbiased brain age in newer samples. Model selection was performed as follows. We divided the training sample (the subsample containing 80 % of the total sample) into 5 folds, and used 4 folds to estimate the parameters of the model ($\widehat{\boldsymbol{\beta}}_{iteration}$) and the remaining fold to calculate $AIC_a(\widehat{\boldsymbol{\beta}}_{iteration})$ and $AIC_b(\widehat{\boldsymbol{\beta}}_{iteration})$. All of the abovementioned splits of the subjects were stratified so the samples had roughly the same chronological age distribution. To ensure reproducibility, the samples were generated with a fixed seed at the beginning of the study.

After cross-validation, we selected the model with the minimum average AIC across fold iterations and fitted it to the whole training data to obtain the final estimates of the parameters, i.e., $\widehat{\boldsymbol{\beta}}_{selected}$. We compared the AICs (a or b) using the relative likelihood, i.e., $RL_m = \exp\left[-\frac{1}{2}(AIC_m - AIC_{selected})\right]$, where AIC_m is the AIC of the $m-th$ model and $AIC_{selected}$ is the minimum AIC. RL_m quantifies how many times the $m-th$ model is more likely than the model with minimum AIC, given the data.

Finally, we corrected the predicted brain ages of the test sample by using $\widehat{\boldsymbol{\beta}}_{selected}$ in Eq. (2) and evaluated the accuracy of the predictions

using the MAE, the correlation $r_{corrected\ brain\ age,age}$, and the coefficient of determination $R^2_{y=x} = 1 - \frac{\sum (y_i - x_i)^2}{\sum (y_i - \bar{y})^2}$, where y_i and x_i is the corrected brain age and chronological age, respectively, for the $i-th$ participant in the test data, i.e., assuming the "perfect" (though incorrect) unbiased model $y = x$. The 95 % confidence interval (CI) of these accuracy measures was calculated using 10,000 bootstraps.

2.7. Evaluating sex-related group differences

To test whether we can detect sex-related difference in brain-PAD in our clinical sample (objective 4) after controlling for the selected covariate model, $f_{\hat{\beta}_{selected}}(age)$, we fitted the models $brain\ PAD \sim f_{\hat{\beta}_{selected}}(age) + sex$ and $brain\ PAD \sim sex$. We also evaluated how much of the variance is explained after adding the covariate terms in $f_{\hat{\beta}_{selected}}(age)$ via an Analysis of Deviance. The deviance for each model was calculated as twice the sum of the log of the squared residuals and a likelihood ratio test, i.e., the difference in deviances, was conducted to compare the two models. This test follows a χ^2 distribution under the null hypothesis that the simpler model $brain\ PAD \sim sex$ is true. On the other hand, to test whether the sex-related difference in brain-PAD was preserved after the bias correction (objective 5), we fitted the model $corrected\ brain\ PAD \sim sex$, where $corrected\ brain\ PAD = f_{\hat{\beta}^{-1}(brain\ age)} - age$. This can be done under the premise that the corrected brain-PAD is unbiased and a covariate term depending on the chronololgical age is not needed. Thus, we also assessed the correlation between corrected brain-PAD and chronological age.

To ensure normality in these regressions, we applied a rank-based inverse normal transformation to the dependent variable *brain-PAD* using the 'Blom' method (parameter c= 3/8 (Downton and Blom, 1961)), modified to preserve the mean of the values. Furthermore, after a first fit, we removed those measurements deemed outliers, based on their Cook's distance being 3 times higher than their sample average (NETER, 1990) and reran the models. Also, after fitting the models, we applied the Shapiro-Wilk test of composite normality (with unspecified mean and variance) on the residuals (for Platykurtic distributions; while the Shapiro-Francia test was used for Leptokurtic distributions) to test whether the normality assumption required for linear models was fulfilled (Shapiro and Wilk, 1965).

3. Results

3.1. Final sample

We received the raw DICOMS from 24,732 MRIs of 1727 patients from the hospital. After removing all non-brain, partial-brain and other images and performing QC, we had a total of 8040 whole-brain MRIs of several modalities [e.g., T2w, FLuid Attenuated Inversion Recovery FLAIR, etc.] from 1543 patients. Fig. 1 reveals that T1ws of nearly isotropic, maximum voxel dimension of 1.2 mm have been increasingly included in clinical MR protocols in recent years. Therefore, our final sample only included the T1w of 840 participants, allowing us to evaluate the accuracy of an existing T1w-based brain age prediction method in our clinical sample.

Patients were scanned with eight different MRI scanners (see Table 1 for summary statistics of some MRI parameters, the number of subjects per scanner and demographics). In the final sample, 554 were females and 286 were males. The average chronological age of the males was 2.7 years older than that of the females (p = 0.042). Also, the chronological age of the total sample ranged from 15 to 95 years, with a median, mean, and standard deviation of 57.5, 54.3 and 18.2 years, respectively.

3.2. Uncorrected brain age predictions

Using DeepBrainNet, we predicted brain age in our sample of 840 clinical T1-weighted MRIs. These predictions are shown in Fig. 2. The bias in the predictions that overestimates younger ages and underestimates older ages is exposed by the lines representing the slope of the linear relation between the chronological and the predicted brain age for each scanner. In order to test our first hypothesis, we evaluated measures of accuracy of these predictions. The MAE (95 % CI) was 8.05 ([7.60, 8.52]) years, $r_{brain\ age,age}$ (95 % CI) was 0.87 ([0.84, 0.88]), and

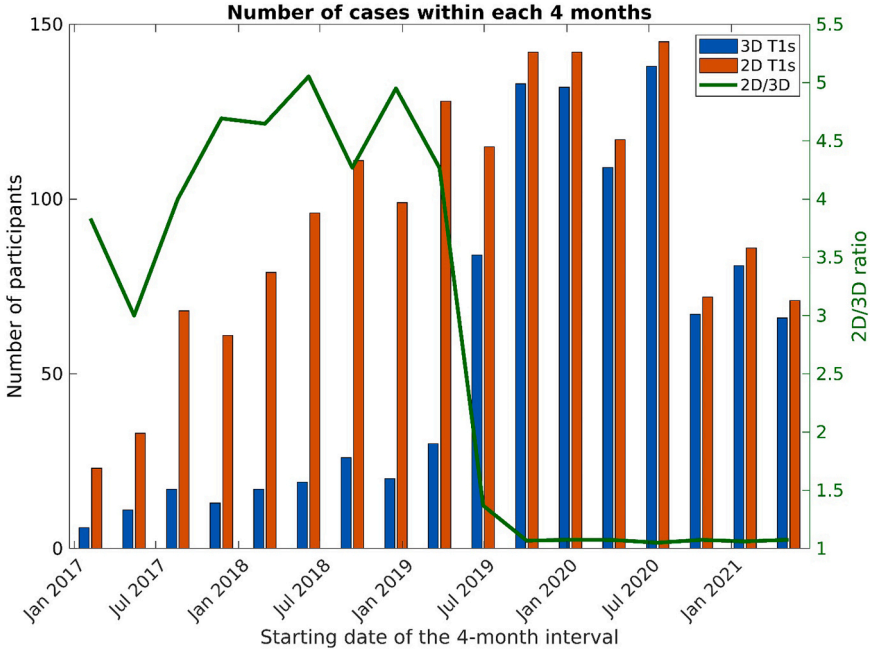


Fig. 1. Comparison of the number of clinical 3D T1w MRIs versus clinical 2D T1w MRIs over time. The data shows that the former sequence is now part of the clinical imaging protocol.

Table 1
Scanners used for MRI acquisition.

Scanner	Manufac-turer	Field Strength [T]	Voxel sizeMean (SD)	Voxel sizeMin.-Max.	TE [ms.] Mean (SD)	AxialCoronalSagittal	Number of subjects	Number of Females	Age (years) Mean (SD)
Aera	Siemens	1.5	[0.9 (0.1), 0.8 (0.2), 0.8 (0.2)]	[0.5–1.0, 0.4–1.0, 0.4–1.2]	2.5 (0.3)	221250	147	90	57.9 (17.3)
Avanto	Siemens	1.5	[0.9 (0.1), 0.9 (0.1), 1.0 (0.1)]	[0.5–1.0, 0.5–1.0, 0.9–1.2]	3.3 (0.6)	581230	181	112	53.3 (18.1)
Prisma	Siemens	3.0	[0.9 (0.1), 0.8 (0.2), 0.8 (0.2)]	[0.4–1.0, 0.4–1.0, 0.4–1.0]	2.3 (0.1)	331290	162	111	54.7 (18.2)
Sola	Siemens	1.5	[0.9 (0.0), 0.8 (0.2), 0.8 (0.2)]	[0.9–0.9, 0.4–0.9, 0.4–0.9]	2.6 (0.2)	1430	44	34	49.4 (18.0)
SignaHDxt	GE	1.5	[0.9 (0.1), 0.9 (0.1), 1.0 (0.1)]	[0.5–1.2, 0.5–1.1, 0.5–1.1]	2.8 (0.3)	12600	72	58	52.5 (19.1)
Skyra	Siemens	3.0	[0.9 (0.1), 0.7 (0.2), 0.7 (0.2)]	[0.5–1.1, 0.4–1.1, 0.4–1.0]	2.3 (0.1)	2220	24	14	58.2 (16.6)
Titan3T	Toshiba	3.0	[0.6 (0.2), 0.9 (0.3), 0.8 (0.3)]	[0.5–1.2, 0.5–1.2, 0.5–1.2]	3.1 (0.6)	12180	30	19	51.5 (19.4)
Verio	Siemens	3.0	[0.9 (0.1), 0.8 (0.2), 0.9 (0.2)]	[0.5–1.2, 0.4–1.0, 0.4–1.2]	3.3 (2.0)	401400	180	116	54.0 (18.2)

Note. Siemens: Siemens Healthineers. GE: General Electric Medical Systems. SD: Standard Deviation. Values were calculated across the final sample. Min.: minimum. Max.: maximum. TR: Repetition time. TE: Echo time.

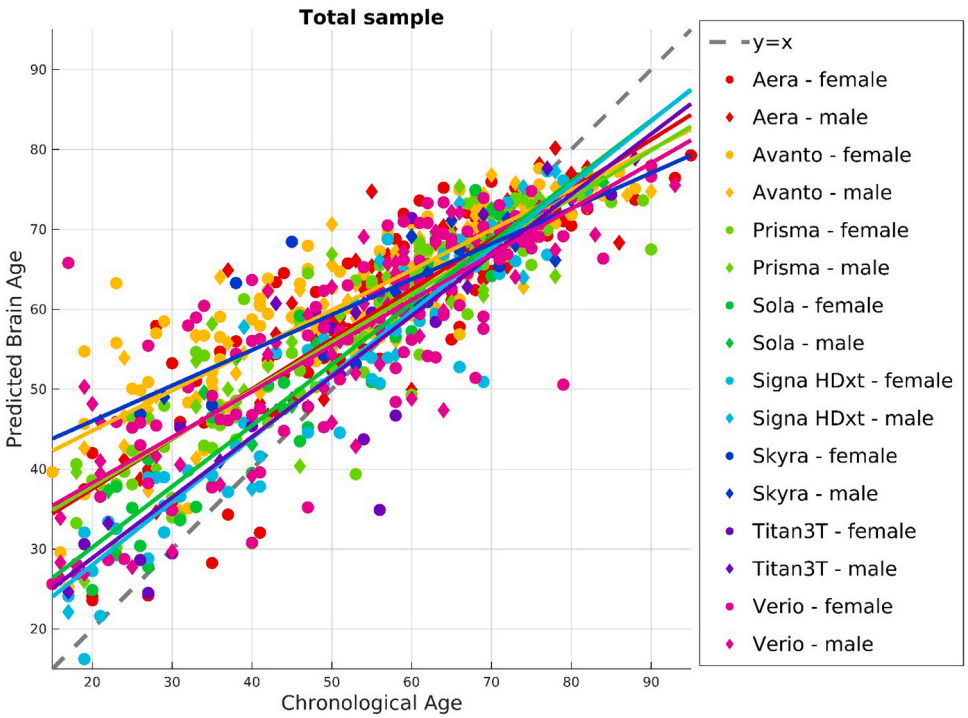


Fig. 2. Brain age prediction for all MRIs in this study. The colored lines represent the slope of the linear relation between the chronological and the predicted brain age for each scanner (see legend for color-scanner relation).

$R^2_{y=x}$ (95 % CI) was 0.33 ([0.21, 0.42]). The results for each scanner are reported in **Table S1** of the Supplemental Materials.

3.3. Characterization of the bias

Before model selection, we did a quantitative characterization of the bias to determine whether is nonlinear in age or scanner-dependent (this responds to part of our objective 1). We did this by fitting the models to the training sample to keep the test sample untouched and avoid influencing our judgement about the models. Table 2 reports the general results of these fits.

Below, we report some important coefficients and tests of nonlinearity and scanner dependence:

Table 2
Results of the regression fits in the training sample (n = 672).

No. Model's $f_\beta(age)$	Degrees of Freedom	RMSE	R^2	Adjusted R^2	F-statisticvs. constant model	p-value
1. *age*	670	6.43	0.743	0.743	1.94e3	4.88e-200
2. age^2	669	6.38	0.748	0.747	992	7.75e-201
3. *age* ∗ *scanner*	656	5.95	0785	0780	160	3.76e-207
4. age^2 ∗ *scanner*	648	5.92	0789	0782	106	1.32e-201

Note. RMSE: Root Mean Squared Error. R^2: Coefficient of determination of the actual fit (this is not $R^2_{y=x}$). Models are written in Wilkinson's notation.

Table 3
AIC and accuracy measures the evaluation folds, averaged across iterations, as well as the corresponding likelihood relative to the selected model.

No. Model's $f_\beta(age)$	Selection of the covariate model		Selection of correction formula		Accuracy of unbiased brain ages		
	AIC_a	REL_a	AIC_b	REL_b	MAE (years)	$r_{\text{corrected brain age, age}}$	$R^2_{y=x}$
1. *age*	884.8*	-	1014.9	0.018	7.80	0.86	0.74
2. age^2	885.1	0.855	1006.9*	-	7.69	0.87	0.75
3. *age* ∗ *scanner*	893.4	0.014	1032.8	2.35e-6	7.50	0.87	0.76
4. age^2 ∗ *scanner*	915.2	2.52e-7	1046.8	2.16e-9	7.64	0.88	0.76

Note. * indicates the selected model. AIC: Akaike Information Criterion. REL: Relative Likelihood. MAE: Mean Average Error. $r_{x,y}$: Correlation between x and y. $R^2_{y=x}$: Coefficient of determination assuming the model $y = x$. Models are written in Wilkinson's notation.

1. $f_\beta(age) = age^2$. The intercept was $\widehat{\beta}_{intercept} = 25.84$ years, and was significantly different from 0 (p = 8.0e-137), and the slope was $\widehat{\beta}_{linear} = 0.6$, and was significantly lower than 1 (p = 1.4e-111).
2. $f_\beta(age) = age^2$. After fitting $f_\beta(age) = \beta_{intercept} + \beta_{linear} age + \beta_{quadatric} age^2 + \varepsilon$ to the whole training dataset, the intercept was $\widehat{\beta}_{intercept} = 19.21$ years, and was significantly different from 0 (p = 8.4e-23), the linear coefficient was $\widehat{\beta}_{linear} = 0.885$, and it was not significantly different from 1 (p = 0.149), and the quadratic coefficient was $\widehat{\beta}_{quadratic} = -2.65e\text{-}3$ and it was significantly different from 0 (p = 6.27e-4).
3. $f_\beta(age) = age * scanner$, we found evidence of a difference in slopes among scanners i.e., $H_0 : \widehat{\beta}_{linear,Aera} = \ldots = \widehat{\beta}_{linear,Verio}$ was rejected (F = 6.63, p = 1.34e-7).
4. $f_\beta(age) = age^2 * scanner$. We found no evidence of scanner-related dependency, i.e., $H_0 : \widehat{\beta}_{linear,Aera} = \ldots = \widehat{\beta}_{linear,Verio}$ (F = 0.92, p = 0.488) and $H_0 : \widehat{\beta}_{quadratic,Aera} = \ldots = \widehat{\beta}_{quadratic,Aera}$ (F = 0.63, p = 0.734) were not rejected. For this model, the main nonlinear effect was also not significant, i.e., $H_0 : \langle \widehat{\beta}_{quadratic,scanner} \rangle_{scanner} = 0$ was rejected (F = 1.87, p = 0.172).

3.4. Selecting the best covariate model for group-level analyses

We performed 5-fold cross-validation in 80 % of the sample (training sample). To avoid biases due to domain mismatches, all data groups (folds and test set) were created having similar distributions of ages (see Figs. S1–3 in the Supplemental Materials). In order to test our second hypothesis, we computed the values of AIC_a, averaged across the evaluation folds, for the four regression models. The first column of Table 3 shows these AIC_a values (see next section for the remaining columns). Having the minimum average AIC_a, we selected model 1, i.e., *brain age* ∼ *age*, as the best covariate model.

3.5. Unbiased (corrected) brain age predictions

In order to test our third hypothesis, we used the same procedure described in the previous section, but this time to evaluate AIC_b. Having the minimum average AIC_b, we selected model 2, i.e., *brain age* ∼ age^2, to calculate the unbiased brain ages via Eq. (2): *corrected brain age* = $f_{\hat{\beta}^{-1}(brain\ age)}$. For each uncorrected brain age, this inversion yielded two real quadratic roots, and one was discarded for being unrealistic. The second column of Table 3 shows the average AIC_b, and the third column shows, for illustration purposes, the prediction performance of the corrected brain ages, also averaged across evaluation folds, if complexity of the model is not taken into account.

We then reported the generalization accuracy of the bias-corrected brain age using the held-out 20 % of the sample (test sample). The plot of the corrected brain age versus chronological age in this subsample is shown in Fig. 3. The MAE (95 % CI) of these predictions was 8.12 ([7.19, 9.29]) years, $r_{corrected\ brain\ age,age}$ (95 % CI) was 0.88 ([0.82, 0.91]), $R^2_{y=x}$ (95 % CI) was 0.76 ([0.67, 0.83]). Also, the average corrected brain-PAD was not significantly different from zero (one-sample t-test, p = 0.421); while the correlation between the corrected brain-PAD, i.e., *corrected brain age* − *age*, and chronological age was −0.02 and not significantly different from zero (p = 0.595). More results for the corrected brain age in the test sample for each scanner can be found in **Table S2** of the Supplemental Materials (**Table S3** shows the same but for the uncorrected predictions for comparison).

Additionally, we found evidence of a residual moderation by scanner of the relationship between the chronological age and the corrected brain age, i.e., $H_0 : \widehat{\beta}_{linear,Aera} = \ldots = \widehat{\beta}_{linear,Verio}$ was rejected (F = 2.34, p = 0.0272). This omnibus difference may be explained by a difference in slope between the Signa HDxt and the Avanto (p = 0.0029, uncorrected), the Prisma (p = 0.0073, uncorrected) or the Verio (p = 0.0292, uncorrected). However, we do not have evidence of this since none of these comparisons survived Bonferroni correction across all pairs.

3.6. Sex-related differences in clinical brain-PAD

Since we selected model 1 as the covariate model, we fitted *brain PAD* ∼ *age* + *sex* to our data to test our fourth hypothesis. We found that brain-PAD was 1.4 years (1.75 years after outlier removal) significantly (p = 0.002) higher for males than for women. This is shown in Fig. 4. We also fitted *brain PAD* ∼ *sex* to our data and found no significant sex-related difference in brain-PAD (ΔPAD = 0.42 years, p = 0.546). When comparing the deviances between this model and the null hypothesis model *brain PAD* ∼ *sex* via a likelihood ratio test, the p-

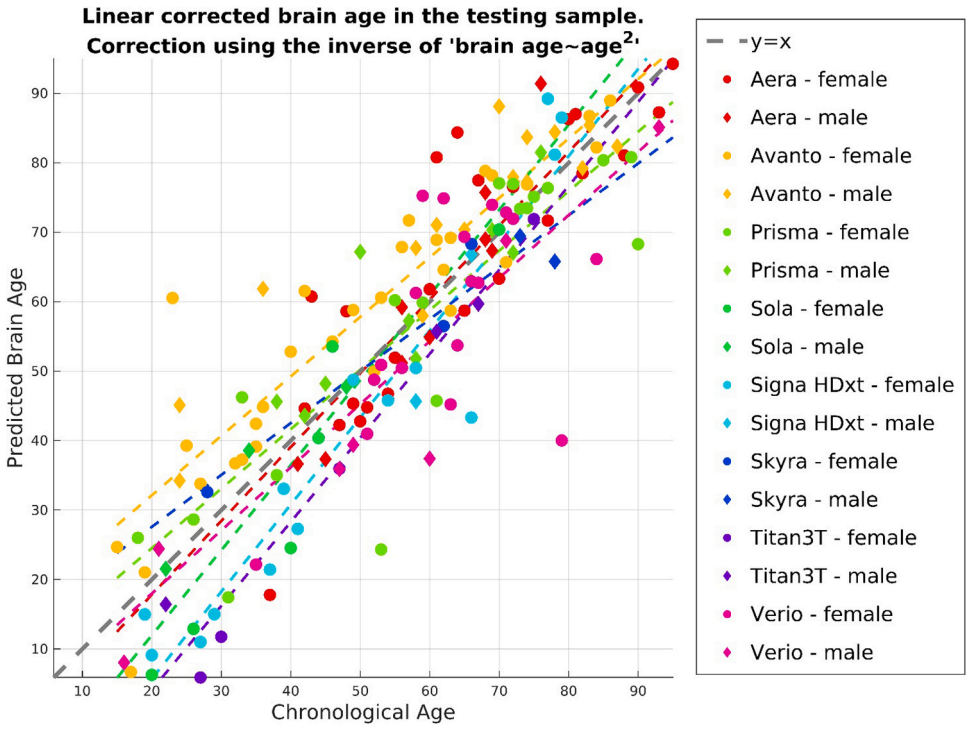


Fig. 3. Brain age prediction, corrected for the linear bias using the model *brain age* ~ *age*^{2} (in Wilkinson's notation) or the MRIs in the test (held-out) sample. The colored lines represent the slope of the linear relation between the chronological and the corrected predicted brain age for each scanner (see legend for color-scanner relation).

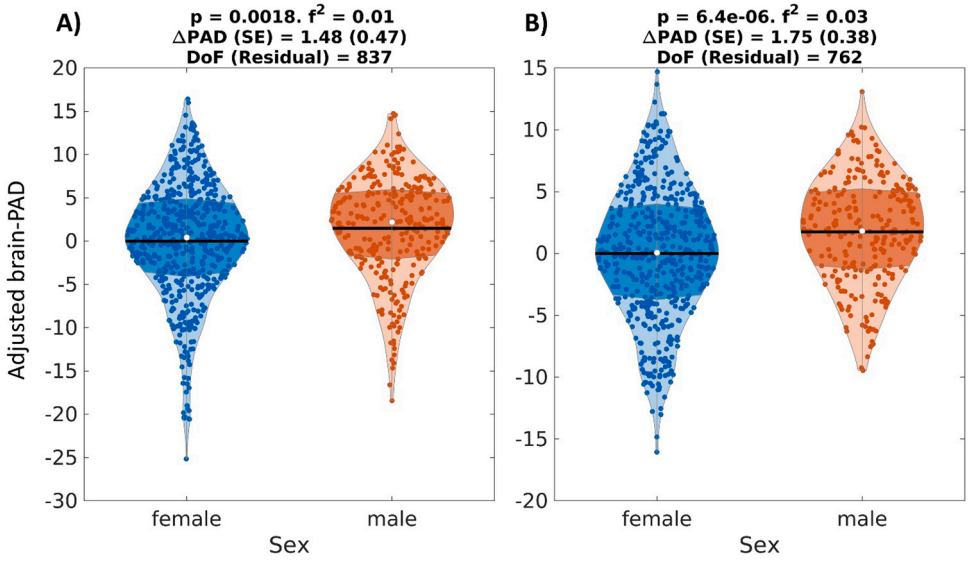


Fig. 4. Differences in brain-PAD between sexes, adjusting for age for A) the whole sample and B) after removing outliers. Within the violin plots, the shaded area is the interquartile region, the white dot indicates the median and the black horizontal line is the mean. Effect sizes, as quantified by the Cohen's f^2, are also shown. DoF: Degrees of Freedom. ΔPAD (in years): difference in brain-PAD across groups. SE (in years): Standard Error.

value was near zero (less than MATLAB's precision limit). These results confirm that adding *age* to the model significantly reduces the variance, allowing the effect of *sex* to be detectable.

On the other hand, since the corrected brain-PAD did not depend on chronological age, i.e., was unbiased, we fitted *corrected brain PAD* ~ *sex* to our data to test our fifth hypothesis. We found that the sex-difference found in the first analysis was preserved, since the corrected brain-PAD was 2.31 years (2.47 years after outlier removal) significantly ($p = 0.0017$) higher for males than for women. This is shown in Fig. 5. As above-mentioned, correcting the bias enables report of unbiased values of the mean of brain ages or brain-PAD for each sex, allowing us to understand what sex is actually deviating from an ontogenetic brain aging. Here, the estimated mean of the corrected brain-PAD in females was not significantly different from zero, i.e., −0.47 with $p = 0.275$, whereas that of males was significantly higher than zero, i.e., 1.73 years with $p = 0.0038$. Using the coefficients of this model, we also determined that the estimated mean of the corrected brain-PAD was not significantly different from zero, i.e., 0.63 years with $p = 0.0854$ (because of imbalance in the size of the sex groups, the estimated mean is not equal to the actual mean) mirroring the results of

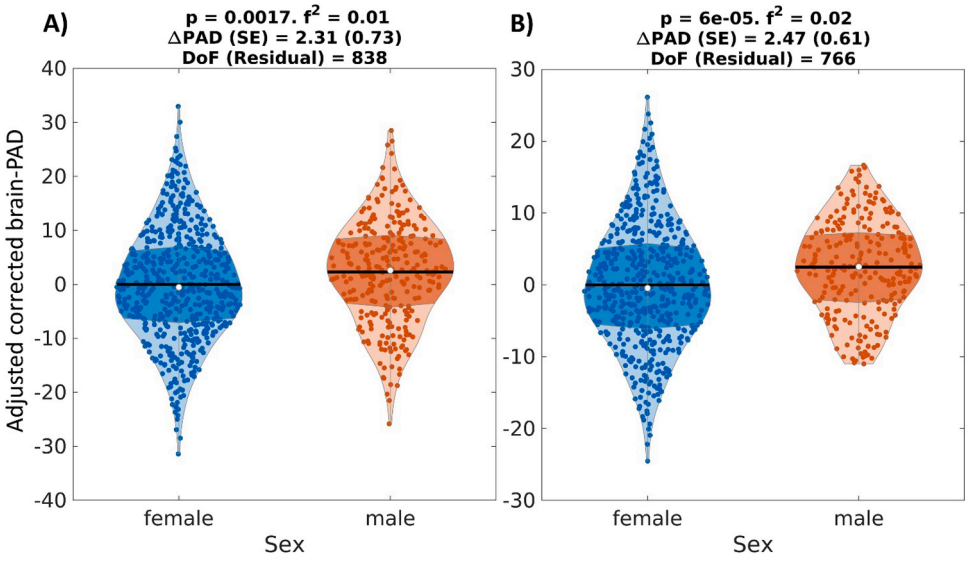


Fig. 5. Differences in corrected brain-PAD between sexes for A) the whole sample and B) after removing outliers. Within the violin plots, the shaded area is the interquartile region, the white dot indicates the median and the black horizontal line is the mean. Effect sizes, as quantified by the Cohen's f^2, are also shown. DoF: Degrees of Freedom. ΔPAD (in years): difference in corrected brain-PAD across groups. SE (in years): Standard Error.

the one-sample t-test above. This suggests that males were the group with older appearing brains.

In all analyses, rejection of composite normality failed after correcting for multiple comparisons ($p > 0.05$).

4. Discussion

In this investigation, we have explored the feasibility of using clinical brain MRIs for brain age prediction based on convolutional neural networks (CNNs), in particular using DeepBrainNet. Our first goal was to compare the performance of these predictions with those from research MRI data in the literature. The MAE (95 % CI) of our clinical data (n = 840) was 8.05 ([7.60, 8.52]) years. This value was higher than the MAE of 4.12 years reported by Bashyam et al. (2020), the developers of DeepBrainNet, in independent (test or out-of-sample) research MRI data (n = 2739). First, we want to point out that one should not purse the lowest MAE by all means. As discussed by Bashyam et al., (2020), attempting to do that could be at the expense of removing clinically relevant brain age deviations. Second, brain age prediction accuracy metrics are study-specific, and cannot be directly compared across different studies (de Lange et al., 2022). Having said that, we discuss several possible reasons why our MAE was higher than that reported by Bashyam et al., (2020) next.

First, we used a clinical sample. Thus, we would expect higher prediction errors associated to the presence of many clinical conditions. Also, our sample was smaller than that of Bashyam et al. (2020), and that could have led to some overestimation of the MAE (de Lange et al., 2022), though our narrower age range (our 15–95 years versus their 3–93 years) may have offset this (de Lange et al., 2022). Moreover, the DeepBrainNet architecture used by Bashyam et al. (2020) to report the out-of-sample MAE was based on the Inception-ResNet-v2 architecture (Szegedy et al., 2017), and they do not report the out-of-sample MAE for the architecture used in the current study (i.e., the VGG16). In relation to the VGG16 architecture, Bashyam et al. (2020) only report the MAE for their training data, and this MAE was higher than the MAE using the Inception-ResNet-v2 architecture on the same training data. This hints at a MAE for the VGG16 architecture used on their out-of-sample likely to be somewhere between their out-of-sample MAE of 4.12 years for the Inception-ResNet-v2 architecture and our MAE of 8.05 years for the VGG16 architecture.

Given that MAE is dependent on brain age distribution, and may vary between studies (de Lange et al., 2022), other accuracy measures like the correlation and the coefficient of determination assuming the model $y = x$ (that we denote as $R^2_{y=x}$) may be more appropriate, depending on the case (de Lange et al., 2022). Unfortunately, Bashyam et al. (2020) did not report these measures for their out-of-sample data. Fortunately, we have reported these measures in our own research T1w data after applying the same MRI preprocessing and DeepBrainNet described in this study (Montesino-Goicolea et al., 2022a; Valdes-Hernandez et al., 2023). For example, for a group of participants with and without chronic musculoskeletal pain (n = 660) the MAE was 6.43 years ([6.07, 6.82]) and $r_{brain\ age,age}$ was 0.86 ([0.83, 0.87]) and $R^2_{y=x}$ was 0.63 ([0.57, 0.68]) (Valdes-Hernandez et al., 2023). While the correlation in that study is comparable to our $r_{brain\ age,age}$= 0.87 ([0.84, 0.88]), $R^2_{y=x}$ was much higher than our value of 0.33 ([0.21, 0.42]), due to the strong bias in the current clinical data. However, the age range of that pain study was narrower, and concentrated around the median age than ours, possibly favoring their $r_{brain\ age,age}$ and $R^2_{y=x}$ to higher values compared to ours (de Lange et al., 2022).

On the other hand, our MAE was higher than that reported by Wood et al. (2022), where brain age was also predicted from clinical MRI, but using other modalities, i.e., T2w and/or DWI MRIs. This could owe to the fact their model was trained directly on a subset of clinical MRIs, while we have used a model that was initially devised for research-oriented MRIs, or maybe because they used a different deep learning model, based on DenseNEt121 (Huang et al., 2017), a 3D model that yielded a less pronounced bias in the predictions. In a recent revision of the performance of deep learning methods for brain age prediction, available as a preprint in (Dörfel et al., 2023), DeepBrainNet showed a slightly higher prediction bias than other CNNs. Future studies involving clinical data should capitalize on the ability of DeepBrainNet to be retrained with small samples via transfer learning (Bashyam et al., 2020), or employ other deep learning models.

We also wanted to characterize the age-related bias provoked by the "regression to the mean" phenomenon that overestimates younger ages and underestimates older ones. We considered the possibility that the bias was dependent on age-related nonlinearities and scanner type. We found that, while these effects are significant if modeled separately, they seem to cancel each other when considered simultaneously. Also, terms quadratic in chronological age could solely explain the bias, since when they are considered in the model, the slope of the bias is not significantly

different from 1. Nevertheless, irrespective of the presence of these nonlinearities and scanner dependencies, we considered more important to maximize the trade-off between generalizable goodness-of-fit and parsimony. That is why we selected models for specific purposes based on the Akaike Information Criterion.

We determined that the best covariate model to account for the bias when performing group-level regression of brain-PAD is the model that linearly depends on chronological age, without any scanner dependency. We remind that accounting for the bias is important to avoid spurious associations due to correlations between chronological age and the independent variable of interest (Beheshti et al., 2019) (a cause of Type I error) and to reduce unwanted age-related variance (Beheshti et al., 2019; de Lange and Cole, 2020) (a cause of Type II error). We exemplified this by exploring the effect of adding chronological age as a covariate when comparing brain-PAD among sexes. This is a good example since both Type II and Type I errors may occur. Indeed, besides the unavoidable age-related bias, chronological age significantly differed among sexes. In addition, we clarify that this selection driven by a maximization of parsimony may only prove useful when the loss in degrees of freedom due to the over-parameterization of the covariate model significantly affects the power of the second-level analysis due to a small sample size. For large enough samples sizes, one may just use the covariate model that delivers the maximum likelihood, as long as the regression model is designed adequately and we are cognizant of the dangers related to stepwise schemes (Smith, 2018).

To accomplish our fourth goal, a model with a nonlinear age-related dependency, but no scanner dependency, was use to propose the bias-correction formula. The formula has no explicit dependency on chronological age and thus does not provoke inflation of the accuracy of the predictions. In fact, the correlation between the brain age and chronological age is the same for the uncorrected and corrected cases. On the other hand, while it is not trivial to theoretically determine the effect of our proposed correction on the MAE, we expect it to be inflated, similar to the way the standard deviation of the corrected brain age scales with the inverse of the slope of the bias when using Cole et al. (2018)'s linear variant (de Lange and Cole, 2020). Thus, the fact that the MAE of our corrected predictions was only 8.12–8.05 = 0.07 years higher than the MAE of the uncorrected ones can be considered an indicator of good performance.

Also, after the bias correction, the correlation between brain-PAD and chronological age disappeared but some moderation by scanner of the association between the chronological ages and the predicted brain ages remained. Marginal evidence (based on uncorrected p-values) suggests that it involved differences between the "Signa HDxt" and some other scanners. This scanner is located in one of the 15 MRI facilities that contributed to this study. It is possible that this residual moderation could be related to some specific characteristics of that particular facility (e.g., scanner, technical, staff, data handlings). Further examination is needed to clarify this.

Our sex-related group analysis provided evidence in favor of our fourth hypothesis. That is, in our clinical sample, when controlling for the bias, males had a more accelerated brain age (by about a year and a half) than females. These results support Beheshti et al. (2021)'s case about the importance of accounting for the bias to successfully detect sex-related differences in brain age gap (Beheshti et al., 2021). Moreover, we found evidence supporting our fifth hypothesis, since these sex-related differences were preserved after bias correction.

Our sex-related results also go along findings in the literature where males have older appearing brains than females. However, there are some alarming discrepancies. For example, Goyal et al. (2019) found that females had brain age gaps lower than males, as measured using Positron Emission Tomography (PET) imaging (Lim et al., 2020), and the former was significantly lower than zero while the latter was not. They attributed it to a mediating role of hormones, sex-related differences in intrinsic cellular and metabolic systems, and immune system sexual dimorphism; involving less loss of cerebral blood flow following puberty, heightened brain glycolysis during young adulthood, less loss of protein synthesis-related gene expression during aging, and a delay in the peak transition point of brain gene expression in females (Goyal et al., 2019). Our results do not support this hypothesis since our difference is explained by males having a significant positive average corrected brain-PAD, and females having an average corrected brain-PAD that is not significantly different from zero. Rather, these values are more in agreement with Beheshti et al. (2021) and Cole et al., (2018b), where males seem to be the group with significantly higher brain age acceleration.

In terms of the size of effects, our sex-related difference in brain-PAD was smaller than that reported by Goyal et al. (2019) using PET (about 3–5 years), similar to Beheshti et al., 2021 (~1.2 years), predicted from fluorodeoxyglucose-PET, and smaller than Cole et al., 2018b(~5.58 years), predicted from MRI. While this could be related to differences in methodology, it could also owe to the clinical nature of our sample. Indeed, the data in the Beheshti et al., 2021 study suggests that neurodegeneration could reverse the sex-related differences in brain age acceleration.

4.1. Limitations and future directions

Our findings support the feasibility of predicting brain age with accuracy from T1w MRIs of relatively high and isotropic resolution. However, some limitations might affect their validity and deserve to be discussed.

First, the current study set out to evaluate the feasibility of predicting accurate brain ages and characterize their bias using MRIs acquired with clinical purposes. However, we have excluded MRIs having evident structural abnormalities, e.g., tumors, stroke lesions, tissue loss. This is because we did not want these obvious outliers to affect these objectives. With a path to feasibility now established, the next step will be to detect significant associations between brain-PAD and the presence and/or severity of structural abnormalities in an independent sample, as well as other clinical diagnoses. Given the exploratory nature of the study and our limited ethical approval (i.e., only de-identified dataset), we did not have enough information about the clinical characteristics of the patients. We are currently working on getting that information for the next study. In general, future studies with well-characterized clinical phenotypes are needed to determine how much of their possible underlying conditions are affecting the estimations.

The study is also affected by some methodological limitations. For example, age was not evenly distributed in our sample. This could have biased model selection towards a model that best corrects brain ages of participants with more frequent chronological ages in the sample. In fact, we re-ran the analysis weighting the observations according to the inverse of the frequency of the occurrence of the chronological ages (bestowing more importance to the lowest and highest ends of the age distribution) and the model including a term quadratic in age without scanner moderation (model 2) had the lowest AIC_a, but closely followed by the model 1 with $RL_a = 0.97$. Furthermore, given all our MRIs are from a single healthcare system, it is unclear whether the brain age estimation generalizes well to other clinical systems, and future studies including multi-institutional samples are needed.

Our sex-related analysis is far from spotless. The use of the same data for both selecting the covariate model and for the group-level analysis itself may raise a red flag of "data leakage". But such circularity would be related only to a deficient selection of the covariate model, which may make some dent on power and sensitivity. A more rigorous approach would be to select the covariate model in the training sample and limit the sex-related group analysis to the test sample. We in fact did this but found no significant differences in brain-PAD among sexes. However, we believe the culprit of this negative result is likely an underpowered test sample, being 5 times smaller than the whole sample (i.e., n = 168), rather than an inappropriate covariate model.

There might be an additional problem related to the sex-related

analysis. The fact that a machine learning algorithm was trained using both males and females obscures sex differences since the algorithm would account for these differences in maximizing accuracy. To overcome this, Goyal et al. (2019) propose to train the machine learning algorithm in one sex and test it in the other (Beheshti et al., 2021). Unfortunately, the online available DeepBrainNet was trained using both sexes together. However, Bashyam et al. (2020) reported that the Mean Absolute Error (MAE) did not exhibit any significant difference between sexes, regardless of whether the model was trained on data from both sexes or solely on data from the opposite sex (Bashyam et al., 2020). This suggests that, in our sample, males may indeed have older appearing brains than females. Nevertheless, we believe these results and those reported in the above-cited studies may need to be revisited in the light of these considerations.

Finally, with the increase in the use of T1w MRIs with relatively high resolution and nearly isotropic voxel sizes in clinical settings, compared to clinical MRIs of lower resolution (e.g., with a 5-mm slice thickness) and other modalities [e.g., T2-weighted or (FLAIR)], which have been hitherto typical in clinical protocols, it is now possible to use current available and well-tested T1w-based brain age methods to predict brain age in clinical MRI data. However, with this recommendation, we are ignoring the supplementary information provided by other modalities. Novel methods, based on a multimodal combination of brain MRI images, like in Wood et al. (2022), must be developed, to leverage the diverse clinical information that multimodal MRI data contains, to increase the accuracy of the predictions, and to boost the relevant sensitivity of predicted brain age to clinical variables of interest. A logical next step is to repurpose DeepBrainNet (via transfer learning) to predict brain age, by training it on a clinical multimodal MRI database.

4.2. Conclusions

In summary, brain age can be predicted from clinical brain T1-weighted MRIs from patients that visit the UFHealth system in North Central Florida, USA. Future studies are needed to test the generalizability of these predictions to other clinical systems and to investigate the ability of the predicted brain age difference in multimodal clinical data to characterize pathological conditions. We stress that predicting brain age at the individual level is the cornerstone of future brain age-based biomarkers and personalized medicine, and it is in clinical settings where brain age biomarkers (and any biomarker in general) are needed the most. Taking the first steps toward that direction, we explored the feasibility of brain age predictions on a clinical population. We conclude recommending more accurate CNNs or transfer-learning retraining to increase the accuracy of the brain age predictions.

CRediT author statement

Pedro A. Valdes-Hernandez: Methodology, Software, Formal Analysis, Data Curation, Writing - Original Draft, Visualization. **Chavier Laffitte Nodarse**: Software, Formal Analysis, Data Curation, Writing - Review & Editing. **James H. Cole**: Conceptualization, Writing - Review & Editing. **Yenisel Cruz-Almeida**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors did not use generative artificial intelligence (AI) or AI-assisted technologies in the writing process.

Funding

This work was supported by NIH/NIA grants R01AG059809 and R01AG067757 (YCA).

Declaration of Competing Interest

Authors have no competing interests to declare.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available because the related project has not concluded, but are available from the corresponding author on reasonable request.

Acknowledgments

We thank Dr. Christos Davatzikos and Vishnu Bashyam for making DeepBrainNet publicly available.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.brainresbull.2023.110811.

References

Avants, B.B., Tustison, N., Song, G., 2009. Advanced normalization tools (ANTS). Insight J. 1–35 (Available). ⟨ftp://ftp3.ie.freebsd.org/pub/sourceforge/a/project/ad/advants/Documentation/ants.pdf⟩.

Bashyam, V.M., Erus, G., Doshi, J., Habes, M., Nasralah, I., Truelove-Hill, M., Srinivasan, D., Mamourian, L., Pomponio, R., Fan, Y., Launer, L.J., Masters, C.L., Maruff, P., Zhuo, C., Völzke, H., Johnson, S.C., Fripp, J., Koutsouleris, N., Satterthwaite, T.D., Wolf, D., Gur, R.E., Gur, R.C., Morris, J., Albert, M.S., Grabe, H. J., Resnick, S., Nick Bryan, R., Wolk, D.A., Shou, H., Davatzikos, C., 2020. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. Brain 143, 2312–2324.

Beheshti, I., Nugent, S., Potvin, O., Duchesne, S., 2019. Bias-adjustment in neuroimaging-based brain age frameworks: a robust scheme. NeuroImage Clin. 24, 102063 https://doi.org/10.1016/j.nicl.2019.102063.

Beheshti, I., Nugent, S., Potvin, O., Duchesne, S., 2021. Disappearing metabolic youthfulness in the cognitively impaired female brain. Neurobiol. Aging 101, 224–229. https://doi.org/10.1016/j.neurobiolaging.2021.01.026.

Butler, E.R., Chen, A., Ramadan, R., Le, T.T., Ruparel, K., Moore, T.M., Satterthwaite, T. D., Zhang, F., Shou, H., Gur, R.C., Nichols, T.E., Shinohara, R.T., 2021. Pitfalls in brain age analyses. Hum. Brain Mapp. 42, 4092–4101.

Chen, C.Le, Kuo, M.C., Chen, P.Y., Tung, Y.H., Hsu, Y.C., Huang, C.W.C., Chan, W.P., Tseng, W.Y.I., 2022. Validation of neuroimaging-based brain age gap as a mediator between modifiable risk factors and cognition. Neurobiol. Aging 114, 61–72. https://doi.org/10.1016/j.neurobiolaging.2022.03.006.

Christman, S., Bermudez, C., Hao, L., Landman, B.A., Boyd, B., Albert, K., Woodward, N., Shokouhi, S., Vega, J., Andrews, P., Taylor, W.D., 2020. Accelerated brain aging predicts impaired cognitive performance and greater disability in geriatric but not midlife adult depression. Transl. Psychiatry 10. https://doi.org/10.1038/s41398-020-01004-z.

Cole, J.H., Franke, K., 2017. Predicting age using neuroimaging: innovative brain ageing biomarkers. Trends Neurosci. 40, 681–690. https://doi.org/10.1016/j.tins.2017.10.001.

Cole, J.H., Ritchie, S.J., Bastin, M.E., Valdés Hernández, M.C., Muñoz Maniega, S., Royle, N., Corley, J., Pattie, A., Harris, S.E., Zhang, Q., Wray, N.R., Redmond, P., Marioni, R.E., Starr, J.M., Cox, S.R., Wardlaw, J.M., Sharp, D.J., Deary, I.J., 2018a. Brain age predicts mortality. Mol. Psychiatry 23, 1385–1392.

Cole, J.H., Ritchie, S.J., Bastin, M.E., Valdés Hernández, M.C., Muñoz Maniega, S., Royle, N., Corley, J., Pattie, A., Harris, S.E., Zhang, Q., Wray, N.R., Redmond, P., Marioni, R.E., Starr, J.M., Cox, S.R., Wardlaw, J.M., Sharp, D.J., Deary, I.J., 2018b. Brain age predicts mortality. Mol. Psychiatry 23, 1385–1392. https://doi.org/10.1038/mp.2017.62.

Dörfel, R.P., Arenas-Gomez, J.M., Fisher, P.M., Ganz, M., Knudsen, G.M., Svensson, J., Plavén-Sigray, P., 2023. Prediction of brain age using structural magnetic resonance imaging: a comparison of accuracy and test-retest reliability of publicly available software packages. bioRxiv 56 (Available). ⟨http://biorxiv.org/content/early/2023/01/27/2023.01.26.525514.abstract⟩, 2023.01.26.525514.

Downton, F., Blom, G., 1961. Statistical estimates and transformed beta-variables. Math. Gaz. 45, 369.

Elliott, M.L., Belsky, D.W., Knodt, A.R., Ireland, D., Melzer, T.R., Poulton, R., Ramrakha, S., Caspi, A., Moffitt, T.E., Hariri, A.R., 2021. Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort. Mol. Psychiatry 26, 3829–3838.

Franke, K., Gaser, C., 2019. Ten years of brainage as a neuroimaging biomarker of brain aging: What insights have we gained? Front. Neurol. 10.

Galton, F., 1886. Regression towards mediocrity in hereditary stature. J. Anthr. Inst. Gt. Br. Irel. 15, 246. https://doi.org/10.2307/2841583.

Goyal, M.S., Blazey, T.M., Su, Y., Couture, L.E., Durbin, T.J., Bateman, R.J., Benzinger, T. L.S., Morris, J.C., Raichle, M.E., Vlassenko, A.G., 2019. Persistent metabolic youth in the aging female brain. Proc. Natl. Acad. Sci. USA 116, 3251–3255.

Huang G., Liu Z., Maaten L.Van Der, Weinberger K.Q. Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society, 2017. pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

Jawinski, P., Markett, S., Drewelies, J., Düzel, S., Demuth, I., Steinhagen-Thiessen, E., Wagner, G.G., Gerstorf, D., Lindenberger, U., Gaser, C., Kühn, S., 2022. Linking Brain Age Gap to Mental and Physical Health in the Berlin Aging Study II. Front Aging Neurosci. 14.

Király, A., Szabó, N., Tóth, E., Csete, G., Faragó, P., Kocsis, K., Must, A., Vécsei, L., Kincses, Z.T., 2016. Male brain ages faster: the age and gender dependence of subcortical volumes. Brain Imaging Behav. 10, 901–910.

Koutsouleris, N., Davatzikos, C., Borgwardt, S., Gaser, C., Bottlender, R., Frodl, T., Falkai, P., Riecher-Rossler, A., Moller, H.-J., Reiser, M., Pantelis, C., Meisenzahl, E., 2014. Accelerated brain aging in schizophrenia and beyond: a neuroanatomical marker of psychiatric disorders. Schizophr. Bull. 40, 1140–1153.

de Lange, A.M.G., Cole, J.H., 2020. Commentary: correction procedures in brain-age prediction. NeuroImage Clin. 26, 24–26.

de Lange, A.M.G., Anatürk, M., Rokicki, J., Han, L.K.M., Franke, K., Alnæs, D., Ebmeier, K.P., Draganski, B., Kaufmann, T., Westlye, L.T., Hahn, T., Cole, J.H., 2022. Mind the gap: performance metric evaluation in brain-age prediction. Hum. Brain Mapp. 43, 3113–3129.

Le, T.T., Kuplicki, R.T., McKinney, B.A., Yeh, H.W., Thompson, W.K., Paulus, M.P., Aupperle, R.L., Bodurka, J., Cha, Y.H., Feinstein, J.S., Khalsa, S.S., Savitz, J., Simmons, W.K., Victor, T.A., 2018. A nonlinear simulation framework supports adjusting for age when analyzing brainAGE. Front. Aging Neurosci. 10, 1–11.

Lim, M., O'Grady, C., Cane, D., Goyal, A., Lynch, M., Beyea, S., Hashmi, J.A., 2020. Threat prediction from schemas as a source of bias in pain perception. J. Neurosci. 40, 1538–1548.

Millar, P.R., Gordon, B.A., Luckett, P.H., Benzinger, T.L.S., Cruchaga, C., Fagan, A.M., Hassenstab, J.J., Perrin, R.J., Schindler, S.E., Allegri, R.F., Day, G.S., Farlow, M.R., Mori, H., Nübling, G., Bateman, R.J., Morris, J.C., Ances, B.M., 2023. Multimodal brain age estimates relate to Alzheimer disease biomarkers and cognition in early stages: a cross-sectional observational study. Elife 12.

Montesino-Goicolea, S., Nodarse, C.L., Cole, J.H., Fillingim, R.B., Cruz-Almeida, Y., 2022a. Brain-predicted age difference mediates the association between self-reported pain and PROMIS sleep impairment in persons with knee osteoarthritis. J. Pain. 23, 40. https://doi.org/10.1016/j.jpain.2022.03.154.

Montesino-Goicolea, S., Valdes-Hernandez, P.A., Cruz-Almeida, Y., 2022b. Chronic musculoskeletal pain moderates the association between sleep quality and dorsostriatal-sensorimotor resting state functional connectivity in community-dwelling older adults. Pain. Res. Manag. 2022, 1–12. https://doi.org/10.1155/2022/4347759.

NETER J. Applied linear statistical models. Regression, Anal variance, Stat Des 1990. Available: ⟨https://ci.nii.ac.jp/naid/10006318572/en/⟩.

Sanford, N., Ge, R., Antoniades, M., Modabbernia, A., Haas, S.S., Whalley, H.C., Galea, L., Popescu, S.G., Cole, J.H., Frangou, S., 2022. Sex differences in predictors and regional patterns of brain age gap estimates. Hum. Brain Mapp. 43, 4689–4698.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples. Biometrika 52, 591–611. https://doi.org/10.1093/biomet/52.3-4.591.

Simonyan K., Vedaldi A., Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. 2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings. 2014. pp. 1–8.

Smith, G., 2018. Step away from stepwise. J. Big Data 5. https://doi.org/10.1186/s40537-018-0143-6.

Studholme C., Hawkes D.J., Hill D.L. A normalised entropy measure for multi-modality image alignment. In: {C}Hanson K.M.{C}, editor. Medical Imaging 1998: Image Processing. 1998, Vol. 3338. pp. 132–143. doi:10.1117/12.310835.

Szegedy C., Ioffe S., Vanhoucke V., Alemi A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI'17. AAAI Press, 2017. pp. 4278–4284.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med Imaging 29, 1310–1320. https://doi.org/10.1109/TMI.2010.2046908.

Valdes-Hernandez, P.A., Laffitte Nodarse, C., Johnson, A.J., Montesino-Goicolea, S., Bashyam, V., Davatzikos, C., Peraza, J.A., Cole, J.H., Huo, Z., Fillingim, R.B., Cruz-Almeida, Y., 2023. Brain-predicted age difference estimated using DeepBrainNet is significantly associated with pain and function—a multi-institutional and multiscanner study. Pain 00. https://doi.org/10.1097/j.pain.0000000000002984.

Wei, R., Xu, X., Duan, Y., Zhang, N., Sun, J., Li, H., Li, Y., Li, Y., Zeng, C., Han, X., Zhou, F., Huang, M., Li, R., Zhuo, Z., Barkhof, F., Cole, H., Liu, J., Brain, Y., 2022. age gap in neuromyelitis optica spectrum disorders and multiple sclerosis. J. Neurol. Neurosurg. Psychiatry 94, 31–37.

Wood, D.A., Kafiabadi, S., Busaidi, A.Al, Guilhem, E., Montvila, A., Lynch, J., Townend, M., Agarwal, S., Mazumder, A., Barker, G.J., Ourselin, S., Cole, J.H., Booth, T.C., 2022. Accurate brain-age models for routine clinical MRI examinations. Neuroimage 249.

Yin, C., Imms, P., Cheng, M., Amgalan, A., Chowdhury, N.F., Massett, R.J., Chaudhari, N. N., Chen, X., Thompson, P.M., Bogdan, P., Irimia, A., 2023. Anatomically interpretable deep learning of brain age captures domain-specific cognitive impairment. Proc. Natl. Acad. Sci. USA 120, 1–11.