PREDICTING THE AGE OF THE BRAIN WITH MINIMALLY PROCESSED T_1-WEIGHTED MRI DATA

Caroline Dartora[a,*], Anna Marseglia[a], Gustav Mårtensson[a], Gull Rukh[b], Junhua Dang[b], J-Sebastian Muehlboeck[a], Lars-Olof Wahlund[a], Rodrigo Moreno[c], José Barroso[d], Daniel Ferreira[a,d,e], Helgi B. Schiöth[b], and Eric Westman[a,f,*], for the Alzheimer's Disease Neuroimaging Initiative[1], the Australian Imaging Biomarkers and Lifestyle flagship study of ageing[2], the Japanese Alzheimer's Disease Neuroimaging Initiative[3], and the AddNeuroMed consortium.

[a]Division of Clinical Geriatrics, Center for Alzheimer Research, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden.

[b]Department of Surgical Sciences, Functional Pharmacology and Neuroscience, Uppsala University, Uppsala, Sweden

[c]Department of Biomedical Engineering and Health Systems, KTH Royal Institute of Technology, Huddinge, Stockholm, Sweden

[d]Department of Psychology, Faculty of Health Sciences, University Fernando Pessoa Canarias, Guía, Gran Canaria, Islas Canarias, Spain

[e]Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA

[f]Department of Neuroimaging, Centre for Neuroimaging Sciences, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.

*Corresponding authors:

Caroline Dartora and Eric Westman

Karolinska Institutet

Department of Neurobiology, Care Sciences and Society (NVS)

Division of Clinical Geriatrics

Neo floor 7 | SE 141 83 Huddinge

Sweden

Email: caroline.dartora@ki.se

[1]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

[2]Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (www.loni.usc.edu/ADNI). The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at www.aibl.csiro.au.

[3]Data used in preparation of this article were obtained from the Japanese Alzheimer's Disease Neuroimaging Initiative (J-ADNI) database deposited in the National Bioscience Database Center Human Database, Japan (Research ID: hum0043.v1, 2016). As such, the investigators within J-ADNI contributed to the design and implementation of J-ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of J-ADNI investigators can be found at:https://humandbs.biosciencedbc.jp/en/hum0043-j-adni-authors.

ABSTRACT

In the last few years, several models trying to calculate the biological brain age have been proposed based on structural magnetic resonance imaging scans (T1-weighted MRIs, T1w), using multivariate methods and artificial intelligence. We aimed to develop and validate a convolutional neural network (CNN) model for brain age prediction (PBA) using minimally processed T1w MRIs. Our model only requires one preprocessing step (i.e., image registration to MNI space), which is an advantage in comparison with previous methods that require more preprocessing steps. We used a multi-cohort dataset of cognitively healthy individuals comprising 16734 MRIs for training and evaluation. To validate our model and its interpretability, we used a multivariate model, Orthogonal Projections to Latent Structures (OPLS), which uses brain segmented cortical thicknesses and volumes. We trained and evaluated the models with the same dataset, and systematically investigated how predictions of the CNN model differ from those of the OPLS model. The validation of our model was made by testing an external dataset. The CNN and the OPLS model achieved a mean absolute error (MAE) in the testing dataset of 3.04 and 4.81 years, respectively. The model's performance in the external dataset was in the typical range of MAE found in the literature for testing sets. The CNN model revealed similar image patterns when grouped by chronological age (CA) and CNN predicted age. No significant differences were found between the oldest and youngest quartiles of age predictions by the CNN in a validation cohort of individuals with CA of 70 years old. Sensitivity maps analysis revealed that the age prediction is based mainly on the ventricles and other CSF spaces, which have been shown in the literature to reflect aging and are in accordance with the most important regions for the prediction in the OPLS model. While both the CNN and the OPLS model demonstrated acceptable performance metrics on a hold-out test set, individual predictions differed substantially, with brain age patterns of the CNN model being more comparable to the chronological age. In conclusion, our CNN model showed results comparable to the literature, using minimally processed images, which may facilitate the future implementation of brain age prediction in research and clinical settings.

Keywords: brain age, neurodegeneration, normal ageing, CNN, UK Biobank

1 INTRODUCTION

In recent years, the concept of an individual's biological age—which can differ from the person's chronological age—has sparked great interest in the medical research community, as ageing is a major risk factor for several age-related health conditions and mortality. There is also a great heterogeneity in health outcomes among individuals of the same chronological age (Jylhävä et al., 2017). During the past decades, research highlighted that the biological ageing process varies between people because of the complex interplay between genetic and environmental factors, such as lifestyle behaviours (Cole et al., 2019, 2017; Fratiglioni et al., 2020). Chronological age is the primary risk factor for most neurodegenerative disorders, including Alzheimer's disease (AD) and vascular dementia (Hou et al., 2019). Like other age-related health conditions, also in the dementia field, there is great heterogeneity in the manifestation of the symptoms as well as underlying brain pathology between people of the same chronological age (Ferreira et al., 2020). Therefore, quantifying the biological age could be a better metric than the traditional chronological age to identify individuals at risk of developing age-related diseases (ARDs) (Cole et al., 2019).

In recent years, parallel advancements in the fields of neuroscience and computational science have enabled researchers to develop novel algorithms to determine the biological age of one of the most important organs during ageing, that is, the brain. The main advantage of a biological marker of brain age is that this could get researchers one step closer to a deeper understanding of the mechanisms underlying pathological cognitive ageing, which culminates in dementia. Dementia is a multifactorial syndrome in which decades of accumulating neuropathology precedes clinical manifestation (Jack et al., 2010). Overtime, the loss of neurons and synapses during the preclinical and prodromal stages will lead to brain atrophy and, therefore, to "older-looking" brains (i.e., when brain age is higher than chronological age) (Bashyam et al., 2020; Cole, 2020; Cole and Franke, 2017; Elliott et al., 2021; Franke and Gaser, 2012; Glorioso et al., 2019; Hwang et al., 2020). In contrast, some individuals will show higher chronological age than brain age, thus showing a "younger-looking" brain, which could reflect a relatively preserved brain structure because of lifelong mechanisms of brain resilience and resistance (e.g., brain reserve) (Cole et al., 2019; Stern et al., 2020). With the current unprecedented growth of elderly population worldwide, the expected dramatic increase of dementia cases (WHO guidelines, 2019), and the lack of effective pharmacological treatments for dementia, a biological marker of brain age could play a key critical role in dementia treatment and prevention (Brusini et al., 2022). In fact, brain age could help to understand the interplay between different cerebral pathophysiological processes contributing to cognitive impairments, paving the way for future advances in the treatment of dementia. Moreover, at the level of primary prevention, brain age could help identify those modifiable factors that contribute throughout life to maintaining the brain health and preserving cognitive functioning.

In the past few years, several brain age models have been developed using different methods to estimate the age of the brain (Baecker et al., 2021a; Bocancea et al., 2021). Typically, a model is trained on neuroimaging data of healthy individuals from one or multiple cohorts covering a large age span. The type of model and the choice of input varies between studies, but most of these previous studies have used preprocessed neuroimaging data. The most used input for predicting brain age models are brain magnetic resonance imaging scans (MRIs), or features derived from structural MRI. In Franke et al. (2010), and Franke and Gaser (2012), a relevance vector regression model was used with tissue maps as input that had undergone dimensionality reduction using principal component analysis. Cole et al. (2017) also used segmented tissue maps but with a Gaussian process model. Hwang et al. (2020) used cortical thickness and volumes from FreeSurfer segmentation as input to a support vector regression model. With the advances in convolutional neural networks (CNN), some recent studies trained CNNs to predict brain age from segmented tissue maps (Liang et al., 2019; Niu et al., 2020) as well as from the voxel intensities of T_1-weighted (T1w) images (Cole et al., 2017; Jonsson et al., 2019). Whereas all these studies achieved a good performance in terms of low mean absolute error (MAE) values on their

respective test sets (typically between 2-6 years), recent work suggests that CNNs can yield superior performance when coupled with large data sets (Bintsi et al., 2020; Kolbeinsson et al., 2019; P. K. Lam et al., 2020). An acknowledged strength of CNNs is that it has shown superior performance in many medical imaging tasks compared to other types of models (Litjens et al., 2017). A potential drawback, however, is that CNNs may suffer from low interpretability, hindering understanding of what the models base their predictions on (Boscolo Galazzo et al., 2022; Hepp et al., 2021; Lombardi et al., 2021). Currently, the use of gradient-based techniques has added some explanation to CNN predictions (Boscolo Galazzo et al., 2022; Hepp et al., 2021; Lombardi et al., 2021), but the interpretation of the differences between brain age and chronological age are still difficult. Orthogonal Projections to Latent Structures (OPLS) models have previously been applied to neuroimaging data in dementia and ageing research for the exact opposite reason, that is, OPLS provides high interpretability (Falahati et al., 2014; Machado et al., 2018).

In this study, we aimed to develop and validate a CNN model for brain age prediction using minimally processed T1w MRIs. We used a large multi-cohort dataset of cognitively healthy individuals comprising more than 16000 T1w MRIs for training and evaluation of the model. To validate our model and its interpretability, we used OPLS applied on brain segmented cortical thickness and volume regional data. We trained and evaluated the models with the same dataset and systematically investigated how predictions of the CNN model differ from those of the OPLS model. As more and more models are being publicly released for the use of the research community, evaluating the model's performance may not be enough making it critical to investigate how and why brain age predictions by these models differ to make them simple for future research and clinical implementation. For this reason, we aimed to develop a CNN model from minimally processed images.

2 METHODS

2.1 DATA AND STUDY POPULATION

Developing a model for predicting "brain age" is dependent on the choice of training data when it comes to defining healthy aging. The ideal dataset for developing a brain age prediction model would include: (1) detailed information and clinical data of study participants in order to be as comprehensive as possible with the selection criteria; (2) a large set of images, which are required to train a CNN model; (3) participants with diverse demographical background and a large, preferably uniform, age distribution to apply the developed model in more datasets (i.e., increase generalizability); (4) images acquired with a wide range of MRI scanners and protocols to improve generalization to new unseen data of the model (Mårtensson et al., 2020), and (5) longitudinal data to ensure that the model does not predict a lower age at a later timepoint. Although initiatives to gather large-scale population-based datasets are ongoing (e.g., UK Biobank), to our knowledge, currently no existing cohort possess all the five characteristics listed above. One possibility is to achieve these five characteristics by combining data from different cohorts in a combined dataset. Combining datasets typically improves points two to five above at the expense of point one and the risk of introducing unwanted correlations in the data (Mårtensson et al., 2020). An example is when combining two datasets A and B, where participants in A are substantially younger than in B and the data in A is acquired with a different MRI protocol.

To partly overcome these limitations, we combined four large well-known ageing cohorts into a large training set, including images acquired from a wide range of scanners and participants with a wide age span, in addition to comprehensive clinical assessments. Two of these cohorts were population-based: the UK Biobank (Sudlow et al., 2015) and the Group of Neuropsychological Studies from the Canary Islands (GENIC, from *Grupo de Estudios Neuropsicologicos de las Islas Canarias*) (Ferreira et al., 2015a). The remaining two cohorts were the Alzheimer's

Disease Neuroimaging Initiative (ADNI) and the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL). Briefly:

- **UK Biobank** (Miller et al., 2016). This is a large, prospective multi-center cohort study conducted across England, Scotland, and Wales. At baseline (2006-2010), about 500 000 participants aged 40-69 years were recruited and extensive phenotypic and genotypic data were collected. Additionally, in 2014 UK biobank imaging study began with a plan to re-invite 100000 participants for brain, heart, and body imaging. The participants underwent 3T brain MRI scans. Within the MRI sample, we selected 13882 individuals without dementia, mild cognitive impairment (MCI), other neurological and psychiatric conditions, and with good-to-excellent self-reported health.
- **GENIC** (Ferreira et al., 2015b). This is a community-based longitudinal cohort of adults aged 29 to 85 years at baseline, who were recruited through advertisements in local schools, primary care health centres, and relatives and acquaintances of the research staff of the University of La Laguna, Tenerife, Canary Islands, Spain. We selected 299 participants without dementia, MCI, neurological, psychiatric, or other systemic diseases as well as no evidence of pathological findings in the 3T brain MRIs (e.g., stroke, tumours, hippocampal sclerosis) or a history of substance abuse.
- **ADNI** (Jack et al., 2008). Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). 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 MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. This longitudinal multi-center study includes more than 2000 individuals in all levels of cognition (from unimpaired to cognitively impaired – including mild stages and dementia), aged 55-90 years, with available 1.5/3T brain MRIs, followed for up over time. Within the available sample, we selected 1489 MRIs of 443 cognitively unimpaired individuals.
- **AIBL** (Rowe et al., 2010). Data was collected by the AIBL study group. AIBL study methodology has been reported previously (Ellis et al., 2009). This is a longitudinal multi-center study including 1100 Australian individuals that are cognitively healthy, MCI, or AD dementia, aged ≥60 years with available 1.5/3T brain MRIs. From this cohort, we selected 957 MRIs from 491 cognitively healthy individuals.

For the analyses in the current study, we selected 16734 T1w MRIs from 15115 cognitively healthy participants from the four cohorts (Figure 1) by applying consistent criteria. The raw MRIs were from both 1.5T (642 MRIs), and 3T (16092 MRIs) scanners. A cognitively healthy status was defined based on the absence of dementia, MCI, and other neurological and psychiatric disorders, as well as self-reported good health (this last when available). Specifically, in the UK Biobank, medical diagnoses were collected through physicians/self-reports or medical records following the International Classification of Disease – 9th or 10th edition (ICD-9 or 10; the full list of the ICD's code used as exclusion criteria are listed in Supplementary Table 1 – section 1). In ADNI and AIBL, subjects diagnosed with MCI or AD were excluded. In the GENIC data, we excluded participants based on the procedure outlined in Machado et al. (2018), designed to be comparable to ADNI. These exclusion criteria included a clinical diagnosis of MCI or dementia, abnormal cognitive performance on extensive neuropsychological assessment, abnormal global cognitive performance defined as a mini-mental state examination (MMSE) score ≤ 24, abnormal findings on MRI according to a neuroradiologist, and evidence of psychiatric or neurological disorder. Distribution plots of the chronological ages in the four cohorts are included as Figure 1 in the Supplementary material (section 2).

Each used dataset in this study was approved by the ethical review boards in the participating sites. At Karolinska Institutet, the use of data for research purposes was approved under DNR 2021-04428.

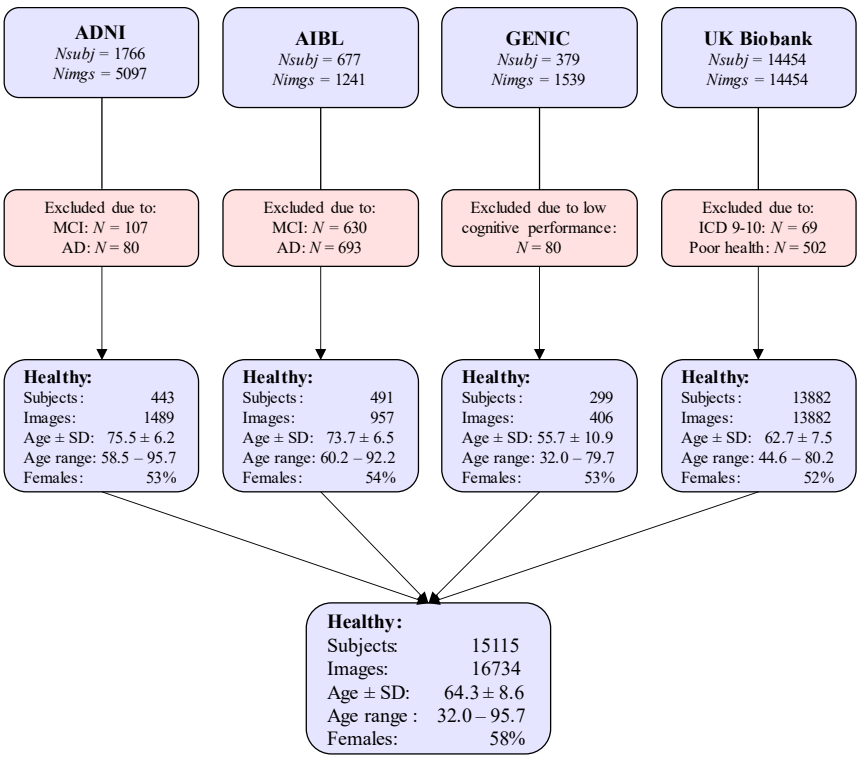

Figure 1. Flowchart describing the data included in this study, which were available in our database system in the time of the study, as well as the exclusion criteria for the different cohorts. Where N=number, *Nsubj*=number of subjects selected in the specific cohort, *Nimg*=number of images selected in the specific cohort, MCI=mild cognitive impairment, AD=Alzheimer disease, and Age ± SD is the mean age and standard deviation. Age ± SD and age range are in years.

2.1.1 External dataset for validation

To validate performance and generalizability of the brain age prediction model, we used two external datasets: the AddNeuroMed, and the Japanese ADNI (J-ADNI). A brief description of both cohorts is presented below. The same inclusion and exclusion criteria used to select the data for training was applied for both cohorts generating and external dataset. To guarantee an unbiased performance evaluation, the external dataset contained images obtained in different scanners than those presented in the training set, and completely unknown to the model. The age distribution within different cohorts is presented in Figure 2.

- **AddNeuroMed** (Birkenbihl et al., 2021; Simmons et al., 2011). AddNeuroMed is part of the InnoMed European Union PF6 programme, designed to develop and validate novel surrogate markers in AD and includes MRI data with other biomarkers and clinical information. A total of 88 cognitively unimpaired individuals (44 females, 44 males), with 149 MRIs, with an average age of 75.1 yrs. (53.0 – 88.9 yrs.), with 1.5T brain MRIs, were included in the external dataset for validation in this study.
- **JADNI** (Iwatsubo et al., 2018). Data used in preparation of this article were obtained from the Japanese Alzheimer's Disease Neuroimaging Initiative (J-ADNI) database deposited in the National Bioscience Database Center Human Database, Japan (Research ID: hum0043.v1, 2016). The J-ADNI was launched in 2007 as a public-private partnership, led by Principal Investigator Takeshi Iwatsubo, MD. The primary goal of J-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 late mild cognitive impairment (MCI) and mild Alzheimer's disease (AD) in the Japanese population. This is a large-scale observational study based on the ADNI cohort in a Japanese population and includes neuroimaging, neuropsychological and biomarker data. From this dataset, we selected 86 individuals (50 female, 36 male), with 413 MRIs, following the inclusion and exclusion criteria

mentioned in section 2.1, with an average age of 68.5 yrs. (60 – 86 yrs.), and with a 1.5T and 3T (385 and 28 MRIs, respectively) brain MRI.

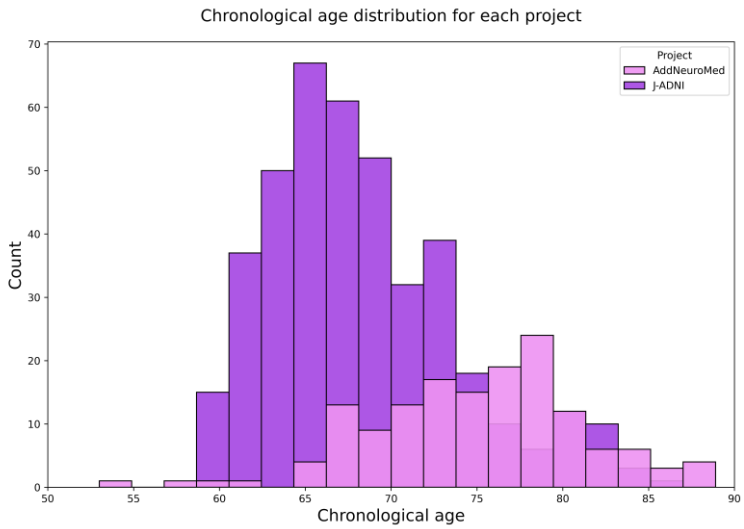

Figure 2. Age distribution for the external test dataset.

2.2 BRAIN AGE MODEL AND ASSUMPTIONS

Brain age is considered a measure that better reflects the current state of an individual's brain than their chronological age. Since this is an abstract concept, there are no ground truth labels available. Instead, researchers usually select healthy subjects within an age span, and their chronological age come to define the brain age of the cohort. We can model brain age as:

$$y_{b,i} = y_{c,i} + \delta_{b,i} \tag{1}$$

where $y_{b,i}$ is the "brain age" of subject i, $y_{c,i}$ his/her chronological age, and $\delta_{b,i}$ his/her "brain age delta", which is hypothesized to vary due to various factors, e.g., lifestyle, genetics, etc. Generally, we are interested in estimating $\delta_{b,i}$ for a subject. In a healthy subpopulation, we expect $\delta_{b,i}$ to be smaller than in, e.g., patients with neurological disorders. Importantly, when training a model to predict chronological age by minimizing $|\hat{y}_i - y_{c,i}|$, where $\hat{y}_i$ is the predicted age by the model, we are implicitly assuming that the average $\delta = 0$, meaning that in, e.g., a subset of 70-year-old subjects, their brain ages will have a distribution with a mean of 70.

Any model will have some degree of noise in its predictions. As we do not have access to ground labels, it is not easy to distinguish between "noise" and "actual" deviation from their chronological age in a predicted brain age delta. We would expect a model with low noise to predict brain ages with smooth trajectories on longitudinal data points when plotted against chronological age. We would also expect a prediction on follow-up data that must not to be younger than on

baseline data (i.e., no reversed aging). These qualities are difficult to quantify in a meaningful way but can serve as a sanity check for the reliability of a "brain age" model.

2.3 MODELS

Two separate models were developed in this project: 1) one model based on CNN and 2) one model based on OPLS (Trygg and Wold, 2002). One notable way in which the models differ is that the CNN bases the age predictions on the voxel intensities in the image, whereas the OPLS model requires prior dimension reduction of the input features.

Data set splits: To not violate the test set and ensure comparability between the models, the data set was randomly split into a training (N_{img} = 15052, N_{subj} = 13612, subsequently split into internal validation sets for the development of each model) and a hold-out test set (N_{img} = 1682, N_{subj} = 1503). If a subject had undergone multiple scans, all their images were assigned to the same set. The test set was evaluated only after a satisfactory performance on the internal validation sets had been attained to reduce the risk of model overfitting for the CNN and OPLS models.

Convolutional Neural Network: We aimed to employ as little preprocessing as possible to make the model accessible and fast. The only preprocessing used was a rigid registration (6 degrees of freedom) to the MNI template using FSL FLIRT 6.0 (FMRIB's Linear Image Registration Tool). Omitting this step yielded worse performance on the development set, despite heavy data augmentation (data not shown). The augmentation on the training set included random scaling, cropping offsets, rotations, affine, and gamma transformations. Each image was cropped to a dimension of 80 x 96 x 80 voxels, with 2 x 2 x 2 mm^3 resolution, thresholded to the 5th and 95th percentile of the voxel values and scaled so that all voxels' values were in the interval [-1,1].

The CNN was implemented using the deep learning framework PyTorch (Paszke et al., 2019). The network architecture of our model was based on the ResNet architecture (He et al., 2016) with 26 layers in total but with 3D instead of 2D kernels. Each convolutional operation is followed by batch normalization and a leaky Rectified Linear Unit (ReLU) activation. The network was trained for 20 epochs with stochastic gradient descent and an initial learning rate of $\lambda = 0.005$ that decreases with a factor of 10 every 5 epochs. With this training scheme, the MAE of the development set had converged without signs of overfitting (data not shown). We use five independent models during development and use the trained networks as an ensemble model on the hold-out test set, where the average prediction was used for future analysis.

To explain the model's brain age prediction, we generated gradient-based saliency maps of subjects with chronological age of 70 ± 1 years. They allow the visualization of the pixels that are important for the prediction based on the computation of the gradient of the prediction with respect to the image.

Orthogonal projections to latent structures model: OPLS is a hypothesis-driven supervised method that has previously been applied to neuroimaging data for classification (Westman et al., 2013, 2011) as well as regression (Machado et al., 2018). The method aims to find the direction in the multidimensional space of the data X (predictor variables) that explains the variance in Y space (the dependent variable, in this case 'age') (Eriksson et al., 2013; Sadeghi-Bazargani et al., 2012). This property can provide good interpretability of the model and an estimation of variable importance.

The OPLS model was conducted using the software SIMCA 15.0 (Sartorius Stedim Biotech). Neuroimaging data was processed through the HiveDB database system (Muehlboeck et al., 2013) using FreeSurfer 6.0.0 (described in detail in Dale et al. (1999), and Fischl et al. (2004), available at https://surfer.nmr.mgh.harvard.edu/). Since existing studies have shown that the thickness of a cortical brain region is generally more sensitive to age-related changes compared to its

volume (Machado et al., 2018; Westman et al., 2013), we selected cortical thickness values together with subcortical and ventricular volumes (adjusted for the estimated total intracranial volume obtained from FreeSurfer) as input to the model. A 7-fold cross-validation scheme was used for model development and training. These seven models were used as an ensemble model on the test set, following the same method as for the CNN model with the hold-out approach.

To explain the model's brain age prediction, we generate the variable importance for projection in the OPLS model. A graphical visualization of the most important regions was also generated including the 15 cortical and subcortical regions with the highest importance for projection.

2.4 ANALYSES

Model performance was assessed using the MAE, defined as:

$$MAE = \frac{1}{N}\sum_{i}^{N}\left|\hat{y}_i - y_{c,i}\right| \tag{2}$$

where $y_{c,i}$ is the chronological age of participant i and $\hat{y}_i$ his/her predicted age. Butler et al. (2021) argue that adjusting for age-dependent brain age delta is problematic and can artificially inflate model performance. This can be illustrated with a model that, regardless of input data, only outputs a single predicted age of, e.g., 70. This will obviously yield a terrible MAE, but when "correcting" for age, the MAE will be 0 between predicted and chronological age

To visualize how the age predictions are related to cortical thickness values, we ran surface group analysis with QDEC (Query, Design, Estimate, Contrast) in FreeSurfer 6.0.0. We used a smoothing kernel of 10 mm width and with sex as a covariate. We grouped subjects of ages 60, 65, 70, 75, and 80 (±1 years) based on chronological and predicted ages. These groups were contrasted to a reference group of 55 ± 1 years-old (based on corresponding age metric) subjects in the general linear model (GLM). Since these groups are of different sizes—and *p*-values are influenced by group size—we present figures overlayed with z-scores. Cortical volume results are included as Supplementary data (Figure 2 – section 3). To visualize the cortical thickness differences between the oldest and youngest quartiles of predictions of individuals with a chronological age of 70±1 in each model, we ran QDEC analysis with sex and age used as covariate and adjusted for false discovery rate (FDR) at a threshold of 0.05. Further, gradient-based sensitivity maps (Smilkov et al., 2017) were calculated for the CNN model, which provides an indication of which voxels were important in individual predictions. We use them to visualize some cases that were predicted to be substantially higher or lower than their chronological age by the CNN model, to understand what features of the brain differ between a younger- and an older-looking brain.

Interpretation of the OPLS model is based on the features' importance for orthogonal projection ranking used by the model to predict brain age. To visualize the important brain regions in the predicted brain age with the OPLS model, features with the highest importance values were presented as a bar plot (ranking plot of all features and their value of importance to the model is presented in the Supplementary material Figure 3 – section 4). The related brain regions were mapped in a visual graph.

3 RESULTS

3.1 MODEL PERFORMANCE

Scatterplots of the (un-adjusted) brain age predictions on the hold-out test set of healthy individuals are shown in Figure 3. The CNN model yielded a lower MAE (3.04 years) than the OPLS model (4.81 years) as compared to participants' chronological age. When comparing the predictions of the two models (CNN vs. OPLS, Fig. 3C), the MAE was higher than for chronological age vs. CNN (Fig. 3A) but lower than for chronological age vs. OPLS (Fig. 3B). There is a distinct age-bias for both models, i.e., younger subjects get predicted as older and vice versa, where the slope of the CNN model predictions is closer to the identity line than that of the OPLS model. Fig. 3D) shows that while there is an association between the brain age deltas of the two models, the size of the variance is substantial.

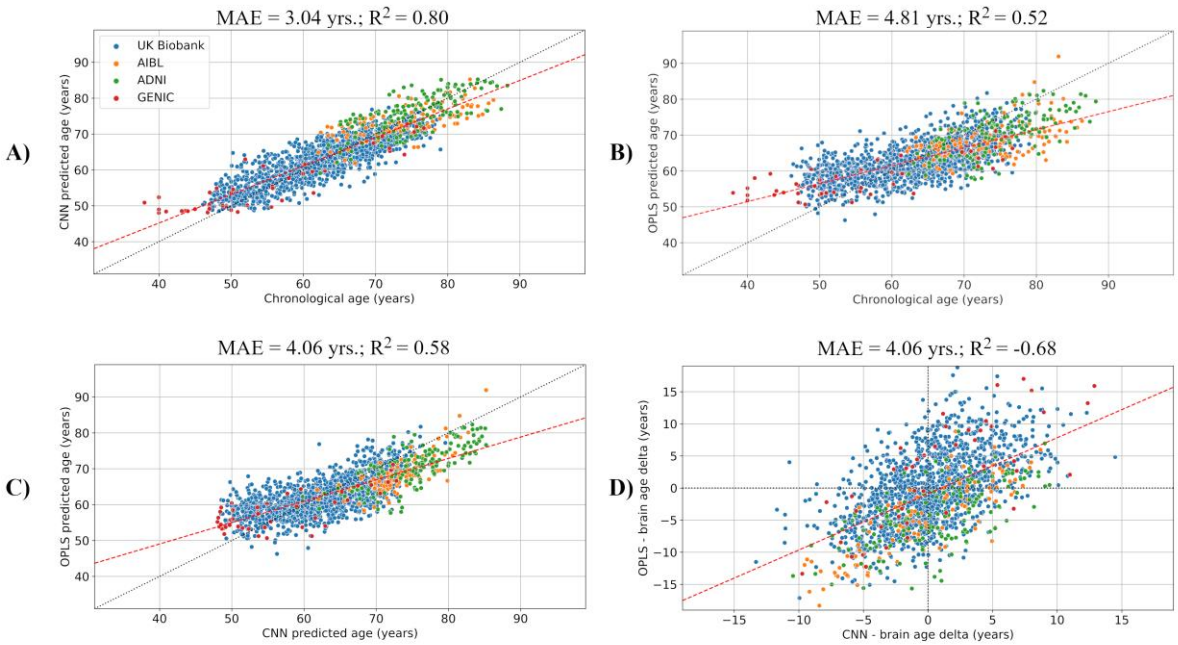

Figure 3. A): Predicted age by the CNN compared to chronological age. B): Predicted age by the OPLS model compared to chronological age. C): Predicted age by the CNN compared to predicted age by the OPLS model. D): Predicted brain age delta by the CNN compared to the OPLS model. Each dot represents a single image in the hold-out test set of healthy individuals, where the colour indicates which cohort it is part of. The dashed red line indicates the linear regression of the predictions. The dashed black line represents the linear relationship between chronological and expected predicted brain age.

In line with previous studies, we also observed an "age-bias" in our predictions where the oldest healthy subjects had systematically lower predictions compared to their chronological age (as more noticeable in supplementary data Figure 4 and 5 – section 5, where we present the predicted age by the CNN and OPLS models compared to chronological age for each cohort individually). While applying an age correction like Beheshti et al. (2019) (for more details see Supplementary material section 6) improved the MAE for both CNN and OPLS models, we chose to report these age-corrected findings as supplementary data only (Supplementary material Figure 6 – section 7).

We further assessed our brain age models' performance by computing the MAE only within the UK biobank test-set and then comparing the achieved MAEs with those from previous studies that evaluated their models in the UK Biobank cohort (Table 1). Our CNN model achieved an MAE of 2.87, which is very similar to the MAEs achieved by CNN models developed in previous studies (ranging between 2.13 and 4.36). The OPLS model achieved an MAE of 4.55, showing a lower performance by a small margin across the publications included in Table 1.

Table 1. Comparison of our models MAE with the literature.

Study	Model	Modality	Preprocessing	MAE
Dartora et al.	CNN (3D ResNet)	T1	Rigid reg.	2.87
Dartora et al.	OPLS	T1	FS cortical values	4.55
Cole (2020)	LASSO	Multimodal	Summary statistics	3.52
Bintsi et al. (2020)	CNN (3D ResNet)	T1	Skull-strip., non-linear MNI reg.	2.64
Bintsi et al. (2020)	CNN (patches)	T1	Skull-strip., non-linear MNI reg.	2.13
Kolbeinsson et al. (2019)	CNN	T1	Skull-strip., non-linear MNI reg.	2.58
P. K. Lam et al. (2020)	R-CNN	T1	Bias-field cor., Skull-strip., Rigid MNI reg.	2.86
P. Lam et al. (2020)	CNN	T1	Bias-field cor., Skull-strip., Rigid MNI reg.	4.36
Gupta et al. (2021)	CNN (slices)	T1	Bias-field cor., Skull-strip., Rigid MNI reg.	2.82
Dinsdale et al. (2021)	CNN (Female population)	T1	UK Biobank pipeline w/ linear registration to MNI	2.86
Dinsdale et al. (2021)	CNN (Male population)	T1	UK Biobank pipeline w/ linear registration to MNI	3.09
Jonsson et al. (2019)	CNN (transfer learning)	T1	Bias-field cor., Skull-strip., DARTEL MNI reg., tissue maps	3.63
Peng et al. (2021)	SFCN	T1	Bias-field cor., Skull-strip., Rigid MNI reg.	2.14
Peng et al. (2021)	CNN	T1	Bias-field cor., Skull-strip., Rigid MNI reg.	2.38
Baecker et al. (2021a)	RVR	T1	Vector of flattened 3D voxel values of cortical regions + PCA	3.66

CNN: convolutional neural network; OPLS: Orthogonal projections to latent structures; PCA: principal component analysis; R-CNN: recurrent CNN; RVR: relevant vector regression; SFCN: simple fully convolutional network.

To obtain a sense of the noise levels from the model and its ability to capture subtle changes in individuals as a result of the ageing process, we plotted longitudinal trajectories in Figure 4 from ADNI and AIBL participants with three or more time-points. A "perfect" model would show smooth and non-declining trajectories, assuming that one's brain age does not vary rapidly or decreases substantially, in our cohort of healthy individuals. Visually, both models seem to generate smooth predictions that increase with chronological age, with some noisy predictions that deviate from the trajectory. However, we observed that the OPLS predictions did not follow the identity line as closely as the CNN predictions, showing higher prediction noise compared to the CNN model.

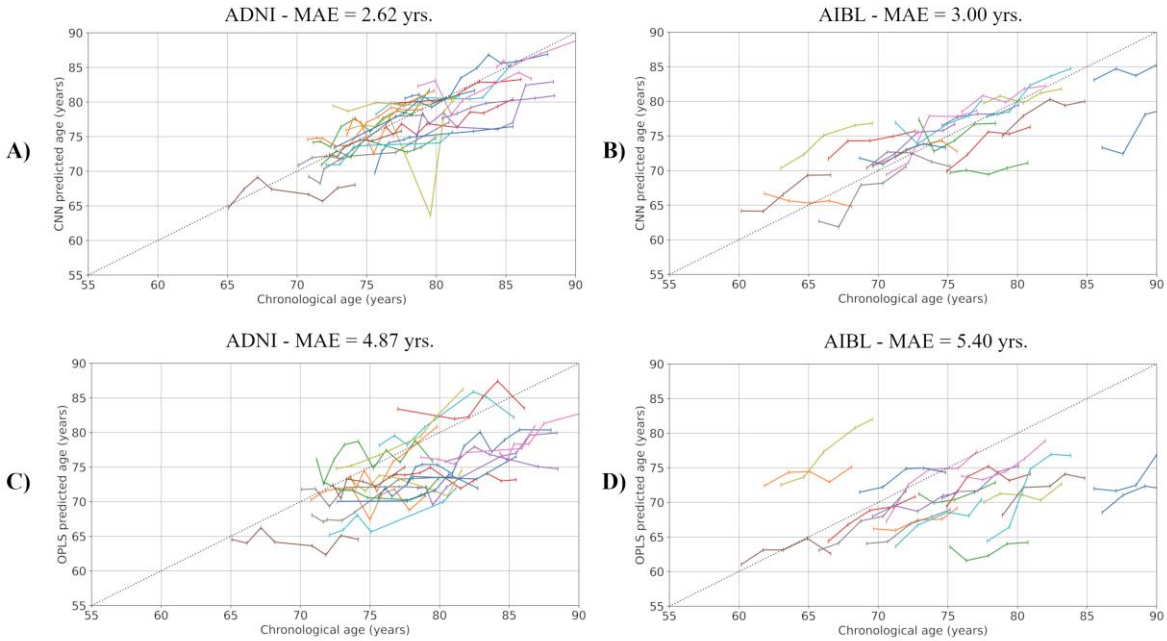

Figure 4. Age prediction on individuals in the test set with three or more timepoints. Each line represents one individual and the vertical dashes at what chronological age the images were acquired at. A): Predictions from the CNN model on ADNI data. B): Predictions from the CNN model on AIBL data. C): Predictions from the OPLS model on ADNI data. D): Predictions from the OPLS model on AIBL data. Abbreviations: Convolutional Neural network (CNN); Orthogonal projections to latent structures (OPLS).

3.2 AGEING PROCESS AND MODEL DIFFERENCES

In Figure 5, we show z-score maps of cortical thickness across different age groups contrasted against participants with chronological and predicted age of 55±1 years, respectively. All age groups displayed similar atrophy patterns compared to 55-year-old individuals, with diffuse atrophy progressing with age. The maps from the OPLS predictions indicated earlier atrophy progression, with notable posterior atrophy already for 60-year-old individuals, whereas the maps from the CNN predicted ages more closely resembled those based on chronological age. Interestingly, one region medially in the primary visual cortex consistently showed increased thickness in older groups.

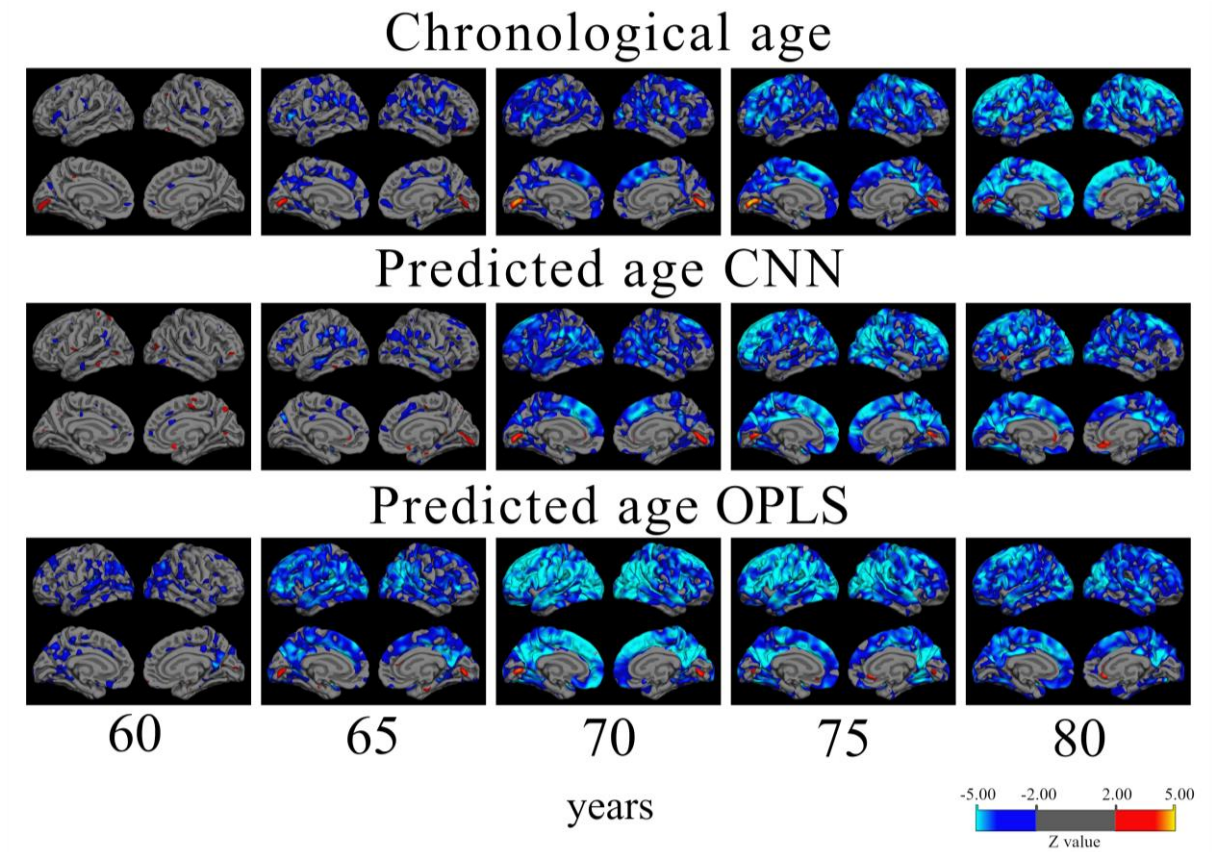

Figure 5. Cortical thickness z-stat maps of cross-sectional differences between a reference group of 55-year-old and older subjects defined by chronological (top), brain age predicted by CNN (middle), and OPLS (bottom). E.g., in the middle row, in the column denoted "70", individuals predicted as 70.0±1.0 years old by the CNN model were compared to those predicted as 55.0±1.0 years. The QDEC analyses were done with a 10 mm smoothing kernel and with sex added as a covariate. Blue colours represent lower cortical thickness (increased atrophy) in the age group compared to the 55-year-old group (with increased atrophy going from light to dark blue colour), and red colours represent an increased cortical thickness compared to the younger group. Data is Z-value transformed therefore representing standard deviations from the 55-years-old group, uncorrected for false discovery rate.

To understand what differs between younger- and older-looking brains on an individual level, we calculated gradient-based sensitivity maps from the CNN model predictions for individuals with chronological age of 70 ± 1, shown in Figure 6 (corresponding axial and sagittal views are available as Supplementary data, section 8, Figure 7 and Figure 8, respectively). We observed that the model predictions were sensitive to ventricular and other CSF spaces, mainly next to the insula. The OPLS model also listed the volumes of lateral and inferior-lateral ventricles as the variables with the highest importance in the predictions (Figure 7 and Supplementary data Figure 3 – section 4).

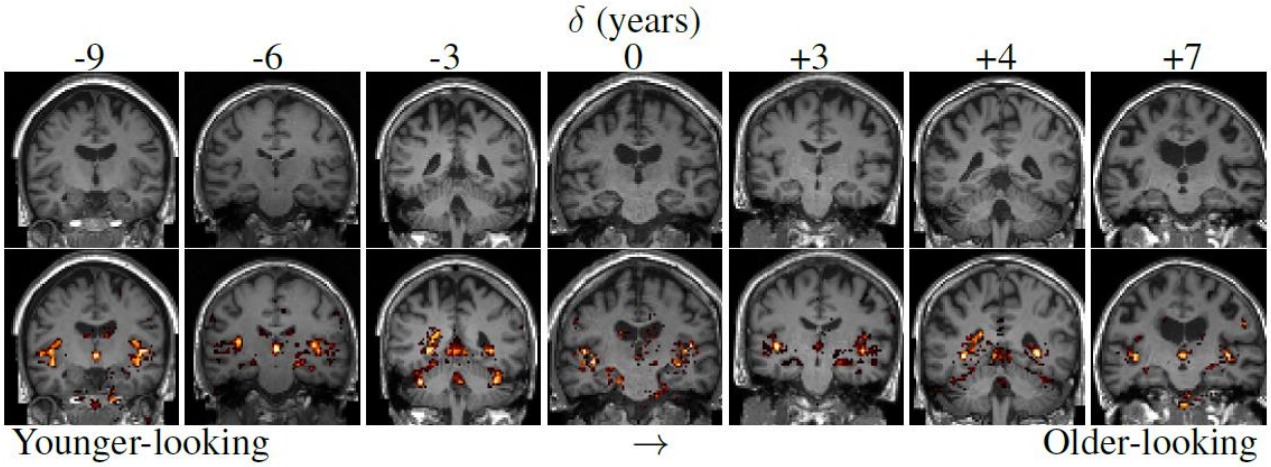

Figure 6. Gradient-based sensitivity maps from seven individuals in UK Biobank with chronological age of 70±1 years but predicted as younger to older (*left* to *right*, brain age delta δ indicated above). The bottom row shows the image above overlayed with the strongest 1% of the values of the sensitivity maps. The coronal slice chosen was the slice with the greatest activation in each image.

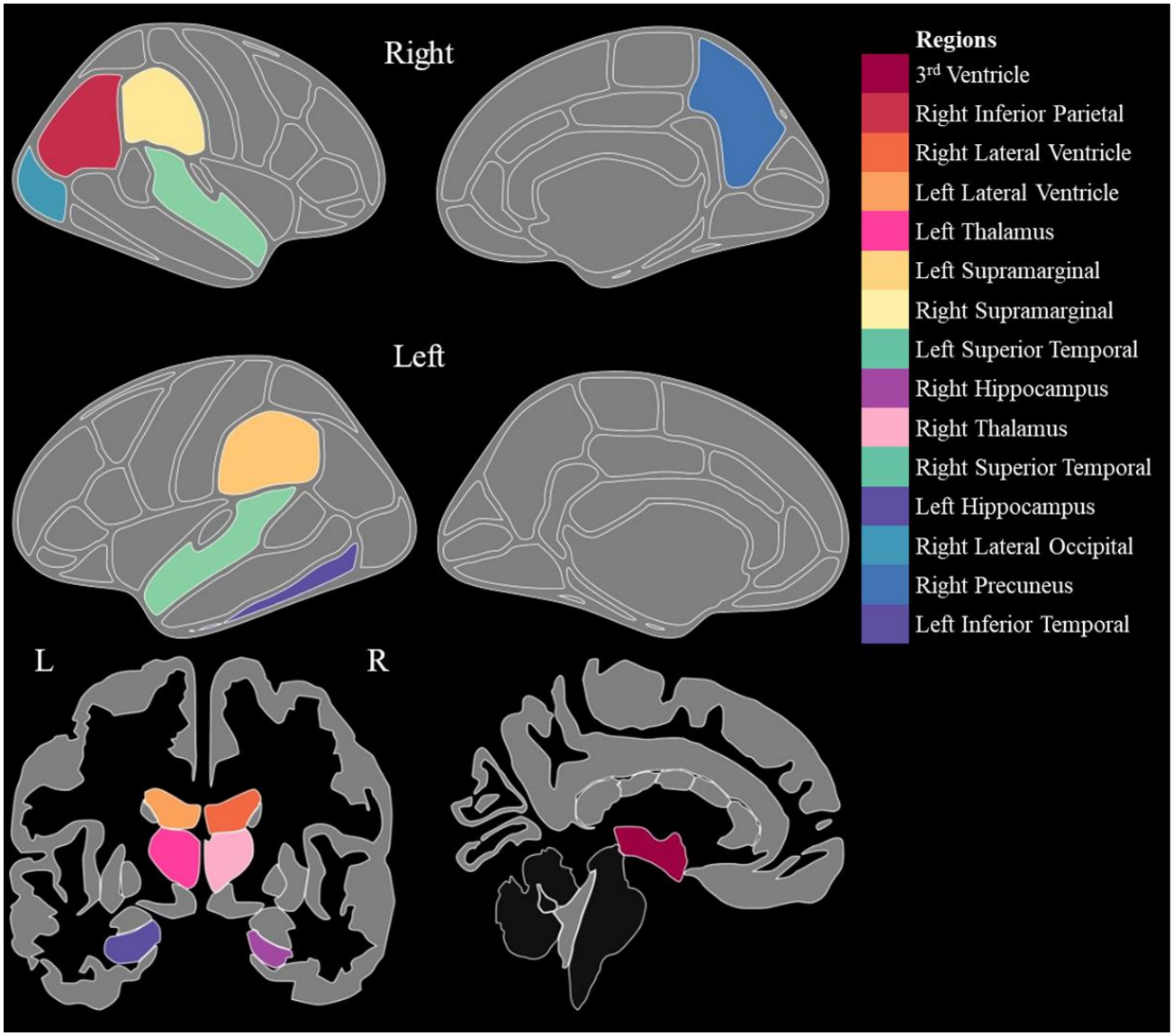

Figure 7. Illustration of the cortical and subcortical brain regions with the highest importance for the predictions in the OPLS model. The presented colours are for illustration purposes and does not represent a real value.

To visualize differences between older- and younger-looking individuals on a cross-sectional level for a particular age, we ran QDEC analyses on the oldest- vs. youngest-looking quartiles for the CNN and OPLS predictions for individuals with chronological age of 70 ± 1 year. *p*-value maps are shown in Figure 8 with thresholds adjusted for false discovery rate at $p < 0.05$. The *p*-value maps based on the OPLS model predictions show widespread significant differences in cortical thickness. However, the CNN model predictions showed only a few small significant regions.

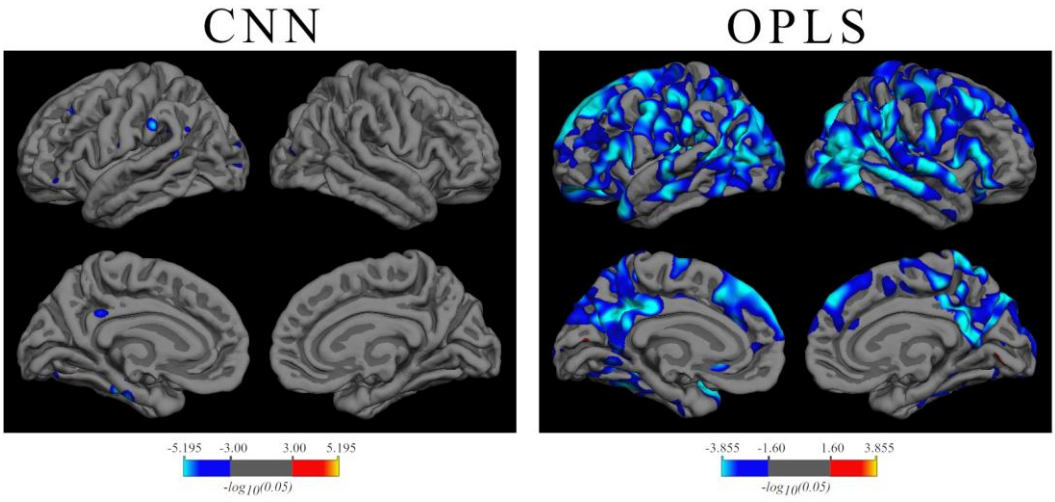

Figure 8. Maps of significant differences in cortical thickness (adjusted for FDR at a threshold of 0.05) between the oldest and youngest quartiles of predictions of individuals with a chronological age of 70±1 to each model. Sex and age were used as covariates in the analysis. *Left*: Quartiles calculated from CNN model predictions. *Right*: Quartiles-based OPLS model predictions. Blue colours represent increased cortical atrophy of the oldest quartiles compared to the younger quartiles of predicted age of individuals with chronological age 70±1, and red colours represent an increased cortical thickness compared to the youngest quartile group.

3.3 VALIDATION IN EXTERNAL DATASET

To validate the models, we used AddNeuroMed and J-ADNI cohorts as external test datasets. The results of both cohorts used as external test datasets are shown in Table 2, Figure 9, and Figure 10.

Table 2. Comparison of average chronological age and predicted brain age by the CNN and OPLS models in the validation cohorts. The values are shown as mean ± standard deviation, and range.

Cohort	**Chronological age**	**Predicted age**		**MAE**	
		CNN	OPLS	CNN	OPLS
AddNeuroMed	75.1 ± 6.1 (53.0 ± 88.9 yrs.)	76.8 ± 1.5 (61.0 – 87.6 yrs.)	76.1 ± 5.6 (62.5 – 92.9 yrs.)	3.96	4.26
J-ADNI	68.5 ± 5.50 (60.0 – 86.0 yrs.)	73.8 ± 1.6 (62.9 – 83.3 yrs.)	71.5 ± 5.0 (57.0 – 82.5 yrs.)	6.06	5.19

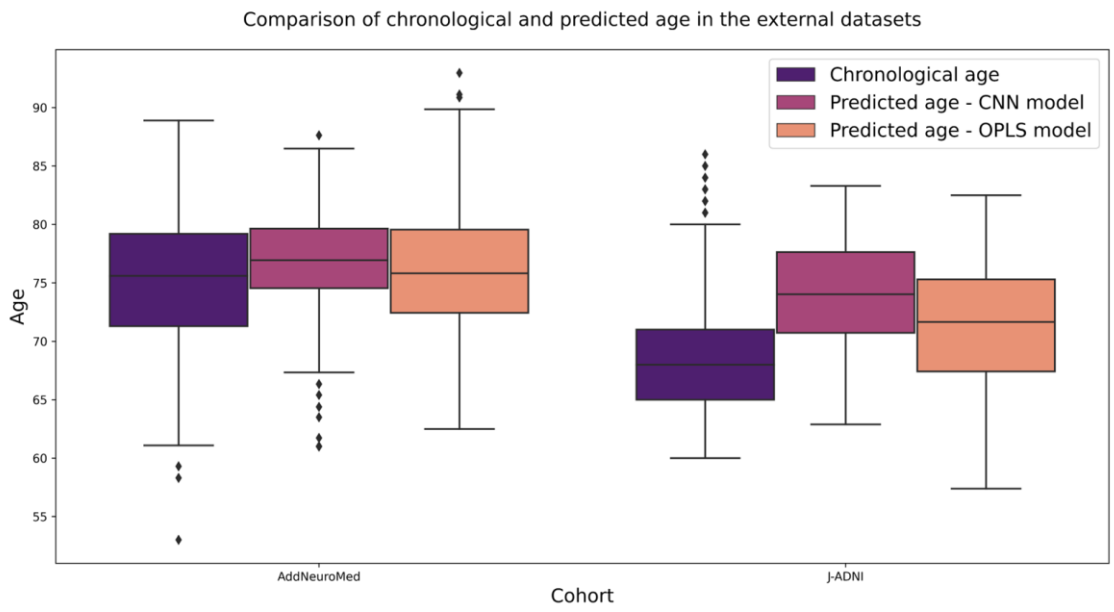

Figure 9. Comparison of chronological and predicted age by the CNN and OPLS model for the validation dataset.

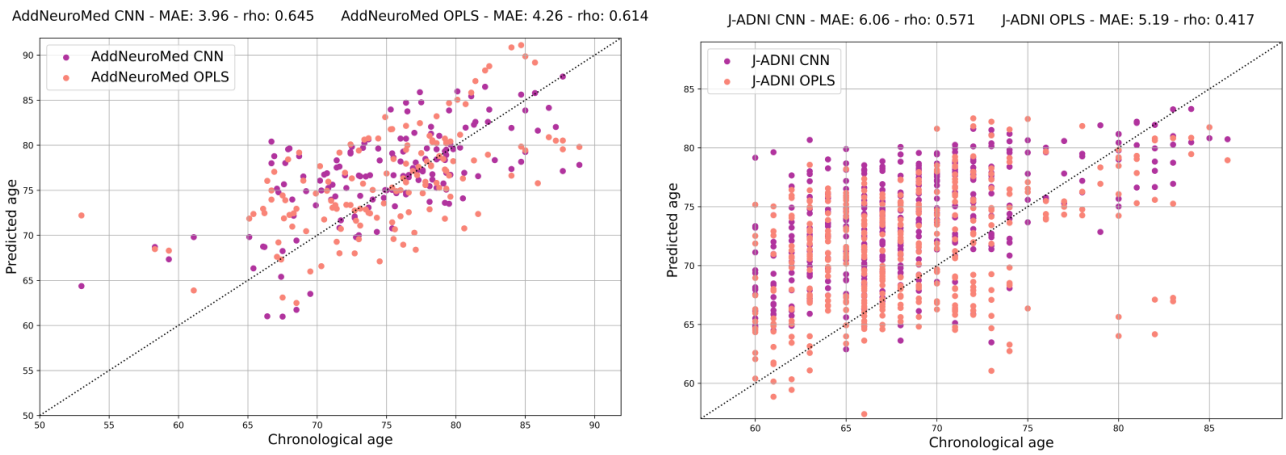

Figure 10. Mean absolute error between chronological and predicted brain age for the CNN and the OPLS models for AddNeuroMed and J-ADNI cohorts.

4 Discussion

In this study, we developed a model to predict the person's biological brain age from raw MRIs using CNN, i.e., a technique within the deep learning framework. We used six cohorts of healthy individuals covering a wide age span. We also predicted brain age using a well-known multivariate approach, that is, OPLS, by training the same sets of MRIs. Finally, we systematically compared the two different brain age prediction models (CNN vs. OPLS), and we compared our models' performance with those reported in the literature for the UK Biobank. Overall, both the CNN and the OPLS models achieved reasonable MAEs that were comparable with those obtained from previous studies, indicating that our models have a good performance. However, on a hold-out test set from the UK-Biobank, our CNN vs. OPLS predictions differed substantially on an individual level, indicating that the CNN model provided a more accurate prediction of brain age. Therefore, any future analysis using the brain age predictions might yield different results and conclusions depending on the choice of model.

4.1 MODEL PERFORMANCE

Both models showed MAE levels comparable to those reported in previous literature, which are typically in the range 2.13 – 8 (Baecker et al., 2021b, 2021a; Bintsi et al., 2020; Cole, 2020; Dinsdale et al., 2021; Gupta et al., 2021; Jonsson et al., 2019; Kolbeinsson et al., 2019; P. K. Lam et al., 2020). The CNN model performed substantially better than the OPLS model on the test set. The lower limit of MAE score is unknown and depends on the inter-subject variability in the test set, as well as the age distribution. By using stricter exclusion criteria for what is considered "healthy", the variability and theoretical MAE lower bound decreases. This makes comparisons between studies of models evaluated on different datasets challenging. Nonetheless, it should be noted that our OPLS model's MAE is still at a level that is comparable to many previously proposed models using similar prediction methods (Cole et al., 2017; Hwang et al., 2020; P. Lam et al., 2020). Still, several studies have trained and evaluated their model on the UK Biobank cohort, which enables fair comparisons (Table 1). However, this restricts the model to cross-sectional image data from a "homogeneous" population from the United Kingdom (UK), acquired in a standard equipment (Siemens Magnetom Skyra Syngo MR D13) with 3T MRI following the same protocol, which is not the reality for datasets and clinical/research settings. In Table 1, which presents the comparison of our results with the literature applied only to UK Biobank images, we observed that even using several time-consuming image preprocessing steps, none of the models achieved an MAE smaller than 2.13. The OPLS model is limited by relying on values of cortical thickness and subcortical gray matter and ventricular volumes for its predictions, as opposed to a CNN that can leverage more complex information from the whole image. The performance of the OPLS model (and any model relying on preprocessed data) is limited by the performance of the image processing software, the resolution of the anatomical atlas, and the selection of features to use as input. The CNN model does not require pre-segmented data and can learn from data what in the image is relevant in aging, including the shape of a structure, which may not be captured by a summary metric such as the volume. On the other hand, the performance of CNNs trained on medical images from one cohort may produce systematically different predictions on images outside the training data distribution (Mårtensson et al., 2020). While well-established neuroimaging tools also suffer from some degree of inter-scanner variability (Guo et al., 2019), it is possible to visually quality control the segmentation results and assess if the input data is valid.

The LASSO model proposed by Cole (2020) showed a lower MAE (3.52) than our OPLS model but used additional information from T_2^*-weighted and fMRI scans as input. When only using measurements from a T1w scan, the MAE was closer (4.14) to the OPLS model's results in our study. Anatürk et al. (2021) proposed the use of the XGBoost model, which achieved a MAE of 3.11, but using T1w, T_2-weighted fluid-attenuated inversion recovery, and diffusion-weighted MRIs. Like the OPLS model, Boyle et al. (2021) used different datasets with gray matter features of T1w MRI to predict brain age using the ElasticNET model, which showed a worse MAE than our OPLS model in the training and test datasets (MAEs 7.28 – 7.60, respectively).

Our CNN model performed slightly worse, yet comparable to most of the CNN models presented in Table 1 and in the state-of-the-art presented by Baecker, Garcia-Dias, et al. (2021). We believe that this could be related to that the input data to our CNN model has not been preprocessed to the same degree as in the previous publications (e.g., no skull-stripping or bias-field correction). The motivation for this trade-off is that more extensive preprocessing limits the accessibility of a model. It also requires that all preprocessing steps are successful prior to feeding it as input to the model, which can limit the usage in some cohorts. Similar to Cole et al. (2017), we tried to keep the images as "raw" as possible, only using rigid registration (6 degrees of freedom) as a preprocessing step. When Cole et al. (2017) used (segmented)

tissue maps instead, the model performed better. The difference between our CNN's MAE (2.87) to the baseline model's MAE in Bintsi et al. (2020) – also a 3D ResNet – of 2.64 is likely attributed to this (training and testing our CNN model only using UK Biobank resulted in a MAE of 2.84). Their patch-based version performed substantially better (MAE = 2.13), which is even lower than the MAE our model performs on the training set. The results in Bintsi et al. (2020) suggest that additional regularization, or even multi-scanner data harmonization, or curriculum learning procedures may improve the performance of our proposed CNN model.

Our longitudinal results (Figure 4) showed another potential confounder in subsequent analyses of brain age predictions: noise in the predictions. As it can be observed, for some subjects, the brain age differed more than five years between two scanning sessions that were one year apart. It seems unlikely that this is a biological phenomenon, but rather attributed to the input data being noisy or of low quality. Most datasets do not have an abundance of longitudinal data to be able to sort "bad" predictions from "good" predictions. If the data and group sizes are large, some level of noise is acceptable and might not affect the interpretation of the results. However, this issue may have a non-negligible effect when analyse data sets that have small group sizes or running longitudinal analysis with few follow-up scans. This is important to keep in mind when conducting studies related to brain age in the future, for example when investigating the association between brain age and neurological disorders that are less common in the population.

Bashyam et al. (2020) showed that a "less-fitted" age prediction model can be better at discriminating healthy controls from AD patients. However, we argue that the primary interest in the ability to predict brain age is not for diagnosing individuals with clinical symptoms of neurodegenerative disorders, as we can create specialized and more accurate models for this purpose. Beyond the diagnostic, predicted brain age models can have important applications in ageing research and, particularly, for the prevention of dementia. For example, brain age models can open new avenues to understand the direct influences of age-related biological processes (e.g., inflammation) on the brain and cognition. At the same time, identifying the modifiable psychosocial and lifestyle factors that promote brain health can help set up intervention and/or strategies aiming at maintaining brain health, thus preventing timely cognitive impairment and dementia. This was shown in a study with rats where a control group was compared to a group under a long-term healthy lifestyle (related to a dietary restriction and enriched environment) (Brusini et al., 2022). That study showed that despite not showing significantly different brain age scores at baseline, the healthy lifestyle group suggests that lifestyle-related intervention approaches may slow down brain ageing.

While both ageing and AD can lead to overall atrophy, the pattern of atrophy in an AD patient is vastly different from that of a healthy individual. Thus, a brain of an AD patient is likely far from the distribution of the training data (comprised of healthy, cognitively unimpaired individuals), meaning that this input is not valid to a reliable and "well- fitted" brain age model (or in other words: there is no reason that excessive atrophy in a key region in AD, e.g., the hippocampus – not represented in the training data – should make the model predict a higher brain age.). Therefore, we recommend that the target population of brain age models comprises individuals who are cognitively healthy or, at maximum, show only subtle cognitive deficits. Many questions are still open on what are the processes that characterize non-pathological ageing vs. those that are typical of pathological ageing. The use of predicted brain age models could pave the way to shed light on this matter. For example, future studies could quantify the degree of brain atrophy in normal ageing vs. that of pathological ageing and identify the lifelong drivers of such processes. Also, they could explore to what extent co-occurring brain pathologies contribute to brain atrophy and subsequent cognitive impairment, as well as the role of specific neuropathological changes. Moreover, based on the differences between predicted brain age and chronological age, different resilience mechanisms in ageing could be investigated. All the above are only a few examples of the vast applications of brain age prediction models in ageing research, which already addresses other diseases, disorders, and

lifestyle habits like epilepsy (Hwang et al., 2020), and smoking and alcohol consumption (Ning et al., 2020), among others (Franke and Gaser, 2019).

4.2 AGEING PROCESS AND MODEL DIFFERENCES

The cross-sectional analysis of the cortical differences between individuals in different chronological/brain age groups (Figure 5 and Supplementary Figure 2) revealed similar patterns when grouped by chronological age and CNN predicted age. This was reasonable given the low MAE score and that ageing is associated with diffuse and widespread atrophy demonstrated in multiple studies (Eavani et al., 2018; Varikuti et al., 2018). The OPLS model showed substantial differences already in 60-year-old individuals. This is probably attributed to its predictions being explicitly based on cortical thickness values from the Desikan atlas, causing some degree of circularity and at flatter slope compared to the CNN model as can be seen in Figure 3Figure 3. That is, the group predicted as, e.g., 65 years-old by the OPLS model have a higher average chronological age than in the case of the CNN model. Further, the OPLS results can be interpreted as that the ageing process past 55 years is characterized initially by atrophy in the posterior cortex, eventually leading to widespread cortical atrophy. This interpretation does not align with the CNN results, which seems to start with unspecific atrophy that becomes dominated by gray matter loss in the frontal lobe (Figure 2 in supplementary material), spreading to temporo-parietal regions, being more in accordance with the age-related differences shown in chronological age in our study.

We observed no significant differences on the cross-sectional maps for older- and younger-looking 70-year-old individuals according to the CNN model (Figure 8). The sensitivity maps shown in Figure 6 indicate that the CNN model based its predictions on CSF spaces next to the insula and ventricular widening. This can be caused by diffuse and unspecific atrophy, which is less likely to yield significant differences in the QDEC analysis shown in Figure 8. Cole (2020) reported that the two T1w measures that were important for age prediction were the normalized gray matter volume and the volume of the fourth ventricle, which is in line with our findings. The OPLS model considered the third and the latero-inferior horns of the lateral ventricles as important variables for the brain age prediction (Supplementary data Figure 3). However, the fourth ventricle itself did not show so much importance for the prediction (Supplementary data Figure 3). The model proposed by Bintsi et al. (2020) also indicated the importance of the ventricles in age prediction, together with the hippocampal region. Since the OPLS model is bounded to use cortical thickness values (together with subcortical gray matter and ventricular volumes), it may not be able to capture subtle global changes. However, this dimensionality reduction has a regularizing effect that – at the expense of performance – can increase interpretability and be sufficient to capture ageing patterns. Nevertheless, due to the higher MAE, it is possible to observe significant differences on the cross-sectional maps for older- and younger-looking 70-year-old individuals in the OPLS model (Figure 8). These differences appear mainly in parieto-occipital and temporal regions, with a cluster present in the frontal lobe, and shows the higher variability in the OPLS model predictions.

Interestingly, one region medially in the primary visual cortex and another region in the anterior cingulate showed higher thickness in older groups in both CNN and OPLS models, when compared with the younger 55 years old group (Figure 5). The shrinkage of the visual cortex is still widely discussed in the literature, with a handful of studies showing from cortical thinning of the visual cortex to the sparing of this region, suggesting that cortical thickness of the visual cortex is use-dependent (e.g., central visual loss with an increase of peripheral vision use) instead of age-related (Burge et al., 2016; Griffis et al., 2016; Jorge et al., 2020). The thicker cortex in the anterior cingulate area has been reported previously in the literature (Ferreira et al., 2014) and has recently been related to brain resistance (Salvadó et al., 2021). Based on these findings, one could suggest that individuals with the diffuse pattern of atrophy observed in Figure 5 may need to

keep spared the strategic anterior cingulate region to stay in the range of normal cognition, as predicted by the compensation hypothesis of healthy ageing (Salvadó et al., 2021).

While many different brain-age prediction models have been presented in the literature, few have looked at actual differences in their predictions. If we believe in the concept of "brain age" – a latent or underlying measure that better reflects the state of the brain than chronological age – how do we interpret the situation where two models produce vastly different brain age scores (see Supplementary Material Figures 4 and 5 for the individual plots of prediction for each model)? A simple explanation is that two models with different training procedures and input data will not yield the same age predictions. This raises the question of which of the two brain age scores is "more correct". As there is no definitive answer to this, researchers need to interpret the findings considering the choice of model, choice of input, preprocessing and regularization methods, inclusion criteria of the cohort, and performance metrics such as MAE. These choices will also need to consider the available computational resources and the amount of data, as more complex models like CNN tend to need more computational resources and to overfit when trained with small datasets. Also, it is also needed to consider all sources of bias that can be generated in the model training (e.g., preprocessing steps, choice of brain template, normalization methods etc.) in the analysis. As the field is moving towards studying the relationship between brain age gap and neurological disorders or cognitive measures (Cole, 2020; Cole et al., 2017; Elliott et al., 2021; Franke and Gaser, 2012; Hwang et al., 2020), this has become even more important to interpret results within and across studies.

4.3 VALIDATION IN EXTERNAL DATASET

The validation of our model was made by testing an external dataset with data of cognitively unimpaired individuals from two different cohorts: AddNeuroMed, based on a European population, and J-ADNI, based on an Asian population. The external dataset was unknown to the model and presented different general characteristics from those used in the training and testing dataset to evaluate the performance and generalizability of the model. Both models presented MAE between 3.96, and 6.06 yrs., and are in the typical range of MAE found in the literature for testing sets (see section 4.1). Predictions from the J-ADNI cohort, presented higher a MAE and average predicted brain age than the AddNeuroMed cohort.

The choice of using AddNeuroMed and J-ADNI cohorts as external datasets was performed to evaluate our model in an inhomogeneous environment. Our external datasets have a different age distribution, different ethnicity, and it mostly uses images acquired in 1.5T scanners, whereas the training dataset was based mainly on 3T MRI (96.2 % of all images used for training and testing). The range of the predicted age by the CNN and OPLS models is comparable to the range of chronological age. Also, as shown in the training-testing dataset, the predicted ages by both models were higher than the chronological age, as expected.

The AddNeuroMed predictions were similar to the CNN and OPLS models and comparable to the chronological age. A smaller MAE was observed in the CNN model compared to the OPLS model. This was also seen in the training-testing step. Unexpectedly, in the J-ADNI cohort, the calculated MAE for the OPLS model was smaller than for the CNN model. A potential explanation could be the difference in ethnicity which has been discussed elsewhere (Choi et al., 2020; Kang et al., 2020; Tang et al., 2018). Different ethnical populations can present differences in brain morphology. Our training model only includes data from Europe and North America, while J-ADNI is an Asian cohort. The unexpected difference with a lower MAE for the OPLS model can be due to that some of the variations disappears due to the preprocessing steps presented in the OPLS model. Further studies need to evaluate how the inclusion of different populations influences brain age prediction models based on minimally processed MRIs, but it is likely that the training model needs to include more

diverse data, i.e., different types of cohorts, as well as a wider range of scanners as well as imaging protocols (Mårtensson et al., 2020).

4.4 STRENGTHS AND LIMITATIONS

This study was performed as an effort to attend the highlighted points related to an ideal image database to develop brain age prediction models. We used detailed information and clinical data for inclusion criteria of many individuals (more than 16000 T1w MRIs) from different parts of the world (Asia, Australia, Europe, and America) and with a diverse number of MRI scanners covering 1.5T and 3T field strengths, which increases the generalizability of the model. We also used all the longitudinal data available and showed that our model presents a reasonable age prediction congruent with the individual timeline, highlighting some discrepant cases (Figure 4). Also, our focus on the cognitive "healthy" status, rather than on the overall health status (for example, excluding from the training set age-related diseases that affect the body's organs other than the brain) is a strength of our study as it makes a clear separation between the outcomes (brain age) and what leads to that outcome (risk factors, e.g., chronic cardiometabolic disease and risk factors, affective/mood disorders).

Another strength of this work is the use of minimal image processing prior to the input of the image to the CNN model. This can facilitate the implementation in research and clinical centres, as the only needed processing is the registration to MNI space that usually takes a few seconds to process and is available in several open brain image processing software, e.g., FSL and FreeSurfer.

Some limitations need to be acknowledged. First, the large dataset included in the current study hindered the possibility of performing an extensive quality control (QC). For the CNN model, this would mean inspecting that the rigid registration was sufficient. For the OPLS model, it would involve performing QC on all FreeSurfer segmentations – a task that would take months to do manually. The random cases we inspected suggested that the overall quality of the segmentations was sufficient, but tools to automate the QC process – such as Brusini et al. (2020) and Klapwijk et al. (2019) – will be necessary for future studies on this size of data. Regardless of the lack of extensive QC in the images, our model showed robust findings with only slightly worse performance when compared to previously published works (Table 1).

Second, the differences in the distribution of chronological age and the number of subjects in each cohort can lead to overfitting of some cohort specific information and characteristics (e.g., the model could learn that images from GENIC are generally in the lower age span). However, the large size of the training set, heavy data augmentation, and only running 20 epochs in training helped minimize the possibility of overfitting. Third, although we defined the cognitively "healthy" status as consistently as possible across the cohorts, the clinical and cognitive assessments relied on similar as well as different sources in each cohort. For instance, we excluded individuals with neurological disorders, which were identified with the ICD-9/10 codes in the UK-Biobank, but not in the other cohorts because of the lack of availability of such data. This might have led to misclassifying some cases with subtle cognitive deficits as cognitively healthy, leading to possibly higher MAE estimation. Finally, it is worth noting that there might be sociodemographic differences between the cohorts, since the recruitment of participants happened in different geographical areas (J-ADNI: Japan, AIBL: Australia, ADNI: North America, UK Biobank, GENIC, AddNeuroMed: Europe). However, this is not necessarily a limitation, but rather a strength. Indeed, the developed algorithm (see Data Availability section) could be applied in future research, in which the biological age of the brain is a focus, to some extent independently of the cohort characteristics, thus increasing the

generalizability of the model. Future studies need to further test this hypothesis as well as the impact of different cultural backgrounds on the estimation of brain age.

5 CONCLUSION

In this study, we developed a predicted brain age CNN model and systematically compared it to a widely used multivariate model based on orthogonal projections to latent structures. Both models were trained and evaluated on the same cohort, comprising more than 16000 T1w MRIs. We validate our findings in an external dataset with two different cohorts, that differ mainly in the MRI field strength when compared to the training and hold-out dataset. While both models demonstrated acceptable performance metrics on a hold-out test set, individual predictions differed substantially, with brain age patterns of the CNN model being more comparable to those presented in chronological age. The consensus between older- and younger-looking brains was small between the models, and subsequent interpretation of the predictions varied depending on the choice of model. These findings demonstrate the importance of understanding the impact of the choice of model in interpreting data based on brain age predictions. Nonetheless, our CNN model showed similar results to state-of-the-art methods in the hold-out and external dataset, using minimally processed images, which facilitates the future implementation of brain age prediction in research and particularly in clinical settings. This accessibility may promote testing and implementing brain age for diagnosis, prognosis, and monitoring of treatment response in different age-related brain disorders.

6 DATA AVAILABILITY

The full code for training the CNN is available at https://bitbucket.org/ericwestman/brainage-prediction-mri/src/master/. We include the trained weights for researchers to apply to their own neuroimaging data sets, together with scripts for training their own model on other datasets.

7 ACKNOWLEDGEMENTS

The authors wish to thank the Swedish Foundation for Strategic Research (SSF), The Swedish Research Council (VR), the Strategic Research Programme in Neuroscience at Karolinska Institutet (StratNeuro), Swedish Brain Power, Center for Innovative Medicine (CIMED), the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, Swedish Brain Foundation, Swedish Alzheimer Foundation, Åke Wiberg Foundation, Stiftelsen Olle Engkvist Byggmästare, the joint research funds of KTH Royal Institute of Technology and Stockholm County Council (HMT), David and Astrid Hagelén Stiftelsen, Loo och Hans Ostermans stiftelse, Birgitta och Sten Westerberg, Vinnova and Digital Futures for financial support, and the NIA-supported Collaboratory on Research Definitions for reserve and resilience in cognitive ageing and dementia, and the Swedish Brain Research Foundation. The Titan X Pascal used for this research was donated by the NVIDIA Corporation.

This research has been conducted using the UK Biobank Resource under Application Number 30172. The UK Biobank study was approved by the National Health Service National Research Ethics Service (11/NW/0382). More information about UK Biobank is available at www.ukbiobank.ac.uk.

The AddNeuroMed study was supported by InnoMed, (Innovative Medicines in Europe) an Integrated Project funded by the European Union of the Sixth Framework program priority FP6-2004-LIFESCIHEALTH-5, Life Sciences, Genomics and Biotechnology for Health.

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

J-ADNI was supported by the following grants: Translational Research Promotion Project from the New Energy and Industrial Technology Development Organization of Japan; Research on Dementia, Health Labor Sciences Research Grant; Life Science Database Integration Project of Japan Science and Technology Agency; Research Association of Biotechnology (contributed by Astellas Pharma Inc., Bristol-Myers Squibb, Daiichi-Sankyo, Eisai, Eli Lilly and Company, Merck-Banyu, Mitsubishi Tanabe Pharma, Pfizer Inc., Shionogi & Co., Ltd., Sumitomo Dainippon, and Takeda Pharmaceutical Company), Japan, and a grant from an anonymous Foundation

REFERENCES

Anatürk, M., Kaufmann, T., Cole, J.H., Suri, S., Griffanti, L., Zsoldos, E., Filippini, N., Singh-Manoux, A., Kivimäki, M., Westlye, L.T., Ebmeier, K.P., de Lange, A.M.G., 2021. Prediction of brain age and cognitive age: Quantifying brain and cognitive maintenance in aging. Hum Brain Mapp 42, 1626–1640. https://doi.org/10.1002/hbm.25316

Baecker, L., Dafflon, J., da Costa, P.F., Garcia-Dias, R., Vieira, S., Scarpazza, C., Calhoun, V.D., Sato, J.R., Mechelli, A., Pinaya, W.H.L., 2021a. Brain age prediction: A comparison between machine learning models using region- and voxel-based morphometric data. Hum Brain Mapp 42, 2332–2346. https://doi.org/10.1002/hbm.25368

Baecker, L., Garcia-Dias, R., Vieira, S., Scarpazza, C., Mechelli, A., 2021b. Machine learning for brain age prediction: Introduction to methods and clinical applications. EBioMedicine. https://doi.org/10.1016/j.ebiom.2021.103600

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. https://doi.org/10.1093/brain/awaa160

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

Bintsi, K.M., Baltatzis, V., Kolbeinsson, A., Hammers, A., Rueckert, D., 2020. Patch-Based Brain Age Estimation from MR Images, in: Lecture Notes in Computer Science. Springer Science and Business Media Deutschland GmbH, pp. 98–107. https://doi.org/10.1007/978-3-030-66843-3_10

Birkenbihl, C., Westwood, S., Shi, L., Nevado-Holgado, A., Westman, E., Lovestone, S., Hofmann-Apitius, M., 2021. ANMerge: A Comprehensive and Accessible Alzheimer's Disease Patient-Level Dataset. J Alzheimers Dis 79, 423–431. https://doi.org/10.3233/JAD-200948

Bocancea, D.I., van Loenhoud, A.C., Groot, C., Barkhof, F., van der Flier, W.M., Ossenkoppele, R., 2021. Measuring Resilience and Resistance in Aging and Alzheimer Disease Using Residual Methods: A Systematic Review and Meta-analysis. Neurology. https://doi.org/10.1212/WNL.0000000000012499

Boscolo Galazzo, I., Cruciani, F., Brusini, L., Salih, A., Radeva, P., Storti, S.F., Menegaz, G., 2022. Explainable Artificial Intelligence for Magnetic Resonance Imaging Aging Brainprints: Grounds and challenges. IEEE Signal Process Mag 39, 99–116. https://doi.org/10.1109/MSP.2021.3126573

Boyle, R., Jollans, L., Rueda-Delgado, L.M., Rizzo, R., Yener, G.G., McMorrow, J.P., Knight, S.P., Carey, D., Robertson, I.H., Emek-Savaş, D.D., Stern, Y., Kenny, R.A., Whelan, R., 2021. Brain-predicted age difference score is related to specific cognitive functions: a multi-site replication analysis. Brain Imaging Behav 15, 327–345. https://doi.org/10.1007/s11682-020-00260-3

Brusini, I., Ferreira Padilla, D., Barroso, J., Skoog, I., Smedby, O., Westman, E., Wang, C., Se, C., 2020. A deep learning-based pipeline for error detection and quality control of brain MRI segmentation results.

Brusini, I., MacNicol, E., Kim, E., Smedby, Ö., Wang, C., Westman, E., Veronese, M., Turkheimer, F., Cash, D., 2022. MRI-derived brain age as a biomarker of ageing in rats: validation using a healthy lifestyle intervention. Neurobiol Aging 109, 204–215. https://doi.org/10.1016/J.NEUROBIOLAGING.2021.10.004

Burge, W.K., Griffis, J.C., Nenert, R., Elkhetali, A., Decarlo, D.K., Ver Hoef, L.W., Ross, L.A., Visscher, K.M., 2016. Cortical thickness in human V1 associated with central vision loss. Nature Publishing Group. https://doi.org/10.1038/srep23268

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. https://doi.org/10.1002/hbm.25533

Choi, Y.Y., Lee, J.J., Choi, K.Y., Seo, E.H., Choo, I.H., Kim, H., Song, M.K., Choi, S.M., Cho, S.H., Kim, B.C., Lee, K.H., 2020. The Aging Slopes of Brain Structures Vary by Ethnicity and Sex: Evidence From a Large Magnetic Resonance Imaging Dataset From a Single Scanner of Cognitively Healthy Elderly People in Korea. Front Aging Neurosci 12, 233. https://doi.org/10.3389/FNAGI.2020.00233/BIBTEX

Cole, J.H., 2020. Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors. Neurobiol Aging 92, 34–42. https://doi.org/10.1016/j.neurobiolaging.2020.03.014

Cole, J.H., Franke, K., 2017. Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. Trends Neurosci. https://doi.org/10.1016/j.tins.2017.10.001

Cole, J.H., Marioni, R.E., Harris, S.E., Deary, I.J., 2019. Brain age and other bodily 'ages': implications for neuropsychiatry. Mol Psychiatry. https://doi.org/10.1038/s41380-018-0098-1

Cole, J.H., Poudel, R.P.K., Tsagkrasoulis, D., Caan, M.W.A., Steves, C., Spector, T.D., Montana, G., 2017. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. Neuroimage 163, 115–124. https://doi.org/10.1016/j.neuroimage.2017.07.059

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179–194. https://doi.org/10.1006/NIMG.1998.0395

Dinsdale, N.K., Bluemke, E., Smith, S.M., Arya, Z., Vidaurre, D., Jenkinson, M., Namburete, A.I.L., 2021. Learning patterns of the ageing brain in MRI using deep convolutional networks. Neuroimage 224. https://doi.org/10.1016/j.neuroimage.2020.117401

Eavani, H., Habes, M., Satterthwaite, T.D., An, Y., Hsieh, M.K., Honnorat, N., Erus, G., Doshi, J., Ferrucci, L., Beason-Held, L.L., Resnick, S.M., Davatzikos, C., 2018. Heterogeneity of structural and functional imaging patterns of advanced brain aging revealed via machine learning methods. Neurobiol Aging 71, 41–50. https://doi.org/10.1016/j.neurobiolaging.2018.06.013

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., Caspi, • Avshalom, 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. https://doi.org/10.1038/s41380-019-0626-7

Ellis, K.A., Bush, A.I., Darby, D., de Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P., Masters, C., Milner, A., Pike, K., Rowe, C., Savage, G., Szoeke, C., Taddei, K., Villemagne, V., Woodward, M., Ames, D., Acosta, O., Ames, J., Agarwal, M., Bahar-Fuchs, A., Baxendale, D., Bechta-Metti, K., Bevage, C., Bevege, L., Bourgeat, P., Brown, B., Clarnette, R., Cowie, T., Crowley, K., Currie, A., El-Sheikh, D., Dickinson, K., Farrow, M., Faux, N., Fripp, J., Fowler, C., Gupta, V., Jones, G., Khoo, J., Killedar, A., Killeen, N., Tae, W.K., Kotsopoulos, E., Lachovitzki, R., Li, Q.X., Liang, X., Lucas, K., Lui, J., Martins, G., Montague, C., Moore, L., Muir, A., O'Halloran, C., O'Keefe, G., Panayiotou, A., Paton, A., Paton, J., Peiffer, J., Pejoska, S., Pertile, K., Porter, L., Price, R., Raniga, P., Rees, G., Rembach, A., Rimajova, M., Robins, P., Ronsisvalle, E., Rumble, R., Rodrigues, M., Salvado, O., Sach, J., Samuel, M., Sittironnarit, G., Taddei, T., Trivedi, D., Trounson, B., Tsikkos, M., Walker, S., Ward, V., Yastrubetskaya, O., 2009. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr 21, 672–687. https://doi.org/10.1017/S1041610209009405

Eriksson, L., Johansson, E., Kettenah-Wold, N., Trygg, J., Wikström, C., Wold, S., 2013. Multi- and Megavariate Data Analysis.

Falahati, F., Westman, E., Simmons, A., 2014. Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. J Alzheimers Dis 41, 685–708. https://doi.org/10.3233/JAD-131928

Ferreira, D., Correia, R., Nieto, A., Machado, A., Molina, Y., Barroso, J., 2015a. Cognitive decline before the age of 50 can be detected with sensitive cognitive measures. Psicothema 27, 216–222. https://doi.org/10.7334/PSICOTHEMA2014.192

Ferreira, D., Correia, R., Nieto, A., Machado, A., Molina, Y., Barroso, J., 2015b. Cognitive decline before the age of 50 can be detected with sensitive cognitive measures. Psicothema 27, 216–222. https://doi.org/10.7334/psicothema2014.192

Ferreira, D., Molina, Y., Machado, A., Westman, E., Wahlund, L.O., Nieto, A., Correia, R., Junqué, C., Díaz-Flores, L., Barroso, J., 2014. Cognitive decline is mediated by gray matter changes during middle age. Neurobiol Aging 35, 1086–1094. https://doi.org/10.1016/J.NEUROBIOLAGING.2013.10.095

Ferreira, D., Nordberg, A., Westman, E., 2020. Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. Neurology. https://doi.org/10.1212/WNL.0000000000009058

Fischl, B., Salat, D.H., Van Der Kouwe, A.J.W., Makris, N., Ségonne, F., Quinn, B.T., Dale, A.M., 2004. Sequence-independent segmentation of magnetic resonance images. https://doi.org/10.1016/j.neuroimage.2004.07.016

Franke, K., Gaser, C., 2019. Ten years of brainage as a neuroimaging biomarker of brain aging: What insights have we gained? Front Neurol. https://doi.org/10.3389/fneur.2019.00789

Franke, K., Gaser, C., 2012. Longitudinal Changes in Individual BrainAGE in Healthy Aging, Mild Cognitive Impairment, and Alzheimer's Disease. GeroPsych (Bern) 25, 235–245. https://doi.org/10.1024/1662-9647/a000074

Franke, K., Ziegler, G., Klöppel, S., Gaser, C., 2010. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. Neuroimage 50, 883–892. https://doi.org/10.1016/j.neuroimage.2010.01.005

Fratiglioni, L., Dekhtyar, S., Fratiglioni, L., Marseglia, A., Dekhtyar, S., 2020. Ageing without dementia: can stimulating psychosocial and lifestyle experiences make a difference? 19, 533–543. https://doi.org/10.1016/S1474-4422(20)30039-9

Glorioso, C.A., Pfenning, A.R., Lee, S.S., Bennett, D.A., Sibille, E.L., Kellis, M., Guarente, L.P., 2019. Rate of brain aging and APOE ϵ4 are synergistic risk factors for Alzheimer's disease. Life Sci Alliance 2. https://doi.org/10.26508/lsa.201900303

Griffis, J.C., Burge, W.K., Visscher, K.M., 2016. Age-dependent cortical thinning of peripheral visual field representations in primary visual cortex. Front Aging Neurosci 8, 248. https://doi.org/10.3389/fnagi.2016.00248

Guo, C., Ferreira, D., Fink, K., Westman, E., Granberg, T., 2019. Repeatability and reproducibility of FreeSurfer, FSL-SIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. Eur Radiol 29, 1355–1364. https://doi.org/10.1007/S00330-018-5710-X

Gupta, U., Lam, P.K., Steeg, G. Ver, Thompson, P.M., 2021. Improved brain age estimation with slice-based set networks, in: Proceedings - International Symposium on Biomedical Imaging. IEEE Computer Society, pp. 840–844. https://doi.org/10.1109/ISBI48211.2021.9434081

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

Hepp, T., Blum, D., Armanious, K., Schölkopf, B., Stern, D., Yang, B., Gatidis, S., 2021. Uncertainty estimation and explainability in deep learning-based age estimation of the human brain: Results from the German National Cohort MRI study. Computerized Medical Imaging and Graphics 92, 101967. https://doi.org/10.1016/j.compmedimag.2021.101967

Hou, Y., Dan, X., Babbar, M., Wei, Y., Hasselbalch, S.G., Croteau, D.L., Bohr, V.A., 2019. Ageing as a risk factor for neurodegenerative disease. Nat Rev Neurol 15, 565–581. https://doi.org/10.1038/S41582-019-0244-7

Hwang, G., Hermann, B., Nair, V.A., Conant, L.L., Dabbs, K., Mathis, J., Cook, C.J., Rivera-Bonet, C.N., Mohanty, R., Zhao, G., Almane, D.N., Nencka, A., Felton, E., Struck, A.F., Birn, R., Maganti, R., Humphries, C.J., Raghavan, M., DeYoe, E.A., Bendlin, B.B., Prabhakaran, V., Binder, J.R., Meyerand, M.E., 2020. Brain aging in temporal lobe epilepsy: Chronological, structural, and functional. Neuroimage Clin 25. https://doi.org/10.1016/j.nicl.2020.102183

Iwatsubo, T., Iwata, A., Suzuki, K., Ihara, R., Arai, H., Ishii, K., Senda, M., Ito, K., Ikeuchi, T., Kuwano, R., Matsuda, H., Sun, C.K., Beckett, L.A., Petersen, R.C., Weiner, M.W., Aisen, P.S., Donohue, M.C., 2018. Japanese and North American Alzheimer's Disease Neuroimaging Initiative studies: Harmonization for international trials. Alzheimer's and Dementia 14, 1077–1087. https://doi.org/10.1016/J.JALZ.2018.03.009

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L.G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging 27, 685–691. https://doi.org/10.1002/JMRI.21049

Jack, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 9, 119–128. https://doi.org/10.1016/S1474-4422(09)70299-6

Jonsson, B.A., Bjornsdottir, G., Thorgeirsson, T.E., Ellingsen, L.M., Walters, G.B., Gudbjartsson, D.F., Stefansson, H., Stefansson, K., Ulfarsson, M.O., 2019. Brain age prediction using deep learning uncovers associated sequence variants. Nat Commun 10. https://doi.org/10.1038/S41467-019-13163-9

Jorge, L., Canário, N., Quental, H., Bernardes, R., Castelo-Branco, M., 2020. Is the Retina a Mirror of the Aging Brain? Aging of Neural Retina Layers and Primary Visual Cortex Across the Lifespan. Front Aging Neurosci 11. https://doi.org/10.3389/fnagi.2019.00360

Jylhävä, J., Pedersen, N.L., Hägg, S., 2017. Biological Age Predictors. EBioMedicine 21, 29–36. https://doi.org/10.1016/J.EBIOM.2017.03.046

Kang, D.W., Wang, S.M., Na, H.R., Park, S.Y., Kim, N.Y., Lee, C.U., Kim, D., Son, S.J., Lim, H.K., 2020. Differences in cortical structure between cognitively normal East Asian and Caucasian older adults: a surface-based morphometry study. Sci Rep 10. https://doi.org/10.1038/S41598-020-77848-8

Klapwijk, E.T., van de Kamp, F., van der Meulen, M., Peters, S., Wierenga, L.M., 2019. Qoala-T: A supervised-learning tool for quality control of FreeSurfer segmented MRI data. Neuroimage 189, 116–129. https://doi.org/10.1016/J.NEUROIMAGE.2019.01.014

Kolbeinsson, A., Kossaifi, J., Panagakis, Y., Bulat, A., Anandkumar, A., Tzoulaki, I., Matthews, P., 2019. Robust deep networks with randomized tensor regression layers. ArXiv.

Lam, P., Zhu, A.H., Gari, I.B., Jahanshad, N., Thompson, P.M., 2020. 3D Grid-Attention Networks for Interpretable Age and Alzheimer's Disease Prediction from Structural MRI.

Lam, P.K., Santhalingam, V., Suresh, P., Baboota, R., Zhu, A.H., Thomopoulos, S.I., Jahanshad, N., Thompson, P.M., 2020. Accurate brain age prediction using recurrent slice-based networks, in: BioRxiv. Cold Spring Harbor Laboratory, p. 32. https://doi.org/10.1117/12.2579630

Liang, H., Zhang, F., Niu, X., 2019. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. Hum Brain Mapp 40, 3143–3152. https://doi.org/10.1002/hbm.24588

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med Image Anal 42, 60–88. https://doi.org/10.1016/J.MEDIA.2017.07.005

Lombardi, A., Diacono, D., Amoroso, N., Monaco, A., Tavares, J.M.R.S., Bellotti, R., Tangaro, S., 2021. Explainable Deep Learning for Personalized Age Prediction With Brain Morphology. Front Neurosci 15, 578. https://doi.org/10.3389/fnins.2021.674055

Machado, A., Barroso, J., Molina, Y., Nieto, A., Díaz-Flores, L., Westman, E., Ferreira, D., 2018. Proposal for a hierarchical, multidimensional, and multivariate approach to investigate cognitive aging. Neurobiol Aging 71, 179–188. https://doi.org/10.1016/j.neurobiolaging.2018.07.017

Mårtensson, G., Ferreira, D., Granberg, T., Cavallin, L., Oppedal, K., Padovani, A., Rektorova, I., Bonanni, L., Pardini, M., Kramberger, M.G., Taylor, J.P., Hort, J., Snædal, J., Kulisevsky, J., Blanc, F., Antonini, A., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Lovestone, S., Simmons, A., Aarsland, D., Westman, E., 2020. The reliability of a deep learning model in clinical out-of-distribution MRI data: A multicohort study. Med Image Anal 66. https://doi.org/10.1016/j.media.2020.101714

Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L.R., Griffanti, L., Douaud, G., Okell, T.W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., Matthews, P.M., Smith, S.M., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci 19, 1523–1536. https://doi.org/10.1038/nn.4393

Muehlboeck, J.S., Westman, E., Simmons, A., 2013. TheHiveDB image data management and analysis framework. Front Neuroinform 7. https://doi.org/10.3389/FNINF.2013.00049

Ning, K., Zhao, L., Matloff, W., Sun, F., Toga, A.W., 2020. Association of relative brain age with tobacco smoking, alcohol consumption, and genetic variants. Sci Rep 10. https://doi.org/10.1038/S41598-019-56089-4

Niu, X., Zhang, F., Kounios, J., Liang, H., 2020. Improved prediction of brain age using multimodal neuroimaging data. Hum Brain Mapp 41, 1626–1643. https://doi.org/10.1002/HBM.24899

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Adv Neural Inf Process Syst 32.

Peng, H., Gong, W., Beckmann, C.F., Vedaldi, A., Smith, S.M., 2021. Accurate brain age prediction with lightweight deep neural networks. Med Image Anal 68, 101871. https://doi.org/10.1016/j.media.2020.101871

Rowe, C.C., Ellis, K.A., Rimajova, M., Bourgeat, P., Pike, K.E., Jones, G., Fripp, J., Tochon-Danguy, H., Morandeau, L., O'Keefe, G., Price, R., Raniga, P., Robins, P., Acosta, O., Lenzo, N., Szoeke, C., Salvado, O., Head, R., Martins,

R., Masters, C.L., Ames, D., Villemagne, V.L., 2010. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiol Aging 31, 1275–1283. https://doi.org/10.1016/J.NEUROBIOLAGING.2010.04.007

Sadeghi-Bazargani, H., Fathi, Savadi-Oskouei, Farhoudi, M., Mehdizadeh Esfanjani, 2012. OPLS statistical model versus linear regression to assess sonographic predictors of stroke prognosis. Neuropsychiatr Dis Treat 8, 387. https://doi.org/10.2147/ndt.s33991

Salvadó, G., Ferreira, D., Operto, G., Cumplido-Mayoral, I., Arenaza-Urquijo, E.M., Cacciaglia, R., Falcon, C., Vilor-Tejedor, N., Minguillon, C., Groot, C., van der Flier, W.M., Barkhof, F., Scheltens, P., Ossenkoppele, R., Kern, S., Zettergren, A., Skoog, I., Hort, J., Stomrud, E., van Westen, D., Hansson, O., Molinuevo, J.L., Wahlund, L.O., Westman, E., Gispert, J.D., Beteta, A., Cañas, A., Crous-Bou, M., Deulofeu, C., Dominguez, R., Emilio, M., Fauria, K., de Echevarri, J.M.G., Grau-Rivera, O., Hernandez, L., Huesa, G., Huguet, J., Marne, P., Menchón, T., Milà-Alomà, M., Polo, A., Pradas, S., Sala-Vila, A., Sánchez-Benavides, G., Soteras, A., Suárez-Calvet, M., Vilanova, M., 2021. The protective gene dose effect of the APOE ε2 allele on gray matter volume in cognitively unimpaired individuals. Alzheimer's & Dementia. https://doi.org/10.1002/ALZ.12487

Simmons, A., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., Kåoszewska, I., Wahlund, L.O., Soininen, H., Lovestone, S., Evans, A., Spenger, C., 2011. The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease: experience from the first 24 months. Int J Geriatr Psychiatry 26, 75–82. https://doi.org/10.1002/GPS.2491

Smilkov, D., Thorat, N., Kim, B., Viégas, F., Wattenberg, M., 2017. SmoothGrad: removing noise by adding noise.

Stern, Y., Arenaza-Urquijo, E.M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., Ewers, M., Franzmeier, N., Kempermann, G., Kremen, W.S., Okonkwo, O., Scarmeas, N., Soldan, A., Udeh-Momoh, C., Valenzuela, M., Vemuri, P., Vuoksimaa, E., Urquiljo, E.M.A., Cantillon, M., Clouston, S.A.P., Estanga, A., Gold, B., Habeck, C., Jones, R., Kochhann, R., Lim, Y.Y., Martínez-Lage, P., Morbelli, S., Okonkwo, O., Ossenkoppele, R., Pettigrew, C., Rosen, A.C., Song, X., Van Loenhoud, A.C., 2020. Whitepaper: Defining and investigating cognitive reserve, brain reserve and brain maintenance. Alzheimers Dement 16, 1305. https://doi.org/10.1016/J.JALZ.2018.07.219

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., Collins, R., 2015. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med 12, e1001779. https://doi.org/10.1371/journal.pmed.1001779

Tang, Y., Zhao, L., Lou, Y., Shi, Y., Fang, R., Lin, X., Liu, S., Toga, A., 2018. Brain structure differences between Chinese and Caucasian cohorts: A comprehensive morphometry study. Hum Brain Mapp 39, 2147–2155. https://doi.org/10.1002/HBM.23994

Trygg, J., Wold, S., 2002. Orthogonal projections to latent structures (O-PLS). J Chemom 16, 119–128. https://doi.org/10.1002/CEM.695

Varikuti, D.P., Genon, S., Sotiras, A., Schwender, H., Hoffstaedter, F., Patil, K.R., Jockwitz, C., Caspers, S., Moebus, S., Amunts, K., Davatzikos, C., Eickhoff, S.B., 2018. Evaluation of non-negative matrix factorization of grey matter in age prediction. Neuroimage 173, 394–410. https://doi.org/10.1016/J.NEUROIMAGE.2018.03.007

Westman, E., Aguilar, C., Muehlboeck, J.S., Simmons, A., 2013. Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. Brain Topogr 26, 9–23. https://doi.org/10.1007/s10548-012-0246-x

Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J.S., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Spenger, C., Wahlund, L.O., 2011. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. Neuroimage 54, 1178–1187. https://doi.org/10.1016/J.NEUROIMAGE.2010.08.044

WHO guidelines, 2019. Risk Reduction Of Cognitive Decline And Dementia, WHO.