medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



# A DEEP LEARNING MODEL FOR BRAIN AGE PREDICTION USING MINIMALLY PRE-PROCESSED T1W-IMAGES AS INPUT

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 the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or 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 aging (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 the analysis or writing of this report. AIBL researchers are listed at https://aibl.csiro.au/.

[3]Data used in the 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 the 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.

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .




## 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 developed and validated a convolutional neural network (CNN)-based biological brain age prediction model that uses only one T1w MRI pre-processing step to simplify implementation and increase accessibility in research settings. Our model only requires rigid image registration to the MNI space, which is an advantage compared to previous methods that require more pre-processing steps, such as feature extraction. We used a multicohort dataset of cognitively healthy individuals (age range = 32.0 – 95.7 yrs.) comprising 17296 MRIs for training and evaluation. We compared our model using hold-out (CNN1) and cross-validation (CNN2-4) approaches. To verify generalizability, we used two external datasets with different population and MRI scan characteristics to evaluate the model. To demonstrate its usability, we included the external dataset's images in the cross-validation training (CNN3). To ensure that our model used only the brain signal on the image, we also predicted brain age using skull-stripped images (CNN4). The trained models achieved a mean absolute error of 2.99, 2.67, 2.67, and 3.08 yrs. for the CNN1-4, respectively. The model's performance in the external dataset was in the typical range of mean absolute error (MAE) found in the literature for testing sets. Adding the external dataset to the training set (CNN3), overall, MAE is unaffected, but individual cohort MAE improves (2.25 to 5.63 years). Salience maps of predictions reveal that periventricular, temporal, and insular regions are the most important for age prediction. We provide indicators for using biological (predicted) brain age as a metric for age correction in neuroimaging studies as an alternative to the traditional chronological age. In conclusion, using different approaches, our CNN-based model showed good performance using only one T1w brain MRI pre-processing step. The proposed CNN model is made publicly available for the research community to be easily implemented and used to study aging and age-related disorders.


*Keywords: brain age, neurodegeneration, normal aging, CNN, UK Biobank*

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



## 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 significant risk factor for several age-related health conditions and mortality. There is also substantial heterogeneity in health outcomes among individuals of the same chronological age (Jylhävä et al., 2017). During the past decades, the research highlighted that the biological aging 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 significant 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 (Cole et al., 2019; Tian et al., 2023).

Parallel advancements in neuroscience and computational science have enabled researchers to develop novel algorithms to determine the biological age of the brain. A biological marker of brain age will enable us to adjust neuroimaging studies for the person's biological age instead of chronological age, minimising anatomical and functional heterogeneities present in groups of healthy individuals. Another advantage is that this could lead to a deeper understanding of pathological ageing mechanisms, which can culminate in dementia. Dementia is a multifactorial syndrome in which decades of accumulating neuropathology precedes clinical manifestation (Jack et al., 2010). The loss of neurons and synapses during the preclinical and prodromal stages can lead to brain atrophy and, therefore, to "older-looking" brains (when biological brain age, i.e. predicted 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 the biological brain age, thus showing a "younger-looking" brain, which could reflect relatively preserved brain structures (e.g., brain maintenance and/or cognitive reserve) (Cole et al., 2019; Stern et al., 2020). With the unprecedented growth of the elderly population worldwide, the expected increase in dementia cases (WHO guidelines, 2019), and while waiting for effective pharmacological treatments against dementia, a biological marker of brain age could play a key role in dementia prevention (Brusini et al., 2022).

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). These prior studies used machine (Cole et al., 2017; Franke et al., 2010; Franke and Gaser, 2012; Hwang et al., 2021) and deep learning (Bintsi et al., 2020; Cole et al., 2017; Gupta et al., 2021; Jonsson et al., 2019; Kolbeinsson et al., 2019; P. Lam et al., 2020a; Liang et al., 2019; Niu et al., 2020) approaches, achieving good performance in terms of mean absolute error (MAE) – between 2 and 6 years. However, the model type and the input choice varied across these studies that used pre-processed magnetic resonance imaging (MRI) data (T1-weighted, T1w), going through normalisation, corrections, segmentation steps, or even image feature extraction. Such a chain of steps is challenging to implement in research and, in the long term, in clinical settings due to time- and resource-consuming constraints.

Typically, a model is trained on neuroimaging data of healthy individuals from one or multiple cohorts covering a large age span. Developing a model for predicting "biological brain age" depends on the choice of training data when defining healthy aging. The "ideal" dataset 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 (Sajedi and Pardakhti, 2019); (3) participants with a diverse demographical background and a large, preferably

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



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 generalisation 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 time point. Although initiatives to gather large-scale population-based datasets are ongoing (e.g., UK Biobank), to our knowledge, currently, no existing cohort possesses all the five characteristics listed above.

In this study, we aimed to develop and validate a CNN model, based on brain images, that uses only one pre-processing step (i.e., rigid registration of T1w MRIs to the Montreal Neurological Institute – MNI – template space) for brain age prediction. This minimal pre-processing feature has the advantage and strength of simplifying the model's implementation and increasing accessibility in research settings. When publicly available, the model can be quickly used for any T1w MRI scan without time- and resource-consuming pre-processing steps. To evaluate our model, we used a large dataset of cognitively healthy individuals from six cohorts to address the "ideal" dataset criteria. The CNN model was compared using hold-out and cross-validation approaches. To verify the model's generalizability, we tested the abovementioned approaches using two external datasets containing different scanners and demographic characteristics from the training set. Furthermore, we included the two cohorts used as external datasets in the cross-validation loop to verify the model's usability with different cohorts. Finally, we employed the cross-validated model to predict brain age in skull-stripped images to ensure our model accurately predicted based on the brain image signal. We then evaluated the model's performance using two external datasets.

## 2 METHODS

### 2.1 STUDY POPULATION

For this study, we included 17296 T1w MRIs from 15289 (1176 are 1.5T and 16120 are 3T) cognitively healthy participants from six cohorts: the Alzheimer's Disease Neuroimaging Initiative (ADNI), the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL), the AddNeuroMed, the Group of Neuropsychological Studies from the Canary Islands (GENIC, from *Grupo de Estudios Neuropsicologicos de las Islas Canarias*), the Japanese ADNI, and the UK Biobank (Figure 1). The description of each cohort is in the Supplementary Material, Section A. A cognitively healthy status was defined based on the absence of dementia, mild cognitive impairment, and other neurological and psychiatric disorders. Further, individuals had to have a clinical dementia rating (CDR) score equal to zero, or mini-mental state examination (MMSE) score $\geq 24$, or self-reported good health (this last when available), or ICD-9 or 10 (details on the used ICD codes can is in Supplementary Material, Section B), depending on the available data in each cohort.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1. Flowchart describing the data included in this study, which were available in our database system at the time of the study, for the different cohorts. Where age ± SD is the mean age and standard deviation, age ± SD and age range are in years.

## 2.2 Convolutional Neural Networks

The implemented CNN uses 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 kernels. Each convolutional operation is followed by batch normalisation and a Rectified Linear Unit (ReLU) activation. The network was trained for 20 epochs with stochastic gradient descent and an initial learning rate of $\lambda$=0.002 that decreases with a factor of 10 every five epochs. We used five independent models during the CNN development and the trained networks as an ensemble model. Data augmentation included random scaling, cropping offsets, rotations (-5 to 5 degrees), affine, and gamma transformations (ranging from 0.5 to 2). Each image was cropped to a dimension of 80 x 96 x 80 voxels, with 2 x 2 x 2 mm³ 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].

We streamline the image pre-processing to improve the model's accessibility and processing speed. The sole pre-processing step involved a rigid registration (with six degrees of freedom) to the MNI template space using FSL FLIRT 6.0 (FMRIB's Linear Image Registration Tool). A rigid registration is more straightforward and quicker than an elastic registration. Omitting this step resulted in worse performance on the development set, despite the use of heavy data augmentation (data not shown). Figure 2 shows a schematic representation of the CNN model.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



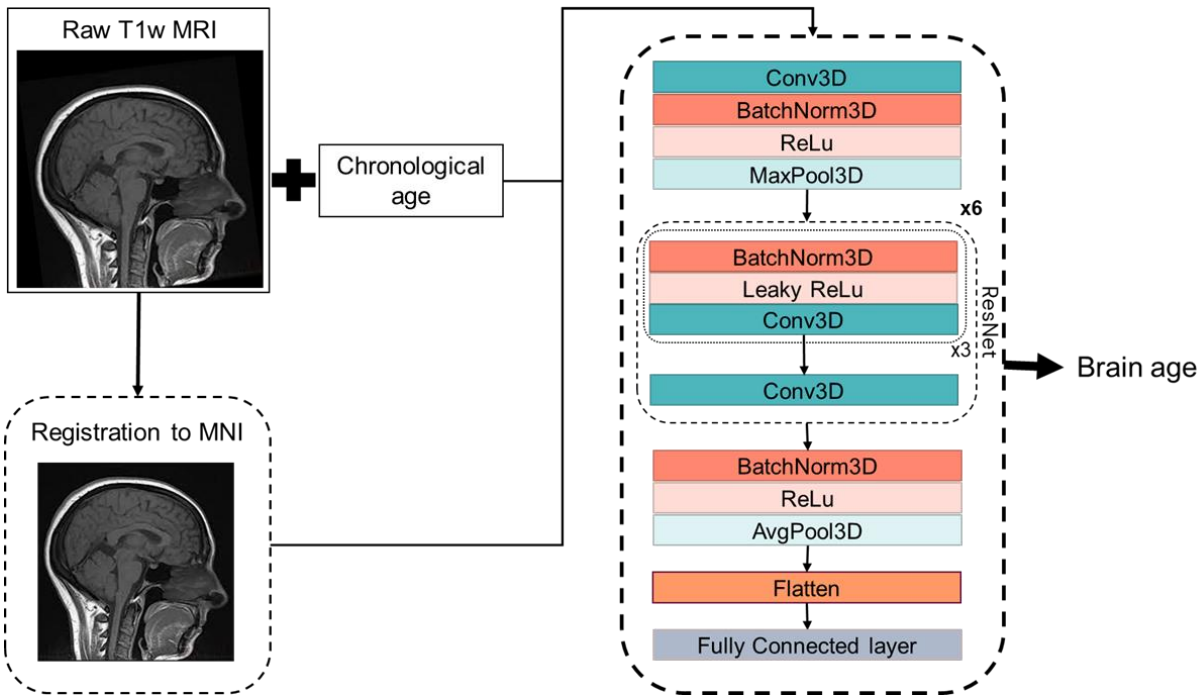


Figure 2. Simplified scheme of our 3D CNN. Where the input is a T1w brain MRI registered to the MNI space, the information on the subject's chronological age, and the output is the predicted brain age of the subject. Conv3D = 3D convolution: BatchNorm3d: 3D batch normalisation; (Leaky)ReLu: (leaky) Rectified Linear Unit; MaxPool3D: 3D max pooling; AvgPool3D: 3D average pooling; FC: fully connected layer; ResNet: residual network block.

Four separate models were developed in this project: one model based on a hold-out approach of CNN (CNN1) and three models based on a cross-validation approach (CNN2-4). Figure 3 displays the evaluation process's scheme of our CNN model. Each one of the turquoise rectangles represents 1/10 parts of the primary dataset, composed of 16734 raw MRIs from ADNI, AIBL, GENIC and UKB cohorts. Light blue rectangles indicate the 149 MRIs from AddNeuroMed cohort, whereas lilac rectangles indicate the 413 MRIs from J-ADNI. The CNN1 model is based on a hold-out approach with the training (80%), development (10%) and test (10%) datasets indicated by the arrows. CNN2 and CNN4 models incorporated all data from ADNI, AIBL, GENIC and UKB cohorts in their cross-validation loop, while AddNeuromed and J-ADNI were used for external validation. The CNN3 model was similar to CNN2 and 4, except that AddNeuroMed and J-ADNI were also incorporated in the cross-validation loop – thus, no external datasets were used for model evaluation.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



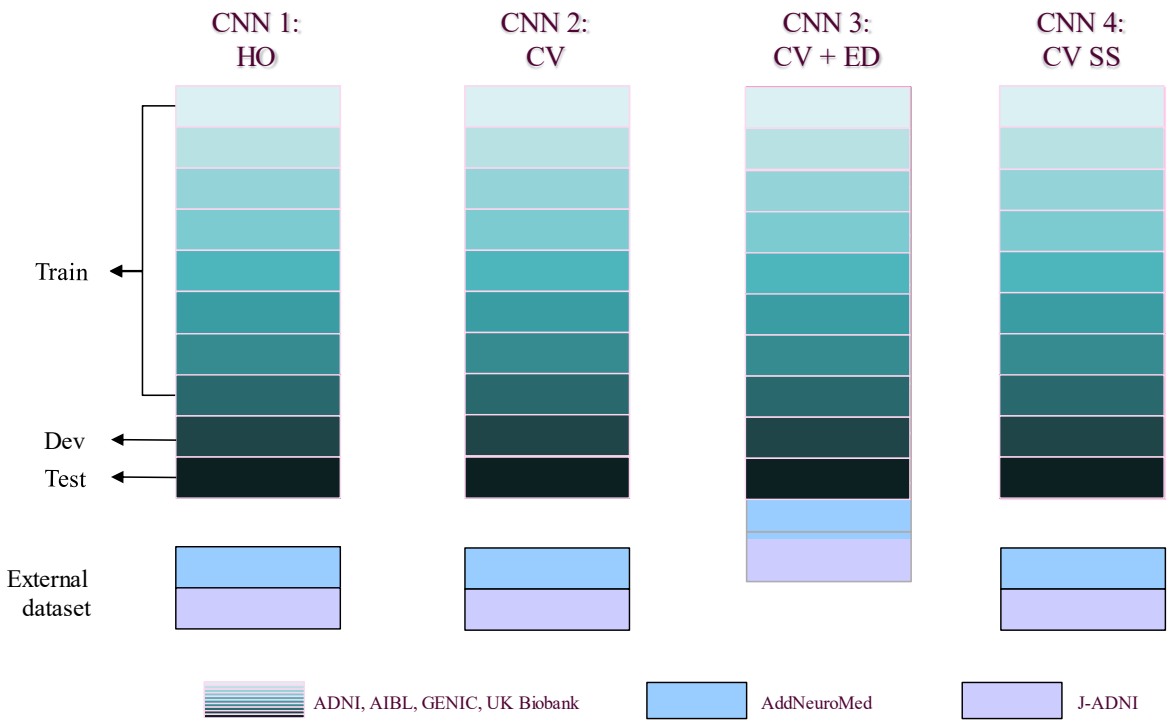


Figure 3. Scheme of evaluation of our CNN model in this study. We primarily used ADNI, AIBL, GENIC, and UK Biobank to develop our model (turquoise scale). CNN1 works in a hold-out approach (data split: 80% train, 10% development, 10% test set, each set of the data indicated by arrows). CNN2 was trained as a 10-folds cross-validation model, using the data of the four primary cohorts (turquoise scale) in the training loop. To evaluate the performance of these two models, we used AddNeuroMed (light blue) and J-ADNI (lilac) as external test sets. In CNN3, we added the external test sets in the 10-fold cross-validation. For comparison reasons, we evaluated our CNN2 scheme with skull stripped T1w MRIs (CNN4). *HO = hold-out; CV = cross-validation; ED = external test set; SS = skull-stripped images.*

### 2.2.1 Hold-out approach

Our CNN-based model was first trained in a hold-out fashion (CNN1). To not violate the test set and ensure comparability between the models, the primary dataset, composed of 16734 MRIs from ADNI, AIBL, GENIC and UKB cohorts, was randomly split into a training ($N_{img}$ = 15052, $N_{subj}$ = 13612, subsequently split into internal validation set for the development of each model) and a hold-out test ($N_{img}$ = 1682, $N_{subj}$ = 1503) set. If subjects had undergone multiple scans, all their images were assigned to the same set. The test set was evaluated after a satisfactory performance on the internal validation set to reduce the risk of model overfitting. The data distribution in train, development, and test sets for each cohort can be found in Supplementary Material, Section C. After training the CNN1 model, we applied this model in the AddNeuroMed and J-ADNI cohorts, to assess the model performance and generalizability in external datasets.

### 2.2.2 Cross-validated approach

To allow comparability to the hold-out model (CNN1), we used the same 16734 MRIs from ADNI, AIBL, Genic, and UK Biobank in the cross-validation approach training loop (CNN2). Stratification by cohort was applied in splitting the 10-fold for training and testing. After a 10-fold cross-validation, the trained model was evaluated in AddNeuroMed and J-ADNI (external cohorts). Furthermore, to ensure that our model's prediction was based on the brain and not on other features (e.g., head shape), we trained a 10-fold cross-validation model using skull-stripped brain images (CNN4). For CNN4 model images input, Freesurfer 6.0.0 was used to perform skull-stripping, applying the algorithm recon-all and selecting the image generated before brain parcellation (brain.finalsurfs.mgz). Images were motion and bias-corrected, transformed to Tailarach space, intensity normalised, and skull stripped. To reduce the size of the final processed image and for comparative reasons, all skull-stripped images were rigidly registered to the MNI space. Similarly to CNN2, also

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



the skull-stripped CNN4 model was externally validated in AddNeuroMed and J-ADNI. Finally, to ensure the cross-validated model's usability when including more diverse data, we trained a last model, CNN3, that included all cohorts (ADNI, AIBL, Genic, and UK Biobank plus AddNeuroMed and J-ADNI) to the ensemble of images within the 10-fold cross-validation.

## 2.3 ANALYSES

### 2.3.1 Model performance

Model performance was assessed using the MAE, defined as:

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

Where $y_{c,i}$ is the chronological age of participant $i$ and $\hat{y}_i$ the predicted age. MAE's values close to zero, indicates model's good performance, with predicted brain age being similar or almost equal to chronological age but considering aging heterogeneity, is unknown the lower limits for a model's MAE. Adjusting for age-dependent predicted brain age differences from the chronological age (brain age difference/gap - BAG) is problematic and can artificially inflate model performance (Butler et al., 2021). This can be illustrated with a model that, regardless of input data, only outputs a single predicted age of, e.g., 70. This will yield a suboptimal MAE, but when "correcting" for age, the MAE will be 0 between predicted and chronological age.

### 2.3.2 Relevant regions for brain age prediction

To explain the model's brain age prediction, we generated 3D gradient-based saliency maps of each subject using the SmoothGrad (Smilkov et al., 2017) algorithm. Salience maps visualizes the important voxels in individual predictions based on the computation of the gradient of the prediction with respect to the smoothed image. For gradient computation, we used the image with 15% noise added. The 3D gradient maps were averaged through the whole image sample, and only 1% of the higher salient values are shown to verify the most critical regions (Levakov et al., 2020; Mouches et al., 2022). For individual extrapolation, we plotted the 1% normalised higher values of the salience maps and overlayed them onto an arbitrary T1w brain MRI. The salience maps are presented according to their brain age difference in the CNN1 model, calculated from the chronological age, from -8 to +8 yrs. of difference from chronological age.

### 2.3.3 Differences in cortical thickness across age groups based on chronological and predicted brain age

To analyse the influence of correcting an individual's brain age in neuroimaging studies, we visualise how brain 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 full width at half maximum of 10 mm, used sex as a covariate and adjusted for false discovery rate at a threshold of 0.05. 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 chronological or predicted brain age) individuals in the general linear model. Since these groups are of different sizes—and *p*-values are influenced by group size—we present figures overlayed with z-scores.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



# 3 RESULTS

## 3.1 MODEL PERFORMANCE

Scatterplots of the brain age predictions on the CNN1 (hold-out) approach for the training and testing set of healthy individuals are shown in Figure 4. The results show that there is a strong correlation between chronological and predicted brain age in training and test sets.

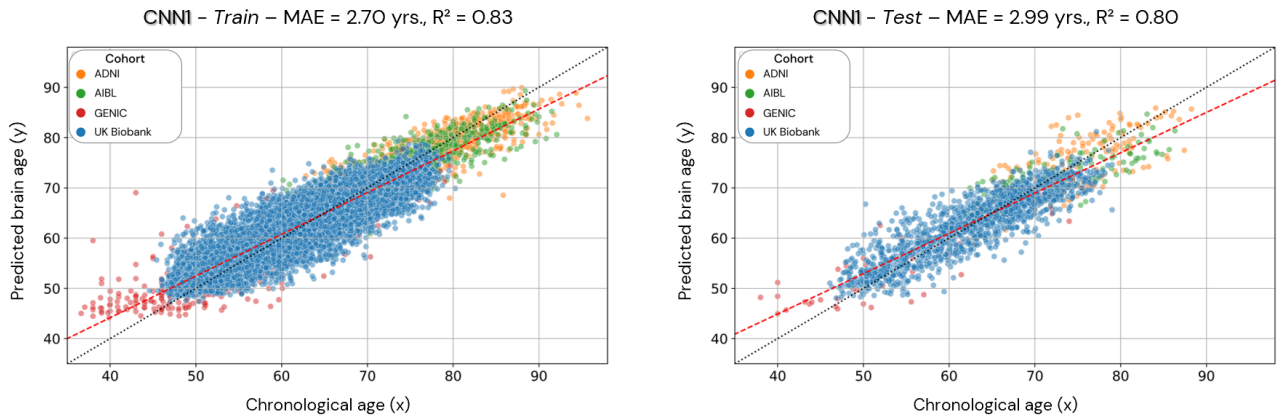


Figure 4. Scatterplots of the predicted brain age in the CNN1 model (hold-out approach) in the training and test datasets. Each coloured dot represents an individual, and each colour is a different cohort. The red line is a linear regression based on the predicted brain age.

Scatterplots of the brain age predictions on the CV approach for the CNN2 (CNN1 cohorts), CNN3 (CNN1 cohorts + J-ADNI and AddNeuroMed), and CNN4 (skull-stripped images from CNN1 cohorts) are shown in Figure 5. For CNN2 and CNN3 models, results show a strong correlation between chronological and predicted brain age with a MAE within the lower range of previously published models. However, CNN4 present a less strong correlation and higher MAE within the presented models in this study.

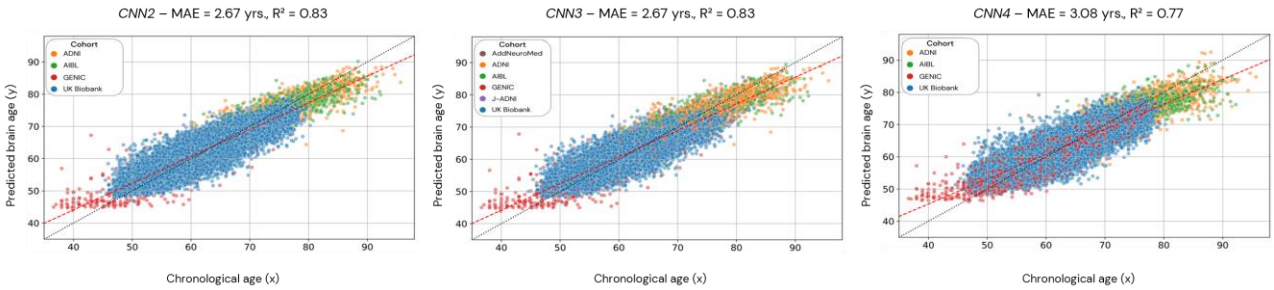


Figure 5. Scatterplots of the predicted brain age in the CNN2, 3 and 4 (cross-validation approach). Each dot represents an individual, and the colour code used for each cohort is presented in the legend. CNN2 was run using 4 cohorts, CNN3 with 6 cohorts, and CNN4 with 4 cohorts but with skull-stripped images. The red line is the linear regression based on the predicted brain age.

Figure 6 (A) illustrates CNN1 (on the left) and CNN2 (on the right) correlations between the predicted brain age (y-axis) by each model in the data from the test set of the CNN1 model, with the chronological age (x-axis). Figure 6 (B) illustrates the correlation between the predicted brain age by CNN1 (x-axis) and CNN2 (y-axis) using the training (on the left) and test (on the right) dataset splits of CNN1. Figure 6 (C) shows the correlations between the brain age difference (BAG) estimated by the two models (CNN1 on the x-axis and CNN2 on the y-axis) in the training and test datasets splits of CNN1.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



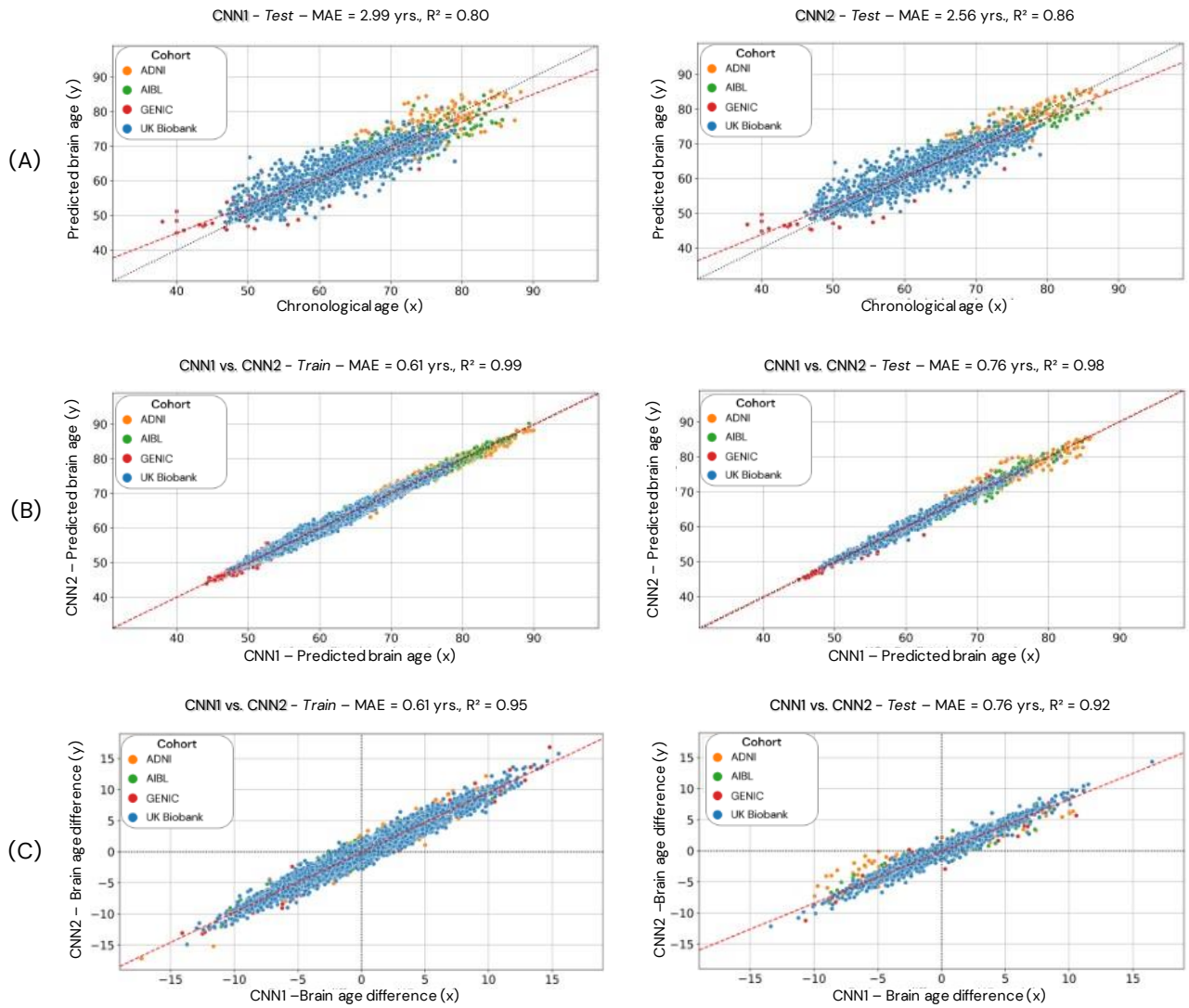


Figure 6. Comparison of the CNN1 (hold-out approach) and CNN2 (cross-validation approach) models. Data used for the comparison is based only on the individuals used in the test dataset of CNN1. In A, we present the test dataset's predicted brain age (y-axis) for both models correlated to chronological age (x-axis). In B, we present the correlation between the predicted brain age estimated by the two models (CNN1 on the x-axis and CNN2 on the y-axis) in the training (left) and test (right) datasets splits of CNN1. In C, we present the correlation between the brain age difference estimated by the two models (CNN1 on the x-axis and CNN2 on the y-axis) in the training and test datasets.

We also evaluate the performance of our model in the external datasets AddNeuroMed and J-ADNI (CNN1, 2 and 4). The age prediction distribution is shown in Figure 7, and shows the variability in age prediction for each one of the trained models when applied to unseen data (external dataset)Figure 7.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



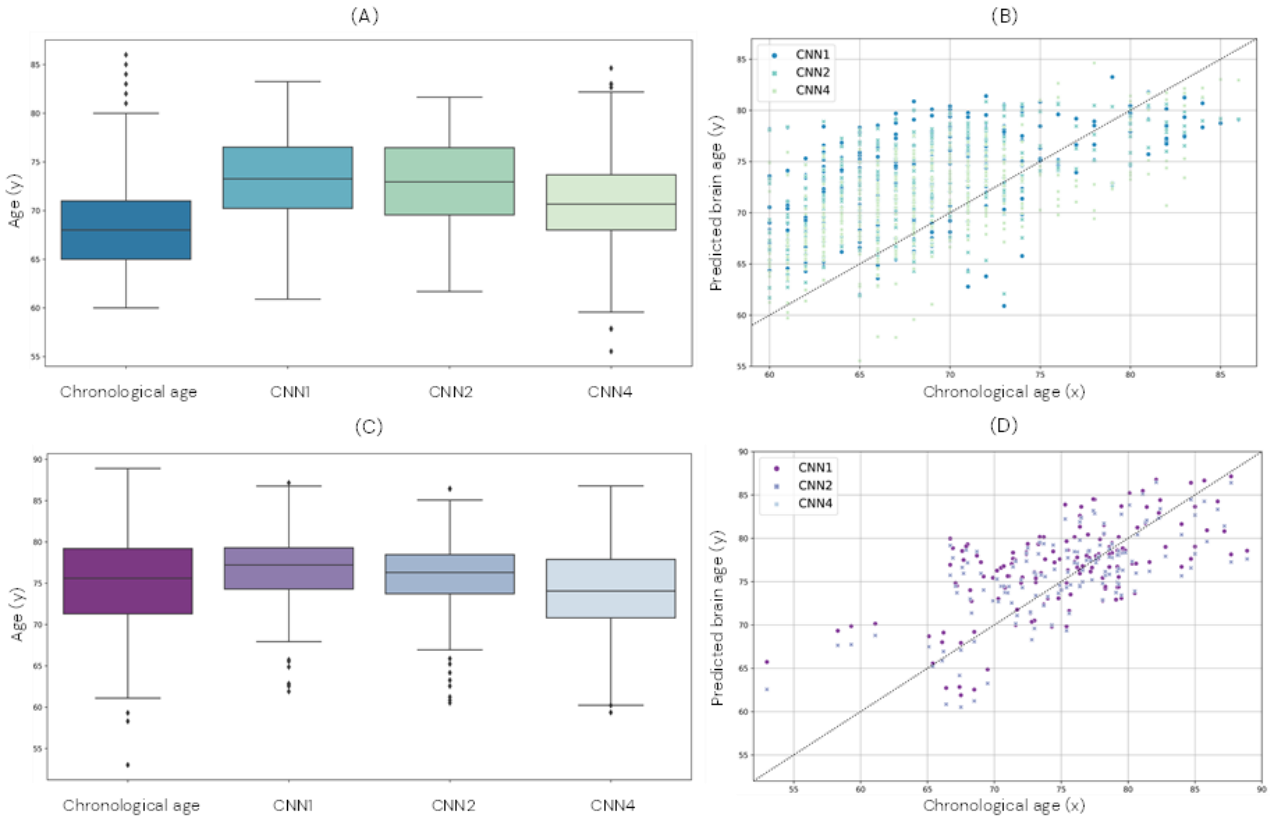


Figure 7. Comparison of brain age prediction distribution for chronological and predicted brain age in the CNN1, 2 and 4 for the AddNeuroMed (A and B) and J-ADNI (C and D) cohort. In A and C, we present the boxplot of the predicted brain age for each model compared to the chronological age. In B and D, the predicted brain age distribution is compared to the chronological age.

The calculated MAE of each cohort for all models is presented in Table 1. Normalized MAE and coefficient of determination for each cohort of the trained models can be found in Supplementary Material, Section D, Table D-1 and D-2.

Table 1. Calculated MAE for each cohort for each of the trained models.

| | **MAE per cohort (years)** | | | | | |
|---|---|---|---|---|---|---|
| **Model** | ADNI | AIBL | GENIC | UK Biobank | AddNeuroMed | J-ADNI |
| CNN1[1] | 2.56 | 2.85 | 4.20 | 2.75 | *4.03* | *5.63* |
| CNN2[2] | 2.40 | 2.69 | 3.96 | 2.66 | *3.67* | *5.30* |
| CNN3[2] | 2.41 | 2.69 | 3.98 | 2.67 | 3.02 | 2.25 |
| CNN4[1] | 2.98 | 3.33 | 4.42 | 3.03 | *3.64* | *4.14* |

The numbers in italic (AddNeuroMed and J-ADNI cohorts, for models CNN1, 2 and 4) were calculated as external datasets to the model. [1]Statistical differences were founded (p-value < 0.05) comparing the CNNs absolute error; [2]No statistical differences were founded in the comparison between absolute error of CNN2 and CNN3 (p-value = 0.37). Statistical differences between the models were calculated using Friedmans test with Conover's *post hoc* test, adjusting for multiple comparisons with the Bonferroni method.

To understand how our model performed compared to existing models, we further assessed our brain age models' performance only within the UK biobank. Then we compared the achieved MAEs with previous studies that evaluated their models in the UK Biobank cohort (Table 2). Our CNN models achieved MAEs ranging between 2.66 and 3.03 yrs. These are very similar to the MAEs achieved by CNN models developed in previous studies (ranging between 2.13 and 4.36). We also present the coefficient of determination between predicted brain age and the identity line for all the available studies.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



Table 2. Comparison of our models MAE with the literature.

| Study | Model | Modality | Pre-processing | MAE | $R^2$ |
|---|---|---|---|---|---|
| Dartora et al. | CNN1 | T1 | Rigid reg. | 2.75 | 0.88 |
| Dartora et al. | CNN2 | T1 | Rigid reg. | 2.66 | 0.89 |
| Dartora et al. | CNN3 | T1 | Rigid reg. | 2.67 | 0.89 |
| Dartora et al. | CNN4 | T1 | Bias-field and motion cor., Skull-strip., Rigid MNI reg. | 3.03 | 0.85 |
| Bintsi et al. (2020) | CNN (3D ResNet) | T1 | Skull-strip., non-linear MNI reg. | 2.64 | 0.77 |
| Bintsi et al. (2020) | CNN (patches) | T1 | Skull-strip., non-linear MNI reg. | 2.13 | 0.85 |
| 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 | 0.87 |
| P. Lam et al. (2020) | CNN | T1 | Bias-field cor., Skull-strip., Rigid MNI reg. | 4.36 | 0.66 |
| 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 | 0.87 |
| Dinsdale et al. (2021) | CNN (Male population) | T1 | UK Biobank pipeline w/ linear registration to MNI | 3.09 | 0.86 |
| Jonsson et al. (2019) | CNN (transfer learning) | T1 | Bias-field cor., Skull-strip., Dartel MNI reg., tissue maps | 3.63 | 0.61 |
| Peng et al. (2021) | SFCN | T1 | Bias-field cor., Skull-strip., Rigid MNI reg. | 2.14 | 0.39 |
| Peng et al. (2021) | CNN | T1 | Bias-field cor., Skull-strip., Rigid MNI reg. | 2.38 | - |

CNN: convolutional neural network; CNN1: our CNN model in a hold-out approach; CNN2: our CNN model in a cross-validation approach; CNN3: our CNN model in a cross-validation approach and including two cohorts in the dataset; CNN4: our CNN model in a cross-validation approach, using skull-stripped images; PCA: principal component analysis; R-CNN: recurrent CNN; SFCN: simple fully convolutional network; $R^2$=coefficient of determination. Dartoral et. al. refers to the current work.

To understand the noise levels from our models and their ability to capture subtle changes as a result of the ageing process, we plotted longitudinal trajectories for the participants with more than one-time point (Figure 8).

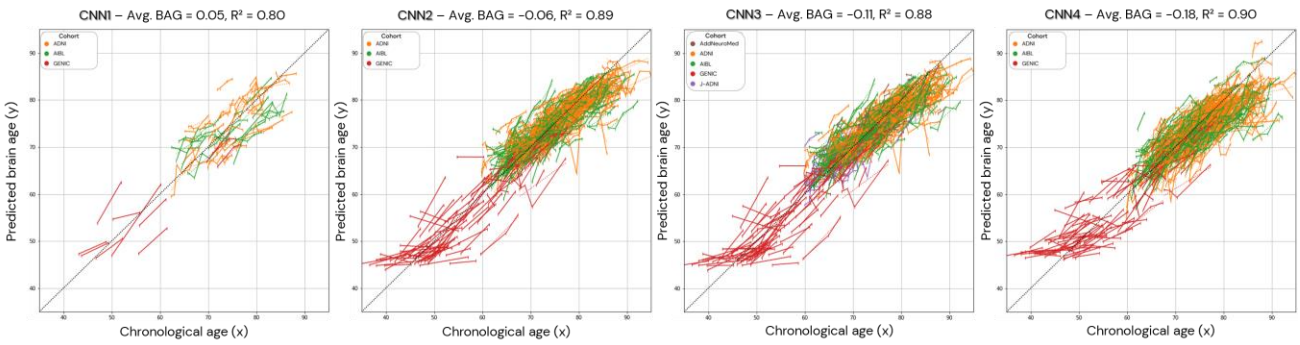


Figure 8. Brain age prediction in longitudinal trajectories for all the models. The average brain age gap/difference (Avg. BAG) in the presented population was calculated and shows a trend towards zero to the brain age difference between predicted and chronological age. In CNN1, only the individuals with longitudinal data in the test set are presented.

### 3.2 RELEVANT REGIONS FOR BRAIN AGE PREDICTION

For the explicability of our model, the salience map of each individual prediction was generated. The averaged overlayed salience maps for each CNN model are presented in Figure 9. Complementary view of salience maps slices is presented in Supplementary Material Section E, Figure E-1.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



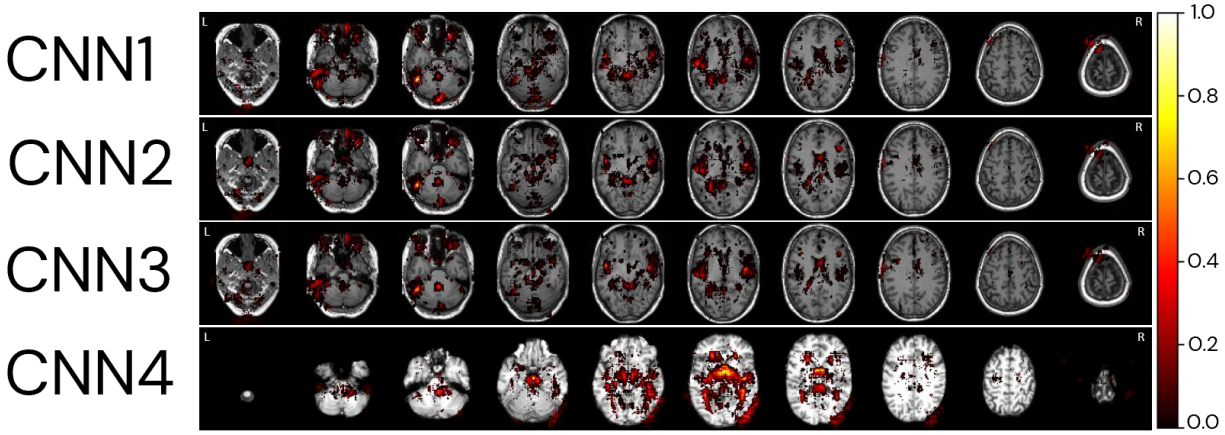


Figure 9. Relevant regions for age prediction. The absolute values of the salience maps for each model were averaged through the whole MRI sample and normalised between 0 and 1 for better visualisation. The SmoothGrad method was used to generate all salience maps. Overlayed salience map colours are normalised for each individual between 0 and 1.

To identify regions of importance for predicting the biological brain age, 18 individuals were randomly selected according to the following criteria: being in the test dataset of the CNN1, being between 64 and 66 yrs., having an age range difference between chronological and predicted brain age in the CNN1 model between -8 and +8 yrs. (Figure 10). The brain age gap for each individual, predicted in each model, is presented in Supplementary Material, Section F, Table F-1.

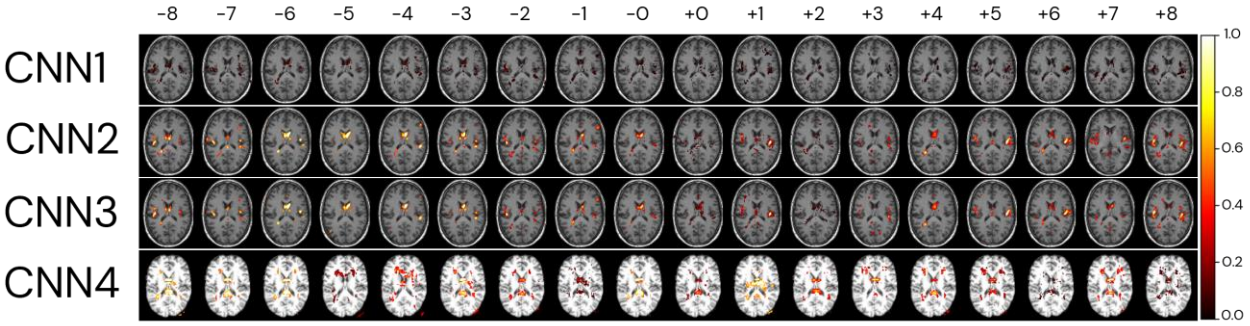


Figure 10. Relevant regions for the brain age prediction based on 1% of the salience maps of individuals with an average age of 65 yrs. overlayed to a random T1w MRI sample for each of the four CNN models. The individuals were randomly selected in the test dataset used in the CNN1, according to their brain age difference and proximity to the average age of our sample (65 yrs.). Each row represents the salience maps of one CNN model, where the columns show the same individual's salience map in each of the four models. The SmoothGrad method was used to generate all salience maps. The brain age difference is shown in the top row of the image, going from -8 to +8 yrs. of difference. Overlaid salience map colours are normalised for each individual between 0 and 1.

### 3.3 Differences in cortical thickness across age groups based on chronological and predicted brain age

Surface group analysis for cortical thickness was run in all available images with QDEC in Freesurfer 6.0.0. Individuals with 60, 65, 70, 75, and 80 (±1 years) chronological and biological (predicted) age were grouped and compared with a group of 55±1 years-old individuals. The analysis used sex as a covariate and was adjusted for a false discovery rate with a threshold of 0.05. Figure 11 shows the age-related differences in cortical thickness for chronological and biological (predicted) age groups by the CNN2 model. The number of individuals in each age group is presented in Supplementary Material, section G, Table G-1.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



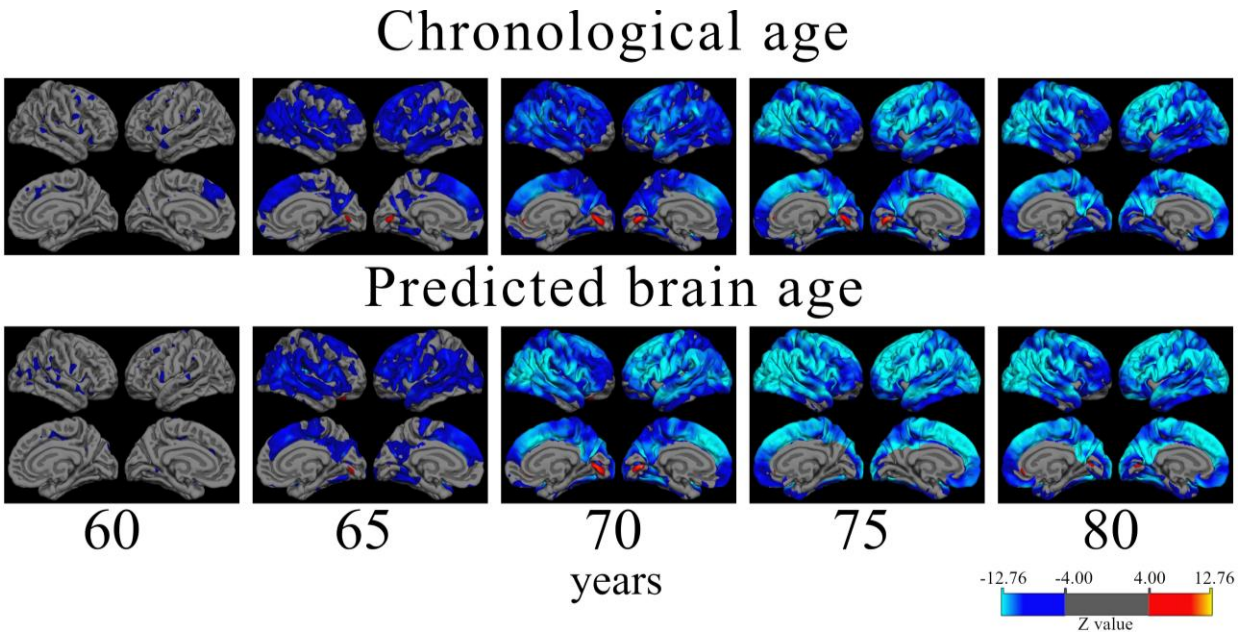


Figure 11. Atrophy patterns in aging according to chronological and predicted brain age in the cross-validation approach (CNN2). Analysis shows cortical thickness z-stat maps corrected for false discovery rate with a threshold of 0.05 of cross-sectional differences between a reference group of 55 (±1.0) years old and older subjects. Individuals were grouped by age ±1.0 years. The QDEC analyses were done with a 10 mm smoothing kernel and sex added as a covariate. Blue colours represent lower cortical thickness (increased atrophy) in the age group compared to the 55-year-old group, and red colours represent an increased cortical thickness compared to the younger group. It is possible to verify that the atrophy patterns start in parietal-frontal regions from the 65-years old group. The approach also captures the increased thickness (in red) around the lingual gyrus.

## 4 Discussion

In this study, we developed a CNN-based model using only one pre-processing step (rigid registration to MNI space) to predict the person's biological brain age from T1w images, to be easily implemented and used. Further, the study was performed as an effort to attend to the highlighted points related to an "ideal" image database (which includes diverse image data) to develop brain age prediction models. Finally, our model evaluation was performed using different approaches – hold-out (CNN1) and cross-validation (CNN2) – and generalizability was tested in external datasets. The usability (CNN3) of the model was also assessed by adding more data (from two external datasets) in the training loop, and to ensure that the previously trained models were using the brain signal for age prediction, we also trained the model using skull-stripped images (CNN4).

A key strength of this work is the use of minimally processed images as input for the CNN model, which makes it feasible for implementation in research and, in the long term, clinical settings. Our model requires only the registration to MNI space, which typically takes a few seconds and can be easily performed using an open brain image processing software such as FSL or FreeSurfer. Additional pre-processing steps would increase the likelihood of image exclusion during quality control and pre-processing. This would limit the model's performance as well as the possibility of using cohorts with a small sample size. CNN models can learn from the image data, including structure shape, which may not be captured by summary metrics such as volume or segmented tissue maps, without requiring pre-segmented data (Liang et al., 2019; Niu et al., 2020). Our focus on a one-step pre-processing, using a rigid registration to the MNI space template previously implemented by Cole et al. (2017), is to ensure the accessibility to our model in future research.

This study includes cohorts to attend to the highlighted points related to, what we believe to be, 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 scanners, which increases the model usability due to its

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



generalisation to new unseen data (Mårtensson et al., 2020). We also used all the longitudinal data available and showed that our model presents a reasonable age prediction congruent with the individual timeline**Error! Reference source not found.**. 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 outcome (brain age) and what leads to that outcome (risk factors, e.g., chronic cardiometabolic disease and risk factors, affective/mood disorders, etc). Further, to validate the generalizability of the models, we used external datasets of cognitively unimpaired individuals from two cohorts: AddNeuroMed, a cross-European study designed to find biomarkers for Alzheimer's disease, and J-ADNI, the Japanese version of the ADNI dataset. We chose to use these cohorts because of differences in age distribution, e.g., AddNeuroMed average age is around ten years older than the average of our total sample, different ethnicity, e.g., our sample is composed mainly of European and North-American individuals, and both cohorts mostly use images acquired in 1.5T scanners, whereas the training dataset was based mainly on 3T MRI, i.e., 96.2 % of all images used for training and testing in the CNN1 approach.

Due to the complexity and time-consuming nature of the training networks with large amounts of data, the hold-out method used in CNN1 is the most common approach in the literature (Cole et al., 2017; Dinsdale et al., 2021b; Gupta et al., 2021; Jonsson et al., 2019; Kolbeinsson et al., 2021; Lam et al., 2020; Peng et al., 2021; Ren et al., 2022). Our CNN1 model achieved an MAE of 2.99 yrs. in the test set, which agrees with the MAE reported in the literature for hold-out test sets, using 1 to 3 cohorts (MAE of 2.14 – 4.65 yrs.) (Cole et al., 2017; Dinsdale et al., 2021b; Gupta et al., 2021; Jonsson et al., 2019; Kolbeinsson et al., 2021; Lam et al., 2020; Peng et al., 2021; Ren et al., 2022). Our calculated MAE is also in the range of the available MAEs of the CNN models (Bintsi et al., 2020; Dinsdale et al., 2021b; Gupta et al., 2021; Jonsson et al., 2019; Kolbeinsson et al., 2021; Lam et al., 2020; Peng et al., 2021) available in the literature that used hold-out approaches and only data from the UK Biobank but performing several pre-processing steps as opposed to our CNN model. The current debate surrounding the development of brain age models lacks a thorough evaluation of CNN-based brain age models across different external datasets. In such scenario, it is difficult and relatively unfair to compare the performance of different models not using the same evaluation dataset (Sajedi and Pardakhti, 2019). To overcome such a challenge, we compared the performance of our CNN in its different approaches related to the two used external datasets, AddNeuroMed and J-ADNI, showing that CNN1's MAE performance is still within a range of 1.5 years when compared to the other CNNs.

Deep learning models, like CNNs, tested in out-of-distribution data, results in performance dropping, increasing underestimation in new unseen data, e.g., clinical data. However, including more diverse and variable data to the model's training, increases models' robustness/reliability (Mårtensson et al., 2020). One way of including variability in the model is running it in a cross-validation fashion. We have done this in CNN2, where the same data used for training, validating, and testing CNN1 was used for training a 10-fold cross-validation model (CNN2). In general, the CNN2 model had a better performance than CNN1. The correlation with chronological age was higher, the MAE was smaller in CNN2 in all cases of age prediction (training and test sets), and calculated brain age difference (BAG) when comparing both models. Comparing the performance in the different cohorts used in the development of this work, we see a tendency of individual smaller MAEs in CNN2. To confirm the robustness of the model, CNN2 was also evaluated in external datasets, confirming the decrease in underestimation, proved by the increase of the model's performance in AddNeuroMed and J-ADNI cohorts when compared to CNN1. Even though CNN1 and CNN2 have similar performance in age prediction, the increased variability in the model's training reduced the average and calculated MAE for each cohort. This is further proved by CNN3, where CNN2 and CNN3 performed similarly ($p>0.05$). Both models presented the same MAE of 2.67

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



years and a coefficient of determination of 0.83 but different performances in out-of-distribution data. Specifically, evaluating the predictions in the J-ADNI cohort in CNN3, it is possible to verify that adding the cohort to the training set decreased the predicted MAE for this cohort, going from an MAE of 5.63 in CNN1 to 2.25 in CNN3. This reinforces our hypothesis that the small number of data does not increase the absolute error across all the cohorts (AddNeuroMed and J-ADNI represent ~3% of the total amount of images used in CNN3) but increases the performance in a specific cohort. This increases the model's usability, as data with more variability will be included in the training set, ameliorating the age prediction in external cohorts (Mårtensson et al., 2020). Therefore, additional tests with a greater variety of cohorts are necessary to understand how wide the model's generalizability is and how it can change by adding more data to the training set.

To ensure that our model was using only brain signals from the T1w MRI and that predictions were not depending on the head morphology or bone tissue, we trained and evaluated CNN4. This model is comparable to CNN2, apart from that skull-stripped brain images were used as input. Regarding performance, CNN2 presented a smaller MAE and better correlation to the chronological age than CNN4. The same tendency is also present in evaluating data only from the UK Biobank. This shows that our CNN works better with minimally pre-processed brain MRIs when compared to heavily pre-processed images. However, for evaluating the external dataset, J-ADNI, CNN4 presented a smaller MAE than CNN2. We hypothesise that using heavily pre-processed images as in CNN4 could remove or decrease the effects of bias field and skull size/format and partially mitigate the inhomogeneity present in the external dataset.

The lower limit of the MAE score is unknown and depends on both the inter-subject variability and age distribution of both training and test datasets. By using stricter exclusion criteria for what is considered "cognitively healthy", the variability and theoretical MAE lower bound decrease. This makes comparisons between studies of models evaluated on different datasets challenging. Several studies have trained and evaluated their model on the UK Biobank cohort, which enables rough comparisons. However, this restricts the model to cross-sectional image data (at least for the first wave of the UK Biobank data) from a "homogeneous" population from the United Kingdom acquired in 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. Also, 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). Comparing our results with the literature applied only to UK Biobank images, we observed that using several time-consuming image pre-processing steps, none of the models achieved an MAE smaller than 2.13. The CNN4, which uses skull-stripped images, showed the worst performance within our different approaches using the same CNN architecture. For a more accurate comparison of the model's performance using the MAE metric, normalised MAE should be used. However, not all the selected papers for comparison presented the average age of the used subgroup of UK Biobank data, limiting the calculation of a normalised MAE. For future comparisons to our work, the normalised MAE for all four different approaches is presented in Supplementary Material, Section F.

Essentially, all our trained models showed MAE levels comparable to those reported in previous literature, which is typically in the range of 2.13 – 6 (Baecker et al., 2021b, 2021a; Bintsi et al., 2020; Cole, 2020; Dinsdale et al., 2021a; Gupta et al., 2021; Jonsson et al., 2019; Kolbeinsson et al., 2019; Lam et al., 2020; Sajedi and Pardakhti, 2019; Tanveer et al., 2022). This indicates that our models have a good performance with the advantage of requiring only one pre-processing step.

A "perfect" model for brain age prediction in cognitive unimpaired individuals should show smooth and non-declining trajectories within the same individual at different time points, assuming that a healthy person's brain age does not vary rapidly or decreases/increases substantially. Visually, our CNN model in different approaches seems to generate smooth

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



predictions that increase with chronological age, with some noisy predictions deviating from the trajectory. For quantitative analysis, we also calculate the mean age gap (BAG) of all individuals, the mean squared error and $R^2$. The models present mean brain age differences between chronological and predicted brain ages smaller than 0.5 yrs. However, the predictions also shows a potential confounder in subsequent analyses of brain age predictions: the noise. As it can be observed, for some subjects, fluctuations between the brain age differed some years between two scanning sessions that were one year apart. It seems unlikely that this is a biological phenomenon but instead attributed to the input data being noisy or low quality. Most datasets do not have abundant longitudinal data to sort "bad" predictions from "good" predictions. If the data and group sizes are extensive, some level of noise is acceptable and might not affect the interpretation of the results. The maximum average difference between chronological and brain age of all longitudinal plots was -0.11, with $R^2$'s higher than 0.88. However, this effect can have a non-negligible impact when analysing data sets with small group sizes or running longitudinal analyses with few follow-up scans. This is important to remember when conducting future studies related to brain age, for example, when investigating the association between brain age and neurological disorders with a low prevalence in the population.

The analysis of important brain regions for prediction was realised qualitatively by the plots of the salience maps. In general, the averaged salience map throughout the whole image sample shows to agree with the highlighted regions in the analysis of the brain age difference (BAG) for the same chronological age (65 years). The analysis of brain age differences (BAG) showed similar important regions for prediction, with variations in importance intensity between them. We hypothesise that the salience map shows negligible importance for these specific regions due to the fewer images used to train the CNN1. It is important to highlight that the brain age differences shown in the top row of Figure 10 are based on the CNN1 predictions for the plotted individuals. In general, the CNN4 model shows brain age differences with higher variability than compared with the other three models (CNN1-3). More studies are necessary to understand how this model presents more significant variation, using the same image data but with more pre-processing steps. The intensity of importance (higher importance showed in yellow colours) also shows to be more significant to predict age as younger (e.g., -6 yrs.) than chronological age, mainly close to the ventricles (Bintsi et al., 2021) and insular cortex (Lee et al., 2022). For prediction as older, greater importance is given to the right side of the insular cortex (Lee et al., 2022) as frontal-occipital regions. The salience maps show regions symmetrical and asymmetrical, mainly on the left side of the brain (Roe et al., 2021) and around the ventricles (Bintsi et al., 2021), as important for age prediction. Left brain asymmetries with aging are a typical pattern found in aging studies of cortical thickness (Frangou et al., 2022; Koelkebeck et al., 2014), cortical volume and surface area (Koelkebeck et al., 2014). In agreement with Lee et al. (2022), who plotted salience maps for different decades, regions with a higher contribution for age prediction were in the insular cortex (from 30 – 50 yrs. groups), ventricular boundary (50-60 yrs. group), and cerebellum (90-100 yrs. group).

Interestingly, regions around the eyes were selected in the three non-skull-stripped models (CNN1-3). We hypothesise that changes in the soft tissue and liquid surrounding the orbital space, as the bony orbit, which has differences in sex (i.e., men usually have greater skeleton size than women) (Erkoç et al., 2015), as well as general dimensions in orbital structures (Rana et al., 2022), could be being used in the model to predict age. For the model using skull-stripped brain images as input (CNN4), right regions close to the cerebellum and occipital lobe, outside the brain, were selected as important for prediction. We believe that the increased noise used for the SmoothGrad in a region that could have a higher neurodegeneration load could be leading to prediction importance outside the brain in this model. More studies are necessary to understand the highlight of this region for the prediction. However, we believe this could be an artefact generated by the SmoothGrad method due to the addition of noise in the image for the construction of the maps. Future

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



studies could use different methods to generate salience maps and use different methods to generate skull-stripped images to verify if this outstanding region (right outside the skull region) continues to be highlighted for prediction in this model.

One of the theoretical uses of biological age is for age correction in neuroimaging studies. The hypothesis behind it is that correcting neuroimaging studies for the biological (predicted) age of the individuals will better handle the heterogeneity that we see in aging, incorporating diversity in longitudinal brain trajectories due to lifestyle, environmental or even biological factors (Cole et al., 2019; Tian et al., 2023). For the analysis of chronological and biological (predicted) age atrophy pattern differences, we can observe differences mainly from the older groups (as of 70 yrs.). In the comparison of the group of 70 yrs. in the chronological and biological (predicted) brain age, atrophied areas of higher statistical significance are present in the mid-frontal to parietal-occipital regions of the group based on biological age. This agrees with the work of Thambisetty et al. (2010), where an anterior-posterior gradient in age-related brain atrophy was found, with frontal-parietal regions showing a greater decline. The groups with 75 and 80 yrs. have more similar atrophy patterns, with higher spread to the parietal lobe in the biological (predicted) brain age. Interestingly, a region of greater thickness in the oldest groups, compared to the reference group (55 yrs.), was found in the primary visual cortex, located in the calcarine sulcus. 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. These studies suggest that the visual cortex thickness is use-dependent instead of age-related (Burge et al., 2016; Griffis et al., 2016; Jorge et al., 2020). We hypothesise that this can be a cohort confounding effect, even individuals of 75 yrs. being present in all cohorts, but in larger amounts in ADNI and AIBL. Differences in atrophy patterns between individuals grouped by their chronological and biological (predicted) brain age need to be further studied. However, our results already show different tendencies in atrophy patterns between them. Correcting for biological (predicted) brain age in neuroimaging studies could be one step further in understanding heterogeneity present in aging and be used in early diagnosis of neurological diseases, prognosis and even monitoring of treatment response, being one step further to precision medicine.

Some limitations need to be acknowledged. The large dataset in the current study hindered the possibility of performing extensive quality control. For the CNN model, this would mean inspecting that the performed rigid registration was adequate. The random cases we inspected suggested that the overall quality of the segmentations was sufficient. However, tools to automate the quality control process – such as Brusini et al. (2020) and Klapwijk et al. (2019) – will be necessary for future studies on this data size. Regardless of the lack of extensive quality control of the images, our model showed robust findings with only slightly worse performance when compared to previously published works.

Studies also need to evaluate how the inclusion of different populations influences brain age prediction models based on minimally processed MRIs. However, the training model likely needs to include more diverse data, i.e., different types of cohorts, e.g., from Asia, Africa, and Latin America, as well as a broader range of scanners (from 1.5T to 7T) as well as imaging protocols (Mårtensson et al., 2020). Further, the differences in the distribution of chronological age and the number of subjects in each cohort can lead to overfitting 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 training set, heavy data augmentation, and only running 20 epochs in training helped minimise the possibility of overfitting.

Even though we defined the cognitively "healthy" status as consistently as possible across the cohorts, some variation exists but we acknowledge the clinical and cognitive assessments relied on similar procedures across cohorts. 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, and UK Biobank, GENIC, and AddNeuroMed: Europe). However, this is not necessarily a limitation but rather a strength. Indeed, the developed algorithm could be applied in future research, in which the biological age of the brain is a focus, to some extent

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



independent of the cohort characteristics, thus increasing the generalizability of the model. Future studies need to further test this hypothesis and the impact of different cultural backgrounds on the estimation of brain age.

## 5 CONCLUSION

In this study, we developed a CNN-based model to predict biological brain age using raw T1w MRI registered to the MNI space, with the goal of accessibility and simplicity in implementation. The model was systematically evaluated using different approaches, comprising several datasets of cognitively healthy individuals with different scans and population characteristics as well as using cross-sectional and longitudinal data. Our CNN-based model provides accurate results compared to state-of-art methods. The generalizability and usability of the model were tested using external datasets with different demographic characteristics, MRI protocols, and MRI scanners, proving the robustness of the model. In addition, we present the important regions for brain age prediction. We also provide indicators for the use of biological (predicted) brain age as a metric for age correction in neuroimaging studies as an alternative to the traditional chronological age based on the differences in cortical atrophy. Finally, the model's code and trained CNN weights are made publicly available for the research community to quickly implement and use in their research to study ageing and age-related brain disorders.

## 6 DATA AVAILABILITY

The complete code for using the CNN is available at https://github.com/westman-neuroimaging-group/brainage-prediction-mri. We include all the developed models and their trained weights for researchers to apply on their own neuroimaging data sets (with and without skull-stripped T1w MRI brain images) and scripts for training their model on other datasets. When new image data is available from different cohorts, the CNN will be re-trained, and trained weights and performance of the model will be updated in the page.

## 7 ACKNOWLEDGEMENTS

The authors wish to thank the Swedish Research Council (VR), the Strategic Research Programme in Neuroscience at Karolinska Institutet (StratNeuro), the Center for Innovative Medicine (CIMED), the Foundation for Geriatric Diseases at Karolinska Institutet, the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Brain Foundation, the Swedish Alzheimer Foundation, the Åke Wiberg Foundation, the Olle Engkvist Byggmästare Foundation, the joint research funds of KTH Royal Institute of Technology and Stockholm County Council (HMT), Swedish Parkinson Foundation, King Gustaf V:s and Queen Victorias Foundation, David and Astrid Hageléns Foundation, Loo and Hans Ostermans Foundation, Vinnova and Digital Futures for financial support, and the NIA-supported Collaboratory on Research Definitions for reserve and resilience in cognitive ageing and dementia. 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.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



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 organisation 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.

GENIC was funded by the Estrategia de Especialización Inteligente de Canarias RIS3 de la Consejería de Economía, Industria, Comercio y Conocimiento del Gobierno de Canarias, co-funded by the Programa Operativo FEDER Canarias 2014–2020 (ProID2020010063), and Universidad Fernando Pessoa Canarias. Data used in the preparation of this article was taken from the GENIC database (Group of Neuropsychological Studies of the Canary Islands, University of La Laguna, Spain. Principal investigators: José Barroso and Daniel Ferreira, contact: daniel.ferreira.padilla@ki.se). The following collaborators contributed to the GENIC-database but did not participate in the analysis or writing of this report (in alphabetic order by family name): Nira Cedrés, Rut Correia, Patricia Díaz, Eloy García, Lissett González, Aída Figueroa, Nerea Figueroa, Teodoro González, Zaira González, Cathaysa Hernández, Edith Hernández, Nira Jiménez, Judith López, Cándida Lozano, Alejandra Machado, María Mata, Yaiza Molina, Antonieta Nieto, Roraima Yánez-Pérez, María Sabucedo, Elena Sirumal, Marta Suárez, Manuel Urbano, and Pedro Velasco.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



## REFERENCES

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

Bintsi, K.-M., Baltatzis, V., Hammers, A., Rueckert, D., 2021. Voxel-Level Importance Maps for Interpretable Brain Age Estimation, in: Reyes, M., Henriques Abreu, P., Cardoso, J., Hajij, M., Zamzmi, G., Rahul, P., Thakur, L. (Eds.), Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data. Springer International Publishing, Cham, pp. 65–74.

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

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



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

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

Dinsdale, N.K., Bluemke, E., Smith, S.M., Arya, Z., Vidaurre, D., Jenkinson, M., Namburete, A.I.L., 2021b. Learning patterns of the ageing brain in MRI using deep convolutional networks. Neuroimage 224, 117401. https://doi.org/10.1016/j.neuroimage.2020.117401

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

Erkoç, M.F., Öztoprak, B., Gümüş, C., Okur, A., 2015. Exploration of orbital and orbital soft-tissue volume changes with gender and body parameters using magnetic resonance imaging. Exp Ther Med 9, 1991. https://doi.org/10.3892/ETM.2015.2313

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

Frangou, S., Modabbernia, A., Williams, S.C.R., Papachristou, E., Doucet, G.E., Agartz, I., Aghajani, M., Akudjedu, T.N., Albajes-Eizagirre, A., Alnæs, D., Alpert, K.I., Andersson, M., Andreasen, N.C., Andreassen, O.A., Asherson, P., Banaschewski, T., Bargallo, N., Baumeister, S., Baur-Streubel, R., Bertolino, A., Bonvino, A., Boomsma, D.I., Borgwardt, S., Bourque, J., Brandeis, D., Breier, A., Brodaty, H., Brouwer, R.M., Buitelaar, J.K., Busatto, G.F., Buckner, R.L., Calhoun, V., Canales-Rodríguez, E.J., Cannon, D.M., Caseras, X., Castellanos, F.X., Cervenka, S., Chaim-Avancini, T.M., Ching, C.R.K., Chubar, V., Clark, V.P., Conrod, P., Conzelmann, A., Crespo-Facorro, B., Crivello, F., Crone, E.A., Dale, A.M., Dannlowski, U., Davey, C., de Geus, E.J.C., de Haan, L., de Zubicaray, G.I., den Braber, A., Dickie, E.W., Di Giorgio, A., Doan, N.T., Dørum, E.S., Ehrlich, S., Erk, S., Espeseth, T., Fatouros-Bergman, H., Fisher, S.E., Fouche, J.P., Franke, B., Frodl, T., Fuentes-Claramonte, P., Glahn, D.C., Gotlib, I.H., Grabe, H.J., Grimm, O., Groenewold, N.A., Grotegerd, D., Gruber, O., Gruner, P., Gur, R.E., Gur, R.C., Hahn, T., Harrison, B.J., Hartman, C.A., Hatton, S.N., Heinz, A., Heslenfeld, D.J., Hibar, D.P., Hickie, I.B., Ho, B.C., Hoekstra, P.J., Hohmann, S., Holmes, A.J., Hoogman, M., Hosten, N., Howells, F.M., Hulshoff Pol, H.E., Huyser, C., Jahanshad, N., James, A., Jernigan, T.L., Jiang, J., Jönsson, E.G., Joska, J.A., Kahn, R., Kalnin, A., Kanai, R., Klein, M., Klyushnik, T.P., Koenders, L., Koops, S., Krämer, B., Kuntsi, J., Lagopoulos, J., Lázaro, L., Lebedeva, I., Lee, W.H., Lesch, K.P., Lochner, C., Machielsen, M.W.J., Maingault, S., Martin, N.G., Martínez-Zalacaín, I., Mataix-Cols, D., Mazoyer, B., McDonald, C., McDonald, B.C., McIntosh, A.M., McMahon, K.L., McPhilemy, G., Meinert, S., Menchón, J.M., Medland, S.E., Meyer-Lindenberg, A., Naaijen, J., Najt, P., Nakao, T., Nordvik, J.E., Nyberg, L., Oosterlaan, J., de la Foz, V.O.G., Paloyelis, Y., Pauli, P., Pergola, G., Pomarol-Clotet, E., Portella, M.J., Potkin, S.G., Radua, J., Reif, A., Rinker, D.A., Roffman, J.L., Rosa, P.G.P., Sacchet, M.D., Sachdev, P.S., Salvador, R., Sánchez-Juan, P., Sarró, S., Satterthwaite, T.D., Saykin, A.J., Serpa, M.H., Schmaal, L., Schnell, K.,

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



Schumann, G., Sim, K., Smoller, J.W., Sommer, I., Soriano-Mas, C., Stein, D.J., Strike, L.T., Swagerman, S.C., Tamnes, C.K., Temmingh, H.S., Thomopoulos, S.I., Tomyshev, A.S., Tordesillas-Gutiérrez, D., Trollor, J.N., Turner, J.A., Uhlmann, A., van den Heuvel, O.A., van den Meer, D., van der Wee, N.J.A., van Haren, N.E.M., van 't Ent, D., van Erp, T.G.M., Veer, I.M., Veltman, D.J., Voineskos, A., Völzke, H., Walter, H., Walton, E., Wang, L., Wang, Y., Wassink, T.H., Weber, B., Wen, W., West, J.D., Westlye, L.T., Whalley, H., Wierenga, L.M., Wittfeld, K., Wolf, D.H., Worker, A., Wright, M.J., Yang, K., Yoncheva, Y., Zanetti, M. V., Ziegler, G.C., Thompson, P.M., Dima, D., 2022. Cortical thickness across the lifespan: Data from 17,075 healthy individuals aged 3–90 years. Hum Brain Mapp 43, 431–451. https://doi.org/10.1002/HBM.25364

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

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

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., Abdulkadir, A., Erus, G., Habes, M., Pomponio, R., Shou, H., Doshi, J., Mamourian, E., Rashid, T., Bilgel, M., Fan, Y., Sotiras, A., Srinivasan, D., Morris, J.C., Marcus, D.S., Albert, M.S., Bryan, N., Resnick, S.M., Nasrallah, I.M., Davatzikos, C., Wolk, D.A., 2021. Disentangling Alzheimer's disease neurodegeneration from typical brain aging using MRI and machine learning. Alzheimer's & Dementia 17, e051532. https://doi.org/10.1002/alz.051532

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

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



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

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

Koelkebeck, K., Miyata, J., Kubota, M., Kohl, W., Son, S., Fukuyama, H., Sawamoto, N., Takahashi, H., Murai, T., 2014. The contribution of cortical thickness and surface area to gray matter asymmetries in the healthy human brain. Hum Brain Mapp 35, 6011–6022. https://doi.org/10.1002/HBM.22601

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

Kolbeinsson, A., Kossaifi, J., Panagakis, Y., Bulat, A., Kumar, A.A., Tzoulaki, I., Matthews, P.M., 2021. Tensor dropout for robust learning. IEEE Journal on Selected Topics in Signal Processing 15, 630–640. https://doi.org/10.1109/JSTSP.2021.3064182

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

Lam, P., Zhu, A.H., Gari, I.B., Jahanshad, N., Thompson, P.M., 2020b. 3D Grid-Attention Networks for Interpretable Age and Alzheimer's Disease Prediction from Structural MRI. https://doi.org/10.48550/arXiv.2011.09115

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

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: Proc. SPIE 11583, 16th International Symposium on Medical Information Processing and Analysis. p. 32. https://doi.org/10.1117/12.2579630

Lee, J., Burkett, B., Min, H.-K., Graff-radford, J., Schwarz, C., Petersen, R., 2022. Deep learning-based brain age prediction in normal aging and dementia. Nature Aging 2022 2:5 2, 412–424. https://doi.org/10.1038/s43587-022-00219-7

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



Levakov, G., Rosenthal, G., Shelef, I., Raviv, T.R., Avidan, G., 2020. From a deep learning model back to the brain—Identifying regional predictors and their relation to aging. Hum Brain Mapp 41, 3235–3252. https://doi.org/10.1002/hbm.25011

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

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

Mouches, P., Wilms, M., Rajashekar, D., Langner, S., Forkert, N.D., 2022. Multimodal biological brain age prediction using magnetic resonance imaging and angiography with the identification of predictive regions. Hum Brain Mapp 43, 2554–2566. https://doi.org/10.1002/HBM.25805

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

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

Rana, K., Juniat, V., Rayan, A., Patel, S., Selva, D., 2022. Normative measurements of orbital structures by magnetic resonance imaging. Int Ophthalmol 42, 3869–3875. https://doi.org/10.1007/S10792-022-02407-1/TABLES/7

Ren, B., Wu, Y., Huang, L., Zhang, Z., Huang, B., Zhang, H., Ma, J., Li, B., Liu, X., Wu, G., Zhang, J., Shen, L., Liu, Q., Ni, J., 2022. Deep transfer learning of structural magnetic resonance imaging fused with blood parameters improves brain age prediction. Hum Brain Mapp 43, 1640–1656. https://doi.org/10.1002/HBM.25748

Roe, J.M., Vidal-Piñeiro, D., Sørensen, Ø., Brandmaier, A.M., Düzel, S., Gonzalez, H.A., Kievit, R.A., Knights, E., Kühn, S., Lindenberger, U., Mowinckel, A.M., Nyberg, L., Park, D.C., Pudas, S., Rundle, M.M., Walhovd, K.B., Fjell, A.M., Westerhausen, R., Masters, C.L., Bush, A.I., Fowler, C., Darby, D., Pertile, K., Restrepo, C., Roberts, B., Robertson, J., Rumble, R., Ryan, T., Collins, S., Thai, C., Trounson, B., Lennon, K., Li, Q.X., Ugarte, F.Y., Volitakis, I., Vovos, M., Williams, R., Baker, J., Russell, A., Peretti, M., Milicic, L., Lim, L., Rodrigues, M., Taddei, K., Taddei, T., Hone, E., Lim, F., Fernandez, S., Rainey-Smith, S., Pedrini, S., Martins, R., Doecke, J., Bourgeat, P., Fripp, J., Gibson, S., Leroux, H., Hanson, D., Dore, V., Zhang, P., Burnham, S., Rowe, C.C., Villemagne, V.L., Yates, P., Pejoska, S.B., Jones, G., Ames, D., Cyarto, E., Lautenschlager, N., Barnham, K., Cheng, L., Hill, A., Killeen, N., Maruff, P., Silbert, B., Brown, B., Sohrabi, H., Savage, G., Vacher, M., 2021. Asymmetric thinning of the cerebral cortex across the adult lifespan is accelerated in Alzheimer's disease. Nat Commun 12. https://doi.org/10.1038/s41467-021-21057-y

Sajedi, H., Pardakhti, N., 2019. Age Prediction Based on Brain MRI Image: A Survey. J Med Syst. https://doi.org/10.1007/s10916-019-1401-7

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.,

medRxiv preprint doi: https://doi.org/10.1101/2022.09.06.22279594; this version posted May 31, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



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

Tanveer, M., Ganaie, M.A., Beheshti, I., Goel, T., Ahmad, N., Lai, K.-T., Huang, K., Zhang, Y.-D., Ser, J. Del, Lin, C.-T., 2022. Deep Learning for Brain Age Estimation: A Systematic Review.

Thambisetty, M., Wan, J., Carass, A., An, Y., Prince, J.L., Resnick, S.M., 2010. Longitudinal changes in cortical thickness associated with normal aging. Neuroimage 52, 1215–1223. https://doi.org/10.1016/j.neuroimage.2010.04.258

Tian, Y.E., Cropley, V., Maier, A.B., Lautenschlager, N.T., Breakspear, M., Zalesky, A., 2023. Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. Nat Med. https://doi.org/10.1038/s41591-023-02296-6

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