Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker

James H Cole[1], Rudra PK Poudel[2], Dimosthenis Tsagkrasoulis[3], Matthan WA Caan[4], Claire Steves[5], Tim D Spector[5], Giovanni Montana[2,3*]

Author affiliations:

[1]Computational, Cognitive & Clinical Neuroimaging Laboratory, Division of Brain Sciences, Imperial College London, London, UK.

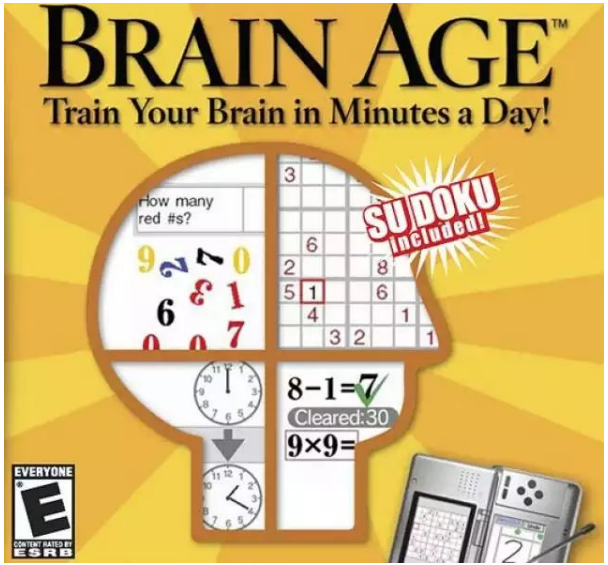
[2]Department of Biomedical Engineering, King's College London, London, UK.

[3]Department of Mathematics, Imperial College London, London, UK.

[4]Department of Radiology, Academic Medical Center, Amsterdam.

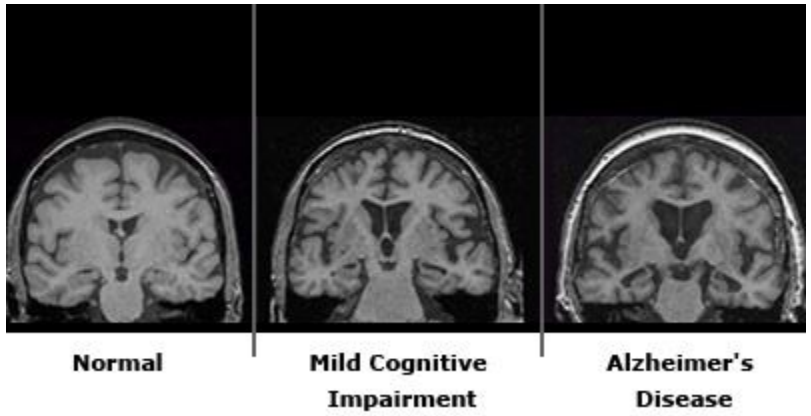
[5]Department of Twin Research & Genetic Epidemiology, King's College London, London, UK.

Brain Age



http://www.vgchartz.com/articles_media/images/brain-age-train-your-brain-in-minutes-a-day-1.jpg

Brain Age



http://www.mayo.edu/research/labs/aging-dementia-imaging/overview

Motivation

"By training models on healthy individuals, brain-based predictions of age can then be made in independent clinical samples. If 'brain-predicted age' is greater than an individual's chronological age, this is thought to reflect some aberrant accumulation of age related changes to the brain"

Data

Dataset	Size	M/F	Mean Age
Brain-Age Normative Control (BANC)	2001	1016/985	36.95 ± 18.12
UK Adult Twin Registry Sub-Study*	62	0/62	61.86 ± 8.36
Study of Reliability of MRI (STORM)†	20	12/8	34.05 ± 8.71
Between Scanner Reliability Study‡	11	7/4	30.88 ± 6.16

Data

Dataset	Size	M/F	Mean Age
Brain-Age Normative Control (BANC)	2001	1016/985	36.95 ± 18.12
UK Adult Twin Registry Sub-Study*	62	0/62	61.86 ± 8.36
Study of Reliability of MRI (STORM)†	20	12/8	34.05 ± 8.71
Between Scanner Reliability Study‡	11	7/4	30.88 ± 6.16

* 27 monozygotic, 4 dizygotic
† Imperial College London, 28.35 ± 1.09 days between scans.
‡ Imperial College London, Academic Medical Center Amsterdam, 68.17 ± 92.23 days between scans. Eight participants were first scanned in Amsterdam, the remaining three in London.

Preprocessing

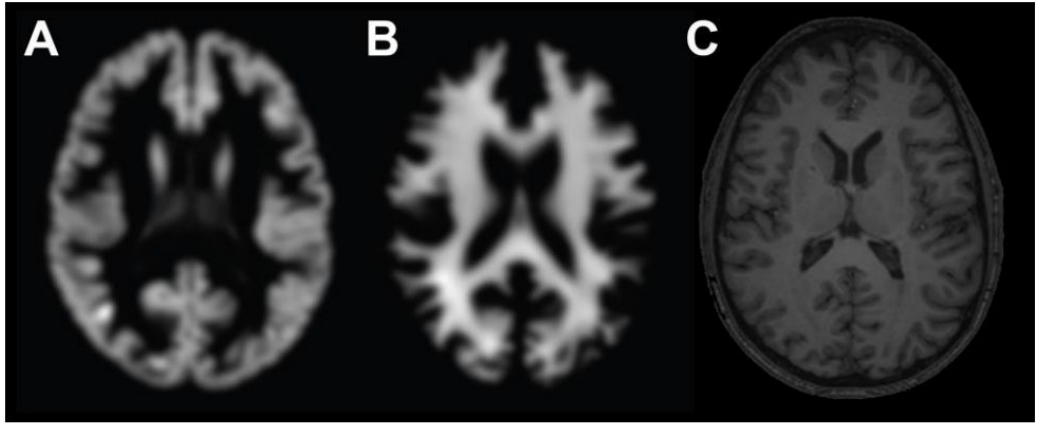
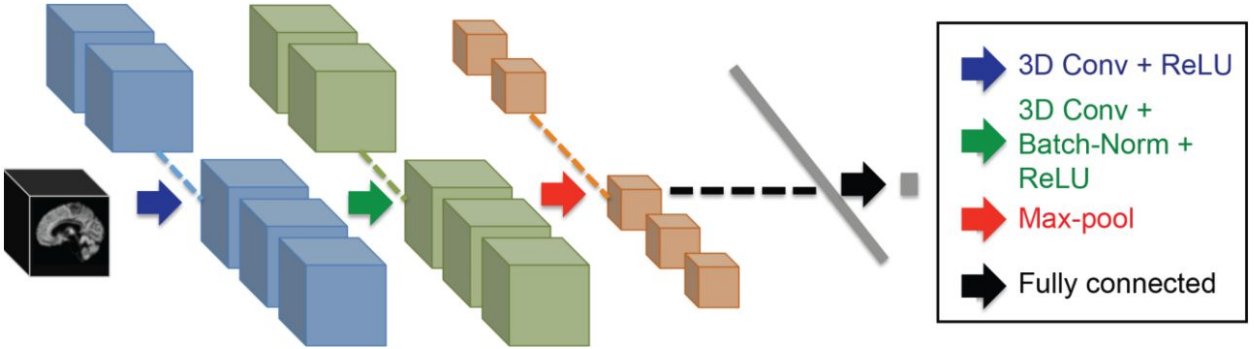


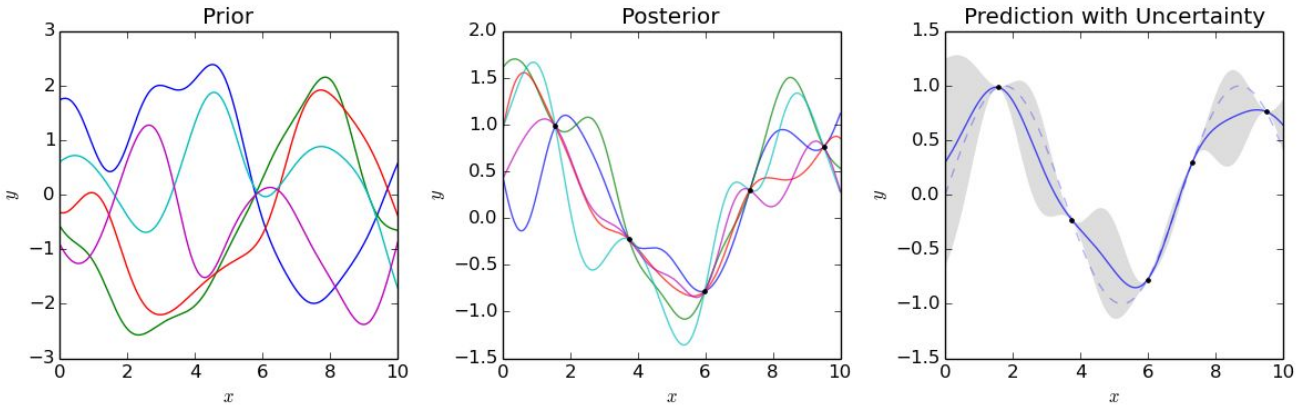
Fig. 1. Examples of neuroimaging input data for use in age prediction models. A) Grey matter volumetric map, normalised to MNI152 space using SPM DARTEL for non-linear registration, 4mm smoothing and modulation, in axial view. B) White matter volumetric map, normalised to MNI152 space, in axial view. C) Raw, or minimally-processed, T1-weighted MRI, rigidly-registered to MNI152 space and resampled to a common voxel space.

Models

Convolutional Neural Network



Gaussian Process Regression



(top) Figure 2, Page 6
(bottom) https://en.wikipedia.org/wiki/Gaussian_process#/media/File:Gaussian_Process_Regression.png

Prediction Results

Table 1. Chronological age prediction accuracy

Method	Input data	MAE (years)	r	R^2	RMSE
CNN	GM	4.16	0.96	0.92	5.31
	WM	5.14	0.94	0.88	6.54
	GM+WM	4.34	0.96	0.91	5.67
	Raw	4.65	0.94	0.88	6.46
GPR	GM	4.66	0.95	0.89	6.01
	WM	5.88	0.92	0.84	7.25
	GM+WM	4.41	0.96	0.91	5.43
	Raw	11.81	0.57	0.32	15.10

MAE = mean absolute error, r = Pearson's r from correlation between chronological age and brain-predicted age, RMSE = root mean squared error, GM = Grey Matter, WM = White Matter.

Reliability Results

Heiritability

Table 2. Heritability estimates from the AE SEM models for different brain-predicted age methods

Method	GM	WM	GM+WM	Raw
Unadjusted				
CNN	0.74 ± 0.09	0.78 ± 0.07	0.84 ± 0.05	0.62 ± 0.10
GPR	0.78 ± 0.07	0.81 ± 0.06	0.82 ± 0.06	0.64 ± 0.10
With age-correction				
CNN	0.55 ± 0.11	0.65 ± 0.10	0.66 ± 0.09	0.50 ± 0.12
GPR	0.55 ± 0.11	0.60 ± 0.10	0.58 ± 0.11	0.64 ± 0.10

Heritability estimates are given by $h^2 = \frac{a^2}{a^2+e^2}$, where *a* and *e* are the path coefficients of the A and E variance components in the SEM model, ± the standard errors of the estimates. GM = grey matter, WM = white matter, CNN = convolutional neural network, GPR = Gaussian processes regression.

Reliability Results

Technical Confounders

Table 3. Within-scanner and between-scanner reliability estimates of brain-predicted age difference

Method	Dataset	GM	WM	GM+WM	Raw
CNN	Within	0.90 [0.76, 0.96]	0.97 [0.90, 0.99]	0.90 [0.77, 0.96]	0.94 [0.86, 0.98]
	Between	0.83 [0.49, 0.95]	0.51 [-0.08, 0.84]	0.85 [0.55, 0.96]	0.66 [0.17, 0.89]
GPR	Within	0.96 [0.90, 0.98]	0.98 [0.94, 0.99]	0.97 [0.92, 0.99]	0.99 [0.97, 0.99]
	Between	0.96 [0.88, 0.99]	0.77 [0.12, 0.94]	0.92 [0.74, 0.98]	0.56 [-0.02, 0.86]

All figures in the table are intraclass correlation coefficients (ICC) and 95% confidence intervals. GM = grey matter, WM = white matter, CNN = convolutional neural network, GPR = Gaussian processes regression.

Reliability Results

Technical Confounders

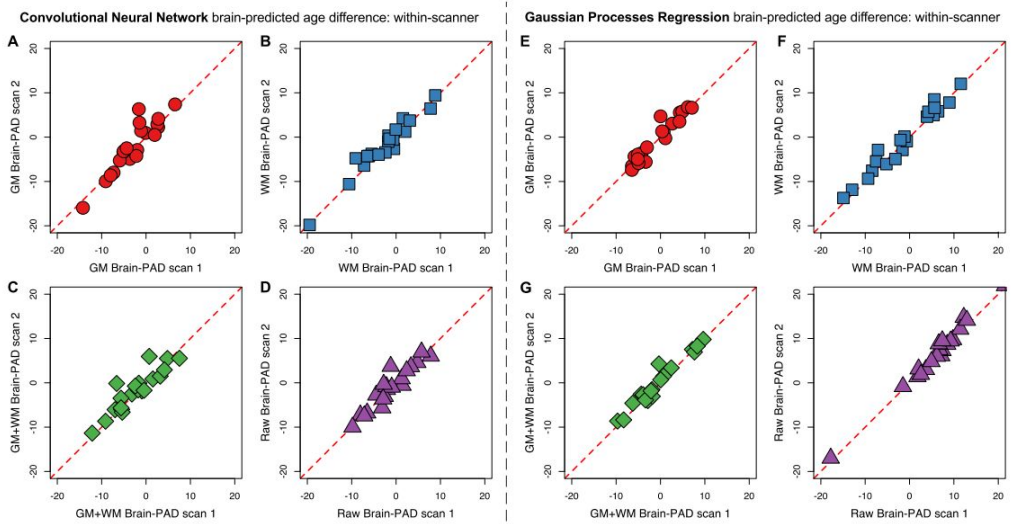


Fig. 4. Within-scanner reliability for Convolutional Neural Networks and Gaussian Processes Regression. Figure shows the correspondence between brain-predicted age difference (Brain-PAD) based on scans acquired four weeks apart on the same scanner (Siemens Verio 3T) on N = 20 individuals, with scan 1 on the x-axis and scan 2 (after four weeks) on the y-axis for all plots. A) Brain-PAD score based on GM maps using CNN. B) Brain-PAD score based on WM maps using CNN. C) Brain-PAD scored based on GM and WM maps combined using CNN. D) Brain-PAD scored based on raw T1-MRI using CNN. E) Brain-PAD score based on GM maps using Gaussian Processes Regression (GPR). F) Brain-

Reliability Results

Technical Confounders

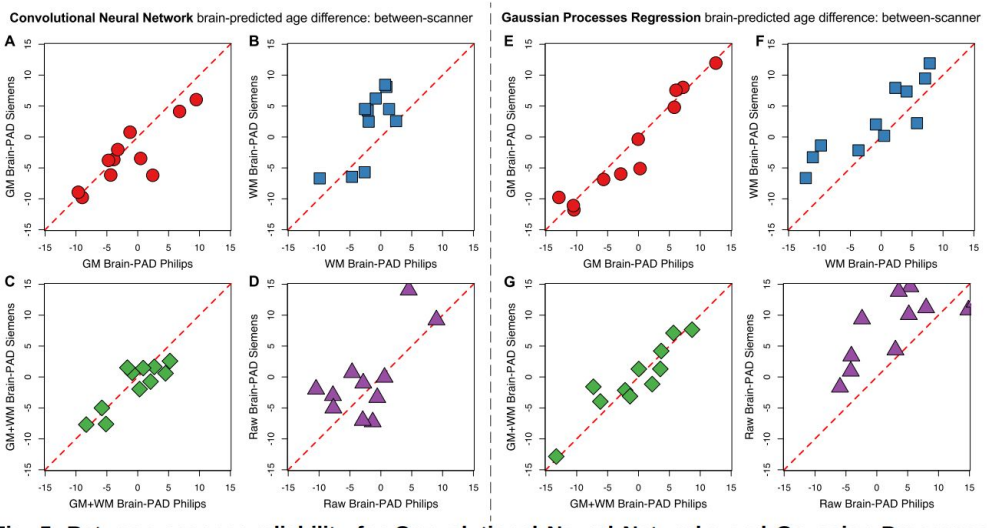


Fig. 5. Between-scanner reliability for Convolutional Neural Networks and Gaussian Processes Regression. Figure shows the correspondence between brain-predicted age difference (Brain-PAD) scores based on scans acquired on two different scanner systems (Siemens Verio 3T and Philips Intera 3T) in N = 11 individuals, with the Philips scan on the x-axis and Siemens scan on the y-axis for all plots. A) Brain-PAD score based on GM maps using CNN. B) Brain-PAD score based on WM maps using CNN. C) Brain-PAD scored based on GM and WM maps combined using CNN. D) Brain-PAD scored based on raw T1-MRI using CNN. E) Brain-PAD score based on GM maps using Gaussian Processes Regression (GPR). F) Brain-PAD score based on WM maps using GPR. G) Brain-PAD score based on GM and WM maps combined using GPR. The red dashed line in all plots is the line of identity.

Conclusions

Pros:

- Nice evaluation of reliability and generalizability.
- Heritability presents a novel way to assess biologic underpinnings, could be applied more broadly.
- Model offers real practical utility.

Cons:

- Age confounding heritability estimation?
- No additional comparisons beyond GPR and CNN.
- Population reliability (e.g. outside UK) not assessed.