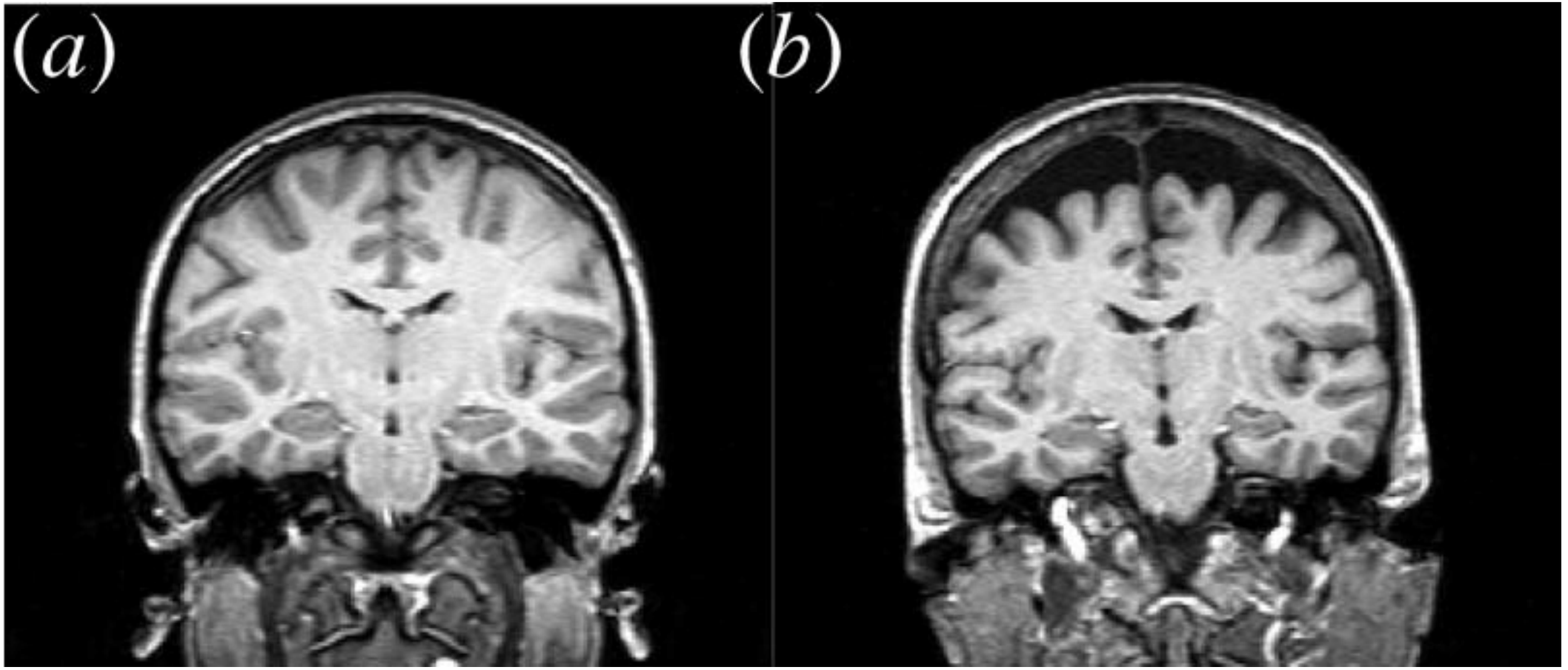
Brain age prediction using deep learning uncovers associated sequence variants

Jonsson, B. A., *et al*. (2019).
Nature communications, 10(1), 5409.

presented by 10801128 Chun-Hung, Chen

Introduction

Can MRI differ young and elderly brain?



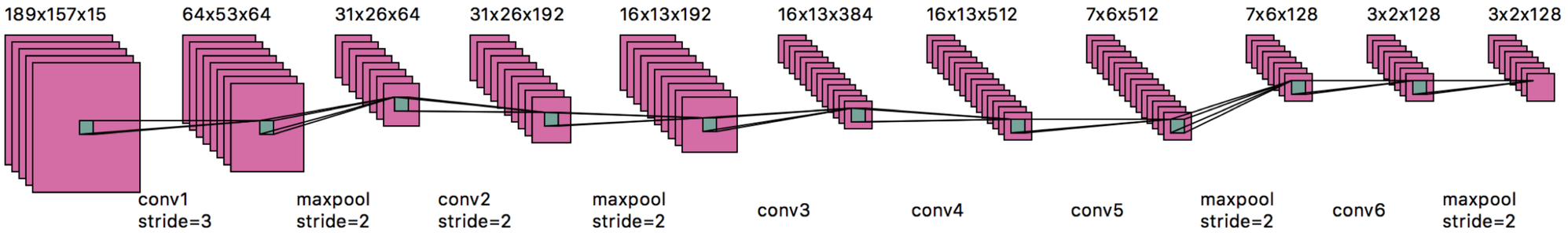
Shan, Z. Y., *et al*. (2005). Selective atrophy of left hemisphere and frontal lobe of the brain in old men. The journals of gerontology. Series A, Biological sciences and medical sciences, 60(2), 165–174.

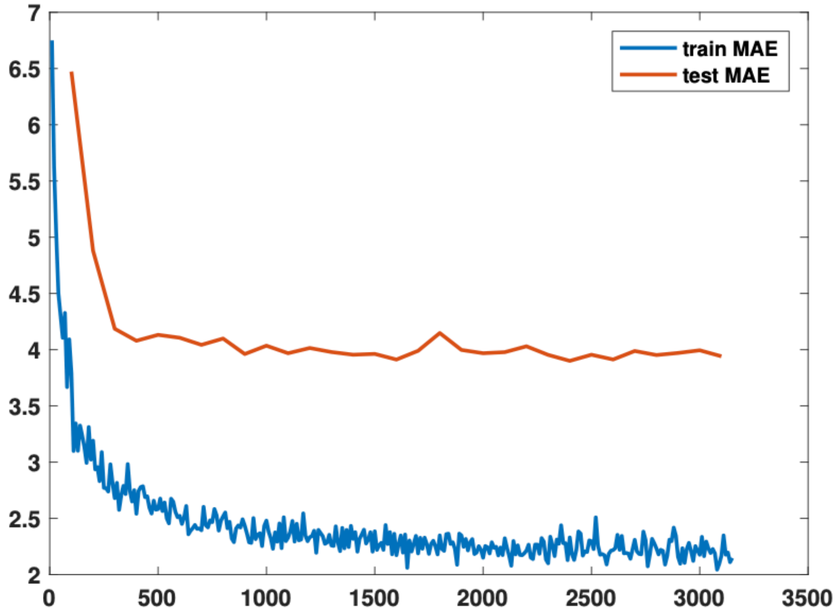
What is needed for brain age prediction?

In “traditional” way, the following features were preferred to be extracted:

- Principal components
- Cortical thickness and surface curvature
- Volume of GM, WM, and CSF

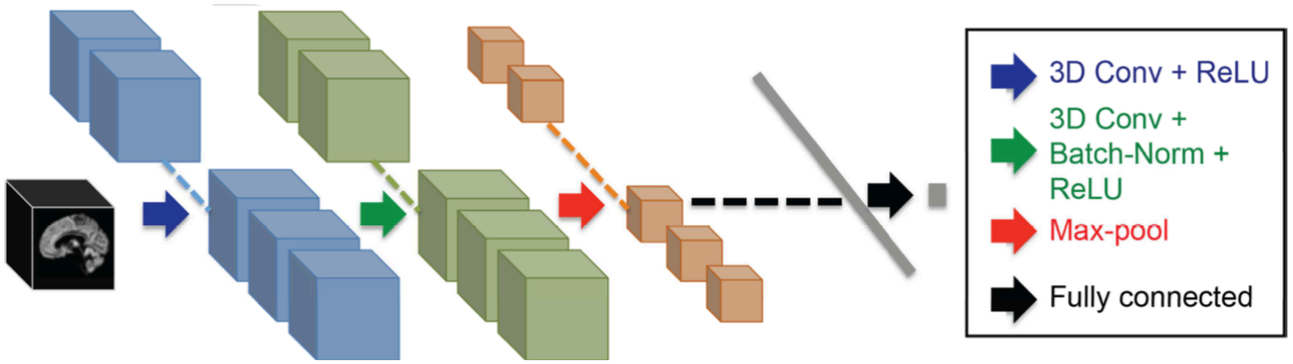
Brain age prediction is not something new.



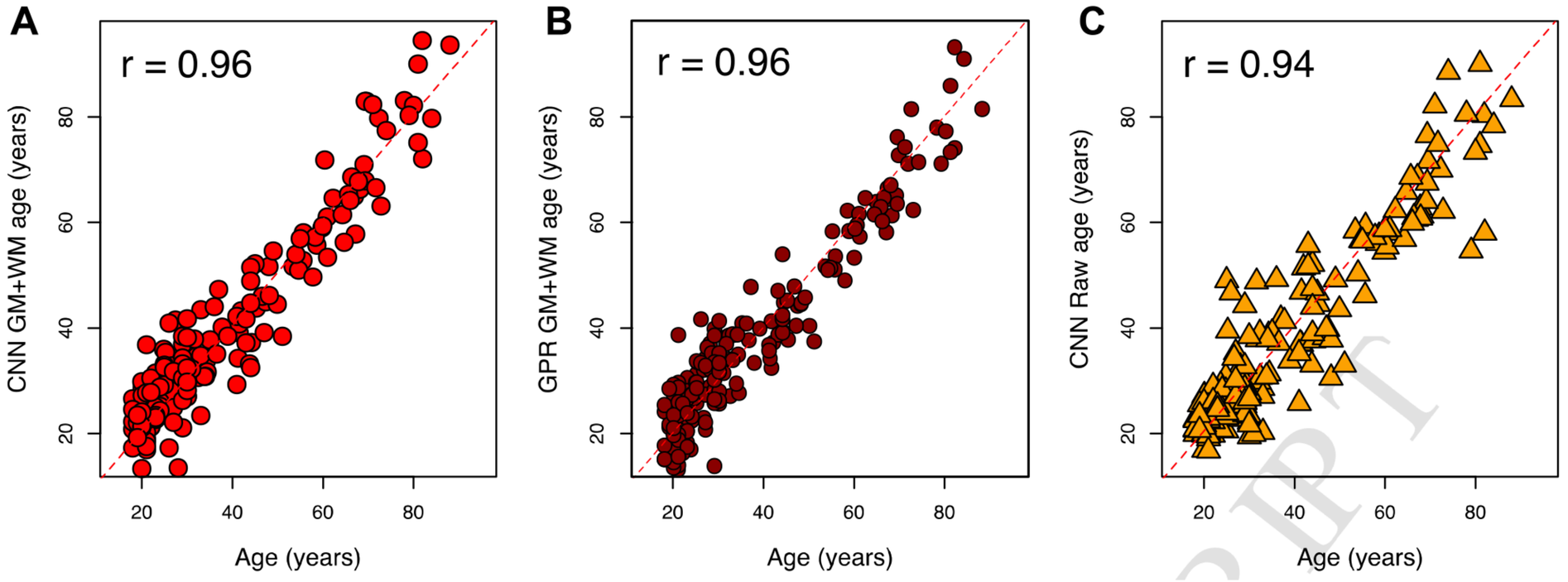


T. Huang, *et al*. (2017)
Age estimation from brain MRI images using deep learning. 2017 IEEE 14th International Symposium on Biomedical Imaging, pp. 849-852.

Brain age prediction is not something new.



Cole, J. H., *et al.* (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. NeuroImage, 163, 115–124.



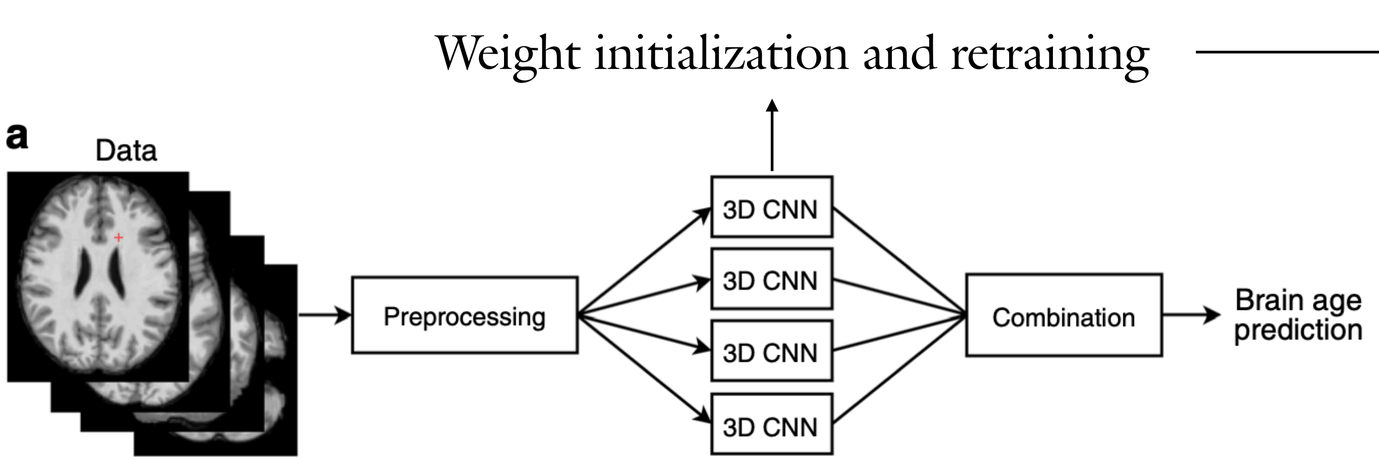
Why PAD matters?

PAD, standing for predicted age difference, is proved to correlate mental and physical fitness.

Also, it has been demonstrated to be “heritable” and to have a “polygenic overlap with brain disorders” such as schizophrenia, bipolar disorder, multiple sclerosis, and Alzheimer’s disease.

Methods

Overview



Weight initialization and retraining → PAD ↔ performance on neuropsychological tests

PAD → genetic analysis → sequence variants

Datasets

Three T1-weighted MRIs datasets:

- Icelandic dataset
- the UK Biobank dataset
- the IXI dataset

Icelandic dataset(for training)

- Scans from 1264 healthy subjects aged between 18 and 75 years
- 1.5T Philips Achieva:

  TR = 8.6 ms, TE = 4.0 ms, FA = 8°, 170 slices, slice thickness = 1.2 mm,

  acquisition matrix = 192 * 192, FOV = 240 * 240 mm
- 1.5T Siemens Aera:

  TR = 2400 ms, TE = 3.54 ms, FA = 8°, 160 slices, slice thickness = 1.2 mm,

  acquisition matrix = 192 * 192, FOV = 240 * 240 mm

the UK Biobank dataset(for transfer learning)

- Scans from 15040 healthy subjects aged between 46 and 79 years old with 3T Siemens Skyra
- On account of that the presence of undetected population structure can lead to both false positive results and failure to detect genuine associations in genetic association studies, in an effort to combat this our analysis was constrained to 12378 individuals of white British ancestry.

the IXI dataset(for transfer learning)

- Scans from 544 healthy subjects aged between 20 and 86 years old with a Philips 3T system, a Philips 1.5T system and a GE 1.5T system

Preprocessing

1. The input data were inhomogeneity corrected. (with CAT12)
2. The skull and other non-brain elements were removed. (with CAT12)
3. The images were registered into the standard MNI space using the deformable registration algorithm DARTEL.

Three types of images were generated:

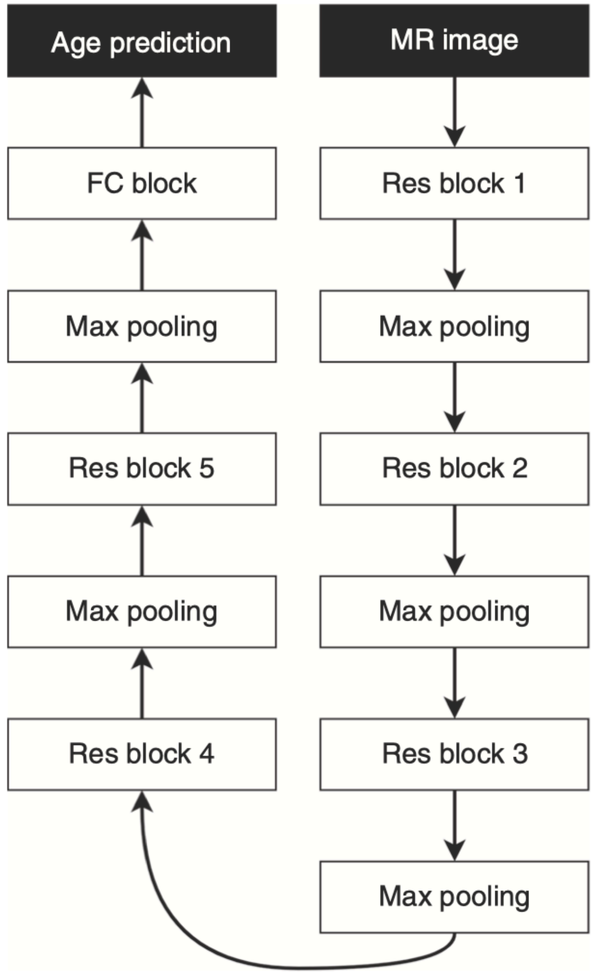
1. MNI-registered image 2. Jacobian map 3. GM and WM soft segmented image

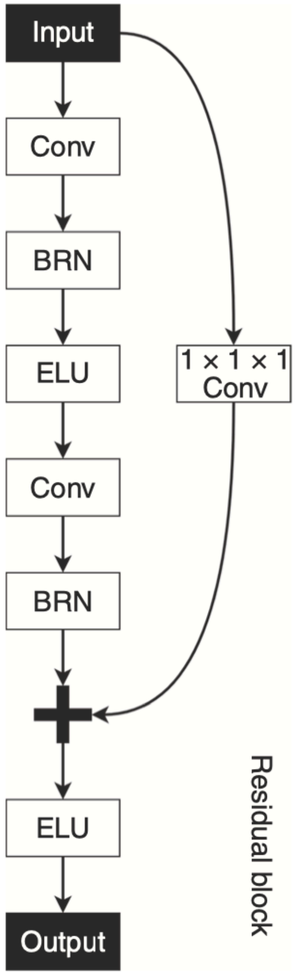
Data augmentation

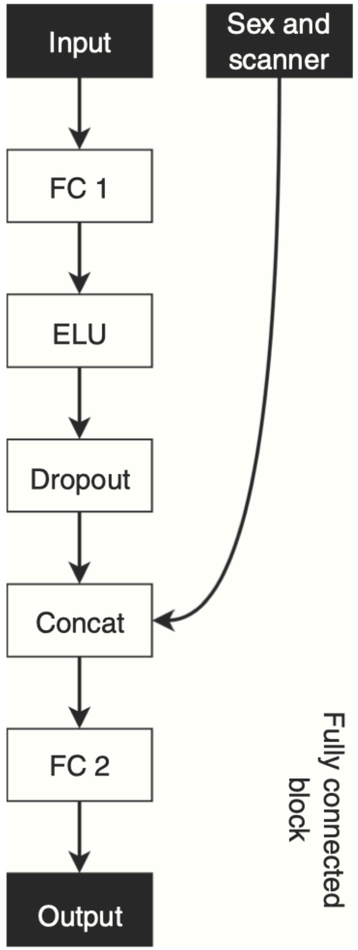
Applying a coordinate transformation to a random subset of the training data, consisting of a combined 3D rotation and a 3D translation

- Rotation angles were between $-40°$ and $40°$ with equal probability
- Translation distance, for each direction, was selected between -10 and 10 voxels with equal probability

CNN architecture







Parameters of each sublayer

3D Convolution:

size = 3*3*3; stride = 1*1*1; activation = ELU;

MaxPooling:

size = 3*3*3; stride = 2*2*2;

MLP:

input size=10240; # of hidden units = 256; output size = 1; activation = ELU

Dropout:

keep rate = 0.8

Parameters of architecture

Loss function = MAE(mean absolute error);

Optimizer = Adam with learning rate = 0.001, decay = 10^-6, $\beta 1 = 0.9$, $\beta 2 = 0.999$;

Regularizer = l2 with $\lambda = 5 * 10^{-5}$;

Batch size = 4;

Epochs = 100;

Difference with Cole

Cole:

VGGNet-like architecture

- vanishing gradient problem limits the potential depth of the network

No input about sex and scanner

Jonsson:

ResNet-like architecture

- no such depth limits
- smoother loss surfaces

Including sex and scanner as input

Using deformation information encoded in Jacobian maps to predict brain age

Pearson correlation coefficient

In order to test for association between PAD and performance on neuropsychological tests

Adjusting the PAD for age at imaging visit, age^2, sex, age*sex, (age^2)*sex, total intracranial volume, the first 40 principal components from genetic ancestry analysis, head motion, genotyping array, imaging center, and assessment center where neuropsychological tests were conducted. The adjustments was performed using linear regression.

Results

Combining CNN outputs improves prediction accuracy

	Type	Method	Val MAE	Val R^2	Test MAE	Test R^2	No. I
(A)	T1-weighted	CNN	**3.996**	**0.810**	**4.006**	**0.829**	1815
	Jacobian	CNN	4.801	0.710	4.804	0.758	1815
	Gray matter	CNN	4.766	0.721	4.641	0.776	1815
	White matter	CNN	4.676	0.735	4.189	0.812	1815
(B)	MV (T1 and JM)	CNN	4.102	0.803	3.919	0.841	1815
	MV (GM and WM)	CNN	4.172	0.790	3.674	0.849	1815
	MV (T1, JM, and GM)	CNN	3.964	0.813	3.838	0.847	1815
	MV (T1, JM, GM, and WM)	CNN	3.845	**0.849**	3.584	0.849	1815
	LRB (T1, JM, GM, and WM)	CNN	**3.581**	0.847	**3.388**	**0.872**	1815
(C)	SBM	RR	5.268	0.689	5.176	0.697	1320
	VBM	GPR	4.278	0.781	4.317	**0.766**	1794
	SM	RR	4.898	0.722	4.937	0.728	1815
	MV (SBM, VBM, and SM)	GPR/RR	4.008	0.808	3.940	0.761	1246
	LRB (SBM, VBM, and SM)	GPR/RR	**3.906**	**0.812**	**3.849**	**0.766**	1246

Transfer Learning

	IXI			UK Biobank		
TL Used	Val MAE	Val R^2	Val set size	Test MAE	Test R^2	Test set size
No	6.420	0.778	104	8.494	−0.630	12395
Yes	**4.149**	**0.907**	104	**3.631**	**0.614**	12395

The best results are shown in bold
S subjects, *TL* transfer learning, *val* validation

From weight initialization to PAD

1. Train four CNNs on the Icelandic dataset on the four previously mentioned image types.

2. Freeze convolutions layers and train the CNNs on the IXI dataset (transfer learning step).

3. Predict brain age in the UK Biobank dataset using CNNs, combine the predictions with majority voting and calculate PAD values.

the UK Biobank with MAE equal to 4.6, 5.5, 5.4, and 4.9 years in four repeats.

The reason why the error is higher here compared with the original results is that we did not reinitialize and retrain the CNNs in cases were the optimization got stuck in a poor local minimum or a saddle point.

Associations between PAD and performance on neuropsychological tests

Training with the IXI dataset for transfer learning in the UK Biobank dataset

The result is no evidence of association between PAD and performance on the fluid intelligence, numeric memory, pairs matching, and prospective memory tests but worse performance on the digit substitution test (DSST), trail making tests (TMTs), and the reaction time test.

Neuropsychological test	**PAD correlation**	**95% CI**	***P* Value**	**No. subjects**
DSST	−0.080	(−0.104, −0.054)	4.3e−11	6849
TMT B	0.076	(0.051, 0.103)	3.1e−09	6076
TMT A	0.053	(0.027, 0.078)	3.8e−05	6076
TMT B - A	0.050	(0.024, 0.075)	1.3e−04	5918
Reaction time	0.030	(0.012, 0.047)	7.9e−04	12387

Genome-wide association study

	rs Number (GRCh38)	Position (min/maj)	Allele	MAF (%)	Effect	95% CI	*P* Value
(A)	rs2435204	chr17:45910839	G/A	26.6	0.11	(0.08, 0.14)	1.4e−12
	rs1452628	chr1:214966544	T/A	36.2	−0.08	(−0.10, −0.05)	2.3e−09
(B)	rs2790099	chr6:45475612	C/T	36.0	−0.06	(−0.09, −0.03)	8.9e−06
	rs6437412	chr3:194747684	C/T	28.2	−0.06	(−0.09, −0.04)	6.8e−06
	rs2184968	chr6:126439848	C/T	46.0	0.05	(0.03, 0.08)	7.5e−05
(C)	rs2435204	chr17:45910839	G/A	26.6	0.08	(0.03, 0.13)	1.5e−03
	rs1452628	chr1:214966544	T/A	36.2	−0.07	(−0.12, −0.03)	8.8e−04
(D)	rs2790099	chr6:45475612	C/T	36.0	−0.07	(−0.11, −0.02)	2.9e−03
	rs6437412	chr3:194747684	C/T	28.2	−0.05	(−0.09, 0.00)	4.9e−02
	rs2184968	chr6:126439848	C/T	46.0	0.06	(0.02, 0.10)	2.9e−03

(A, B) Association between sequence variants and PAD for 12378 subjects in discovery set. (A) Genome-wide significant sequence variants. (B) Sequence variants associated with structural MRI brain phenotypes that also associate with PAD. (C, D) Association between sequence variants and PAD for 4456 subjects from the replication set. (C) Genome-wide significant sequence variants. (D) Sequence variants associated with structural MRI brain phenotypes that also associate with PAD. Note that the reported effect sizes are for PAD normalized to unit variance. Before normalization the standard deviation of PAD was ~4 years. Thus the associated lowering of the protective allele of rs1452628 is approximately −0.32 years

CI confidence interval, *MAF* minor allele frequency

Discussion