



OPEN

## Predicting 'Brainage' in late childhood to adolescence (6-17yrs) using structural MRI, morphometric similarity, and machine learning

Daniel Griffiths-King<sup>1</sup>, Amanda G. Wood<sup>1,2,3,4</sup> & Jan Novak<sup>1,4</sup>✉

Brain development is regularly studied using structural MRI. Recently, studies have used a combination of statistical learning and large-scale imaging databases of healthy children to predict an individual's age from structural MRI. This data-driven, predicted 'Brainage' typically differs from the subjects chronological age, with this difference a potential measure of individual difference. Few studies have leveraged higher-order or connectomic representations of structural MRI data for this Brainage approach. We leveraged morphometric similarity as a network-level approach to structural MRI to generate predictive models of age. We benchmarked these novel Brainage approaches using morphometric similarity against more typical, single feature (i.e., cortical thickness) approaches. We showed that these novel methods did not outperform cortical thickness or cortical volume measures. All models were significantly biased by age, but robust to motion confounds. The main results show that, whilst morphometric similarity mapping may be a novel way to leverage additional information from a T1-weighted structural MRI beyond individual features, in the context of a Brainage framework, morphometric similarity does not provide more accurate predictions of age. Morphometric similarity as a network-level approach to structural MRI may be poorly positioned to study individual differences in brain development in healthy participants in this way.

Developmental neuroscience has embraced neuroimaging studies of brain structure to characterize brain maturation and to understand how this gives rise to cognitive development. Developmental neuroimaging studies have highlighted distinct developmental trajectories of specific cortical tissues such as white matter (WM) and grey matter (GM), across different regions of the cortex<sup>1</sup>. The volume of cortical GM specifically shows an 'inverted U', nonlinear trajectory<sup>1-4</sup>, with pre-pubertal expansion of the cortical GM<sup>5</sup> followed by a post-pubertal sustained loss of GM volume (despite synaptic density plateauing after puberty according to molecular and cellular evidence<sup>3</sup>). Brain maturation has specific regional trajectories; peak GM density and reductions in GM volume occur earliest in primary function areas, somatosensory and primary motor cortices, and latest in higher-order association areas, dorsolateral prefrontal cortex and superior temporal gyrus for instance<sup>1</sup>. Cortical thickness maturation also shows a similar trajectory, with generalized reductions over time<sup>6-8</sup>, in line with what would be expected from models of synaptic pruning and myelination<sup>8</sup>. These longstanding findings show, given these measures vary as a function of age, that an individual's chronological age may be deduced from an MRI scan of their brain.

This is the premise of the Brainage framework, the idea that multivariate patterns of brain structure in large samples of MRI from healthy children are related to age and, by using data-driven or machine learning approaches, that association can be learnt. By applying these learnt patterns to new data, we can predict the age of an individual based on their MRI (see<sup>9,10</sup> for review). This apparent age, or more commonly termed "Brainage", is akin to a reading age; it reflects the current observed status of the brain in terms of morphometry in comparison to 'typical' norms of brain structural development.

The Brainage of any individual is unlikely to be perfectly aligned to their chronological (actual) age. Differences between Brainage and chronological age may reflect normal-variation or individual differences between children. This metric of difference and/or perturbation is typically referred to as Brainage $\Delta$  (delta), the calculated

<sup>1</sup>Aston Institute of Health and Neurodevelopment, College of Health and Life Sciences, Aston University, Birmingham B4 7ET, UK. <sup>2</sup>School of Psychology, Faculty of Health, Melbourne Burwood Campus, Deakin University, Geelong, VIC, Australia. <sup>3</sup>Murdoch Children's Research Institute, Melbourne, VIC, Australia. <sup>4</sup>These authors jointly supervised this work: Amanda G. Wood and Jan Novak. ✉email: j.novak@aston.ac.uk

difference between chronological age and apparent/predicted Brainage<sup>9,11</sup>. In the case of diseased populations, this measure allows us to estimate the perturbation that the disease state has upon brain development and aging (i.e.,<sup>11</sup>). For instance, a Brainage $\Delta$  of zero would be indicative of an individual following a normative developmental trajectory, whilst a higher or greater Brainage $\Delta$  would represent advanced brain (and possibly cognitive) aging, a state of perturbation from the typical developmental trajectory<sup>12</sup>. Brain development (specifically grey matter change) follows highly ‘programmed’ trajectories<sup>5,13–15</sup> (driven in part by genetics<sup>16–19</sup>). Therefore, neurological disruption to the ongoing development of the brain during this period is likely to potentially symptomatic, impacting on future brain and cognitive maturation. Therefore, an approach for quantifying typical brain maturation will likely hold benefit in understanding atypical developmental patterns that hold clinical implication<sup>20,21</sup>.

In recent years, data-driven estimations of the brain’s apparent age have been calculated using machine-learning approaches to detect latent patterns associated with age across several neuroimaging modalities (including structural (sMRI), diffusion (dMRI) and functional MRI (fMRI)). Utilising machine learning approaches in this way, can consider the multivariate complexities of the neurodevelopmental trajectories of these meso-scale measures. However, using regional-level data as independent features to predict age may ignore potentially relevant, higher-order multicollinearities between regions. Connectomic approaches<sup>22</sup>, that consider the brains’ network-level organisation, may therefore hold greater potential in predicting Brainage. Typically, connectomic approaches would utilise diffusion and functional MRI data, but these can suffer from quality issues associated with EPI sequences<sup>23</sup>, and also have long acquisition times and may therefore be less tolerable in clinical populations and paediatrics<sup>24</sup>.

This study proposes connectomic approaches to sMRI data as a potential method to use in the Brainage framework. Previous approaches utilising individual morphometric measurements from sMRI in Brainage prediction in paediatrics<sup>20,25,26</sup> have achieved comparable prediction accuracies to those multimodal studies incorporating additional modalities with sMRI<sup>27,28</sup> (although, methodological differences preclude meaningful direct comparison across these paediatric studies). However, relatively few studies leverage connectomics-level analyses of sMRI data for Brainage prediction. Corps and Reikik<sup>23</sup> utilised morphometric data (curvature, cortical thickness, sulcal depth) to produce multi-view morphological brain networks. Briefly, they investigated multiple feature-networks where ‘connections’ are the dissimilarity (absolute difference) between regions in terms of absolute feature values. This produced multiple feature networks as predictive variables. No previous studies have leveraged network-level approaches to sMRI to generate single morphometric networks as variables in predictive models of Brainage. Given the abundance of available sMRI data (T1w MRI is more readily acquired across clinical and research contexts compared to other MRI modalities and therefore may find greater translatability and application), and the effective Brainage prediction previously achieved using this data<sup>27,28</sup>, there is strong rationale in trying to leverage additional predictive information from this modality of MRI, using a connectomic approach.

In the current paper, we employed a morphometric similarity mapping approach<sup>29</sup> to combine multiple features into a single network, capturing higher-order morphometric organisation across the cortex. Previously these networks have been shown to be sensitive to neurodevelopmental abnormalities<sup>30</sup>.

Specifically, this paper generated normative Brainage models using connectomic approaches to sMRI, as outlined in King and Wood<sup>24</sup>, leveraging both network-level approaches whilst restricting necessary MRI sequences to a T1w sMRI. This approach may better account for absolute dissimilarity due to scaling (as in Corps and Reikik<sup>23</sup>) and instead capture those relationships that are indicative of coordinated cortical development and maturation<sup>29</sup>.

The current study evaluates the use of T1w morphometric similarity mapping, as a novel approach in predicting Brainage in a cohort of typically developing children. This study investigates whether Brainage prediction methods are more accurate when using morphometric similarity measures of the developing cortex, compared to individual morphometric measures. No previous study has benchmarked these novel Brainage approaches (using morphometric similarity) against more typical, single morphometric feature approaches.

## Results

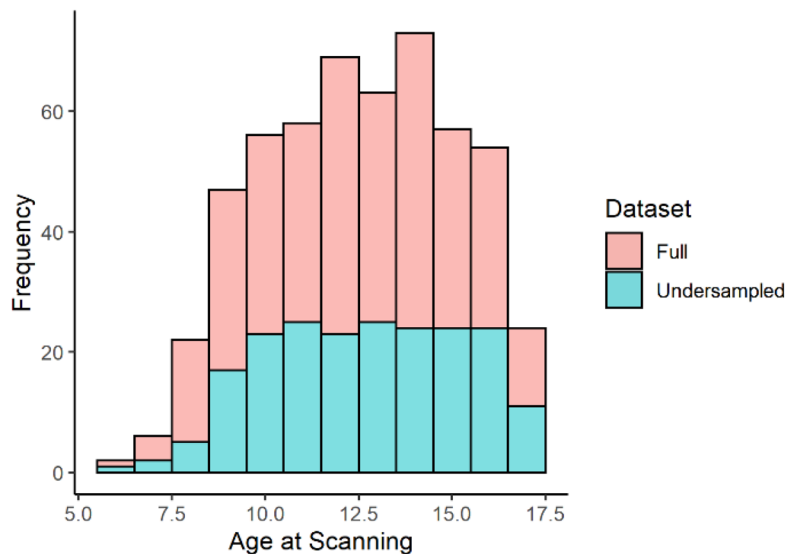
**Dataset.** This study employed data from healthy controls from the open-access Autism Brain Imaging Data Exchange cohort (ABIDE, Di Martino, Yan<sup>31</sup>) data from the Pre-processed Connectome Project (PCP, Bellec, Yan<sup>32</sup>, for full details see Pre-processed Connectome Project website <http://preprocessed-connectomes-project.org/>). Healthy controls were included that were < 17 years old and met strict quality control criteria (outlined below). After applying these criteria, the remaining cases from the ABIDE dataset used in the current analyses consisted of 327 healthy controls, with a mean age of  $12.4 \pm 2.5$  years. (see Table 1). We utilized the Freesurfer<sup>33</sup> processed outputs supplied by the PCP. This provides cortical morphometry measures across regions of the Desikan-Killiany atlas<sup>34</sup>.

**Model evaluation.** To evaluate the Brainage models derived from different morphometric feature sets, specifically morphometric similarity, compared to individual morphometric features, the ABIDE cohort were divided into training and independent test cohorts in the ratio of 3:1 ( $n = 245$  &  $n = 82$  respectively). The training cohort was further subdivided into an internal training and validation cohort at a ratio of 5:1 ( $n = 204$  &  $n = 41$ ). Selection of the training-set was pseudo-random to enable under sampling based on age (see Fig. 1).

Prediction utilized 10 different feature-sets (See Table 2); i–vii) each of the individual morphometric features, viii) all features, ix) nodal-level strength of morphometric similarity and x) edge-level weights of the morphometric similarity, with each model having chronological age at scanning as the dependent variable. Across models, performance was evaluated based upon reducing mean absolute error (MAE) and maximizing predictive  $R^2$ .

	Entire cohort	Internal training cohort	Internal validation cohort	Test cohort
n	327	204	41	82
Mean age (yrs. $\pm$ SD)	12.4 $\pm$ 2.5	12.7 $\pm$ 2.5	12.3 $\pm$ 2.6	11.8 $\pm$ 2.4
Min. age (yrs.)	6.5	6.5	7.3	7.3
Max. age (yrs.)	16.9	16.9	16.9	16.6
Sex (M:F)	259:68	164:40	31:10	64:18
Mean IQ (IQ $\pm$ SD)	110 $\pm$ 15 <sup>a</sup>	110 $\pm$ 13 <sup>b</sup>	111 $\pm$ 11 <sup>c</sup>	109 $\pm$ 14 <sup>d</sup>

**Table 1.** Demographic information from entire cohort (n = 327 from ABIDE dataset) and the internal training and validation cohorts and the independent test cohort. <sup>a</sup>Available for n = 308, <sup>b</sup>available for n = 192, <sup>c</sup>available for n = 40, <sup>d</sup>available for n = 76.



**Figure 1.** Age of participants in the entire cohort versus those in the training cohort. This graph highlights the under sampling of the training set based on age. It was important to ensure that the training cohort did not disproportionately represent any one specific age group, just because of the greater frequency of that age group in the full dataset. A ‘flatter’ distribution of age was selected in the training set by under sampling age bands that are ‘overrepresented’ in the overall dataset. This was less successful at the extreme ‘tails’ of childhood (approximately less than 8 years and greater than 16.5 years) where less data was available to sample from. N.B for visualization purposes, this graph is re-binned to bins of width 1 year.

Feature set	n features
Individual morphometric features	
1. Surface area	68 <sup>a</sup>
2. Curvature index	68
3. Folding index	68
4. Gaussian curvature	68
5. Mean curvature	68
6. Cortical thickness	68
7. Cortical volume	68
8. All individual features	476 <sup>b</sup>
Morphometric similarity	
9. Morphometric similarity nodal strength	68
10. Morphometric similarity edge weight	2278 <sup>c</sup>

**Table 2.** Features sets used to produce Brainage models. N.B. <sup>a</sup> Number of regions in the Desikan Killiany atlas, <sup>b</sup> Number of regions time number of individual morphometric features, <sup>c</sup> Number of off-diagonal elements of the connectivity matrix.

**ML algorithm and kernel selection.** Brainage prediction was conducted across two, kernel-based regression approaches; a) Gaussian Processes Regression (GPR) and b) Relevance Vector Regression (RVR). These were selected as both are commonly used in the literature<sup>9, 10, 35</sup>, and non-linear and/or kernel-based algorithms typically outperform linear approaches (likely due to the multicollinearity in morphometric measures<sup>36, 37</sup>). Two different kernels were tested for each algorithm: a) laplacedot (Laplace radial basis kernel) and b) RBFdot (Gaussian radial basis function). Internal training and validation were conducted to select the machine learning algorithm and kernel, based upon performance when trained on the internal training cohort and evaluated on the internal validation set. The hold out test cohort was not included in this process. Figure 2 details this further.

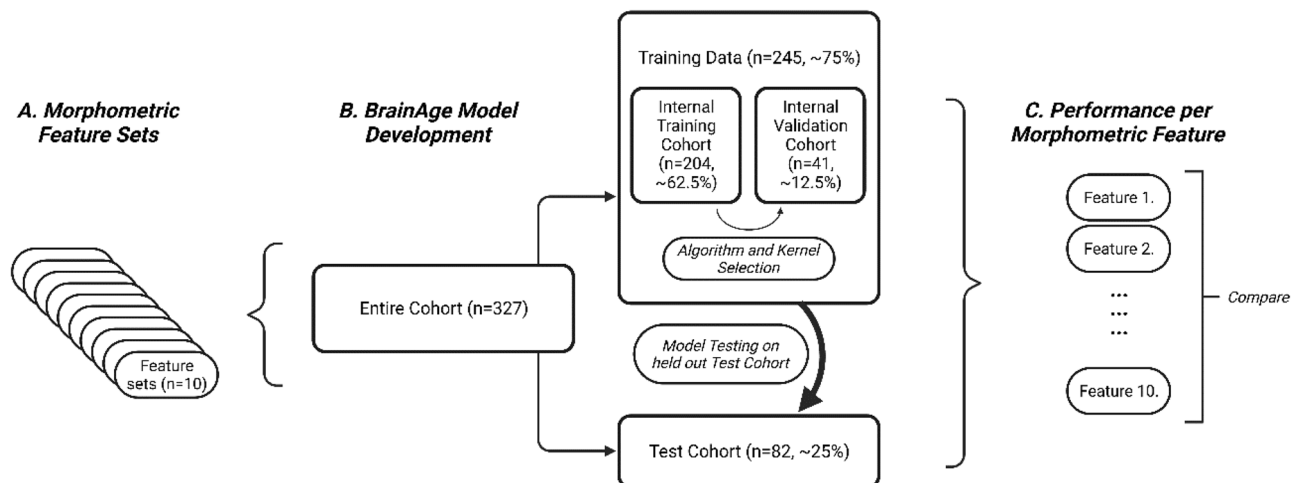
Table 3 highlights the performance of each model in both the internal training and validation sets. For all feature sets, Gaussian processes regression (paired with either the laplacedot or RBFdot) seemed to perform best on the validation set. The model (algorithm + kernel) which performed best on the internal validation set for each feature set was evaluated on the independent test cohort to estimate performance for each feature set.

**Model evaluation on independent test cohort.** Models were trained on the training cohort ( $n = 245$ ) and then evaluated on the independent test cohort ( $n = 82$ ). Table 4 highlights the results of this model testing, with data plotted in Fig. 3. Evaluations suggest that Gaussian and mean curvature performed poorest, with prediction worse than a model of just the mean ( $R^2 = -0.05$  &  $-0.09$  respectively). Morphometric Similarity edge weights, cortical volume, thickness and all individual features performed strongest ( $R^2 = 0.19, 0.29, 0.37$  &  $0.39$  respectively). Based on random resampling of the data (training/testing cohorts), we calculated mean predicted  $R^2$  of models and 95% confidence intervals (CI) of these values. Only models based upon Morphometric Similarity edge weights, cortical volume, thickness and all individual features had 95% CI that did not cross predictive  $R^2 = 0$ . Performance across the resampling for these models was variable, as can be seen in the 95% CI.

Null models were produced by permuting age in the training cohort and evaluating on actual testing data. The mean predictive  $R^2$  values of the resampled models, and the distribution of  $R^2$  values from the permuted 'null' cases allowed calculation of  $p$ -values, where models performed above random noise in the data. Again, only models based upon Morphometric Similarity edge weights, cortical volume, thickness and all individual features produced models which performed significantly above null models.

**Prediction using density thresholded morphometric similarity.** Given that correlation-derived networks may represent both 'real' statistical associations and potential noisy/spurious associations<sup>38</sup> we also tested prediction based upon edge-level Morphometric Similarity, thresholded at an individual-level, at multiple network densities; from top 5% edges to 50% in steps of 5%. For all densities, in terms of both predicted  $R^2$  and MAE, GuassPrc outperformed the RVM algorithm in internal validation procedures. Irrespective of kernel, prediction performed equally well (to 2dp) for all densities tested (All models 5–50% density: MAE = 1.73 yrs, Pred.  $R^2 = 0.32$ ) on the internal validation. Given the predictive accuracy remains constant even when the network is thresholded to enforce greater sparsity, this suggests that the top 5% of edges in terms of weight are those that are most sensitive to individual differences due to age. As the performance did not change compared to the original, unthresholded network, we used the unthresholded network for the presented analyses here, including performance evaluation on the hold out testing cohort.

**Potential biases in Brainage $\Delta$ .** Brainage $\Delta$  was calculated as the absolute difference between chronological (actual) actual and predicted Brainage, in the testing cohort. This indexes the degree to which an individual diverges from age-expected, brain development (combined with model error). As expected, due to the 'healthy' nature of the participants, many of these values were close to zero, although there was large variability (across



**Figure 2.** Flowchart detailing the use of data in the current study for internal validation, training, and testing. Figure created with BioRender.com.

Feature	Algorithm	Kernel	Internal training (n = 204)		Internal validation (n = 41)		Feature	Algorithm	Kernel	Internal training (n = 204)		Internal validation (n = 41)	
			MAE (yrs.)	Pred. R <sup>2</sup>	MAE (yrs.)	Pred. R <sup>2</sup>				MAE (yrs.)	Pred. R <sup>2</sup>	MAE (yrs.)	Pred. R <sup>2</sup>
Surface area	GuassPrc	laplacedot	<b>1.40</b>	<b>0.55</b>	<b>1.96</b>	<b>0.06</b>	Cortical thickness	GuassPrc	laplacedot	<b>1.07</b>	<b>0.72</b>	<b>1.44</b>	<b>0.53</b>
		rbfdot	1.40	0.55	1.96	0.06			rbfdot	1.07	0.72	1.44	0.53
	RVM	laplacedot	1.51	0.45	2.28	-0.19		RVM	laplacedot	1.24	0.64	1.96	0.09
		rbfdot	1.47	0.48	2.23	-0.17			rbfdot	1.20	0.65	1.95	0.10
Curvature index	GuassPrc	laplacedot	<b>1.40</b>	<b>0.57</b>	<b>2.06</b>	<b>0.02</b>	Cortical volume	GuassPrc	laplacedot	<b>1.23</b>	<b>0.64</b>	<b>1.56</b>	<b>0.37</b>
		rbfdot	1.41	0.56	2.06	0.02			rbfdot	1.24	0.63	1.56	0.37
	RVM	laplacedot	1.41	0.49	2.54	-1.20		RVM	laplacedot	1.42	0.52	1.68	0.27
		rbfdot	1.28	0.57	2.59	-1.22			rbfdot	1.44	0.50	1.66	0.28
Folding index	GuassPrc	laplacedot	<b>1.32</b>	<b>0.62</b>	<b>1.97</b>	<b>0.09</b>	All individual features	GuassPrc	laplacedot	1.05	0.74	1.37	0.49
		rbfdot	1.29	0.63	1.97	0.09			rbfdot	<b>1.06</b>	<b>0.73</b>	<b>1.37</b>	<b>0.49</b>
	RVM	laplacedot	1.32	0.52	3.22	-2.05		RVM	laplacedot	1.26	0.60	1.43	0.51
		rbfdot	1.39	0.48	3.02	-1.72			rbfdot	1.24	0.62	1.40	0.53
Gaussian curvature	GuassPrc	laplacedot	<b>1.36</b>	<b>0.60</b>	<b>1.82</b>	<b>0.23</b>	Morphometric similarity: nodal strength	GuassPrc	laplacedot	1.40	0.56	1.77	0.31
		rbfdot	1.34	0.61	1.82	0.23			rbfdot	<b>1.39</b>	<b>0.56</b>	<b>1.77</b>	<b>0.31</b>
	RVM	laplacedot	1.23	0.59	3.85	-3.41		RVM	laplacedot	1.32	0.57	2.01	0.05
		rbfdot	1.25	0.58	3.53	-3.01			rbfdot	1.18	0.64	2.03	0.03
Mean curvature	GuassPrc	laplacedot	<b>1.40</b>	<b>0.56</b>	<b>1.91</b>	<b>0.19</b>	Morphometric similarity: edge weights	GuassPrc	laplacedot	1.20	0.69	1.73	0.32
		rbfdot	1.42	0.55	1.91	0.19			rbfdot	<b>1.20</b>	<b>0.68</b>	<b>1.73</b>	<b>0.32</b>
	RVM	laplacedot	1.44	0.49	2.20	-0.10		RVM	laplacedot	0.36	0.96	1.82	0.31
		rbfdot	1.62	0.37	2.14	-0.04			rbfdot	0.33	0.97	1.82	0.31

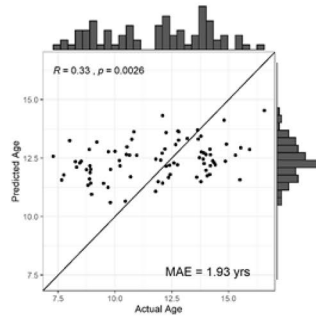
**Table 3.** Performance of Brainage models trained on different feature sets and assessed on Internal Training and Validation cohorts. Performance on internal validation cohort informed the selection of ML algorithm and kernel. Pred. R<sup>2</sup>, Predicted R<sup>2</sup>; GuassPrc, Gaussian processes regression; RVM, Relevance Vector Machine; laplacedot, Laplace radial basis kerne; rbfdot, Gaussian radial basis function. Negative Pred. R<sup>2</sup> values (in red) represent where performance was poorer than prediction using only the mean. Bold indicates for each feature set the combination of algorithm and kernel which produced the most favourable results in the validation set (based on predicted R<sup>2</sup> as the evaluation metric).

Feature	Training cohort (n = 245)		Test cohort (n = 82)				
	MAE (yrs.)	Pred. R <sup>2</sup>	MAE (yrs.)	Pred. R <sup>2</sup>	Mean <sup>a</sup> pred. R <sup>2</sup>	Pred. R <sup>2</sup> (95% CI) <sup>b</sup>	p-value <sup>c</sup>
Surface area	1.41	0.54	1.93	0.04	0.04	(-0.06 to 0.15)	0.029
Curvature index	1.39	0.57	1.91	0.02	-0.03	(-0.16 to 0.08)	0.233
Folding index	1.35	0.60	1.92	0.06	0.04	(-0.11 to 0.15)	0.025
Gaussian curvature	1.36	0.59	2.03	-0.05	0.02	(-0.09 to 0.13)	0.067
Mean curvature	1.38	0.57	1.99	-0.09	0.01	(-0.10 to 0.12)	0.076
Cortical thickness	1.07	0.72	1.48	0.37	0.41	(0.28 to 0.52)	<0.001
Cortical volume	1.21	0.65	1.62	0.29	0.26	(0.13 to 0.38)	<0.001
All individual features	1.03	0.74	1.48	0.39	0.38	(0.27 to 0.47)	<0.001
Morphometric similarity: nodal strength	1.35	0.58	1.94	0.02	0.10	(-0.02 to 0.23)	0.005
Morphometric similarity: edge weights	1.16	0.70	1.73	0.19	0.23	(0.12 to 0.31)	<0.001

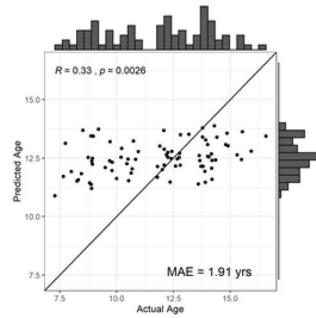
**Table 4.** Performance of Brainage models trained on different feature sets and assessed on Training and Test samples. Pred. R<sup>2</sup>, Predicted R<sup>2</sup>, CI, Confidence Interval. Training sample represents the combination of both training and validation samples. Negative predicted R<sup>2</sup> values (in red) represent where performance was poorer than prediction using only the mean. <sup>a</sup>Mean and <sup>b</sup>Confidence intervals of predictive R<sup>2</sup> values are based upon 100 random partitions (training/ testing cohorts) of the data. <sup>c</sup>p-value derived from 1000 permutations of age at scanning in the full training sample (see methods below, bold = significant at  $\alpha < 0.05/10$ ).

**Individual Morphometric Features**

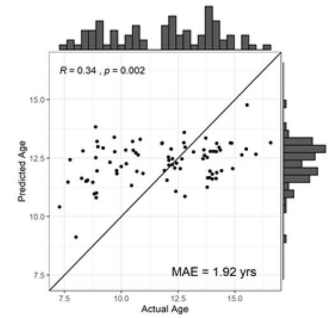
**Surface Area**



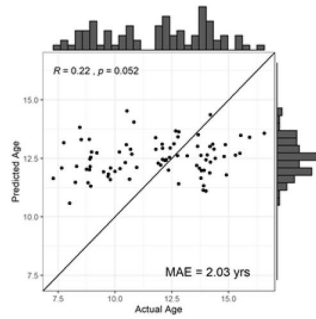
**Curvature Index**



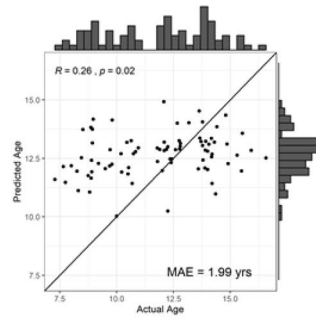
**Folding Index**



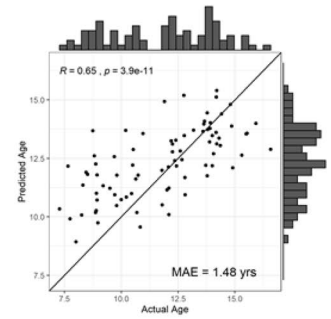
**Gaussian Curvature**



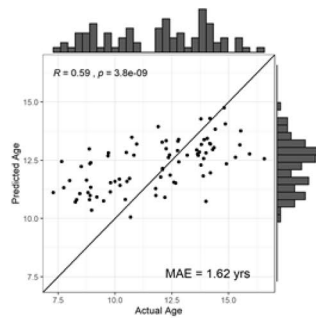
**Mean Curvature**



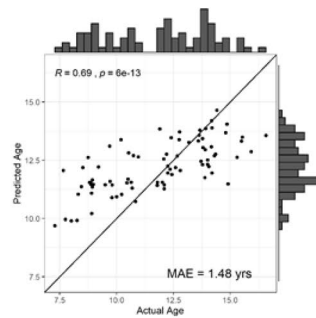
**Cortical Thickness**



**Cortical Volume**

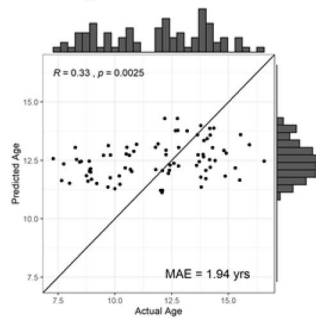


**All Individual Features**

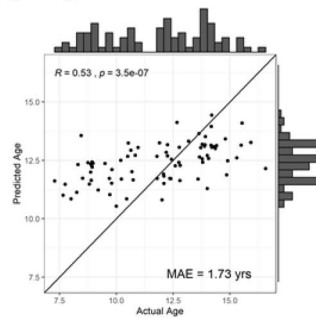


**Morphometric Similarity Networks**

**Nodal Strength**



**Edge Weights**



**Figure 3.** Performance of Brainage prediction on independent testing cohort, for each of the feature sets, including (a) individual morphometric features and (b) network features based on Morphometric Similarity. Chronological age is plotted against the age predicted by the model. Plotted line is where actual age = predicted age ( $x = y$ ), which would represent perfect prediction.

all feature sets; mean (SD) = 0.62(2.11), median = 0.44). Whilst the variation in Brainage $\Delta$  at the group-level was similar across models, Fig. 4 also shows that, at an individual participant level there was large variability in

Brainage $\Delta$  between models. That is to say, a participant with a high Brainage $\Delta$  for cortical thickness, could have a smaller or close-to-zero Brainage $\Delta$  for another morphometric measure. Visual inspection of performance on test data highlighted that across many of the feature sets there was a flatter gradient in the data (actual vs predicted age) compared to the line of  $x = y$  (perfect prediction) suggesting an overestimation of age in younger children and an under estimation of age in the older adolescent participants. This was further seen in Fig. 4 when individual Brainage $\Delta$  profiles (across models) when divided by age group (childhood, early adolescence, middle adolescence). Given age-related bias in previous studies<sup>12, 39, 40</sup>, we controlled for age in the remaining analyses of Brainage $\Delta$  using partial correlations, a correction approach which has been identified to be as effective as correcting the predicted values or correcting the model for these biases<sup>41</sup>.

#### (A) Potential Biases in Brainage $\Delta$ : Sex.

Potential sex differences in Brainage $\Delta$  estimation were investigated using linear models controlling for actual age. Across all models, the effect of sex did not meet significance.

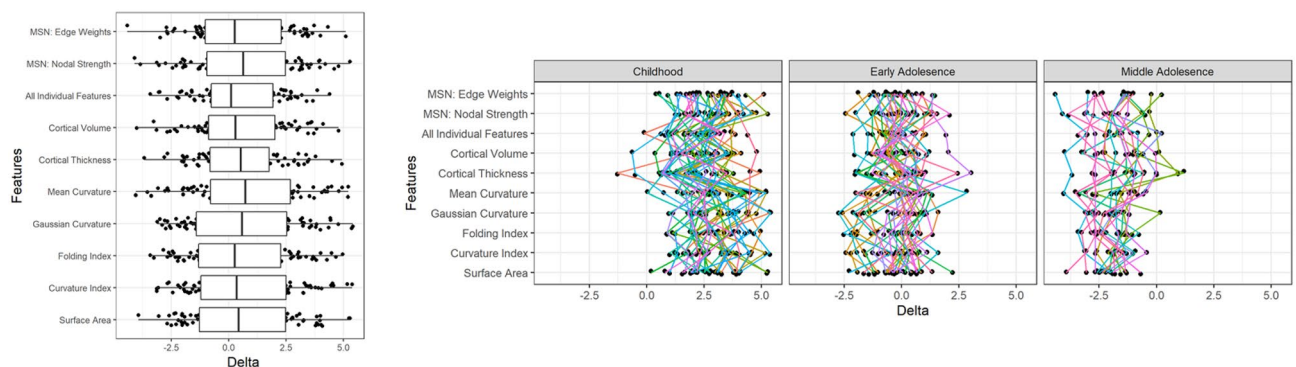
#### (B) Potential Biases in Brainage $\Delta$ : Motion.

To evaluate potential bias in the models from motion we used the Entropy Focus Criterion (EFC<sup>42</sup>) as a proxy for motion derivable from T1w images. EFC uses the Shannon entropy of voxel intensities to typically quantify the amount of motion present<sup>43</sup>, specifically through the sensitivity to motion-induced artifacts (e.g., ghosting and blurring induced by head motion). MRI autofocusing techniques based on EFC optimisation have been shown to reduce motion artifacts effectively<sup>42</sup>.

Average correlation between Brainage $\Delta$  and EFC across the models was close to zero ( $\bar{r} = -0.03$ ), with no correlation reaching statistical significance (Table 5).

**Exploratory relationship with cognition.** To evaluate Brainage $\Delta$  as a putative measure of meaningful variation due to individual differences, we investigated relationships between Brainage $\Delta$  and individual IQ using partial correlations (using actual age as a confound to address age-bias in the Brainage $\Delta$  measure). A limited number ( $n = 56$ ) of children in the test sample had valid measures of IQ. Neither morphometric similarity derived Brainage $\Delta$  or that calculated from models of individual features correlated with IQ. Whilst still non-significant, the greatest effect was found with surface area.

**Combining models for Brainage prediction.** Exploratory analyses were conducted to combine the feature sets from the best performing Brainage models to investigate whether models provided incremental increase in Brainage prediction by predicting unique variance in age. Combining cortical thickness, cortical volume and morphometric similarity edge weights, as the best performing individual features using Gaussian processes regression (with rbf kernel), training on the training cohort resulted in comparable performance to the best performing models seen in Table 4 (MAE = 1.06 yrs., Pred  $R^2 = 0.74$ ). On the independent testing cohort, performance dropped significantly (MAE = 1.59 yrs, Pred.  $R^2 = 0.31$ ), performing better than the Cortical volume and morphometric similarity weight models but still outperformed by the cortical thickness model. The Brainage $\Delta$  estimates from this model (thickness + volume + morphometric similarity edge weights) were still biased by age (Pearson's  $r = -0.93$ ,  $p < 0.00001$ ), with no discernible relationship to individual differences in cognition (Pearson's  $r = -0.17$ ,  $p = 0.23$ ).



**Figure 4.** Plots showing Brainage $\Delta$  for the testing cohort; across each model based on different feature sets (left) and when divided into developmental periods of childhood (5–11 years), early adolescence (11–14 yrs) and middle adolescence (14–17 years).

Covariate	n	Feature set	Pearson		Spearman Rho	
			R	p	r	p
Motion <sup>a</sup>	82	Surface area	-0.05	0.65	-0.16	0.16
		Curvature index	-0.04	0.74	0.01	0.93
		Folding index	0.06	0.57	0.04	0.71
		Gaussian curvature	0.03	0.76	0.05	0.67
		Mean curvature	0.01	0.94	0.04	0.75
		Cortical thickness	-0.06	0.62	-0.08	0.48
		Cortical volume	-0.23	0.04	-0.06	0.61
		All individual features	-0.07	0.51	-0.11	0.31
		Morphometric similarity: nodal strength <sup>b</sup>	0.04	0.73	-0.08	0.47
		morphometric similarity: edge weights <sup>b</sup>	-0.03	0.77	-0.06	0.61
IQ <sup>a</sup>	56	Surface area	0.23	0.08	0.24	0.08
		Curvature index	0.10	0.47	0.09	0.50
		Folding index	-0.05	0.72	-0.06	0.66
		Gaussian curvature	0.03	0.83	-0.07	0.60
		Mean curvature	-0.04	0.76	0.02	0.89
		Cortical thickness	-0.04	0.77	0.00	0.99
		Cortical volume	0.00	0.99	0.08	0.57
		All Individual features	-0.14	0.30	-0.17	0.23
		Morphometric similarity: nodal strength <sup>b</sup>	0.05	0.69	0.07	0.62
		Morphometric similarity: edge weights <sup>b</sup>	-0.06	0.65	0.05	0.71
Covariate	n	Feature set	Statistic	P		
Sex <sup>a</sup>	82	Surface area	-1.46	0.15		
		Curvature index	1.53	0.13		
		Folding index	0.33	0.74		
		Gaussian curvature	0.14	0.89		
		Mean curvature	1.25	0.22		
		Cortical thickness	-0.56	0.57		
		Cortical volume	1.64	0.10		
		All individual features	-0.16	0.88		
		Morphometric similarity: nodal strength <sup>b</sup>	-0.39	0.70		
		Morphometric similarity: edge weights <sup>b</sup>	-0.32	0.75		

**Table 5.** Statistical associations between Brainage $\Delta$  and covariates for each feature set. <sup>a</sup>Controlling for actual age, <sup>b</sup> derived from unthresholded Morphometric Similarity, Bold indicates those tests significant at Bonferroni corrected  $\alpha$ -level = .000833.

## Discussion

To our knowledge, this is the first study to construct Brainage models derived from network-level descriptions of neuroanatomical organization across the cortex. These models using morphometric similarity as a basis for predicting chronological age did not outperform non-network models, using 'standard' morphometric features.

Specifically, the morphometric similarity model was outperformed (in terms of lowest MAE and highest predicted  $R^2$ ) by models which included all individual structural features, followed by cortical thickness and volumetric models. The morphometric similarity edge weight model did, however, perform significantly better than null models on testing data, suggesting that these Brainage models are capturing 'real' patterns of variation indicative of age. The fact it was outperformed by a model where all individual features were entered separately indicated that the morphometric similarity model does not necessarily capture additional variation. One could argue that the morphometric similarity nodal-level model represents a more efficient model compared to the all-features model (given the smaller size of the feature set)—but given the 68-feature cortical thickness model also outperformed the morphometric similarity model, it would not be a more effective data reduction approach either.

The best performing (individual) structural feature for age prediction in this study was cortical thickness. Conversely, in a previous report of lifespan (8–96 yrs) Brainage prediction, in the 8–18 yr old group, across all approaches using either cortical area, thickness or volume, the greatest performance (i.e. lowest mean prediction error) was actually seen using brain volume model<sup>44</sup>. However, across the six prediction techniques investigated in<sup>44</sup>, cortical thickness models outperformed cortical volume models in 3/6 methods. This similar performance is maybe unsurprising given that cortical thickness and surface area both independently contribute to volume measurements<sup>45</sup> and that volume measurements can be estimated from the product of cortical thickness and surface area measurements at all locations across the cortical mantle (in surface-based approaches)<sup>46</sup>. The findings



of this analysis, alongside previous reports<sup>20, 21, 25</sup>, highlight the sensitivity and importance of cortical thickness in late childhood to adolescent development compared to surface area and curvature-based cortical measures or even novel methods of morphometric similarity.

All other tested structural features (Surface Area, Curvature Index, Folding Index, Gaussian Curvature, Mean Curvature) did not significantly outperform null models. This suggests that these measures are less sensitive to developmental changes within the window of late childhood to adolescence (6–17 yrs). For example, surface area and gyrification index measures may be more relevant to the developmental changes found from the third trimester to the early post-natal period as evidenced in imaging of term and preterm infants<sup>47</sup>. Certainly gross (rather than ROI) structural measures highlight that differing rates of change (and peak velocity of change) vary across developmental periods, with different measures being potentially more discriminatory over these periods<sup>48</sup>. Therefore, current results would highlight that these additional features are less important during this developmental window.

We also found that combining best performing models (cortical thickness, volume, and morphometric similarity edge weights) resulted in a drop in performance compared to the cortical thickness model. Whilst not a direct statistical comparison, this suggests that these models do not capture independent variance in relation to age. This seems to disagree with previous work<sup>20</sup> which found that joint covariation across multiple structural features predicted variance in age independently from variance in individual features. It is unclear to the degree that the joint and distinct variation features used in<sup>20</sup> are directly comparable to representations learnt by the machine learning techniques in this study, or even how they relate to the morphometric similarity approach outlined here and in<sup>29</sup>. Comparison of these different approaches to data ‘fusion’ across morphometric measures will be required to reconcile the potential differences between these study findings.

As well as feature sets affecting Brainage estimation, the machine learning or prediction workflow is also a key factor. This study found GPR to outperform the RVR approach. These methods were selected as they have been shown to outperform other linear approaches<sup>36</sup>, including in paediatrics<sup>37</sup>. On the surface, our finding seems to contradict other, comparative analyses of machine learning models in predicting Brainage using morphometric data who found RVR to systematically outperform GPR<sup>49</sup>. However, the one scenario in which GPR did outperform RVR in<sup>49</sup>, was in the test case with the smallest number of participants, closer to that of the sample size used here. Therefore, machine learning model will be an important consideration for future use cases.

Currently, only two other study predicted Brainage from sMRI in the ABIDE cohort<sup>23, 50</sup>. Using a complex network approach to T1w MRI, in 7–20 yr olds, Bellantuono et al.<sup>50</sup> achieved a MAE of 1.53 years using deep learning models. The slightly larger age range means that the MAE are not entirely comparable with the current study, although the present study has outperformed this. It is important to note that the network approach to T1w MRI in this study modelled correlation grey-levels of the image rather than structural metrics.

When Brainage $\Delta$  was calculated for the test cohort, there was great variability in of an individual’s delta values for each of the feature sets; there appeared to be little consistency in these values between models. The varying individual profiles of Brainage $\Delta$  has two possible explanations. Firstly, Brainage $\Delta$  represents the combined measure of individual variance from the expected developmental trajectory plus the error in the normative age model. It therefore may be the case that the random error in each of the models is resulting in variance in Brainage $\Delta$ , across feature sets, at the individual level. This could have potential implications for the comparison of studies utilizing the Brainage measure if there is limited consistency in these measures within an individual participant. Alternatively, a potentially more interesting explanation, is that each Brainage model is indexing relevant divergences/individual differences in different aspects of cortical architecture, resulting in between model variance in Brainage $\Delta$ . This could prove to be useful in neurological conditions that influence difference aspects of brain development/organization in the paediatric brain, for instance a Brainage model based upon MRI measure of white matter may be more sensitive to differences from normative brain development in acute demyelinating disorders such as multiple sclerosis. In this scenario, multiple Brainage $\Delta$ ’s from different features, or even imaging modalities could be used, as potential biomarkers of clinically relevant outcomes.

However, it is difficult to statistically test each of these explanations (model error vs meaningfully different divergences) because there are a limited number of models used in any one study. Future meta-analytic research could compare within-participant Brainage $\Delta$  values across feature sets, whilst controlling for the MAE of the model themselves, in order to isolate ‘real’ within-individual variation in the Brainage $\Delta$  measure. Future studies could also use multiple (even multi-modal) Brainage models and use the feature specific Brainage $\Delta$ ’s as individual predictors in regression models, to assess unique predictive variance offered by each feature.

A strength of the current study was the extensive assessments of the morphometric similarity model, in the context of the Brainage framework across multiple analyses;

- (a) we tested against individual structural features, in a held-out testing cohort,
- (b) we assessed robustness of performance in terms of sampling (assessing the 95% CI of performance) and against meaningful null models and,
- (c) we investigated correlations between Brainage $\Delta$  and biases/cognition in the independent testing sample.

As noted by<sup>50</sup>, ABIDE is also a particularly challenging dataset for the estimation of Brainage, due to the number of different sites and acquisition protocols. For future Brainage studies of development, this high bar should at least be maintained, with future improvements seen by validating on an entirely independent dataset (for example as seen in<sup>20</sup>).

An outstanding question for future research is whether there is need for models such as morphometric similarity as the popularity for deep learning/machine learning approaches become more prevalent. Fisch et al.<sup>51</sup> report the results of the Predictive Analytic Competition (2019) for predicting chronological age from structural

neuroimaging. They highlight the high-performing nature of neural networks for deep machine learning within the Brainage framework. Morphometric Similarity models the covariance structure of anatomical MRI features in a way which is constrained by anatomy (either using ROIs or voxels for instance) typically using a very specific, linear approach to these covariances/similarity (Pearson's correlation coefficients). The morphometric similarity model has been shown to capture biologically meaningful information<sup>29</sup> however, imposing such a model as an anatomical-prior may be redundant in analysing larger sample sizes with machine learning approaches. The machine learning/deep learning approaches that are becoming more popular in the neuroimaging literature, when fed all the individual features which are used to construct the morphometric similarity network (as we have done here), should be able to recover any covariance between structural features (even beyond linear relationships) that is captured by the morphometric similarity network approach. This may be supported by the results reported here, with greater performance seen for a model using all features compared to the morphometric similarity models.

One way in which studies have tried to identify the functional relevance of Brainage $\Delta$  is through associations with outcomes such as cognition. However, we found no relationship between Brainage as a measure of individual-difference and cognition in this typically developing cohort. Interestingly, more accurate Brainage models did not hold any greater associations with cognition. These results suggest that, when these models are generalized to 'novel' cases (in this situation the testing cohort), the resultant Brainage $\Delta$  measures do not hold information pertinent to individual differences in cognition. Ball et al.<sup>52</sup> also reported no significant relationship between individual-level Brainage $\Delta$  (derived from voxel-based cortical thickness, volume and surface area) and cognitive abilities (as measured by the NIH Toolbox Cognition Battery). They hypothesized that this may be due to the methods they utilised which maximized the captured age-related variance in neuroanatomical measures, and that cognition-related variance (non-age related) may be captured by a different, orthogonal pattern of neuroanatomical correlates. However other studies have also found no convincing relationship between Brainage and cognition in typical developing children<sup>25,53</sup>. Of those that did find a relationship in developing cohorts<sup>54,55</sup>, these associations were small to moderate in size and thus likely require large sample sizes to reliably detect<sup>53</sup>. Finding neurodevelopmental outcomes for which this approach offers meaningful insight is key in providing functional utility of this approach as a relevant biomarker, but it seems that cognition is an unlikely candidate for this. It is perhaps unsurprising though, given that these models have been optimized for more accurate predictions of age, rather than optimized to capture phenotypes of interest<sup>56</sup>. A more appropriate approach for future studies, may be to establish models that directly capture variance in cognitive ability, rather than capturing an indirect biomarker<sup>52,56</sup>.

Whilst the current study investigated morphometric similarity in the Brainage framework as an indirect marker of cognition, the current finding also adds to an increasing literature which questions the direct association between the meso-scale organization of morphometry across the cortex (as captured by morphometric similarity) and measures of cognitive ability. Outside of the Brainage framework, we also found no relationship between these measures and cognitive abilities<sup>24</sup>, failing to replicate the findings of<sup>29</sup>. However, a recent study of adolescence has highlighted the predictive validity of morphometric similarity across cognition/intelligence and psychiatric symptoms<sup>57</sup>, and so this is still very much an open area of research.

## Conclusion

Overall, whilst these network models of sMRI, using morphometric similarity, seem to mature as a function of age in typical neurodevelopment<sup>29</sup>, and capture meaningful variation indicative of chronological age in the Brainage framework, these networks are not most sensitive to the changes across childhood compared to other, more simplistic features, for instance cortical thickness measures.

## Methods

**Materials and data availability.** The data used in this research was acquired through the public Autism Brain Imaging Data Exchange (ABIDE, Di Martino, Yan<sup>31</sup>) database. Specifically, we used the ABIDE data release as shared by the Preprocessed Connectome Project (PCP, Bellec, Yan<sup>32</sup>). For full details and access see Pre-processed Connectome Project website <http://preprocessed-connectomes-project.org/>. Results and meta-data of the current study are available on request from Dr Griffiths-King. The R code is also available from the authors upon request, however all open-source packages used in the study are listed here: *data.table*, *scales*, *psych*, *ggplot2*, *neuroCombat*, *ggseg*, *dplyr*, *ggpubr*, *ggExtra*, *kernelab*, *ppcor*, *PupillometryR*, *tidyr*.

**Ethics statement.** The database has de-identified all the patient health information associated with the data. A favourable ethical opinion was granted by Aston University Research Ethics Committee (UREC) for the secondary analysis of the ABIDE datasets (no. 1309).

**Participants.** The ABIDE dataset consists of a large sample of 532 individuals with autism spectrum disorders and 573 typical controls, composed of MRI (functional and structural) and phenotypic information for each subject, accumulated across 17 independent neuroimaging sites. The scan procedures and parameters are described in more detail on the ABIDE website (<http://fcon-1000.projects.nitrc.org/indi/abide/>). We applied four inclusion criteria to this dataset, only including subjects who; a) passed a strict MRI quality control of raw structural MRI (see below), b) were recorded as controls within the ABIDE database, c) at time of scan were aged < 17 years old and d) had pre-processed Freesurfer data available as part of the PCP data release. This resulted in a total cohort of  $n = 327$ . Group demographics can be seen in Table 1. The ABIDE cohort had a mean IQ of approximately 110, as measured across multiple age-appropriate IQ tests (See ABIDE documentation for details).

**Data quality check.** The PCP data release includes image quality metrics (IQMs) which provide quantitative ratings of the quality of the raw T1-weighted (T1w) MR images. These are calculated using the Quality Assessment Protocol software (QAP, Shehzad, Giavasis<sup>58</sup>). The ABIDE dataset includes data from 17 recruitment sites, and such there is potential for ‘batch effects’ on QA metrics<sup>43</sup>. We used the six spatial anatomical QA measures. Hence, all QA metrics were centred (mean subtracted) and scaled (divided by standard deviation) within sites, then recoded to increased values representing greater quality. This results in metrics which can be compared between sites. For each subject, QA metrics were coded as failed if they had a Z score below  $-1.5$  (indicating quality which was 1.5SD below the mean). We included subjects if they had zero or one QA metric that fell below this quality metric. Of the ABIDE cases who were recorded as a) controls and b) being younger than 17 years of age at scanning ( $n = 361$ ), 14 subjects were removed due to having greater than one QA metric fall below the 1.5SD cut off (20 participants also had no Freesurfer data available, resulting in the final ABIDE dataset of  $n = 327$ ). Further details of the automated QA measures which are included can be found here: [http://preprocessed-connectomes-project.org/abide/quality\\_assessment.html](http://preprocessed-connectomes-project.org/abide/quality_assessment.html) and <http://preprocessed-connectomes-project.org/quality-assessment-protocol>.

**Structural MRI processing with freesurfer.** 3D tissue segmentation and estimation of morphometric features from T1w MR images was conducted using an established pipeline (Freesurfer version 5.1; details are published elsewhere Fischl, van der Kouwe<sup>59</sup>, see Fischl<sup>33</sup> for review). Briefly, T1w images were stripped of non-brain tissues<sup>60</sup>, GM/WM boundaries were tessellated and topology was automatically corrected<sup>61, 62</sup>. Finally, deformation of this surface was performed, to optimally define the pial (Cerebro-spinal fluid/GM) and white (GM/WM) surfaces using maximum shifts in intensity gradients to define boundaries of these tissue classes<sup>63–65</sup>. Morphometric measures were derived from these surface-based models using standard methods from the Free-surfer recon-all pipeline<sup>33</sup>.

**Data harmonization.** Multi-site imaging data harmonization was conducted using the neuroComBat package<sup>66, 67</sup>, an R implementation of the ComBat method<sup>68</sup> for removing batch-effects (i.e. site-effects) in neuroimaging data. This was applied to the participant by ROI matrix for each morphometric feature individually, to remove site effects found in the ABIDE data, whilst protecting biological variation due to age. Fortin and colleagues have shown this approach to be effective in removing site effects in multi-site imaging data even when the biological covariate of interest (in this case age) is not balanced across sites<sup>66, 67</sup>. These site-corrected morphometric measures were used for a) estimation of morphometric similarity networks and b) for the individual feature models.

**Estimating morphometric similarity.** Previously morphometric similarity was estimated from morphometric features measured in-vivo by both structural and diffusion MRI<sup>29</sup>. However, we highlighted significant correspondence between this multimodal morphometric similarity and that estimated with only features obtainable from a T1w MRI<sup>24</sup> and recent papers have similarly adopted this T1w-only approach<sup>69</sup>, as does the current study.

To estimate morphometric similarity, the nodes for network construction were the ROIs from the Desikan-Killiany atlas<sup>34</sup>. At an individual-level, the seven morphometric features estimated for each node can be expressed as a set of  $n$  vectors of length 7, with each vector as a different anatomical region ( $n = 68$ ), and each element of the vector a different morphometric measure. To normalize measures within this length 7 vector, each morphometric feature is demeaned, and SD scaled across the 68 regions, using Z-scores. A correlation matrix was generated for each participant, where each element of the matrix is the correlation between the feature vectors for every possible pairwise combination of regions. This correlation matrix represents the morphometric similarity derived meso-scale cortical organisation for each participant. This was an unthresholded matrix.

For each node/ROI, we calculated both nodal degree and nodal strength. Nodal degree was the number of edges that had survived thresholding for each node. Normalised nodal strength was calculated as the ‘magnitude’ of morphometric similarity for each node. This is defined as the sum of the morphometric similarity weights of all of the edges of node  $i$ <sup>70</sup>, normalised by the degree of the node (nodes with a higher number of edges will by definition have a greater magnitude of morphometric similarity). We also calculated the average nodal strength across the network to provide a global measure of the magnitude of morphometric similarity.

In subsequent exploratory analyses we investigated the thresholded matrix across multiple network densities (threshold  $x = 5–40$  in increments of 5), retaining only  $x\%$  strongest absolute values of morphometric similarity across the graph. This has the effect of removing potential false-positive ‘edges’ of morphometric similarity. Metrics were calculated as per the unthresholded matrix.

**Sampling for training, validation and testing samples.** The ABIDE cohort were divided into training, and test samples in the ratio of 6:2 ( $n = 245$ , &  $n = 82$  respectively). Sampling for the training sample was selected pseudo-randomly, via stratified under sampling based upon age. The entire sample was binned into 0.5 yr bins dependent on age at scanning, up to the cut-off criteria of 17 years. Bins for ages 6–9 yrs were collapsed due to the much lower participant numbers in this lower tail of the distribution. Participants were equally sampled from each bin to derive the final training sample size. The training cohort was further subdivided into an internal training and validation cohort at a ratio of 5:1 ( $n = 204$  &  $n = 41$ ).

**Brainage prediction models.** Brainage prediction was conducted across two, kernel-based regression approaches using the Kernlab package in R<sup>71</sup>; a) Gaussian Processes Regression (GPR) and b) Relevance Vector

Regression (RVR). These were selected as these are both commonly used in the Brainage literature<sup>9,10</sup>. Two different kernels were tested for each algorithm: a) laplacedot (Laplace radial basis kernel) and b) RBFdot (Gaussian radial basis function). Algorithm and kernel selection was conducted based upon performance on the internal validation set.

A GPR/RVR model was defined, with chronological age as the dependent variable and the morphometric data (for each of the feature sets) as the independent variables, to build a model of ‘healthy’ structural brain development. Prediction utilized 10 different feature-sets; i–vii) each of the individual morphometric features, viii) all features, ix) nodal-level strength of the morphometric similarity graph and x) edge-level weights of the morphometric similarity graph. Final model evaluation was conducted based upon performance on the independent test cohort.

In all cases, performance was evaluated based upon reducing mean absolute error (MAE) and maximizing predictive  $R^2$ . Standard linear regression  $R^2$  is a biased estimate of model performance especially at lower performances<sup>72</sup>, whereas predicted  $R^2$  is more appropriate for quantifying regression accuracy<sup>73</sup>, calculated as;

$$\text{Predicted } R^2 = 1 - \text{Normalised MSE} \quad (1)$$

where the normalised MSE (Mean Squared Error) can be expanded to;

$$\text{Predicted } R^2 = 1 - \frac{\text{MSE (Predicted Value, Observed Value)}}{\text{MSE (Observed Value, Mean Value)}} \quad (2)$$

## Robustness of brainage models

**Robustness to sampling.** To assess robustness of models to the sampling partitions of the data, mean and confidence intervals of predictive  $R^2$  values are calculated. We carried out 100 random partitions (training/testing cohort) of the data and repeated analyses to generate a vector of 100 predictive  $R^2$  values for the testing set from which we can take a mean metric and assess the 95% confidence interval (see Table 4).

**NHST of models.** To assess the ‘real effect’ of models in comparison to ‘null’ models, we used permutation testing to conduct null hypothesis significance testing (NHST). We established the null hypothesis as no meaningful patterns in the data between age and feature sets in training, and thus poor performance on the test cohort. To derive such models, we permuted ( $n = 1000$ ) the dependent variable of age in the training cohort and reran the models. These were then evaluated on the testing cohort where the true actual age was used and the ‘null’  $R^2$  was calculated. The mean predictive  $R^2$  values of the resampled models (above), and the distribution of  $R^2$  values from the permuted ‘null’ cases allowed calculation of  $p$ -values where the frequency of instances in the distribution where the mean predictive  $R^2$  was greater than that of the null models. Significance of  $p$ -values was assessed at the level of Bonferroni corrected  $\alpha < 0.005$ , corrected over the 10 models.

## Data availability

The data used in this research was acquired through the public Autism Brain Imaging Data Exchange (ABIDE, Di Martino, Yan<sup>31</sup>) database. Specifically, we used the ABIDE data release as shared by the Preprocessed Connectome Project (PCP, Bellec, Yan<sup>32</sup>). For full details see Pre-processed Connectome Project website <http://preprocessed-connectomes-project.org/>. The R code is also available from the authors upon request, however all open-source packages used in the study are listed here: *data.table*, *scales*, *psych*, *ggplot2*, *neuroCombat*, *ggseg*, *dplyr*, *ggpubr*, *ggExtra*, *kernlab*, *ppcor*, *PupillometryR*, *tidyr*.

Received: 13 February 2023; Accepted: 10 September 2023

Published online: 20 September 2023

## References

- Giedd, J. N. & Rapoport, J. L. Structural MRI of pediatric brain development: What have we learned and where are we going?. *Neuron* **67**(5), 728–734 (2010).
- Giedd, J. N. Structural magnetic resonance imaging of the adolescent brain. *Ann. N. Y. Acad. Sci.* **1021**, 77–85 (2004).
- Gilmore, J. H. *et al.* Regional gray matter growth, sexual dimorphism, and cerebral asymmetry in the neonatal brain. *J. Neurosci.* **27**(6), 1255–1260 (2007).
- Knickmeyer, R. C. *et al.* A structural MRI study of human brain development from birth to 2 years. *J. Neurosci.* **28**(47), 12176–12182 (2008).
- Mills, K. L. *et al.* Structural brain development between childhood and adulthood: Convergence across four longitudinal samples. *Neuroimage* **141**, 273–281 (2016).
- Herting, M. M. *et al.* A longitudinal study: Changes in cortical thickness and surface area during pubertal maturation. *Plos One* **10**(3), e0119774 (2015).
- Nie, J. *et al.* Longitudinal development of cortical thickness, folding, and fiber density networks in the first 2 years of life. *Hum. Brain Mapp.* **35**(8), 3726–3737 (2014).
- Whitaker, K. J. *et al.* Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proc. Natl. Acad. Sci. U. S. A.* **113**(32), 9105–9110 (2016).
- Cole, J. H. & Franke, K. Predicting age using neuroimaging: Innovative brain ageing biomarkers. *Trends Neurosci.* **40**(12), 681–690 (2017).
- Cole, J. H., Franke, K. & Cherbuin, N. *Quantification of the Biological Age of the Brain Using Neuroimaging*. *Osfpreprints*, (2018).
- Cole, J. H. *et al.* Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Ann. Neurol.* **77**(4), 571–581 (2015).
- Beheshti, I. *et al.* Bias-adjustment in neuroimaging-based brain age frameworks: A robust scheme. *J. NeuroImage Clin.* **24**, 102063 (2019).

13. Bataille, D., Edwards, A. D. & O’Muircheartaigh, J. Annual research review: Not just a small adult brain: Understanding later neurodevelopment through imaging the neonatal brain. *J. Child Psychol. Psychiatry* **59**(4), 350–371 (2018).
14. Raznahan, A. *et al.* How does your cortex grow?. *J. Neurosci.* **31**(19), 7174–7177 (2011).
15. Shaw, P. *et al.* Neurodevelopmental trajectories of the human cerebral cortex. *J. Neurosci.* **28**(14), 3586–3594 (2008).
16. Schmitt, J. E. *et al.* A multivariate analysis of neuroanatomic relationships in a genetically informative pediatric sample. *Neuroimage* **35**(1), 70–82 (2007).
17. Teeuw, J. *et al.* Genetic influences on the development of cerebral cortical thickness during childhood and adolescence in a dutch longitudinal twin sample: The brainscale study. *Cereb. Cortex* **29**(3), 978–993 (2019).
18. Wallace, G. L. *et al.* A pediatric twin study of brain morphometry. *J Child Psychol. Psychiatry* **47**(10), 987–993 (2006).
19. Fjell, A. M. *et al.* Development and aging of cortical thickness correspond to genetic organization patterns. *Proc. Natl. Acad. Sci. U. S. A.* **112**(50), 15462–15467 (2015).
20. Zhao, Y. *et al.* Brain age prediction: Cortical and subcortical shape covariation in the developing human brain. *Neuroimage* **202**, 116149 (2019).
21. Dosenbach, N. U. *et al.* Prediction of individual brain maturity using fMRI. *Science* **329**(5997), 1358–1361 (2010).
22. Bullmore, E. & Sporns, O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**(3), 186–198 (2009).
23. Corps, J. & Reikik, I. Morphological brain age prediction using multi-view brain networks derived from cortical morphology in healthy and disordered participants. *Sci. Rep.* **9**(1), 9676 (2019).
24. King, D. J. & Wood, A. G. Clinically feasible brain morphometric similarity network construction approaches with restricted magnetic resonance imaging acquisitions. *Netw. Neurosci.* **4**(1), 274–291 (2020).
25. Khundrakpam, B. S. *et al.* Prediction of brain maturity based on cortical thickness at different spatial resolutions. *Neuroimage* **111**, 350–359 (2015).
26. Franke, K. *et al.* Brain maturation: Predicting individual BrainAGE in children and adolescents using structural MRI. *Neuroimage* **63**(3), 1305–1312 (2012).
27. Ball, G., Beare, R. & Seal, M. L. Charting shared developmental trajectories of cortical thickness and structural connectivity in childhood and adolescence. *Hum. Brain Mapp.* **40**(16), 4630–4644 (2019).
28. Brown, T. T. *et al.* Neuroanatomical assessment of biological maturity. *Curr. Biol.* **22**(18), 1693–1698 (2012).
29. Seidlitz, J. *et al.* Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation. *Neuron* **97**(1), 231–247 (2018).
30. Seidlitz, J. *et al.* Transcriptomic and cellular decoding of regional brain vulnerability to neurodevelopmental disorders. *bioRxiv* **2**, 406 (2019).
31. Di Martino, A. *et al.* The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* **19**(6), 659–667 (2014).
32. Cameron, C. *et al.* The neuro Bureau preprocessing Initiative: Open sharing of preprocessed neuroimaging data and derivatives. *Front. Neuroinform.* <https://doi.org/10.3389/conf.fninf.2013.09.00041> (2013).
33. Fischl, B. FreeSurfer. *Neuroimage* **62**(2), 774–781 (2012).
34. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**(3), 968–980 (2006).
35. Baecker, L. *et al.* Brain age prediction: A comparison between machine learning models using region- and voxel-based morphometric data. *Hum. Brain Mapp.* **42**(8), 2332–2346 (2021).
36. More, S. *et al.* Brain-age prediction: A systematic comparison of machine learning workflows. *bioRxiv* **42**, 2332 (2022).
37. Modabbernia, A. *et al.* Systematic evaluation of machine learning algorithms for neuroanatomically-based age prediction in youth. *Hum. Brain Mapp.* **43**(17), 5126–5140 (2022).
38. Fornito, A., Zalesky, A. & Breakspear, M. Graph analysis of the human connectome: Promise, progress, and pitfalls. *Neuroimage* **80**, 426–444 (2013).
39. Liang, H., Zhang, F. & Niu, X. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. *Hum. Brain Mapp.* **40**, 3143–3152 (2019).
40. Zhang, B. *et al.* Age-level bias correction in brain age prediction. *NeuroImage Clin.* **37**, 103319 (2023).
41. de Lange, A. -M. G. & Cole, J. H. Commentary: Correction procedures in brain-age prediction. *J. NeuroImage Clin.* **26**. (2020).
42. Atkinson, D. *et al.* Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. *IEEE Trans. Med. Imaging* **16**(6), 903–910 (1997).
43. Esteban, O. *et al.* MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE* **12**(9), e0184661 (2017).
44. Valizadeh, S. A. *et al.* Age prediction on the basis of brain anatomical measures. *Hum. Brain Mapp.* **38**(2), 997–1008 (2017).
45. Winkler, A. M. *et al.* Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage* **53**(3), 1135–1146 (2010).
46. Winkler, A. M. *et al.* Joint analysis of cortical area and thickness as a replacement for the analysis of the volume of the cerebral cortex. *Cereb. Cortex* **28**(2), 738–749 (2018).
47. Shimony, J. S. *et al.* Comparison of cortical folding measures for evaluation of developing human brain. *Neuroimage* **125**, 780–790 (2016).
48. Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604**(7906), 525–533 (2022).
49. Han, J. *et al.* Brain age prediction: A comparison between machine learning models using brain morphometric data. *Sensors (Basel)* **22**(20), 8077 (2022).
50. Bellantuono, L. *et al.* Predicting brain age with complex networks: From adolescence to adulthood. *Neuroimage* **225**, 117458 (2021).
51. Fisch, L. *et al.* Editorial: Predicting chronological age from structural neuroimaging: The predictive analytics competition 2019. *Front. Psychiatry* **12**, 710932 (2021).
52. Ball, G. *et al.* Modelling neuroanatomical variation during childhood and adolescence with neighbourhood-preserving embedding. *Sci. Rep.* **7**(1), 17796 (2017).
53. Ball, G. *et al.* Individual variation underlying brain age estimates in typical development. *Neuroimage* **235**, 118036 (2021).
54. Erus, G. *et al.* Imaging patterns of brain development and their relationship to cognition. *Cereb. Cortex* **25**(6), 1676–1684 (2015).
55. Lewis, J. D. *et al.* T1 white/gray contrast as a predictor of chronological age, and an index of cognitive performance. *Neuroimage* **173**, 341–350 (2018).
56. Teterova, A. & Pat, N. *The (Limited?) Utility of Brain Age as a Biomarker for Capturing Cognitive Decline* (elif Sciences Publications Ltd, 2023).
57. Xinran, W. *et al.* Morphometric dis-similarity between cortical and subcortical areas underlies cognitive function and psychiatric symptomatology: A preadolescence study from ABCD. *Mol. Psychiatry* **28**(3), 1146–1158. <https://doi.org/10.1038/s41380-022-01896-x> (2022).
58. Zarrar, S. *et al.* The preprocessed connectomes project quality assessment protocol - A resource for measuring the quality of MRI data. *Front. Neurosci.* <https://doi.org/10.3389/conf.fnins.2015.91.00047> (2015).
59. Fischl, B. *et al.* Automatically parcellating the human cerebral cortex. *Cereb. Cortex* **14**(1), 11–22 (2004).
60. Segonne, F. *et al.* A hybrid approach to the skull stripping problem in MRI. *Neuroimage* **22**(3), 1060–1075 (2004).

61. Segonne, F., Pacheco, J. & Fischl, B. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Trans. Med. Imaging* **26**(4), 518–529 (2007).
62. Fischl, B., Liu, A. & Dale, A. M. Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans. Med. Imaging* **20**(1), 70–80 (2001).
63. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis – I. Segmentation and surface reconstruction. *Neuroimage* **9**(2), 179–194 (1999).
64. Dale, A. M. & Sereno, M. I. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction - A linear-approach. *J. Cogn. Neurosci.* **5**(2), 162–176 (1993).
65. Fischl, B. & Dale, A. M. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U.S.A.* **97**(20), 11050–11055 (2000).
66. Fortin, J. P. *et al.* Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage* **167**, 104–120 (2018).
67. Fortin, J. P. *et al.* Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* **161**, 149–170 (2017).
68. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**(1), 118–127 (2007).
69. He, N. *et al.* Predicting human inhibitory control from brain structural MRI. *Brain Imag. Behav.* **14**(6), 2148–2158. <https://doi.org/10.1007/s11682-019-00166-9> (2019).
70. Fornito, A., Zalesky, A. & Bullmore, E. *Fundamentals of Brain Network Analysis* (Academic Press, 2016).
71. Karatzoglou, A., Smola, A., Hornik, K. in *kernelab – An S4 Package for Kernel Methods in R.* (2019).
72. Scheinost, D. *et al.* Ten simple rules for predictive modeling of individual differences in neuroimaging. *Neuroimage* **193**, 35–45 (2019).
73. Poldrack, R. A., Huckins, G. & Varoquaux, G. Establishment of best practices for evidence for prediction: A review. *JAMA Psychiatry* **7**(5), 534 (2019).

## Acknowledgements

Autism Brain Imaging Data Exchange I data were made available by Adriana Di Martino supported by (NIMH K23MH087770) and the Leon Levy Foundation. Primary support for the work by Michael P. Milham and the INDI team in relation to this data was provided by gifts from Joseph P. Healy and the Stavros Niarchos Foundation to the Child Mind Institute, as well as by a NIMH award to MPM (NIMH R03MH096321). For this work, DGK was supported by a grant from Aston College of Health and Life Sciences to JN and DGK and a grant from Birmingham Children's Hospital Research Foundation (BCHRF) to AW.

## Author contributions

D.G.-K. conceptualised and designed the study, developed methods, analysed the data, and drafted the initial manuscript. J.N. and A.W. reviewed and edited the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to J.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023