Structural connectome fingerprinting and age prediction in pediatric development: assessing voxel- and surface-based white matter connectivity

Noor B Al-Sharif^a, Etienne St-Onge^b, Jacob W Vogel^a, Maxime Descoteaux^{b,*}, Alan C Evans^{a,*}

^a*MCIN, Montreal Neurological Institute, McGill University, Montreal, Canada*

^b*SCIL, Sherbrooke University, Sherbrooke, Canada*

Abstract

Mapping structural white matter connectivity is a challenge, with many barriers to accurate representation. Here, we assessed the replicability and reliability of two connectome-generating methods, voxel- or surface-based, using test-retest analyses, fingerprinting and age prediction. The two connectomic methods are initiated by the same state-of-the-art dMRI processing pipeline before diverging at the tractography and connectome-generating steps using either voxels or surfaces. While both methods performed very well across all analyses, voxel-based connectomes performed marginally better than surface-based connectomes. Notably, structural connectomes derived from either method demonstrate reliably accurate representations of both individuals and their chronological age, comparable to similar analyses employing multi-modal features. The difference in methodological performance could be attributed to a number of method-specific features but ultimately show that cutting-edge tractography with robust dMRI processing produces reliable white matter connectivity measures.

Keywords: connectivity, diffusion-weighted imaging, surfaces, tractography

1. Introduction

The majority of post-natal brain development occurs in white matter (WM), namely through the pruning, refinement or development of axonal

*These authors should be considered joint senior author.

connections (Sowell et al., 2004; Lenroot & Giedd, 2006; Giedd & Rapoport, 2010). These microstructural changes underlie larger transformations in both structural and functional whole-brain connectivity (Lebel et al., 2008). Despite advancements in research methods at multiple levels of inquiry, we are still establishing a ‘map’ of how developmental changes manifest anatomically, in complement to well-described cognitive development (Goldman-Rakic, 1987; Nagy et al., 2004; Barrouillet, 2015). Abnormalities in both structural and functional connectivity have been suggested to underlie myriad neurological conditions, disorders and diseases (O’Sullivan et al., 2001; Pievani et al., 2014; Kana et al., 2014). Having a clear, comprehensive reference of how the brain typically looks during development will help to better understand deviations from the norm.

Structural connectivity, a representation of anatomical WM connections, is a measure designed to assess how the brain is internally connected (Jones, 2008; Jbabdi et al., 2015; Sotiropoulos & Zalesky, 2019). Due to a number of methodological limitations, however, characterizing brain connectivity *in vivo* is notoriously difficult (Jones et al., 2013; Jeurissen et al., 2019). Imaging methods typically employed to determine *in vivo* connectivity, such as T1- or diffusion-weighted magnetic resonance imaging (dMRI), lack the resolution acuity to depict axon-specific connections. Therefore, estimations of structural connectivity are made using diffusion tensor imaging (DTI) such as fractional anisotropy (FA), radial, mean and axial diffusivity (RD, MD, AD), as well as tractography (Basser & Jones, 2002). The latter contributes to one of the most common measures of WM connectivity: streamline count, the number of tractogram streamlines which connect two brain regions. In addition to scanning resolution, the accuracy of these measures is also impeded by a number of issues in tractogram representation such as crossing fibers and the gyral crown bias (Schilling et al., 2018; Girard et al., 2014).

One of the most common methods to measure structural connectivity is the voxel-based approach, which uses a predefined atlas in 3D voxel space to delineate anatomical regions. Tractograms aligned within the same space represent connections between regions of interest (ROIs) when a streamline crosses a labelled voxel region. Though voxel-wise methods are widespread and fairly simple to execute and interpret, using voxels to represent connectivity does present some anatomical limitations. These include structural distortions introduced by nonlinear registration, poor cortical coverage and bias against complex or short-range streamlines (Jbabdi & Johansen-Berg, 2011; Van Essen et al., 2014). Recent advances in cortical surface modeling

42 have improved our ability to represent anatomical connectivity in ways that
43 cannot be captured as efficiently or accurately using voxel-wise approaches
44 (Smith et al., 2012; St-Onge et al., 2018).

45 Surface-based connectivity uses cortical mesh surfaces, instead of vox-
46 els, as the seeding points for initializing and terminating streamlines (Smith
47 et al., 2012; St-Onge et al., 2018). Enhancing tractography with surfaces
48 attempts to bypass certain disadvantages of traditional tractography, gen-
49 erating more precise and uniform endpoint placement across a surface. By
50 using information from the cortical surface directly in creating the connec-
51 tome, surface-based connectivity also offers the potential to integrate other
52 grey matter (GM) features such as cortical thickness, GM volume and surface
53 area in connectomic analyses.

54 A key component of a method’s utility lies in its stability, which is often
55 described in the contexts of reliability and reproducibility. These are quan-
56 tifiable properties which determine how well a method’s results can be repli-
57 cated and extrapolated to other datasets. The former illustrates a method’s
58 inherent accuracy while the latter depicts its broader real-world relevance.
59 While methods can be validated through replication in another dataset, we
60 can assess our methods within the same dataset by relying on the use of
61 same-session repeated acquisitions.

62 Here, to evaluate and compare the reliability and reproducibility of both
63 voxel- and surface-based connectivity methods, we employed three analy-
64 ses: test-retest similarity/distance measures, fingerprinting and age predic-
65 tion. Using corresponding pairs of imaging and connectivity data, a simi-
66 larity/distance metric can be measured to evaluate the similarity between
67 two data files from the same subject. The same metric can be calculated
68 between subjects, allowing the computation of inter- and intra-subject sim-
69 ilarity/distance averages and variance for each method. Another practical
70 and popular application of test-retest analysis is fingerprinting (Finn et al.,
71 2015), a method which assesses how well subjects can be identified using only
72 their imaging data. This is determined by how accurately a subject’s scan
73 correlates to itself or to another scan from the same subject, as compared
74 to the rest of the dataset. Fingerprinting allows us to understand both how
75 reliable the results are and also how well measures of structural connectiv-
76 ity represent unique individuals. Thirdly, we use age prediction as a tool
77 to validate the biological relevance and reliability of structural connectivity
78 methods. Determining how well a subject’s age can be predicted using struc-
79 tural connectivity illustrates whether such measures accurately and reliably

80 represent normative age across development (Cole et al., 2017; Díaz-Arteche
81 & Rakesh, 2020). While using white matter-derived features in such data-
82 driven analyses is a budding field, few studies have focused specifically on the
83 use of white matter in age prediction during pediatric development (Brown
84 et al., 2012; Han et al., 2014; Erus et al., 2015; Niu et al., 2020). We exe-
85 cute the test-retest and fingerprinting analyses using same-session test-retest
86 data for 560 subjects from a pediatric imaging dataset and the age predic-
87 tion in 813 subjects from the same study. These data are preprocessed with
88 *Tractoflow-pve* (Al-Sharif et al., 2020; Theaud et al., 2020), a state-of-the-art
89 diffusion processing pipeline from the Sherbrooke Connectivity Imaging Lab
90 (SCIL), after which the connectomic mapping is completed using two differ-
91 ent strategies: voxel- or surface-based. Due to the surface-based method’s
92 more nuanced tractogram construction, we hypothesized that its connectivity
93 representations should provide improved results over voxel-based measures
94 in accurately depicting structural connectivity.

95 2. Methods

96 Anatomical MRI data were processed initially through one common dMRI
97 processing pipeline, *Tractoflow-pve*, before subsequent processing to generate
98 surface- and voxel-based connectomes. Processing steps are briefly described
99 below; full details are described in (Al-Sharif et al., 2020), which also provides
100 links to the data and processed derivatives.

101 2.1. Participants

102 Participants, aged 3-22 years, were recruited as part of the Pediatric Imag-
103 ing Neurocognition and Genetics (PING) study in nine major metropolises
104 across the United States (Jernigan et al., 2016). Here, 821 subjects’ data
105 were used in total (426 male), 560 of which had at least two same session
106 acquisitions for use in the test-retest analyses. In the test-retest group, 8
107 subjects were missing complete demographic information including age but
108 were retained as these features were irrelevant to the analyses. These subjects
109 were excluded from the age prediction analyses, bringing the total sample for
110 that analysis to 813 subjects. Of these 813 subjects, the mean age was 12.75
111 years (SD = 4.85 years) and in the test-retest group, the mean age for the
112 552 subjects with complete demographic data was 13.35 years (SD = 4.61
113 years). Further details are provided in Jernigan et al. (2016).

114 2.2. MR acquisition

115 As part of the PING study data collection, multi-modal images were
 116 acquired using harmonized protocols on four 3T MR scanners (Siemens Trio-
 117 Tim, GE Discovery, GE Signa, Philips Achieva) at nine sites. T1w scans in-
 118 cluded a motion-reduction protocol to minimize motion artefacts commonly
 119 found in pediatric datasets. DMRI acquisitions consisted of 32-direction
 120 scans, including 2 baseline b0 volumes, with a single shell b-value of 1000.
 121 From PING, we initially gained access to a total of 954 subjects with both
 122 T1w and dMRI volumes, of which 634 had at least two or more repeated
 123 same-session scans. Subjects were quality assessed at multiple steps of pro-
 124 cessing and some were excluded for major scanning artefacts, incomplete
 125 demographic information or failed preprocessing. In total, 1543 unique acqui-
 126 sitions were used in this study, from 821 subjects, 560 of which had multiple
 127 same-session scans.

128 2.3. T1w and dMRI processing

129 *Tractoflow* (Theaud et al., 2020), developed by SCIL, is a state-of-the-
 130 art dMRI processing pipeline which takes raw dMRI, complementary T1w
 131 MRI and bvec/bval files as input, and generates processed diffusion met-
 132 ric derivatives including tractography. For our purposes, we modified the
 133 original *Tractoflow* pipeline to include the whole-brain mask and partial vol-
 134 ume estimation (PVE) maps for GM, WM, CSF and subcortical GM from
 135 CIVET 2.1 (Ad-Dab'bagh et al., 2006) along with the initial dMRI input.
 136 The pipeline is openly available through the NIMH Data Archive and GitHub
 137 as *Tractoflow-pve* (Al-Sharif et al., 2020).

138 Through *Tractoflow-pve*, T1w images are processed using their given cor-
 139 responding CIVET 2.1-produced brain masks and PVE segmentation files.
 140 T1w images are masked, de-noised and N4 bias-field corrected before un-
 141 dergoing nonlinear registration to the diffusion image. Their complemen-
 142 tary PVE and binary mask files also undergo the same image registration.
 143 Diffusion-weighted volumes and corresponding b0 images files are masked, de-
 144 noised, bias-field and topup-eddy corrected along with their bvec and bval
 145 files. Diffusion tensor imaging (DTI) metrics are extracted and computed
 146 from the dMRI volume to generate FA, RD, AD, and MD image volumes, as
 147 well as fibre orientation distribution function (fODF) volumes.

148 2.4. *Tractography*

149 Within *Tractoflow-pve*, structural tractograms were generated using Par-
 150 ticle Filtering Tractography (PFT) (Girard et al., 2014), which employs WM-
 151 GM-CSF tissue maps to demarcate “include” and “exclude” regions of the
 152 brain. “Include” regions define areas of WM where valid, anatomical stream-
 153 lines are likely to reach. The resulting tractograms are an approximate rep-
 154 resentation of WM macrostructure and connections within the brain. A sub-
 155 sequent processing pipeline called *Surface-Enhanced Tractography (SET)* (St-
 156 Onge et al., 2018), was employed to generate surface-based tractograms and
 157 structural connectomes. An adaptation of the PFT algorithm, SET uses sur-
 158 faces instead of tissue maps to create streamlines initialized at the GM-WM
 159 boundary of cortical and subcortical surfaces. Cortical surfaces are output
 160 by CIVET 2.1 directly whereas subcortical surfaces, including the brainstem,
 161 were converted from CIVET 2.1 ANIMAL segmentation labels (Collins et al.,
 162 1999) using a marching cube algorithm. SET combines these surfaces with
 163 dMRI output from *Tractoflow-pve* to reconstruct WM connectivity.

164 2.5. *Template construction*

165 To avoid using an adult brain template for this pediatric dataset, a study-
 166 specific common space T1w template was created using Advanced Normal-
 167 ization Tools’ (ANTs) iterative multi-variate template construction script
 168 (Avants et al., 2011). The resulting transformation files were used to reg-
 169 ister the voxel-wise DKT atlas into native subject T1w and dMRI space to
 170 generate the voxel-based connectivity matrices, as described below.

171 2.6. *Connectomes*

172 Voxel-based connectomic matrices were generated using a Desikan-Killiany-
 173 Tourville (DKT) (Desikan et al., 2006) segmentation label aligned to each
 174 subject’s native diffusion space. To ensure the same subcortical regions were
 175 used in both surface- and voxel-based methods, the DKT cortical atlas was
 176 combined with subcortical labels from CIVET 2.1 ANIMAL segmentation in
 177 each subject’s T1w space before registration to diffusion space. The dMRI-
 178 registered segmentation label and the subject’s tractogram were used to cre-
 179 ate acquisition-specific connectivity matrices. Connectivity is represented
 180 as the number of tractography streamlines with endpoints within a ROI as
 181 delineated by the DKT label. Surface-based connectomic matrices were gen-
 182 erated using SET by counting the number of streamlines which intersect
 183 two cortical or subcortical regions of a surface. Surfaces are delineated into

184 ROIs using cortical labels in the CIVET DKT surface atlas. Connectomes
185 were subsequently normalized by the sum of their total streamline count to
186 facilitate comparison between methods.

187 *2.7. Analysis*

188 *2.7.1. Test-retest*

189 To assess the variability of our reconstructed connectivity matrices, we
190 computed the distance between two acquisitions from the same subject (inter-
191 subject) and other subjects (intra-subject). This metric informs how similar
192 two matrices are to each other and also how distinguishable they are from
193 matrices of a different subject. For both approaches, we calculated the mean
194 L1 distance for both inter- and intra-subject comparisons and determined
195 the ratio of inter/intra distance (Aggarwal et al., 2001; Allen et al., 2014; St-
196 Onge et al., 2018). A lower L1 distance indicates more similarity between two
197 connectomes, where a distance of 0 would represent two identical matrices.
198 Thus, a higher ratio indicates a better ability both to match subjects to
199 themselves (intra L1) and also distinguish them from others (inter L1). The
200 L1 distance between two connectivity matrices (A, B) is defined by the sum
201 of the absolute difference between each connection weight (c), $\sum_{\forall c} |A_c - B_c|$.
202 Results from each method were compared using a t-test.

203 *2.7.2. Fingerprinting*

204 In addition to the above assessments, we used data from the 560 subjects
205 with multiple same-session scans to implement a fingerprinting analysis. The
206 first and second scans of all subjects with at least two acquisitions were ex-
207 tracted as follows into two separate sets of data. First, each connectome’s
208 diagonal was set to zeros before symmetrizing and normalizing the matrices.
209 Symmetrized matrices are created by averaging the connectivity between cor-
210 responding pairwise connections, and matrices were normalized by dividing
211 each connection by the sum of all pairwise connections within the matrix.
212 Data from each connectome were then vectorized, or flattened, and appended
213 to a 2D matrix for either “first” or “second” scans. To determine the simi-
214 larity between a “first” scan and all other “second”, or repeated, scans, we
215 ran a Pearson’s correlation between each acquisition’s connectome and the
216 connectomes of every other repeated acquisition in the dataset. Fingerprint-
217 ing is considered successful if the maximum correlation value results from the
218 comparison of two scans from the same subject. The fingerprinting accuracy

219 between the voxel- and surface-based methods was then compared with a χ^2
220 test.

221 2.7.3. Age prediction

222 While the previous approaches assess and compare the reproducibility of
223 the two diffusion pipelines, we also wished to assess and compare how well
224 data prepared using these methods performed in real-world applications. We
225 chose to submit these data to a machine learning pipeline designed to pre-
226 dict participant age using purely neuroimaging information. Age prediction
227 was chosen as it is a fairly easy problem with wide application across the
228 neuroimaging field (Cole et al., 2017; F da Costa et al., 2020; Niu et al.,
229 2020; Han et al., 2014; Tønnesen et al., 2020; Liem et al., 2017; Erus et al.,
230 2015; Díaz-Arteche & Rakesh, 2020), and because it is relevant to the biol-
231 ogy of neurodevelopment. Our goal was to establish the maximum reliable
232 age prediction accuracy (and minimum age prediction error) using basic ma-
233 chine learning approaches and the preprocessed imaging data as input. The
234 machine learning pipeline was applied separately for datasets preprocessed
235 using voxel-wise and surface-based approaches, and output from the best
236 model from each dataset were compared. We were also interested in whether
237 additionally adding cortical thickness data could improve predictions.

238 As above, the diagonal of each subject’s tractographic connectivity matrix
239 was set to 0s, the matrices were symmetrized and normalized. For subjects
240 with more than one same-session scan, matrices derived from all acquisitions
241 were averaged such that each subject had only one connectivity matrix. The
242 lower triangle of each matrix was then extracted, flattened, and appended
243 to one another, creating a 2D subject x (connectivity) feature matrix. A
244 similar subject x feature matrix was created for cortical thickness (CTx)
245 by appending vectors for each subject containing the cortical thickness of
246 each brain region. Both types of feature matrices were normalized using a
247 standard scaler. To partially account for differences in scanner models, we
248 included dummy coded variables for each of the scanner models into the
249 feature matrices. Finally, we set aside data from 20% of the participants as
250 a final “lock-box set”, and ensured that the proportion of subjects younger
251 than 8, 8-12, and 13+ was equal between the lock-box set and the rest of the
252 data.

253 The following analysis was conducted using the python package scikit-
254 learn (Pedregosa et al., 2011). We next built a grid search that cycled through
255 different combinations of pipeline steps, machine learning estimators, imaging

256 features and hyperparameters. Specifically, the following pipeline aspects
257 were varied:

- 258 • Input features: WM connectivity only, CTx only, WM and CTx fea-
259 tures
- 260 • Feature dimension reduction: None, PCA (100 components), only fea-
261 tures with a significant (FDR $p < 0.05$) correlation with age (f_regression)
- 262 • Estimators and hyperparameters:
 - 263 – LassoCV
 - 264 – Support vector regressor (SVR) w/ Linear kernel w/ C chosen
265 from the array[0.0001,0.001,0.1,1,10]
 - 266 – Random forest regressor (RFR) w/ max depth selected from the
267 array [3, 5, 7, 10, 15, 20, 40, 50, None]

268 Hyperparameters for all models were chosen using a nested cross-validated
269 grid search (GridSearchCV), except for Lasso, which has a built-in cross-
270 validation framework for determining the value for the alpha parameter. All
271 possible combinations of these pipeline inputs were modeled with the excep-
272 tion that no PCA was applied for models using CTx only features as the
273 number of components exceeded the number of input features.

274 For each pipeline, thirty 80/20 training/validation splits were created
275 from the data. For each split, the Pipeline was fit to the training data and
276 used to predict the age of unseen individuals in the testing data. The mean
277 squared error (MSE) and R^2 score were recorded for each split in order to
278 establish a range of uncertainty within the predictions of each pipeline. For
279 both the voxel-wise and SET processed data separately, the best performing
280 model was selected based on having the lowest MSE and/or highest R^2 score.
281 In the event of more than one model with equivalent scores, the model with
282 the lowest number of features was selected. Finally, the best performing
283 model was evaluated on the left out 20% lock-box set, and this performance
284 was compared across the two WM preprocessing pipelines. To validate that
285 the selected model's performance was not influenced by the specific lock-
286 box sample, 100 new train-test splits were generated from the total dataset
287 and trained using the only selected model. The distribution of R^2 and MSE
288 results on the new 100 test sets were then compared between both voxel

289 and surface methods. To assess whether subcortical regions were driving
290 prediction results, i.e. regions in which SET might be less effective, the
291 analyses were repeated using only cortical ROIs, again comparing R^2 and
292 MSE results.

293 2.7.4. *Resampling*

294 To test whether our results were robust to arbitrary decisions in stream-
295 line seeding count, we resampled all connectomes for each method at various
296 counts of resulting streamlines. This allowed us to evaluate whether the test-
297 retest variability of a method would change over a range of total streamline
298 counts. The original voxel-based connectomes were initialized with a total
299 of 50 million (M) seeds while the surface-based connectomes totalled 25M
300 seeds. However, the resulting amount of reconstructed streamlines varies per
301 subject and averaged to *5M for PFT and *10M for SET. For this reason, re-
302 sulting connectomes were resampled at specific total streamline counts: 100k,
303 250k, 500k, 1M, 1.5M, and 2M. In addition, to resample the connectomes by
304 seed count, the number of streamline count was weighted by the percentage
305 of seeds resulting in a streamline. This resampling was done at 100k, 250k,
306 500k, 1M, 1.5M, 2M, 3M, 5M, 7.5M, 10M, and 15M total seeds. We then
307 re-ran the fingerprinting and age prediction analyses with the resampled con-
308 nectomes, grouped by preprocessing method (voxel or surface), resampling
309 reference (seeds or streamlines), and matrix resolution (seed or streamline
310 count), and compared the results between corresponding groups of voxel and
311 surface connectomes.

312 3. Results

313 3.1. *Test-retest*

314 Test-retest L1 distance (Table 1) for voxel-derived connectomes showed
315 an intra-subject distance average = 0.144 (0.061 SD) that was significantly
316 lower ($T = -73.4$, $p < 0.0001$) than that of surface-based connectomes at
317 0.231 (0.061 SD). Alternatively, voxel connectomes displayed an inter-subject
318 distance of 0.418 (0.021 SD) which was also lower than that of the surface
319 connectomes at 0.554 (0.032 SD). The subsequent intra/inter distance ratio
320 for voxel connectomes was 2.900 and that of the surface connectomes was
321 2.394. Similar results were obtained using L2 and χ^2 distance.

L1 distance <i>method</i>	intra-subj		inter-subj		ratio
	mean	SD	mean	SD	
SET	0.231	0.061	0.554	0.032	2.394
VOX	0.144	0.027	0.418	0.021	2.900

Table 1: L1 distance metric results where intra-subject distance measures similarity of data from same subject and inter-subject distance measures similarity between data from different subjects. Intra/inter ratio indicates better ability to self-identify and distinguish from others.

3.2. Fingerprinting

Fingerprinting analyses using voxel-based connectomes were able to match subjects to their own scans with 99.8% accuracy, correctly identifying 559 subjects out of 560 (Figure 1). This performance was significantly better ($\chi^2 = 29.6$, $p < 0.0001$) than that of surface-based connectome fingerprinting showed 94.6% accuracy, correctly identifying 530 subjects.

Figure 1: **a.** Correlation of each subject to itself for both connectomic methods, determined by a Pearson's r . Subjects whose highest correlation resulted from a correct self-identification are marked in orange, incorrect instances are in blue. **b.** Fingerprint accuracy per method, as a percent of correctly identified matches.

3.3. Age prediction

After iteratively training and testing multiple models with 80% of the total dataset, we evaluated the average test-set R^2 and MSE of each model

331 to determine the best fit for both connectomic methods (Figure 2). We also
 332 prioritised using models with fewer features. The best fitting model using
 333 voxel-based connectomes was Lasso Select using both GM (CTx) and WM
 334 (connectivity) features, with an average R^2 of 0.804 (range = 0.728-0.833,
 335 mean MSE = 4.48, Figure 2a and c). Using CTx features alone resulted
 336 in an average R^2 of 0.711 (0.635-0.765) and using WM connectivity alone
 337 resulted in an average R^2 of 0.733 (0.637-0.797). We then applied this model
 338 to the final left out 20% lock-box set for voxel-based connectomes, resulting
 339 in an R^2 of 0.824 (MSE = 4.46, MAE = 2.11, Figure 2e). Feature weights for
 340 which voxel-wise connections contributed most to the age prediction model
 341 are depicted in Figure 3a.

342 The best fitting model for surface-based connectomes was also Lasso Se-
 343 lect using both GM and WM features with a mean R^2 of 0.802 (range =
 344 0.756-0.840, mean MSE = 4.52) (Figure 2b and d). Using GM alone resulted
 345 in an average R^2 of 0.711 (range = 0.635-0.765), while using WM connec-
 346 tivity alone resulted in an average R^2 of 0.726 (range = 0.633-0.779). In
 347 the final lock-box set, this model resulted in an R^2 of 0.797 (MSE = 5.14,
 348 MAE = 2.27) (Figure 2f). Feature weights for which surface-wise connections
 349 contributed most to the age prediction model are depicted in Figure 3b.

350 R^2 and MSE distributions from the 100 train-test samples using only
 351 the best performing model showed that age prediction performance was not
 352 significantly different between voxel and surface methods (R^2 : $p = 0.42$,
 353 MSE: $p = 0.41$). These results were unchanged when using cortical ROIs
 354 only.

355 3.4. Resampling

356 While the accuracy of the fingerprinting analyses varied slightly with re-
 357 sampling, the results from voxel-based connectomes (seeds mean = 0.997,
 358 streamlines mean = 0.998) were still consistently higher than those of the
 359 surface-based connectomes (seeds mean = 0.943, streamlines mean = 0.945),
 360 at all resolutions (Figure 4, Table 2). Age prediction results also varied little
 361 from the full resolution connectomes, with voxel-based connectome predic-
 362 tions (seeds mean $R^2 = 0.814$, streamlines mean $R^2 = 0.811$) almost exactly
 363 equal to those of the surface-based connectomes (seeds mean $R^2 = 0.810$,
 364 streamlines mean $R^2 = 0.810$), regardless of the seed or streamline resolu-
 365 tion.

Figure 2: Average results for all models trained and tested using connectomes from both methods: a and b show results using voxel-based connectomes; d and e show results using surface-based connectomes. Each machine learning pipeline was tested using WM connectivity only (blue), CT only (orange) and ALL features (green). Pipeline with the highest average R^2 and lowest average MSE when applied to test set (indicated by red arrows) was selected as model to evaluate on lock-box data (c and f).

Seeds	R^2		Accuracy (%)	
	mean	range	mean	range
VOX	0.814	(0.803, 0.821)	0.997	(0.996, 0.998)
SET	0.810	(0.797, 0.818)	0.944	(0.941, 0.946)
Streamlines	R^2		Accuracy (%)	
	mean	range	mean	range
VOX	0.811	(0.808, 0.815)	0.998	(0.996, 0.998)
SET	0.810	(0.801, 0.814)	0.945	(0.943, 0.946)

Table 2: Average age prediction (R^2) and fingerprinting (accuracy) results for resampled connectomes, grouped by resampling reference (seeds or streamlines) per connectomic method.

Figure 3: Feature weights: connectivity features which contributed most strongly to the age prediction model for a. voxel-based connectomes and b. surface-based connectomes, depicted as an increase with age within that connection in red and a decrease in blue.

Figure 4: Mean fingerprinting accuracy and age prediction R^2 results across all resampled connectomes, using either seeds (a,c) or streamlines (b,d) as the resampling reference. Inset shows close-up of differences in fingerprinting accuracy between methods.

4. Discussion

The quantification of whole-brain white matter structure presents a process rife with pitfalls and erroneous estimation. To assess the methods available for overcoming these limitations when measuring structural connectivity, we employed both traditional and more recently developed preprocessing techniques. Despite the potential advancements offered by using Surface-Enhanced Tractography, connectomes from voxel-based pipelines outperformed surface-based connectomes in reproducibility metrics and produced equivalent results in age prediction. While results varied slightly between the two, both pipelines generated connectivity derivatives which were highly reliable and reproducible.

Both pipelines were able to match subjects to their own data with very high accuracy, though using voxel-based connectomes achieved slightly better results than those of surface-based connectomes. The higher L1 inter/intra ratio for voxel-based connectomes indicated that voxel-based connectomes were more easily distinguished solely based on the connectivity matrix similarity. This result was unexpected as we predicted that with more enhanced, biologically accurate streamline coverage provided by SET, subjects would have more nuanced and individualised tractograms and subsequent connectomes. However, SET's higher frequency streamline improvements may have introduced more noise into the connectomes, increasing the L1 distance, making them harder to distinguish from each other. The enhancements may also have been negated by the use of a broad brain parcellation such as DKT,

rendering these improvements inconsequential for this particular analysis. Alternatively, the voxel-based connectomes might be subject to subject-specific deformities introduced by nonlinear registration, which would make an individual more easily matched to their own data. These results outperform the first instance of connectomic fingerprinting using rsfMRI which presented 89% accuracy using whole-brain connectomes (Finn et al., 2015). Parker et al. (2016) first showed that anatomical white matter data can be used as fingerprints based on tract shape, with 93% accuracy, which both our connectome groups surpassed. Another study of the Human Connectome Project (Kumar et al., 2018) compared imaging data from multiple modalities and found that, while using FA from dMRI showed better fingerprinting capabilities than T1w features alone, a combination multi-modal features offered the best “separability”. Here, the voxel-based connectomes outperformed any of the dMRI metrics used on their own in Kumar et al. (2018) and matched the results using multi-modal combinations. White matter features used in fingerprinting may outperform resting-state connectivity due to their reliance on defined anatomical structure rendering better reliability between scans, as opposed to state-dependent connectomes. Notably, regardless of preprocessing pipeline, differences in individual structural connections as represented by tractography are measurable to the extent that they can be used reliably for unique individual identification.

While the age prediction results on the lock-box set was qualitatively higher using voxel-based connectomes, the range of results from an iterative sample show that this difference is not significant. Given the wide age range of the PING dataset, the defining drivers of age prediction may lie in changes within more principal, long-range connectivity patterns, such as those that transverse the subcortical structures, and not in areas of nuanced local connectivity, which are more likely to have been enhanced by SET. Subject specificity to such an individualized degree at the local level would be irrelevant if the larger effects of age were determined by other regions or measures. When subcortical ROIs were removed from the analysis, however, the results remained as is. This might suggest these results are similarly impacted by SET-induced high frequency noise or DKT-induced flattening of high frequency data, as in the fingerprinting. Interestingly, the plotted weights of connectivity features which contributed to the prediction model differ widely between voxel- and surface-based methods (Fig. 3). Voxel-based prediction models show sparse feature plots, with few regions contributing to the model, whereas the surface-enhanced connectivity plots are far denser,

likely due to SET's on-average increased streamline count and cortical coverage across the whole brain. However, this may also indicate SET's ability to incorporate more whole-brain connectivity features into its connectomic representation as more connections are generated in SET's tractograms. SET might improve modeling and representation of other connectivity measures in other contexts, especially when coupled with grey matter features, but these potential improvements could be far too small to have an impact in our current analyses. SET may also introduce other sources of error that prevented it from outperforming voxel-based methods. Other studies in pediatric or lifespan age prediction have shown comparable results, the best of which consistently use multi-modal models including more features than we have employed here (Brown et al., 2012; Erus et al., 2015; Liem et al., 2017; Han et al., 2014; Niu et al., 2020; Cole et al., 2017). However, we do show that even with fairly basic preprocessing tools and limited measures, high age prediction accuracy can still be achieved.

Hypothesizing that SET's advanced tractogram modeling might outperform voxel-based methods with fewer streamlines, i.e. at lower sampled connectomes, we resampled all connectomes at varying levels of seed and streamline count. Resampling had almost no effect on either the fingerprinting or age prediction results. Fingerprinting accuracy using voxel-based connectomes was nearly identical at every instance of resampling by streamlines. Marginally less consistency is seen when resampling by seeding point, though the range in accuracy is the same as in streamline resampling; streamline resampling appeared to be most consistent using between 500k and 1.5m streamlines. Surface-based connectomes also show stable results regardless of resampling, with the most consistent results when resampling by seeds between 1m and 3m seeding points. In either case, these results suggest that highly accurate identification of individual subjects using structural connectomes can be consistently achieved with an average minimum of 1m seeding points, forgoing the need for more computationally demanding processing at higher resolutions. However, the potential increased connectivity and higher frequency impacts of SET may not be captured in our current analysis.

5. Conclusion

In summary, our results indicate that structural connectomes, regardless of methodology, can be used reliably for subject-specific identification and capture biologically relevant characteristics of development. Ultimately,

however, the strength of these connectomes may lie in the dMRI processing pipeline that initiates both voxel- and surface-based methods. With a robust, reliable processing pipeline such as *Tractoflow*, resulting tractograms lead to reliable and informative structural connectomes. Using voxel-based atlases to create connectomes is a common and quick technique that may be sufficient for high-level analyses. Surface-based methods will take more time and computational resources, but may be justifiable if subsequent analyses call for integrating grey matter or other surface-based features such as cortical thickness.

Data Availability Statement

The data that support the findings of this study are available in NIMH Data Archive (NDA) at 10.15154/1519178, study ID #932 ([dataset] Al-Sharif et al., 2020). These data were derived from the following resources also available in NDA as Pediatric Imaging, Neurocognition, and Genetics (PING) study ID #2607.

Acknowledgements

Supported by grants from Brain Canada (238990, 243030), CFREF/HBHL Innovative Ideas (247613), Coutu Research Fund (241177), CFREF/HBHL Discovery (247712), Université de Sherbrooke Research Chair in NeuroInformatics and NSERC Discovery Grant.

Conflict of Interest

Maxime Descoteaux is co-founder of Imeka Solutions Inc. Other authors declare they have no actual or potential competing financial interests.

References

Ad-Dab'bagh, Y., Lyttelton, O., Muehlboeck, J., Lepage, C., Einarson, D., Mok, K., Ivanov, O., Vincent, R., Lerch, J., Fombonne, E. et al. (2006). The civet image-processing environment: a fully automated comprehensive pipeline for anatomical neuroimaging research. In *Proceedings of the 12th annual meeting of the organization for human brain mapping*. Florence, Italy volume 2266.

- 493 Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the surprising
494 behavior of distance metrics in high dimensional space. In *International*
495 *conference on database theory* (pp. 420–434). Springer.
- 496 [dataset] Al-Sharif, N. B., St-Onge, E., Theaud, G., Evans, A. C., & De-
497 scoteaux, M. (2020). Diffusion-weighted derivatives of ping. NIMH Data
498 Archive. doi:10.15154/1519178.
- 499 Al-Sharif, N. B., St-Onge, E., Theaud, G., Evans, A. C., & Descoteaux, M.
500 (2020). Processing the diffusion-weighted magnetic resonance imaging of
501 the ping dataset. *bioRxiv*, .
- 502 Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., &
503 Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the
504 resting state. *Cerebral cortex*, *24*, 663–676.
- 505 Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C.
506 (2011). A reproducible evaluation of ants similarity metric performance in
507 brain image registration. *Neuroimage*, *54*, 2033–2044.
- 508 Barrouillet, P. (2015). Theories of cognitive development: From piaget to
509 today.
- 510 Basser, P. J., & Jones, D. K. (2002). Diffusion-tensor mri: theory, experi-
511 mental design and data analysis—a technical review. *NMR in Biomedicine:*
512 *An International Journal Devoted to the Development and Application of*
513 *Magnetic Resonance In Vivo*, *15*, 456–467.
- 514 Brown, T. T., Kuperman, J. M., Chung, Y., Erhart, M., McCabe, C., Ha-
515 gler Jr, D. J., Venkatraman, V. K., Akshoomoff, N., Amaral, D. G., Bloss,
516 C. S. et al. (2012). Neuroanatomical assessment of biological maturity.
517 *Current biology*, *22*, 1693–1698.
- 518 Cole, J. H., Poudel, R. P., Tsagkrasoulis, D., Caan, M. W., Steves, C., Spec-
519 tor, T. D., & Montana, G. (2017). Predicting brain age with deep learning
520 from raw imaging data results in a reliable and heritable biomarker. *Neu-*
521 *roImage*, *163*, 115–124.
- 522 Collins, D. L., Zijdenbos, A. P., Baaré, W. F., & Evans, A. C. (1999). Ani-
523 mal+ insect: improved cortical structure segmentation. In *Biennial Inter-*
524 *national Conference on Information Processing in Medical Imaging* (pp.
525 210–223). Springer.

- 526 F da Costa, P., Dafflon, J., & Lopez Pinaya, W. H. (2020). Brain-age pre-
527 diction using shallow machine learning: Predictive analytics competition
528 2019. *Frontiers in Psychiatry*, 11, 1367.
- 529 Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C.,
530 Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T.
531 et al. (2006). An automated labeling system for subdividing the human
532 cerebral cortex on mri scans into gyral based regions of interest. *Neuroim-
533 age*, 31, 968–980.
- 534 Díaz-Arteche, C., & Rakesh, D. (2020). Using neuroimaging to predict
535 brain age: insights into typical and atypical development and risk for psy-
536 chopathology. *Journal of Neurophysiology*, 124, 400–403.
- 537 Erus, G., Battapady, H., Satterthwaite, T. D., Hakonarson, H., Gur, R. E.,
538 Davatzikos, C., & Gur, R. C. (2015). Imaging patterns of brain develop-
539 ment and their relationship to cognition. *Cerebral cortex*, 25, 1676–1684.
- 540 Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun,
541 M. M., Papademetris, X., & Constable, R. T. (2015). Functional connec-
542 tome fingerprinting: identifying individuals using patterns of brain con-
543 nectivity. *Nature neuroscience*, 18, 1664–1671.
- 544 Giedd, J. N., & Rapoport, J. L. (2010). Structural mri of pediatric brain
545 development: what have we learned and where are we going? *Neuron*, 67,
546 728–734.
- 547 Girard, G., Whittingstall, K., Deriche, R., & Descoteaux, M. (2014). To-
548 wards quantitative connectivity analysis: reducing tractography biases.
549 *Neuroimage*, 98, 266–278.
- 550 Goldman-Rakic, P. S. (1987). Development of cortical circuitry and cognitive
551 function. *Child development*, (pp. 601–622).
- 552 Han, C. E., Peraza, L. R., Taylor, J.-P., & Kaiser, M. (2014). Predicting
553 age across human lifespan based on structural connectivity from diffusion
554 tensor imaging. In *2014 IEEE Biomedical Circuits and Systems Conference
555 (BioCAS) Proceedings* (pp. 137–140). IEEE.
- 556 Jbabdi, S., & Johansen-Berg, H. (2011). Tractography: where do we go from
557 here? *Brain connectivity*, 1, 169–183.

Jbabdi, S., Sotiropoulos, S. N., Haber, S. N., Van Essen, D. C., & Behrens, T. E. (2015). Measuring macroscopic brain connections in vivo. *Nature neuroscience*, 18, 1546.

Jernigan, T. L., Brown, T. T., Hagler Jr, D. J., Akshoomoff, N., Bartsch, H., Newman, E., Thompson, W. K., Bloss, C. S., Murray, S. S., Schork, N. et al. (2016). The pediatric imaging, neurocognition, and genetics (ping) data repository. *Neuroimage*, 124, 1149–1154.

Jeurissen, B., Descoteaux, M., Mori, S., & Leemans, A. (2019). Diffusion mri fiber tractography of the brain. *NMR in Biomedicine*, 32, e3785.

Jones, D. K. (2008). Studying connections in the living human brain with diffusion mri. *cortex*, 44, 936–952.

Jones, D. K., Knösche, T. R., & Turner, R. (2013). White matter integrity, fiber count, and other fallacies: the do’s and don’ts of diffusion mri. *Neuroimage*, 73, 239–254.

Kana, R. K., Uddin, L. Q., Kenet, T., Chugani, D., & Müller, R.-A. (2014). Brain connectivity in autism. *Frontiers in human neuroscience*, 8, 349.

Kumar, K., Toews, M., Chauvin, L., Colliot, O., & Desrosiers, C. (2018). Multi-modal brain fingerprinting: a manifold approximation based framework. *NeuroImage*, 183, 212–226.

Lebel, C., Walker, L., Leemans, A., Phillips, L., & Beaulieu, C. (2008). Microstructural maturation of the human brain from childhood to adulthood. *Neuroimage*, 40, 1044–1055.

Lenroot, R. K., & Giedd, J. N. (2006). Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. *Neuroscience & biobehavioral reviews*, 30, 718–729.

Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Masouleh, S. K., Huntenburg, J. M., Lampe, L., Rahim, M., Abraham, A., Craddock, R. C. et al. (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment. *Neuroimage*, 148, 179–188.

Nagy, Z., Westerberg, H., & Klingberg, T. (2004). Maturation of white matter is associated with the development of cognitive functions during childhood. *Journal of cognitive neuroscience*, 16, 1227–1233.

- 590 Niu, X., Zhang, F., Kounios, J., & Liang, H. (2020). Improved prediction of
591 brain age using multimodal neuroimaging data. *Human Brain Mapping*,
592 41, 1626–1643.
- 593 O’Sullivan, M., Jones, D. K., Summers, P., Morris, R., Williams, S., &
594 Markus, H. (2001). Evidence for cortical “disconnection” as a mechanism
595 of age-related cognitive decline. *Neurology*, 57, 632–638.
- 596 Parker, G. D., Evans, G. J., & Jones, D. K. (2016). Structural fingerprinting
597 of the human brain: How unique is tract shape to the individual. *ISMRM*.
- 598 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel,
599 O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J.,
600 Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Édouard Duches-
601 nay (2011). Scikit-learn: Machine learning in python. *Journal of Machine*
602 *Learning Research*, 12, 2825–2830.
- 603 Pievani, M., Filippini, N., Van Den Heuvel, M. P., Cappa, S. F., & Frisoni,
604 G. B. (2014). Brain connectivity in neurodegenerative diseases—from phe-
605 notype to proteinopathy. *Nature Reviews Neurology*, 10, 620.
- 606 Schilling, K., Gao, Y., Janve, V., Stepniewska, I., Landman, B. A., & An-
607 derson, A. W. (2018). Confirmation of a gyral bias in diffusion mri fiber
608 tractography. *Human brain mapping*, 39, 1449–1466.
- 609 Smith, R. E., Tournier, J.-D., Calamante, F., & Connelly, A. (2012).
610 Anatomically-constrained tractography: improved diffusion mri stream-
611 lines tractography through effective use of anatomical information. *Neu-
612 roimage*, 62, 1924–1938.
- 613 Sotiropoulos, S. N., & Zalesky, A. (2019). Building connectomes using diffu-
614 sion mri: why, how and but. *NMR in Biomedicine*, 32, e3752.
- 615 Sowell, E. R., Thompson, P. M., & Toga, A. W. (2004). Mapping changes
616 in the human cortex throughout the span of life. *The Neuroscientist*, 10,
617 372–392.
- 618 St-Onge, E., Daducci, A., Girard, G., & Descoteaux, M. (2018). Surface-
619 enhanced tractography (set). *NeuroImage*, 169, 524–539.

- 620 Theaud, G., Houde, J.-C., Boré, A., Rheault, F., Morency, F., & Descoteaux,
621 M. (2020). Tractoflow: A robust, efficient and reproducible diffusion mri
622 pipeline leveraging nextflow & singularity. *NeuroImage*, 218, 116889.
- 623 Tønnesen, S., Kaufmann, T., de Lange, A.-M. G., Richard, G., Doan, N. T.,
624 Alnæs, D., van der Meer, D., Rokicki, J., Moberget, T., Maximov, I. I.
625 et al. (2020). Brain age prediction reveals aberrant brain white matter in
626 schizophrenia and bipolar disorder: A multisample diffusion tensor imaging
627 study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*,
628 5, 1095–1103.
- 629 Van Essen, D. C., Jbabdi, S., Sotiropoulos, S. N., Chen, C., Dikranian,
630 K., Coalson, T., Harwell, J., Behrens, T. E., & Glasser, M. F. (2014).
631 Mapping connections in humans and non-human primates: aspirations
632 and challenges for diffusion imaging. In *Diffusion MRI* (pp. 337–358).
633 Elsevier.