Predicting age and clinical risk from the neonatal connectome

Yassine Taoudi-Benchekroun[1], Daan Christiaens[1,2], Irina Grigorescu[1], Oliver Gale-Grant[1,3] Andreas Schuh[4], Maximilian Pietsch[1], Andrew Chew[1], Nicholas Harper[1], Shona Falconer[1], Tanya Poppe[1], Emer Hughes[1], Jana Hutter[1], Anthony N Price[1], J-Donald Tournier[1], Lucilio Cordero-Grande[1,5], Serena J Counsell[1], Daniel Rueckert[4,6], Tomoki Arichi[1,7,8], Joseph V Hajnal[1], A David Edwards[1,9]*, Maria Deprez[1]*, Dafnis Batalle[1,3]*†

[1] Centre for the Developing Brain, School of Imaging Sciences & Biomedical Engineering, King's College London, London, United Kingdom
[2] Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium
[3] Department of Forensic and Neurodevelopmental Science, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
[4] Biomedical Image Analysis Group, Department of Computing, Imperial College London, London, United Kingdom
[5] Biomedical Image Technologies, ETSI Telecomunicación, Universidad Politécnica de Madrid & CIBER-BBN, Madrid, Spain
[6] Institute for Artificial Intelligence and Data Science in Medicine and Healthcare, Faculty of Medicine and Informatics, Technical University of Munich, Munich, Germany
[7] Department of Bioengineering, Imperial College London, London, United Kingdom
[8] Children's Neurosciences, Evelina London Children's Hospital, Guy's and St Thomas' NHS Trust, London, United Kingdom
[9] MRC Centre for Neurodevelopmental Disorders, King's College London, London, United Kingdom

*Contributed equally

†Corresponding author:

Dr. Dafnis Batalle
Department of Forensic and Neurodevelopmental Science
Institute of Psychiatry, Psychology & Neuroscience
King's College London
16, De Crespigny Park, SE5 8AF, London, UK
+44-(0)-207-848-0922 | dafnis.batalle@kcl.ac.uk

Abstract

The development of perinatal brain connectivity underpins motor, cognitive and behavioural abilities in later life. Diffusion MRI allows the characterisation of subtle inter-individual differences in structural brain connectivity. Individual brain connectivity maps (connectomes) are by nature high in dimensionality and are complex to interpret. Machine learning methods are a powerful tool to uncover properties of the connectome which are not readily visible, and can give us clues as to how and why individual developmental trajectories differ.

In this manuscript we used Deep Neural Networks and Random Forests to predict demographic and neurodevelopmental characteristics from neonatal structural connectomes in a large sample of babies (n = 524) from the developing Human Connectome Project. We achieved an accurate prediction of post menstrual age (PMA) at scan in term-born infants (Mean absolute error (MAE) = 0.72 weeks, r = 0.83 and p<0.001). We also achieved good accuracy when predicting gestational age at birth in a cohort of term and preterm babies scanned at term equivalent age (MAE = 2.21 weeks, r = 0.82, p<0.001). We subsequently used sensitivity analysis to obtain feature relevance from our prediction models, with the most important connections for prediction of PMA and GA found to be predominantly thalamocortical. From our models of PMA at scan for infants born at term, we computed a brain maturation index (*predicted age minus actual age*) of individual preterm neonates and found a significant correlation between this index and motor outcome at 18 months corrected age. Our results demonstrate the applicability of machine learning techniques in analyses of the neonatal connectome, and suggest that a neural substrate for later developmental outcome is detectable at term equivalent age.

1. INTRODUCTION

Magnetic Resonance Imaging (MRI) allows a broad range of in vivo insights about the structure and function of the human brain. Diffusion MRI in particular enables the characterization of microstructural changes in the orientation and organisation of major white matter tracts, the systematic description of whole-brain structural networks: the human *connectome* (Honey et al., 2010; Sporns et al., 2005).

During the perinatal period, the brain undergoes significant changes and consolidation of structural connectivity, which are thought to underpin the expansion of motor, cognitive and behavioural abilities (Johnson, 2001). Since the inception of *connectomics* (Hagmann 2005; Sporns, Tononi, and Kötter 2005) several studies have tried to characterise early development of the structural connectome (Fan et al. 2011; Hagmann et al. 2010). Subtle alterations in the development of brain connectivity have been suggested to underlie atypical neurodevelopmental outcome in populations with perinatal risk factors, such as children born preterm (Batalle et al., 2018). Preterm birth comprises approximately 11% of all births, and is the main global cause of death and disability in children under 5 years of age (Blencowe et al., 2012), as well as representing one of the most pervasive perinatal risk factors for atypical neurodevelopment (Wood et al., 2000). It has been associated with an increased risk of developing neurodevelopmental conditions such as motor, visuospatial and sensorimotor delay (Marlow et al., 2007), inattention, anxiety and social difficulties (Johnson and Marlow, 2014), autism spectrum (Johnson et al., 2010), cerebral palsy (Marlow et al., 2005) or psychiatric disorders in adulthood such as depression and bipolarity (Nosarti et al., 2012). Therefore, improving our understanding of how preterm birth affects structural brain development remains an important goal.

With this goal in mind, the developing Human Connectome Project (dHCP) has collected demographic and MRI data from a large cohort of term and preterm-born neonates. The dHCP dataset comprises structural, diffusion and functional MRI data with high spatial, angular, and temporal resolutions. Features of this project include: advances in hardware (Hughes et al., 2017) and protocols for neonatal diffusion MRI acquisition (Hutter et al. 2018); the use of multiband techniques to accelerate acquisition time combined with approaches to correct motion (Cordero-Grande et al. 2016; Cordero-Grande et al. 2019; Christiaens et al. 2021); and the development of state-of-the-art preprocessing pipelines for neonatal MRI data (Bozek et al., 2018; Christiaens et al. 2018; Bastiani et al. 2019; Fitzgibbon et al. 2020; Makropoulos et al., 2018). Together, these have significantly improved neonatal MRI acquisition and reconstruction methods.

Despite the resultant advances in data quality, studying the neonatal connectome remains challenging. Indeed, many methodological issues hamper the interpretation of the connectome (Sporns, 2013) including the difficulty of detecting origins and termination of connections (Jbabdi and Johansen-Berg, 2011) and returning a high number of false positive streamlines (Maier-Hein et al., 2017). However, the development of state-of-the-art diffusion MRI pipelines have partially addressed some of these issues (Makropoulos et al., 2018; Christiaens et al., 2021), and machine learning approaches are now well suited to discover complex underlying patterns in the structural connectome. Deep neural networks,

which are known for their ability to model complex non-linear multivariate relationships, can help uncover hidden patterns in the connectome, and potentially detect atypical patterns of connectivity in individual subjects.

In adult brain connectivity research, a number of studies have used machine learning and deep learning to study the structural connectome (see (Brown and Hamarneh, 2016) for a summary). Some studies also used similar approaches in neonates: Kawahara and colleagues have developed BrainNetCNN, a Convolutional Neural Network composed of edge to edge, edge to node, and node to graph convolution filters on structural connectivity to predict post menstrual age (PMA) at scan and cognitive performance from the structural connectome (Kawahara et al., 2017). A recent study by Girault and colleagues similarly focused on using the structural connectome at birth to predict cognitive abilities (Mullen score) at age 2 with dense neural networks (Girault et al., 2019). However, little is known about the predictive power of the connectome in a large normative neonatal population such as that of the dHCP.

A promising method in adult and neonatal neuroscience is the study of the "brain maturation index" (also known as "brain delta" or "predicted age difference") corresponding to the apparent age of the subject as compared to the norm (Dosenbach et al. 2010; Cao et al. 2015; Jonsson et al. 2019; Liem et al. 2017; Smith et al. 2019). By training regression models to fit the age of subjects from large normative imaging datasets, we can predict the age of individual subjects and compute the difference between the prediction and subject's true age. This difference gives information about brain maturation and its divergence from the population norm. As such, in adults, a positive (predicted age > true age) difference is interpreted as demonstrating accelerated ageing, and is associated with disorders such as cognitive impairment (Liem et al., 2017), schizophrenia (Koutsouleris et al., 2014) or diabetes (Franke et al., 2013). In neurodevelopment, studying the brain maturation index is therefore extremely relevant for preterm born infants where neurodevelopmental delays and psychiatric disorders often occur (Brown et al., 2017; Galdi et al., 2020; Rasmussen et al., 2017).

In this work, we propose the use of two different machine learning algorithms - Random Forests (RF) and Deep Neural Networks (DNN) - to predict PMA at scan (i.e. brain maturation), and gestational age (GA) at birth (i.e. degree of prematurity), from the neonatal structural connectome in a large sample of neonates scanned at term equivalent age. Additionally, we use sensitivity analysis to obtain feature relevance in the prediction models thereby identifying the connections and brain regions impacting the predictions. Finally, using models obtained for the term-born cohort, brain maturation index was computed for our preterm-born cohort and subsequently correlated to neurodevelopment at 18 months.

2. Methods and Materials

2.1 Participants

All participants were part of the dHCP, approved by the National Research Ethics Service West London committee (14/LO/1169).

524 infants (240 female and 284 male), born between 23^{+0} weeks and 42^{+2} week of gestation, underwent MRI between 37^{+1} weeks and 45^{+1} weeks as part of the dHCP (2nd data release). The participant gestational age at birth (GA) and postmenstrual age at scan (PMA) distribution is presented in Figure 1A-B. Full participant clinical information is presented in Table 1.

The Bayley III Scales of Infant and Toddler Development (BSID-III) (Bayley 2006) were collected at 18 months corrected age and available for 314 infants including 50 preterm-born infants. We used scores of motor (fine and gross), communication (expressive and receptive) and cognitive (raw) score. Assessments were carried out by experienced paediatricians or psychologists. Detailed assessment distributions are presented in Table 1.

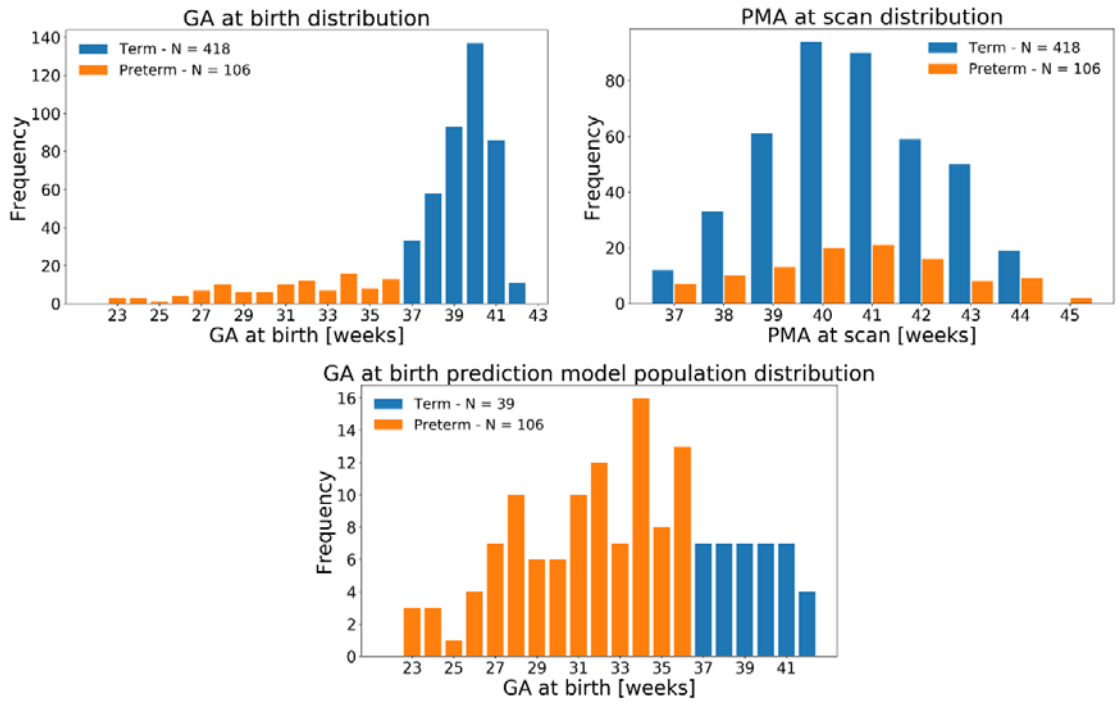


Figure 1. Distribution of (A) GA at birth and (B) PMA at scan of full cohort (N=524). (C) GA at birth of cohort used for predicting GA at birth.

2.2 MRI acquisition

All scans were collected in the Evelina Newborn Imaging Centre based on the Neonatal Intensive Care Unit, St Thomas hospital London using a Philips Achieva 3T scanner (Best, NL) and the dHCP neonatal brain imaging system which includes a 32 channel receive neonatal head coil (Rapid Biomedical GmbH, Rimpar, DE) (Hughes et al., 2017). Informed written parental consent was obtained prior to imaging. Positioning of all infants was done with a lightweight protective “shell”, which was positioned on an MRI safe trolley to ease

transportation. Immobilization of the infants in the shell was done using bead filled inflatable pads (Pearltec, Zurich, CH). In addition to the pads, acoustic protection included earplugs moulded from a silicone-based putty (President Putty, Coltene Whaledent, Mahwah, NJ, USA) placed in the external auditory meatus and neonatal earmuffs (MiniMuffs, Natus Medical Inc, San Carlos, CA, USA). To avoid sudden sound changes which might wake up the infant, the MRI software was modified in order to gradually increase the noise from 0 to the average operating point (Hughes et al., 2017). All scans were supervised by a paediatrician or neonatal nurse experienced in MRI procedures; vital signs including pulse oximetry, temperature and electrocardiography data were monitored throughout data acquisition. All infants were scanned during natural unsedated sleep following feeding.

T2-weighted images were acquired using a Turbo spin echo sequence with parameters TR = 12s and TE = 156ms, SENSE factor 2.11 (axial) and 2.54 (sagittal) with overlapping slices (resolution = 0.8 × 0.8 × 1.6 mm^3). Super-resolution methods (Kuklisova-Murgasova et al., 2012) as well as motion correction methods (Cordero-Grande et al., 2018) were combined to maximise precision and resolution of T2-weighted images (resolved to 0.8 × 0.8 × 0.8 mm^3). Diffusion weighted imaging was acquired in 300 directions with parameters TR = 3.8s, TE = 90ms, SENSE factor 1.2, multiband factor 4, resolution = 1.5 × 1.5 × 3mm^3 (with 1.5mm slice overlap), diffusion gradient encoding: b=0 s/mm (n=20), b=400 s/mm (n=64), b=1000 s/mm (n=88), b=2600 s/mm (n=128), and using interleaved phase encoding (Hutter et al., 2018b).

2.3 Pre-processing and connectome generation

Tissue segmentation of T2-weighted volumes was performed using a neonatal specific segmentation pipeline (Makropoulos et al., 2014) and template (Schuh et al. 2018). Parcellation of 90 cortical and subcortical regions (Shi et al., 2011) adapted to the dHCP weekly age-dependant high-resolution bespoke template (Schuh et al. 2018) was propagated to each subject's T2w native space through non-linear registration based in a diffeomorphic symmetric image normalization method (SyN) available in ANTS software (Avants et al., 2011), using T2w contrast and tissue segmentation as input channels. Tissue maps and atlas parcellation were propagated from each T2w native space to each subject's diffusion native space with a rigid registration using b=0 volumes as target. All rigid registrations were performed with IRTK software (Schnabel et al. 2001). Details of the 90 cortical and subcortical regions are presented in Supplementary Table 1.

Diffusion MRI was reconstructed at an effective resolution of 1.5mm isotropic and denoised using a patch-based estimation of the diffusion signal based on random matrix theory (Veraart et al., 2016). Gibbs ringing was suppressed (Kellner et al., 2016) and B0 field map estimated from b=0 volumes in order to correct magnetic susceptibility-induced distortion using FSL Topup (Andersson et al., 2003). Data was corrected for slice-level motion and distortion in a data-driven q-space representation using a bespoke spherical harmonics and radial decomposition (SHARD) basis of rank 89 corresponding to spherical harmonics of order lmax=0,4,6,8 for each respective shell, with registration operating at a reduced rank of 22 (Christiaens et al., 2018). DWI intensity inhomogeneity field correction was performed using the ANTs N4 algorithm (Tustison et al., 2010). Tools and pipelines implemented in MRtrix3 (Tournier et al., 2019) were used for quantitative analysis of the diffusion MRI data. Developing neonatal brain tissue undergoes rapid changes in cellular properties and water content that can be to a first approximation captured by a non-negative linear combination

of anisotropic signal from relatively mature WM and from isotropic free fluid (Pietsch et al., 2019). We use data from 20 healthy full term control babies from our sample to extract a set of two representative WM (Tournier et al., 2013) and fluid-like (Dhollander et al., 2016, Dhollander et al., 2018) signal fingerprint (*response functions*) that are used to deconvolve each subject's diffusion signal into a fibre orientation distribution (FOD) image, capturing WM-GM-like signal, and scalar fluid density image using the multi-tissue multi-shell constrained spherical deconvolution technique (Jeurissen et al. 2014). Residual intensity inhomogeneity was corrected, and component densities calibrated using a multi-tissue log-domain intensity normalisation (Raffelt, et al. 2017). Resulting normalised WM-GM-like FODs were used to generate 10 million streamlines with an anatomically constrained probabilistic tractography (ACT) (Smith et al., 2012) with biologically accurate weights (SIFT2) (Smith, et al. 2015). The fibre density SIFT2 proportionality coefficient (μ) for each subject was obtained to achieve inter-subject connection density normalisation. The structural connectivity network of each infant was then constructed by calculating the μ × SIFT2-weighted sum of streamlines connecting each pair of regions (thus built as a symmetric adjacency matrix of size 90x90).

In addition, we used 73 structural connectivity matrices obtained from an independent dataset (Batalle et al., 2017) to test the design of the initial hyperparameters and architecture for predictive algorithms presented in sections 2.4.3 and 2.4.4.

2.4 Prediction of age at scan and age at birth

All analyses on this section were performed using Python 3.7. The machine learning library Scikit Learn (Pedregosa et al., 2011) was used for training the RF algorithm. The deep learning framework Keras (version 2.0.3) (Chollet et al., 2015) was used to train the deep learning models.

2.4.1 Feature set

As the structural connectome is presented as a symmetric adjacency matrix (in our case of size 90x90, with 90 brain regions) the lower triangle of the matrix contains all information. We thus extracted and reshaped the lower triangle of each subject's structural connectome S_i as a 1D vector X_i with number of connectivity elements n = 4005, thus leading to the ensemble X of connectivity vectors across N subjects:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_N \end{bmatrix}, with\ X_i = \left[x_{i,1}, x_{i,2}, \dots, x_{i,n}\right];\ x_{i,j} \in \mathbb{R}^+\ and\ n = 4005$$

We normalized each data point across the training sets, and normalized the testing set with the training normalization values using a min-max normalization:

$$\tilde{x}_{i,j} = \frac{x_{i,j} - Min(X)}{Max(X) - Min(X)}\ with\ Min(X)\ and\ Max(X) \in \mathbb{R}^+$$

Thus, assuming that testing data also falls between previous ranges, our training and testing data has the following form:

$$\tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \dots \\ \tilde{X}_N \end{bmatrix}, with\ \tilde{X}_i = \left[\tilde{x}_{i,1}, \tilde{x}_{i,2}, \dots, \tilde{x}_{i,n}\right];\ \tilde{x}_{i,j} \in [0,1]\ and\ n = 4005$$

2.4.2 Prediction models

We carried a set of predictions of demographic information on different population samples using different regression algorithms. In each case, we fitted a regression model f to predict a variable Y representing demographic information (e.g. GA at birth or PMA at scan) for subjects in our dataset. We thus had prediction *Y'* as follows:

$$Y' = f(\tilde{X});\ with\ Y' = [y_1', y_2', \dots, y_N']^T$$

We computed the regressor f that minimizes $| Y' - Y |^2$. We used two different supervised machine learning regression algorithms to do this: RF and DNN.

2.4.3. Random Forests regression

RF are an ensemble learning method for classification and regression based on constructing a multitude of decision trees *(weak learners)* which are individually trained through the technique of "bagging". RF makes predictions by averaging the prediction of each individual tree, hence acting as a *strong learner* (Breiman L., 2001). For optimal performance, two main hyperparameters should be tuned: the number of trees (estimators) in the forest and the maximum depth of each tree. The number of trees determines the smoothness of the decision boundary, and the depth corresponds to the maximum number of levels allowed for each tree. RF regressors' performance often depends on finding the optimal value for these to ensure that there is no overfitting or underfitting.

Here we use the RF regressor from the Scikit Learn *RandomForestRegressor* implementation (Pedregosa et al., 2011). The RF were trained using mean squared error (MSE) as loss function. Hyperparameters were tuned separately for the PMA at scan and GA at birth prediction by performing a grid search on a set of 73 structural connectomes from an independent dataset (Batalle et al., 2017). This allowed us to choose optimal parameters without overfitting our model to the studied data. Hyperparameters used are presented in sections 2.4.6 and 2.4.7.

2.4.4. Deep Neural Networks regression

Deep (Fully Connected) Neural Networks (DNN) are universal function approximators whose parameters can be trained to model complex nonlinear relationships between features and labels via backpropagation (Rumelhart et al., 1986). However, the performance of a DNN also depends on the hyperparameters: design choices are mainly related to the architecture of the network (the layer types, number of layers and number of nodes per layers, activation functions at each layer), the loss function, and the training method (the number of epochs, the optimization function and its parameters).

The DNN in this work were implemented using the deep learning library Keras (version 2.0.3) (Chollet et al., 2015). As performing a grid search to find the best model hyperparameters is computationally expensive when training DNN, we started with a basic architecture built from previous work and common DNN knowledge (Smith 2018), and subsequently optimized these via manual refinement architecture search. To avoid overfitting the model to the data used in this paper, this was done on the same set of 73 structural connectomes from an independent sample as was used for the RF training, independently for the GA at birth and PMA at scan prediction. For both prediction tasks, the models were trained using MSE as a loss function and the Adam optimizer (Kingma and Ba, 2015), albeit with different learning rates. Further details on the network architecture are included in sections 2.4.6 and 2.4.7.

2.4.5 Training and evaluation of the models

To assess the performance of the prediction models, the evaluation metric was calculated on test data excluded from training and hyperparameter tuning. We split the dataset into k groups (folds) and fit the model k times. Each time, one group is used to evaluate performance, while the rest of the groups are used for training and validation. The evaluation scheme is presented in Figure 2B.

We split the data into k=5 groups (folds), with 20% of data used for testing at each fold. The remaining 80% of the data were further split for training (65%) to fit the models and validation (15%) to tune the hyperparameters. Min-max normalisation presented in section 2.4.1 is fitted on the training/validation set, where normalization parameters are saved and applied to the test set.

We added a bias correction as previously described (Smith et al. 2019; Peng et al. 2019) to correct age dependency of the training residuals. Briefly, we used a linear model $Y' = f(X) = \alpha Y + \beta$ to obtain an unbiased estimate of Y' as $\hat{Y} = \frac{Y' - \beta}{\alpha}$, where the parameters α and β are estimated during training (on both the combination of training and validation set) and are thus applied directly to the test set. We obtained our final corrected prediction $\hat{Y}_i$ for each structural connectome as follows:

$$\hat{Y}_i = \frac{f(\tilde{X}_i) - \beta}{\alpha} = \frac{Y_i' - \beta}{\alpha} = F(\tilde{X}_i); \quad \alpha, \beta \in \mathbb{R}$$

The final performance is calculated by averaging test-set performance over the 5 folds. We used mean absolute error (MAE) as our evaluation metric, calculated on each test set *k* as follows:

$$MAE_k = \frac{1}{N_k} \sum_{\forall i \in S_k} \left| Y_i - \hat{Y}_i \right|$$

Where N_k is the number of subjects belonging to test set *k* (S_k) and Y_i and $\hat{Y}_i$ are actual and predicted outcome of subject *i*. In addition, we also evaluate MSE and R^2 scores for each test set k, which are calculated as follows:

$$MSE_k = \frac{1}{N_k} \sum_{\forall i \in S_k} \left(Y_i - \hat{Y}_i\right)^2$$

$$R_k^2 = 1 - \frac{\sum_{\forall i \in S_k} \left(Y_i - \hat{Y}_i\right)^2}{\sum_{\forall i \in S_k} (Y_i - \bar{Y})^2}$$

Where $\bar{Y}$ is the mean actual age of test set k. We also calculated Pearson's Correlation (r_k) and p-value (p_k) between actual (Y) and predicted output $(\hat{Y})$ for each test set. As we obtain a prediction for every subject (albeit with different models) we can compute the MAE_{tot}, MSE_{tot}, R_{tot}^2, r_{tot} and p_{tot} by considering the predictions of the 5 test sets (see Figure 2A). Finally, we assessed the presence of heteroscedasticity in our predictions by comparing the variance σ_{err}^2 of the absolute error – a lower variance signifies more homoscedastic predictions.

2.4.6 Prediction of PMA at scan in term-born infants

To build a model of typical development of connectivity we used the full cohort of 418 term-born babies (GA at birth >= 37) with PMA at scan between 37 and 45 weeks.

We first predicted PMA at scan from the vectorised and normalized structural connectome $\tilde{X}$ using RF regressor model. Optimal parameters of the model (max depth = 250, number of estimators = 30) were found by performing a grid search in an independent dataset (see section 2.4.3). We trained each fold on N≈335 samples (80%) including a validation set. We then tested the model on the remaining set (N≈83, 20%) in each fold, thus being able to predict age at scan on all 418 structural connectomes of term infants (see Figure 2A).

In a similar fashion, we also trained a regression DNN to predict PMA at scan from the vectorised and normalized structural connectome $\tilde{X}$. This DNN comprises one input layer with 4005 input nodes, 7 hidden layers, 6 activation layers (ReLu), 5 batch normalisation layers and one output layer with one node. Training was done for 50 epochs with learning rate of 0.007 and remaining parameters with default value. Detailed structure of the architecture of this DNN is provided in Figure 2D.

We applied the previously described bias correction method on both DNN and RF, by fitting α and β for each model f_k using both the training and validation set; thus reaching 5 distinct models $F_1, F_2, \ldots, F_5$ for both the DNN and RF methods.

2.4.7 Prediction of GA at birth

To assess the effect of preterm birth on structural connectivity we trained a prediction model for GA at birth from $\tilde{X}$ with both DNN and RF in a similar fashion as previously described for prediction of PMA at scan.

Since the dHCP cohort has significantly more term-born than preterm-born infants, there is a "class imbalance" in the GA distribution that may skew the model prediction. We therefore randomly selected a sub-sample of term subjects that had, on average, equal

density of subjects on each GA at birth weekly bin. Our 106 preterm infants were distributed in 15 different GA at birth bins (22w-23w; 23w-24w … 36w-37w), thus providing an average of 7 infants per age category. We kept all 106 preterm infants and randomly sampled 7 infants from each of the term age categories (37w-38w, 38w-39w, … 41w-42w) and the 4 subjects born 42w-43w (as only 4 were available between 42w and 43w), for a total of 39 term infants. This resulted in a total of 145 infants with a balanced distribution (see Figure 1C).

We first attempted prediction of GA at birth from the vectorised and normalized structural connectome $\tilde{X}$ using RF with optimal parameters max depth = 300, number of estimators = 50. For each fold, we trained using N≈116 samples (80%) including a validation set. We then tested the model on the remaining set (N≈29, 20%) in each fold, predicting GA at birth for all 145 structural connectomes considered.

In a similar fashion, we also trained a DNN to predict GA at birth. This DNN consists of one input layer, 6 hidden layers, 6 activation layers (ReLu), one dropout layer, and one output layer. 120 epochs were used for training, with learning rate 0.003 and remaining parameters at default value. Detailed information on the architecture is provided in Figure 2E.

We applied the bias correction method on both DNN and RF by fitting α and β for each model h_i from the validation set; thus reaching 5 distinct models $H_1, H_2, \ldots, H_5$ for both the DNN and RF methods.

2.4.8 Brain maturation index

We defined brain maturation index δ (also called brain age or predicted age difference in the literature) as the difference between the predicted age $\hat{Y}$ and true age Y of a subject n (Dosenbach et al., 2010): $\delta_i = \hat{Y}_i - Y_i$

We developed a model of typical brain development by training 5 models to predict PMA at scan on term-born infants only (section 2.4.6). Prediction of PMA at scan for each preterm subject was computed by taking the mean of the predictions from each of the 5 DNN trained models F_k from each cross-validation partition:

$$\hat{Y}_i = \frac{1}{5}\sum_{k=1}^{5} F_k\left(\tilde{X}_i\right) = G\left(\tilde{X}_i\right)$$

Following this, we computed the brain maturation index δ_i of each preterm subject:

$$\delta_i = \hat{Y}_i - Y_i = G\left(\tilde{X}_i\right) - Y_i$$

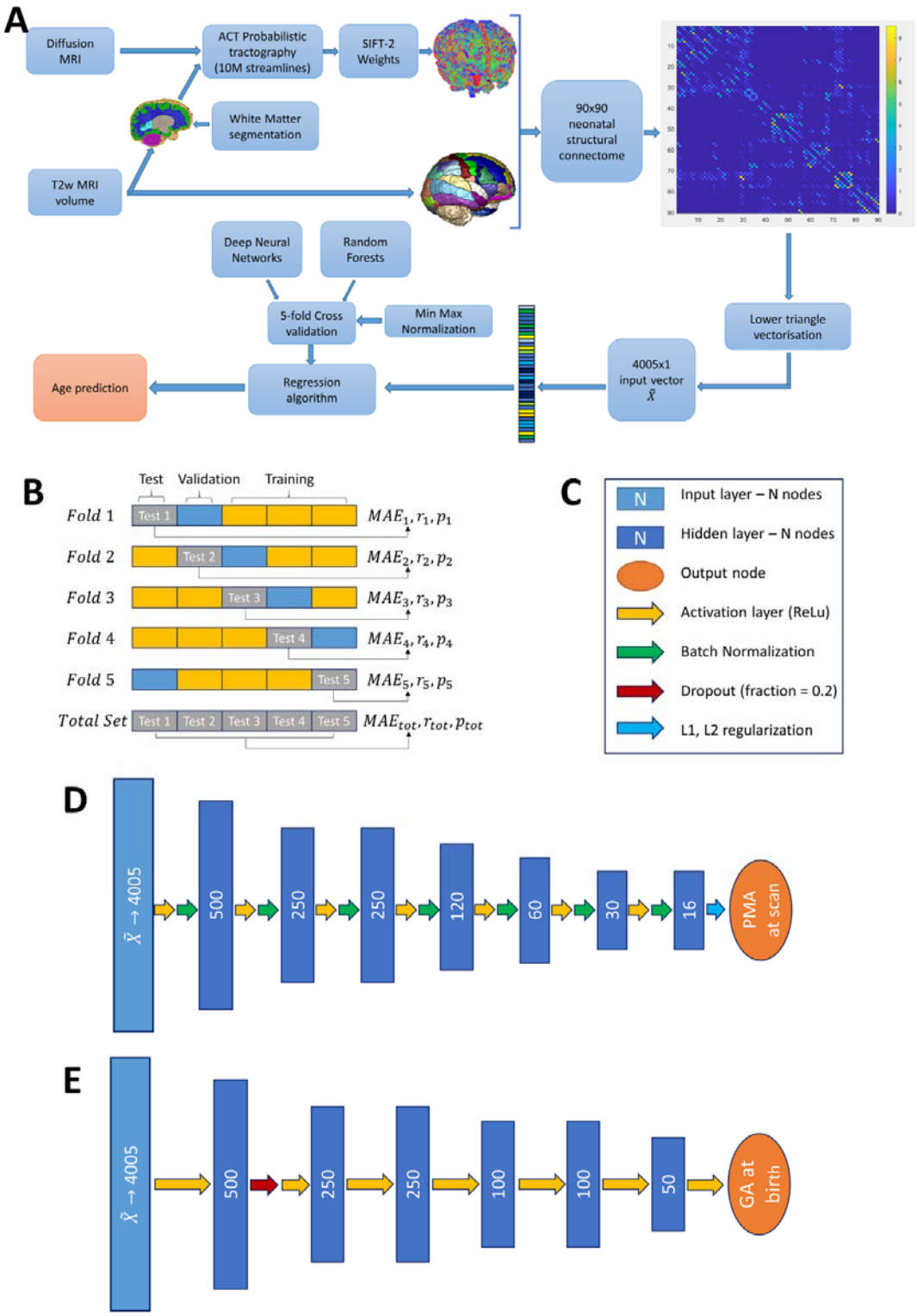


Figure 2. (A) Pipeline for age prediction from MRI. (B) Cross Validation protocol. (C) Legend for DNN architecture components. (D) PMA at scan DNN architecture. (E) GA at birth DNN architecture.

2.5 Sensitivity analysis to obtain feature relevance

The interpretation of models was computed using feature sensitivity analysis (Saltelli 2002) in a similar fashion for both the RF and DNN models on both predictions of GA and PMA (thus obtaining 4 different interpretations or "relevance maps"). We iteratively zeroed out every one of the 4005 connections and computed the magnitude of the change in predictions when zeroing out each feature compared to prediction from the unaltered connectome. The higher the magnitude of the change, the more impacted the model is to the given feature in the computing predictions. For each model, this was computed across all subjects and all folds; we took the mean magnitude of change across all folds as final feature of importance. We thereby obtained for each model features of importance for each of the 4005 connections. Following this step, we were also able to identify the most relevant brain regions, by summing the relevance across each column of each of the 4 relevance maps obtained. We used BrainNetViewer (Xia et al, 2013) to visualize the different relevance maps.

2.6 Statistical methods

Differences between term and preterm cohorts on all relevant characteristics were assessed by computing a two tailed independent t-test or chi-square test as appropriate. The association between brain maturation index δ_i and neurodevelopmental outcomes was assessed with Pearson's Correlation coefficient for all preterm infants having both δ_i and BSID-III developmental outcomes at age 18 months corrected age. All outcomes were corrected for socio economic status, captured by the English Index of Multiple Deprivation (IMD) Rank. The IMD is a geographical measure that summarizes information from 38 different factors such as income, employment, education, crime rates and health situation for all postcode areas in England (Index of Multiple Deprivation, 2015). A lower IMD rank indicates a lower level of deprivation. We used the mother's postcode at the time of birth to calculate this. All p-values presented are uncorrected for multiple comparisons

2.7 Data availability

The imaging and collateral data from the dHCP can be downloaded by registering at https://data.developingconnectome.org/

Structural connectivity networks and code used to predict age at birth and age at scan are available in https://github.com/CoDe-Neuro/Predicting-age-and-clinical-risk-from-the-neonatal-connectome

3. Results

3.1 Sample characteristics

There were no significant differences in PMA at scan and male/female proportion between term and preterm neonates in this study. For the subjects for which 18 months BSID-III follow-up neurodevelopmental assessment was available, there were no significant differences in outcomes between term and preterm infants. However, a significant group difference (p=0.002) was found in IMD scores, with term infants showing significantly higher deprivation than preterm infants. Detailed cohort characteristics are provided in Table 1 and Figure 1. Neurodevelopmental outcome details are provided in Table 1.

Table 1. Detailed sample and outcome characteristics.

	Term Born N = 418	Preterm born N = 106	p-value*
Gestational age at birth [weeks $^{+days}$]	Median = 40^{+1} IQR = 39^{+0} – 40^{+6}	Median = 32^{+2} IQR = 28^{+5} – 34^{+4}	<0.001
Postmenstrual age at scan [weeks $^{+days}$]	Median = 41^{+0} IQR = 39^{+6} – 42^{+2}	Median = 41^{+0} IQR = 39^{+5} – 42^{+2}	0.771
Sex, no. of female (%)	195 (47%)	45 (42%)	0.439
BSID-III, no. (% of total population)	322 (77%)	77 (73%)	0.343
Gestational age at birth (weeks $^{+days}$) of subjects with BSID-III data available	Median = 39^{+6} IQR = 39^{+1} – 40^{+5}	Median = 31^{+6} IQR = 28^{+4} – 34^{+5}	< 0.001
Postmenstrual age at scan (weeks $^{+days}$) of subjects with BSID-III data available	Median = 41^{+1} IQR = 39^{+6} – 42^{+2}	Median = 41^{+1} IQR = 39^{+5} – 42^{+3}	0.957
IMD Score of subjects with BSID-III data available	Mean = 13839 STD = 7387	Mean = 16465 STD = 7805	0.005
Fine Motor	Mean = 20.69 STD = 3.12	Mean = 20.22 STD = 3.36	0.241
Gross Motor	Mean = 17.34 STD = 2.37	Mean = 16.75 STD = 2.40	0.052
Cognitive Score	Mean = 10.08 STD = 2.15	Mean = 9.79 STD = 2.59	0.303
Expressive communication	Mean = 17.26 STD = 3.62	Mean = 16.89 STD = 4.03	0.434
Receptive Communication	Mean = 18.36 STD = 4.29	Mean = 18.51 STD = 5.26	0.800

*p-values computed with two tailed independent t-test or chi-square test as appropriately.

3.2 Prediction of PMA at scan

3.2.1 Prediction of PMA at scan with RF

We trained a RF regressor to fit for PMA at scan from vectorised and normalized structural connectome $\tilde{X}$ on term infants only. We obtained $MAE_{tot} = 0.84$ weeks, $MSE_{tot} = 1.10$, R^2_{tot} = 0.61, $\sigma^2_{err} = 1.10$ with correlation between true and predicted $r_{tot} = 0.79$ $p_{tot} < 0.001$. Figure 3A shows true PMA vs predicted PMA on each of the 5 cross-validation folds. Detailed results of each fold are presented in Figure 3C.

3.2.2 PMA at scan prediction with DNN

Similarly, we trained a DNN regressor to fit for PMA at scan from vectorised and normalized structural connectome $\tilde{X}$ on term infants only. We obtained $MAE_{tot} = 0.72$ weeks, $MSE_{tot} = 0.94$, R^2_{tot} = 0.67, $\sigma^2_{err} = 0.94$ with correlation between true and predicted $r_{tot} = 0.83$ and $p_{tot} < 0.001$. Figure 3B shows true PMA vs predicted PMA on each of the 5 cross-validation folds. Detailed results of each fold are presented in Figure 3D.

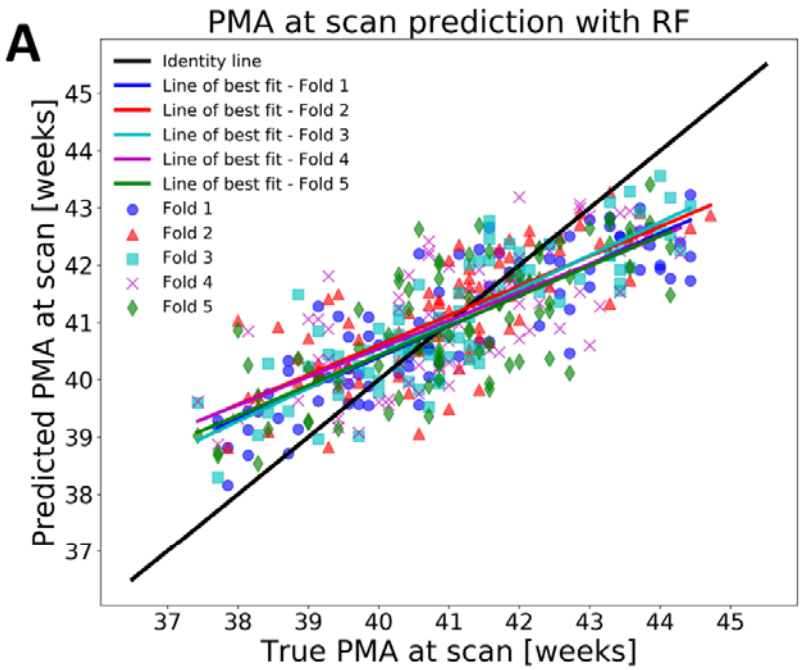
3.3 Prediction of GA at birth

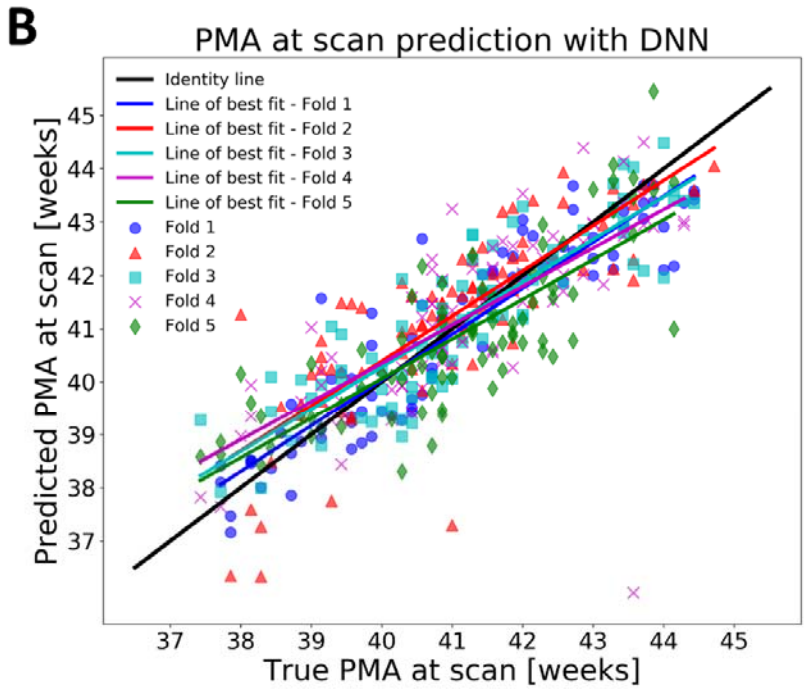
3.3.1 Prediction of GA at birth with RF

We trained a RF regressor to fit GA at birth from vectorised and normalized structural connectome $\tilde{X}$ on balanced data (145 infants). We obtained $MAE_{tot} = 2.76$ weeks, $MSE_{tot} = 12.95$, R^2_{tot} = 0.43, $\sigma^2_{err} = 12.93$ with correlation between true and predicted $r_{tot} = 0.67$ and $p_{tot} < 0.001$. Figure 4A shows true GA at birth vs predicted GA at birth on each of the 5 folds fold. Detailed results of each fold are presented in Figure 4B.

3.3.2 Prediction of GA at birth with DNN

Similarly, we trained a DNN from vectorised and normalized structural connectome $\tilde{X}$ on balanced data. We obtained $MAE_{tot} = 2.21$ weeks, $MSE_{tot} = 8.90$, R^2_{tot} = 0.61, $\sigma^2_{err} = 8.86$ with correlation between true and predicted $r_{tot} = 0.82$ and $p_{tot} < 0.001$. Figure 4B shows true GA vs predicted GA on each of the 5 folds. Detailed results of each fold are presented in Figure 4D.





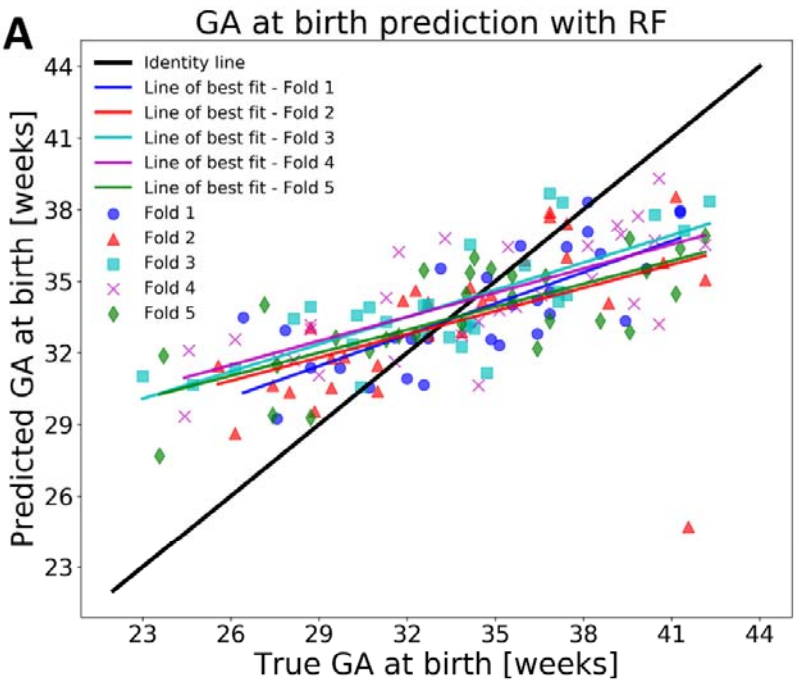
C

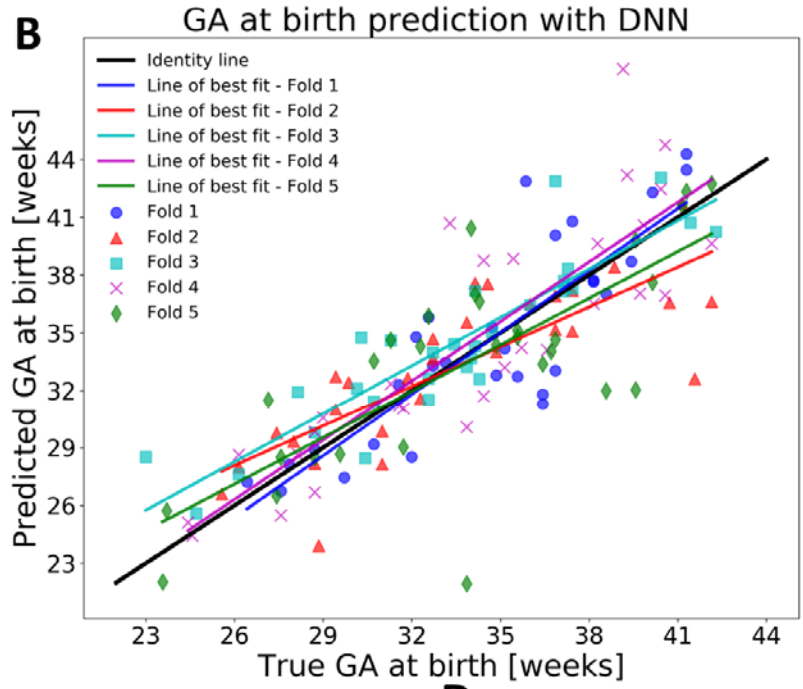
Fold by fold results - PMA Corrected Prediction with RF					
	MAE	MSE	R^2	Correlation	p value
Fold 1	0.8687	1.1703	0.6872	0.8637	<<0.001
Fold 2	0.8111	1.0505	0.5766	0.7714	<<0.001
Fold 3	0.7581	0.8899	0.6707	0.8307	<<0.001
Fold 4	0.8800	1.2175	0.5213	0.7238	<<0.001
Fold 5	0.8767	1.1925	0.5384	0.7372	<<0.001
Total	0.8387	1.1036	0.6083	0.7886	<<0.001

D

Fold by fold results - PMA Corrected Prediction with DNN					
	MAE	MSE	R^2	Correlation	p value
Fold 1	0.6227	0.6036	0.8387	0.9180	<<0.001
Fold 2	0.8014	1.0740	0.5671	0.8066	<<0.001
Fold 3	0.6606	0.6832	0.7472	0.8665	<<0.001
Fold 4	0.7368	1.3695	0.4615	0.7285	<<0.001
Fold 5	0.7769	0.9875	0.6178	0.8068	<<0.001
Total	0.7195	0.9424	0.6655	0.8276	<<0.001

Figure 3. Detailed results of prediction of PMA at scan on term cohort. True vs Predicted with (A) RF, (B) DNN. Fold by fold result with (C) RF, (D) DNN.





C

Fold by fold results - GA Corrected Prediction with RF					
	MAE	MSE	R^2	Correlation	p value
Fold 1	2.2933	8.2487	0.5007	0.7446	<<0.001
Fold 2	2.5409	17.026	0.2226	0.4984	0.00593
Fold 3	2.6691	10.707	0.4854	0.7570	<<0.001
Fold 4	3.2971	14.605	0.4581	0.7264	<<0.001
Fold 5	3.0124	14.141	0.4454	0.7299	<<0.001
Total	2.7626	12.946	0.4255	0.6670	<<0.001

D

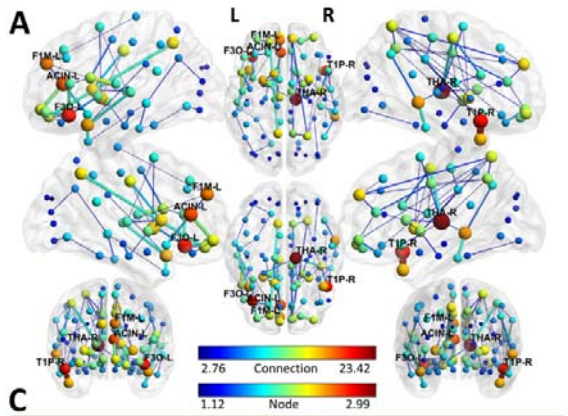
Fold by fold results - GA Corrected Prediction with DNN					
	MAE	MSE	R^2	Correlation	p value
Fold 1	2.1004	7.1552	0.5669	0.8536	<<0.001
Fold 2	2.1143	8.0324	0.6332	0.8000	<<0.001
Fold 3	1.6632	5.3014	0.7452	0.8958	<<0.001
Fold 4	2.4973	10.3184	0.6172	0.8611	<<0.001
Fold 5	2.6632	13.6840	0.4634	0.7544	<<0.001
Total	2.2077	8.8983	0.6051	0.8195	<<0.001

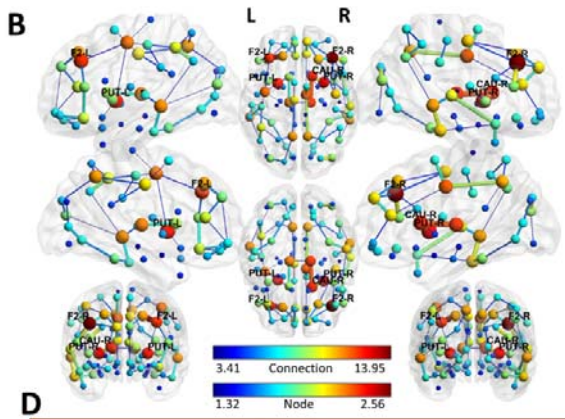
Figure 4. Detailed results of prediction of GA at birth. True vs Predicted with (A) RF, (B) DNN. Fold by fold result with (C)RF, (D)DNN.

3.3 Sensitivity analysis

We computed sensitivity analysis on each of the 4 models as described in the methods section, thereby obtaining 4 distinct connection relevance maps. From the relevance maps, we identified the 5 most relevant connections and brain regions (nodes) from each model, which are shown in Figure 6 C, D, G and H. Detailed relevance maps are shown in Figures 6 A, B, E and F). For visualization purposes, we computed and showed the z-score of each connection and region relevance instead of the raw relevance (magnitude of change) computed and show the 80 connections with higher relevance for each model (~2 centile).

Visualisation of relevant connections shows different patterns across models, though similar types of connections and regions were detected as important. The thalamus, temporal lobe, frontal cortex, cingulate gyrus and putamen regions were identified as particularly important regions in all four models and were also involved in relevant connections. Connections within the frontal cortex, and between the frontal cortex and temporal and parietal lobe, cingulate and precuneus were highly relevant. Though not statistically significant, right hemisphere nodes were on average more relevant across all 4 models.



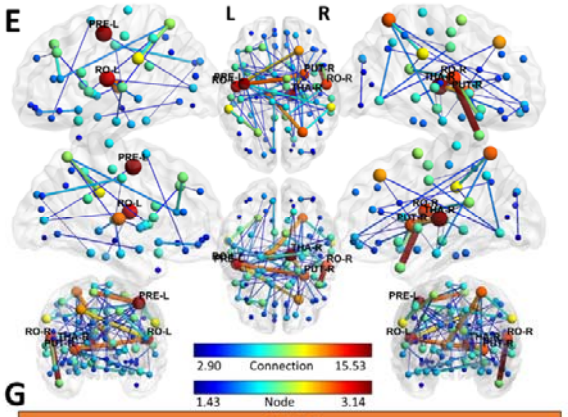


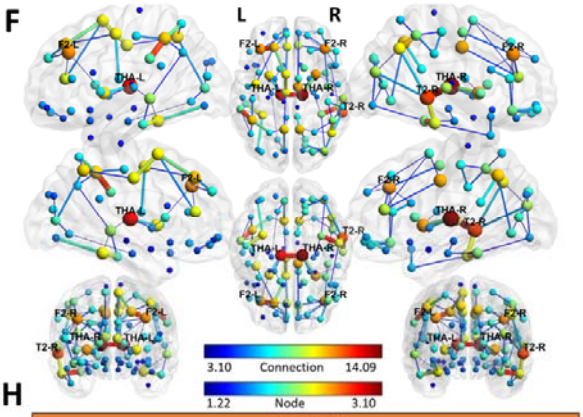
C

PMA RF		
Most important connections		Most important Nodes
Temporal pole: superior temporal gyrus- Right	Temporal pole:middle temporal gyrus- Right	Thalamus - Right (**THA-R**)
Inferior frontal gyrus,opercular part - Left	Rolandic operculum - Left	Temporal pole: superior temporal gyrus- Right (**T1P-R**)
Superior frontal gyrus, orbital- Right	Middle frontal gyrus,orbital part- Right	Inferior frontal gyrus,orbital part - Left (**F3O-L**)
Superior frontal gyrus,medial orbital - Left	Anterior cingulate and paracingulate gyri - Left	Anterior cingulate and paracingulate gyri-Left (**ACIN-L**)
Inferior frontal gyrus,triangular part - Left	Inferior frontal gyrus,orbital part - Left	Superior frontal gyrus, medial - Left (**F1M-L**)

D

PMA DNN		
Most important connections		Most important Nodes
Superior Frontal Gyrus, Dorsolateral - Left	Middle frontal gyrus - Left	Middle frontal gyrus - Right (**F2-R**)
Caudate nucleus- Right	Thalamus- Right	Lenticular nucleus, putamen- Right (**PUT-R**)
Middle frontal gyrus - Right	Inferior frontal gyrus,triangular part- Right	Lenticular nucleus, putamen - Left (**PUT-L**)
Superior occipital gyrus- Right	Middle occipital gyrus- Right	Caudate nucleus - Right (**CAU-R**)
Middle temporal gyrus- Right	Inferior temporal gyrus- Right	Middle frontal gyrus - Left (**FR-L**)





G

GA RF		
Most important connections		Most important Nodes
Rolandic operculum- Right	Temporal pole:middle temporal gyrus- Right	Thalamus- Right (**THA-R**)
Precentral gyrus - Left	Supplementary motor area- Right	Precentral gyrus - Left (**PRE-L**)
Rolandic operculum - Left	Lenticular nucleus, putamen- Right	Rolandic operculum - Left (**RO-L**)
Superior parietal gyrus- Right	Thalamus - Left	Rolandic operculum- Right (**RO-R**)
Lenticular nucleus, putamen- Right	Thalamus- Right	Lenticular nucleus, putamen- Right (**PUT-R**)

H

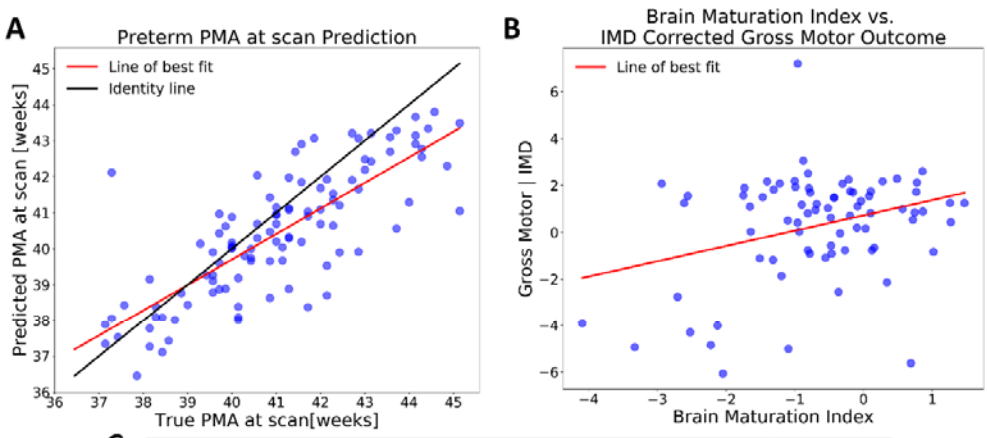
GA DNN		
Most important connections		Most important Nodes
Superior temporal gyrus- Right	Middle temporal gyrus- Right	Thalamus- Right (**THA-R**)
Thalamus - Left	Thalamus- Right	Lenticular nucleus, putamen - Right (**PUT-R**)
Fusiform gyrus- Right	Inferior temporal gyrus- Right	Middle temporal gyrus- Right (**T2-R**)
Inferior parietal, but supra-marginal and angular gyri-Left	Supramarginal gyrus - Left	Middle frontal gyrus - Left (**F2-L**)
Superior Frontal Gyrus, Dorsolateral - Left	Middle frontal gyrus - Left	Middle frontal gyrus - Right (**F2-R**)

Figure 5. Relevance of connections and nodes for predictions of PMA with RF (A) and DNN (B), as well as GA with RF (E) and DNN (F); colour scale of each connection and region is the z-value of the computed relevance. Tables showing five most relevant connections and regions for predictions of PMA with RF (C) and DNN (D), as well as GA with RF (G) and DNN (H).

3.4 Brain Maturation Index

We computed the predicted PMA at scan of each preterm-born infant scanned at term-equivalent age by averaging out the 5 predictions from the DNN term trained models. We obtained MAE of 1.01 weeks on prediction of 106 preterm infants ($MSE_{tot} = 2.24$, R^2_{tot} = 0.42, $\sigma^2_{err} = 1.45$), with correlation between true and predicted age between $r = 0.79$, ($p < 0.001$). True vs predicted age is presented in Figure 6A.

We computed the brain maturation index δ_i from each prediction. Brain maturation index (δ_i) was significantly correlated with BSID-III gross motor scale at 18 months corrected age ($r = 0.3751$, $p = 0.0008$, Figure 6 B and C).



C

	Outcome	Sample size	Correlation with prediction residuals	
			Correlation	p-value
BSID-III outcomes	Fine Motor	77	0.1073	0.3529
	Gross Motor	**77**	**0.3751**	**0.0008**
	Receptive Communication	77	0.1648	0.1519
	Expressive Communication	77	0.2623	0.0211
	Cognitive scaled score	77	0.2138	0.0618

Figure 6. Association of Brain Maturation Index with BSID-III outcomes in preterm-born infants. (A) True vs predicted PMA at scan for preterm infants. (B) Brain maturation index δ vs BSID-III Gross motor outcome corrected for IMD. (C) Detailed correlation and p-values of preterm brain maturation index and BSID-III outcomes – statistically significant results surviving multiple comparisons correction are highlighted in bold red.

4. Discussion

We have used machine learning to uncover structural brain connectivity patterns associated with trajectories of early life development from high dimensionality neonatal MRI data. This enabled us to accurately predict PMA at scan in a large sample of term-born infants, and accurately predict GA at birth in a sample of preterm infants from scans at term equivalent age. The connectome features driving our predictions are of biological importance: in the preterm cohort we have shown significant correlation between a connectome derived brain maturation index and motor development at 18 months corrected age.

We achieved high accuracy in our prediction of PMA at scan in term born babies, reaching a low MAE of 0.72 weeks ($MSE_{tot} = 0.94$, $R^2_{tot} = 0.67$) and a high correlation between true and predicted age ($r_{tot} = 0.83$; $p_{tot} < 0.001$). While the structural connectome presents several important challenges to study brain connectivity (Campbell and Pike, 2014) including high numbers of false positive streamlines (Maier-Hein et al., 2017), our results suggest that there is reliable information present to capture the subtle changes associated with weekly development. There have been several studies evaluating white matter microstructural and connectivity changes during the first days after birth. Indeed, it has been found that the postnatal period is marked by further dendritic arborization, refinement of existing intracortical connections, and an increase in synaptogenesis which results in an abundance of connections (for a review see (Keunen et al., 2017)). Some important changes have also been found in the structural connectome in the early postnatal period, mainly an increase in integration (the ease with which different brain regions communicate) and segregation (presence of clusters, i.e., capacity for specialised processing) (Batalle et al., 2017). These changes are likely captured by DNN and underlie the accurate prediction of PMA at scan. To our knowledge only a few studies have attempted the prediction of PMA at scan from the structural connectome. Using an RF method Brown and colleagues achieved prediction precision of 1.6 weeks using a sample of moderate and very preterm infants (77 scans, GA at birth between 24 and 32 weeks, scanned between 27 and 45 weeks) (Brown et al., 2017). Kawahara et al used a cohort of 115 preterm infants (born between 24 and 32 weeks gestational age) and reported a MAE of 2.17 weeks and *r* between true and predicted age of 0.87. Although widespread connections contributed to their predictive models, one (between the Right Lingual Gyrus and Fusiform Gyrus) was deemed to be of high importance (Kawahara et al., 2017). In addition to an improved MAE, using a sensitivity analysis approach we have shown a complex pattern of connections used for prediction of PMA (Figure 5A-D). The most important connections identified were thalamocortical or between frontal and temporal lobes, although widespread connections across the whole brain contributed to prediction. This is in keeping with a large body of previous literature on early life brain development: thalamocortical connections are developing late in the third trimester (Batalle et al., 2018), and are sensitive to disruption due to preterm birth or perinatal pathology (Ball et al., 2012, Ball et al. 2013, Batalle et al., 2017). Thalamic maturation continues in the first postnatal days (Kostovic and Jovanov-Milosevic, 2015), and

it is therefore not surprising that inter-individual differences in thalamic connectivity appear to be important in the prediction of PMA at scan. We identified the superior frontal cortex and frontal orbital gyrus as two of the most important nodes in our PMA at scan prediction, which is in line with literature showing that connections to these regions are developing rapidly at term equivalent age (van den Heuvel et al., 2015; Kawahara et al., 2017). Right sided connections and nodes appear slightly more predictive of PMA than those on the left (although this difference was not significant). Structural asymmetry is well established by term equivalent age, although right sided connections are less efficient (Ratnarajah et al., 2013). One previous group has suggested that right sided connections are more important in predictions of neurodevelopment (Girault et al., 2019) and it is reasonable to hypothesize that inter-individual variation in the rate of development may also be important in prediction of PMA.

We have also demonstrated good ability to predict GA at birth from the neonatal connectome (DNN performance of $MAE_{tot} = 2.21\ weeks$; $MSE_{tot} = 8.90$; $R^2_{tot} = 0.61$; $r_{tot} = 0.82$; $p_{tot} < 0.001$). The nodes and connections of greatest importance to our models were mostly in and between the thalamus, frontal, and temporal lobes, and showed a slight right sided dominance (Figure 5E-H). The DNN model outperformed the RF model in prediction of GA at birth (Figures 4 A-D), which may be partially attributed to the difference in the fundamental mechanics of the models used and the high nonlinearity of the data. The effects of low GA at birth on structural brain development are widespread and complex: of particular relevance to our findings preterm birth has been associated with reduced structural (Ball et al., 2013) and functional thalamocortical connectivity (Toulmin et al., 2015). Organisation of fronto-temporal connectivity may also be impaired by preterm birth although the reported pattern of preterm associated changes is heterogeneous, including a higher proportion of cortical-cortical connections (Ball et al., 2014), changes in global network efficiency (Brown et al., 2014; Lee et al., 2019) and decreased fronto-limbic connectivity (de Almeida et al., 2021). It is worth noting that there is considerable overlap in the features we found to be more important for predicting PMA and GA at birth, and the critical regions used for prediction (thalamus, frontal and temporal lobes) are some of the most frequently cited in other studies of early life connectivity (van den Heuvel et al., 2015; Pandit et al., 2015). It may be the case that the rate of development (and thus potential for inter individual difference) is highest for these connections in the neonatal period, which may explain why they are most readily identified by machine learning techniques.

Brain maturation indices, which compare a predicted age to true age are a useful tool to characterise individual variations in the maturational trajectory of brain connectivity (Cao et al., 2015). In adults, a positive δ (predicted age > true age) is viewed as undesirable, and possibly indicative of the emergence of cognitive decline (Jonsson et al., 2019). In neonates however the opposite is likely true - a negative δ (predicted age < true age) may be indicative of developmental delay. We computed δ for all preterm infants (n=50) who had completed an 18-month follow-up visit and found a positive correlation between brain maturation index and BSID-III gross motor score (r=0.375, p<0.001). There is a large body of research on the effects of prematurity on motor development – lower GA at birth is universally associated with lower BSID-III motor scores (Greene et al., 2012; Velikos et al.,

2015; Ahn et al., 2017). There is also some evidence that individual motor development differences in preterm children are in part due to individual brain connectivity profiles – diffusion measures in the cingulum (Schadl et al., 2017), corpus collosum (Lean et al., 2019), and global white matter diffusion tensor imaging measures (Girault et al., 2019) have all been associated with neurodevelopmental outcomes up to 2 years of age. Thalamocortical connectivity, key to the prediction of both PMA at scan and GA at birth in our study, has previously been associated with early life cognitive development (Ball et al., 2015; Eixarch et al., 2016). Two recent studies also using the dHCP dataset have used normative modelling, and specifically individual deviations from population growth trajectories, to associate imaging and outcome (Dimitrova et al., 2020; O'Muircheartaigh et al., 2020). The correlation observed here between our brain maturation index and cognitive development is therefore in keeping with existing knowledge and gives our model biological validity. However, since brain maturation index is obtained as the error in a model trained with term-born infants only, it is reasonable to expect larger model error (more negative δ) linked to reduced GA at birth (extreme prematurity), which is in turn associated with poorer neurodevelopmental outcomes. Hence, it is difficult to disentangle whether brain maturation index is just capturing features linked to the structural connectivity phenotype of preterm babies, or if it is capturing information related to motor outcome at 18 months that is embedded in the structural connectome at birth.

Our analysis uses well established machine learning techniques; however, it is important to be aware of their limitations. We have assessed the performance of RF and DNN for prediction of key developmental characteristics in a large sample of neonates. To our knowledge this is the first application of these techniques in a large cohort of neonates. A previous study predicting PMA from the structural connectome used a convolutional neural network approach (BrainNetCNN) to extract information from edge-to-edge, edge-to-node and node-to-node data from connectomes(Kawahara et al., 2017). The spatial distribution of adjacency matrices is not necessarily reflective of brain region locality and connectivity characteristics, so in this study, when predicting age directly from connectome data we instead chose to use RF and DNN, which don't require data known to have local correlations. Using this approach, we achieved better performance compared to RF on age prediction from the neonatal structural connectome. Prediction of PMA at scan was highly accurate with RF ($MAE_{tot} = 0.84$; $MSE_{tot} = 1.10$; $R^2_{tot} = 0.61$; $\sigma^2_{err} = 1.10$; $r = 0.79$; $p < 0.001$), but DNN achieved better performance ($MAE_{tot} = 0.72$; $MSE_{tot} = 0.94$; $R^2_{tot} = 0.67$; $\sigma^2_{err} = 0.94$; $r = 0.83$; $p < 0.001$), with a more homoscedastic distribution of predictions over each of the 5 cross-validation folds (Figure 3). The improved performance of DNN over RF was more evident for prediction of GA at birth, with a better performance ($MAE_{tot} = 2.21$; $MSE_{tot} = 8.90$; $R^2_{tot} = 0.61$; $\sigma^2_{tot} = 8.86$; $r = 0.82$; $p < 0.001$) and more homoscedastic distribution of predictions on each fold with DNN over RF ($MAE_{tot} = 2.76$; $MSE_{tot} = 12.95$; $R^2_{tot} = 0.43$; $\sigma^2_{tot} = 12.93$; $r = 0.67$; $p < 0.001$) (Figure 4A-D). Another important consideration in machine learning MRI studies is prediction bias. As it can be observed in Figures 3A-B and 4A-B, prediction bias is still present after correction was applied (as described in section 2.4.5): since we fitted the bias correction values on the training set and blindly applied to the respective test sets (to

prevent overfitting), the bias cannot be perfectly corrected. This is especially true for Random Forests – thus showing the higher stability and performance of DNN on unseen data. We chose to downsample the term cohort to achieve balance between age categories. Although this reduced our sample size, it allowed us to avoid a class imbalance problem, which could have caused a systematic positive bias for preterm infants (predicted GA > true GA). The relatively good performance of our model suggests that the impact of preterm birth on brain connectivity development is important and clearly apparent in the neonatal structural connectome. We achieved high performance in the prediction of PMA at scan in our preterm cohort by averaging the predictions from all 5 term trained DNN models ($MAE_{tot} = 1.01$; $MSE_{tot} = 2.24$; $R^2_{tot} = 0.42$; $\sigma^2_{tot} = 1.45$; $r = 0.79$; $p < 0.001$). Although there was high accuracy in prediction and correlation between true and predicted age for preterm infants, predictions were on average inferior to those obtained for PMA at scan for term infants (Figure 6A). This is expected as preterm neonates are known to have specific differences in structural connectivity when compared with their term counterparts which may have reduced the generalizability of the predictive model (Ball et al., 2012; Batalle et al., 2017; Smyser et al., 2010). It is also important to mention that many methods have been proposed to interpret machine learning models and gain insight into the inner functioning of trained models (for a review, see (Molnar et al., 2020) and (Xie et al., 2020)). As we used both RF and DNN in this work, we chose to use a standard method of sensitivity analysis, as it could easily be applied in the same manner to both RF and DNN models instead of using other interpretation methods specifically tuned for each model. Furthermore, sensitivity analysis allows to visualize feature importance throughout the entire data set as opposed to in individual samples. This allowed us to easily compare network features contributing to the prediction of GA and PMA for both RF and DNN models.

Although the dHCP dataset is one of the largest neonatal MRI resources currently available a key next step will be to investigate whether our results generalise to other cohorts. The regression algorithms were built and trained with the specific MRI acquisition protocol, brain parcellation and connectome generation methods developed for that project, which may limit how translatable our findings are. To enable this, our predictive algorithms have been made publicly available (https://github.com/CoDe-Neuro/Predicting-age-and-clinical-risk-from-the-neonatal-connectome). As the data was normalized prior to training, given a similar parcellation, we expect that significant correlation between true and predicted GA at birth or PMA at scan should be obtained if tested on different data. Future work could also focus on implementation of more modern machine learning approaches – for example geometric deep learning, which is specifically designed for data in non-Euclidean space (Bronstein et al., 2017). There is also increasing interest in predicting outcomes at age 18-24 months directly from neonatal brain connectivity, as done in (Girault et al., 2019). We have implemented our own version of their method and tested on the sub-set of our data set with available BSID-III at 18 months of age. However, no significant prediction capacity was reached with our data (data not shown). This might be due to different developmental outcome (Mullen scale instead of BSID-III), as well as differences in the pre-processing pipeline, or differences in the sample size and characteristics.

5. Conclusion

We have used DNN to accurately predict PMA and GA at Birth from the neonatal connectome and have used sensitivity analysis to describe the brain features most important in our models. We have additionally computed a brain maturation index which is associated with future motor development. We achieved a MAE of 0.72 weeks in predicting PMA at scan, demonstrating that the neonatal structural connectome contains key developmental information. Furthermore, prediction of GA at birth with MAE of 2.21 weeks shows that the patterns characteristic of prematurity are clearly present in the neonatal connectome, and can be uncovered with machine learning approaches. Several connectivity patterns identified as relevant for our models are in line with findings from existing neurodevelopmental studies. Lastly, brain maturation index was significantly correlated to BSID-III motor outcome at 18 months, suggesting the potential of this approach to develop biomarkers for prediction of atypical neurodevelopment.

6. Funding

This work was supported by the European Research Council under the European Union Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 319456. The authors acknowledge infrastructure support from the National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre (BRC) at South London, Maudsley NHS Foundation Trust, King's College London and the NIHR-BRC at Guys and St Thomas' Hospitals NHS Foundation Trust (GSTFT). The authors also acknowledge support in part from the Wellcome Engineering and Physical Sciences Research Council (EPSRC) Centre for Medical Engineering at King's College London [WT 203148/Z/16/Z] and the Medical Research Council (UK) [MR/K006355/1 and MR/L011530/1]. Additional sources of support included the Sackler Institute for Translational Neurodevelopment at King's College London, the European Autism Interventions (EU-AIMS) trial and the EU AIMS-2-TRIALS, a European Innovative Medicines Initiative Joint Undertaking under Grant Agreements No. 115300 and 777394, the resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (Grant FP7/2007–2013). DC is supported by a Flemish Research Foundation (FWO) Fellowship [12ZV420N]. OGG is supported by a grant from the UK Medical Research Council / Sackler Foundation [MR/P502108/1]. MD is supported by the Academy of Medical Sciences Springboard Award [SBF004\1040]. TA is supported by a MRC Clinician Scientist Fellowship [MR/P008712/1] and translation support award [MR/V036874/1]. ADE, TA, JOM received support from the Medical Research Council Centre for Neurodevelopmental Disorders, King's College London [MR/N026063/1]. DB received support from a Wellcome Trust Seed Award in Science [217316/Z/19/Z]. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

References

Ahn, S.H. and Kim, S.A., (2017), Assessment of preterm infants using the Bayley-III scales in Korea. *Annals of rehabilitation medicine*, *41*(5), p.843.

Andersson, J. L., Skare, S. and Ashburner, J. (2003), 'How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging', *Neuroimage* **20**(2), 870–888.

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A. and Gee, J. C. (2011), 'A reproducible evaluation of ants similarity metric performance in brain image registration', *Neuroimage* **54**(3), 2033–2044.

Ball, G., Aljabar, P., Zebari, S., Tusor, N., Arichi, T., Merchant, N., Robinson, E. C., Ogundipe, E., Rueckert, D., Edwards, A. D. et al. (2014), 'Rich-club organization of the newborn human brain', *Proceedings of the National Academy of Sciences* **111**(20), 7456–7461.

Ball, G., Boardman, J. P., Rueckert, D., Aljabar, P., Arichi, T., Merchant, N., Gousias, I. S., Edwards, A. D. and Counsell, S. J. (2012), 'The effect of preterm birth on thalamic and cortical development', *Cerebral cortex* **22**(5), 1016–1024.

Ball, G., Pazderova, L., Chew, A., Tusor, N., Merchant, N., Arichi, T., Allsop, J.M., Cowan, F.M., Edwards, A.D. and Counsell, S.J., (2015). Thalamocortical connectivity predicts cognition in children born preterm. *Cerebral cortex*, *25*(11), pp.4310-4318.

Ball, G., Boardman, J.P., Aljabar, P., Pandit, A., Arichi, T., Merchant, N., Rueckert, D., Edwards, A.D. and Counsell, S.J., (2013). The influence of preterm birth on the developing thalamocortical connectome. *Cortex*, *49*(6), pp.1711-1721.

Bastiani, M., Andersson, J. L., Cordero-Grande, L., Murgasova, M., Hutter, J., Price, A. N., Makropoulos, A., Fitzgibbon, S. P., Hughes, E., Rueckert, D. et al. (2019), 'Automated processing pipeline for neonatal diffusion mri in the developing human connectome project', *NeuroImage* **185**, 750–763.

Batalle, D., Edwards, A. D. and O'Muircheartaigh, J. (2018), 'Annual research review: not just a small adult brain: understanding later neurodevelopment through imaging the neonatal brain', *Journal of Child Psychology and Psychiatry* **59**(4), 350–371.

Batalle, D., Hughes, E. J., Zhang, H., Tournier, J.-D., Tusor, N., Aljabar, P., Wali, L., Alexander, D. C., Hajnal, J. V., Nosarti, C. et al. (2017), 'Early development of structural networks and the impact of prematurity on brain connectivity', *Neuroimage* **149**, 379–392.

Bayley, N. (2006), *Bayley scales of infant and toddler development*, PsychCorp, Pearson.

Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A.-B., Narwal, R., Adler, A., Garcia, C. V., Rohde, S., Say, L. et al. (2012), 'National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications', *The lancet* **379**(9832), 2162–2172.

Bozek, J., Makropoulos, A., Schuh, A., Fitzgibbon, S., Wright, R., Glasser, M. F., Coalson, T. S., O'Muircheartaigh, J., Hutter, J., Price, A. N. et al. (2018), 'Construction of a neonatal cortical surface atlas using multimodal surface matching in the developing human connectome project', *NeuroImage* **179**, 11–29.

Breiman, L. (2001), 'Random forests', *Machine learning* **45**(1), 5–32.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A. and Vandergheynst, P. (2017), 'Geometric deep learning: going beyond euclidean data', *IEEE Signal Processing Magazine* **34**(4), 18–42.

Brown, C. J. and Hamarneh, G. (2016), 'Machine learning on human connectome data from mri', *arXiv preprint arXiv:1611.08699* .

Brown, C. J., Moriarty, K. P., Miller, S. P., Booth, B. G., Zwicker, J. G., Grunau, R. E., Synnes, A. R., Chau, V. and Hamarneh, G. (2017), Prediction of brain network age and factors of delayed maturation in very preterm infants, *in* 'International Conference on Medical Image Computing and Computer-Assisted Intervention', Springer, pp. 84–91.

Brown, C.J., Miller, S.P., Booth, B.G., Andrews, S., Chau, V., Poskitt, K.J. and Hamarneh, G., (2014). Structural network analysis of brain development in young preterm neonates. *Neuroimage*, *101*, pp.667-680.

Campbell, J. S. and Pike, G. B. (2014), 'Potential and limitations of diffusion mri tractography for the study of language', *Brain and language* **131**, 65– 73.

Cao, B., Mwangi, B., Hasan, K. M., Selvaraj, S., Zeni, C. P., Zunta-Soares, G. B. and Soares, J. C. (2015), 'Development and validation of a brain maturation index using longitudinal neuroanatomical scans', *Neuroimage* **117**, 311–318.

Chollet, F. et al. (2015), 'Keras', https://github.com/fchollet/keras.

Christiaens, D., Cordero-Grande, L., Hutter, J., Price, A. N., Deprez, M., Hajnal, J. V. and Tournier, J.-D. (2018), 'Learning compact *q*-space representations for multi-shell diffusion-weighted mri', *IEEE transactions on medical imaging* **38**(3), 834–843.

Christiaens, D., Cordero-Grande, L., Pietsch, M., Hutter, J., Edwards, A. D., Deprez, M., Hajnal, V. and Tournier, J. (2018), 'Multi-shell shard reconstruction from scattered slice diffusion mri data in the neonatal brain', *ISMRM (Paris)* .

Christiaens, D., Cordero-Grande, L., Pietsch, M., Hutter, J., Price, A.N., Hughes, E.J., Vecchiato, K., Deprez, M., Edwards, A.D., Hajnal, J.V. and Tournier, J.D., (2021). Scattered slice SHARD reconstruction for motion correction in multi-shell diffusion MRI. *Neuroimage*, *225*, p.117437.

Cordero-Grande, L., Christiaens, D., Hutter, J., Price, A. N. and Hajnal, J. V. (2019), 'Complex diffusion-weighted image estimation via matrix recovery under general noise models', *NeuroImage* **200**, 391–404.

Cordero-Grande, L., Hughes, E. J., Hutter, J., Price, A. N. and Hajnal, J. V. (2018), 'Three-dimensional motion corrected sensitivity encoding reconstruction for multi-shot multi-slice

mri: application to neonatal brain imaging', *Magnetic resonance in medicine* **79**(3), 1365–1376.

Cordero-Grande, L., Teixeira, R. P. A., Hughes, E. J., Hutter, J., Price, A. N. and Hajnal, J. V. (2016), 'Sensitivity encoding for aligned multishot magnetic resonance reconstruction', *IEEE Transactions on Computational Imaging* **2**(3), 266–280.

Counsell, S., Rutherford, M., Cowan, F. and Edwards, A. (2003), 'Magnetic resonance imaging of preterm brain injury', *Archives of Disease in Childhood-Fetal and Neonatal Edition* **88**(4), F269–F274.

de Almeida, J.S., Meskaldji, D.E., Loukas, S., Lordier, L., Gui, L., Lazeyras, F. and Hüppi, P.S., (2021). Preterm birth leads to impaired rich-club organization and fronto-paralimbic/limbic structural connectivity in newborns. *NeuroImage*, *225*, p.117440.

Dhollander, T., Raffelt, D. and Connelly, A. (2016), Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion mr data without a co-registered t1 image, *in* 'ISMRM Workshop on Breaking the Barriers of Diffusion MRI', Vol. 5, p. 5.

Dhollander, T., Raffelt, D. and Connelly, A. (2018), Accuracy of response function estimation algorithms for 3-tissue spherical deconvolution of diverse quality diffusion mri data, *in* 'Proceedings of the Joint Meeting of the European Society for Magnetic Resonance in Medicine and Biology and the International Society of Magnetic Resonance in Medicine, Paris, France', Vol. 1569.

Dimitrova, R., Pietsch, M., Christiaens, D., Ciarrusta, J., Wolfers, T., Batalle, D., Hughes, E., Hutter, J., Cordero-Grande, L., Price, A. N. et al. (2020), 'Heterogeneity in brain microstructural development following preterm birth', *Cerebral Cortex* **30**(9), 4800–4810.

Dosenbach, N. U., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., Nelson, S. M., Wig, G. S., Vogel, A. C., Lessov-Schlaggar, C. N. et al. (2010), 'Prediction of individual brain maturity using fmri', *Science* **329**(5997), 1358–1361.

Eixarch, E., Muñoz-Moreno, E., Bargallo, N., Batalle, D. and Gratacos, E., (2016). Motor and cortico-striatal-thalamic connectivity alterations in intrauterine growth restriction. *American journal of obstetrics and gynecology*, *214*(6), pp.725-e1.

Fan, Y., Shi, F., Smith, J. K., Lin, W., Gilmore, J. H. and Shen, D. (2011), 'Brain anatomical networks in early human brain development', *Neuroimage* **54**(3), 1862–1871.

Fitzgibbon, S. P., Harrison, S. J., Jenkinson, M., Baxter, L., Robinson, E. C., Bastiani, M., Bozek, J., Karolis, V., Grande, L. C., Price, A. N. et al. (2020), 'The developing human connectome project (dhcp) automated resting-state functional processing framework for newborn infants.', *BioRxiv* p. 766030.

Foulder-Hughes, L. and Cooke, R. (2003), 'Motor, cognitive, and behavioural disorders in children born very preterm', *Developmental Medicine & Child Neurology* **45**(2), 97–103.

Franke, K., Gaser, C., Manor, B. and Novak, V. (2013), 'Advanced brainage in older adults with type 2 diabetes mellitus', *Frontiers in aging neuroscience* **5**, 90.

Galdi, P., Blesa, M., Stoye, D. Q., Sullivan, G., Lamb, G. J., Quigley, A. J., Thrippleton, M. J., Bastin, M. E. and Boardman, J. P. (2020), 'Neonatal morphometric similarity mapping for predicting brain age and characterizing neuroanatomic variation associated with preterm birth', *NeuroImage: Clinical* **25**, 102195.

Girault, J. B., Munsell, B. C., Puechmaille, D., Goldman, B. D., Prieto, J. C., Styner, M. and Gilmore, J. H. (2019), 'White matter connectomes at birth accurately predict cognitive abilities at age 2', *Neuroimage* **192**, 145– 155.

Greene, M.M., Patra, K., Nelson, M.N. and Silvestri, J.M., (2012). Evaluating preterm infants with the Bayley-III: patterns and correlates of development. *Research in developmental disabilities*, *33*(6), pp.1948-1956.

Hagmann, P. (2005), 'From diffusion mri to brain connectomics', Technical report, EPFL.

Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V. J., Meuli, R., Thiran, J.-P. and Grant, P. (2010), 'White matter maturation reshapes structural connectivity in the late developing human brain', *Proceedings of the National Academy of Sciences* **107**(44), 19067– 19072.

Honey, C. J., Thivierge, J.-P. and Sporns, O. (2010), 'Can structure predict function in the human brain?', *Neuroimage* **52**(3), 766–776.

Hughes, E. J., Winchman, T., Padormo, F., Teixeira, R., Wurie, J., Sharma, M., Fox, M., Hutter, J., Cordero-Grande, L., Price, A. N. et al. (2017), 'A dedicated neonatal brain imaging system', *Magnetic resonance in medicine* **78**(2), 794–804.

Hutter, J., Christiaens, D. J., Schneider, T., Cordero-Grande, L., Slator, P. J., Deprez, M., Price, A. N., Tournier, J.-D., Rutherford, M. and Hajnal, J. V. (2018), 'Slice-level diffusion encoding for motion and distortion correction', *Medical image analysis* **48**, 214–229.

Hutter, J., Tournier, J. D., Price, A. N., Cordero-Grande, L., Hughes, E. J., Malik, S., Steinweg, J., Bastiani, M., Sotiropoulos, S. N., Jbabdi, S. et al. (2018), 'Time-efficient and flexible design of optimized multishell hardi diffusion', *Magnetic resonance in medicine* **79**(3), 1276–1292.

Jbabdi, S. and Johansen-Berg, H. (2011), 'Tractography: where do we go from here?', *Brain connectivity* **1**(3), 169–183.

Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A. and Sijbers, J. (2014), 'Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data', *NeuroImage* **103**, 411–426.

Johnson, M. H. (2001), 'Functional brain development in humans', *Nature Reviews Neuroscience* **2**(7), 475–483.

Johnson, S., Hollis, C., Kochhar, P., Hennessy, E., Wolke, D. and Marlow, N. (2010), 'Autism spectrum disorders in extremely preterm children', *The Journal of pediatrics* **156**(4), 525–531.

Johnson, S. and Marlow, N. (2014), Growing up after extremely preterm birth: lifespan mental health outcomes, *in* 'Seminars in Fetal and Neonatal Medicine', Vol. 19, Elsevier, pp. 97–104.

Jonsson, B. A., Bjornsdottir, G., Thorgeirsson, T., Ellingsen, L. M., Walters, G. B., Gudbjartsson, D., Stefansson, H., Stefansson, K. and Ulfarsson, M. (2019), 'Brain age prediction using deep learning uncovers associated sequence variants', *Nature communications* **10**(1), 1–10.

Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., Zwicker, J. G. and Hamarneh, G. (2017), 'BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment', *NeuroImage* **146**, 1038–1049.

Kellner, E., Dhital, B., Kiselev, V. G. and Reisert, M. (2016), 'Gibbs-ringing artifact removal based on local subvoxel-shifts', *Magnetic resonance in medicine* **76**(5), 1574–1581.

Keunen, K., Counsell, S. J. and Benders, M. J. (2017), 'The emergence of functional architecture during early brain development', Neuroimage **160**, 2–14.

Kingma, D. P. and Ba, J. (2014), 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980* .

Koutsouleris, N., Davatzikos, C., Borgwardt, S., Gaser, C., Bottlender, R., Frodl, T., Falkai, P., Riecher-Rössler, A., Möller, H.-J., Reiser, M. et al. (2014), 'Accelerated brain aging in schizophrenia and beyond: a neuroanatomical marker of psychiatric disorders', *Schizophrenia bulletin* **40**(5), 1140–1153.

Kuklisova-Murgasova, M., Quaghebeur, G., Rutherford, M. A., Hajnal, J. V. and Schnabel, J. A. (2012), 'Reconstruction of fetal brain mri with intensity matching and complete outlier removal', *Medical image analysis* **16**(8), 1550–1564.

Lean, R.E., Han, R.H., Smyser, T.A., Kenley, J.K., Shimony, J.S., Rogers, C.E., Limbrick, D.D. and Smyser, C.D., (2019). Altered neonatal white and gray matter microstructure is associated with neurodevelopmental impairments in very preterm infants with high-grade brain injury. *Pediatric research*, *86*(3), pp.365-374.

Lee, J.Y., Park, H.K. and Lee, H.J., (2019). Accelerated small-world property of structural brain networks in preterm infants at term-equivalent age. *Neonatology*, *115*(2), pp.99-107.

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.

Maier-Hein, K. H., Neher, P. F., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q. et al. (2017), 'The challenge of mapping the human connectome based on diffusion tractography', *Nature communications* **8**(1), 1–13.

Makropoulos, A., Gousias, I. S., Ledig, C., Aljabar, P., Serag, A., Hajnal, J. V., Edwards, A. D., Counsell, S. J. and Rueckert, D. (2014), 'Automatic whole brain mri segmentation of the developing neonatal brain', *IEEE transactions on medical imaging* **33**(9), 1818–1831.

Makropoulos, A., Robinson, E. C., Schuh, A., Wright, R., Fitzgibbon, S., Bozek, J., Counsell, S. J., Steinweg, J., Vecchiato, K., Passerat-Palmbach, J. et al. (2018), 'The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction', *Neuroimage* **173**, 88–112.

Marlow, N., Hennessy, E. M., Bracewell, M. A., Wolke, D. et al. (2007), 'Motor and executive function at 6 years of age after extremely preterm birth', *Pediatrics* **120**(4), 793–804.

Marlow, N., Wolke, D., Bracewell, M. A. and Samara, M. (2005), 'Neurologic and developmental disability at six years of age after extremely preterm birth', *New England journal of medicine* **352**(1), 9–19.

Molnar, C., Casalicchio, G. and Bischl, B., (2020), September. Interpretable machine learning–a brief history, state-of-the-art and challenges. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 417-431). Springer, Cham.

Nosarti, C., Reichenberg, A., Murray, R. M., Cnattingius, S., Lambe, M. P., Yin, L., MacCabe, J., Rifkin, L. and Hultman, C. M. (2012), 'Preterm birth and psychiatric disorders in young adult life', *Archives of general psychiatry* **69**(6), 610–617.

O'Muircheartaigh, J., Robinson, E. C., Pietsch, M., Wolfers, T., Aljabar, P., Grande, L. C., Teixeira, R. P., Bozek, J., Schuh, A., Makropoulos, A. et al. (2020), 'Modelling brain development to detect white matter injury in term and preterm born neonates', *Brain* **143**(2), 467–479.

Pandit, A.S., Robinson, E., Aljabar, P., Ball, G., Gousias, I.S., Wang, Z., Hajnal, J.V., Rueckert, D., Counsell, S.J., Montana, G. and Edwards, A.D., (2014). Whole-brain mapping of structural connectivity in infants reveals altered connection strength associated with growth and preterm birth. *Cerebral cortex*, *24*(9), 2324-2333.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011), 'Scikit-learn: Machine learning in python', *the Journal of machine Learning research* **12**, 2825–2830.

Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A. and Smith, S. M. (2019), 'Accurate brain age prediction with lightweight deep neural networks', *BioRxiv* .

Pietsch, M., Christiaens, D., Hutter, J., Cordero-Grande, L., Price, A. N., Hughes, E., Edwards, A. D., Hajnal, J. V., Counsell, S. J. and Tournier, J.-D. (2019), 'A framework for multi-component analysis of diffusion mri data over the neonatal period', *NeuroImage* **186**, 321–337.

Raffelt, D., Dhollander, T., Tournier, J.-D., Tabbara, R., Smith, R. E., Pierre, E. and Connelly, A. (2017), Bias field correction and intensity normalisation for quantitative analysis of apparent fibre density, *in* 'Proc. Intl. Soc. Mag. Reson. Med', Vol. 25, p. 3541.

Rasmussen, J. M., Kruggel, F., Gilmore, J. H., Styner, M., Entringer, S., Consing, K. N., Potkin, S. G., Wadhwa, P. D. and Buss, C. (2017), 'A novel maturation index based on neonatal diffusion tensor imaging reflects typical perinatal white matter development in humans', *International Journal of Developmental Neuroscience* **56**, 42–51.

Ratnarajah, N., Rifkin-Graboi, A., Fortier, M.V., Chong, Y.S., Kwek, K., Saw, S.M., Godfrey, K.M., Gluckman, P.D., Meaney, M.J. and Qiu, A., (2013). Structural connectivity asymmetry in the neonatal brain. *Neuroimage*, *75*, pp.187-194.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986), 'Learning representations by back-propagating errors', *Nature* **323**(6088), 533–536.

Saltelli, A., (2002). Sensitivity analysis for importance assessment. *Risk analysis*, *22*(3), pp.579-590.

Schadl, K., Vassar, R., Cahill-Rowley, K., Yeom, K.W., Stevenson, D.K. and Rose, J., (2018), Prediction of cognitive and motor development in preterm children using exhaustive feature selection and cross-validation of near-term white matter microstructure. *NeuroImage: Clinical*, *17*, pp.667-679.

Schnabel, J. A., Rueckert, D., Quist, M., Blackall, J. M., Castellano-Smith, A. D., Hartkens, T., Penney, G. P., Hall, W. A., Liu, H., Truwit, C. L. et al. (2001), A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations, *in* 'International Conference on Medical Image Computing and Computer-Assisted Intervention', Springer, pp. 573–581.

Schuh, A., Makropoulos, A., Robinson, E. C., Cordero-Grande, L., Hughes, E., Hutter, J., Price, A. N., Murgasova, M., Teixeira, R. P. A., Tusor, N. et al. (2018), 'Unbiased construction of a temporally consistent morphological atlas of neonatal brain development', *bioRxiv* p. 251512.

Shi, F., Yap, P.-T., Wu, G., Jia, H., Gilmore, J. H., Lin, W. and Shen, D. (2011), 'Infant brain atlases from neonates to 1-and 2-year-olds', *PloS one* **6**(4), e18746.

Smith, Leslie N. 'A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay.' arXiv preprint arXiv:1803.09820 (2018).

Smith, R. E., Tournier, J.-D., Calamante, F. and Connelly, A. (2012), 'Anatomically-constrained tractography: improved diffusion mri streamlines tractography through effective use of anatomical information', *Neuroimage* **62**(3), 1924–1938.

Smith, R. E., Tournier, J.-D., Calamante, F. and Connelly, A. (2015), 'Sift2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography', *Neuroimage* **119**, 338–351.

Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E. and Miller, K. L. (2019), 'Estimation of brain age delta from brain imaging', *Neuroimage* **200**, 528–539.

Smith, T., Noble, M., Noble, S., Wright, G., McLennan, D., Plunkett, E. et al. (2015), 'The english indices of deprivation 2015', *London: Department for Communities and Local Government* .

Smyser, C. D., Inder, T. E., Shimony, J. S., Hill, J. E., Degnan, A. J., Snyder, A. Z. and Neil, J. J. (2010), 'Longitudinal analysis of neural network development in preterm infants', *Cerebral cortex* **20**(12), 2852– 2862.

Sporns, O. (2013), 'The human connectome: origins and challenges', *Neuroimage* **80**, 53–61.

Sporns, O., Tononi, G. and Kötter, R. (2005), 'The human connectome: a structural description of the human brain', *PLoS Comput Biol* **1**(4), e42.

Toulmin, H., Beckmann, C.F., O'Muircheartaigh, J., Ball, G., Nongena, P., Makropoulos, A., Ederies, A., Counsell, S.J., Kennea, N., Arichi, T. and Tusor, N., (2015). Specialization and integration of functional thalamocortical connectivity in the human infant. *Proceedings of the National Academy of Sciences*, *112*(20), pp.6485-6490.

Tournier, J.-D., Calamante, F. and Connelly, A. (2013), 'Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging', *NMR in Biomedicine* **26**(12), 1775–1786.

Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H. and Connelly, A. (2019), 'Mrtrix3: A fast, flexible and open software framework for medical image processing and visualisation', *NeuroImage* **202**, 116137.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A. and Gee, J. C. (2010), 'N4itk: improved n3 bias correction', *IEEE transactions on medical imaging* **29**(6), 1310–1320.

Van Den Heuvel, M.P. and Sporns, O., 2011. Rich-club organization of the human connectome. *Journal of Neuroscience*, *31*(44), pp.15775-15786.

Van Den Heuvel, M.P., Kersbergen, K.J., De Reus, M.A., Keunen, K., Kahn, R.S., Groenendaal, F., De Vries, L.S. and Benders, M.J., (2015). The neonatal connectome during preterm brain development. *Cerebral cortex*, *25*(9), pp.3000-3013.

Velikos, K., Soubasi, V., Michalettou, I., Sarafidis, K., Nakas, C., Papadopoulou, V., Zafeiriou, D. and Drossou, V., (2015). Bayley-III scales at 12 months of corrected age in preterm infants: Patterns of developmental performance and correlations to environmental and biological influences. *Research in developmental disabilities*, *45*, pp.110-119.

Veraart, J., Novikov, D. S., Christiaens, D., Ades-Aron, B., Sijbers, J. and Fieremans, E. (2016), 'Denoising of diffusion mri using random matrix theory', *Neuroimage* **142**, 394–406.

Wood, N. S., Marlow, N., Costeloe, K., Gibson, A. T. and Wilkinson, A. R. (2000), 'Neurologic and developmental disability after extremely preterm birth', *New England Journal of Medicine* **343**(6), 378–384.

Xia, M., Wang, J. and He, Y., (2013). BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one*, *8*(7), p.e68910.

Xie, N., Ras, G., van Gerven, M. and Doran, D. (2020), 'Explainable deep learning: A field guide for the uninitiated', *arXiv preprint arXiv:2004.14545* .

List of figures:

Figure 1. Distribution of (A) GA at birth and (B) PMA at scan of full cohort (N=524). (C) GA at birth of cohort used for predicting GA at birth.

Figure 2. (A) Pipeline for age prediction from MRI. (B) Cross Validation protocol. (C) Legend for DNN architecture components. (D) PMA at scan DNN architecture. (E) GA at birth DNN architecture.

Figure 3. Detailed results of prediction of PMA at scan on term cohort. True vs Predicted with (A) RF, (B) DNN. Fold by fold result with (C) RF, (D) DNN.

Figure 4. Detailed results of prediction of GA at birth. True vs Predicted with (A) RF, (B) DNN. Fold by fold result with (C) RF, (D) DNN.

Figure 5. Visualization with BrainNetViewer of the relevance of connections and nodes for predictions of PMA with RF (A) and DNN (B), as well as GA with RF (E) and DNN (F); colour scale of each connection and region is the z-value of the computed relevance. Tables showing five most relevant connections and regions for predictions of PMA with RF (C) and DNN (D), as well as GA with RF (G) and DNN (H).

Figure 6. Association of Brain Maturation Index with BSID-III outcomes in preterm-born infants. (A) True vs predicted PMA at scan for preterm infants. (B) Brain maturation index δ vs BSID-III Gross motor outcome corrected for IMD. (C) Detailed correlation and p-values of preterm brain maturation index and BSID-III outcomes – statistically significant results highlighted in bold red.