bioRxiv preprint doi: https://doi.org/10.1101/2023.10.16.562598; this version posted October 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Surface Generative Modelling of Neurodevelopmental Trajectories

Abdulah Fawaz, Saga N. B. Masui, Logan Z. J. Williams, Simon Dahan, A. David Edwards and Emma C. Robinson

***Abstract*— Cortical neurodevelopment is sensitive to disruption following preterm birth, with lasting impact on cognitive outcomes. The creation of generative models of neurodevelopment could aid clinicians in identifying at-risk subjects but is complicated by the degree of subject variability in cortical folding, and significant heterogeneity in the effect of preterm birth. In this work, we propose a graph convolutional generative adversarial network (GAN) and a training scheme to simulate neonatal cortical surface developmental trajectories. The proposed model is used to smoothly modify two cortical phenotypes: post-menstrual age at scan (PMA) and gestational age at birth (GA) on data from the developing Human Connectome Project (dHCP) [1]. The synthetic images were validated with an independently trained regression network, and compared against follow up scans, indicating that the model can realistically age individuals whilst preserving subject-specific cortical morphology. Deviation between simulated 'healthy' scans, and preterm follow up scans generated a metric of individual atypicality, which improved prediction of 18-month cognitive outcome over GA alone.**

***Index Terms*— Graph Neural Network, GAN, cortical surface, generative modelling, neurodevelopment**

I. Background

During the third trimester of pregnancy, the human brain develops significantly in terms of microstructure and functional organisation. This process is known to be sensitive to factors such as preterm birth. Multiple studies have reported significant changes in both cortical microstructure and morphology, including changes to mean diffusivity [2]–[4], cortical thickness and surface area [3], [5]–[7], and reduced white and grey matter volumes [7]–[9]. Atypical cortical neurodevelopment is known to be correlated to poorer cognitive outcomes including Attention-Deficit Hyperactivity Disorder (ADHD), Autism Spectrum Disorders (ASD) and cerebral palsy [10]–[13].

Generative modelling of cortical neurodevelopment could aid clinicians by improving mechanistic understanding of the causes of disease, or be used to identify patients for early clinical intervention. Computational modelling of cortical neurodevelopment faces numerous challenges, particularly due to heterogeneity of cortical organisation and folding patterns, among even healthy subjects. This limits the accuracy of population-based analyses based on diffeomorphic registration, which is insufficient to normalise all sources of variability, and so under which residual misalignments remain [14], [15]. Furthermore, the impact of preterm birth has itself been shown to be heterogeneous [3] - meaning that neurodevelopmental trajectories are unique to individuals.

One increasingly popular approach for modelling heterogeneous brain phenotypes is normative modelling, which attempts to predict typical patterns of neurodevelopment which then act as a reference to understand deviations from the norm. Gaussian Process Regression normative models have found some success in characterising the impact of prematurity on cortical neurodevelopment [16], [17], but remain reliant on image registration leading to high uncertainty at the cortex where intersubject variability is highest [3].

Deep generative modelling has emerged as a powerful tool in overcoming these challenges, leveraging the ability of convolutional neural networks (CNNs) to model complex tasks independent of image registration. Deep generative modelling has already been used to simulate age-related disease progression on volumetric neuroimaging data [18]. For example, Bass, Silva, Sudre, *et al.* [19], [20] used a VAE-GAN [21], with separate content and attribute encodings, to disentangle class specific features of disease, from class irrelevant features of cortical shape variation, to map patterns of disease related brain atrophy in individuals with mild cognitive impairment (MCI) or Alzheimer's disease (AD). Others models have sought to improve interpretability by incorporating confounding factors, such as age: for example, Ravi, Alexander, Oxtoby, *et al.* [22] proposed a Degenerative Adversarial NeuroImage Net (DaniNet) to simulate AD progression, conditioned on age, disease state and a biological model of disease progression. More recent developments include the use of latent diffusion

A.F is supported by an Engineering and Physical Sciences Research Council (EPSRC) United Kingdom Doctoral Training Partnership (DTP) - EPSRC DTP 2018 [EP/R513064/1]. L.Z.J.W is supported by funding from the Commonwealth Scholarship Commission, United Kingdom. E.C.R is supported by an Academy of Medical Sciences/the British Heart Foundation/the Government Department of Business, Energy and Industrial Strategy/the Wellcome Trust Springboard Award [SBF003/1116], a Wellcome Collaborative Award [215573/Z/19/Z] and an MRC Methodology grant MR/V03832X/1. A.D.E. receives support from the Medical Research Council Centre for Neurodevelopmental Disorders, King's College London [grant MR/N026063/1].

A.F, S.N.B.M, S.D, L.Z.J.W., and E.C.R are affiliated with Department of Biomedical Engineering, School of Biomedical Engineering and Imaging Sciences, King's College London, UK. L.Z.J.W, E.C.R and A.D.E. are also affiliated with Centre for the Developing Brain, School of Biomedical Engineering and Imaging Sciences, King's College London, UK. A.D.E is affiliated with; the Department for Forensic and Neurodevelopmental Sciences, and the MRC Centre for Neurodevelopmental Disorders, King's College London, UK.

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.16.562598; this version posted October 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

modelling, to generate age-conditioned brain images with significantly higher perceptual quality than those produced with GANs [23]. Other approaches have sought to go beyond interpretation towards explanation through incorporation of concepts from causal inference [24].

The closest related work to this paper is the age progression/regression model described in Xia, Chartsias, Tsaftaris, *et al.* [25], which uses a GAN, directly conditioned on age difference, to generate subject-specific brain age progression maps (on 2D MRI slices) without the need for longitudinal data. In followup work, the same authors extended their model by further conditioning on health state, to simulate brain ageing with and without MCI or AD [26].

The above models, whilst powerful, are suitable only for 2D images and/or 3D volumes of the brain, yet surface-based analysis of the cortex has been shown to be advantageous over volume-based due to its ability to more accurately encode geodesic distances along the cortex, and hence model patterns of cortical organisation [14], [27]. However, the application of deep learning to surfaces is non-trivial, as surfaces belong to a class of data domains that includes graphs, point clouds and manifolds, which are mathematically inconsistent with foundational aspects of traditional CNNs [28].

Geometric deep learning (gDL) is a nascent field generalising deep learning to a range of non-Euclidean domains. A thorough review of these methods and the theory behind them can be found in the literature [29], but they have already begun to find application on the cortical surface; for cortical surface segmentation [30], [31], regression of phenotypes [32], [33], and cortical surface registration [34], [35]. Comparative studies of gDL models on the cortical surface [33], [36] have highlighted that these models often compromise on either computational feasibility, network expressivity, or fidelity to the symmetry of the underlying data [36]. Generative gDL models have been proposed for drug discovery [37], 3D shape manipulation [38] and point cloud generation [39], but to our knowledge are yet to be applied to generative modelling of cortical imaging features.

In this paper we propose an extension to our preliminary work [40], which trained a CycleGAN to translate cortical appearance of preterm scans to appear like healthy term controls. As the model was conditioned purely on discrete classes, it was was not able to *continuously* age brains. This work therefore proposes a graph convolutional GAN, capable of simulating neonatal cortical surface neurodevelopmental trajectories, conditioned on continuous age phenotypes: post-menstrual age at scan (PMA) and gestational age at birth (GA). To achieve this, we take inspiration from existing work [25], [26] and directly condition our generator on the desired age difference to learn a cortical surface difference map. Our work incorporates the following specific contributions:

- We describe a model that integrates graph convolutions into a modified CycleGAN framework that can produce fully controllable predictions across the entire neurodevelopmental trajectory of the neonates. Model design is validated through an ablation study, with hyperparameter tuning.
- We perform experiments on the developing human connectome project (dHCP) dataset [41] demonstrating the utility of our model in generating realistic, age-accurate and subject-specific predictions of cortical maturation and the effect of changes in degree of prematurity.
- We demonstrate the utility of our method by identifying biomarkers that correlate to improved predictions of cognitive outcomes of preterm subjects.

II. Methods

A. Problem Statement

We denote a neonatal cortical surface feature map as $\boldsymbol{x}_{s,b}$, for a subject with PMA $= s$, and GA $= b$. We define cortical neurodevelopment as an additive process, and model changes in s and b separately:

$$\boldsymbol{x}_{s+\Delta s,b} = \boldsymbol{x}_{s,b} + \boldsymbol{d}_{\boldsymbol{x},\Delta s} \tag{1}$$

$$\boldsymbol{x}_{s,b+\Delta b} = \boldsymbol{x}_{s,b} + \boldsymbol{d}_{\boldsymbol{x},\Delta b} \tag{2}$$

where, $\boldsymbol{d}_{\boldsymbol{x},\Delta s}$ is a difference map on $\boldsymbol{x}$ due to a Δs change in s. We use **bold** notation to denote vectors throughout. The aim of our model is to learn $\boldsymbol{d}$ from $\boldsymbol{x}$, by conditioning the generator and discriminator on age values (s,b) and age difference values ($\Delta s, \Delta b$) as detailed in section II-D. Learning age difference maps, over directly aging images, was found to be advantageous as it helps to preserve subject identity, since intrasubject differences, due to age, are expected to be much smaller than intersubject variation due to different cortical folding patterns. This was verified this through ablation (Sec IV-A).

B. Graph Convolutional Networks (GCNs)

The traditional form of convolution used in 2D and 3D imaging is invalid on cortical surface meshes due to their irregular local structure and the need for rotational, not translational, equivariance to correctly represent positional variation of features. In this paper we propose to use a form of graph convolutional operator which is invariant to the numbering and ordering of vertices [42]:

$$y^i = W_1 x^i + W_2 \sum_{j \in N(i)} e_{i,j} x^j \tag{3}$$

Here, $e_{i,j}$ is the adjacency matrix, W_1, W_2 are learned filter weights, x^i denotes the ith vertex of $\boldsymbol{x}$, and $\boldsymbol{y}$ is the output of the convolutional operation. Since the sum operation is applied over all the neighbours of point i before a single learned weight (W_2) is applied, the output of the network is unchanged by variation in either the number or ordering of the neighbours.

C. Proposed Model

Continuous modelling of neurodevelopmental phenotypes is implemented through the training of an n-step cycle GAN. At each step (Fig 1A) two operations are implemented: 1) an age transformation operation (blue arrow) that uses the generator to age the input image by a specified age difference; and 2) an image discrimination operation (red arrow) that evaluates the accuracy of the simulated ageing through a binary cross entropy loss (L_{AGE}) estimating the probability

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.16.562598; this version posted October 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

that the synthetic image is a realistic image of the target age. The network uses a closed cycle of n of these operations, where the image is transformed to a number of intermediate ages before closing the cycle, by transforming the image back to the original age and evaluating a reconstruction loss L_{RECON} that enforces cycle consistency i.e that the final image $\boldsymbol{x}_n$ is equivalent to the initial image $\boldsymbol{x}_0$. This takes the form of a smooth L1 loss:

$$L_{RECON} = \begin{cases} 0.5(\boldsymbol{x}_n - \boldsymbol{x}_0)^2 & \text{if } |\boldsymbol{x}_n - \boldsymbol{x}_0| < 1 \\ |\boldsymbol{x}_n - \boldsymbol{x}_0| = 0.5 & \text{otherwise} \end{cases} \quad (4)$$

Examples of 2, 3 and n-cycle cases are shown in Fig 1B.

D. Architecture

The architecture of the model is shown in Fig 2. In each case, the conditional generator and discriminator are graph convolutional networks (GCNs), with convolutions implemented using Eq 3 and convolutional layers interleaved with ReLu activation functions. The input data are sphericalised icosahedral cortical feature maps of resolution 40962 vertices (see section III-A). These are downsampled between convolutional layers, to lower resolution icospheres of resolution 10242, 2562 and 642 vertices, using an efficient hexagonal pooling scheme [36]. Upsampling of the decoder network is implemented using bilinear interpolation.

The generator is composed of an encoder, age modifier component, and a decoder. The encoder extracts latent features, which the age modifier transforms according to the conditioned age difference (Δb or Δs), and the decoder generates an age difference map.

The discriminator takes the form of a GCN classifier, which is conditioned on a target age (s, or b) as shown in Fig 2. The output is a probability representing the confidence that the input image is a realistic image of that specific target age. In both the generator and discriminator, the conditioned variables, age difference and target age, are represented as floats, in contrast to the discrete ordinal encodings approach of prior methods [25], [26]. This allows our model to perform age progression or age regression depending on the sign of the age difference, and is intuitive as neural networks are known to (mostly) approximate smooth continuous functions [43], where we expect the modification of the image encoding to smoothly vary with age.

III. Experimental Setup

A. Data and Augmentations

1) Dataset: All experiments were run using neuroimaging data from the publicly available dHCP dataset* [1], [44]–[48]. The dataset comprises preterm and term subjects, cross-sectionally scanned between 24-45 weeks post menstrual age (PMA), and gestational ages at birth (GA) between 21-40 weeks. Table I describes the dataset and the subsets used in this paper. In the dHCP, the preterm group typically undergo a scan either shortly after birth, or at term-equivalent age, but a subset of 45 subjects are scanned twice for which longitudinal data is available.

Two different cognitive outcome measures were available for each subject, measured at around 18 months (592 ± 74 days at assessment): QChat [49] and Bayley-III [50]. QChat is a self-reported questionnaire based metric that assesses Autism Spectrum Disorder risk, whereas Bayley-III is a composite metric of language and motor development, derived from behavioural observations. These are also shown in table I.

2) Preprocessing: The full details of image reconstruction and preprocessing pipelines are described in [1], [47] and references therein. In brief, motion corrected and reconstructed T2w and T1w sMRI images were passed through the dHCP structural pipeline† [47], which performed Draw-EM tissue segmentation [51], surface extraction [48] and inflation [27] to return vertex-matched inner (white matter), outer (pial), midthickness, inflated and spherical surfaces. This process generated a number of univariate surface feature maps of which we use sulcal depth and cortical myelination defined as T1w/T2w ratio maps [52]. Individual sphericalised maps were then registered to the dHCP 40-week neonatal symmetric cortical surface template [53] using multimodal surface matching (MSM) [54], [55], driven by sulcal depth features. Aligned features were then resampled to a regular $40,962$-vertex icosphere using barycentric interpolation, implemented using Human Connectome Project (HCP) workbench software [56].

B. Implementation

The data were augmented offline with non-linear warps to produce realistic variations in the data and improve network generalisation. These were produced by the scheme described in Fawaz, Williams, Alansary, *et al.* [36]. We used the myelination and sulcal depth features maps as our input channels. These were normalised to a mean of 0 and a standard deviation of 1. Train, validation and test sets were split evenly between preterms and terms, with ratios 0.8:0.1:0.1. Target ages for simulation were sampled at random from between 28 and 45 weeks. Graph convolutions were implemented using PyTorch Geometric [57]. Models were trained with an Adam Optimiser with learning rate of 0.001 for 100 epochs. Our code can be found at `https://github.com/Abdulah-Fawaz/continuous_cGAN`.

C. Experiments

In this paper, we evaluate our proposed model through two experiments. The first to simulate changes with respect to healthy cortical maturation during the third trimester, and the second reflects the impact of prematurity on cortical neurodevelopment.

1) Simulating Healthy Cortical Neurodevelopment: In our first experiment we make the simplifying assumption that scans acquired from preterm neonates, shortly after birth, are approximately healthy. We then model the effect of increasing PMA by training on cross-sectional data acquired from all

*http://www.developingconnectome.org

†https://github.com/BioMedIA/dhcp-structural-pipeline

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.16.562598; this version posted October 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

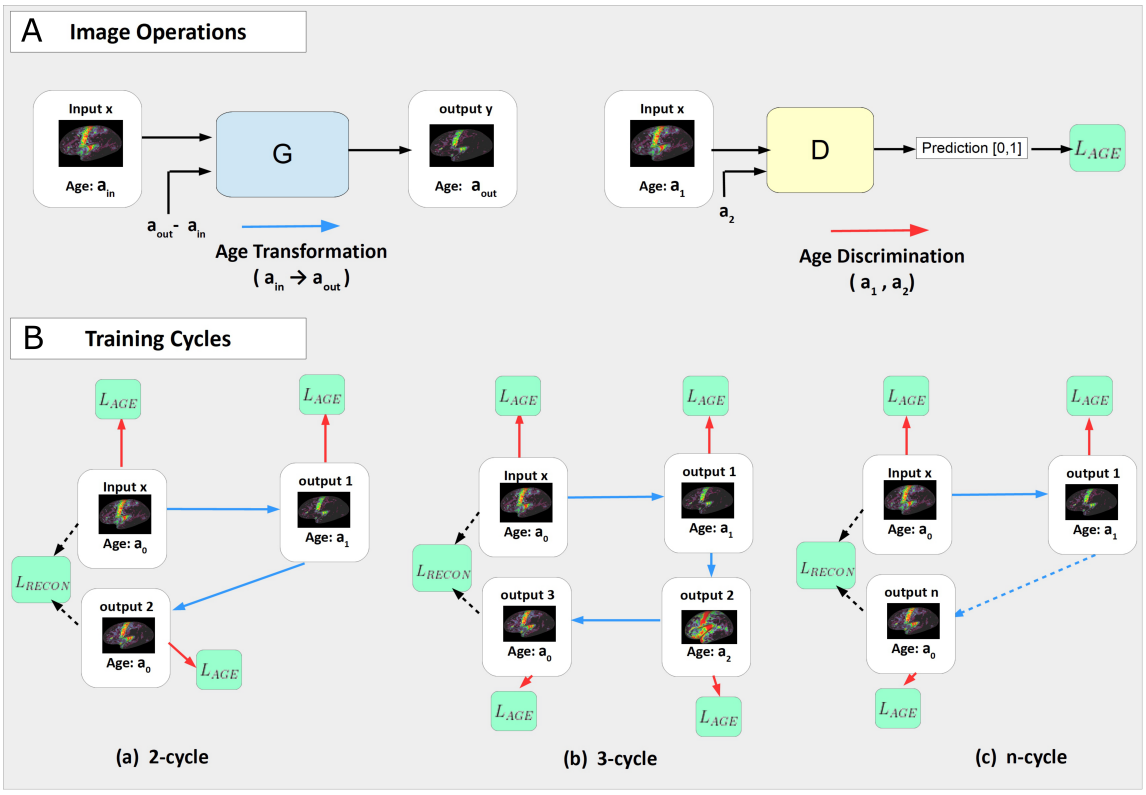

Fig. 1: The proposed model consists of a n-step closed training cycle (B) of image operations (A) regularised by a cycle loss. During each step in a closed cycle, an input image is transformed to a new age (blue arrow) and an age loss is evaluated (red arrow) to measure the accuracy of the new image to the target age. The number of transformations, n, in a cycle is a hyperparameter. Training cycles of lengths 2, 3 and n are shown.

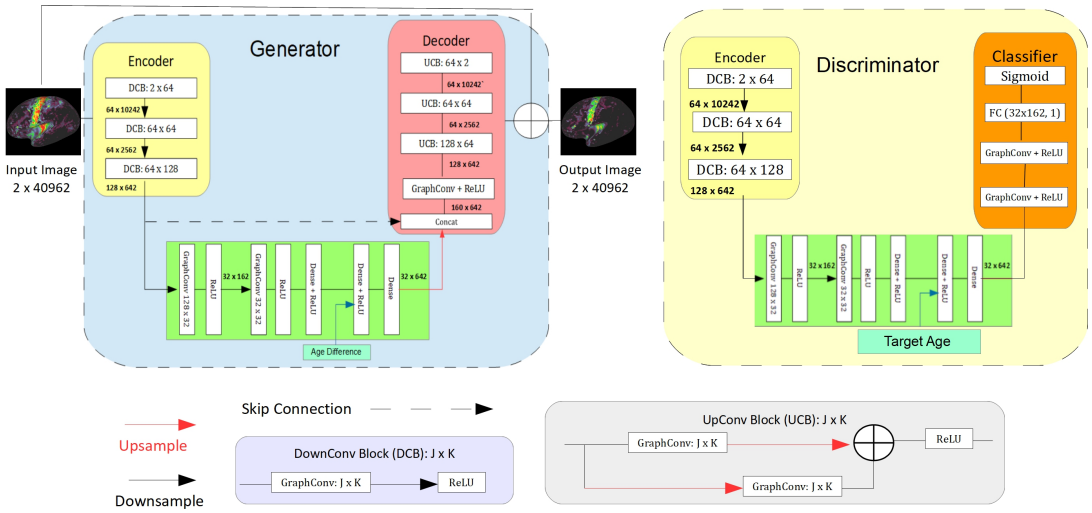

Fig. 2: The generator consists of three sections: an encoder (yellow), an age modifier (green) and a decoder (red). The encoder performs feature extraction on an input image. The age modifier takes in the desired age difference and transforms the latent vector to one of the desired age. Original and transformed feature vectors are then combined by concatenation and passed to the decoder, which generates a difference map, adding this to the original image to obtain the transformed image. The aim of the discriminator is to determine if the input image is a realistic image of a specific target age. It takes the form of a GCN conditioned on a specific target age, that generates a probability that the input image corresponds to a real image of the target age.

preterm neonates' first scans, and all scans from term controls. For this dataset (described in table I), it is assumed that PMA≈GA; thus, GA need not be modelled explicitly, and Δb can be set to zero in Eq (1). To train the model on this

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.16.562598; this version posted October 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

	Preterms					Terms				
Dataset	**Count**	**PMA**	**GA**	**QChat**	**Bayley**	**Count**	**PMA**	**GA**	**QChat**	**Bayley**
Full	161	37.0 ± 4.0	31.9 ± 3.5	31.0 ± 10.2	99.7 ± 12.2	419	41.1 ± 1.7	40.0 ± 1.2	30.4 ± 8.9	100.4 ± 10.9
Exp1	111	33.96 ± 2.27	31.7 ± 3.37	32.0 ± 9.7	99.4 ± 11.9	"	"	"	"	"
Exp2	95	40.75 ± 2.08	32.06 ± 3.65	29.8 ± 10.7	100.1 ± 12.6	"	"	"	"	"
Longitudinal	45	33.8 ± 2.0 (first scan) 41.2 ± 1.4 (second scan)	31.0 ± 3.3	30.7 ± 9.3	99.7 ± 10.8	N/A	N/A	N/A	N/A	N/A

TABLE I: Breakdown of the datasets used in this paper. GA and PMA measured in weeks. Figures show mean and standard deviation.

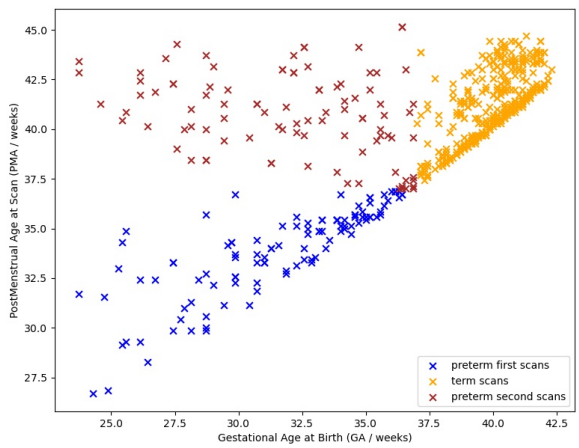

Fig. 3: A plot to show the distribution of images in our dataset.

dataset, the generator and discriminator are conditioned on Δs and s respectively.

2) Simulating the Impact of Preterm Birth: The aim of the second experiment was to simulate the impact of GA, and therefore degree of prematurity, from scans acquired around term-equivalent age. Preterm neonates first scans were excluded from the data, resulting in the dataset shown in table I. While changes with respect to healthy cortical maturation are known to be relatively stark and well-characterised in the literature [3], [6], the effect of preterm birth is relatively subtle and heterogeneous [17]. Further, the effect observed in the dHCP data set is confounded by individual's PMA at scan. To account for both varying GA and PMA, the experiment is therefore modified to add PMA to both the generator and discriminator as a confound (concatenated as a float). Training is then conditioned on b, and Δb. Additionally, a final information preservation loss L_{IP} is applied between successive images to encourage the network to learn changes that are proportional to the difference in GA:

$$L_{IP}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}) = |\boldsymbol{x}_n - \boldsymbol{x}_{n+1}| \cdot \exp(-|b_{n+1} - b_n|) \quad (5)$$

where image $\boldsymbol{x}_i$ has GA of b_i. This is to reflect that the effect of preterm birth on the brain (varying GA at birth whilst holding PMA constant) is smaller than the variation due to changes in PMA at scan for constant GA.

D. Evaluation

For quantitative evaluation, three different metrics were used:

(A) 'Age MAE' - or the mean absolute error (MAE) between the apparent age of a synthetic image and the targeted age. Here apparent age was determined from an independently trained gDL regression model, implemented with 4 layers of MoNet [58] convolutions, ReLU activations and a final fully connected layer (previously described in [36]). Training was performed on the same train/val/test splits as for the proposed image generation model. This returned a baseline age MAE (for ground truth data) of 0.68 ± 0.19 for PMA and 1.5 ± 0.38 weeks on the more difficult task of predicting GA.

(B) 'Subject specificity' - this evaluates the similarity between an original input scan and synthetic (aged) images. This was measured from three similarity metrics: the peak signal-to-noise ratio (PSNR), the mean square error (MSE), and the structural similarity index measure (SSIM). Note that SSIM was modified to function on the icosahedron by adapting the requisite Gaussian filtering operations to the surface.

(C) Prediction of preterm outcome - clinical utility of the model of healthy cortical ageing was validated for prediction of cognitive outcome at 18 months. Here, a metric summarising deviation from typical development was derived by estimating the mean square error (MSE) between each individual's ground truth follow up scan, and a simulated 'healthy' scan at the same age. This was compared against QChat and Bayley-III test scores. Since GA alone is already correlated with cognitive outcomes for preterm subjects, we measure the changes in this correlation from a GA-only baseline, to linear regression of these scores, estimated from a combination of GA + deviation metric.

E. Ablation

Prior to running all experiments we ran ablation analyses, on the task of simulating healthy cortical neurodevelopment. Experimental design was validated against a baseline CycleGAN, trained to translate examples between two discrete classes: term (PMA $\geq$ 37 weeks) and preterm (PMA $<$ 37 weeks) [40]. Model parameters were then evaluated by investigating the impact of changing training cycle length (from $n = [2, 3, 4, 5]$), and evaluating how well the model performed when trained to learn images directly instead of age difference maps - to test the hypothesis that learning difference maps improves subject specificity (run for the n=3 cycle only).

IV. Results

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.16.562598; this version posted October 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Model	Age Accuracy (MAE / weeks)	Subject Specificity		
		SSIM	PSNR	MSE
CycleGAN [40]	7.66 ± 2.11	0.80	**25.6**	177
2-cycle	2.70 ± 0.70	0.74	17.3	203
3-cycle	1.02 ± 0.2	**0.82**	17.9	**167**
3-cycle w/o diff maps	1.14 ± 0.46	0.66	13.2	251
4-cycle	**0.8 ± 0.38**	0.70	16.1	231
5-cycle	1.55 ± 0.50	0.74	16.1	208

TABLE II: A table comparing the performance of different generative models on the simulation of typical cortical developmental trajectories, as measured by the accuracy of synthesised images to target PMA and subject specificity.

A. Ablation Study

The results of the ablation study are shown in table II. The CycleGAN baseline produced a high subject specificity, but a very poor age accuracy (7.66 weeks MAE) which is expected due to the discrete nature of the image conditioning. The model with the highest age accuracy (0.8 weeks MAE) was the 4-cycle model, but it also demonstrated the lowest subject specificity. The $n = 3$ model gave the next best performance in age accuracy (1.0 weeks MAE) whilst retaining the overall highest subject specificity, and thus was the best performing model overall. The same model, without difference maps, witnessed a large drop in subject specificity, confirming the hypothesis that generating translations using difference maps significantly preserves individual's cortical morphology without compromising on the precision of the age generation. Going forward all experiments therefore use the 3-cycle network and generate difference maps.

B. Simulating Healthy Cortical Development

On simulation of healthy cortical neurodevelopment, data generated from the network returned an age MAE of 1.02 ± 0.2 weeks. This compares well against the performance of the baseline regression network on ground truth data (Fig. 4) with some greater error at the extremes of the distribution, where fewer data samples were available.

Examples of the images generated by the model across time are shown in Fig. 5 for two example preterm individuals. It can be seen, through comparisons to population average templates [53], [59], that changes in myelination (top) and sulcal depth (bottom) follow expected neurodevelopmental trends. Myelination increases most strongly along the central sulcus and major sulci deepen and grow as PMA increases. More subtle features, such as myelination of the middle temporal (MT) area, and the emergence and growth of smaller sulci along the temporal lobe are present. However, this cannot simply be explained as the network generating a population average model of development. Visual inspection of the cortical folding patterns for each subject show highly divergent patterns of folding in the frontal lobe which are preserved during simulated ageing. Follow up scans are shown for reference, but these are not expected to match precisely as our model is only simulating healthy cortical neurodevelopment without taking into account the impact of preterm birth. This property of identity preservation was further investigated through statistical comparison of each preterm individual's follow up scan

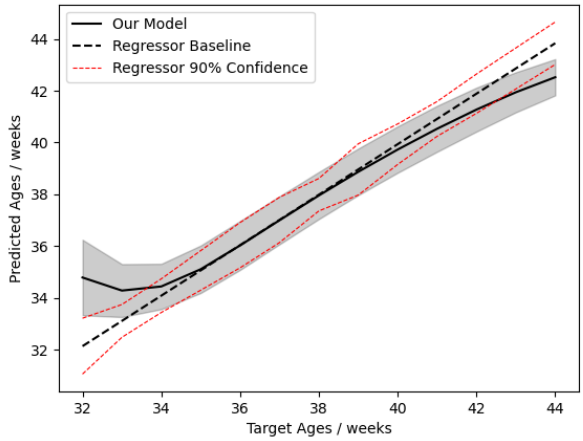

Fig. 4: A comparison of age MAE for simulated and ground truth cortical feature maps. Here the X-axis reports target (or ground truth) PMA and the Y-axis reports PMA predicted from the baseline regressor. Performance on generated images is shown with a solid line, compared to ground truth (dashed). The grey region indicates the uncertainty in the age accuracy of the GCN model on generated images, and the red dashed lines show the 90% confidence interval of the regressor.

with their simulated images. The mean square error (MSE) difference between the sulcal depth feature maps of a subject's follow up scan and its own simulated image was lower than any other subject (aged to the same PMA) for 42 of the 45 subjects for which longitudinal data was available (93.3%). The average margin between the correct subject's generated image and the next-closest subject was 18.4 ± 13.3 MSE. Sulcal depth features maps were used because they contain an individual's unique folding pattern, and are therefore a good marker of subject preservation.

Further, we find that the proposed biomarker (MSE deviation between ground truth and simulated follow up scans) increases the r^2 correlation between predicted and ground truth 18-month cognitive test scores (for the same 45 left out individuals) with QChat correlation increasing slightly from 0.30 to 0.32, and the Bayley-III correlation increasing more substantially from 0.41 to 0.53. Poorer performance on the Qchat may be explained by the fact that it is primarily used for screening ASD, but there are no-known individuals at high risk of ASD in the dHCP dataset [1].

C. Simulating the Impact of Preterm Birth

Age MAE for the second experiment was 1.3 ± 0.8 weeks. While this is larger than for the PMA experiment, it remains close to the 1.5 ± 0.38 obtained from ground truth data (Fig 7) indicating that the performance of the regressor is likely a limiting factor in evaluating the true performance of the generative model. Nevertheless, there is a clear trend that points to our model being able to monotonically alter the apparent GA with target GA.

Examples of the images generated by the model for different GA are shown in Fig 6. The first row shows an example

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.16.562598; this version posted October 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

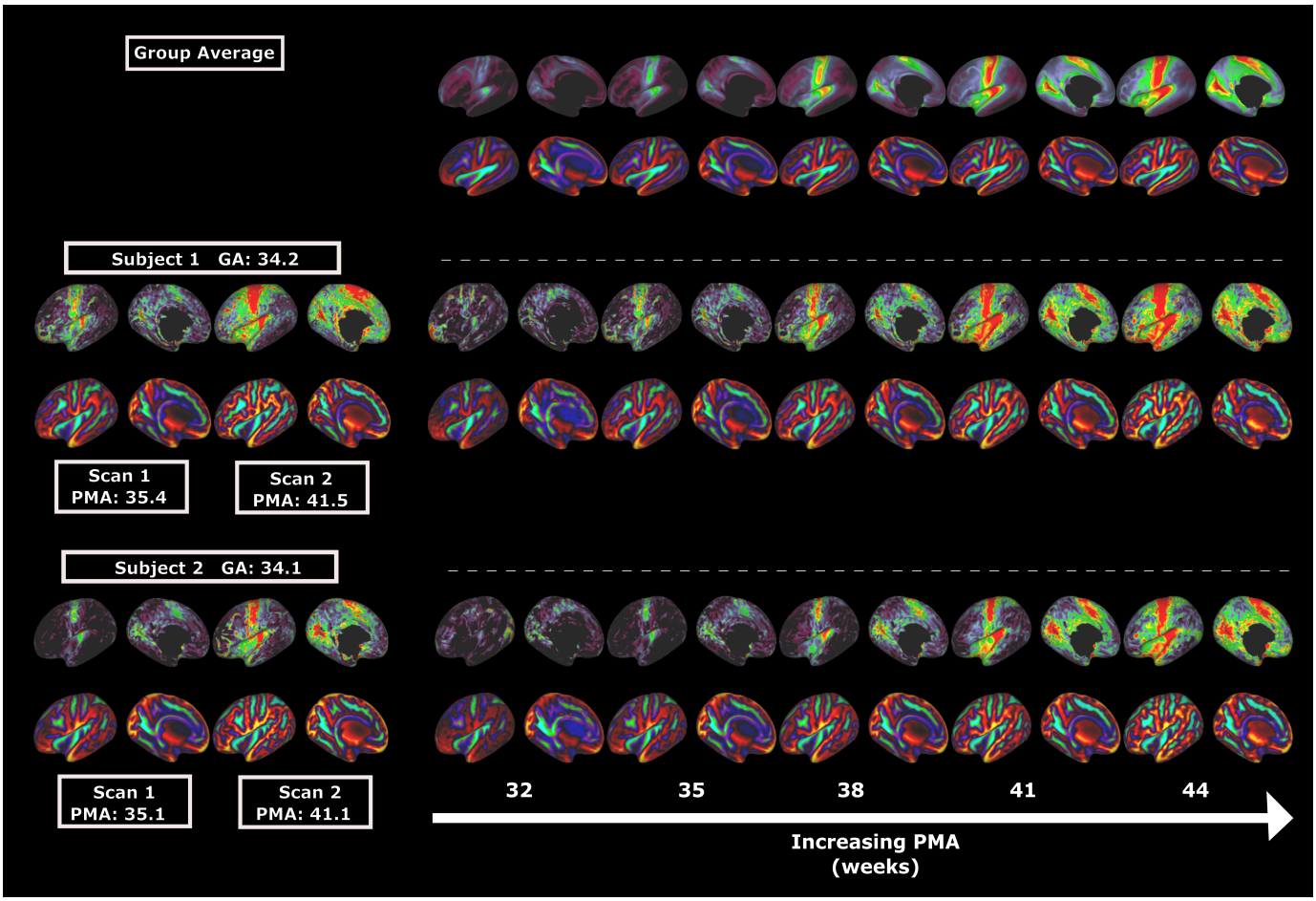

Fig. 5: Group average templates for the dHCP dataset (top), a subject of GA 34.2 weeks, and scans at PMA 35.4 weeks and 41.5 weeks (middle, left) and a second subject of GA 34.1 weeks and PMA 35.1 weeks and 41.1 weeks (bottom, left) alongside a set of sample synthetic images of (columns 2-6) generated by the model to represent the same subjects at PMA 32, 35, 38, 41 and 44 weeks respectively. The model predicts an increase in myelination and a change in sulci depth consistent with those seen in the group average templates, while preserving individual cortical folding patterns.

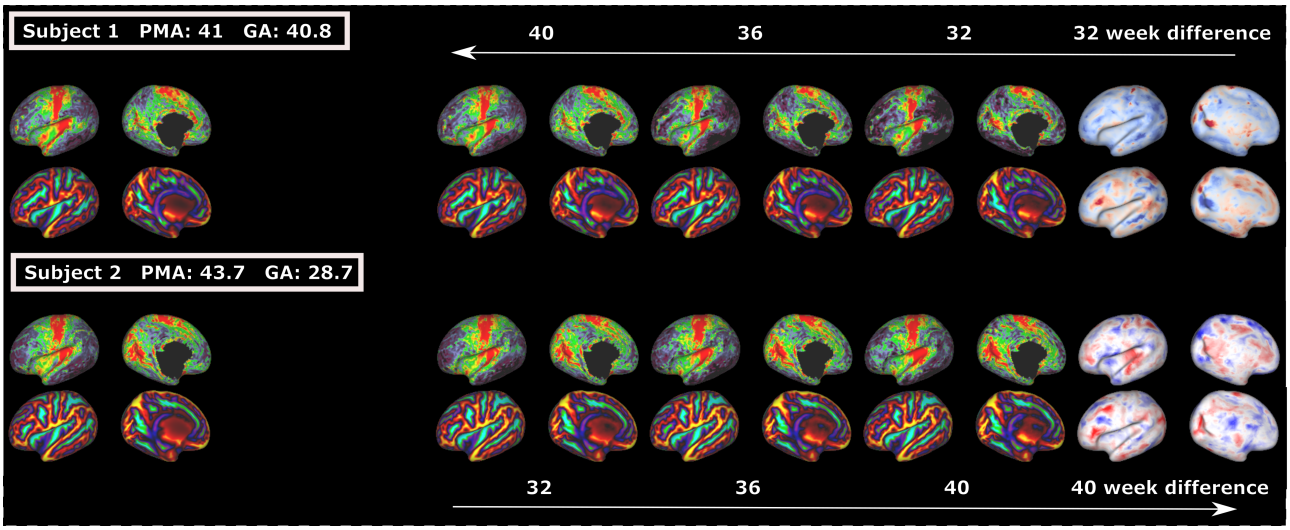

Fig. 6: (Top) A subject of original GA 40.8 weeks and PMA 41.0 weeks with simulated decrease of GA to 40, 36 and 32 weeks. (Bottom) A preterm subject of original GA 28.7 weeks and PMA 43.7 weeks aged up to a GA of 40 weeks. Images in the final column show difference maps between the original images and the final synthetic images.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.16.562598; this version posted October 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

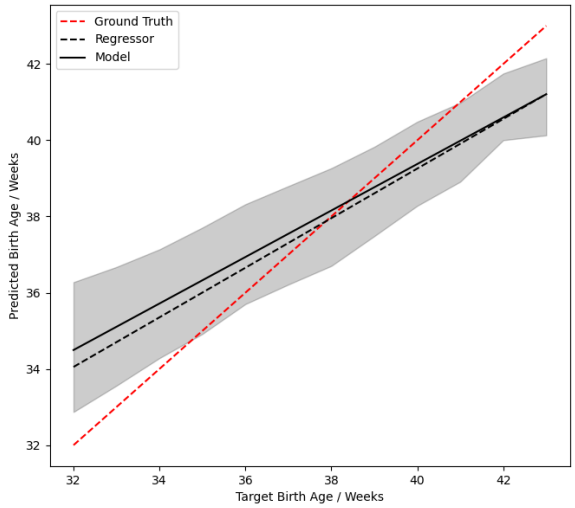

Fig. 7: A plot showing the predicted gestational age (GA) vs input target age of our GCN model (solid) compared against the baseline regressor accuracy (dashed). The grey region indicates the standard error of the GCN model predictions, and the red dashed line shows the ground truth.

term-control subject with GA 40.8 weeks (PMA at scan 41.0 weeks) as the model simulates decreasing GA to 32 weeks. Conversely, the second example shows a preterm subject with GA 28.7 weeks and PMA at scan of 43.7 weeks, altered so that their GA appears 40 weeks.

In contrast to the rapid changes that occur as PMA is increased from 32 to 44 weeks, differences in GA are associated with more subtle changes. To aid interpretation, the final column shows difference maps between the original images and the final synthetic images. These show a global reduction in myelination across most of the cortical surface as GA decreases - in line with a large body of existing research that links preterm birth with delayed myelination [47], [60], [61]. The decrease in myelination is not total, however, and certain regions of the preterm brain are known to display higher levels of myelination suspected to be due to early exposure to stimuli [62], and these too are present in our difference maps.

The impact of decreasing GA on sulcal depth is less well-documented, but is known to impact shallowing of the sulci and gyri [6] and this is reflected in the correspondence between the folding pattern and the difference maps. Regional analysis [3], [6] of the cortical surface of preterms identified relatively shallow temporal and cingulate sulci as regions most consistently affected by preterm birth - the two regions most strongly highlighted by our model.

V. Conclusion

Deep generative modelling on surfaces presents challenges on two fronts. Modelling of surfaces is complicated by the lack of a global coordinate system over the surface on which to define convolutions [63], and difficulty defining regular up/down sampling [29]. Deep generative models must contend with the limited availability of paired data in medical imaging, retaining subject identity during image generation, whilst simultaneously modelling cortical heterogeneity across populations.

In this paper, we have presented a novel deep generative model of neonatal cortical surface neurodevelopment that solves both sets of challenges by integrating state of the art methods from geometric deep learning and Euclidean generative modelling. The model was thoroughly validated through quantitative and qualitative assessments, focusing on the accuracy of age prediction for generated images, and the preservation of subject identity. Our results demonstrate the success of our model in predicting changes associated with cortical maturation and degree of prematurity. Furthermore, the model was used to propose a novel biomarker that improved the accuracy with which it is possible to predict cognitive outcomes of at risk individuals. Longterm, such an approach could be utilised in a clinical environment to identify subjects for clinical intervention.

Moving forward, there are several opportunities for further exploration and improvement. Our model could be applied to other forms of MRI surface data beyond sMRI, e.g. diffusion-weighted imaging or functional MRI. Additionally, conducting studies incorporating larger datasets would allow us to validate our findings across diverse populations and improve the generalisability of our model, particularly as it was observed that our model accuracy decreased in regions of lower data availability.

More fundamentally, there was an underlying assumption that scans acquired from preterm births near the time of birth were healthy. This is highly unlikely to be satisfied in practice since underlying risk factors, genetic and environmental, are likely to be implicated in the occurrence of preterm birth. In future, this assumption could be relaxed with the use of fetal scans to represent healthy subjects at lower PMA. Another avenue of further work, inspired by Xia, Chartsias, Wang, *et al.* [26], would be to condition the model on additional discrete factors such as disease state, sex, or birth weight, to further disentangle how demographic and clinical factors impact neurodevelopment.

References

[1] A. D. Edwards, D. Rueckert, S. M. Smith, *et al.*, "The developing human connectome project neonatal data release," *Frontiers in neuroscience*, vol. 16, p. 886 772, 2022.

[2] M. Bouyssi-Kobar, M. Brossard-Racine, M. Jacobs, J. Murnick, T. Chang, and C. Limperopoulos, "Regional microstructural organization of the cerebral cortex is affected by preterm birth," *NeuroImage: Clinical*, vol. 18, pp. 871–880, 2018.

[3] R. Dimitrova, M. Pietsch, J. Ciarrusta, *et al.*, "Preterm birth alters the development of cortical microstructure and morphology at term-equivalent age," *bioRxiv*, 2021.

[4] G. Ball, L. Srinivasan, P. Aljabar, *et al.*, "Development of cortical microstructure in the preterm human brain," *Proceedings of the National Academy of Sciences*, vol. 110, no. 23, pp. 9541–9546, 2013.

[5] M. Ajayi-Obe, N. Saeed, F. Cowan, M. A. Rutherford, and A. D. Edwards, "Reduced development of cerebral cortex in extremely preterm infants," *The Lancet*, vol. 356, no. 9236, pp. 1162–1163, 2000.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.16.562598; this version posted October 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[6] E. Engelhardt, T. E. Inder, D. Alexopoulos, *et al.*, "Regional impairments of cortical folding in premature infants," *Annals of neurology*, vol. 77, no. 1, pp. 154–162, 2015.

[7] A. Makropoulos, P. Aljabar, R. Wright, *et al.*, "Regional growth and atlasing of the developing human brain," *Neuroimage*, vol. 125, pp. 456–478, 2016.

[8] G. Ball, J. P. Boardman, D. Rueckert, *et al.*, "The effect of preterm birth on thalamic and cortical development," *Cerebral cortex*, vol. 22, no. 5, pp. 1016–1024, 2012.

[9] D. K. Thompson, S. K. Warfield, J. B. Carlin, *et al.*, "Perinatal risk factors altering regional brain structure in the preterm infant," *Brain*, vol. 130, no. 3, pp. 667–677, 2007.

[10] N. Marlow, D. Wolke, M. A. Bracewell, and M. Samara, "Neurologic and developmental disability at six years of age after extremely preterm birth," *New England journal of medicine*, vol. 352, no. 1, pp. 9–19, 2005.

[11] L. W. Doyle and S. Saigal, "Long-term outcomes of very preterm or tiny infants," *NeoReviews*, vol. 10, no. 3, e130–e137, 2009.

[12] R. H. Pickler, J. M. McGrath, M. B. A. Reyna, *et al.*, "A model of neurodevelopmental risk and protection for preterm infants," *The Journal of perinatal & neonatal nursing*, vol. 24, no. 4, p. 356, 2010.

[13] J. J. Volpe, "Brain injury in premature infants: A complex amalgam of destructive and developmental disturbances," *The Lancet Neurology*, vol. 8, no. 1, pp. 110–124, 2009.

[14] M. F. Glasser, T. S. Coalson, E. C. Robinson, *et al.*, "A multimodal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, 2016.

[15] E. Dohmatob, G. Varoquaux, and B. Thirion, "Inter-subject registration of functional images: Do we need anatomical images?" *Frontiers in neuroscience*, vol. 12, p. 64, 2018.

[16] J. O'Muircheartaigh, E. C. Robinson, M. Pietsch, *et al.*, "Modelling brain development to detect white matter injury in term and preterm born neonates," *Brain*, vol. 143, no. 2, pp. 467–479, 2020.

[17] R. Dimitrova, M. Pietsch, D. Christiaens, *et al.*, "Heterogeneity in brain microstructural development following preterm birth," *Cerebral Cortex*, vol. 30, no. 9, pp. 4800–4810, 2020.

[18] V. Wegmayr, M. Hörold, and J. M. Buhmann, "Generative aging of brain mri for early prediction of mci-ad conversion," in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, IEEE, 2019, pp. 1042–1046.

[19] C. Bass, M. da Silva, C. Sudre, P.-D. Tudosiu, S. Smith, and E. Robinson, "Icam: Interpretable classification via disentangled representations and feature attribution mapping," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7697–7709, 2020.

[20] C. Bass, M. da Silva, C. Sudre, *et al.*, "Icam-reg: Interpretable classification and regression with feature attribution for mapping neurological phenotypes in individual scans," *arXiv preprint arXiv:2103.02561*, 2021.

[21] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," in *International conference on machine learning*, PMLR, 2016, pp. 1558–1566.

[22] D. Ravi, D. C. Alexander, N. P. Oxtoby, and A. D. N. Initiative, "Degenerative adversarial neuroimage nets: Generating images that mimic disease progression," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*, Springer, 2019, pp. 164–172.

[23] W. H. Pinaya, P.-D. Tudosiu, J. Dafflon, *et al.*, "Brain imaging generation with latent diffusion models," in *MICCAI Workshop on Deep Generative Models*, Springer, 2022, pp. 117–126.

[24] P. Sanchez and S. A. Tsaftaris, "Diffusion causal models for counterfactual estimation," *arXiv preprint arXiv:2202.10166*, 2022.

[25] T. Xia, A. Chartsias, S. A. Tsaftaris, and A. D. N. Initiative, "Consistent brain ageing synthesis," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22*, Springer, 2019, pp. 750–758.

[26] T. Xia, A. Chartsias, C. Wang, S. A. Tsaftaris, A. D. N. Initiative, *et al.*, "Learning to synthesise the ageing brain without longitudinal data," *Medical Image Analysis*, vol. 73, p. 102 169, 2021.

[27] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[28] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," *arXiv preprint arXiv:2104.13478*, 2021.

[29] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.

[30] K. Gopinath, C. Desrosiers, and H. Lombaert, "Segrecon: Learning joint brain surface reconstruction and segmentation from images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2021, pp. 650–659.

[31] F. Zhao, Z. Wu, L. Wang, *et al.*, "Spherical deformable u-net: Application to cortical surface parcellation and development prediction," *IEEE transactions on medical imaging*, vol. 40, no. 4, pp. 1217–1228, 2021.

[32] S. Dahan, A. Fawaz, L. Z. J. Williams, *et al.*, "Surface vision transformers: Attention-based modelling applied to cortical analysis," *arXiv preprint arXiv:2203.16414*, 2022.

[33] V. Vosylius, A. Wang, C. Waters, *et al.*, "Geometric deep learning for post-menstrual age prediction based on the neonatal white matter cortical surface," in *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis*, Springer, 2020, pp. 174–186.

[34] M. A. Suliman, L. Z. Williams, A. Fawaz, and E. C. Robinson, "A deep-discrete learning framework for spherical surface registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2022, pp. 119–129.

[35] F. Zhao, Z. Wu, F. Wang, *et al.*, "S3reg: Superfast spherical surface registration based on deep learning," *IEEE Transactions on Medical Imaging*, vol. 40, no. 8, pp. 1964–1976, 2021.

[36] A. Fawaz, L. Z. Williams, A. Alansary, *et al.*, "Benchmarking geometric deep learning for cortical segmentation and neurodevelopmental phenotype prediction," *bioRxiv*, 2021.

[37] J. Xiong, Z. Xiong, K. Chen, H. Jiang, and M. Zheng, "Graph neural networks for automated de novo drug design," *Drug Discovery Today*, vol. 26, no. 6, pp. 1382–1393, 2021.

[38] M. Eisenberger, D. Novotny, G. Kerchenbaum, *et al.*, "Neuromorph: Unsupervised shape interpolation and correspondence in one go," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7473–7483.

[39] D. W. Shu, S. W. Park, and J. Kwon, "3d point cloud generative adversarial network based on tree structured graph convolutions," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 3859–3868.

[40] A. Fawaz, L. Z. Williams, A. D. Edwards, and E. Robinson, *A deep generative model of neonatal cortical surface development*, 2022. [Online]. Available: `https://arxiv.org/abs/2206.07542`.

[41] S. P. Fitzgibbon, S. J. Harrison, M. Jenkinson, *et al.*, "The developing human connectome project (dhcp) automated resting-state functional processing framework for newborn infants," *NeuroImage*, vol. 223, p. 117 303, 2020.

[42] C. Morris, M. Ritzert, M. Fey, *et al.*, "Weisfeiler and leman go neural: Higher-order graph neural networks," in *Proceedings*

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.16.562598; this version posted October 17, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

of the AAAI conference on artificial intelligence, vol. 33, 2019, pp. 4602–4609.

[43] S. Liang and R. Srikant, "Why deep neural networks for function approximation?" *arXiv preprint arXiv:1610.04161*, 2016.

[44] E. J. Hughes, T. Winchman, F. Padormo, *et al.*, "A dedicated neonatal brain imaging system," *Magnetic resonance in medicine*, vol. 78, no. 2, pp. 794–804, 2017.

[45] L. Cordero-Grande, E. J. Hughes, J. Hutter, A. N. Price, and J. V. Hajnal, "Three-dimensional motion corrected sensitivity encoding reconstruction for multi-shot multi-slice mri: Application to neonatal brain imaging," *Magnetic resonance in medicine*, vol. 79, no. 3, pp. 1365–1376, 2018.

[46] M. Kuklisova-Murgasova, G. Quaghebeur, M. A. Rutherford, J. V. Hajnal, and J. A. Schnabel, "Reconstruction of fetal brain mri with intensity matching and complete outlier removal," *Medical image analysis*, vol. 16, no. 8, pp. 1550–1564, 2012.

[47] A. Makropoulos, E. C. Robinson, A. Schuh, *et al.*, "The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction," *Neuroimage*, vol. 173, pp. 88–112, 2018.

[48] A. Schuh, A. Makropoulos, R. Wright, *et al.*, "A deformable model for the reconstruction of the neonatal cortex," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, IEEE, 2017, pp. 800–803.

[49] C. Allison, S. Baron-Cohen, S. Wheelwright, *et al.*, "The q-chat (quantitative checklist for autism in toddlers): A normally distributed quantitative measure of autistic traits at 18–24 months of age: Preliminary report," *Journal of autism and developmental disorders*, vol. 38, no. 8, pp. 1414–1425, 2008.

[50] C. A. Albers and A. J. Grieve, "Test review: Bayley, n.(2006). bayley scales of infant and toddler development–third edition. san antonio, tx: Harcourt assessment," *Journal of Psychoeducational Assessment*, vol. 25, no. 2, pp. 180–190, 2007.

[51] A. Makropoulos, I. S. Gousias, C. Ledig, *et al.*, "Automatic whole brain mri segmentation of the developing neonatal brain," *IEEE transactions on medical imaging*, vol. 33, no. 9, pp. 1818–1831, 2014.

[52] M. F. Glasser and D. C. Van Essen, "Mapping human cortical areas in vivo based on myelin content as revealed by t1-and t2-weighted mri," *Journal of Neuroscience*, vol. 31, no. 32, pp. 11 597–11 616, 2011.

[53] J. Bozek, A. Makropoulos, A. Schuh, *et al.*, "Construction of a neonatal cortical surface atlas using multimodal surface matching in the developing human connectome project," *NeuroImage*, vol. 179, pp. 11–29, 2018.

[54] E. C. Robinson, S. Jbabdi, M. F. Glasser, *et al.*, "Msm: A new flexible framework for multimodal surface matching," *Neuroimage*, vol. 100, pp. 414–426, 2014.

[55] E. C. Robinson, K. Garcia, M. F. Glasser, *et al.*, "Multimodal surface matching with higher-order smoothness constraints," *Neuroimage*, vol. 167, pp. 453–465, 2018.

[56] D. Marcus, J. Harwell, T. Olsen, *et al.*, "Informatics and data mining tools and strategies for the human connectome project," *Frontiers in neuroinformatics*, vol. 5, p. 4, 2011.

[57] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch Geometric," in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[58] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model cnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017.

[59] L. Z. Williams, S. P. Fitzgibbon, J. Bozek, *et al.*, "Structural and functional asymmetry of the neonatal cerebral cortex," *Nature Human Behaviour*, pp. 1–14, 2023.

[60] S. J. Counsell, M. A. Rutherford, F. M. Cowan, and A. D. Edwards, "Magnetic resonance imaging of preterm brain injury," vol. 88, no. 4, F269–F274, 2003.

[61] R. Schiffmann and M. S. van der Knaap, "Invited article: An mri-based approach to the diagnosis of white matter disorders," *Neurology*, vol. 72, no. 8, pp. 750–759, 2009.

[62] M. P. Pittet, L. Vasung, P. S. Huppi, and L. Merlini, "Newborns and preterm infants at term equivalent age: A semi-quantitative assessment of cerebral maturity," *NeuroImage: Clinical*, vol. 24, p. 102 014, 2019.

[63] F. Zhao, S. Xia, Z. Wu, *et al.*, "Spherical u-net on cortical surfaces: Methods and applications," *CoRR*, vol. abs/1904.00906, 2019. arXiv: `1904.00906`. [Online]. Available: `http://arxiv.org/abs/1904.00906`.