# Neurodevelopmental Phenotype Prediction:
# A State-of-the-Art Deep Learning Model

**Dániel Unyi**                                                                    UNYI.DANIEL@TMIT.BME.HU

**Bálint Gyires-Tóth**                                                             TOTH.B@TMIT.BME.HU

*Department of Telecommunications and Media Informatics*

*Budapest University of Technology and Economics*

*Műegyetem rkp. 3., H-1111 Budapest, Hungary*

## Abstract

A major challenge in medical image analysis is the automated detection of biomarkers from neuroimaging data. Traditional approaches, often based on image registration, are limited in capturing the high variability of cortical organisation across individuals. Deep learning methods have been shown to be successful in overcoming this difficulty, and some of them have even outperformed medical professionals on certain datasets. In this paper, we apply a deep neural network to analyse the cortical surface data of neonates, derived from the publicly available Developing Human Connectome Project (dHCP). Our goal is to identify neurodevelopmental biomarkers and to predict gestational age at birth based on these biomarkers. Using scans of preterm neonates acquired around the term-equivalent age, we were able to investigate the impact of preterm birth on cortical growth and maturation during late gestation. Besides reaching state-of-the-art prediction accuracy, the proposed model has much fewer parameters than the baselines, and its error stays low on both unregistered and registered cortical surfaces.

**Keywords:** Biomarker Detection, Neuroimaging, Deep Neural Network, Cortical Surface, Developing Human Connectome Project

## 1. Introduction

The goal of neuroimaging is collecting visual data of the nervous system, and analysing this data in order to better understand its structure and function. High-resolution images are primarily acquired through magnetic resonance imaging (MRI), which allows non-invasive and non-radioactive measurements on the patient. Large, publicly available MRI datasets and the corresponding processing pipelines (Elam et al., 2021; Glasser et al., 2013) enable the research community to work on novel machine learning models, aimed to improve our understanding of pathological conditions, and to assist medical experts in diagnosis, prognosis and therapeutic suggestions.

Deep learning models have two crucial advantages over other machine learning models: 1) they can automatically identify the most important biomarkers, automating the meticulous work known as feature selection 2) they can exploit geometric priors, like the spatial correlation of 2D/3D MRI images or surfaces. With state-of-the-art accuracy, novel deep learning methods provide valuable contributions to the field of neuroimaging, for instance the detection of neurodegenerative diseases (Farooq et al., 2017; Barbaroux et al., 2020), segmentation of brain tumors (Pereira et al., 2016), or parcellation of the cerebral cortex (Fawaz et al.,

2021). Some of the methods are even able to outperform medical professionals on certain datasets (Liu et al., 2019).

The cerebral cortex is a thin, highly convoluted neural tissue, constituting the outer layer of the brain. Billions of neurons live within its folds and grooves, facilitating complex cognitive processes which ultimately make us humans and ourselves. Due to its vital role and complexity, it is intensively studied by the machine learning research community. The most common approaches for processing cortical MRI images are convolutional neural networks (He et al., 2016), and more recently, vision transformers (Dosovitskiy et al., 2020). Other approaches involve the extension of convolutions to manifolds (Bronstein et al., 2021), based on the reconstruction of the cortical surface as a polygon mesh.

In this paper, we extend a previous line of work, organised around the Developing Human Connectome Project (dHCP). We apply a deep neural network to identify neurodevelopmental biomarkers, and based on these biomarkers, to predict scan age and birth age of preterm and term neonates. During late gestation, rapid cortical growth and maturation occurs. Since preterm birth interrupts essential cellular and molecular processes, it increases the risk of developing neuropsychiatric conditions in later life. The main contributions of this work are the following:

- Our deep learning model achieves state-of-the-art accuracy in scan age prediction and birth age prediction.

- Using SHAP values, we investigate which myelinated cortical areas has the most significant effect onto the model's final predictions.

- We describe the slight modifications we made to our model, to win an international machine learning challenge in this topic.

## 2. Related Work

The cerebral cortex is the highly convoluted outer layer of the brain, often reconstructed from structural MRI scans as a polygon mesh. The deep learning based analysis of polygon meshes requires the generalisation of convolution from Euclidean grids to manifolds, which led to the development of geometric deep learning (gDL) and intrinsic mesh CNNs (Bronstein et al., 2021).

Many previous works applied gDL techniques directly on the mesh-based representation of the cortical surface. Cucurull et al. (2018) framed cerebral cortex parcellation as a mesh segmentation task. By utilizing graph neural networks (Defferrard et al., 2016; Veličković et al., 2017), they segmented the Broca area of 100 participants from the Human Connectome Project, and significantly improved the Dice score over algorithmic baselines and multilayer perceptrons (which processed the cortical features independently, just as we did in this paper). A similar approach was taken by Gopinath et al. (2019) to parcellation: the aligned spectral embedding of the cortical meshes were concatenated to the other cortical features, before feeding them into a graph neural network. They segmented the whole cortical surface of 101 subjects from the Mindboggle dataset, and achieved better Dice score than the FreeSurfer algorithm (Fischl et al., 2004), one of the most popular automatic parcellation tools.

Others prefer to inflate the cortical meshes and project them onto a sphere, which is commonly referred to as "sphericalisation" (Makropoulos et al., 2018). There are various gDL approaches explicitly designed for spherical manifolds (Cohen et al., 2018; Defferrard et al., 2020; Zhao et al., 2019). According to Barbaroux et al. (2020), spherical CNNs show superior performance to traditional CNNs in classifying Alzheimer's dis-

ease versus cognitively normal, on 589 patients from the ADNI-1 cohort. Zhao et al. (2019) proposed Spherical U-Net, by replacing all operations in the standard U-Net (Ronneberger et al., 2015) with equivalent spherical operations. They demonstrated the capability of their model not only in parcellation, but in predicting the cortical feature development of 370 infants in their first year after birth.

Recently, a large body of work was organised around the Developing Human Connectome Project (dHCP) - a publicly available dataset of cortical surfaces, acquired from preterm and term neonates. Fawaz et al. (2021) benchmarked a collection of gDL methods, comparing them with each other and ResNet (He et al., 2016) on sphericalised cortical surfaces. They aimed to predict the postmenstrual age (PMA) at scan and gestational age (GA) at birth. On registered surfaces, the encoder of Spherical U-Net (Zhao et al., 2019) performed the best, but this performance drastically dropped on unregistered surfaces. On unregistered surfaces, the rotationally equivariant Spherical CNN (Cohen et al., 2018) and MoNet (Monti et al., 2017) were the best performing models. The authors briefly investigated whether sphericalisation holds any disadvantage, but the corresponding results were not conclusive. In the follow-up work by Dahan et al. (2022), they adapted vision transformers to sphericalised cortical surfaces (SiTs), representing data as a sequence of triangular patches. SiTs scored similarly or better than gDL methods in both prediction tasks, and could gain benefit from pretraining on ImageNet or on the same dataset using self-supervised learning (e.g. masked patch prediction). To encourage the endeavours of developing novel methods regarding GA regression at birth, the authors organised a MICCAI 2022 Challenge with deadline of 8th July, 2022, leading up to our current paper.

## 3. Methods

### 3.1. Data

Our initial dataset was derived from the publicly available third release of the Developing Human Connectome Project (dHCP). The reconstructed T1 and T2-weighted magnetic resonance images (Kuklisova-Murgasova et al., 2012; Hughes et al., 2017; Cordero-Grande et al., 2018) were processed according to the dHCP structural pipeline[1] (Makropoulos et al., 2018), which performed tissue segmentation (Makropoulos et al., 2012, 2014, 2016), surface extraction (Schuh et al., 2017), and inflation and projection to a sphere. This process also generates cortical feature descriptors, including three descriptors of cortical geometry (sulcal depth, curvature, thickness) and the T1w/T2w ratio, highly correlated to cortical myelination (Glasser et al., 2013).

The dataset contains the cortical surface of 588 neonatal subjects, with accompanying metadata such as gestational age (GA) at birth, postmenstrual age (PMA) at scan, sex, birthweight, head circumference, and radiology score. Table 1 and Figure 1 show insight into the dataset for both tasks. Some of the preterm neonates were scanned twice: once around birth and once at term-equivalent age. Since our goal is to predict scan age as a marker of healthy development, and to explore the impact of preterm birth on term-equivalent development, we excluded the later preterm scans from scan age regression and the earlier preterm scans from birth age regression (also important for the direct comparison with previous methods).

We preprocessed the initial dataset as follows. For experiments in native space (i.e. unregistered data), we used each subject's native sphere, while in template space (i.e. registered data), we used the transform from

---

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

Table 1: Dataset description.

| Dataset | Total | Maturity | | Sex | |
|---|---|---|---|---|---|
| | | Preterm | Term | Male | Female |
| PMA at scan | 530 | 111 | 419 | 288 | 242 |
| GA at birth | 514 | 95 | 419 | 283 | 231 |

each subject's native surface to the dHCP symmetric 40-week surface template. The 4 cortical features (sulcal depth, curvature, thickness, and T1w/T2w ratio) were resampled from the native/template sphere to the 6-th order icosphere with barycentric interpolation, implemented by the Human Connectome Project workbench software (Marcus et al., 2011). This pipeline results 40962 equally spaced vertices per subject, with 4 cortical features per vertex. Right hemispheres were mirrored onto the sagittal plane, such that their orientation matched with left hemispheres. We computed the mean and standard deviation of the 4 features in the train set, and standardized the input data every time before training or inference.

### 3.2. Model

We used a multilayer perceptron (MLP) in each of our experiments. The architecture is shown on Fig. 2. We applied the following layer sequence four times: $\{Linear() \rightarrow Tanh() \rightarrow BatchNorm()\}$. After that, we performed mean pooling along the vertex dimension, and added a regression head $Linear() \rightarrow Tanh() \rightarrow Linear()$ to produce the final predictions. Each hidden layer had 16 units, resulting fewer than 1400 trainable parameters. We had 4 input features in scan age prediction, but 5 in birth age prediction, since in this case we also incorporated scan age as a confound. The model was trained using mean squared error (MSE) as loss function, and Adam as optimizer with learning rate 0.001. Batch size was set to 32. Training was halted when the validation loss had not decreased in the last 200 epochs, and the model with the lowest validation loss was restored.

## 4. Results and Discussion

### 4.1. State-of-the-Art Results

For each of the four tasks (scan age prediction and birth age prediction in both native and template space), we trained an MLP model four times with different initial weights, and monitored the mean absolute error (MAE) in weeks[2]. For direct comparison with previous models (Fawaz et al., 2021; Dahan et al., 2022), we report the best MAE score achieved on the test set, alongside the standard deviation across all four trainings, in Table 2. We compared our MLP to the following geometric deep learning methods: S2CNN (Cohen et al., 2018), ChebNet (Defferrard et al., 2016), GConvNet (Kipf and Welling, 2016), Spherical U-Net (Zhao et al., 2019), MoNet (Monti et al., 2017), and the following surface vision transformers: SiT-0 was not pretrained, SiT-1 was pretrained on ImageNet, and SiT-2 was pretrained via masked patch prediction on the dHCP dataset (Dahan et al., 2022). Overall, our MLP model consistently outperformed the other models in native space, with MAE of 0.54 weeks in scan age prediction and 1.08 weeks in birth age prediction. It was also

---

2. Code to reproduce the results is publicly available at https://github.com/daniel-unyi-42/Neurodevelopmental-Phenotype-Prediction
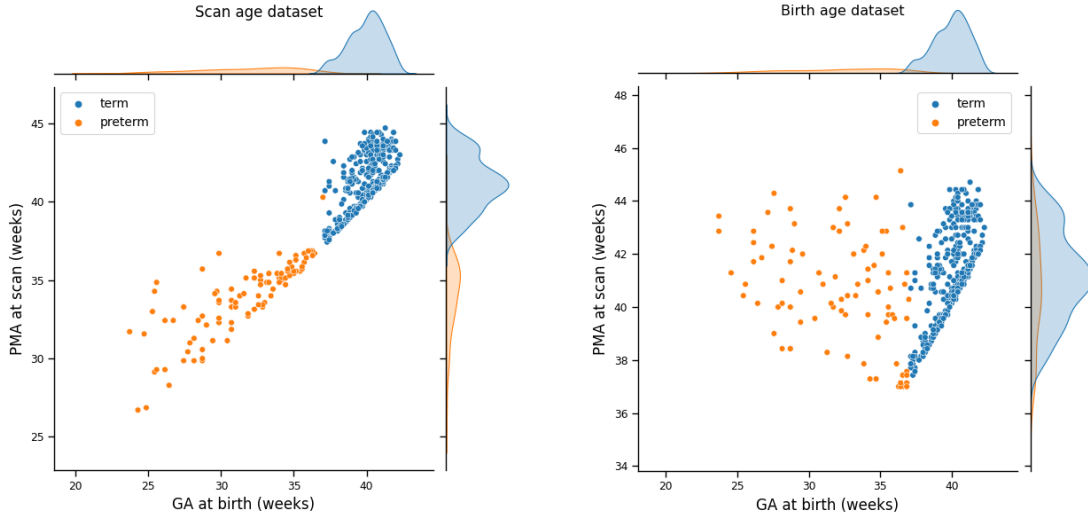
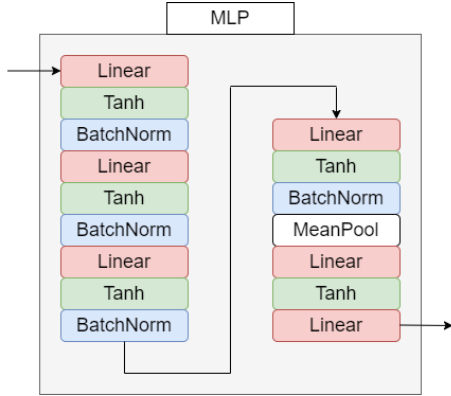Figure 1: Distribution of the scan age and birth age datasets.



Figure 2: Our proposed architecture.

competitive compared to the best performing models in template space, with the exception of Spherical U-Net in birth age prediction, which still leads the race by a large margin. Incorporating scan age as a confound in birth age prediction yielded lower error, in agreement with the findings of Dahan et al. (2022). We also performed 10-fold cross validation, and report the mean and standard deviation across the 10 different test folds in Table 2. We think this is a fairer way to evaluate and compare future models trained on this dataset.

It is important to note that our model has less than 1400 trainable parameters, a fraction of the several million parameters in previous models. We attributed the success of our lightweight MLP model to two reasons: 1) it can generalise better on unseen data 2) it was easier to optimise its hyperparameters (for instance, it could not train as well with asymmetric activation functions as with tanh). However, it is possible that message passing models inherently overfit this dataset: changing the first 4 linear layers to the convolutional layers of ChebNet, GConv-Net, or MoNet results an immediate drop in performance, regardless of the small number of parameters. Using an NVIDIA TITAN V GPU, our model requires only an average of 834 ms training time per epoch and 76 ms inference time.

## 4.2. Identifying Neurodevelopmental Biomarkers

Since our model was successful at detecting biomarkers, it is worth to investigate how it concluded the predicted values of scan age and birth age. For this purpose, we

283

Table 2: Best **M**ean **A**bsolute **E**rror and standard deviation across 4 runs are reported (in weeks). Our MLP model achieves better results than surface vision transformers (Dahan et al., 2022) and geometric deep learning methods (Fawaz et al., 2021), with only a fraction of trainable parameters. We also report the result of 10-fold cross validation in the last row.

| Method | Scan age | | | Birth age | | |
|---|---|---|---|---|---|---|
| | **Template** | **Native** | **Avg** | **Template** | **Native** | **Avg** |
| S2CNN | 0.63±0.02 | 0.73±0.25 | 0.68 | 1.35±0.68 | 1.52±0.60 | 1.44 |
| ChebNet | 0.59±0.37 | 0.77±0.49 | 0.68 | 1.57±0.15 | 1.70±0.36 | 1.64 |
| GConvNet | 0.75±0.13 | 0.75±0.26 | 0.75 | 1.77±0.26 | 2.30±0.74 | 2.04 |
| Spherical UNet | 0.57±0.18 | 0.87±0.50 | 0.72 | **0.85±0.17** | 2.16±0.57 | 1.51 |
| MoNet | 0.57±0.02 | 0.61±0.05 | 0.59 | 1.44±0.08 | 1.58±0.06 | 1.51 |
| SiT-0 | 0.60±0.02 | 0.76±0.03 | 0.68 | 1.14±0.12 | 1.44±0.03 | 1.29 |
| SiT-1 | 0.59±0.03 | 0.71±0.02 | 0.65 | 1.15±0.05 | 1.69±0.03 | 1.42 |
| SiT-2 | 0.55±0.04 | 0.63±0.06 | 0.59 | 1.13±0.02 | 1.47±0.08 | 1.30 |
| **MLP(ours)** | **0.50±0.05** | **0.54±0.01** | **0.52** | 1.08±0.05 | **1.08±0.17** | **1.08** |
| **MLP-cv(ours)** | 0.55±0.05 | 0.55±0.03 | 0.55 | 1.00±0.12 | 1.12±0.22 | 1.06 |
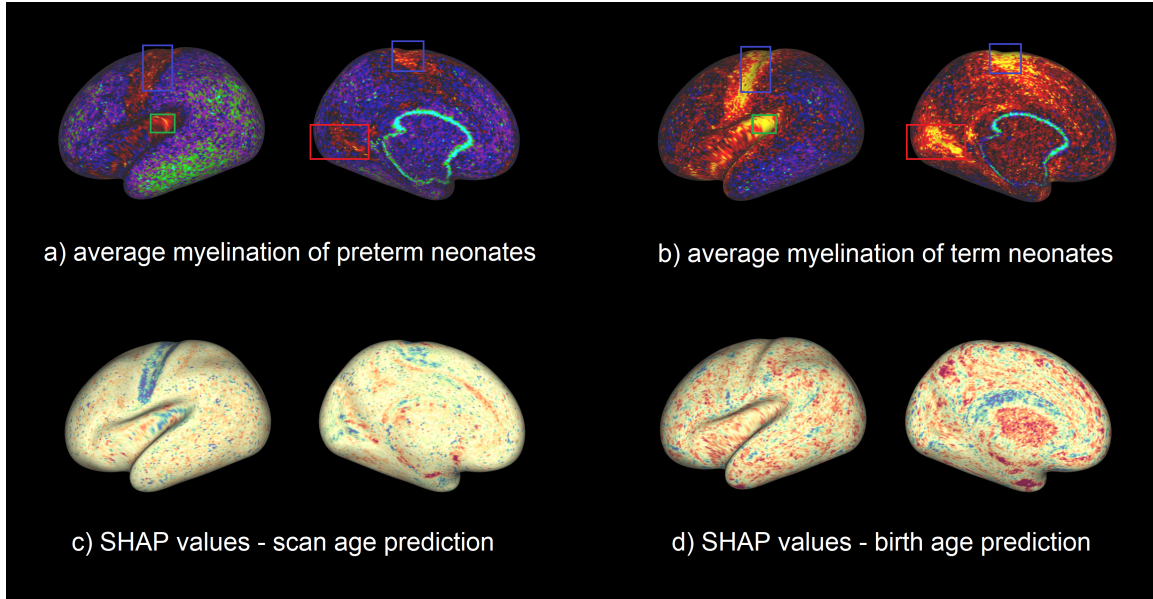


Figure 3: a) and b) show the average cortical myelination of preterm and term neonates. The increase of myelin levels can be seen in the visual (red box), auditory (green box), and somatosensory (blue box) areas. c) and d) highlight the SHAP values in scan age and birth prediction, with red and blue colours indicating that the given area is important in the model's prediction making.

used the DeepExplainer method (Shrikumar et al., 2017) of the SHapley Additive ExPlanations (SHAP) library[3] (Lundberg and Lee, 2017). SHAP assigns a so-called SHAP value to each feature, indicating its importance in the model's prediction making.

We averaged the T1w/T2w ratio of preterm (GA<37 weeks) and term (GA>37 weeks) neonates separately over the validation and test sets, to visualise the dependence of cortical myelination on maturity. Figure 3 highlights the difference of myelin levels in the visual, auditory and somatosensory areas between the two gestational periods. SHAP values indicate that the latter two areas were indeed important indicators in scan age prediction, marking the process of healthy development. In birth age prediction, other parts of the cortex become more important, including the superior frontal gyrus, the praecuneus, or the temporal pole. These areas are responsible for high-level cognitive functions, like certain aspects of consciousness, memory, or visuospatial processing, and are likely impacted by preterm birth.

### 4.3. SLCN: A MICCAI 2022 Challenge

Our solution ranked first on the MICCAI 2022 Challenge called Surface Learning for Clinical Neuroimaging. The goal of this challenge was to elicit submissions of explainable machine learning methods, which are able to identify neurodevelopmental biomarkers from cortical surface data. Thereby, submissions were evaluated on the task of birth age regression. The same train and validation sets (acquired from the dHCP) were provided to the participants that we used throughout this paper, but the test set was kept hidden during the challenge. Metadata was also provided for subjects in the train and validation

sets: the PMA at scan, sex, birthweight, and head circumference. Since the PMA at scan was not available for test set subjects, we had to work out a less straightforward deconfounding strategy. The key observation was that predicting the confound and incorporating it into the loss function consistently decreased the error on the test set. Predicting the birthweight variable and incorporating it into the loss also yielded better results: the standard deviation decreased among validation errors (i.e. evaluations of the same model with different initial weights). This is understandable, since it is strongly correlated with our target variable. In conclusion, the only modification we made to our MLP model was adding 3 output units instead of 1. These 3 output units correspond to the GA at birth, the PMA at scan, and the birthweight of the subject; and since we incorporated all 3 variables into the loss function, we required the MLP model to predict all of them as accurately as possible. Our best test MAE was 1.226 in template space and 1.386 in native space.

## 5. Conclusion

In this work, we applied a deep neural network to predict neurodevelopmental phenotypes, based on the cortical surface data of preterm and term neonates. We achieved state-of-the-art results compared to previous baselines, demonstrating that models with fewer priors and fewer parameters can be more effective on certain datasets. We also identified some myelinated cortical areas that are biomarkers of healthy development, and others that are possible biomarkers of preterm birth.

Regarding future research, we plan to dissipate the debate around whether or not sphericalisation leads to the loss of useful information (as we mentioned, results in Fawaz et al. (2021) are not conclusive). We are

---

3. https://github.com/slundberg/shap

going to apply intrinsic mesh CNNs on the high-resolution, native cortical meshes, and identify the cortical areas whose geometry may affect the predictions. We also plan to explore further metadata like biological sex, utilizing it both as confound and as target variable.

## Acknowledgments

## References

Hugo Barbaroux, Xinyang Feng, Jie Yang, Andrew F. Laine, and Elsa D. Angelini. Encoding human cortex using spherical CNNs - A study on Alzheimer's disease classification. *IEEE 17th International Symposium on Biomedical Imaging (ISBI 2020)*, pages 234–241, 2020.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Taco S. Cohen, Mario Geiger, Jonas Koehler, , and Max Welling. Spherical CNNs. *International Conference on Learning Representations (ICLR)*, 2018.

Lucilio Cordero-Grande, Emer J. Hughes, Jana Hutter, Anthony N. Price, and Joseph V. Hajnal. 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, 2018.

Guillem Cucurull, Konrad Wagstyl, Arantxa Casanova, Petar Veličković, Estrid Jakobsen, Michal Drozdzal, Adriana Romero, Alan Evans, and Yoshua Bengio. Convolutional neural networks for mesh-based parcellation of the cerebral cortex. *Medical Imaging with Deep Learning (MIDL)*, 2018.

Simon Dahan, Abdulah Fawaz, Logan Z. J. Williams, Chunhui Yang, Timothy S. Coalson, Matthew F. Glasser, A. David Edwards, Daniel Rueckert, and Emma C. Robinson. Surface vision transformers: Attention-based modelling applied to cortical analysis. *Medical Imaging with Deep Learning (MIDL)*, 2022.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

Michaël Defferrard, Martino Milani, Frédérick Gusset, and Nathanaël Perraudin. DeepSphere: a graph-based spherical CNN. *arXiv preprint arXiv:2012.15000*, 2020.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Jennifer Stine Elam, Matthew F.Glasser, Michael P.Harms, Stamatios N. Sotiropoulos, Jesper L. R. Andersson, Gregory C.Burgess, Sandra W. Curtiss, Robert Oostenveld, Linda J. Larson-Prior, Jan-Mathijs Schoffelen, Michael R.Hodge, Eileen A.Cler, Daniel M.Marcus,

Deanna M. Barch, Essa Yacoub, Stephen M. Smith, Kamil Ugurbil, and David C. Van Essen. The Human Connectome Project: A retrospective. *NeuroImage*, 244, 2021.

Ammarah Farooq, SyedMuhammad Anwar, Muhammad Awais, and Saad Rehman. A deep CNN based multi-class classification of Alzheimer's disease using MRI. *IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 1–6, 2017.

Abdulah Fawaz, Logan Z. J. Williams, Amir Alansary, Cher Bass, Karthik Gopinath, Mariana da Silva, Simon Dahan, Chris Adamson, Bonnie Alexander, Deanne Thompson, Gareth Ball, Christian Desrosiers, Hervé Lombaert, Daniel Rueckert, A. David Edwards, and Emma C. Robinson. Benchmarking geometric deep learning for cortical segmentation and neurodevelopmental phenotype prediction. *bioRxiv preprint 2021.12.01.470730*, 2021.

Bruce Fischl, André van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H. Salat, Evelina Busa, Larry J. Seidman, Jill Goldstein, David Kennedy, Verne Caviness, Nikos Makris, Bruce Rosen, and Anders M. Dale. Automatically parcellating the human cerebral cortex. *Cerebral cortex*, 14(1):11–22, 2004.

Matthew F. Glasser, Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L.Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R. Polimeni, David C. Van Essen, and Mark Jenkinson. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

Karthik Gopinath, Christian Desrosiers, and Herve Lombaert. Graph convolutions on spectral embeddings for cortical surface parcellation. *Medical image analysis*, 54: 297–305, 2019.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

Emer J. Hughes, Tobias Winchman, Francesco Padormo, Rui Teixeira, Julia Wurie, Maryanne Sharma, Matthew Fox, Jana Hutter, Lucilio Cordero-Grande, Anthony N. Price, Joanna Allsop, Jose Bueno-Conde, Nora Tusor, Tomoki Arichi, A. D. Edwards, Mary A. Rutherford, Serena J. Counsell, , and Joseph V. Hajnal. A dedicated neonatal brain imaging system. *Magnetic Resonance in Medicine*, 78(2): 794–804, 2017.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Maria Kuklisova-Murgasova, Gerardine Quaghebeur, Mary A. Rutherford, Joseph V. Hajnal, and Julia A. Schnabel. Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. *Medical Image Analysis*, 16(8): 1550–1564, 2012.

Xiaoxuan Liu, Livia Faes, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, Gabriella Moraes, Mohith Shamdas, Christoph Kern, Joseph R Ledsam, Martin K Schmid, Konstantinos Balaskas, Eric J Topol, Prof Lucas M Bachmann, Pearse A Keane, and Prof Alastair K Denniston. A comparison of deep learning performance against health-care pro-

fessionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet Digital Health*, 1(6):271–297, 2019.

Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 2017.

Antonios Makropoulos, Christian Ledig, Paul Aljabar, Ahmed Serag, Joseph V. Hajnal, A. David Edwards, Serena J. Counsell, and Daniel Rueckert. Automatic tissue and structural segmentation of neonatal brain MRI using expectation-maximization. *MICCAI Grand Challenge on Neonatal Brain Segmentation*, pages 9–15, 2012.

Antonios Makropoulos, Ioannis S. Gousias, Christian Ledig, Paul Aljabar, Ahmed Serag, Joseph V. Hajnal, A. David Edwards, Serena J. Counsell, and Daniel Rueckert. Automatic whole brain MRI segmentation of the developing neonatal brain. *IEEE Transactions on Medical Imaging*, 33(9):1818–1831, 2014.

Antonios Makropoulos, Paul Aljabar, Robert Wright, Britta Hüning, Nazakat Merchant, Tomoki Arichi, Nora Tusor, Joseph V. Hajnal, A. David Edwards, Serena J.Counsell, and Daniel Rueckert. Regional growth and atlasing of the developing human brain. *Neuroimage*, 125:456–478, 2016.

Antonios Makropoulos, Emma C. Robinson, Robert Wright Andreas Schuh, Sean Fitzgibbon, Jelena Bozek, Serena J. Counsell, Johannes Steinweg, Katy Vecchiato, Jonathan Passerat-Palmbach, Gregor Lenz, Filippo Mortari, Tencho Tenev, Eugene P. Duff, Matteo Bastiani, Lucilio Cordero-Grande, Emer Hughes, Nora Tusor, Jacques Donald Tournier, Jana Hut-

ter, Anthony N. Price, Rui Pedro A.G. Teixeira, Maria Murgasova, Suresh Victor, Christopher Kelly, Mary A. Rutherford, Stephen M. Smith, A. David Edwards, Joseph V. Hajnal, Mark Jenkinson, and Daniel Rueckert. The Developing Human Connectome Project: A minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*, 173:88–112, 2018.

Daniel S. Marcus, John Harwell, Timothy Olsen, Michael Hodge, Matthew F. Glasser, Fred Prior, Mark Jenkinson, Timothy Laumann, Sandra W. Curtiss, and David C. Van Essen. Informatics and data mining tools and strategies for the Human Connectome Project. *Frontiers in Neuroinformatics*, 5:4, 2011.

Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model CNNs. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5115–5124, 2017.

Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A. Silva. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, 35(5):1240–1251, 2016.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

Andreas Schuh, Antonios Makropoulos, Robert Wright, Emma C. Robinson, Nora Tusor, Johannes Steinweg, Emer Hughes, Lucilio Cordero Grande, Anthony Price, Jana Hutter, Joseph V. Hajnal, and Daniel Rueckert. A deformable model for the reconstruction of the neonatal cortex. *IEEE*

*14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 800–803, 2017.

Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning important features through propagating activation differences. *International Conference on Machine Learning*, pages 3145–3153, 2017.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Fenqiang Zhao, Shunren Xia, Zhengwang Wu, Dingna Duan, Li Wang, Weili Lin, John H. Gilmore, Dinggang Shen, and Gang Li. Spherical U-Net on cortical surfaces: methods and applications. *International Conference on Information Processing in Medical Imaging*, pages 855–866, 2019.