

SuD-CoTAN: Sulcal Depth-guided Anatomically Consistent Fetal Cortical Surface Reconstruction

Irina Grigorescu^{1,2} 

IRINA.GRIGORESCU@KCL.AC.UK

Jiaxin Xiao²

JIAXIN.1.XIAO@KCL.AC.UK

Yourong Guo²

YOURONG.GUO@KCL.AC.UK

Vanessa Kyriakopoulou¹

VANESSA.KYRIAKOPOULOU@KCL.AC.UK

Alena Uus^{1,2}

ALENA.UUS@KCL.AC.UK

Vyacheslav Karolis^{1,2}

SLAVA.KAROLIS@KCL.AC.UK

Kaili Liang²

KAILI.LIANG@KCL.AC.UK

Mohamed A. Suliman²

MOHAMED.SULIMAN@KCL.AC.UK

Qiang Ma³

Q.MA20@IMPERIAL.AC.UK

Daniel Rueckert³

D.RUECKERT@IMPERIAL.AC.UK

Bernhard Kainz³

B.KAINZ@IMPERIAL.AC.UK

A. David Edwards¹

AD.EDWARDS@KCL.AC.UK

Joseph V. Hajnal¹

JO.HAJNAL@KCL.AC.UK

Mary Rutherford¹

MARY.RUTHERFORD@KCL.AC.UK

Maria Deprez^{*1,2}

MARIA.DEPREZ@KCL.AC.UK

Emma C. Robinson^{*1,2}

EMMA.ROBINSON@KCL.AC.UK

¹ Department of Early Life Imaging, School of Biomedical Engineering & Imaging Sciences, King's College London, London, United Kingdom

² Department of Biomedical Computing, School of Biomedical Engineering & Imaging Sciences, King's College London, London, United Kingdom

³ BioMedIA, Department of Computing, Imperial College London, London, UK

Editors: Under Review for MIDL 2026

Abstract

Accurate and anatomically consistent fetal cortical surface reconstruction is essential for studying early brain development, yet existing methods often lack reliable vertex-wise correspondence and fail to harmonise their outputs across heterogeneous MRI datasets. We introduce Sulcal Depth-guided CoTAN (SuD-CoTAN), a learning-based framework that fits anatomically and topologically consistent cortical meshes directly to T2-weighted MRI and performs alignment to age-matched templates in one single step. All models are trained exclusively on normative samples from the developing Human Connectome Project (dHCP) and evaluated within-sample and on a different acquisition protocol. Results show that SuD-CoTAN generalises to new datasets in ways that harmonise global morphometric properties by better capturing the surface geometry of individual cases; its template fitting is precise, delivering vertex-wise anatomical correspondences that result in sharp weekly averages of sulcal depth and curvature maps in template space. This supports direct vertex-wise Gaussian Process regression of neurodevelopmental trends without a need for any additional registration. Collectively, this whole pipeline runs in ~3 seconds. This suggests that SuD-CoTAN offers promise as a screening tool for cortical malformations during fetal development.

Keywords: fetal MRI, cortical surface reconstruction, deep learning

* Contributed equally

1. Introduction

Characterising fetal cortical development is central to detecting conditions where the folding process deviates from typical trajectories. In the fetal period, disruptions to cortical folding are associated with a range of neurodevelopmental disorders, making precise characterisation of surface geometry critical for quantifying typical and atypical maturation (Garcia et al., 2025; Story et al., 2021; Kyriakopoulou et al., 2014). However, obtaining surfaces that are both anatomically meaningful and harmonised across magnetic resonance imaging (MRI) protocols, remains challenging, particularly in fetal MRI.

Current approaches for surface extraction often rely on deformation of the template surface mesh in order to preserve topology of the cortical surfaces (Schuh et al., 2017; Zöllei et al., 2020; Ma et al., 2022) or implicit representations (Cruz et al., 2021; Wang et al., 2023; Gopinath et al., 2023); however, these approaches cannot guarantee vertex-wise anatomical correspondences across the population. These correspondences need to be estimated post-hoc, using spherical projection and registration (Robinson et al., 2014, 2018; Besenczi et al., 2024), to enable accurate detection of subtle cortical abnormalities through high-resolution (vertex-level) comparisons. Additionally, changes in scanner and reconstruction protocols translate to differences in apparent resolution, contrast, and partial volume, resulting in inconsistent cortical surface measures across different acquisition protocols. Classical neonatal pipelines (Makropoulos et al., 2018) can produce high-quality surfaces for neonatal MRI, yet they fail to generalise to fetal scans, where lower resolution leads to partial-volume effects that blur fine cortical structures (see Figure 1).

Current best practice requires lengthy processing, involving several error-prone steps, including intensity-based tissue segmentation (Scott et al., 2011; Gholipour et al., 2017; Makropoulos et al., 2018; Uus et al., 2021), deformable mesh fitting at the boundaries, inflation (for visualisation), and projection to the sphere followed by surface registration (Robinson et al., 2014, 2018; Besenczi et al., 2024). This makes them highly non-practical for use in the clinic. Ideally, surface reconstruction should instead be robust: generating anatomically and topologically correct meshes, that are harmonised across acquisition protocols and automatically aligned to a normative reference cohort for precision detection of outliers.

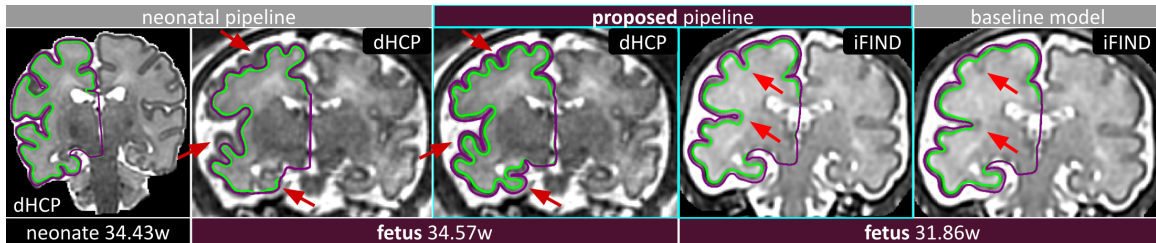


Figure 1: Traditional neonatal surface-reconstruction pipelines do not perform well on fetal MRI (see red arrows), while our proposed SuD-CoTAN framework is able to produce anatomically plausible surfaces for fetal scans from both dHCP and iFIND. The *baseline* deep learning models (see Section 3.2) underperform on the unseen iFIND dataset.

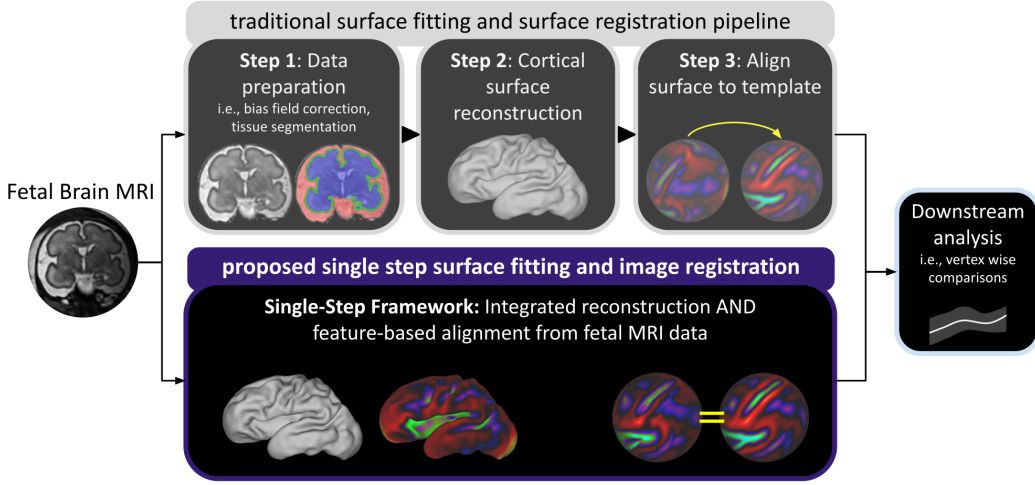


Figure 2: Traditional pipelines (top row) rely on several separate image processing steps: tissue segmentation, followed by surface fitting, followed by spherical alignment to a population-average template for vertex-wise analysis. By contrast, our method (bottom row) integrates template alignment directly into the reconstruction process, producing analysis-ready, template-aligned surfaces directly from raw fetal MRI in a single step.

Contributions: We introduce **Sulcal Depth-guided CoTAN (SuD-CoTAN)**, a learning-based framework for fetal cortical surface reconstruction. **SuD-CoTAN** builds on CoTAN framework (Ma et al., 2023), which produces topologically consistent surfaces. Unlike CoTAN, **SuD-CoTAN** also provides vertex-wise anatomical consistency, where reconstructed surfaces that are directly aligned to weekly fetal templates. Additionally, **SuD-CoTAN** is robust to variations in acquisition protocol, producing consistent surface measures across different fetal datasets. We evaluate our approach on fetal MRI scans from both normative samples from the developing Human Connectome Project (dHCP) and controls from the Intelligent Fetal Imaging and Diagnosis (iFIND) clinical research study, demonstrating geometric accuracy and cross-dataset harmonisation. Results show the method achieves superior surface precision and enhanced harmonisation relative to baseline methods, with sufficient coherence across acquisitions to support the high-resolution (per-vertex) normative modeling of sulcal depth.

2. Methods

Our proposed single-step surface reconstruction and template alignment framework is illustrated in Figure 2. While traditional pipelines depend on brain tissue segmentations to extract cortical surfaces, and rely on spherical registration to templates to enable vertex-wise comparisons, our method operates on fetal MRI directly, and integrates template alignment into the reconstruction, to produce analysis-ready, template-aligned surfaces from one processing step. This pipeline is described in Figure 3, with an architecture that builds from

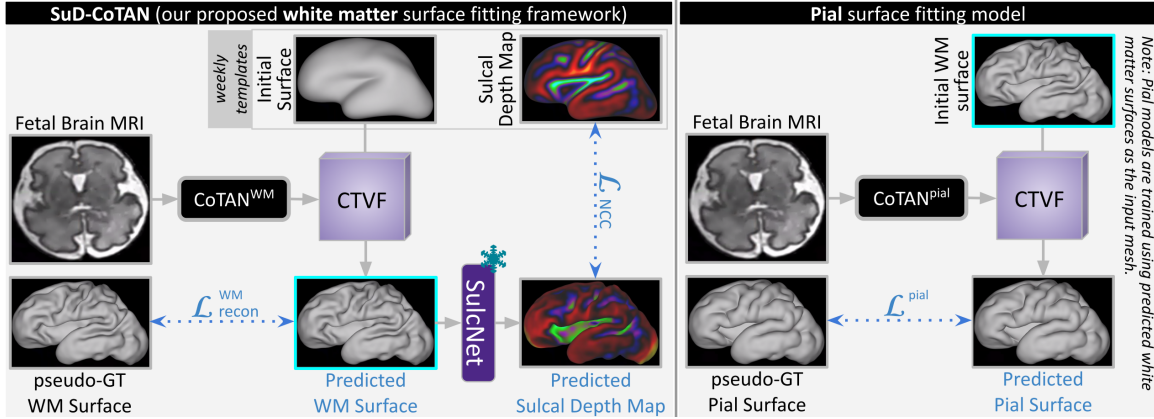


Figure 3: Overview of the **SuD-CoTAN** framework. Left: The model takes fetal brain MRI as input and deforms a gestational-age-specific template surface into a subject-specific WM prediction. A pre-trained **SulcNet** provides sulcal-depth guidance to ensure consistent deformations across subjects. Right: For pial surface reconstruction, the initial meshes are subject-specific predicted WM surfaces and are subsequently deformed into pial surfaces as proposed in (Ma et al., 2024).

CoTAN (Ma et al., 2023) but introduces key methodological and data processing advances that: 1) introduce a novel **SulcNet** module that estimates sulcal depth during training to guide and constrain template alignment; 2) adapt training with augmentations to achieve harmonisation; and 3) allow CoTAN to be run successfully for fetal data for the first time.

2.1. Generation of pseudo-ground truth surfaces and initial templates

CoTAN (Ma et al., 2023) relies on the availability of pseudo-ground truth (pGT) surfaces to supervise training. However, these are non-trivial to obtain (Ma et al., 2024), due to the lower image quality of fetal MRI stemming from challenges associated with scanning free-moving fetuses within the maternal body, where volumetric data need to be reconstructed from stacks of acquired 2D slices (Wright et al., 2018; Cordero-Grande et al., 2019; Kuklisova-Murgasova et al., 2012; Uus et al., 2025). Since acquisition and development impact image contrast and tissue properties, this has required the development of bespoke tissue segmentation algorithms (Uus et al., 2023) adapted from our neonatal protocols (Makropoulos et al., 2014, 2016) through extensive manual correction; followed by adaption of surface-fitting modules (Schuh et al., 2017) to extract approximate fetal pGT surfaces. Initial templates were then generated through use of Multimodal Surface Matching (MSM) (Robinson et al., 2014, 2018; Besenczi et al., 2024) using the methods described in (Kuklisova-Murgasova et al., 2011; Serag et al., 2012; Bozek et al., 2018; Williams et al., 2023; Karolis et al., 2023). This generates separate templates for each week of gestation (from 25–36 weeks) through applying temporally adaptive kernel regression with repeated alignment and averaging. Once defined, all examples were aligned to the 36-week space through template-to-template alignment registrations.

2.2. SuD-CoTAN: Sulcal Depth-guided Cortical Surface Reconstruction

Our proposed **SuD-CoTAN** takes as input a fetal T_2 -weighted (T_2w) MRI scan and the gestational age (GA) of the subject, and predicts a conditional time-varying velocity field (CTVF) that deforms the closest-in-age initial surface into the subject-specific WM (inner) surface (see Figure 3). To enable single-step surface fitting while implicitly enforcing surface-template correspondence, **SuD-CoTAN** integrates **SulcNet** as a feature-based guidance module. Specifically, through minimizing the normalised cross correlation (NCC) between the predicted sulcal depth map and an age-matched template sulcal depth map, **SulcNet** encourages anatomically consistent deformations. For pial (outer) surface reconstruction, a second CoTAN-based network is trained following the formulation introduced in (Ma et al., 2024). In this stage, the initial surface is given by the predicted WM surface from the pre-trained **SuD-CoTAN** model, which is then further deformed via a learned CTVF to match the pGT pial surfaces.

2.3. SulcNet: Sulcal Depth Prediction Network

SulcNet itself is a spherical U-Net (Monti et al., 2017; Zhao et al., 2019), adapted from (Suliman et al., 2022), that takes the 3D coordinates of the deformed WM surface as input and outputs average surface convexity (or sulcal depth) at each vertex. The input to this network corresponds to vertex coordinates from white matter meshes that have been resampled to the resolution and mesh topology of a sixth-order icosphere (with 40,962 vertices) - taking advantage of implicit vertex correspondence of spheres and anatomical (WM) surfaces to define convolutional kernels on the sphere, while learning from coordinates that reflect the native WM geometry. The architecture then follows a six-stage encoder-decoder design, with each resolution defined by progressively coarser/finer icospheres (Suliman et al., 2025), learning MoNet convolutions (Monti et al., 2017) with LeakyReLU activations ($\alpha = 0.2$), hexagonal mean-pooling and transpose-convolutions for upsampling (Zhao et al., 2019). A compact feedforward layer (Suliman et al., 2022) produces the final sulcal depth prediction per vertex. These generated sulcal depth maps are then used to guide template matching through integration of a sulcal depth loss (see Section 2.4).

2.4. Optimisation

SuD-CoTAN is trained with a combination of four losses (Ma et al., 2023, 2024): a bidirectional Chamfer (bi-Chamfer) distance loss (\mathcal{L}_{cd}) (Bongratz et al., 2022) between the predicted and pGT vertices, and three smoothness penalties: a normal consistency loss \mathcal{L}_{nc} , constraining cosine similarity between adjacent face normals, the mesh Laplacian loss \mathcal{L}_{lap} to promote smoothness, and an edge length loss \mathcal{L}_{edge} to discourage irregular or excessively stretched edges. During training, the pre-trained **SulcNet** outputs a sulcal depth map ($SD_{predicted}$) from each predicted WM surface, which is then compared with that of the closest-in-age template ($SD_{template}$) using NCC. The total WM loss becomes:

$$\mathcal{L}^{WM} = \mathcal{L}_{cd} + \lambda_{lap}\mathcal{L}_{lap} + \lambda_{nc}\mathcal{L}_{nc} + \lambda_{edge}\mathcal{L}_{edge} + \lambda_{SD}\mathcal{L}_{NCC}(SD_{predicted}, SD_{template}) \quad (1)$$

For pial surfaces, we follow the previously proposed method from (Ma et al., 2024), where the Chamfer distance is replaced with a single-directional Chamfer loss \mathcal{L}_{cd-1d} between

predicted and pGT pial surfaces. Two additional penalties are applied: an inflation loss \mathcal{L}_{infl} , which constrains surface inflation to follow normal directions (Ma et al., 2024), and a pial-outside-WM loss \mathcal{L}_{pow} (Appendix B), which penalizes pial vertices whose displacement vectors point inward relative to WM normals. The total pial loss becomes:

$$\mathcal{L}^{pial} = \mathcal{L}_{cd-1d} + \lambda_{lap}\mathcal{L}_{lap} + \lambda_{nc}\mathcal{L}_{nc} + \lambda_{edge}\mathcal{L}_{edge} + \lambda_{infl}\mathcal{L}_{infl} + \lambda_{pow}\mathcal{L}_{pow} \quad (2)$$

SulcNet uses a Smooth L1 loss and it is trained independently on pGT meshes and sulcal depth maps derived from normative samples using our classical surface processing pipeline (Uus et al., 2023; Schuh et al., 2017; Makropoulos et al., 2018; Schuh et al., 2017), then frozen during optimisation of the **SuD-CoTAN** network. The final model was selected based on the lowest validation mean squared error (MSE) and NCC.

Training. WM and pial models were trained for 400 epochs using the Adam optimizer. Following (Ma et al., 2023, 2024), we first pre-train the models using relatively large loss weights, then fine-tune them with smaller weights (Appendix C). During training, we apply MONAI (Cardoso et al., 2022) augmentations designed to mimic realistic variations encountered in fetal MRI, namely, bias-field inhomogeneities, Gaussian noise, Gaussian smoothing, gamma contrast adjustments, and random histogram shifts (see Appendix D for more details on how these were parameterised). **SulcNet** was trained for 1000 epochs using an Adam optimizer, incorporating geometric augmentations consisting of small random translations and rotations of the meshes (see Appendix D for more details).

3. Experiments and Results

3.1. Fetal MRI Data Selection and Preprocessing

We evaluate our framework on two fetal T_2w MRI datasets, using one for training (dHCP) and the other exclusively for testing (iFIND). All scans were motion corrected and reconstructed using slice-to-volume registration (SVR) to 0.5mm isotropic resolution (Kuklisova-Murgasova et al., 2012; Wright et al., 2018; Cordero-Grande et al., 2019; Uus et al., 2025). For all experiments, T_2w images were affinely aligned to a 36-week fetal brain atlas (Uus et al., 2021), and in this study we only consider the left hemisphere.

Training Dataset. The training and validation dataset consists of 210 dHCP scans (20.86 - 38.29 weeks GA) acquired on a 3T Philips Achieva MRI system with a 32-channel cardiac coil ($T_E = 250\text{ms}$, $1.1 \times 1.1\text{mm}$ in-plane resolution, and 2.2mm slice thickness) using a dedicated protocol (Price et al., 2019). We use 200 subjects for training and 10 for validation (see Table A.1, “Training” rows).

Testing Dataset. Additional 20 dHCP subjects were reserved for testing, together with 173 iFIND subjects (22.29 - 31.86 weeks GA) scanned on a 1.5T Philips Ingenia MRI system using a 28-channel torso coil ($T_E = 80\text{ms}$, $1.25 \times 1.25\text{mm}$ in-plane resolution, and 2.5mm slice thickness). We select 20 iFIND subjects, age-matched to the dHCP test set, and generate pGT surfaces using the same strategy. We use these dHCP and iFIND subjects to evaluate the performance of our proposed model (see Table A.1, “Evaluation” rows). For subsequent cortical metric analyses, we select a subset of 140 subjects from the entire dHCP dataset to match the GAs of the 173 iFIND subjects (see Table A.1, “Feature Extraction” rows).

Table 1: White matter and pial surface predictions on the dHCP and iFIND test datasets, showing ASSD and HD between the predicted and the pGT surfaces, and the ratio of SIF in the predicted surfaces. Last row shows average standard deviation across mean sulcal depth maps calculated from the predicted white matter surfaces. Values in bold represent best performing models ($p < .05$).

	Metric	bline	bline+SD	bline+aug	SuD-CoTAN
WM	ASSD ↓	0.16±0.10	0.18±0.10	0.14±0.10	0.15±0.10
	HD ↓	0.40±0.30	0.44±0.30	0.29±0.10	0.30±0.10
	SIF(%) ↓	0.00±0.00	0.01±0.01	0.01±0.03	0.0003±0.001
Pial	ASSD ↓	0.15±0.10	0.17±0.10	0.14±0.01	0.14±0.01
	HD ↓	0.36±0.14	0.41±0.19	0.30±0.07	0.30±0.07
	SIF(%) ↓	0.001±0.001	0.03±0.09	0.03±0.09	0.01±0.02
SD	$\overline{\sigma(\text{SD})}$ ↓	1.16	1.13	1.20	1.14

3.2. Ablation Study

To evaluate the contribution of each component in our framework, we performed an ablation study on both WM and pial surface reconstruction (Table 1). For WM surfaces, we train four models: (i) a *baseline* model (*i.e.*, a CoTAN model using weekly templates as initial surfaces and no augmentation); (ii) *baseline + augmentation*, incorporating intensity and contrast augmentations; (iii) *baseline + SD*, which adds **SulcNet** guidance without augmentation; and (iv) our proposed **SuD-CoTAN**, combining augmentations with **SulcNet** guidance. For pial surface reconstruction, we train four corresponding models.

We assess geometric accuracy (average symmetric surface distances (ASSD) and 90th percentile of Hausdorff distances (HD)), mesh quality (self-intersecting faces (SIF)), and the stability of predicted sulcal depth maps in the template space, quantified as the average standard deviation across mean sulcal depth (SD) maps computed from the predicted WM surfaces ($\overline{\sigma(\text{SD})}$). All results were calculated on combined dHCP and iFIND test sets.

We observe that sulcal-depth guided models (**SuD-CoTAN** the *baseline+SD* model) achieved the lowest sulcal-depth variability, indicating that **SulcNet** guidance improves anatomical consistency without compromising surface integrity. Removing augmentation significantly increases geometric errors due to poor generalisability to iFIND dataset (see iFIND example in Figure 1), highlighting its importance for cross-dataset generalisation. Mesh quality remained high across all settings, with SIF equal to zero for the *baseline*, nearly zero for **SuD-CoTAN**, and only slightly higher in the other models.

3.3. SulcNet: Evaluation of Predicted Sulcal Depth Maps

We next evaluated the **SulcNet** module to quantify its accuracy in sulcal depth prediction and thus suitability for guiding surface reconstruction. Figure 4 presents qualitative (left) and quantitative (right) results. Visually, the two representative examples (28.14 and 33.86 weeks GA) show strong agreement between ground-truth and predicted sulcal depth maps. Moreover, **SulcNet** achieved a low MSE of $0.32 \pm 0.17\text{mm}$ and a high NCC of 0.99 ± 0.01 , indicating strong correspondence between predicted and pGT sulcal depth maps, suggest-

ing that **SulcNet** provides the consistency and accuracy required for reliable, age-aware guidance during cortical surface reconstruction.

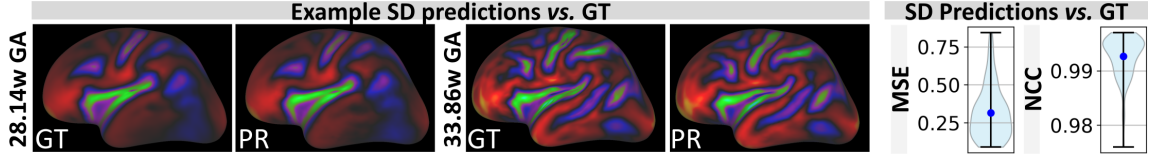


Figure 4: Predictions from **SulcNet** showing (left): two examples comparing pGT *vs.* predictions (PR); and textitright: quantitative evaluation of MSE and NCC between pGT and predicted for dHCP test examples. Blue dots represent the mean.

3.4. Population-level Consistency and Developmental Trajectories

To assess whether **SuD-CoTAN** produces more developmentally consistent population-level representations, we generated age-stratified average and standard deviation sulcal depth maps from the 313 dHCP and iFIND subjects (Table A.1, “Feature Extraction” rows) across gestational ages 22–32 weeks (Figure 5). Figure 5A presents example average and standard deviation sulcal depth maps projected onto inflated surfaces. Visually, mean sulcal depth maps are better defined with lower standard deviation when sulcal depth guidance is used (**SuD-CoTAN**). Figure 5B plots the distribution of variance (left) and the mean sharpness (right) across the surface, for all GAs. These results show that **SuD-CoTAN** yields higher sharpness and lower variability relative to the *baseline + augmentation* model. For similar results on curvature, see Appendix E.

To further quantify developmental coherence, we modeled sulcal depth trajectories across the 313 dHCP and iFIND subjects (Table A.1 “Feature Extraction”) using vertex-wise Gaussian Process (GP) regression, fitted using the GPyTorch variational framework (Gardner et al., 2018), using a combination of linear and Radial Basis Function (RBF) kernels. We compared sulcal depth outputs from the *baseline + augmentation* and **SuD-CoTAN** models, and assessed accuracy using the mean absolute error (MAE) between predicted and observed sulcal depth. Figure 6 showcases representative trajectories at two anatomical locations (gyral crown in orange and sulcal fundus in blue), alongside whole-brain GP means and MAE maps. Relative to *baseline + augmentation*, **SuD-CoTAN** yields lower MAE and trajectories with clearer separation between gyral and sulcal developmental trends, representing a deepening of the sulcal fundus and a rising gyral crown with GA. These patterns indicate that the proposed model better captures anatomically meaningful developmental trajectories at the population level.

3.5. Harmonisation of Surface-based Metrics

To evaluate cross-dataset harmonisation, we analysed global cortical metrics from the predicted surfaces of the same 313 dHCP and iFIND subjects, including mean cortical thickness, mean absolute sulcal depth, mean absolute curvature, and average surface area. Figure 7 shows the average metrics as a function of GA corresponding to the proposed

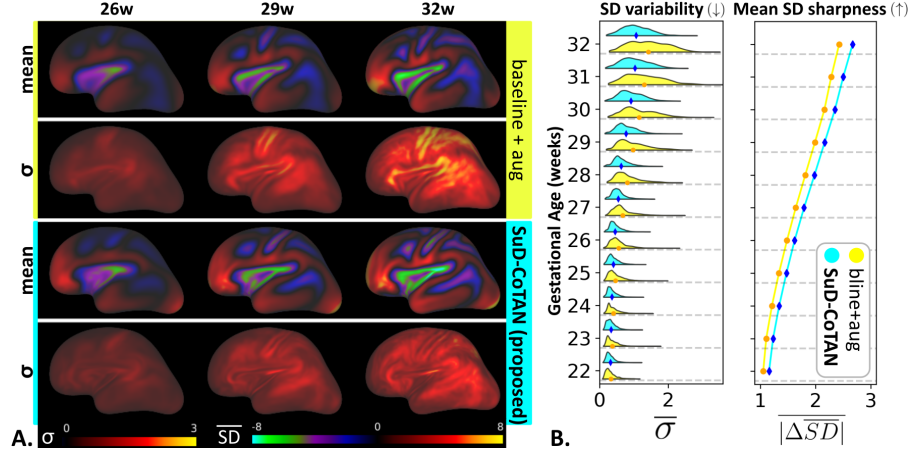


Figure 5: Sulcal depth maps across gestational ages. **A.** Example mean and standard deviation sulcal depth maps for subjects scanned at 26, 29 and 32 weeks for the *baseline + augmentation* model (first two rows) and our proposed **SuD-CoTAN** model (last two rows). **B.** Sulcal depth (SD) variability (lower is better) and the mean absolute Laplacian of the average sulcal depth map (as a proxy for sharpness, where higher is better) for each gestational age, comparing the *baseline + augmentation* (yellow) and our proposed model (cyan).

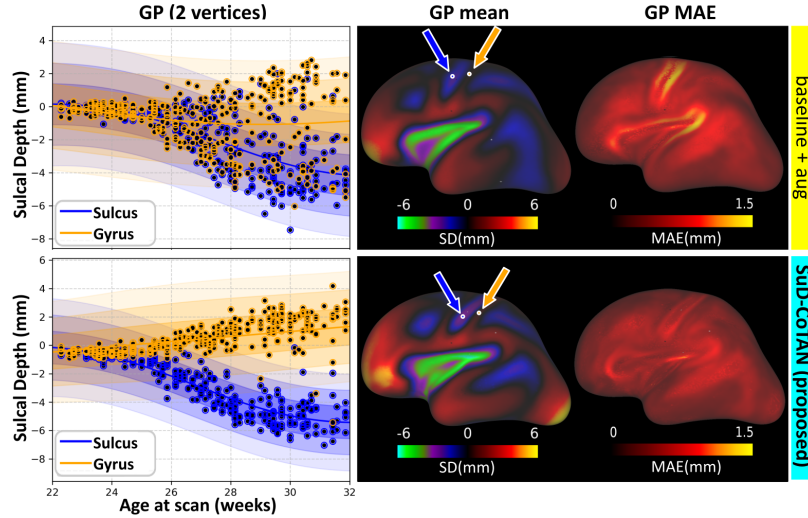


Figure 6: GP-derived sulcal depth trajectories. Representative developmental trajectories at a gyral crown (orange) and a sulcal fundus (blue), alongside GP mean maps and MAE maps for the *baseline + augmentation* and **SuD-CoTAN** models.

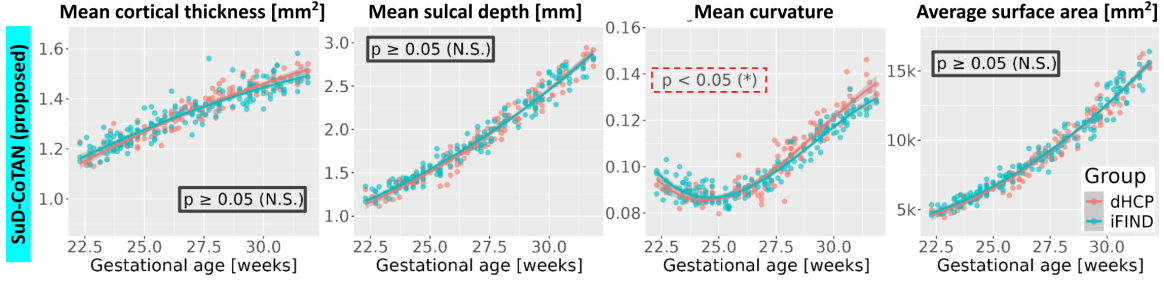


Figure 7: Global average cortical metrics as a function of gestational age. Dashed red boxes indicate metrics showing a statistically significant cohort effect between dHCP (red) and iFIND (blue) predicted surfaces, while black boxes mark metrics where this cohort effect is non-significant.

SuD-CoTAN model (see Appendix F for the other three models). Polynomial regression controlling for GA revealed no significant cohort effects in terms of cortical thickness, sulcal depth, and surface area, while mean curvature still shows a minor cohort difference. As opposed to the *baseline* and *baseline+SD* models, **SuD-CoTAN** showed the strongest improvement in harmonising cortical thickness, an encouraging result given its reliance on accurate white and pial surface reconstructions. Notably, despite using sulcal-depth guidance, the *baseline + SD* model did not harmonise mean sulcal depth across datasets, highlighting the importance of augmentation.

4. Discussion and Conclusions

This study proposed **SuD-CoTAN**, a sulcal-depth-guided framework for fetal cortical surface reconstruction, delivering anatomically informed surfaces with improved geometric accuracy, sharper population-level maps, and more consistent developmental trajectories. Harmonisation analysis further showed that the method reduces cohort differences in cortical metrics, producing more comparable measurements across dHCP and iFIND. GP modelling of sulcal depth trajectories reveals clearer developmental patterns, with deepening sulci and rising gyral crowns over gestation, supporting the biological plausibility of the reconstructed surfaces. These results indicate that anatomically informed guidance provides stable and biologically meaningful constraints that improve cross-dataset generalisation without compromising surface quality. Despite these benefits, a key limitation of this work stems from the substantial natural variability in cortical folding. The human cortex exhibits diverse folding patterns, including branching, splitting, and tertiary folds, that cannot be fully captured by registration to a single population-average template (Guo et al., 2025). Future work will explore strategies that better accommodate individual folding diversity. Overall, our proposed model represents a promising step toward robust, harmonised fetal cortical surface analysis suitable for multi-site studies and future large-scale developmental neuroimaging applications.

References

- Renato Besenczi, Yourong Guo, and Emma C Robinson. High performance groupwise cortical surface registration with multimodal surface matching. In *International Workshop on Biomedical Image Registration*, pages 321–332. Springer, 2024.
- Fabian Bongratz, Anne-Marie Rickmann, Sebastian Pölsterl, and Christian Wachinger. Vox2cortex: Fast explicit reconstruction of cortical surfaces from 3d mri scans with geometric deep neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20773–20783, 2022.
- Jelena Bozek, Antonios Makropoulos, Andreas Schuh, Sean Fitzgibbon, Robert Wright, Matthew F Glasser, Timothy S Coalson, Jonathan O’Muircheartaigh, Jana Hutter, Anthony N Price, et al. Construction of a neonatal cortical surface atlas using multimodal surface matching in the developing human connectome project. *NeuroImage*, 179:11–29, 2018.
- M Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, et al. Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701*, 2022.
- L Cordero-Grande, AN Price, EJ Hughes, R Wright, MA Rutherford, and JV Hajnal. Automating motion compensation in 3t fetal brain imaging: localize, align and reconstruct. In *ISMRM*, volume 2019, page 1000, 2019.
- Rodrigo Santa Cruz, Leo Lebrat, Pierrick Bourgeat, Clinton Fookes, Jurgen Fripp, and Olivier Salvado. Deepcsr: A 3d deep learning approach for cortical surface reconstruction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 806–815, 2021.
- Kara E Garcia, Kaylin Taylor, Meghana Bhaskara, Clemente Velasco-Annis, John Vieth, Joshua Garrett, Jessica Patel, Benjamin Pointer, Sha Cao, Jane W Newburger, et al. Longitudinal magnetic resonance imaging reveals differences in cortical expansion in fetuses with congenital heart defects. *Cerebral Cortex*, 35(8):bhaf220, 2025.
- Jacob Gardner, Geoff Pleiss, Kilian Q Weinberger, David Bindel, and Andrew G Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. *Advances in neural information processing systems*, 31, 2018.
- Ali Gholipour, Caitlin K Rollins, Clemente Velasco-Annis, Abdelhakim Ouaalam, Alireza Akhondi-Asl, Omur Afacan, Cynthia M Ortinau, Sean Clancy, Catherine Limperopoulos, Edward Yang, et al. A normative spatiotemporal mri atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Scientific reports*, 7(1):476, 2017.
- Karthik Gopinath, Douglas N Greve, Sudeshna Das, Steve Arnold, Colin Magdamo, and Juan Eugenio Iglesias. Cortical analysis of heterogeneous clinical brain mri scans for large-scale neuroimaging studies. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 35–45. Springer, 2023.

- Yourong Guo, Mohamed A Suliman, Logan ZJ Williams, Renato Besenczi, Kaili Liang, Simon Dahan, Vanessa Kyriakopoulou, Grainne M McAlonan, Alexander Hammers, Jonathan O’Muircheartaigh, et al. Motifs of brain cortical folding from birth to adulthood: structural asymmetry and folding-functional links. *bioRxiv*, pages 2025–03, 2025.
- Vycheslav Karolis, Logan Williams, Vanessa Kyriakopoulou, Jelena Bozek, Alena Uus, Antonios Makropoulos, Andreas Schuh, Lucilio Cordero Grande, Emer Hughes, Anthony Price, et al. Developing human connectome project spatio-temporal surface atlas of the fetal brain. 2023.
- Maria Kuklisova-Murgasova, Paul Aljabar, Latha Srinivasan, Serena J Counsell, Valentina Doria, Ahmed Serag, Ioannis S Gousias, James P Boardman, Mary A Rutherford, A David Edwards, et al. A dynamic 4D probabilistic atlas of the developing brain. *NeuroImage*, 54(4):2750–2763, 2011.
- 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.
- Vanessa Kyriakopoulou, Deniz Vatansever, Samia Elkommos, Sarah Dawson, Amy McGuinness, Joanna Allsop, Zoltán Molnár, Joseph Hajnal, and Mary Rutherford. Cortical overgrowth in fetuses with isolated ventriculomegaly. *Cerebral Cortex*, 24(8):2141–2150, 2014.
- Qiang Ma, Liu Li, Emma C Robinson, Bernhard Kainz, Daniel Rueckert, and Amir Alansary. CortexODE: Learning cortical surface reconstruction by neural ODEs. *IEEE Transactions on Medical Imaging*, 42(2):430–443, 2022.
- Qiang Ma, Liu Li, Vanessa Kyriakopoulou, Joseph V Hajnal, Emma C Robinson, Bernhard Kainz, and Daniel Rueckert. Conditional temporal attention networks for neonatal cortical surface reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 312–322. Springer, 2023.
- Qiang Ma, Liu Li, Emma C Robinson, Bernhard Kainz, and Daniel Rueckert. Weakly supervised learning of cortical surface reconstruction from segmentations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 766–777. Springer, 2024.
- 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 Hünning, Nazakat Merchant, Tomoki Arichi, Nora Tusor, Joseph V Hajnal, A David Edwards, Serena J Counsell, et al. Regional growth and atlasing of the developing human brain. *Neuroimage*, 125:456–478, 2016.

- Antonios Makropoulos, Emma C Robinson, Andreas Schuh, Robert Wright, Sean Fitzgibbon, Jelena Bozek, Serena J Counsell, Johannes Steinweg, Katy Vecchiato, Jonathan Passerat-Palmbach, et al. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*, 173:88–112, 2018.
- 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. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5115–5124, 2017.
- Anthony N Price, Lucilio Cordero-Grande, Emer Hughes, Suzanne Hiscocks, Elaine Green, Laura McCabe, Jana Hutter, Giulio Ferrazzi, Maria Deprez, Thomas Roberts, et al. The developing human connectome project (dhcp): fetal acquisition protocol. In *Proceedings of the annual meeting of the International Society of Magnetic Resonance in Medicine (ISMRM)*, volume 244. International Society for Magnetic Resonance in Medicine (ISMRM), 2019.
- Emma C Robinson, Saad Jbabdi, Matthew F Glasser, Jesper Andersson, Gregory C Burgess, Michael P Harms, Stephen M Smith, David C Van Essen, and Mark Jenkinson. Msm: a new flexible framework for multimodal surface matching. *Neuroimage*, 100:414–426, 2014.
- Emma C Robinson, Kara Garcia, Matthew F Glasser, Zhengdao Chen, Timothy S Coalson, Antonios Makropoulos, Jelena Bozek, Robert Wright, Andreas Schuh, Matthew Webster, et al. Multimodal surface matching with higher-order smoothness constraints. *Neuroimage*, 167:453–465, 2018.
- Andreas Schuh, Antonios Makropoulos, Robert Wright, Emma C Robinson, Nora Tusor, Johannes Steinweg, Emer Hughes, Lucilio Cordero Grande, Anthony Price, Jana Hutter, et al. A deformable model for the reconstruction of the neonatal cortex. In *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, pages 800–803. IEEE, 2017.
- Julia A Scott, Piotr A Habas, Kio Kim, Vidya Rajagopalan, Kia S Hamzelou, James M Corbett-Detig, A James Barkovich, Orit A Glenn, and Colin Studholme. Growth trajectories of the human fetal brain tissues estimated from 3d reconstructed in utero mri. *International Journal of Developmental Neuroscience*, 29(5):529–536, 2011.
- Ahmed Serag, Vanessa Kyriakopoulou, Mary A Rutherford, A David Edwards, Joseph V Hajnal, Paul Aljabar, Serena J Counsell, James Boardman, and Daniel Rueckert. A multi-channel 4d probabilistic atlas of the developing brain: application to fetuses and neonates. *Annals of the BMVA*, 2012(3):1–14, 2012.
- Lisa Story, Alice Davidson, Prachi Patkee, Bobbi Fleiss, Vanessa Kyriakopoulou, Kathleen Colford, Srividhya Sankaran, Paul Seed, Alice Jones, Jana Hutter, et al. Brain volumetry in fetuses that deliver very preterm: an mri pilot study. *NeuroImage: Clinical*, 30:102650, 2021.

- Mohamed A Suliman, Logan ZJ Williams, Abdulah Fawaz, and Emma C Robinson. A deep-discrete learning framework for spherical surface registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 119–129. Springer, 2022.
- Mohamed A Suliman, Logan ZJ Williams, Abdulah Fawaz, and Emma C Robinson. Un-supervised multimodal surface registration with geometric deep learning. *Medical Image Analysis*, page 103821, 2025.
- Alena Uus, Irina Grigorescu, Maximilian Pietsch, Dafnis Batalle, Daan Christiaens, Emer Hughes, Jana Hutter, Lucilio Cordero Grande, Anthony N Price, Jacques-Donald Tournier, et al. Multi-channel 4d parametrized atlas of macro-and microstructural neonatal brain development. *Frontiers in Neuroscience*, 15:661704, 2021.
- Alena Uus, Sara Neves Silva, Jordina Aviles Verdera, Kelly Payette, Megan Hall, Kathleen Colford, Aysha Luis, Helena Sousa, Zihan Ning, Thomas Roberts, et al. Scanner-based real-time three-dimensional brain+ body slice-to-volume reconstruction for t2-weighted 0.55-t low-field fetal magnetic resonance imaging. *Pediatric Radiology*, 55(3):556–569, 2025.
- Alena U. Uus, Vanessa Kyriakopoulou, Antonios Makropoulos, Abi Fukami-Gartner, Daniel Cromb, Alice Davidson, Lucilio Cordero-Grande, Anthony N. Price, Irina Grigorescu, Logan Z. J. Williams, Emma C. Robinson, David Lloyd, Kuberan Pushparajah, Lisa Story, Jana Hutter, Serena J. Counsell, A. David Edwards, Mary A. Rutherford, Joseph V. Hajnal, and Maria Deprez. Bounti: Brain volumetry and automated parcellation for 3d fetal mri. August 2023. doi: 10.7554/elife.88818.1. URL <http://dx.doi.org/10.7554/eLife.88818.1>.
- Li Wang, Zhengwang Wu, Liangjun Chen, Yue Sun, Weili Lin, and Gang Li. iBEAT V2.0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction. *Nature protocols*, 18(5):1488–1509, 2023.
- Logan ZJ Williams, Sean P Fitzgibbon, Jelena Bozek, Anderson M Winkler, Ralica Dimitrova, Tanya Poppe, Andreas Schuh, Antonios Makropoulos, John Cupitt, Jonathan O’Muircheartaigh, et al. Structural and functional asymmetry of the neonatal cerebral cortex. *Nature human behaviour*, 7(6):942–955, 2023.
- Robert Wright, Bishesh Khanal, Alberto Gomez, Emily Skelton, Jacqueline Matthew, Jo V Hajnal, Daniel Rueckert, and Julia A Schnabel. Lstm spatial co-transformer networks for registration of 3d fetal us and mr brain images. In *International Workshop on Preterm, Perinatal and Paediatric Image Analysis*, pages 149–159. Springer, 2018.
- 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. In *International Conference on Information Processing in Medical Imaging*, pages 855–866. Springer, 2019.

Lilla Zöllei, Juan Eugenio Iglesias, Yangming Ou, P Ellen Grant, and Bruce Fischl. Infant freesurfer: An automated segmentation and surface extraction pipeline for t1-weighted neuroimaging data of infants 0–2 years. *Neuroimage*, 218:116946, 2020.

Appendix A. Dataset Split

Table A.1: Number of scans used for training and evaluating our models, as well as cortical feature extraction.

Task	Dataset	Data Split	Subjects	GA (weeks)
Training	dHCP	Train	200	29.1 (± 3.88)
	dHCP	Valid	10	28.9 (± 6.03)
Evaluation	dHCP	Test	20	28.5 (± 3.79)
	iFIND	Test	20	27.9 (± 3.14)
Feature Extraction	dHCP	Train/Valid/Test	140	27.3 (± 2.49)
	iFIND	Test	173	26.7 (± 2.74)

Appendix B. Pial-outside-WM Loss

Letting v_i^{wm} and v_i^{pial} denote corresponding WM and pial vertices, and n_i^{wm} the unit WM normal, we define:

$$\mathcal{L}_{pow} = \frac{1}{N} \sum_{i=0}^{N-1} \text{ReLU} \left(-n_i^{wm} \cdot \widehat{(v_i^{pial} - v_i^{wm})} \right), \quad (\text{B.1})$$

where $\widehat{(\cdot)}$ denotes vector normalisation. The ReLU ensures that only inward-pointing displacements (*i.e.*, violations) contribute to the loss.

Appendix C. Training loss weights used for SuD-CoTAN

Table C.1: Weights for training SuD-CoTAN.

Surface Type	Loss Term	Pre-training	Fine-tuning
WM	λ_{lap}	0.5	0.1
	λ_{nc}	5×10^{-4}	1×10^{-4}
	λ_{edge}	5×10^{-4}	1×10^{-4}
	λ_{SD}	10	1
Pial	λ_{lap}	0.5	0.1
	λ_{nc}	5×10^{-4}	1×10^{-4}
	λ_{edge}	5×10^{-4}	1×10^{-4}
	λ_{infl}	2.5	2.5
	λ_{pow}	2.5	2.5

Appendix D. Data augmentations during training

Table D.1: Data augmentations applied during model training.

Model	Augmentation	Parameters
SuD-CoTAN	Bias-field augmentation	degree = 3; coeff. $\sim \mathcal{U}(-0.5, 0.5)$
	Gaussian noise	$\sigma \sim \mathcal{U}(0.0, 0.02)$
	Gaussian smoothing	$\sigma \sim \mathcal{U}(0.5, 1.5)\text{mm}$
	Gamma correction	$\gamma \sim \mathcal{U}(0.75, 1.25)$
	Histogram shifts	3–5 random control points
SulcNet	Translation	$t_x, t_y, t_z \sim \mathcal{U}(-3, 3)\text{mm}$
	Rotation	$r_x, r_y, r_z \sim \mathcal{U}(-15, 15)^\circ$

Appendix E. Average Curvature

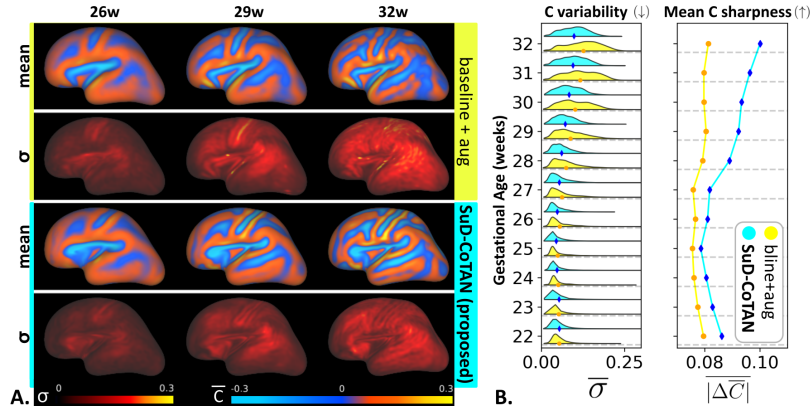


Figure E.1: Curvature maps across gestational ages. **A.** Example average and standard deviation curvature maps for subjects scanned at 26, 29 and 32 weeks for the *baseline + augmentation* model (first two rows) and our proposed **SuD-CoTAN** model (last two rows). **B.** Curvature (C) maps variability (lower is better) and the mean absolute Laplacian of the average curvature map (as a proxy for sharpness, where higher is better) for each gestational age, comparing the *baseline + augmentation* (yellow) and our proposed model (cyan).

Appendix F. Harmonisation of Surface-based Metrics

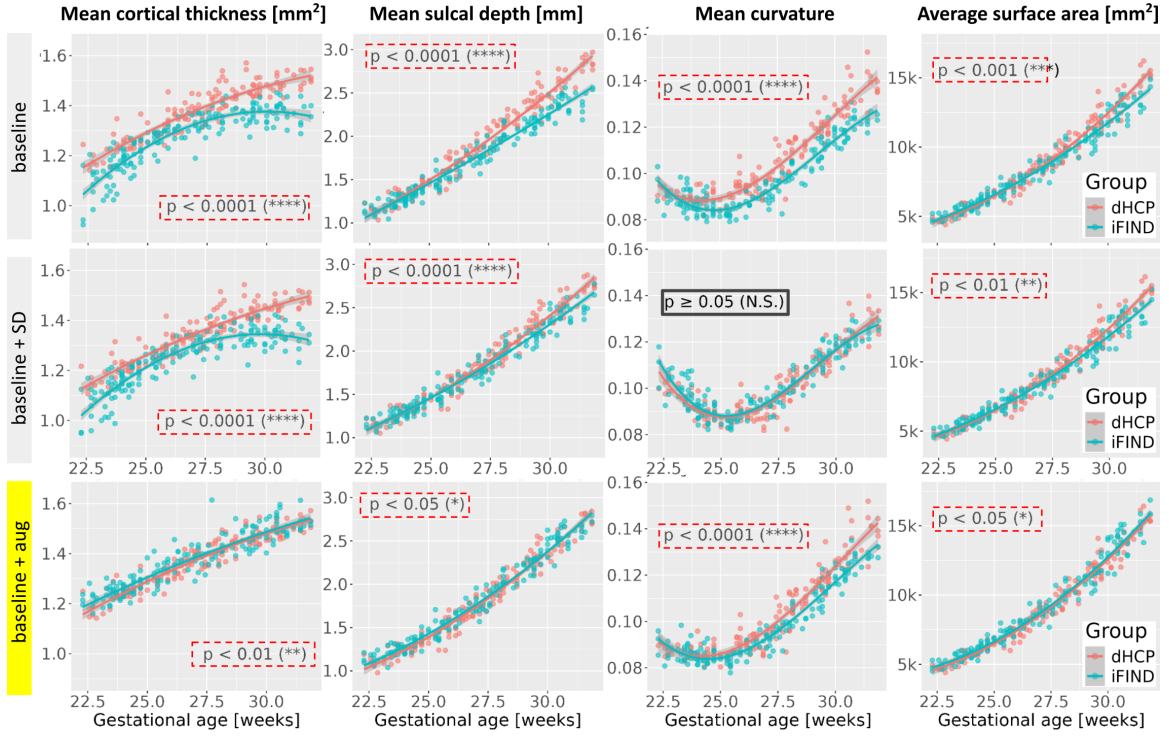


Figure F.1: Global average cortical metrics as a function of gestational age. Dashed red boxes indicate metrics showing a statistically significant cohort effect between dHCP (red) and iFIND (blue) predicted surfaces.