Personalized phenotype encoding and prediction of pathological development from cross-sectional images.

Connor Elkhill*, Ines. A Cruz-Guerrero, PhD*, Jiawei Liu, MS*, Marius George Linguraru, DPhil†, Allyson Alexander, MD§, Brooke French, MD¶, Antonio R. Porras, PhD*‖
*Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO
†Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Hospital, Washington, DC
§Department of Pediatric Neurosurgery, Children's Hospital Colorado, Aurora, CO
¶Department of Pediatric Plastic and Reconstructive Surgery, Children's Hospital Colorado, Aurora, CO
‖Departments of Pediatrics, Surgery and Biomedical Informatics, School of Medicine, Aurora, CO
Email: connor.2.elkhill@cuanschutz.edu

***Abstract*—The prediction of anatomical development plays a crucial role in pediatric treatment selection and planning. We present a novel deep learning architecture to make personalized predictions of pediatric normative and pathologic head development using only cross-sectional data. We designed growth predictor that learns the anatomical effects of age and sex in the presence of pathology and a novel phenotype encoder that utilizes domain adversarial training to create age- and sex-independent representations of patient phenotypes. We combined these modules to instantiate patient phenotypes to specific ages for personalized anatomical predictions conditioned to cranial pathology. We trained our models using standardized head segmentations generated from cross-sectional CT images and 3D photograms and evaluated model performance using an independent longitudinal dataset of normative subjects and children with craniosynostosis. The model achieved a head surface growth prediction error of 4.93 ± 2.29 mm and a volumetric error 0.17 ± 0.11 L in patients with cranial pathology, and 4.61 ± 3.28 mm and 0.27 ± 0.19 L for normative subjects, demonstrating state-of-the-art accuracy. Our method is the first to create age- and sex-agnostic phenotypical representations and enable personalized predictions of pathological development from only cross-sectional data.**

***Index Terms*—predictive growth, craniosynostosis, generative adversarial network, domain adversarial training, pediatric pathological development**

I. Introduction

Children undergo a rapid cranial and brain growth during the first few years of life that is crucial for their cognitive development [1]. Craniosynostosis is a condition where one or more of the cranial sutures fuse prematurely and constrain cranial and brain development, often requiring surgical intervention [2]. However, current clinical standards to evaluate developmental anomalies rely on normative growth charts of simple metrics such as head circumference or intracranial volume [3], [4] that cannot characterize or predict development of patients with pathology [5], [6].

Given the scarcity of longitudinal datasets in the pediatric population, several methods to predict cranial growth using only cross-sectional training datasets have recently been proposed. Liu et al [7] created a normative reference of cranial growth based on age and sex using principal component analysis and temporal regression. However, this method only identified average growth trajectories in the normative pediatric population without any personalization. Porras et al [8] presented a personalized predictive model of cranial growth, but the optimization method used was not computationally feasible for large datasets. Recently, Liu et al [9] presented a data-driven model of cranial suture growth trained only with cross-sectional data, and demonstrated state-of-the-art accuracy to predict pathologic development in patients with craniosynostosis. However, this method required the observation of the cranial sutures from computed tomography (CT) images, an image modality that requires harmful radiation exposure [10]. In a different domain, Xia et al [11] utilized generative adversarial networks to learn subject-specific trajectories from a cross-sectional dataset of magnetic resonance (MR) images to predict changes in the brain morphology of patients with Alzheimer's disease. However, their paired training scheme prevented significant anatomical temporal changes, which provided good results in adults but is not realistic in children [3].

As an alternative to CT imaging, 3D photogrammetry has become a popular radiation-free and low-cost clinical alternative to evaluate pediatric cranial malformations [12]. However, it can only image the head surface and hence, prior personalized methods [9] relying on the identification of the cranial sutures are not feasible. We present a novel deep learning architecture designed to make personalized predictions of normative and pathologic development trained using head surface information from both cross-sectional CT images and 3D photograms, enabling radiation-free prediction of pediatric development in a clinical setting. We propose a temporal growth predictor model that learns age- and sex-specific anatomical distributions in presence or absence of pathology in the pediatric population. We also present a novel

phenotype encoder that utilizes domain adversarial training to generate age- and sex-agnostic latent representations of patient phenotypes. These modules work collaboratively to generate personalized predictions of anatomical development by instantiating patient-specific phenotype representations to specific ages conditioned by the presence of pathology. The architecture was trained using only cross-sectional data from normative children and patients with craniosynostosis under the age of 10 years and was evaluated using an independent longitudinal dataset.

II. Materials and Methods

A. Data

After IRB approval at University of Colorado Anschutz Medical Campus (IRB #20-1563), we collected two independent retrospective datasets: (1) a cross-sectional dataset used for model training; and (2) a longitudinal dataset used for performance evaluation. **Dataset 1** includes 2,672 cross-sectional CT images (N=2,262) and 3D photograms (N=410) of two patient populations: 2,020 normative subjects without cranial pathology (1081 male, 939 female, age 3.14 ± 3.05 years, range 0-10 years) and 652 patients with craniosynostosis (384 male, 268 female, age 0.64 ± 1.04 years, range 0-8.8 years). Dataset 2 includes two groups of longitudinal images: 61 CT image pairs from 51 normative subjects (28 male, 23 female) with average age at first image 2.24 ± 2.22 years and age at second image 3.55 ± 2.71 years (five subjects had two image pairs); and 75 pairs of 3D photograms from 75 patients with craniosynostosis (58 male, 17 female), with age of 0.61 ± 1.07 years at the first image and 0.85 ± 1.18 years at the second image.

B. Standard representation of the head surface

We used publicly available methods to segment the head surface from CT images and 3D photograms and create standardized 2D anatomical representations [8]. In summary, Hounsfeld units thresholding and the marching cubes algorithm [14] were used to segment the head surface from a CT image. Four cranial landmarks were automatically identified at the glabella, two temporal processes of the dorsum sellae and the opisthion in CT images. In 3D photogrammetry, a similar publicly available method was utilized to identify a series of anatomical landmarks on the head and face [13]. Using their respective landmarks, each image was aligned to a standardized template image annotated with both sets of landmarks. The calvaria was segmented from the rest of the head at the cranial base using the plane defined by the nasion, left and right tragion in the standardized template (Fig 1b). Finally, the head surface segmented from either image modality was sampled in spherical coordinates to create a standard 2D representation [8] Fig 1c.

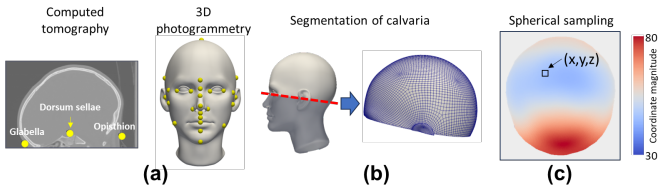

Fig. 1. Standard representation of the head surface. (a) CT image or 3D photogram annotated with anatomical landmarks. Detailed annotations of landmarks for 3D photograms can be found in [13]. (b) Segmented head surface using the naso-tragal plane. (c) Two-dimensional standard representation of the head surface using spherical sampling, where each pixel contains the corresponding location in the head surface (X, Y, Z).

C. Model Architecture

We propose a neural network architecture composed of two main modules as presented in 2 and described below.

Personalized phenotype encoder (PPE). The goal of this module is to create a vectorized representation of the head phenotype of a patient independently from age and sex. To accomplish this, we designed a phenotype encoder that uses unsupervised representation learning and the convolutional architecture presented in [15] to provide a latent patient-specific phenotype representation $\mathbf{l}$ based on a real head shape observation $\mathbf{X}$. Unlike previous work, we incorporate an additional phenotype discriminator (PD, purple in 2) and use a domain adversarial training scheme [16] (see details in Eq. 1) to promote a phenotype representation $\mathbf{l}$ that is independent from age or sex. Hence, the encoder will learn to generate latent phenotype representations $\mathbf{l}$ that cannot be used by the PD to identify the age and sex of the patient.

Growth predictor (GP). This module uses an adversarial training scheme with the goal of generating a realistic head surface anatomy from a latent representation of a head phenotype $\mathbf{l}$ (generated by the PPE) together with information about age, sex and pathology by leveraging the learned anatomical distributions of specific patient groups in the training dataset. The GP and PPE work collaboratively to create personalized latent phenotype representations that are age- and sex-agnostic and that, when combined with specific age, sex, and pathology information, can represent the head anatomy of a specific patient. The predictor follows the convolutional architecture proposed by [15] but it was modified to include the conditions

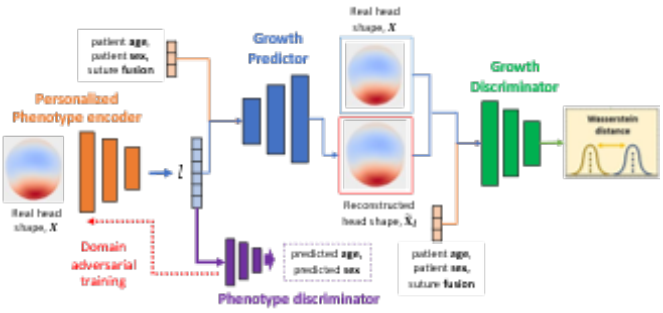

Fig. 2. Proposed neural network architecture as described in Section II-C. The personalized phenotype encoder generates a patient-specific latent representation, l, from head shape, independent of age or sex. A growth predictor, trained on population data, predicts head anatomies based on age, sex, and suture fusion status. The trained model can then generate personalized predictions at any age.

of patient age (encoded as a continuous variable), sex (binary encoded) and suture fusion status (one-hot encoded). Finally, the growth discriminator (GD) (Fig 2, in green) enables adversarial training of both the GP and PPE. The discriminator also utilizes the convolutional architecture proposed in [15] to compute the Wasserstein distance [17] and distinguish between real head shapes and head shapes reconstructed from **l**.

D. Optimization

We defined the prediction of age and sex from the PD as $S', A' = PD(l; \theta_{PD})$ respectively, where θ_{PD} are the learned parameters of the PD and **l** is the latent representation generated from the PPE using real input image **X**. We computed the cross entropy between S and S' and KL divergence between the continuous distributions of A and A', where A and S are the true patient age and sex. The PPE and PD modules are trained using the following adversarial loss function:

$$L_{PPE} = \max_{PPE} \min_{PD} S * log(S') + (1 - S) * log(1 - S') + A * \log \frac{A}{A'} \quad (1)$$

To train the GP module, we used an adversarial scheme with the following Wasserstein loss function [17].

$$L_{GP} = \mathbb{E}_{\tilde{\mathbf{X}}_l \sim P_l | \mathbf{X} \sim P_x} GD(\tilde{\mathbf{X}}_l | A, S, C) - GD(\mathbf{X} | A, S, C) + \lambda_{grad} \mathbb{E}_{\hat{X} \sim P_{\hat{X}}} [\|\nabla_{\hat{X}} GD(\hat{X}, A, S, C, \Theta_{GD})\|_2 - 1]^2, \quad (2)$$

where P_l and P_x are the probability distributions of the reconstructed images from the GP and P_X is the probability distribution of the real input images. The second term in Eq. 2 is the gradient penalty aimed to optimize discriminator performance and enable calculation of Wasserstein distance [17]. For each set of input images **x** and $\tilde{\mathbf{x}}$ to GD, we uniformly sampled from the distribution $P_{\hat{X}}$ random points $\hat{X}$ between the real input image distribution P_X and reconstructed image distribution $P_{\hat{X}_l}$ and computed the L2-norm of the gradient with respect to the model parameters Θ_{GD}. This penalty is weighted in the loss using the hyperparameter λ_{grad}. To train our model, the two loss functions presented in Eq. 1 and Eq. 2 were combined with an additional term quantifying the reconstruction error between the original image **X** and the reconstructed image $\tilde{X}_l$ to promote accurate image predictions. The loss function used during training was:

$$L = L_{GP} + \lambda_{PPE} L_{PPE} + \lambda_{rec} \left\| \tilde{\mathbf{X}}_l - \mathbf{X} \right\|_2 \quad (3)$$

where λ_{PPE} and λ_{rec} are hyperparameters balancing the contribution of the domain-adversarial and reconstruction losses, respectively.

III. Experiments and Results

A. Image preprocessing and training details

We randomly divided Dataset 1 into training (90%, N=1,712) and validation (10%, N=190) sets. We processed each image as described in Section 2.2 and represented them using an image resolution of 64×64 pixels. To facilitate training, we normalized all image pixel components to the range [-1,1] and age to the range [0,1]. Sex was encoded as 1 for male and 0 for female and suture fusion status was one-hot encoded for each of the following sutures: sagittal, metopic, right coronal, and left coronal. Patients with lambdoid suture fusion were excluded due to insufficient data from this rare form of craniosynostosis. Prior to model training and to facilitate convergence, the GP was first warmed up [18] for 500 epochs. Specifically, we initialized the GP and GD parameters to generate realistic anatomies using randomly sampled noise from a standard normal distribution as synthetic phenotype representations l using only the loss function from Eq. 2. Once initialized, we trained the entire network using Eq. 3 and the Adam optimizer for no more than 5,000 epochs, stopping upton convergence of the validation loss or no improvements after 100 consecutive epochs. All experiments were performed on an Intel i9-10900X CPU with 32 GB RAM and an NVIDIA Quadro RTX 4000 GPU using Python 3.11 and Pytorch 2.0.1. Additional details of all hyperparameters utilized by our model, along with the trained network can be found at: [GITHUB LINK AVAILABLE UPON ACCEPTANCE]

B. Performance Evaluation

We evaluated the predictive accuracy of our model using Dataset 2. Each longitudinal image pair was organized in a bi-directional fashion resulting in 272 total pairs (122 images of normative subjects, 150 images of patients with craniosynostosis), which allowed evaluating predictions from the younger image to the older image and vice-versa. We computed the surface prediction error as the point-wise Euclidean distance in millimeters between each point in the predicted and true images and we evaluated difference in volume between the predicted and true head shape. Table I shows a comparison between the performance of our proposed network and other methods including our own ablation studies: our proposed network without the use of domain-adversarial loss, our proposed network without the condition of pathology to make predictions, a publicly available PCA-based regression model of normative head development in the pediatric population [7], a conditional DCGAN [15], and the model of brain aging progression available from [11]. All models in Table I were trained on Dataset 1 and evaluated using Dataset 2.

TABLE I
MEAN SURFACE PREDICTION AND VOLUMETRIC ERRORS FOR PATIENTS IN DATASET 2. P-VALUES ESTIMATED USING A PAIRED WILCOXON SIGNED-RANK TEST BETWEEN THE PROPOSED NETWORK AND OTHER METHODS.

Model	Surface prediction error (mm)	Volumetric Error (L)
Proposed model	4.84 ± 2.63	0.19 ± 0.15
without adversarial loss	5.22 ± 2.72 ($p < 0.005$)	0.20 ± 0.17 ($p = 0.951$)
without condition	6.37 ± 2.56 ($p < 0.005$)	0.21 ± 0.16 ($p < 0.005$)
PCA Model	4.75 ± 4.03 ($p = 0.192$)	0.29 ± 0.24 ($p < 0.005$)
DCGAN	5.38 ± 2.71 ($p < 0.005$)	0.19 ± 0.15 ($p = 0.332$)
BrainAging	6.80 ± 4.42 ($p < 0.005$)	0.27 ± 0.21 ($p < 0.005$)

IV. Discussion

We present a novel deep learning network trained using only cross-sectional data to make personalized predictions of

anatomical development for patients with and without pathology. We demonstrated through evaluation presented in Table I that our proposed network can infer anatomical development with state-of-the-art accuracy.

We found that the PCA-based regression model of normative development provides the lowest average surface prediction error as compared to other methodologies, which we hypothesize is due to the substantially higher image resolution employed by this model (500×500 image resolution) [7]. Despite the lower image resolution of our proposed method (64×64 image resolution), the differences in surface prediction error between our proposed method and the PCA-based regression model are not significant.

We also improve performance using our proposed domain-adversarial loss function, suggesting it enables more accurate instantiation of patient phenotypes to various stages of development. Performance is also reduced when the condition of pathological status is removed, since pathology is an important mediator of pediatric development [2]. Our model also achieved significantly improved surface prediction errors compared to a Conditional-DCGAN [15], which only utilizes randomly sampled noise that is not personalized and Brain Ageing GAN [11], because it assumes longitudinal anatomical changes are not significant, which is not realistic in pediatric development.

Our proposed method provides similar volumetric errors to Conditional-DCGAN which we hypothesize is due to the shared training procedure and structure of the growth predictor found in both models. Importantly, our network produces significantly lower volumetric errors than the PCA-based regression normative model, due to its inability to make prediction of pathological development present in patients with craniosynostosis that are included within Dataset 2. Our model also produces lower volume errors compared to the Brain Ageing GAN, as it assumes limited anatomical changes between images. One limitation of our study is the limited sizes of our longitudinal evaluation datasets, since these datasets are rare in children.

Unlike previous works, the presented model does not make strigent statistical assumptions and can consider pathologic development (i.e., [7], [8]), and can create age- and sex-independent latent representations of patient phenotypes using domain-adversarial training [16] to predict personalized development. Our proposed network can be used to study pediatric phenotypes associated with pathology in diverse datasets of children with varying age and sex.

V. Conclusion

We presented a novel deep learning network using a growth predictor that learns age- and sex-specific anatomical distributions of pediatric head shape and a novel phenotype encoding module to generate age- and sex-independent latent phenotype representations to predict personalized anatomical development in children. The presented model showed state-of-the-art predictive accuracy evaluated on an independent longitudinal dataset. Our age- and sex-independent phenotype representations could be leveraged to study associations between patient phenotypes and pathology in diverse pediatric datasets.

References

[1] M. Cao, H. Huang, and Y. He, "Developmental connectomics from infancy through early childhood," *Trends in Neurosciences*, vol. 40, pp. 494–506, 8 2017.

[2] I. M. Mathijssen, "Updated guideline on treatment and management of craniosynostosis," *Journal of Craniofacial Surgery*, vol. 32, pp. 371–450, 1 2021.

[3] P. Meyer-Marcotty, H. Bohm, C. Linz, J. Kochel, A. Stellzig-Eisenhauer, and T. Schweitzer, "Three-dimensional analysis of cranial growth from 6 to 12 months of age," *The European Journal of Orthodontics*, vol. 36, pp. 489–496, 10 2014.

[4] J. D. Rollins, J. S. Collins, and K. R. Holden, "United states head circumference growth reference charts: Birth to 21 years," *The Journal of Pediatrics*, vol. 156, pp. 907–913.e2, 6 2010.

[5] R. Breakey, P. G. Knoops, A. Borghi, N. Rodriguez-Florez, J. O'Hara, G. James, D. J. Dunaway, S. Schievano, and N. O. Jeelani, "Intracranial volume and head circumference in children with unoperated syndromic craniosynostosis," *Plastic and reconstructive surgery (1963)*, vol. 142, no. 5, pp. 708e–717e, 2018. publisher-place: United States publisher: United States: American Society of Plastic Surgeons.

[6] P. A. Thakkar, K. Yagnik, N. T. Parmar, R. R. Das, and U. P. Thakkar, "Observer variability in head circumference measurement using routine versus non-stretchable tapes in children," *Journal of Nepal Paediatric Society*, vol. 37, pp. 238–243, 6 2018.

[7] J. Liu, C. Elkhill, S. LeBeau, B. French, N. Lepore, M. G. Linguraru, and A. R. Porras, "Data-driven normative reference of pediatric cranial bone development," *Plastic and Reconstructive Surgery - Global Open*, vol. 10, p. e4457, 8 2022.

[8] A. R. Porras, R. Keating, J. Lee, and M. G. Linguraru, "Predictive statistical model of early cranial development," *IEEE Transactions on Biomedical Engineering*, vol. 69, pp. 537–546, 2 2022.

[9] J. Liu, J. H. Froelicher, B. French, M. G. Linguraru, and A. R. Porras, "Data-driven cranial suture growth model enables predicting phenotypes of craniosynostosis," *Scientific Reports*, vol. 13, p. 20557, 11 2023.

[10] T. Schweitzer, H. Böhm, P. Meyer-Marcotty, H. Collmann, R.-I. Ernestus, and J. Krauß, "Avoiding ct scans in children with single-suture craniosynostosis," *Child's Nervous System*, vol. 28, pp. 1077–1082, 7 2012.

[11] T. Xia, A. Chartsias, C. Wang, and S. A. Tsaftaris, "Learning to synthesise the ageing brain without longitudinal data," *Medical Image Analysis*, vol. 73, p. 102169, 10 2021.

[12] M. S. Kurniawan, P. A. Tio, T. Abdel Alim, G. Roshchupkin, C. M. Dirven, M. M. Pleumeekers, I. M. Mathijssen, and M.-L. C. van Veelen, "3D Analysis of the Cranial and Facial Shape in Craniosynostosis Patients: A Systematic Review," *Journal of Craniofacial Surgery*, vol. 35, no. 3, 2024.

[13] C. Elkhill, J. Liu, M. G. Linguraru, S. LeBeau, D. Khechoyan, B. French, and A. R. Porras, "Geometric learning and statistical modeling for surgical outcomes evaluation in craniosynostosis using 3d photogrammetry," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107689, 10 2023.

[14] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *ACM SIGGRAPH Computer Graphics*, vol. 21, pp. 163–169, 8 1987.

[15] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 1 2016. arXiv:1511.06434 [cs].

[16] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *arXiv.org*, 2016. publisher-place: Ithaca publisher: Ithaca: Cornell University Library, arXiv.org.

[17] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," vol. 30, Curran Associates, Inc., 2017.

[18] T. Grigoryev, A. Voynov, and A. Babenko, "When, why, and which pretrained gans are useful?," 3 2022. arXiv:2202.08937 [cs].