Supplementary Material

Paper ID 1422

1 Dataset Details

Fetal Ultrasound The US volumes used in this study were from the INTERGROWTH-21st Fetal Growth Longitudinal Study [3]. That study aimed to describe fetal growth and neurodevelopment in a geographically diverse, healthy population of women, enrolled before 14 weeks' gestation. The volumes were acquired with a Philips US device (Philips HD-9, Philips Ultrasound, USA) using a curvilinear abdominal transducer. All gestational ages were determined based on the last menstrual period and confirmed with US crown-rump measurements that needed to agree within 7 days.

Adult MRI The adult MR images used in this study were from the IXI dataset, which contains MR images from healthy subjects. The images were acquired across three imaging sites in London, UK, on different imaging systems. We preprocessed all image volumes using the FSL Anat pipeline, the key stages of which are skull stripping, bias field correction, and linear registration to the 1mm MNI template.

2 Baseline Implementations

For all baseline implementations, the learning rate was tuned individually from the following options: 1e-3, 5e-4, 1e-4, 5e-5, 1e-5.

SFCN The SFCN architecture [4] is a classification-based method that is trained for regression with a KL-divergence loss between a Gaussian distribution around the true age, and the predicted class probabilities. The number of days/years per bin was selected using cross-validation based on the validation performance, and selected from (1, 2, and 7 days per week) for US, and from (1, and 2 years per bin) for MRI. For both datasets, the smallest number of days/years per bin resulted in optimal performance and was therefore reported in the main results section. This corresponded to the original SFCN implementation where 1 year per bin was used. The standard deviation of the Gaussian distribution was set to 1 day/year, following [4].

SFCN-reg This model uses the same backbone architecture as proposed in [4] but with the last classification layer replaced by a layer with a single regression output. This model was then trained with a mean squared error loss.

CORAL The model referred to as CORAL in the main text consists of a classification ResNet-18 with a CORAL ordinal regression loss, which has been applied for fetal brain age prediction [2]. In the original work, a bin size of 1 week was used, but we selected the optimal bin size from the same options as for the SFCN architecture. Also for this architecture, the optimal bin size was 1 day/year per bin, and this value has therefore been reported in the main paper.

INSightR-Net We introduced this model in earlier work [1] and implemented it according to the publicly available code at [1]. For an equal comparison, the same number of prototypes was used as in the proposed *ExPeRT*, 100, with 512 channels in the add-on layers. The weights of the loss components were kept equal to the original implementation ($\alpha_{MSE} = 1$, $\alpha_{Clst} = 1$, $\alpha_{PSD} = 10$), and the Δ_l in the $\mathcal{L}_{Clst}$ to 2 days/years (due to the larger label range for the age prediction, the original paper used 0.5).

[1] `https://github.com/lindehesse/INSightR-Net`

3 Additional Results INSightR-Net Baseline

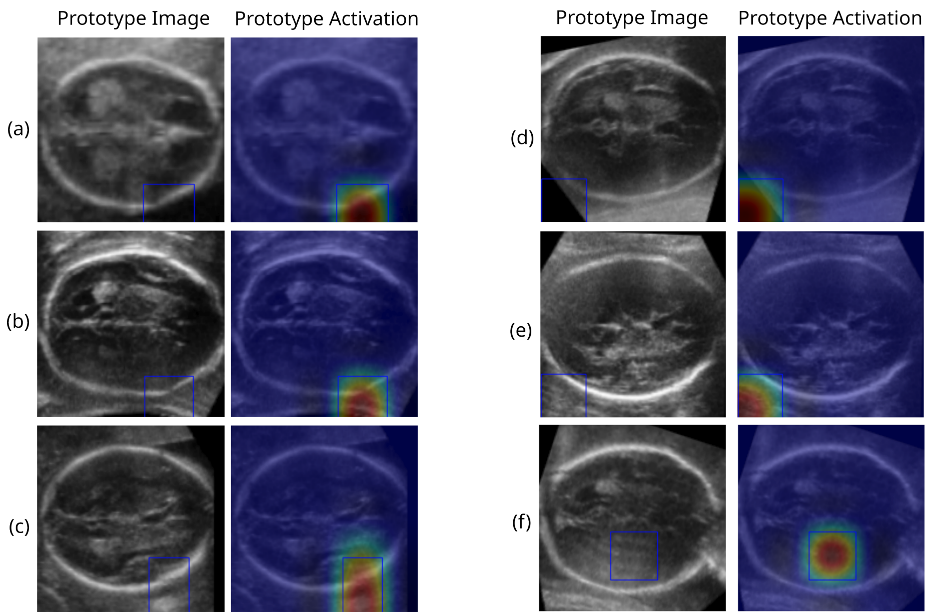

Figure 1: Prototypes recovered from baseline INSightR-Net [1] The shown prototypes were randomly picked from a single trained model. It can be seen that the prototype patches (indicated by the blue boxes) do not recover specific age-discriminative regions but rather background and skull regions.

References

[1] Linde S. Hesse and Ana I.L. Namburete. INSightR-Net: Interpretable Neural Network for Regression Using Similarity-Based Comparisons to Prototypical Examples. In *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pages 502–511. Springer, 2022.

[2] Lok Hin Lee, Elizabeth Bradburn, Rachel Craik, Mohammad Yaqub, Shane A Norris, Leila Cheikh Ismail, Eric O Ohuma, Fernando C Barros, Ann Lambert, Maria Carvalho, et al. Machine learning for accurate estimation of fetal gestational age based on ultrasound images. *NPJ Digital Medicine*, 6(1):36, 2023.

[3] Aris T. Papageorghiou, Eric O. Ohuma, Douglas G. Altman, Tullia Todros, Leila Cheikh Ismail, Ann Lambert, Yasmin A. Jaffer, Enrico Bertino, Michael G. Gravett, Manorama Purwar, J. Alison Noble, Ruyan Pang, Cesar G. Victora, Fernando C. Barros, Maria Carvalho, Laurent J. Salomon, Zulfiqar A. Bhutta, Stephen H. Kennedy, and José Villar. International standards for fetal growth based on serial ultrasound measurements: The Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *Lancet*, 384(9946):869–879, 2014.

[4] Han Peng, Weikang Gong, Christian F Beckmann, Andrea Vedaldi, and Stephen M Smith. Accurate brain age prediction with lightweight deep neural networks. *Medical image analysis*, 68:101871, 2021.