

Optimizing Performance of Transformer-based Models for Fetal Brain MR Image Segmentation

Nicolò Pecco

Pasquale Anthony Della Rosa, PhD

Matteo Canini, PhD

Gianluca Nocera, MD

Paola Scifo, PhD

Paolo Ivo Cavoretto, MD, PhD

Massimo Candiani, MD

Andrea Falini, MD

Antonella Castellano, MD, PhD*

Cristina Baldoli, MD*

From the Neuroradiology Unit and CERMAC (N.P., P.A.D.R., M. Canini, G.N., A.F., A.C., C.B.), Department of Nuclear Medicine (P.S.), and Department of Obstetrics and Gynecology (P.I.C., M. Candiani), IRCCS Ospedale San Raffaele, Via Olgettina 58-60, 20132 Milan, Italy; and Department of ??????, Vita-Salute San Raffaele University, Milan, Italy (N.P., G.N., M. Candiani, A.F., A.C.). Received XXX; revision requested XXX; revision received XXX; accepted XXX. Supported by the Italian Ministry of Health's Ricerca Finalizzata 2016 (grant RF-2016-02364081; Principal Investigator: P.A.D.R.). **Address correspondence to P.A.D.R.** (email: dellarosa.pasquale@hsr.it).

* A.C. and C.B. are co-senior authors.

<https://doi.org/10.1148/ryai.230229>

Purpose: To test transformer-based models' performance when manipulating pretraining weights, dataset size, input size and comparing the best-model with reference standard and state-of-the-art models for a resting-state functional (rs-fMRI) fetal brain extraction task.

Materials and Methods: An internal retrospective dataset (fetuses = 172; images = 519; collected from 2018–2022) was used to investigate influence of dataset size, pretraining approaches and image input size on Swin-UNETR and UNETR models. The internal and an external (fetuses = 131; images = 561) datasets were used to cross-validate and to assess generalization capability of the best model against state-of-the-art models on different scanner types and number of gestational weeks (GW). The Dice similarity coefficient (DSC) and the Balanced average Hausdorff distance (BAHD) were used as segmentation performance metrics. GEE multifactorial models were used to assess significant model and interaction effects of interest.

Results: Swin-UNETR was not affected by pretraining approach and dataset size and performed best with the mean dataset image size, with a mean DSC of 0.92 and BAHD of 0.097. The Swin-UNETR was not affected by scanner type. Generalization results on the internal dataset showed that Swin-UNETR had lower performances compared with reference standard models and comparable performances on the external dataset. Cross-validation on internal

and external test sets demonstrated better and comparable performance of Swin-UNETR versus convolutional neural network architectures during the late-fetal period (GWs > 25) but lower performance during the midfetal period (GWs ≤ 25).

Conclusion: Swin-UNTER showed flexibility in dealing with smaller datasets, regardless of pretraining approaches. For fetal brain extraction of rs-fMRI, Swin-UNTER showed comparable performance with reference standard models during the late-fetal period and lower performance during the early GW period.

©RSNA, 2024

Swin-UNETR with pre-training weights, dataset size and input size optimization, showed comparable performance over state-of-the-art models in resting-state fMRI fetal brain segmentation, with performance differences between mid-and late-fetal periods.

Key Points

When comparing the performance of Swin-UNETR against UNETR on a retrospective fetal resting-state functional MRI dataset (fetuses = 172; images = 519), the Swin-UNETR model showed the best performance metrics using the mean image voxel size, with a mean Dice similarity coefficient (DSC) of 0.92 and balanced average Hausdorff distance BAHD of 0.097.

The Swin-UNETR best-model showed comparable performance to convolutional neural network models for both 1.5T and 3T scanners, with mean DSCs of 0.918 and 0.923, respectively.

Model performance generalization results showed lower performance of Swin-UNETR compared with the reference standard on the internal dataset and comparable performance on the external dataset, and model cross-validation on both test sets demonstrated better model performance on the internal test set and performance differences between the midfetal and late-fetal periods.

Author contributions

Guarantor of integrity of entire study, **N.P., P.A.D.R., M. Candiani, A.C., C.B.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **N.P., P.A.D.R., G.N., M. Candiani, A.C.**; clinical studies, **P.I.C., M. Candiani, A.F., A.C., C.B.**; experimental studies, **N.P., P.A.D.R., P.S., M. Candiani, A.F., C.B.**; statistical analysis, **N.P., P.A.D.R., P.S., M. Candiani**; and manuscript editing, **N.P., P.A.D.R., G.N., P.S., P.I.C., M. Candiani, A.F., A.C., C.B.**

Conflicts of interest are listed at the end of this article.

Automated extraction of the fetal brain from the surrounding maternal compartment faces substantial challenges due to the considerable change of the brain's shape and size throughout gestational weeks (GW).

Convolutional neural networks (CNNs) (1) have proven to be a powerful tool for segmentation tasks across various medical imaging modalities (2,3). Integration of attention mechanisms (4–7) and Vision-Transformers (8) in CNN-inspired architectures (8–10) has led to the development of transformer-based architectures such as UNETR and Swin-UNETR, showing promising results for image segmentation tasks (11–13). The UNETR is a modified version of the U-Net architecture, which replaces the convolutional layers with multihead self-attention modules (14,15). The Swin-UNETR instead introduces a hierarchical structure with alternating stages of shift-based windows and nonlocal self-attention modules (11).

The main goal of this work was two-fold: 1) to evaluate the performance of Swin-UNETR against UNETR transformer-based architectures on a fetal rs-fMRI brain extraction task while assessing the influence of pretraining weight and dataset size between architectures and testing the influence of image size on the transformer-based “best-model” set-up; and 2) to assess the transformer-based “best-model” performance generalization and cross-validation as compared with a reference standard and state-of-the-art existing models with different architectures and degrees of optimization for the same task (ie, nnU-Net (16); U-Net (17); generative adversarial network (GAN) (18)).

Materials and Methods

Study Sample

Data from 179 fetuses retrospectively collected at San Raffaele hospital (OSR) between 2018–2022 were initially considered for inclusion in the study. Fetuses with central nervous system congenital anomalies, brain parenchymal signal alterations and inaccurate masking were excluded by a senior developmental neuroscientist, P.A.D.R. with more than 20 years of experience in advanced neuroimaging data preprocessing and analyses, resulting in final study sample of 172 fetuses with 519 scans. 77 fetuses were scanned at 1.5T (median GW = 32; min GW = 21, max GW = 37) and 95 at 3T (median GW = 28; min GW = 21, max GW = 36). Fetal GW at scan ranged from 21 to 37 (median GW = 29 ± 3.72). The study protocol was conducted in accordance with the Declaration of Helsinki and approved a priori by the OSR Ethics Committee (registration No. EK Nr.2174/2016). All women provided written informed consent prior to MRI examination. Additionally, an independent sample of 131 fetuses with 561 scans (OpenNeuro.org), was included for external testing. Rs-fMRI acquisition details for the internal and external datasets and specifics about creation, using the RS-FetMRI package (19), and validation of manually generated ground-truth for the internal dataset are reported in the Supplemental Materials—Internal dataset and External dataset.

Models' Development

Models were implemented in Pytorch (<https://pytorch.org/> version 1.13.1) and MONAI (<https://moani.io/> version 1.1.0) on a workstation with Ubuntu (version 18.04; <https://releases.ubuntu.com/18.04/>), with an Intel Xeon Processor R, twenty-cores, 16GB of RAM and an 8×-Tesla-V100 card (NVIDIA-DGX-1; 6 GB). Pretraining weights were downloaded from MONAI (<https://github.com/Project-MONAI/tutorials>). Performances of our

“best-model” setup and architecture were compared with three deep learning models: a) a reference standard U-Net (17) model developed and optimized for a fetal brain extraction task with a different architecture (ie, CNN-1) sourced from (<https://github.com/rauschecker-sugrue-labs/fetal-brain-segmentation>); and two state-of-the-art models, b) nnU-net (16) optimized for brain extraction task-general purposes with a similar architecture that of the reference standard (CNN-2), obtained from the repository at (<https://github.com/MIC-DKFZ/nnUNet>), and c) SeGAN (18) optimized for the same task with a different architecture (ie, GAN) and downloaded from (<https://github.com/YuanXue1993/SegAN>) (Fig 1). For Swin-UNETR, reference standard and state-of-the-art models’ parameters and detailed experimental methodology, please refer to Table E1 and Supplemental Materials.

Data Stratification

The internal dataset (fetuses = 172; images = 519) was divided at fetus-level into training and testing set with 90:10 split. A fixed optimal train-test split ratio (20) was chosen to reach convergence toward the optimum values of the weights in the pretrained approaches. The training set was further split into training and validation sets, an 80:20 split, respectively, with equally distributed fetuses across GW and scanners. The test set (fetuses = 17; images = 47) included one fetus per GW and remained consistent across all experiments. For GWs with a limited number of images, assignment priority was given to the test, validation, and training sets, respectively. The training dataset was progressively downsized to two subsets (ie, 66% (fetuses = 103; images = 310), 33% (fetuses = 52; images = 154)) by dropping fetuses sequentially from the 66% to the 33% subset, keeping GWs and scanner representativeness consistent (see Table E2, Fig 1 and Fig E1). Additionally, the external dataset was used for two distinct purposes. Firstly, the external dataset was split with the same splitting procedure implemented for our internal dataset (ie, train: (fetuses = 94; images = 403); validation: (fetuses = 23; images = 103); test: (fetuses = 14; images = 55) to train and test all of the models on a different data distribution, scan acquisition and image processing, thus assessing models’ generalization capabilities. Secondly, a test set to mirror the gestational week distribution of our internal test set (one fetuses per GW) was created for the cross-validation of models trained on our internal dataset, aimed at testing differences or similarities between model performances at specific GW ranges. Refer to Supplemental Materials for details about the external test-set construction.

Statistical Analysis

Dice similarity coefficient (DSC) and Balanced Average Hausdorff Distance (BAHD) (21) metrics were used to measure the segmentations overlap and shape disparities, by using the ‘EvaluateSegmentation’ tool (22). We used a generalized equation estimation (GEE) approach, as implemented in SPSS Statistics (version 23 by IBM), for assessing a) GW representativeness across dataset size (GEE-1); b) performance differences between Swin-UNETR and UNETR, while evaluating the influence of pretraining and scratch approaches and dataset size (GEE-2); c) the influence of image size on the “best-model” performance (GEE-3); d) model performance generalization, in terms of differences between our “best-model,” nnU-Net, U-Net and GAN performance metrics, while evaluating the influence of scanner type differences in the internal

dataset (GEE-4) and the use of an internal or external dataset (GEE-5); and e) model performance cross-validation, in terms of differences between our “best-model,” nnU-Net, U-Net and GAN performance metrics on internal and external test sets, while comparing models within test sets for each GW and between test sets for midfetal (GWs ≤ 25) and late-fetal periods (GWs > 25) (23) (GEE-6). For detailed descriptions of each GEE model, please refer to Table 1.

We report model effects and only post hoc tests, corrected for sequential Bonferroni comparisons ($P < .05$), of significant factorial interactions of interest for each GEE model based on our experimental aims. Table 2 summarizes significant posthoc mean differences. All multifactorial posthoc comparisons, including both significant and not significant effects, are reported instead in Table E9. Descriptive statistics, including mean, standard errors and 95% CIs for each GEE model are included in Tables E3–E8.

Differences in performance between models were considered statistically significant when both DSC and BAHD metrics were statistically significant.

Data and Code Availability

The internal dataset is publicly accessible at the OSR repository (doi: 10.17632/dyg9dpmgvs.1). The source code of the Swin-UNETR model for the fetal rs-fMRI brain extraction task is publicly accessible on GitHub (<https://github.com/NicoloPecco/Swin-Functional-Fetal-Brain-Segmentation>), and the original code for the Swin-UNETR algorithm is available at (<https://github.com/Project-MONAI/research-contributions/tree/main/SwinUNETR>).

Results

Generalized Equation Estimation (GEE) Approach-Results

GEE-1 showed no significant phase-and dataset size-dependent effects on the number of images representative of each GW (ie, no significant effects for the interactions phase x GW, dataset size x GW and phase x dataset size x GW) (see Table 1-GEE-1).

GEE-2 showed a significant approach-dependent x dataset size-dependent interaction effect on DSC and BAHD metrics differences between Swin-UNETR and UNETR (ie, significant effect for the interaction approach x dataset size x Model) (see Table 1-GEE-2 and GEE-2b). Posthoc tests revealed that for all pairwise comparisons, the Swin-UNETR consistently demonstrated significantly higher performance across all approach x dataset combinations compared with the UNETR (see Summary Table 2-GEE-2c with significant P values). No evidence of differences were observed between pretrain and scratch approaches, nor among diverse dataset sizes for the Swin-UNETR (see Table E9-GEE-2c). UNETR had significantly higher performances on the 66% compared with the 33% subsets in the pretrain approach (P (DSC) = 0.004; P (BAHD) = 0.02), and on the 100% dataset compared with the 66% subset in the scratch approach (P (DSC) = 0.009; P (BAHD) = 0.02) (see Summary Table 2-GEE-2c and Table E9-GEE-2c and Table E3-Fig 2[ID]FIG2[ID]).

GEE-3 showed a significant main effect of image size on DSC and BAHD “best-model” performance metrics (see Table 1-GEE-3 and GEE-3b). Posthoc tests demonstrated that using

the mean dataset image size resulted in higher performances when compared with smaller voxels or larger ones (see Summary Table 2-GEE-3c with significant P values) and no evidence of a difference compared with a voxel size of 2 mm³ (see Table E9 GEE-3c). The smallest image voxel size of 0.50 mm³ resulted in lower performances when compared with larger voxel sizes (see Summary Table 2-GEE-3c with significant P values). No evidence of differences in performance were found for image voxel size of 0.75 mm³ when compared with larger voxel sizes (see Table E9-GEE-3c), except for comparison with the mean image voxel size (P (DSC) = 0.001; P (BAHD) = 0.048). Performance with a voxel size of 3 mm³ was higher than performance with voxel sizes of 4 mm³ (P (DSC) < 0.001; P (BAHD) < 0.001) and 5 mm³ (P (DSC) < 0.001; P (BAHD) < 0.001) (see Summary Table 2-GEE-3c and Table E9-GEE-3c and Table E4-Fig 3 and E2).

GEE-4 showed a significant scanner-dependent effect on DSC and BAHD metrics differences for comparisons between “best-model,” nnU-Net, U-Net and GAN performances on the internal dataset (ie, significant effect for the interaction scanner x model) (see Table 1-GEE-4 and GEE-4b). Posthoc tests revealed that the reference standard demonstrated significantly higher performance metrics against the CNN-2 on both 1.5T (P (DSC) < 0.001; P (BAHD) < 0.001) and 3T (P (DSC) < 0.001; P (BAHD) < 0.001). Swin-UNETR showed no evidence of a difference in performance when compared with CNN-2 and the reference standard on both 1.5 and 3 Tesla scanners (see Table E9 GEE-4c). In addition, when comparing models on the 1.5T scanner, the reference standard model showed higher performances than the GAN model (P (DSC) < 0.001; P (BAHD) = 0.02). The GAN model always exhibited significantly lower performances when compared with other models on 3T (see Summary Table 2-GEE-4c and Table E5-Fig E4).

GEE-5 showed a significant dataset type-dependent effect on DSC and BAHD metrics differences for comparisons between “best-model,” nnU-Net, U-Net and GAN performances, while using the internal or external dataset (ie, significant effect for the interaction dataset type x model) (see Table 1-GEE-5 and GEE-5b). Posthoc tests revealed higher performances of the reference standard model on the internal dataset when compared with other models (see Summary Table 3[ID]TBL3[/ID]-GEE-5c with significant P values) but no evidence of differences when compared with CNN-2 and SWIN on the external dataset (see Table E9-GEE-5c). Our best-model showed no evidence of a difference in performance when compared with the CNN-2 model for both datasets (see Table E9-GEE-5c). Conversely, the GAN model demonstrated significantly lower performance when compared with the reference standard model, the other state-of-the-art model and our best-model for both datasets (see Summary Table 3 GEE-5c and Table E9-GEE-5c and Table E6-Fig E3).

GEE-6 showed a significant test-set type-dependent x GW-dependent interaction effect on DSC and BAHD metrics differences for comparisons between “best-model,” nnU-Net, U-Net and GAN performances (ie, significant effect for the interaction test-set type x GW x model) (see Table 1, GEE-6 and 6b). Posthoc tests for comparisons within test-set type at each GW revealed that when comparing performance of the models on the internal test set against the reference-standard across GWs, the Swin-UNETR consistently demonstrated lower performance throughout the midfetal period except for GW at 23 and for GW at 26, 33 and 36 (see Summary

Table 4[ID]TBL4[ID]-GEE-6c with significant P values). Conversely, at GWs 28 and 31, Swin-UNETR demonstrated higher performance (see Summary Table 4-GEE-6c with significant P values). The CNN-2 had lower performance as compared with the reference-standard for GWs at 25, 27 to 30, 32 to 37 (see Summary Table 4 GEE-6c with significant P values). When comparing the Swin-UNETR with the CNN-2, better performances were observed for the CNN-2 at GWs 22 and 25; while for GWs at 28, 30, 31 and 37 the Swin-UNETR showed better performance (see Summary Table 4 GEE-6c with significant P values). The GAN model always showed significantly lower performances compared with the reference standard for each GW, except at GWs 22, 28, 31, 35 and 37, where no evidence of a difference was observed (see Summary Table 4-GEE-6c with significant P values and Table E9-GEE-6c).

For the external test set, posthoc tests revealed that the Swin-UNETR demonstrated significantly lower performance at GWs 26, 28 and 37 when compared with the reference standard (see Summary Table 4-GEE-6c with significant P values). Conversely, at GWs 31 and 36, the Swin-UNETR demonstrated significantly higher performance (see Summary Table 4-GEE-6c with significant P values). No evidence of differences in both DSC and BAHD metrics was found between reference standard and Swin-UNETR performances across other GWs (see Table E9-GEE-6c). The CNN-2 showed consistently higher performance metrics when compared with the reference-standard at GWs at 30, 31, 32 and 36 but lower performance at GW 26 (see Summary Table 4 GEE-6c with significant P values). No evidence of differences was found when comparing CNN-2 performance with the reference standard for both DSC and BAHD scores across other GWs (see Table E9-GEE-6c). When assessing performance differences between the Swin-UNETR and CNN-2, significantly higher performances were observed for the CNN-2 at GWs 23, 26, 28, 30 and 31; whereas at GW 29, Swin-UNETR showed a significantly higher performance (see Summary Table 4-GEE-6c with significant P values). The GAN always showed significantly lower performance at GWs 23 and 28 when compared with the reference standard (see Summary Table 4 GEE-6c with significant P values).

Posthoc tests for comparisons between test set types for midfetal (GWs ≤ 25) and late-fetal periods (GWs > 25) revealed that, in the midfetal period, almost all comparisons (75%-12/16) between internal and external test sets showed significantly higher performances for all models on the internal dataset. For the late-fetal period, half of the comparisons (50%-24/48) showed a significantly higher performance for all models on the internal dataset except for the CNN-2 at GW 37 (see Summary Table 4-GEE-6c, Table E9-GEE-6c, Tables E7 and E8 and Figs 4, 5).

Discussion

The general objective of this study was two-fold. On the one hand, we aimed at pinpointing the best transformer-based deep learning model architecture (ie, Swin-UNTER versus UNETR) for a fetal brain extraction task using rs-fMR images, while evaluating first the impact of different approaches (pretrain versus scratch) and dataset sizes (100%, 66%, 33%) on different transformer-based architectures and then the impact of rs-fMRI image resolution (upsampling versus downsampling) on the best transformer-based architecture model setup. On the other hand, we aimed at evaluating performance of the best transformer-based architecture model

setup against the most representative alternative architectures (ie, CNNs and GAN) tailored for the same task, in terms of both *model generalization for the task on-hand* (ie, validation of best transformer-based architecture model setup as compared with the other architectures model setup across performances two different datasets) and *model cross-validation for specific task characteristics* (ie, cross-validation of model performances at different GWs across two fetal periods (midfetal and late-fetal) on an internal test-set and external test-set).

First, both UNETR and Swin-UNETR were not affected by the pretraining approach, likely due to the pretraining weights employed for CT, which may not offer significant benefits for the fetal brain rs-fMRI extraction task. The Swin-UNETR did not show any performances difference when trained with different subset sizes, demonstrating flexibility in dealing with smaller datasets. On the contrary, the UNETR showed a difference between the 66% and 33% subset suggesting a more limited capability to manage smaller datasets. In addition, the best transformer-based architecture model setup (ie, Swin-UNETR best-model) achieved higher performances when employing the mean dataset image size as compared with both smaller and larger voxel sizes. Notably, the performances decrease paralleled to upsampled voxel sizes may result from redundant information and artifacts.

Second, our model generalization findings highlight the importance of dataset choice in model training and evaluation, as different fetal rs-fMRI datasets may contain unique images characteristics but also a particular selection of fetuses that can impact the variability on models' performances.

Model cross-validation results on the GW-split internal dataset revealed that fetal brain extraction at earlier GWs presented a heightened challenge for the Swin-UNETR best-model. The specific features of fetal brain-maternal abdomen boundaries at early gestational weeks may provide insight into why the Swin-UNETR best-model performance drops with respect to CNN-2, which shows the highest performance for GWs at the lower tail of the GW sample distribution. CNNs demonstrate effective feature hierarchies, parameter efficiency, and local detail recognition, likely translating, in our study, into optimal model performance metrics also at lower GWs. Notably, nnU-net (CNN-2) is optimized for brain extraction task-general purposes as the Swin-UNETR, thus a more "CNN architecture-general" advantage seems to arise, with CNN-2 being less affected by a) the lower number of scans at the lower tail of the GW distribution (ie 21–23,) and b) the substantial differences between fetal brain anatomic landmarks at the midfetal period (≤ 25 GW) with respect to late-fetal period (ie, > 25 GW) (23).

On the contrary, for GWs at the upper tail of the GW sample distribution, the Swin-UNETR best-model showed a significantly higher performance compared with CNN-2 and comparable performance with the reference-standard model, most likely due to the emergence of fetal brain discriminating features from maternal abdominal tissue, resembling a "more general" brain extraction task. Taken together, these results suggest that, for a fetal brain extraction task, the sensitivity of model performance is closely linked to the representativeness of scans across different GWs. This, in turn, is intricately tied to the diverse features exhibited in fetal images, particularly in delineating boundaries between the fetal brain and maternal abdomen, change in fetal brain size, and contrast variability in the rs-fMRI scans especially in the midfetal period.

Therefore, augmenting dataset representativeness could potentially lead to more comparable results between transformer-based and CNN architectures on the whole range of GWs, augmenting their generalization capabilities (24).

Interestingly, performance differences between models on GWs at the lower tail of the GW sample distribution were instead less evident for the external dataset. Furthermore, for GWs in the midfetal period, 75% of the comparisons between performance metrics on internal and external test sets showed a significantly higher performance of all models on the internal dataset; while in the late-fetal period, the percentage dropped by 25%, highlighting a general model sensitiveness to dataset type and GW representativeness within each dataset. This finding is further confirmed by significantly higher model performance differences observed for earlier GWs when using the Internal test-set, while vanishing for the external test-set, where differences instead remained only at later GWs, with the Swin-UNETR best-model and CNN-2 still showing better performances.

This study had some limitations. First, the images are not consistently and homogeneously represented among each GW considered in the gestational week range, especially at lower gestational weeks ($GW < 25$). This lack in lower GW representativeness could affect models' performance and generalizability and introduce potential biases toward higher represented GWs. Second, our sample can be considered relatively small for deep learning application in radiology and unavailability of open-source databases containing fetal fMRI images limits external validation, preventing from fully addressing advantages or disadvantages associated with either architecture and models. Thus, future research should aim to create publicly available fetal resting state large database with balanced and representative images among GWs to enhance the robustness and generalization of deep learning models for fetal resting state brain extraction tasks across the entire gestational period.

In conclusion, future exploration and development involving transformer-based models, using ad-hoc hyperparameter fine-tuning, contrast enhancing approaches and GW data-specific representativeness augmentation procedures, for this specific fetal brain extraction task holds great potential for translational research on specific prenatal clinical purposes with smaller datasets. In addition, this study adds compelling evidence to the generalization and cross-validation of deep learning architectures for a fetal brain extraction task on rs-fMRI, as the study uses all of the most representative architectures and compares model performances at different fetal stages of development (see (25) for a review) while using datasets with different acquisition characteristics and GW representativeness in sample distributions.

Acknowledgments: The authors would like to thank the reviewers for their exceptionally detailed, thoughtful, and constructive feedback and all the participants for their participation and their motivation. We also want to thank the Service Center for Research (CSTOR) at IRCCS San Raffaele Research Institute for their support on IT setup and implementation.

Disclosures of conflicts of interest: **N.P.** No relevant relationships. **P.A.D.R.** Italian Ministry of Health's "Ricerca Finalizzata 2016" (grant number RF-2016-02364081; Principal Investigator: Pasquale Anthony Della Rosa). **M. Canini** RF-2016-02364081; Principal Investigator: Pasquale Anthony Della Rosa. **G.N.** No relevant relationships. **P.S.** e Italian Ministry of Health's "Ricerca Finalizzata 2016" (grant number RF-2016-02364081; Principal Investigator: Pasquale Anthony Della Rosa). **P.I.C.** No relevant relationships. **M. Candiani** No relevant

relationships. **A.F.** No relevant relationships. **A.C.** No relevant relationships. **C.B.** Italian Ministry of Health's "Ricerca Finalizzata 2016" (grant number RF-2016-02364081; Principal Investigator: Pasquale Anthony Della Rosa).

References

1. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022; 11976–11986.
https://openaccess.thecvf.com/content/CVPR2022/html/Liu_A_ConvNet_for_the_2020s_CVPR_2022_paper.html. Accessed March 28, 2023.
2. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. J Digit Imaging 2017;30(4):449–459.
3. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys 2019;29(2):102–127.
4. Vaswani A, Shazeer N, Parmar N, et al. Attention is All you Need. In: Advances in Neural Information Processing Systems 30 (NIPS 2017). Curran Associates.
https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html. Published 2017. Accessed March 28, 2023.
5. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv 1810.04805 [preprint]
<https://arxiv.org/abs/1810.04805>. Posted October 11, 2018. Accessed March 28, 2023.
6. Fedus W, Zoph B, Shazeer N. Switch transformers: scaling to trillion parameter models with simple and efficient sparsity. J Mach Learn Res 2023;23(1):5232–5270.
<https://dl.acm.org/doi/abs/10.5555/3586589.3586709>.
7. Shamshad F, Khan S, Zamir SW, et al. Transformers in Medical Imaging: A Survey. arXiv 2201.09873 [preprint] <https://arxiv.org/abs/2201.09873>. Posted January 24, 2022. Accessed March 28, 2023.
8. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv 2010.11929 [preprint] <https://arxiv.org/abs/2010.11929>. Posted October 22, 2020. Accessed March 28, 2023.
9. Bello I, Zoph B, Vaswani A, Shlens J, Le QV. Attention Augmented Convolutional Networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019; 3286–3295.
https://openaccess.thecvf.com/content_ICCV_2019/html/Bello_Attention_Augmented_Convolutional_Networks_ICCV_2019_paper.html. Accessed March 28, 2023.
10. Vaswani A, Ramachandran P, Srinivas A, Parmar N, Hechtman B, Shlens J. Scaling Local Self-Attention for Parameter Efficient Visual Backbones. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021; 12894–12904.
https://openaccess.thecvf.com/content/CVPR2021/html/Vaswani_Scaling_Local_Self-

Attention_for_Parameter_Efficient_Visual_Backbones_CVPR_2021_paper.html. Accessed March 28, 2023.

11. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth H, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. arXiv 2201.01266 [preprint] <https://arxiv.org/abs/2201.01266>. Posted January 4, 2022. Accessed March 28, 2023.
12. Tang Y, Yang D, Li W, et al. Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022; 20698–20708.
13. Hatamizadeh A, Tang Y, Nath V, et al. UNETR: Transformers for 3D Medical Image Segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022; 574–584. https://openaccess.thecvf.com/content/WACV2022/html/Hatamizadeh_UNETR_Transformers_for_3D_Medical_Image_Segmentation_WACV_2022_paper.html. Published 2022. Accessed March 28, 2023.
14. Fu J, Liu J, Tian H, et al. Dual Attention Network for Scene Segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019; 3146–3154. https://openaccess.thecvf.com/content_CVPR_2019/html/Fu_Dual_Attention_Network_for_Scene_Segmentation_CVPR_2019_paper.html. Accessed March 28, 2023.
15. Wang X, Girshick R, Gupta A, He K. Non-Local Neural Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018; 7794–7803. https://openaccess.thecvf.com/content_cvpr_2018/html/Wang_Non-Local_Neural_Networks_CVPR_2018_paper.html. Accessed March 28, 2023.
16. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18(2):203–211.
17. Tran CBN, Nedelec P, Weiss DA, et al. Development of Gestational Age-Based Fetal Brain and Intracranial Volume Reference Norms Using Deep Learning. AJNR Am J Neuroradiol 2023;44(1):82–90.
18. Xue Y, Xu T, Zhang H, Long LR, Huang X. SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation. Neuroinformatics 2018;16(3-4):383–392.
19. Pecco N, Canini M, Mosser KHH, et al. RS-FetMRI: a MATLAB-SPM Based Tool for Pre-processing Fetal Resting-State fMRI Data. Neuroinformatics 2022;20(4):1137–1154.
20. Reis HC, Turk V, Khoshelham K, Kaya S. MediNet: transfer learning approach with MediNet medical visual database. Multimedia Tools Appl 2023;82(25):1–44.
21. Aydin OU, Taha AA, Hilbert A, et al. On the usage of average Hausdorff distance for segmentation performance assessment: hidden error when used for ranking. Eur Radiol Exp 2021;5(1):4 [Published correction appears in Eur Radiol Exp 2022;6(1):56.].

22. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 2015;15(1):29.
23. Kostović I, Katušić A, Kostović Srzentić M. Chapter 19 - Linking histology and neurological development of the fetal and infant brain. In: Martin CR, Preedy VR, Rajendram R, eds. Factors Affecting Neurodevelopment: Genetics, Neurology, Behavior, and Diet. Academic Press, 2021; 213–225.
24. Khosla C, Saini BS. Enhancing Performance of Deep Learning Models with different Data Augmentation Techniques: A Survey. In: 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, UK. IEEE, 2020; 79–85.
25. Ciceri T, Squarcina L, Giubergia A, Bertoldo A, Brambilla P, Peruzzo D. Review on deep learning fetal brain segmentation from Magnetic Resonance images. Artif Intell Med 2023;143:102608.

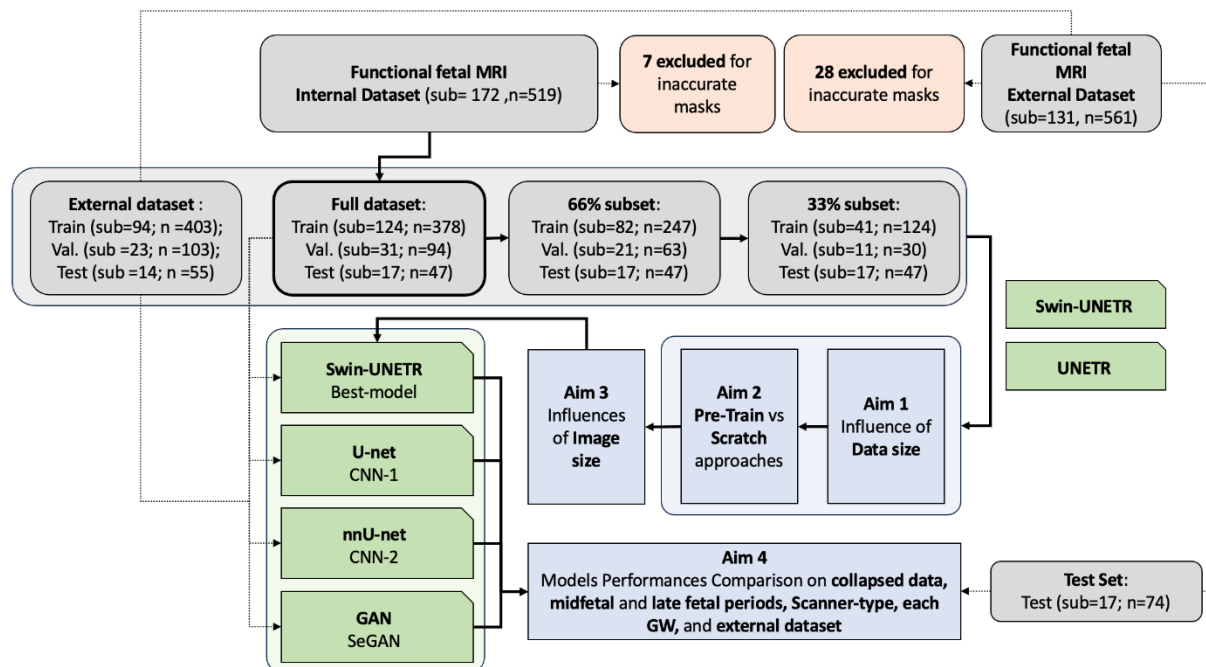


Figure 1: Flow diagram shows internal and external dataset allocation and splits for models training, validation and test sets and with respect to the study aims. CNN = convolutional neural network, GW = gestational week.

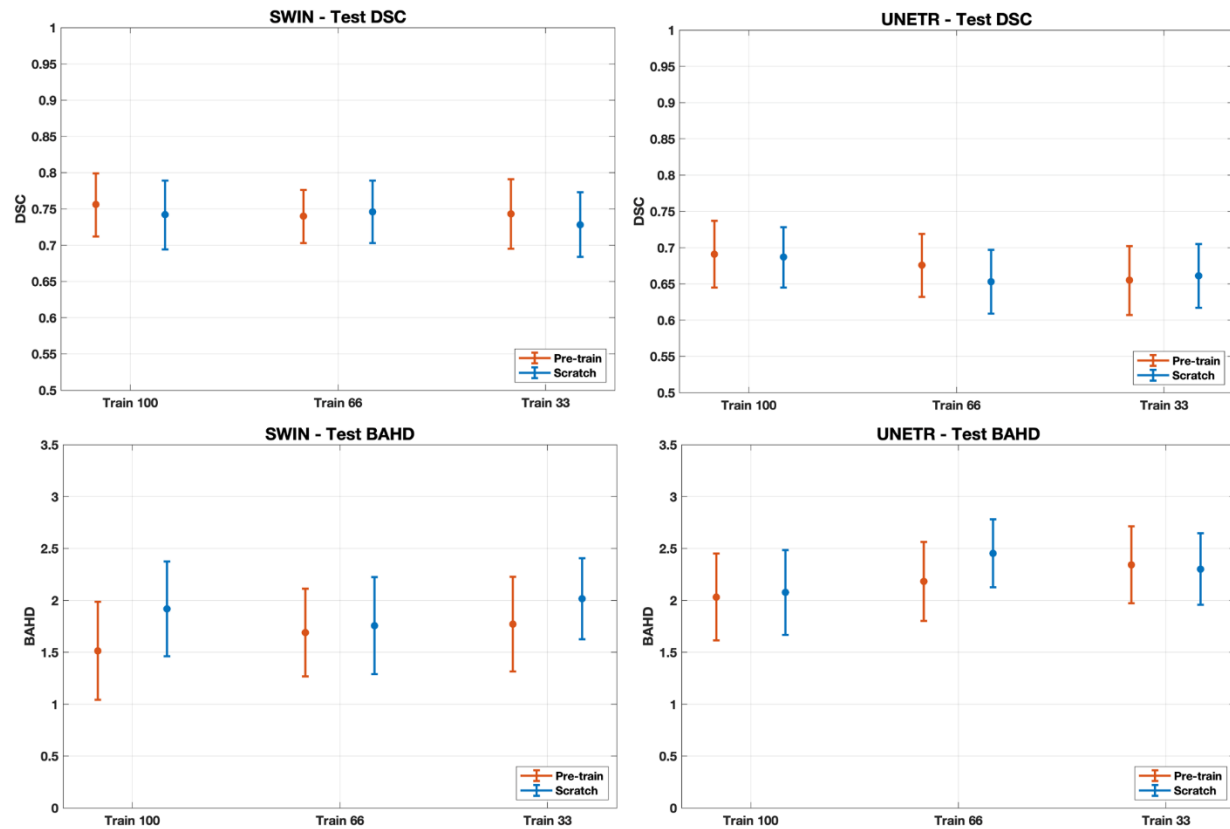


Figure 2: The figure shows two panels with plots comparing the performance of pretraining (orange bars) and scratch (blue bars) training for Swin-UNETR (left) and UNETR (right) models on three dataset configurations: Full, 66%, and 33%. Each panel is further divided into two subpanels: left subpanels show the mean DSC performance, and the right subpanels show the mean BAHD performance. DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance.

Just Accepted papers have undergone full peer review and have been accepted for publication. This article will undergo copyediting, layout, and proof review before it is published in its final version. Please note that during production of the final copyedited article, errors may be discovered which could affect the content.

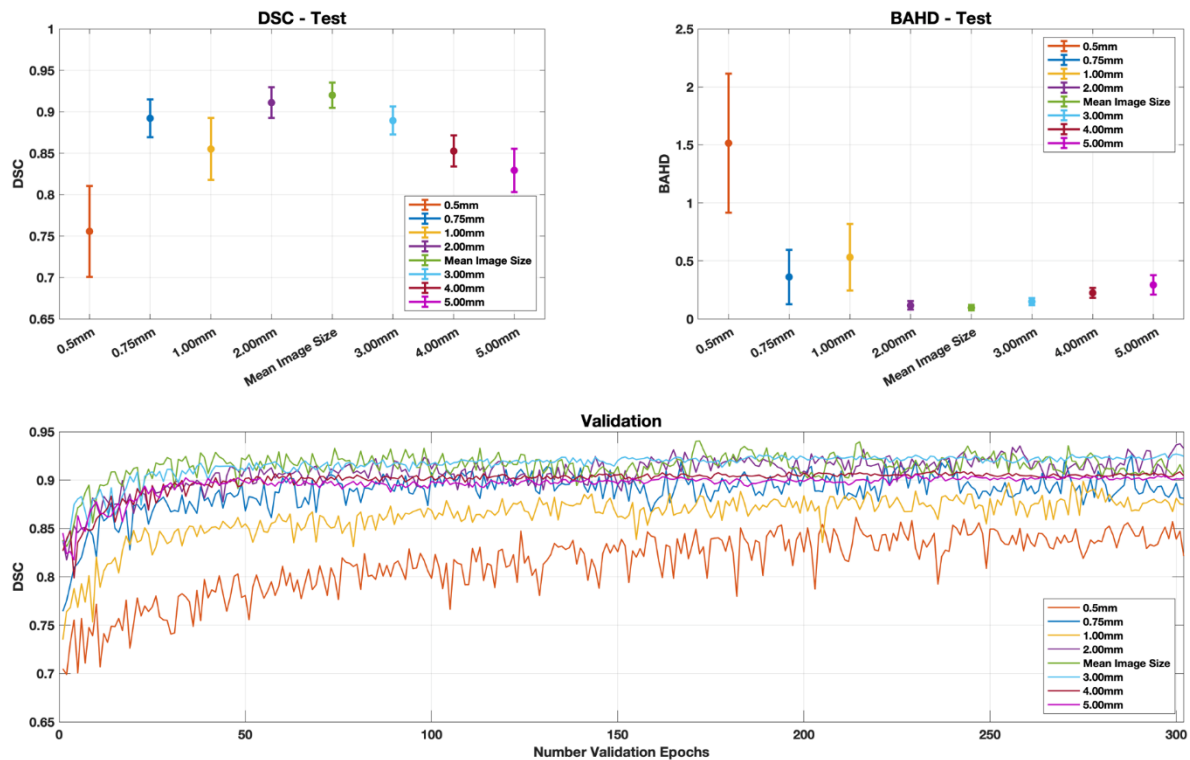


Figure 3: The graphs show the mean DSC (upper-left) and mean BAHD (upper-right) performance on the internal test set for the Swin-UNETR model using eight different input image sizes. The figure below reports DSC values for each validation fold. Color-coding is relative to the image voxel size—See Figure Legends. DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance.

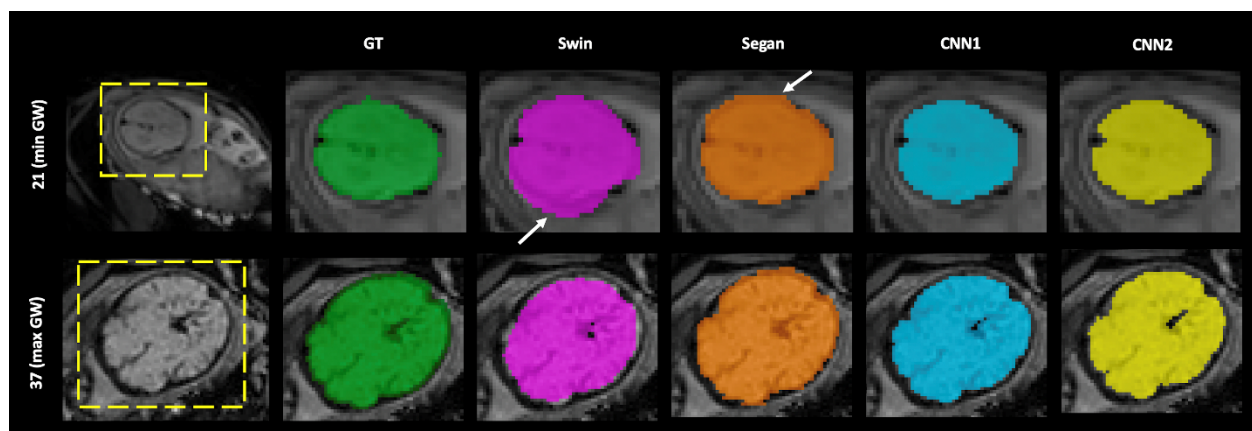


Figure 4: Raw fetal rs-fMRI volumes (first column) showing the 21st GW for the midfetal period (first row) and the 37th GW for the late-fetal period (second row). Segmentation masks are shown, respectively, for ground truth RS-FetMRI (GT) (second

column), the Swin-UNETR model (third column), SeGAN (fourth column), CNN-1 (fifth column) and CNN-2 (sixth column). White arrows indicate the models' segmentation errors (ie, inclusion of nonfetal brain tissue or exclusion of certain regions within the fetal brain). Rs-fMRI = resting state functional MRI, rs-fetMRI = resting state fetal MRI, GW = gestational week.

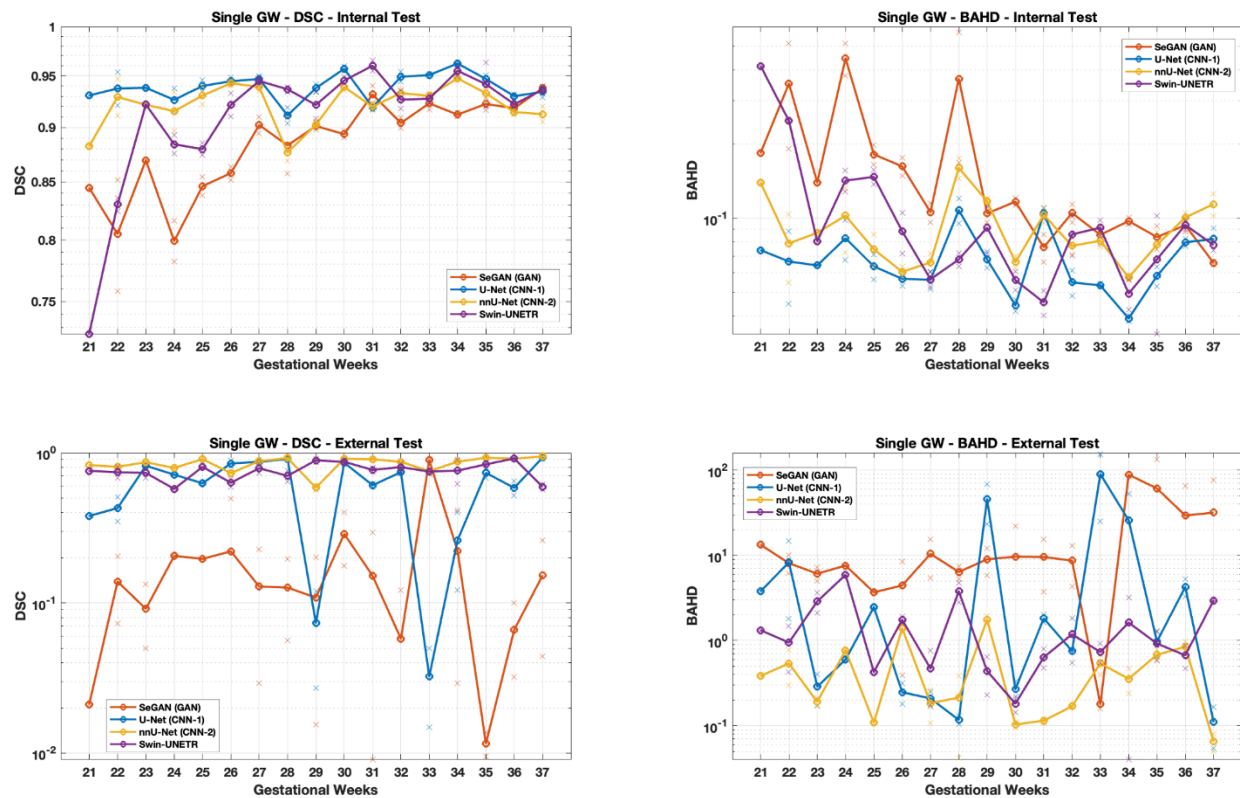


Figure 5: The graphs show the mean DSC (upper-panel) and mean BAHD (lower-panel) for each GW spanning from 21 to 37 for the Swin-UNETR (violet), CNN-1 (blue), CNN-2 (yellow) and SeGAN (orange) for models cross-validation on internal and external test sets. The y-axes scale was log-transformed for visualization purposes of the lower and upper bound CIs for all models DSC and BAHD performance metrics in the same graph. DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance, GW = gestational week

Table 1

Generalized Estimating Equations (GEE) Model Descriptions for Model Effect

Source	Type III		
	Wald Chi-Square	df	Sig.
GEE-1–DV = counts			
(Intercept)	0,327	1	0,567

Train/Validation	0,116	1	0,734
Subset	3,733	2	0,155
GW	0,15	1	0,698
Train/Validation * Subset	1,162	2	0,559
Train/Validation * GW	0,172	1	0,678
Subset * GW	4,49	2	0,106
Train/Validation * Subset * GW	2,2	2	0,333
GEE-2°-DV = DSC			
(Intercept)	1066,024	1	0
UNT/SWIN	135,877	1	0
Pretraining_Scratch	20,665	1	0
Subset	28,477	2	0
UNT/SWIN * Pretraining_Scratch	0	1	0,985
UNT/SWIN * Subset	5,168	2	0,075
Pretraining_Scratch * Subset	0,752	2	0,687
UNT/SWIN * Pretraining_Scratch * Subset	10,788	2	0,005
GEE-2b-DV = BAHD			
(Intercept)	100,897	1	0
UNT/SWIN	48,533	1	0
Pretraining_Scratch	54,956	1	0
Subset	15,477	2	0
UNT/SWIN * Pretraining_Scratch	11,043	1	0,001
UNT/SWIN * Subset	13,084	2	0,001
Pretraining_Scratch * Subset	4,256	2	0,119
UNT/SWIN * Pretraining_Scratch * Subset	9,023	2	0,011
GEE-3°-DV = DSC			
(Intercept)	8733,984	1	0
Image Size	358,889	7	0
GEE-3b-DV = BAHD			
(Intercept)	63,198	1	0
Image Size	151,723	7	0
GEE-4°-DV = DSC			
(Intercept)	85459,996	1	0
Scanner	0,178	1	0,673
Model	246,068	3	0
Scanner * Model	12,251	3	0,007
GEE-4b-DV = BAHD			
(Intercept)	242,021	1	0
Scanner	0,344	1	0,558
Model	106,106	3	0
Scanner * Model	12,66	3	0,005
GEE-5°-DV = DSC			
(Intercept)	246971,793	1	0
IN/EX	19,107	1	0
Model	559,112	3	0
IN/EX * Model	50,762	3	0
GEE-5b-DV = BAHD			
(Intercept)	272,879	1	0

IN/EX	0,016	1	0,898
Model	107,455	3	0
IN/EX * Model	13,619	3	0,003
GEE-6°-DV = DSC			
(Intercept)	174474,296	1	0
GW	47706,991	16	0
IN_TS/EX_TS	7453,432	1	0
Model	4343,381	3	0
GW * IN_TS/EX_TS	8606,754	16	0
GW * Model	3364366,656	46	0
IN_TS/EX_TS * Model	3870,951	3	0
GW * IN_TS/EX_TS * Model	7149872,813	46	0
GEE-6b-DV = BAHD			
(Intercept)	37,201	1	0
GW	15937423,58	16	0
IN_TS/EX_TS	35,178	1	0
Model	311,775	3	0
GW * IN_TS/EX_TS	14238338,92	16	0
GW * Model	27485975116	47	0
IN_TS/EX_TS * Model	290,78	3	0
GW * IN_TS/EX_TS * Model	21758434415	47	0

Note.—GEE models description. DV = Dependent Variable, DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance, UNT = UNETR, SWIN = Swin-UNETR, GW = Gestational Week, Subset = 100%, 66%, 33%, Model = Swin-UNETR, CNN1, CNN2 and GAN, IN = Internal Dataset, EX = External Dataset, IN_TS = Internal test set, EX_TS = External test set.

Table 2

Generalized Estimating Equations (GEE) Models' Significant Results Summarized for DSC and BAHD for Models, Pretraining and Dataset Size Effect (GEE-2c), Image Size Effect (GEE-3c) and Models and Scanner Effect (GEE-4c)

GEE 2c	DSC Mean Diff.	BAHD Mean Diff.
UNT*100%*P-SWIN*100%*P	-0,064*	0,517*
UNT*66%*P-SWIN*66%*P	-0,064*	0,495*
UNT*66%*S-SWIN*66%*S	-0,093*	0,696*
UNT*33%*P-SWIN*33%*P	-0,088*	0,572*
UNT*33%*S-SWIN*33%*S	-0,068*	0,285*
UNT*P*66%-UNT*P*33%	0,020*	-0,160*
UNT*S*100%-UNT*S*66%	0,034*	-0,376*
GEE 3c		
0.5-0.75	-0,137*	1,153*
0.5-1	-0,099*	0,983*
0.5-2	-0,155*	1,397*
0.5-mean image size	-0,164*	1,417*
0.5-3	-0,134*	1,365*
0.5-4	-0,097*	1,290*
0.5-5	-0,074*	1,221*

0.75-mean image size	-0,028*	0,264*
1-2	-0,056*	0,415*
1-mean image size	-0,065*	0,435*
1-3	-0,034*	0,382*
2-3	0,022*	-0,033*
2-4	0,058*	-0,108*
2-5	0,082*	-0,176*
mean image size-3	0,030*	-0,052*
mean image size-4	0,067*	-0,128*
mean image size-5	0,091*	-0,196*
3-4	0,037*	-0,075*
3-5	0,060*	-0,143*
GEE 4c		
1.5T*GAN-1.5T*CNN1	-0,043*	0,072*
1.5T*CNN1-1.5T*CNN2	0,021*	-0,029*
3T*GAN-3T*CNN1	-0,057*	0,111*
3T*GAN-3T*CNN2	-0,046*	0,096*
3T*GAN-3T*SWIN	-0,043*	0,089*
3T*CNN1-3T*CNN2	0,011*	-0,016*

Note.—Significant Posthoc test Comparisons are reported in terms of Mean Difference (Mean Diff.). DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance, UNT = UNETR, SWIN = Swin-UNETR, Subset = 100%, 66%, 33%, *P* = Pretraining, S = Scratch, 1.5T = 1.5 Tesla scanner, 3T = 3 Tesla scanner. See Table E9-GEE-2c, GEE-3c and GEE-4c—for post hoc tests corrected for sequential Bonferroni comparisons, Standard Error and Wald 95% CIs.

Table 3

Generalized Estimating Equations (GEE) Models' Significant Results Summarized for DSC and BAHD for Models and Datasets Effect

GEE 5c	DSC Mean Diff.	BAHD Mean Diff.
IN*GAN-IN*CNN1	-0,049*	0,090*
IN*GAN-IN*CNN2	-0,033*	0,067*
IN*GAN-IN*SWIN	-0,030*	0,061*
IN*CNN1-IN*CNN2	0,017*	-0,023*
IN*CNN1-IN*SWIN	0,019*	-0,029*
EX*GAN-EX*CNN1	-0,068*	0,112*
EX*GAN-EX*CNN2	-0,068*	0,135*
EX*GAN-EX*SWIN	-0,071*	0,139*

Note.—Significant Posthoc test Comparisons are reported in terms of Mean Difference (Mean Diff.). DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance, UNT = UNETR, SWIN = Swin-UNETR, IN = Internal Dataset, EX = External Dataset. See Table E9-GEE-5c—for post hoc tests corrected for sequential Bonferroni comparisons, Standard Error and Wald 95% CIs.

Table 4

Generalized Estimating Equations (GEE) Models' Significant Results Summarized for DSC and BAHD for Model, Gestational Week and Test Sets Effect

GEE 6c	DSC Mean Diff.	BAHD Mean Diff.		DSC Mean Diff.	BAHD Mean Diff.		DSC Mean Diff.	BAHD Mean Diff.
22*IN_TS*CNN1–22*IN_TS*SWIN	0,107*	–0,182*	29*EX_TS*CNN2–29*EX_TS*SWIN	–0,306*	1,325*	22*IN_TS*GAN–22*EX_TS*GAN	0,666*	–7,734*
22*IN_TS*CNN2–22*IN_TS*SWIN	0,099*	–0,17*	30*IN_TS*GAN–30*IN_TS*CNN1	–0,063*	0,073*	23*IN_TS*GAN–23*EX_TS*GAN	0,778*	–5,963*
22*EX_TS*GAN–22*EX_TS*CNN2	–0,668*	7,548*	30*IN_TS*GAN–30*IN_TS*CNN2	–0,045*	0,05*	23*IN_TS*CNN2–23*EX_TS*CNN2	0,057*	–0,103*
22*EX_TS*GAN–22*EX_TS*SWIN	–0,604*	7,139*	30*IN_TS*GAN–30*IN_TS*SWIN	–0,052*	0,061*	23*IN_TS*SWIN–23*EX_TS*SWIN	0,187*	–2,803*
23*EX_TS*GAN–23*EX_TS*CNN1	–0,729*	5,816*	30*IN_TS*CNN1–30*IN_TS*CNN2	0,018*	–0,022*	24*IN_TS*GAN–24*EX_TS*GAN	0,593*	–7,136*
23*EX_TS*GAN–23*EX_TS*CNN2	–0,774*	5,913*	30*IN_TS*CNN2–30*IN_TS*SWIN	–0,006*	0,01*	24*IN_TS*CNN1–24*EX_TS*CNN1	0,211*	–0,516*
23*EX_TS*CNN2–23*EX_TS*SWIN	0,13*	–2,693*	30*EX_TS*CNN1–30*EX_TS*CNN2	–0,053*	0,165*	24*IN_TS*CNN2–24*EX_TS*CNN2	0,122*	–0,667*
24*IN_TS*GAN–24*IN_TS*CNN1	–0,127*	0,364*	30*EX_TS*CNN2–30*EX_TS*SWIN	0,042*	–0,078*	24*IN_TS*SWIN–24*EX_TS*SWIN	0,31*	–5,764*
24*IN_TS*GAN–24*IN_TS*CNN2	–0,116*	0,344*	31*IN_TS*GAN–31*IN_TS*SWIN	–0,028*	0,031*	25*IN_TS*GAN–25*EX_TS*GAN	0,65*	–3,503*
24*IN_TS*GAN–24*IN_TS*SWIN	–0,085*	0,304*	31*IN_TS*CNN1–31*IN_TS*SWIN	–0,041*	0,059*	25*IN_TS*CNN1–25*EX_TS*CNN1	0,313*	–2,405*
24*IN_TS*CNN1–24*IN_TS*SWIN	0,042*	–0,059*	31*IN_TS*CNN2–31*IN_TS*SWIN	–0,04*	0,058*	25*IN_TS*CNN2–25*EX_TS*CNN2	0,023*	–0,034*
25*IN_TS*GAN–25*IN_TS*CNN1	–0,094*	0,118*	31*EX_TS*CNN1–31*EX_TS*CNN2	–0,297*	1,711*	25*IN_TS*SWIN–25*EX_TS*SWIN	0,074*	–0,28*
25*IN_TS*GAN–25*IN_TS*CNN2	–0,084*	0,107*	31*EX_TS*CNN1–31*EX_TS*SWIN	–0,161*	1,19*	26*IN_TS*CNN1–26*EX_TS*CNN1	0,095*	–0,191*
25*IN_TS*CNN1–25*IN_TS*CNN2	0,009*	–0,011*	31*EX_TS*CNN2–31*EX_TS*SWIN	0,136*	–0,521*	26*IN_TS*CNN2–26*EX_TS*CNN2	0,209*	–1,334*
25*IN_TS*CNN1–25*IN_TS*SWIN	0,06*	–0,084*	32*IN_TS*GAN–32*IN_TS*CNN1	–0,045*	0,05*	26*IN_TS*SWIN–26*EX_TS*SWIN	0,29*	–1,667*
25*IN_TS*CNN2–25*IN_TS*SWIN	0,051*	–0,072*	32*IN_TS*GAN–32*IN_TS*SWIN	–0,022*	0,019*	27*IN_TS*CNN1–27*EX_TS*CNN1	0,076*	–0,153*
26*IN_TS*GAN–26*IN_TS*CNN1	–0,087*	0,106*	32*IN_TS*CNN1–32*IN_TS*CNN2	0,016*	–0,022*	28*IN_TS*GAN–28*EX_TS*GAN	0,757*	–6,022*
26*IN_TS*GAN–26*IN_TS*CNN2	–0,085*	0,102*	32*EX_TS*CNN1–32*EX_TS*CNN2	–0,128*	0,584*	28*IN_TS*SWIN–28*EX_TS*SWIN	0,235*	–3,725*
26*IN_TS*SWIN	–0,064*	0,074*	33*IN_TS*GAN–33*IN_TS*CNN1	–0,028*	0,032*	29*IN_TS*GAN–29*EX_TS*GAN	0,793*	–8,891*
26*IN_TS*CNN1–26*IN_TS*SWIN	0,023*	–0,032*	33*IN_TS*CNN1–33*IN_TS*CNN2	0,021*	–0,028*	29*IN_TS*CNN2–29*EX_TS*CNN2	0,317*	–1,646*
26*EX_TS*CNN1–26*EX_TS*CNN2	0,116*	–1,147*	33*IN_TS*CNN1–33*IN_TS*SWIN	0,023*	–0,038*	30*IN_TS*CNN1–30*EX_TS*CNN1	0,098*	–0,224*
26*EX_TS*CNN1–26*EX_TS*SWIN	0,217*	–1,508*	33*EX_TS*GAN–33*EX_TS*CNN2	0,139*	–0,36*	30*IN_TS*CNN2–30*EX_TS*CNN2	0,026*	–0,037*
26*EX_TS*CNN2–26*EX_TS*SWIN	0,102*	–0,361*	33*EX_TS*GAN–33*EX_TS*SWIN	0,144*	–0,554*	30*IN_TS*SWIN–30*EX_TS*SWIN	0,075*	–0,125*
27*IN_TS*GAN–27*IN_TS*CNN1	–0,045*	0,05*	34*IN_TS*GAN–34*IN_TS*CNN1	–0,05*	0,058*	31*IN_TS*CNN1–31*EX_TS*CNN1	0,311*	–1,721*
27*IN_TS*GAN–27*IN_TS*CNN2	–0,037*	0,04*	34*IN_TS*GAN–34*IN_TS*CNN2	–0,035*	0,04*	31*IN_TS*SWIN–31*EX_TS*SWIN	0,19*	–0,591*
27*IN_TS*GAN–27*IN_TS*SWIN	–0,043*	0,049*	34*IN_TS*GAN–34*IN_TS*SWIN	–0,043*	0,048*	32*IN_TS*CNN1–32*EX_TS*CNN1	0,206*	–0,698*
27*IN_TS*CNN1–27*IN_TS*CNN2	0,008*	–0,01*	34*IN_TS*CNN1–34*IN_TS*CNN2	0,015*	–0,018*	32*IN_TS*CNN2–32*EX_TS*CNN2	0,062*	–0,092*

28*IN_TS*CNN1– 28*IN_TS*CNN2	0,035*	–0,052*	35*IN_TS*CNN1– 35*IN_TS*CNN2	0,014*	–0,02*	33*IN_TS*GAN– 33*EX_TS*GAN	0,029*	–0,093*
28*IN_TS*CNN1– 28*IN_TS*SWIN	–0,025*	0,04*	36*IN_TS*CNN1– 36*IN_TS*CNN2	0,015*	–0,021*	33*IN_TS*CNN2– 33*EX_TS*CNN2	0,175*	–0,458*
28*IN_TS*CNN2– 28*IN_TS*SWIN	–0,06*	0,092*	36*IN_TS*CNN1– 36*IN_TS*SWIN	0,008*	–0,014*	33*IN_TS*SWIN– 33*EX_TS*SWIN	0,178*	–0,641*
28*EX_TS*GAN– 28*EX_TS*CNN1	–0,783*	6,274*	36*EX_TS*CNN1– 36*EX_TS*CNN2	–0,331*	3,42*	34*IN_TS*CNN2– 34*EX_TS*CNN2	0,075*	–0,298*
28*EX_TS*GAN– 28*EX_TS*CNN2	–0,796*	6,177*	36*EX_TS*CNN1– 36*EX_TS*SWIN	–0,333*	3,594*	35*IN_TS*CNN1– 35*EX_TS*CNN1	0,212*	–0,916*
28*EX_TS*CNN1– 28*EX_TS*SWIN	0,208*	–3,675*	37*IN_TS*GAN– 37*IN_TS*CNN2	0,025*	–0,048*	35*IN_TS*SWIN– 35*EX_TS*SWIN	0,104*	–0,854*
28*EX_TS*CNN2– 28*EX_TS*SWIN	0,22*	–3,579*	37*IN_TS*CNN1– 37*IN_TS*CNN2	0,022*	–0,032*	36*IN_TS*CNN1– 36*EX_TS*CNN1	0,346*	–4,185*
29*IN_TS*GAN– 29*IN_TS*CNN1	–0,036*	0,037*	37*IN_TS*CNN2– 37*IN_TS*SWIN	–0,024*	0,036*	37*IN_TS*CNN2– 37*EX_TS*CNN2	–0,033*	0,049*
29*IN_TS*CNN1– 29*IN_TS*CNN2	0,035*	–0,049*	37*EX_TS*CNN1– 37*EX_TS*SWIN	0,337*	–2,843*	37*IN_TS*SWIN– 37*EX_TS*SWIN	0,343*	–2,876*
29*EX_TS*GAN– 29*EX_TS*SWIN	–0,783*	8,557*	37*EX_TS*CNN2– 37*EX_TS*SWIN	0,352*	–2,888*			

Note.—Significant Posthoc test Comparisons are reported in terms of Mean Difference (Mean Diff.). DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance, SWIN = Swin-UNETR, IN_TS = Internal test set, EX_TS = External test set. See Table E9-GEE-6c-for post hoc tests corrected for sequential Bonferroni comparisons, Standard Error and Wald 95% CIs.

Supplemental Materials

Materials and Methods

Internal Dataset (Rs-fMRI Image Acquisition)

Fetal MR scanning was performed on a Philips Achieva 1.5T scanner and on a Philips Intera 3T scanner, using a 16-channels body coil. Functional scans (rs-fMRI) acquired on 1.5 scanner consisted of GE EPI scans (TR = 2000 ms, TE = 30 ms, acquisition voxel size $2.25 \times 2.25 \times 3$ mm, # slices = 25, slice gap = 0). Each rs-fMRI scan consisted of 60 volumes lasting 2 s each, for a total scanning time of 2 min per scan. Four to 8 consecutive rs-fMRI sessions (ie, 240–480 scans, covering from 8 to 16 min of continuous brain activity at rest) (1) were acquired for each subject depending on the quality of the scans. Fetal MR performed on a 3T scanner consisted of GE EPI scans (TR = 2000 ms, TE = 30 ms, acquisition voxel size $2.24 \times 2.24 \times 3$ mm, # slices = 28, slice gap = 0). Each rs-fMRI scan consisted of 60 volumes lasting 2 s each, for a total scanning time of 2 min per scan. Four to 8 consecutive rs-fMRI sessions (ie, 240–480 scans, covering from 8 to 16 min of continuous brain activity at rest) were acquired for each subject depending on the quality of the scans. Each rs-fMRI scan consisted of 25 to 28 slices (matrix in-plane resolution range: 80:96 voxels. In-plane voxel resolution was 2.25×2.25 at 1.5T and 2.24×2.24 at 3T (average in-plane resolution = 2.24×2.24 mm²) with slice thickness of 3 mm³.

Internal Dataset (Rs-fMRI Ground Truth)

The internal dataset was processed using the RS-FetMRI package (<https://github.com/NicoloPecco/RS-FetMRI>) (22), a single-subject semiautomatic fetal rs-fMRI image processing protocol with six interdependent modules. It employs the SPM Unified-Segmentation algorithm (23) for fetal brain extraction, deriving session-specific inner-brain masks based on a priori probability maps of GW-specific fetal inner brain space. A set of representative fetal brain functional scans spanning across GW sample distribution was manually segmented in each acquisition plan by C.B., a senior expert fetal and pediatric neuroradiologist with 20 years' experience, using ITK-SNAP (www.itksnap.org) to assess similarity and distance between manual and automatic RS-FetMRI segmentation masks for model ground truth evaluation.

Internal Dataset (Ground Truth Assessment)

Ground Truth Assessment The similarity and distance metrics evaluation between manual segmentation masks and RS-FetMRI automatic masks on the representative set ($n = 47$ scans) showed a DSC of 0.906 (95 CI: 0.900–0.913) and a BAHD of 0.08 (95% CI: 0.07–0.09), demonstrating a high degree of overlap.

External Dataset (Rs-fMRI Image Acquisition)

Fetal MR scanning was performed for 197 fetuses (gestational age 24–39 weeks, $M = 30.9$, $SD = 4.2$) on a Siemens Verio 3T scanner using an abdominal 4-Channel Flex Coil using echo-planar sequence (TR/TE: 2000/30; 4 mm slice thickness, axial, interleaved ascending slice order, 360 volumes) and multiecho resting-state sequences (TR/TEs: 2000/18,34,50). Other ten fetuses (gestational ages 30–36 weeks, $M = 32.7$, $SD = 1.9$) were acquired on a Siemens Skyra 3T using a 32-channel abdominal coil (TR/TE:2000/30; 3 mm slices, 32 slices parallel to the bicommissural plane, 150 volumes) (5). Each rs-fMRI scan consisted of 37 slices and a matrix in-plane resolution of 96×96 voxels with an isotropic voxel resolution of 3.5 mm^3 .

External Dataset (Rs-fMRI Image Processing)

The external dataset, obtained from Openneuro.org, comprises 159 subjects, each undergoing 3 to over 10 sessions, along with their corresponding manually segmented ground truth. A total of 28 subjects were excluded by P.A.D.R. for inaccurate masking, resulting in 131 subjects with 561 scans. Variations in scanning parameters among subjects, sessions, and runs are reported, consisting indiscrepancies in affine matrices and matrix dimensions between image-ground truth pairs (ie, $96 \times 96 \times 37$ or $96 \times 37 \times 96$). To address this issue, the scans' affine matrices for all subjects were replaced with the ground truth's affine matrix. However, 67 subjects still exhibited distinct matrix dimensions between scans and ground truth, consistently in the form of $96 \times 96 \times 37$ for ground truth and $96 \times 37 \times 96$ for the respective scan. To address matrix-dimension inconsistencies, a zero-padding technique was applied to the smaller dimension, accompanied by a crop on the third dimension.

Models' Development

For Swin-UNETR and UNETR random flip, rotation, and intensity scaling parameters were applied with probabilities of 0.1, 0.1, and 0.5, respectively. Patches number was set to 4 with an equal ratio of positive and negative samples. Divisible padding was used to fulfill models dimension requirements. Batch size was set to 1 and data were reshuffled at every epoch (2). Both models used self-supervised pretraining approach (3,4), leveraging unlabeled data to exploit the intrinsic structure without manual annotation. Evaluation number was set to 100 for the full dataset and later adapted to 66 and 33 when assessing datasets representing 66% and 33% of the entire dataset, respectively. The Swin-UNETR model was pretrained on 5050 publicly available CT images (12), while the UNETR was pretrained on 771 CT images from the TCIA-Covid-19 CT dataset (13). Influence of dataset size and pretraining versus scratch approaches were studied by evaluating UNETR and Swin-UNETR on the three datasets (100%, 66% and 33%) and by modifying the weight initialization. Particularly, the transfer learning approach was evaluated by continuing the training process with the pretrained weights (pretraining) against the random weights initialization (scratch). The influence of image size was studied by upsampling the voxel size to 0.5 mm^3 , 0.75 mm^3 , 1 mm^3 and 2 mm^3 and downsampling to 3 mm^3 , 4 mm^3 and 5 mm^3 and by including the mean image size ($2.24 \times 2.24 \times 3 \text{ mm}^3$). Bilinear and nearest neighbor interpolation methods were used for images and ground-truth, respectively.

Dataset Stratification

Due to the absence of external dataset GW information, we utilized the Fetal Brain GA Prediction tool (6) to estimate the GW for each subject. We then randomly generated a test set to mirror the gestational week distribution of our internal test set (one subject per GW), for the cross-validation of models trained on our internal dataset, aimed at testing differences or similarities between model performances and specific GW distributional ranges.

Statistical Analysis

We used an exchangeable working correlation structure for repeated measures and clustered data. We fitted all GEE models with the main effects and interaction effects (ie, full factorial) for multifactorial models.

References

1. Van Dijk KR, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J Neurophysiol* 2010;103(1):297–321.
2. Kerley CI, Cai LY, Tang Y, et al. Batch size: go big or go home? Counterintuitive improvement in medical autoencoders with smaller batch size. *Proc SPIE Int Soc Opt Eng* 2023;12464:124640H.

3. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth H, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. arXiv 2201.01266 [preprint] <https://arxiv.org/abs/2201.01266>. Posted January 4, 2022. Accessed March 28, 2023.
4. Hatamizadeh A, Tang Y, Nath V, et al. UNETR: Transformers for 3D Medical Image Segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022; 574–584. https://openaccess.thecvf.com/content/WACV2022/html/Hatamizadeh_UNETR_Transformers_for_3D_Medical_Image_Segmentation_WACV_2022_paper.html. Published 2022. Accessed March 28, 2023.
5. Rutherford S, Sturmfels P, Angstadt M, et al. Automated brain masking of fetal functional MRI with open data. Neuroinformatics 2022;20(1):173–185.
6. Vahedifard F, Liu X, Marathu KK, et al. Artificial Intelligence Prediction of Gestational Age of Fetal in Brain Magnetic Resonance Imaging versus ultrasound Using three different Biometric Measurements. Preprints 2023, 2023072052. . Posted July 21, 2023. Accessed DATE.

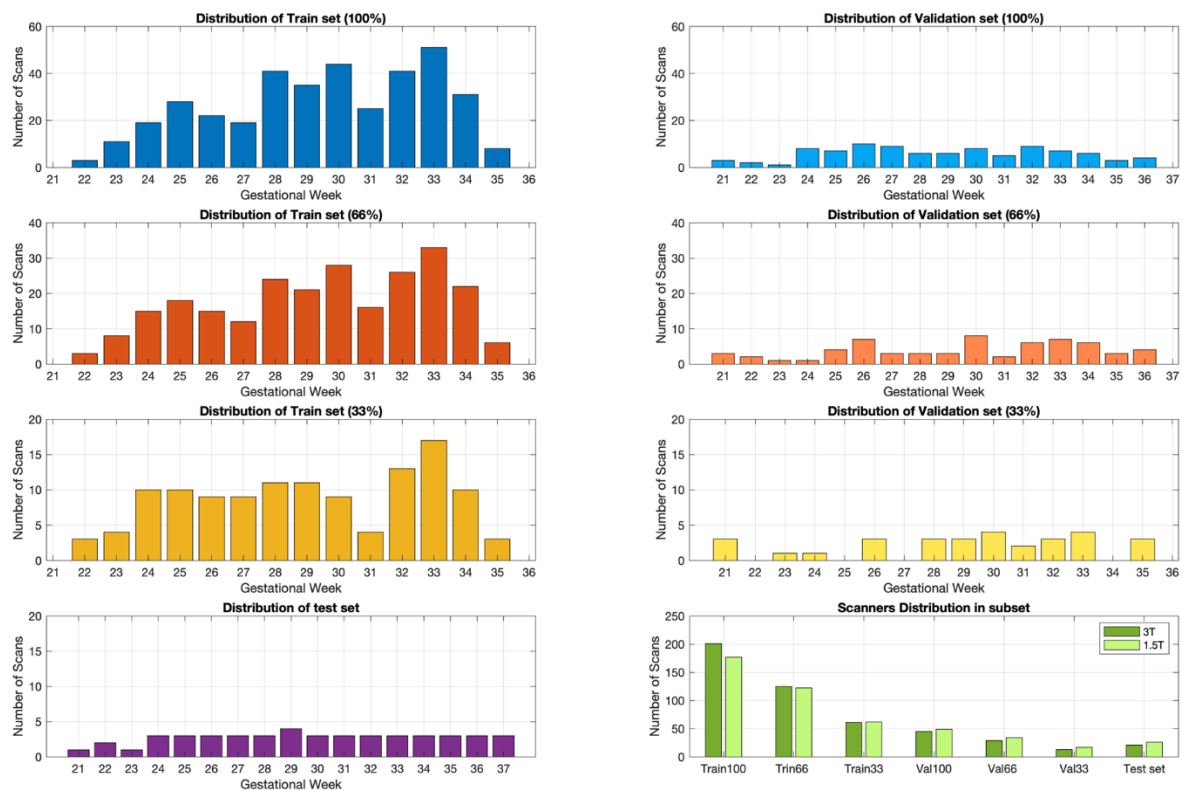


Figure E1: Distribution of fetal images across gestational weeks in train and validation sets, for full dataset, subsets, and test set. The y-axis displays the number of images, while the x-axis shows gestational week ranges. The bars are colored to reflect the proportion of the dataset they represent, with blue representing 100%, orange representing 66%, and yellow representing 33% of the dataset.

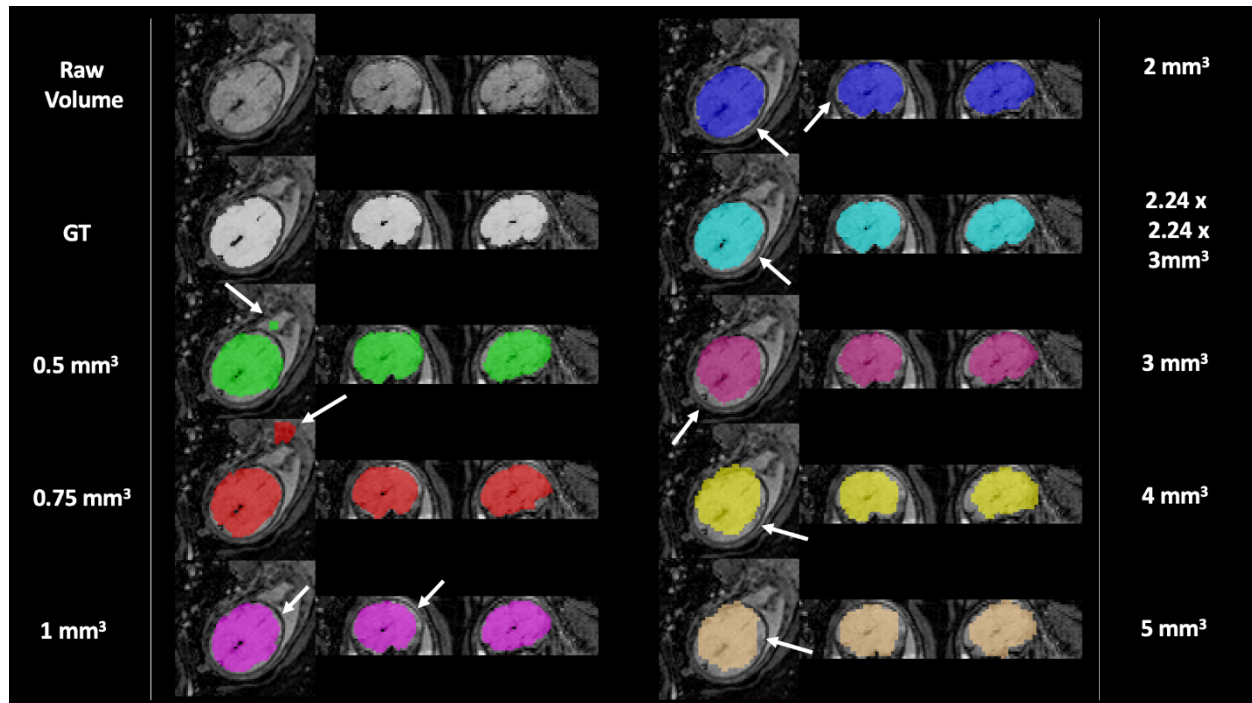


Figure E2: The Figure reports the raw fetal rs-fMRI volume for a 32 GW subject and the SWINs' models-derived brain segmentations for the image size evaluation. White arrows indicate errors (ie, inclusion of nonfetal brain tissue or exclusion of certain regions within the fetal brain).

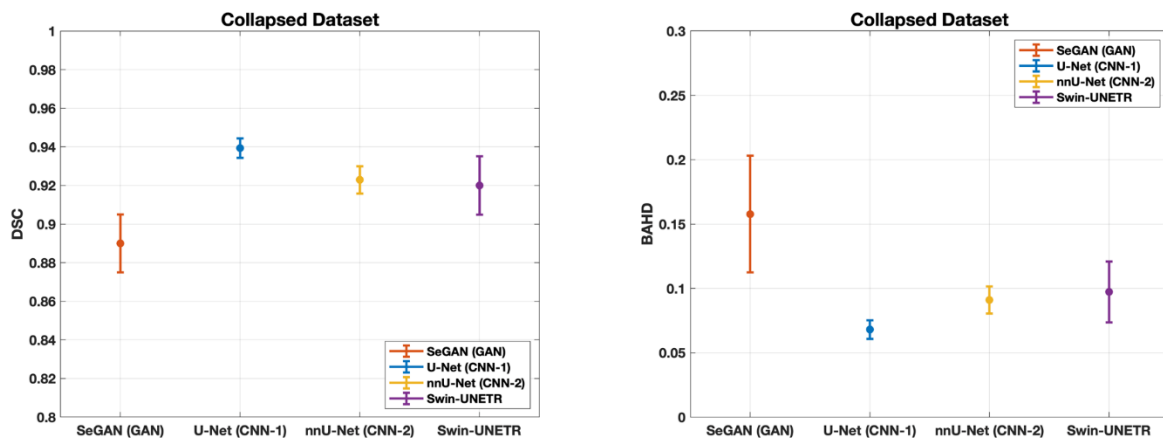


Figure E3: The Figure shows the mean DSC (left) and mean BAHD (right) with CIs performance on the test-set for the SeGAN (orange), CNN-1 (blue), CNN-2 (yellow) and Swin UNETR (violet).

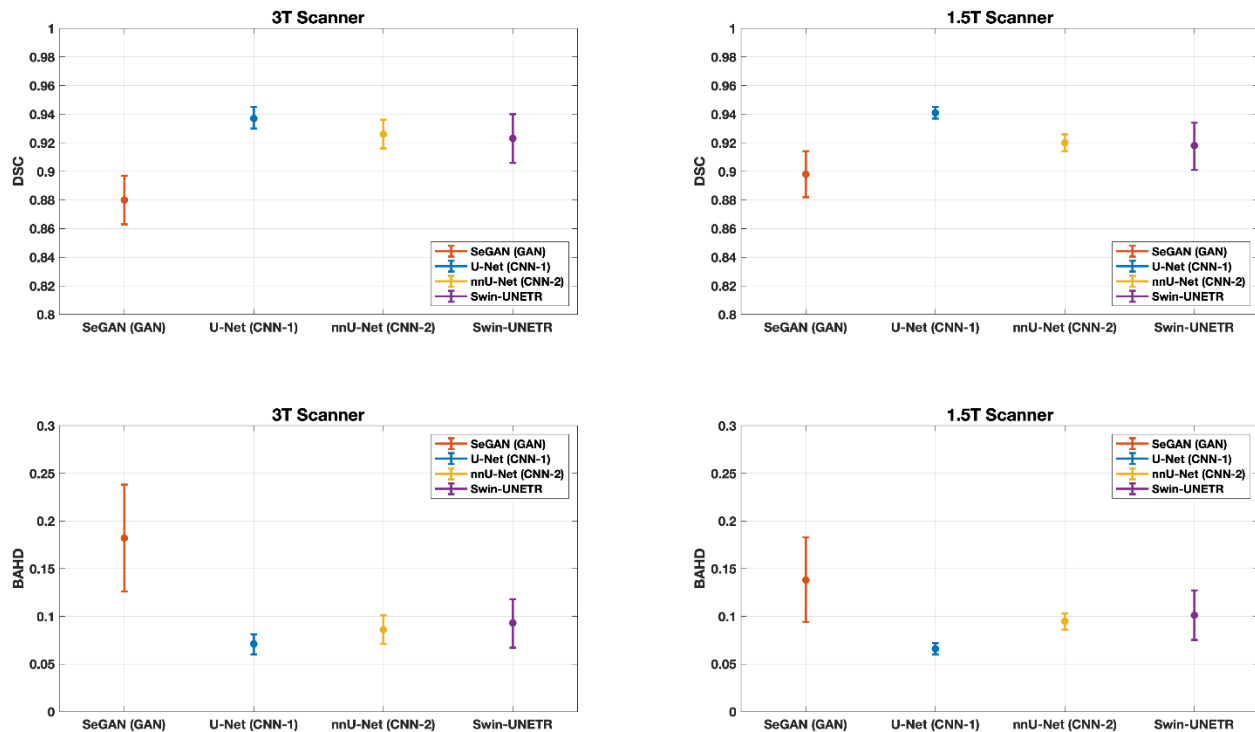


Figure E4: The Figure shows the mean DSC (upper-panel) for 3T (left) and 1.5T (right) and mean BAHD (lower-panel) and CIs for 3T (left) and 1.5T (right) performances on the test set for the SeGAN (orange), CNN-1 (blue), CNN-2 (yellow) and Swin UNETR (violet).

Table E1

Key Hyperparameters for Each Model Tested for Aim 4: SWIN-UNETR, nnU-net; U-Net; and the GAN

Model	Input Images	Batch Size	Learning Rate	Number Iteration (10^3)	Data Augmentation	Loss Function	Training Time (Hours)
Swin	3D	1	10^{-4}	30	Yes	DSC	≈13
nnU-Net	3D	22	10^{-2}	2	Yes	Cross-Entropy	≈20
U-Net	3D	12	10^{-4}	20	Yes	DSC	≈125
GAN	2D	36	$2 * 10^{-4}$	300	Yes	Multiscale L1	≈60

Note.—The table reports information for type of input, batch size, learning rate, number of iterations, data-augmentation usage, the loss function implemented and training time (in hours) for each model.

Results

Table E2

Percentage Overview of Subjects in Train, Validation and Test Set for Internal Dataset, Subsets and External Dataset

Dataset Size	Set	Number of Subjects (%)
	<i>Train</i>	124/172%–72.09%
100%	<i>Val.</i>	31/172%–18.02%
	<i>Train</i>	82/172%–47.67%
66%	<i>Val.</i>	21/172%–12.20%
	<i>Train</i>	41/172%–23.83%
33%	<i>Val.</i>	11/172%–6.39%
	<i>Test</i>	17/172%–9.88%
External Dataset	<i>Train</i>	94/131%–71.75%
	<i>Val.</i>	23/131%–17.55%
	<i>Test</i>	14/131%–10.68%

Note.—Percentage and ratios of fetal rs-fMRI subjects scans for the 100%, 66% and 33% subsets for train, validation and test set and external dataset.

Table E3

Mean DSC and BAHD and 95% CIs for the UNETR and Swin-UNETR for the Full Internal Dataset and 66% and 33% Subsets Using Pretraining and Scratch Approaches

Model	P-S	Subset	DSC				BAHD			
			Mean	Std. Error	95% CI		Mean	Std. Error	95% CI	
					Lower	Upper			Lower	Upper
UNETR	P	100%	0,691	0,024	0,645	0,737	2,031	0,214	1,613	2,450
		66%	0,676	0,022	0,632	0,719	2,183	0,194	1,803	2,563
		33%	0,655	0,024	0,607	0,702	2,343	0,190	1,971	2,714
	S	100%	0,687	0,021	0,645	0,728	2,078	0,208	1,669	2,486
		66%	0,653	0,022	0,609	0,697	2,454	0,167	2,126	2,781
		33%	0,661	0,022	0,617	0,705	2,302	0,176	1,957	2,646
SWIN-UNETR	P	100%	0,756	0,022	0,712	0,799	1,514	0,241	1,042	1,987
		66%	0,740	0,018	0,703	0,776	1,689	0,215	1,267	2,111
		33%	0,743	0,024	0,695	0,791	1,771	0,233	1,315	2,228
	S	100%	0,742	0,024	0,694	0,789	1,918	0,233	1,462	2,374
		66%	0,746	0,022	0,703	0,789	1,757	0,238	1,290	2,224
		33%	0,728	0,023	0,684	0,773	2,017	0,199	1,626	2,407

Note.—Data in parentheses are 95% confidence intervals. DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance, P = Pretrain, S = Scratch.

Table E4**Overview of Model Performance with Variable Image Size**

	DSC				BAHD					
Image Voxel Size (mm)	Mean	Std. Error	95% CI		Mean	Std. Error	95% CI		Cropping pad	Divisible Pad (k)
			Lower	Upper			Lower	Upper		
0.50 × 0.50 × 0.50	0,756	0,022	0,712	0,799	1,514	0,241	1,042	1,987	96 × 96 × 96	–
0.75 × 0.75 × 0.75	0,892	0,009	0,874	0,910	0,361	0,094	0,177	0,546	96 × 96× 96	–
1.00 × 1.00 × 1.00	0,855	0,015	0,826	0,884	0,532	0,116	0,305	0,759	64 × 64 × 64	–
2.00 × 2.00 × 2.00	0,911	0,007	0,896	0,926	0,117	0,015	0,089	0,146	64 × 64 × 64	64
2.24 × 2.24 × 3.00	0,920	0,006	0,908	0,932	0,097	0,010	0,079	0,116	64 × 64 × 64	64
3.00 × 3.00× 3.00	0,889	0,007	0,876	0,903	0,150	0,013	0,125	0,174	64 × 64 × 64	64
4.00 × 4.00 × 4.00	0,853	0,008	0,838	0,867	0,225	0,017	0,192	0,258	64 × 64 × 64	64
5.00 × 5.00 × 5.00	0,829	0,011	0,809	0,850	0,293	0,034	0,227	0,359	64 × 64 × 64	64

Note.—Mean DSC and BAHD Swin-UNETR performance with 95% CIs for different image voxel dimensions, the cropping pad windows and the divisible pad. DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance.

Table E5**Model Performances on 1.5 T and 3 T Scanners**

Scanner	Model	DSC				BAHD			
		Mean	Std. Error	95% CI		Mean	Std. Error	95% CI	
				Lower	Upper			Lower	Upper
1.5T	GAN	0,898	0,008	0,882	0,914	0,138	0,023	0,094	0,183
	CNN1	0,941	0,002	0,937	0,945	0,066	0,003	0,060	0,072
	CNN2	0,920	0,003	0,914	0,926	0,095	0,004	0,086	0,103
	SWIN	0,918	0,008	0,901	0,934	0,101	0,013	0,075	0,127
3T	GAN	0,880	0,009	0,863	0,897	0,182	0,029	0,126	0,238
	CNN1	0,937	0,004	0,930	0,945	0,071	0,005	0,060	0,081
	CNN2	0,926	0,005	0,916	0,936	0,086	0,008	0,071	0,101
	SWIN	0,923	0,009	0,906	0,940	0,093	0,013	0,067	0,118

Note.—Mean DSC and BAHD Swin-UNETR performance with 95% CIs for the Internal and External datasets. DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance.

Table E6**Models Generalizability-Overview of Model Performance with Internal and External Dataset**

Dataset	Model	DSC				BAHD			
		Mean	Std. Error	95% CI		Mean	Std. Error	95% CI	
				Lower	Upper			Lower	Upper
Internal	GAN	0,890	0,006	0,878	0,902	0,158	0,018	0,122	0,193
	CNN1	0,939	0,002	0,935	0,943	0,068	0,003	0,062	0,074
	CNN2	0,923	0,003	0,917	0,928	0,091	0,004	0,083	0,099
	SWIN	0,920	0,006	0,908	0,932	0,097	0,010	0,079	0,116

External	GAN	0,883	0,002	0,879	0,887	0,202	0,028	0,146	0,257
	CNN1	0,951	0,002	0,948	0,954	0,089	0,012	0,065	0,113
	CNN2	0,950	0,004	0,941	0,959	0,067	0,007	0,052	0,081
	SWIN	0,954	0,003	0,948	0,959	0,063	0,006	0,052	0,074

Note.—Mean DSC and BAHD Swin-UNETR performance with 95% CIs for the Internal and External datasets.
DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance.

Table E7

Models Cross-validation – Models DSC and BAHD Performance Metrics on Each Gestational Weeks (GW) as Mean DSC, Standard Error and 95% CI on Internal Test-Set

GW	DSC					BAHD			
	Model	Mean	Std. Error	95% CI		Mean	Std. Error	95% CI	
				Lower	Upper			Lower	Upper
21	SeGAN	0,845	0,000	0,845	0,845	0,184	0,000	0,184	0,184
	CNN1	0,931	0,000	0,931	0,931	0,074	0,000	0,074	0,074
	CNN2	0,882	0,000	0,882	0,882	0,139	0,000	0,139	0,139
	SWIN	0,725	0,000	0,725	0,725	0,416	0,000	0,416	0,416
22	SeGAN	0,805	0,024	0,758	0,852	0,353	0,082	0,191	0,514
	CNN1	0,937	0,008	0,921	0,954	0,067	0,011	0,045	0,089
	CNN2	0,929	0,009	0,912	0,947	0,079	0,012	0,055	0,103
	SWIN	0,830	0,003	0,825	0,836	0,249	0,001	0,247	0,250
23	SeGAN	0,870	0,000	0,870	0,870	0,140	0,000	0,140	0,140
	CNN1	0,938	0,000	0,938	0,938	0,064	0,000	0,064	0,064
	CNN2	0,922	0,000	0,922	0,922	0,087	0,000	0,087	0,087
	SWIN	0,922	0,000	0,922	0,922	0,081	0,000	0,081	0,081
24	SeGAN	0,799	0,009	0,782	0,816	0,447	0,034	0,379	0,514
	CNN1	0,926	0,006	0,915	0,937	0,083	0,008	0,068	0,098
	CNN2	0,915	0,009	0,897	0,934	0,102	0,015	0,073	0,132
	SWIN	0,884	0,004	0,876	0,893	0,142	0,007	0,129	0,156
25	SeGAN	0,846	0,004	0,838	0,855	0,182	0,009	0,165	0,199
	CNN1	0,940	0,003	0,934	0,946	0,064	0,004	0,056	0,071
	CNN2	0,931	0,004	0,922	0,939	0,075	0,006	0,064	0,086
	SWIN	0,880	0,003	0,874	0,885	0,147	0,005	0,137	0,157
26	SeGAN	0,858	0,003	0,852	0,864	0,162	0,007	0,149	0,176
	CNN1	0,945	0,002	0,941	0,948	0,057	0,002	0,053	0,060
	CNN2	0,943	0,001	0,941	0,945	0,061	0,001	0,058	0,063
	SWIN	0,922	0,006	0,910	0,933	0,088	0,008	0,072	0,105
27	SeGAN	0,902	0,004	0,895	0,910	0,106	0,005	0,096	0,115
	CNN1	0,947	0,002	0,944	0,950	0,056	0,002	0,052	0,061
	CNN2	0,939	0,002	0,935	0,943	0,066	0,003	0,060	0,072
	SWIN	0,945	0,002	0,941	0,949	0,056	0,002	0,053	0,060
28	SeGAN	0,883	0,013	0,857	0,909	0,369	0,102	0,168	0,570
	CNN1	0,911	0,004	0,904	0,919	0,108	0,006	0,095	0,120
	CNN2	0,877	0,004	0,869	0,884	0,160	0,007	0,145	0,175
	SWIN	0,937	0,001	0,934	0,939	0,068	0,002	0,064	0,072

29	SeGAN	0,901	0,002	0,898	0,905	0,105	0,004	0,096	0,113
	CNN1	0,938	0,002	0,934	0,942	0,068	0,003	0,063	0,073
	CNN2	0,903	0,001	0,900	0,905	0,117	0,002	0,114	0,120
	SWIN	0,922	0,007	0,908	0,935	0,091	0,010	0,072	0,111
30	SeGAN	0,894	0,002	0,891	0,897	0,117	0,002	0,113	0,121
	CNN1	0,957	0,001	0,955	0,959	0,044	0,001	0,042	0,047
	CNN2	0,939	0,001	0,936	0,942	0,066	0,002	0,062	0,070
	SWIN	0,945	0,002	0,941	0,949	0,056	0,002	0,051	0,061
31	SeGAN	0,932	0,004	0,923	0,940	0,076	0,005	0,066	0,086
	CNN1	0,919	0,001	0,917	0,921	0,105	0,003	0,100	0,110
	CNN2	0,920	0,002	0,916	0,924	0,103	0,004	0,096	0,110
	SWIN	0,960	0,003	0,955	0,965	0,045	0,003	0,040	0,051
32	SeGAN	0,904	0,003	0,899	0,909	0,105	0,004	0,096	0,114
	CNN1	0,949	0,003	0,944	0,954	0,055	0,003	0,048	0,061
	CNN2	0,933	0,002	0,929	0,937	0,077	0,003	0,071	0,083
	SWIN	0,927	0,004	0,918	0,935	0,086	0,008	0,070	0,102
33	SeGAN	0,923	0,003	0,917	0,929	0,086	0,004	0,078	0,093
	CNN1	0,951	0,000	0,950	0,951	0,053	0,000	0,053	0,054
	CNN2	0,930	0,002	0,927	0,934	0,081	0,002	0,076	0,086
	SWIN	0,927	0,002	0,924	0,931	0,091	0,004	0,085	0,098
34	SeGAN	0,912	0,001	0,910	0,914	0,097	0,002	0,094	0,100
	CNN1	0,962	0,001	0,961	0,963	0,039	0,001	0,038	0,040
	CNN2	0,948	0,001	0,946	0,949	0,057	0,001	0,056	0,059
	SWIN	0,955	0,003	0,950	0,961	0,049	0,004	0,042	0,056
35	SeGAN	0,922	0,003	0,916	0,928	0,084	0,005	0,075	0,093
	CNN1	0,947	0,002	0,943	0,951	0,058	0,003	0,053	0,063
	CNN2	0,933	0,003	0,926	0,939	0,078	0,005	0,068	0,088
	SWIN	0,942	0,011	0,920	0,963	0,068	0,017	0,034	0,102
36	SeGAN	0,918	0,002	0,914	0,922	0,094	0,003	0,088	0,099
	CNN1	0,930	0,001	0,928	0,932	0,080	0,001	0,077	0,083
	CNN2	0,915	0,001	0,912	0,917	0,101	0,002	0,096	0,105
	SWIN	0,922	0,002	0,919	0,925	0,094	0,002	0,090	0,098
37	SeGAN	0,938	0,001	0,936	0,940	0,066	0,000	0,065	0,067
	CNN1	0,934	0,003	0,929	0,939	0,082	0,004	0,074	0,091
	CNN2	0,913	0,004	0,905	0,920	0,114	0,006	0,102	0,126
	SWIN	0,936	0,001	0,934	0,939	0,078	0,002	0,074	0,082

Note.—The table reports mean, standard error and 95% CI DSC and BAHD score for the Swin UNETR best-model, gold-standard and state-of-the-art models for all gestational weeks for the internal test set.

Table E8

Models Cross-validation-models DSC and BAHD Performance Metrics on Each Gestational Weeks (GW) as Mean DSC, Standard Error and 95% CI on External Test Set

GW	DSC					BAHD			
	Model	Mean	Std. Error	95% CI		Mean	Std. Error	95% CI	
				Lower	Upper			Lower	Upper

21	SeGAN	0,021	0,000	0,021	0,021	13,341	0,000	13,341	13,341
	CNN1	0,380	0,000	0,380	0,380	3,795	0,000	3,795	3,795
	CNN2	0,828	0,000	0,828	0,828	0,385	0,000	0,385	0,385
	SWIN	0,758	0,000	0,758	0,758	1,322	0,000	1,322	1,322
22	SeGAN	0,139	0,034	0,073	0,205	8,086	0,985	6,155	10,017
	CNN1	0,428	0,041	0,348	0,509	8,283	3,305	1,806	14,761
	CNN2	0,807	0,008	0,792	0,823	0,538	0,123	0,297	0,780
	SWIN	0,743	0,032	0,679	0,806	0,948	0,265	0,428	1,467
23	SeGAN	0,091	0,021	0,050	0,133	6,103	0,588	4,951	7,255
	CNN1	0,820	0,022	0,777	0,863	0,287	0,058	0,173	0,401
	CNN2	0,865	0,004	0,858	0,872	0,190	0,010	0,170	0,211
	SWIN	0,735	0,027	0,683	0,787	2,883	0,397	2,104	3,662
24	SeGAN	0,206	0,000	0,206	0,206	7,582	0,000	7,582	7,582
	CNN1	0,715	0,000	0,715	0,715	0,598	0,000	0,598	0,598
	CNN2	0,793	0,000	0,793	0,793	0,770	0,000	0,770	0,770
	SWIN	0,575	0,000	0,575	0,575	5,906	0,000	5,906	5,906
25	SeGAN	0,197	0,000	0,197	0,197	3,684	0,000	3,684	3,684
	CNN1	0,627	0,000	0,627	0,627	2,469	0,000	2,469	2,469
	CNN2	0,908	0,000	0,908	0,908	0,109	0,000	0,109	0,109
	SWIN	0,806	0,000	0,806	0,806	0,427	0,000	0,427	0,427
26	SeGAN	0,220	0,140	-0,054	0,495	4,432	2,063	0,388	8,477
	CNN1	0,849	0,019	0,812	0,887	0,247	0,034	0,181	0,314
	CNN2	0,734	0,002	0,731	0,737	1,395	0,074	1,250	1,539
	SWIN	0,632	0,020	0,593	0,671	1,755	0,081	1,597	1,914
27	SeGAN	0,129	0,051	0,029	0,228	10,414	2,540	5,435	15,393
	CNN1	0,871	0,004	0,864	0,878	0,209	0,021	0,168	0,250
	CNN2	0,880	0,007	0,867	0,893	0,184	0,039	0,107	0,261
	SWIN	0,790	0,029	0,734	0,846	0,469	0,150	0,174	0,764
28	SeGAN	0,127	0,036	0,056	0,197	6,391	0,553	5,306	7,475
	CNN1	0,910	0,003	0,904	0,915	0,117	0,006	0,105	0,130
	CNN2	0,922	0,006	0,910	0,935	0,214	0,088	0,042	0,386
	SWIN	0,702	0,030	0,643	0,761	3,793	0,500	2,813	4,772
29	SeGAN	0,109	0,047	0,016	0,202	8,995	1,603	5,853	12,138
	CNN1	0,074	0,024	0,027	0,120	45,939	11,525	23,352	68,527
	CNN2	0,586	0,012	0,563	0,609	1,763	0,088	1,590	1,937
	SWIN	0,892	0,011	0,869	0,914	0,439	0,106	0,230	0,647
30	SeGAN	0,288	0,057	0,177	0,400	9,660	6,290	-2,668	21,988
	CNN1	0,859	0,007	0,846	0,872	0,268	0,031	0,207	0,330
	CNN2	0,913	0,003	0,908	0,918	0,103	0,004	0,095	0,111
	SWIN	0,871	0,009	0,853	0,888	0,181	0,020	0,142	0,220
31	SeGAN	0,152	0,073	0,009	0,294	9,565	2,982	3,720	15,410
	CNN1	0,608	0,006	0,597	0,620	1,826	0,102	1,627	2,025
	CNN2	0,905	0,002	0,901	0,910	0,115	0,003	0,109	0,121
	SWIN	0,770	0,011	0,748	0,792	0,636	0,081	0,477	0,795
32	SeGAN	0,058	0,033	-0,006	0,122	8,677	2,218	4,331	13,023
	CNN1	0,743	0,005	0,733	0,752	0,753	0,013	0,729	0,778
	CNN2	0,871	0,001	0,868	0,874	0,169	0,000	0,169	0,170
	SWIN	0,803	0,023	0,759	0,847	1,185	0,324	0,549	1,821

33	SeGAN	0,894	0,004	0,887	0,901	0,179	0,011	0,158	0,200
	CNN1	0,032	0,009	0,015	0,050	90,104	33,098	25,234	154,974
	CNN2	0,755	0,018	0,721	0,789	0,539	0,073	0,396	0,682
	SWIN	0,750	0,013	0,724	0,775	0,733	0,094	0,547	0,918
34	SeGAN	0,222	0,098	0,029	0,414	88,535	69,815	-48,299	225,370
	CNN1	0,262	0,071	0,122	0,402	25,801	13,914	-1,471	53,073
	CNN2	0,872	0,004	0,865	0,880	0,355	0,059	0,239	0,471
	SWIN	0,761	0,070	0,624	0,898	1,625	0,809	0,040	3,211
35	SeGAN	0,012	0,001	0,010	0,014	61,112	37,323	-12,040	134,264
	CNN1	0,735	0,028	0,680	0,791	0,974	0,170	0,641	1,306
	CNN2	0,927	0,001	0,925	0,930	0,681	0,054	0,575	0,787
	SWIN	0,838	0,007	0,823	0,852	0,922	0,171	0,587	1,257
36	SeGAN	0,066	0,017	0,032	0,101	29,351	18,390	-6,692	65,394
	CNN1	0,584	0,032	0,522	0,646	4,264	0,504	3,277	5,252
	CNN2	0,915	0,003	0,908	0,922	0,844	0,069	0,710	0,979
	SWIN	0,917	0,002	0,913	0,921	0,671	0,104	0,468	0,874
37	SeGAN	0,153	0,055	0,044	0,261	31,795	22,652	-12,602	76,191
	CNN1	0,930	0,011	0,909	0,952	0,110	0,028	0,055	0,166
	CNN2	0,945	0,004	0,937	0,953	0,065	0,008	0,050	0,080
	SWIN	0,593	0,012	0,570	0,616	2,954	0,043	2,870	3,037

Note.—The table reports mean, standard error and 95% CI DSC and BAHD score for the Swin UNETR best-model, gold-standard and state-of-the-art models for all gestational weeks when tested on the external test set.

Table E9

Generalized Estimating Equations (GEE) Models' Results for DSC and BAHD for Model Effect

	DSC					BAHD				
	Mean Diff.	Std. Error	Seq. Bonf. Sig.	Wald 95% CI for Difference		Mean Diff.	Std. Error	Seq. Bonf. Sig.	Wald 95% CI for Difference	
				Lower	Upper				Lower	Upper
GEE 2c										
UNT*100%*P-SWIN*100%*P	-0,064*	0,013	0	-0,105	-0,02	0,517*	0,121	0,001	0,123	0,911
UNT*100%*S-SWIN*100%*S	-0,055*	0,009	0	-0,085	-0,03	0,16	0,093	1	-0,105	0,424
UNT*66%*P-SWIN*66%*P	-0,064*	0,007	0	-0,089	-0,04	0,495*	0,085	0	0,211	0,778
UNT*66%*S-SWIN*66%*S	-0,093*	0,012	0	-0,133	-0,05	0,696*	0,133	0	0,259	1,134
UNT*33%*P-SWIN*33%*P	-0,088*	0,007	0	-0,112	-0,06	0,572*	0,083	0	0,295	0,849
UNT*33%*S-SWIN*33%*S	-0,068*	0,005	0	-0,084	-0,05	0,285*	0,066	0,001	0,068	0,502
UNT*100%*P-UNT*100%*S	0,004	0,01	1	-0,016	0,025	-0,046	0,096	1	-0,253	0,16
UNT*66%*P-UNT*66%*S	0,023	0,012	0,956	-0,013	0,059	-0,27	0,116	0,381	-0,62	0,079
UNT*33%*P-UNT*33%*S	-0,006	0,003	1	-0,016	0,004	0,041	0,032	1	-0,041	0,123
SWIN*100%*P-SWIN*100%*S	0,014	0,006	0,375	-0,004	0,032	-0,404*	0,075	0	-0,651	-0,156
SWIN*66%*P-SWIN*66%*S	-0,006	0,008	1	-0,025	0,012	-0,068	0,086	1	-0,267	0,13
SWIN*33%*P-SWIN*33%*S	0,014	0,008	0,956	-0,008	0,037	-0,246	0,084	0,096	-0,508	0,016
UNT*P*100%-UNT*P*66%	0,015	0,014	1	-0,018	0,049	-0,152	0,127	1	-0,473	0,169
UNT*P*100%-UNT*P*33%	0,036	0,013	0,165	-0,005	0,078	-0,311	0,123	0,288	-0,692	0,069
UNT*P*66%-UNT*P*33%	0,020*	0,004	0	0,009	0,033	-0,160*	0,047	0,024	-0,309	-0,01

SWIN*P*100%-SWIN*P*66%	0,016	0,006	0,2	-0,003	0,035	-0,174	0,075	0,381	-0,4	0,052
SWIN*P*100%-SWIN*P*33%	0,013	0,006	0,422	-0,004	0,03	-0,257*	0,062	0,001	-0,456	-0,057
SWIN*P*66%-SWIN*P*33%	-0,003	0,009	1	-0,021	0,015	-0,082	0,1	1	-0,313	0,149
UNT*S*100%-UNT*S*66%	0,034*	0,009	0,008	0,005	0,063	-0,376*	0,11	0,02	-0,724	-0,03
UNT*S*100%-UNT*S*33%	0,026	0,009	0,071	-0,001	0,053	-0,224	0,093	0,345	-0,508	0,06
UNT*S*66%-UNT*S*33%	-0,008	0,013	1	-0,037	0,021	0,152	0,112	1	-0,141	0,445
SWIN*S*100%-SWIN*S*66%	-0,004	0,009	1	-0,023	0,014	0,161	0,089	1	-0,101	0,423
SWIN*S*100%-SWIN*S*33%	0,013	0,006	0,553	-0,005	0,032	-0,099	0,085	1	-0,312	0,114
SWIN*S*66%-SWIN*S*33%	0,018	0,009	0,779	-0,009	0,044	-0,26	0,094	0,156	-0,553	0,033
GEE 3c										
0.5-0.75	-0,137*	0,016	0	-0,186	-0,087	1,153*	0,218	0	0,5	1,153*
0.5-1	-0,099*	0,013	0	-0,139	-0,06	0,983*	0,21	0	0,366	0,983*
0.5-2	-0,155*	0,019	0	-0,213	-0,098	1,397*	0,237	0	0,677	1,397*
0.5-mean image size	-0,164*	0,019	0	-0,224	-0,105	1,417*	0,238	0	0,691	1,417*
0.5-3	-0,134*	0,019	0	-0,19	-0,078	1,365*	0,238	0	0,646	1,365*
0.5-4	-0,097*	0,018	0	-0,15	-0,044	1,290*	0,237	0	0,575	1,290*
0.5-5	-0,074*	0,017	0	-0,121	-0,026	1,221*	0,235	0	0,523	1,221*
0.75-1	0,037*	0,008	0	0,014	0,06	-0,171	0,065	0,065	-0,347	-0,171
0.75-2	-0,019*	0,007	0,031	-0,036	-0,001	0,244	0,092	0,065	-0,008	0,244
0.75-mean image size	-0,028*	0,007	0,001	-0,047	-0,009	0,264*	0,093	0,048	0,001	0,264*
0.75-3	0,003	0,007	1	-0,013	0,019	0,211	0,093	0,095	-0,022	0,211
0.75-4	0,040*	0,007	0	0,02	0,059	0,136	0,092	0,278	-0,07	0,136
0.75-5	0,063*	0,008	0	0,04	0,086	0,068	0,096	0,478	-0,12	0,068
1-2	-0,056*	0,011	0	-0,087	-0,024	0,415*	0,11	0,002	0,098	0,415*
1-mean image size	-0,065*	0,012	0	-0,1	-0,03	0,435*	0,112	0,002	0,107	0,435*
1-3	-0,034*	0,012	0,022	-0,066	-0,003	0,382*	0,112	0,008	0,06	0,382*
1-4	0,002	0,011	1	-0,02	0,025	0,307*	0,11	0,049	0,001	0,307*
1-5	0,026	0,011	0,058	-0,001	0,052	0,239	0,11	0,095	-0,027	0,239
2-mean image size	-0,009*	0,003	0,019	-0,017	-0,001	0,02	0,008	0,065	-0,001	0,02
2-3	0,022*	0,004	0	0,01	0,033	-0,033*	0,01	0,01	-0,06	-0,033*
2-4	0,058*	0,004	0	0,045	0,071	-0,108*	0,012	0	-0,146	-0,108*
2-5	0,082*	0,006	0	0,063	0,101	-0,176*	0,028	0	-0,261	-0,176*
mean image size-3	0,030*	0,003	0	0,021	0,04	-0,052*	0,007	0	-0,073	-0,052*
mean image size-4	0,067*	0,004	0	0,054	0,081	-0,128*	0,012	0	-0,166	-0,128*
mean image size-5	0,091*	0,007	0	0,069	0,112	-0,196*	0,029	0	-0,286	-0,196*
3-4	0,037*	0,004	0	0,024	0,049	-0,075*	0,011	0	-0,108	-0,075*
3-5	0,060*	0,007	0	0,039	0,081	-0,143*	0,029	0	-0,229	-0,143*
4-5	0,023*	0,006	0,001	0,007	0,039	-0,068	0,027	0,065	-0,139	-0,068
GEE 4c										
1.5T*GAN-1.5T*CNN1	-0,043*	0,008	0	-0,066	-0,02	0,072*	0,022	0,02	0,006	0,139
1.5T*GAN-1.5T*CNN2	-0,022	0,008	0,086	-0,046	0,001	0,043	0,023	0,685	-0,022	0,109
1.5T*GAN-1.5T*SWIN	-0,02	0,008	0,142	-0,042	0,003	0,037	0,022	0,915	-0,025	0,099
1.5T*CNN1-1.5T*CNN2	0,021*	0,002	0	0,014	0,027	-0,029*	0,003	0	-0,038	-0,02
1.5T*CNN1-1.5T*SWIN	0,023*	0,008	0,048	0	0,046	-0,035	0,013	0,113	-0,074	0,003
1.5T*CNN2-1.5T*SWIN	0,002	0,007	1	-0,012	0,017	-0,006	0,012	1	-0,032	0,019
3T*GAN-3T*CNN1	-0,057*	0,01	0	-0,088	-0,03	0,111*	0,027	0,001	0,026	0,196
3T*GAN-3T*CNN2	-0,046*	0,01	0	-0,077	-0,02	0,096*	0,026	0,005	0,016	0,175
3T*GAN-3T*SWIN	-0,043*	0,004	0	-0,055	-0,03	0,089*	0,026	0,011	0,011	0,168

3T*CNN1-3T*CNN2	0,011*	0,002	0	0,004	0,019	-0,016*	0,004	0	-0,027	-0,004
3T*CNN1-3T*SWIN	0,015	0,009	1	-0,011	0,041	-0,022	0,015	1	-0,063	0,019
3T*CNN2-3T*SWIN	0,003	0,01	1	-0,018	0,025	-0,006	0,016	1	-0,04	0,028
GEE 5c										
IN*GAN-IN*CNN1	-0,049*	0,006	0	-0,068	-0,03	0,090*	0,017	0	0,036	0,144
IN*GAN-IN*CNN2	-0,033*	0,007	0	-0,052	-0,01	0,067*	0,018	0,002	0,014	0,119
IN*GAN-IN*SWIN	-0,030*	0,005	0	-0,045	-0,02	0,061*	0,017	0,007	0,01	0,111
IN*CNN1-IN*CNN2	0,017*	0,002	0	0,011	0,022	-0,023*	0,002	0	-0,031	-0,015
IN*CNN1-IN*SWIN	0,019*	0,006	0,01	0,003	0,036	-0,029*	0,01	0,037	-0,058	-0,001
IN*CNN2-IN*SWIN	0,003	0,006	1	-0,01	0,016	-0,006	0,01	1	-0,028	0,015
EX*GAN-EX*CNN1	-0,068*	0,002	0	-0,074	-0,06	0,112*	0,016	0	0,062	0,162
EX*GAN-EX*CNN2	-0,068*	0,004	0	-0,081	-0,05	0,135*	0,028	0	0,048	0,222
EX*GAN-EX*SWIN	-0,071*	0,003	0	-0,08	-0,06	0,139*	0,028	0	0,051	0,226
EX*CNN1-EX*CNN2	0,001	0,005	1	-0,009	0,01	0,023	0,014	0,793	-0,015	0,06
EX*CNN1-EX*SWIN	-0,003	0,003	1	-0,009	0,004	0,026	0,013	0,445	-0,01	0,063
EX*CNN2-EX*SWIN	-0,003	0,003	1	-0,011	0,004	0,004	0,004	1	-0,006	0,014
GEE 6c										
21*IN_TS*GAN-21*IN_TS*CNN1	-0,086	0	NA	NA	NA	0,11	0	NA	NA	NA
21*IN_TS*GAN-21*IN_TS*CNN2	-0,038	0	NA	NA	NA	0,045	0	NA	NA	NA
21*IN_TS*GAN-21*IN_TS*SWIN	0,12	0	NA	NA	NA	-0,231	0	NA	NA	NA
21*IN_TS*CNN1-21*IN_TS*CNN2	0,048	0	NA	NA	NA	-0,065	0	NA	NA	NA
21*IN_TS*CNN1-21*IN_TS*SWIN	0,205	0	NA	NA	NA	-0,342	0	NA	NA	NA
21*IN_TS*CNN2-21*IN_TS*SWIN	0,157	0	NA	NA	NA	-0,276	0	NA	NA	NA
21*EX_TS*GAN-21*EX_TS*CNN1	-0,359	0	NA	NA	NA	9,545	0	NA	NA	NA
21*EX_TS*GAN-21*EX_TS*CNN2	-0,807	0	NA	NA	NA	12,956	0	NA	NA	NA
21*EX_TS*GAN-21*EX_TS*SWIN	-0,737	0	NA	NA	NA	12,019	0	NA	NA	NA
21*EX_TS*CNN1-21*EX_TS*CNN2	-0,448	0	NA	NA	NA	3,411	0	NA	NA	NA
21*EX_TS*CNN1-21*EX_TS*SWIN	-0,379	0	NA	NA	NA	2,473	0	NA	NA	NA
21*EX_TS*CNN2-21*EX_TS*SWIN	0,069	0	NA	NA	NA	-0,937	0	0	-0,937	-0,937
22*IN_TS*GAN-22*IN_TS*CNN1	-0,132	0,032	0,095	-0,269	0,005	0,286	0,094	1	-0,075	0,647
22*IN_TS*GAN-22*IN_TS*CNN2	-0,124	0,033	0,34	-0,263	0,015	0,274	0,095	1	-0,08	0,627
22*IN_TS*GAN-22*IN_TS*SWIN	-0,025	0,021	1	-0,078	0,028	0,104	0,083	1	-0,109	0,316
22*IN_TS*CNN1-22*IN_TS*CNN2	0,008	0,001	0	0,006	0,011	-0,012	0,001	0	-0,018	-0,006
22*IN_TS*CNN1-22*IN_TS*SWIN	0,107*	0,011	0	0,056	0,159	-0,182*	0,01	0	-0,23	-0,135
22*IN_TS*CNN2-22*IN_TS*SWIN	0,099*	0,012	0	0,047	0,151	-0,17*	0,012	0	-0,223	-0,116
22*EX_TS*GAN-22*EX_TS*CNN1	-0,289*	0,037	0	-0,452	-0,13	-0,197	3,331	1	-6,794	6,4
22*EX_TS*GAN-22*EX_TS*CNN2	-0,668*	0,04	0	-0,848	-0,49	7,548*	1,093	0	2,696	12,399
22*EX_TS*GAN-22*EX_TS*SWIN	-0,604*	0,056	0	-0,858	-0,35	7,139*	1,194	0	1,858	12,419
22*EX_TS*CNN1-22*EX_TS*CNN2	-0,379*	0,046	0	-0,58	-0,18	7,745	3,351	1	-3,243	18,733
22*EX_TS*CNN1-22*EX_TS*SWIN	-0,314	0,066	0,004	-0,594	-0,03	7,336	3,415	1	-3,452	18,123
22*EX_TS*CNN2-22*EX_TS*SWIN	0,065	0,028	1	-0,027	0,156	-0,409	0,184	1	-1,001	0,182
23*IN_TS*GAN-23*IN_TS*CNN1	-0,068	0	NA	NA	NA	0,076	0	NA	NA	NA
23*IN_TS*GAN-23*IN_TS*CNN2	-0,052	0	NA	NA	NA	0,053	0	NA	NA	NA
23*IN_TS*GAN-23*IN_TS*SWIN	-0,052	0	NA	NA	NA	0,059	0	NA	NA	NA
23*IN_TS*CNN1-23*IN_TS*CNN2	0,017	0	NA	NA	NA	-0,023	0	NA	NA	NA
23*IN_TS*CNN1-23*IN_TS*SWIN	0,016	0	NA	NA	NA	-0,016	0	NA	NA	NA
23*IN_TS*CNN2-23*IN_TS*SWIN	0	0	NA	NA	NA	0,007	0	NA	NA	NA
23*EX_TS*GAN-23*EX_TS*CNN1	-0,729*	0,019	0	-0,815	-0,64	5,816*	0,591	0	3,142	8,49

23*EX_TS*GAN-23*EX_TS*CNN2	-0,774*	0,018	0	-0,856	-0,69	5,913*	0,588	0	3,254	8,571
23*EX_TS*GAN-23*EX_TS*SWIN	-0,643*	0,038	0	-0,816	-0,47	3,22	0,882	0,957	-0,616	7,056
23*EX_TS*CNN1-23*EX_TS*CNN2	-0,045	0,019	1	-0,107	0,018	0,097	0,057	1	-0,066	0,259
23*EX_TS*CNN1-23*EX_TS*SWIN	0,085	0,031	1	-0,026	0,197	-2,596	0,389	0	-4,323	-0,869
23*EX_TS*CNN2-23*EX_TS*SWIN	0,13*	0,026	0,002	0,019	0,241	-2,693*	0,401	0	-4,473	-0,912
24*IN_TS*GAN-24*IN_TS*CNN1	-0,127*	0,014	0	-0,19	-0,07	0,364*	0,041	0	0,181	0,547
24*IN_TS*GAN-24*IN_TS*CNN2	-0,116*	0,018	0	-0,195	-0,04	0,344*	0,05	0	0,124	0,565
24*IN_TS*GAN-24*IN_TS*SWIN	-0,085*	0,012	0	-0,137	-0,03	0,304*	0,041	0	0,123	0,486
24*IN_TS*CNN1-24*IN_TS*CNN2	0,011	0,006	1	-0,007	0,029	-0,019	0,011	1	-0,051	0,012
24*IN_TS*CNN1-24*IN_TS*SWIN	0,042*	0,007	0	0,013	0,071	-0,059*	0,009	0	-0,098	-0,021
24*IN_TS*CNN2-24*IN_TS*SWIN	0,031	0,007	0,02	0,001	0,061	-0,04	0,01	0,19	-0,083	0,003
24*EX_TS*GAN-24*EX_TS*CNN1	-0,509	0	NA	NA	NA	6,984	0	0	6,984	6,984
24*EX_TS*GAN-24*EX_TS*CNN2	-0,588	0	NA	NA	NA	6,812	0	NA	NA	NA
24*EX_TS*GAN-24*EX_TS*SWIN	-0,369	0	NA	NA	NA	1,676	0	NA	NA	NA
24*EX_TS*CNN1-24*EX_TS*CNN2	-0,078	0	NA	NA	NA	-0,171	0	NA	NA	NA
24*EX_TS*CNN1-24*EX_TS*SWIN	0,14	0	NA	NA	NA	-5,308	0	NA	NA	NA
24*EX_TS*CNN2-24*EX_TS*SWIN	0,219	0	NA	NA	NA	-5,136	0	NA	NA	NA
25*IN_TS*GAN-25*IN_TS*CNN1	-0,094*	0,006	0	-0,121	-0,07	0,118*	0,011	0	0,067	0,169
25*IN_TS*GAN-25*IN_TS*CNN2	-0,084*	0,007	0	-0,117	-0,05	0,107*	0,013	0	0,05	0,164
25*IN_TS*GAN-25*IN_TS*SWIN	-0,033*	0,006	0	-0,059	-0,01	0,035	0,012	1	-0,01	0,079
25*IN_TS*CNN1-25*IN_TS*CNN2	0,009*	0,001	0	0,003	0,015	-0,011*	0,002	0	-0,02	-0,002
25*IN_TS*CNN1-25*IN_TS*SWIN	0,06*	0,005	0	0,036	0,084	-0,084*	0,007	0	-0,115	-0,052
25*IN_TS*CNN2-25*IN_TS*SWIN	0,051*	0,007	0	0,022	0,079	-0,072*	0,009	0	-0,112	-0,033
25*EX_TS*GAN-25*EX_TS*CNN1	-0,43	0	NA	NA	NA	1,216	0	NA	NA	NA
25*EX_TS*GAN-25*EX_TS*CNN2	-0,711	0	NA	NA	NA	3,576	0	NA	NA	NA
25*EX_TS*GAN-25*EX_TS*SWIN	-0,61	0	NA	NA	NA	3,257	0	NA	NA	NA
25*EX_TS*CNN1-25*EX_TS*CNN2	-0,281	0	NA	NA	NA	2,36	0	NA	NA	NA
25*EX_TS*CNN1-25*EX_TS*SWIN	-0,179	0	NA	NA	NA	2,041	0	NA	NA	NA
25*EX_TS*CNN2-25*EX_TS*SWIN	0,101	0	NA	NA	NA	-0,319	0	NA	NA	NA
26*IN_TS*GAN-26*IN_TS*CNN1	-0,087*	0,005	0	-0,109	-0,07	0,106*	0,009	0	0,066	0,146
26*IN_TS*GAN-26*IN_TS*CNN2	-0,085*	0,002	0	-0,096	-0,07	0,102*	0,006	0	0,075	0,128
26*IN_TS*GAN-26*IN_TS*SWIN	-0,064*	0,009	0	-0,101	-0,03	0,074*	0,015	0,008	0,006	0,142
26*IN_TS*CNN1-26*IN_TS*CNN2	0,002	0,003	1	-0,004	0,008	-0,004	0,003	1	-0,012	0,004
26*IN_TS*CNN1-26*IN_TS*SWIN	0,023*	0,004	0	0,006	0,04	-0,032*	0,007	0,007	-0,061	-0,003
26*IN_TS*CNN2-26*IN_TS*SWIN	0,021	0,007	1	-0,005	0,047	-0,028	0,01	1	-0,064	0,008
26*EX_TS*GAN-26*EX_TS*CNN1	-0,629*	0,121	0,001	-1,149	-0,11	4,185	2,03	1	-2,099	10,469
26*EX_TS*GAN-26*EX_TS*CNN2	-0,513	0,139	0,439	-1,098	0,072	3,038	2,137	1	-2,652	8,728
26*EX_TS*GAN-26*EX_TS*SWIN	-0,412	0,16	1	-0,968	0,145	2,677	2,144	1	-2,803	8,157
26*EX_TS*CNN1-26*EX_TS*CNN2	0,116*	0,017	0	0,04	0,192	-1,147*	0,108	0	-1,631	-0,664
26*EX_TS*CNN1-26*EX_TS*SWIN	0,217*	0,039	0	0,05	0,385	-1,508*	0,115	0	-2,024	-0,993
26*EX_TS*CNN2-26*EX_TS*SWIN	0,102*	0,021	0,005	0,01	0,193	-0,361*	0,007	0	-0,393	-0,329
27*IN_TS*GAN-27*IN_TS*CNN1	-0,045*	0,006	0	-0,069	-0,02	0,05*	0,007	0	0,018	0,081
27*IN_TS*GAN-27*IN_TS*CNN2	-0,037*	0,006	0	-0,063	-0,01	0,04*	0,008	0,002	0,005	0,074
27*IN_TS*GAN-27*IN_TS*SWIN	-0,043*	0,003	0	-0,058	-0,03	0,049*	0,004	0	0,031	0,068
27*IN_TS*CNN1-27*IN_TS*CNN2	0,008*	0,001	0	0,005	0,011	-0,01*	0,001	0	-0,015	-0,005
27*IN_TS*CNN1-27*IN_TS*SWIN	0,002	0,003	1	-0,006	0,009	0	0,004	1	-0,008	0,008
27*IN_TS*CNN2-27*IN_TS*SWIN	-0,006	0,004	1	-0,016	0,004	0,01	0,004	1	-0,004	0,023
27*EX_TS*GAN-27*EX_TS*CNN1	-0,742*	0,049	0	-0,961	-0,52	10,205	2,538	0,237	-0,894	21,304

27*EX_TS*GAN-27*EX_TS*CNN2	-0,751*	0,049	0	-0,97	-0,53	10,23	2,548	0,243	-0,912	21,372
27*EX_TS*GAN-27*EX_TS*SWIN	-0,661*	0,053	0	-0,898	-0,42	9,945	2,418	0,163	-0,641	20,532
27*EX_TS*CNN1-27*EX_TS*CNN2	-0,009	0,007	1	-0,027	0,009	0,025	0,045	1	-0,073	0,123
27*EX_TS*CNN1-27*EX_TS*SWIN	0,081	0,028	1	-0,023	0,185	-0,26	0,157	1	-0,701	0,181
27*EX_TS*CNN2-27*EX_TS*SWIN	0,09	0,031	1	-0,025	0,206	-0,285	0,162	1	-0,752	0,182
28*IN_TS*GAN-28*IN_TS*CNN1	-0,028	0,011	1	-0,067	0,01	0,261	0,097	1	-0,086	0,608
28*IN_TS*GAN-28*IN_TS*CNN2	0,007	0,012	1	-0,02	0,033	0,209	0,098	1	-0,1	0,518
28*IN_TS*GAN-28*IN_TS*SWIN	-0,053*	0,012	0,017	-0,104	-0	0,301	0,1	1	-0,082	0,684
28*IN_TS*CNN1-28*IN_TS*CNN2	0,035*	0,001	0	0,029	0,04	-0,052*	0,003	0	-0,064	-0,041
28*IN_TS*CNN1-28*IN_TS*SWIN	-0,025*	0,003	0	-0,04	-0,01	0,04*	0,005	0	0,019	0,061
28*IN_TS*CNN2-28*IN_TS*SWIN	-0,06*	0,004	0	-0,078	-0,04	0,092*	0,006	0	0,064	0,12
28*EX_TS*GAN-28*EX_TS*CNN1	-0,783*	0,035	0	-0,94	-0,63	6,274*	0,549	0	3,812	8,735
28*EX_TS*GAN-28*EX_TS*CNN2	-0,796*	0,038	0	-0,964	-0,63	6,177*	0,58	0	3,578	8,776
28*EX_TS*GAN-28*EX_TS*SWIN	-0,576*	0,035	0	-0,732	-0,42	2,598	0,654	0,28	-0,257	5,453
28*EX_TS*CNN1-28*EX_TS*CNN2	-0,012	0,005	1	-0,03	0,005	-0,097	0,088	1	-0,315	0,121
28*EX_TS*CNN1-28*EX_TS*SWIN	0,208*	0,029	0	0,083	0,332	-3,675*	0,496	0	-5,884	-1,467
28*EX_TS*CNN2-28*EX_TS*SWIN	0,22*	0,033	0	0,077	0,363	-3,579*	0,575	0	-6,126	-1,032
29*IN_TS*GAN-29*IN_TS*CNN1	-0,036*	0,002	0	-0,044	-0,03	0,037*	0,003	0	0,025	0,049
29*IN_TS*GAN-29*IN_TS*CNN2	-0,001	0,003	1	-0,008	0,005	-0,013	0,006	1	-0,031	0,006
29*IN_TS*GAN-29*IN_TS*SWIN	-0,02	0,005	0,339	-0,043	0,002	0,013	0,007	1	-0,007	0,033
29*IN_TS*CNN1-29*IN_TS*CNN2	0,035*	0,003	0	0,021	0,05	-0,049*	0,004	0	-0,068	-0,03
29*IN_TS*CNN1-29*IN_TS*SWIN	0,016	0,006	1	-0,005	0,038	-0,023	0,009	1	-0,055	0,008
29*IN_TS*CNN2-29*IN_TS*SWIN	-0,019	0,007	1	-0,044	0,007	0,026	0,011	1	-0,011	0,062
29*EX_TS*GAN-29*EX_TS*CNN1	0,035	0,045	1	-0,068	0,138	-36,944	12,506	1	-84,217	10,329
29*EX_TS*GAN-29*EX_TS*CNN2	-0,477*	0,053	0	-0,711	-0,24	7,232	1,669	0,063	-0,083	14,548
29*EX_TS*GAN-29*EX_TS*SWIN	-0,783*	0,058	0	-1,038	-0,53	8,557*	1,684	0,002	1,146	15,967
29*EX_TS*CNN1-29*EX_TS*CNN2	-0,512*	0,034	0	-0,662	-0,36	44,176	11,521	0,478	-6,024	94,376
29*EX_TS*CNN1-29*EX_TS*SWIN	-0,818*	0,032	0	-0,961	-0,68	45,501	11,435	0,278	-4,468	95,469
29*EX_TS*CNN2-29*EX_TS*SWIN	-0,306*	0,013	0	-0,363	-0,25	1,325*	0,101	0	0,873	1,777
30*IN_TS*GAN-30*IN_TS*CNN1	-0,063*	0,002	0	-0,073	-0,05	0,073*	0,003	0	0,06	0,085
30*IN_TS*GAN-30*IN_TS*CNN2	-0,045*	0,003	0	-0,058	-0,03	0,05*	0,004	0	0,034	0,067
30*IN_TS*GAN-30*IN_TS*SWIN	-0,052*	0,003	0	-0,066	-0,04	0,061*	0,004	0	0,042	0,08
30*IN_TS*CNN1-30*IN_TS*CNN2	0,018*	0,002	0	0,009	0,028	-0,022*	0,003	0	-0,033	-0,011
30*IN_TS*CNN1-30*IN_TS*SWIN	0,012	0,003	0,06	0	0,024	-0,012	0,003	0,603	-0,025	0,002
30*IN_TS*CNN2-30*IN_TS*SWIN	-0,006*	0,001	0	-0,009	-0	0,01*	0,001	0	0,008	0,013
30*EX_TS*GAN-30*EX_TS*CNN1	-0,571*	0,06	0	-0,835	-0,31	9,392	6,307	1	-7,671	26,455
30*EX_TS*GAN-30*EX_TS*CNN2	-0,624*	0,057	0	-0,877	-0,37	9,557	6,292	1	-7,585	26,699
30*EX_TS*GAN-30*EX_TS*SWIN	-0,582*	0,059	0	-0,845	-0,32	9,479	6,303	1	-7,633	26,59
30*EX_TS*CNN1-30*EX_TS*CNN2	-0,053*	0,006	0	-0,08	-0,03	0,165*	0,031	0	0,03	0,3
30*EX_TS*CNN1-30*EX_TS*SWIN	-0,011	0,008	1	-0,033	0,011	0,087	0,03	1	-0,025	0,199
30*EX_TS*CNN2-30*EX_TS*SWIN	0,042*	0,008	0	0,01	0,075	-0,078*	0,017	0,012	-0,151	-0,005
31*IN_TS*GAN-31*IN_TS*CNN1	0,012	0,004	1	-0,003	0,028	-0,029*	0,004	0	-0,044	-0,013
31*IN_TS*GAN-31*IN_TS*CNN2	0,012	0,004	1	-0,003	0,027	-0,027*	0,005	0	-0,047	-0,007
31*IN_TS*GAN-31*IN_TS*SWIN	-0,028*	0,002	0	-0,037	-0,02	0,031*	0,003	0	0,018	0,044
31*IN_TS*CNN1-31*IN_TS*CNN2	-0,001	0,001	1	-0,003	0,002	0,001	0,002	1	-0,002	0,005
31*IN_TS*CNN1-31*IN_TS*SWIN	-0,041*	0,002	0	-0,051	-0,03	0,059*	0,001	0	0,056	0,062
31*IN_TS*CNN2-31*IN_TS*SWIN	-0,04*	0,002	0	-0,05	-0,03	0,058*	0,002	0	0,048	0,067
31*EX_TS*GAN-31*EX_TS*CNN1	-0,457*	0,076	0	-0,787	-0,13	7,739	3,04	1	-2,772	18,251

31*EX_TS*GAN-31*EX_TS*CNN2	-0,754*	0,071	0	-1,069	-0,44	9,45	2,98	1	-2,342	21,243
31*EX_TS*GAN-31*EX_TS*SWIN	-0,618*	0,083	0	-0,981	-0,26	8,929	3,06	1	-2,549	20,407
31*EX_TS*CNN1-31*EX_TS*CNN2	-0,297*	0,008	0	-0,332	-0,26	1,711*	0,105	0	1,243	2,179
31*EX_TS*CNN1-31*EX_TS*SWIN	-0,161*	0,008	0	-0,196	-0,13	1,19*	0,083	0	0,819	1,561
31*EX_TS*CNN2-31*EX_TS*SWIN	0,136*	0,013	0	0,079	0,192	-0,521*	0,083	0	-0,89	-0,153
32*IN_TS*GAN-32*IN_TS*CNN1	-0,045*	0,004	0	-0,064	-0,03	0,05*	0,006	0	0,022	0,079
32*IN_TS*GAN-32*IN_TS*CNN2	-0,029	0,004	0	-0,045	-0,01	0,028	0,006	0,062	0	0,056
32*IN_TS*GAN-32*IN_TS*SWIN	-0,022*	0,002	0	-0,032	-0,01	0,019*	0,004	0,005	0,002	0,036
32*IN_TS*CNN1-32*IN_TS*CNN2	0,016*	0,001	0	0,013	0,019	-0,022*	0,001	0	-0,025	-0,02
32*IN_TS*CNN1-32*IN_TS*SWIN	0,022	0,005	0,068	0	0,045	-0,031	0,009	1	-0,069	0,006
32*IN_TS*CNN2-32*IN_TS*SWIN	0,006	0,005	1	-0,007	0,02	-0,009	0,009	1	-0,031	0,013
32*EX_TS*GAN-32*EX_TS*CNN1	-0,685*	0,037	0	-0,85	-0,52	7,924	2,205	1	-1,58	17,428
32*EX_TS*GAN-32*EX_TS*CNN2	-0,813*	0,034	0	-0,963	-0,66	8,508	2,217	0,474	-1,154	18,17
32*EX_TS*GAN-32*EX_TS*SWIN	-0,745*	0,01	0	-0,789	-0,7	7,492	1,893	0,3	-0,775	15,759
32*EX_TS*CNN1-32*EX_TS*CNN2	-0,128*	0,003	0	-0,144	-0,11	0,584*	0,012	0	0,529	0,639
32*EX_TS*CNN1-32*EX_TS*SWIN	-0,06	0,027	1	-0,148	0,027	-0,432	0,312	1	-1,255	0,391
32*EX_TS*CNN2-32*EX_TS*SWIN	0,068	0,024	1	-0,02	0,157	-1,016	0,324	1	-2,289	0,257
33*IN_TS*GAN-33*IN_TS*CNN1	-0,028*	0,003	0	-0,04	-0,02	0,032*	0,004	0	0,015	0,049
33*IN_TS*GAN-33*IN_TS*CNN2	-0,007	0,004	1	-0,019	0,004	0,005	0,006	1	-0,008	0,018
33*IN_TS*GAN-33*IN_TS*SWIN	-0,005	0,005	1	-0,016	0,007	-0,006	0,007	1	-0,022	0,011
33*IN_TS*CNN1-33*IN_TS*CNN2	0,021*	0,002	0	0,013	0,028	-0,028*	0,002	0	-0,037	-0,019
33*IN_TS*CNN1-33*IN_TS*SWIN	0,023*	0,002	0	0,014	0,033	-0,038*	0,003	0	-0,054	-0,023
33*IN_TS*CNN2-33*IN_TS*SWIN	0,003	0,002	1	-0,002	0,008	-0,01	0,003	0,334	-0,022	0,001
33*EX_TS*GAN-33*EX_TS*CNN1	0,862*	0,009	0	0,82	0,903	-89,925	33,104	1	-208,787	28,938
33*EX_TS*GAN-33*EX_TS*CNN2	0,139*	0,018	0	0,059	0,219	-0,36*	0,078	0,019	-0,703	-0,016
33*EX_TS*GAN-33*EX_TS*SWIN	0,144*	0,015	0	0,077	0,212	-0,554*	0,102	0	-1,002	-0,106
33*EX_TS*CNN1-33*EX_TS*CNN2	-0,723*	0,026	0	-0,837	-0,61	89,565	33,086	1	-28,988	208,117
33*EX_TS*CNN1-33*EX_TS*SWIN	-0,717*	0,021	0	-0,809	-0,63	89,371	33,031	1	-28,95	207,692
33*EX_TS*CNN2-33*EX_TS*SWIN	0,005	0,014	1	-0,024	0,035	-0,194	0,078	1	-0,459	0,072
34*IN_TS*GAN-34*IN_TS*CNN1	-0,05*	0,001	0	-0,056	-0,04	0,058*	0,002	0	0,05	0,067
34*IN_TS*GAN-34*IN_TS*CNN2	-0,035*	0,001	0	-0,042	-0,03	0,04*	0,002	0	0,03	0,05
34*IN_TS*GAN-34*IN_TS*SWIN	-0,043*	0,003	0	-0,056	-0,03	0,048*	0,004	0	0,032	0,064
34*IN_TS*CNN1-34*IN_TS*CNN2	0,015*	0	0	0,013	0,017	-0,018*	0,001	0	-0,021	-0,016
34*IN_TS*CNN1-34*IN_TS*SWIN	0,007	0,003	1	-0,003	0,018	-0,01	0,004	1	-0,024	0,004
34*IN_TS*CNN2-34*IN_TS*SWIN	-0,008	0,003	1	-0,018	0,003	0,008	0,004	1	-0,004	0,02
34*EX_TS*GAN-34*EX_TS*CNN1	-0,04	0,152	1	-0,353	0,273	62,735	76,905	1	-115,134	240,603
34*EX_TS*GAN-34*EX_TS*CNN2	-0,65*	0,096	0	-1,067	-0,23	88,18	69,821	1	-90,881	267,242
34*EX_TS*GAN-34*EX_TS*SWIN	-0,539	0,157	1	-1,195	0,117	86,91	70,12	1	-91,926	265,746
34*EX_TS*CNN1-34*EX_TS*CNN2	-0,61*	0,07	0	-0,919	-0,3	25,446	13,859	1	-15,253	66,145
34*EX_TS*CNN1-34*EX_TS*SWIN	-0,499*	0,014	0	-0,561	-0,44	24,175	13,108	1	-14,393	62,744
34*EX_TS*CNN2-34*EX_TS*SWIN	0,111	0,069	1	-0,082	0,304	-1,27	0,752	1	-3,404	0,863
35*IN_TS*GAN-35*IN_TS*CNN1	-0,025*	0,005	0,001	-0,045	-0	0,026	0,007	0,944	-0,005	0,056
35*IN_TS*GAN-35*IN_TS*CNN2	-0,01	0,006	1	-0,028	0,007	0,006	0,008	1	-0,013	0,024
35*IN_TS*GAN-35*IN_TS*SWIN	-0,019	0,009	1	-0,048	0,009	0,016	0,014	1	-0,019	0,05
35*IN_TS*CNN1-35*IN_TS*CNN2	0,014*	0,002	0	0,005	0,023	-0,02*	0,003	0	-0,032	-0,007
35*IN_TS*CNN1-35*IN_TS*SWIN	0,005	0,012	1	-0,021	0,031	-0,01	0,019	1	-0,05	0,031
35*IN_TS*CNN2-35*IN_TS*SWIN	-0,009	0,012	1	-0,036	0,018	0,01	0,018	1	-0,029	0,049
35*EX_TS*GAN-35*EX_TS*CNN1	-0,724*	0,027	0	-0,843	-0,6	60,138	37,391	1	-43,916	164,193

35*EX_TS*GAN-35*EX_TS*CNN2	-0,916*	0,002	0	-0,926	-0,91	60,431	37,335	1	-43,72	164,582
35*EX_TS*GAN-35*EX_TS*SWIN	-0,826*	0,008	0	-0,862	-0,79	60,19	37,208	1	-43,584	163,964
35*EX_TS*CNN1-35*EX_TS*CNN2	-0,192*	0,03	0	-0,321	-0,06	0,293	0,216	1	-0,273	0,858
35*EX_TS*CNN1-35*EX_TS*SWIN	-0,102	0,034	1	-0,232	0,027	0,052	0,336	1	-0,625	0,729
35*EX_TS*CNN2-35*EX_TS*SWIN	0,09*	0,006	0	0,062	0,118	-0,241	0,147	1	-0,653	0,171
36*IN_TS*GAN-36*IN_TS*CNN1	-0,012	0,003	0,425	-0,025	0,002	0,014	0,004	1	-0,003	0,031
36*IN_TS*GAN-36*IN_TS*CNN2	0,004	0,004	1	-0,005	0,012	-0,007	0,005	1	-0,02	0,006
36*IN_TS*GAN-36*IN_TS*SWIN	-0,004	0,004	1	-0,013	0,005	0	0,005	1	-0,01	0,009
36*IN_TS*CNN1-36*IN_TS*CNN2	0,015*	0	0	0,013	0,017	-0,021*	0,001	0	-0,025	-0,017
36*IN_TS*CNN1-36*IN_TS*SWIN	0,008*	0,001	0	0,003	0,012	-0,014*	0,001	0	-0,02	-0,008
36*IN_TS*CNN2-36*IN_TS*SWIN	-0,008*	0,001	0	-0,013	-0	0,007	0,002	0,127	0	0,014
36*EX_TS*GAN-36*EX_TS*CNN1	-0,518*	0,027	0	-0,637	-0,4	25,087	18,534	1	-23,472	73,646
36*EX_TS*GAN-36*EX_TS*CNN2	-0,849*	0,017	0	-0,923	-0,77	28,507	18,356	1	-21,911	78,925
36*EX_TS*GAN-36*EX_TS*SWIN	-0,851*	0,016	0	-0,922	-0,78	28,68	18,326	1	-21,798	79,159
36*EX_TS*CNN1-36*EX_TS*CNN2	-0,331*	0,033	0	-0,476	-0,19	3,42*	0,545	0	1,006	5,834
36*EX_TS*CNN1-36*EX_TS*SWIN	-0,333*	0,032	0	-0,472	-0,19	3,594*	0,451	0	1,582	5,605
36*EX_TS*CNN2-36*EX_TS*SWIN	-0,002	0,002	1	-0,006	0,003	0,173	0,126	1	-0,158	0,505
37*IN_TS*GAN-37*IN_TS*CNN1	0,004	0,004	1	-0,005	0,012	-0,017	0,005	1	-0,037	0,003
37*IN_TS*GAN-37*IN_TS*CNN2	0,025*	0,005	0	0,005	0,045	-0,048*	0,006	0	-0,077	-0,02
37*IN_TS*GAN-37*IN_TS*SWIN	0,002	0,002	1	-0,003	0,007	-0,012*	0,002	0,001	-0,022	-0,002
37*IN_TS*CNN1-37*IN_TS*CNN2	0,022*	0,001	0	0,017	0,026	-0,032*	0,002	0	-0,039	-0,024
37*IN_TS*CNN1-37*IN_TS*SWIN	-0,002	0,001	1	-0,006	0,002	0,005	0,002	1	-0,003	0,012
37*IN_TS*CNN2-37*IN_TS*SWIN	-0,024*	0,002	0	-0,034	-0,01	0,036*	0,004	0	0,018	0,055
37*EX_TS*GAN-37*EX_TS*CNN1	-0,778*	0,062	0	-1,048	-0,51	31,685	22,669	1	-28,331	91,7
37*EX_TS*GAN-37*EX_TS*CNN2	-0,793*	0,059	0	-1,049	-0,54	31,73	22,658	1	-28,294	91,753
37*EX_TS*GAN-37*EX_TS*SWIN	-0,44*	0,059	0	-0,697	-0,18	28,841	22,655	1	-29,398	87,081
37*EX_TS*CNN1-37*EX_TS*CNN2	-0,015	0,008	1	-0,038	0,008	0,045	0,022	1	-0,022	0,112
37*EX_TS*CNN1-37*EX_TS*SWIN	0,337*	0,008	0	0,304	0,371	-2,843*	0,025	0	-2,956	-2,731
37*EX_TS*CNN2-37*EX_TS*SWIN	0,352*	0,009	0	0,311	0,393	-2,888*	0,038	0	-3,058	-2,718
21*IN_TS*GAN-21*EX_TS*GAN	0,824	NA	NA	NA	NA	-13,156	NA	NA	NA	NA
21*IN_TS*CNN1-21*EX_TS*CNN1	0,551	NA	NA	NA	NA	-3,721	NA	NA	NA	NA
21*IN_TS*CNN2-21*EX_TS*CNN2	0,055	NA	NA	NA	NA	-0,245	NA	NA	NA	NA
21*IN_TS*SWIN-21*EX_TS*SWIN	-0,033	NA	NA	NA	NA	-0,906	NA	NA	NA	NA
22*IN_TS*GAN-22*EX_TS*GAN	0,666*	0,041	0	0,479	0,853	-7,734*	0,989	0	-12,138	-3,330
22*IN_TS*CNN1-22*EX_TS*CNN1	0,509*	0,042	0	0,32	0,698	-8,217	3,305	1	-19,492	3,059
22*IN_TS*CNN2-22*EX_TS*CNN2	0,122*	0,012	0	0,068	0,176	-0,459	0,124	0,763	-0,998	0,079
22*IN_TS*SWIN-22*EX_TS*SWIN	0,088	0,032	1	-0,029	0,204	-0,699	0,265	1	-1,634	0,236
23*IN_TS*GAN-23*EX_TS*GAN	0,778*	0,021	0	0,683	0,874	-5,963*	0,588	0	-8,625	-3,302
23*IN_TS*CNN1-23*EX_TS*CNN1	0,118*	0,022	0	0,024	0,213	-0,223	0,058	0,479	-0,476	0,030
23*IN_TS*CNN2-23*EX_TS*CNN2	0,057*	0,004	0	0,04	0,074	-0,103*	0,01	0	-0,15	-0,057
23*IN_TS*SWIN-23*EX_TS*SWIN	0,187*	0,027	0	0,072	0,303	-2,803*	0,397	0	-4,568	-1,038
24*IN_TS*GAN-24*EX_TS*GAN	0,593*	0,009	0	0,554	0,633	-7,136*	0,034	0	-7,291	-6,980
24*IN_TS*CNN1-24*EX_TS*CNN1	0,211*	0,006	0	0,186	0,236	-0,516*	0,008	0	-0,55	-0,481
24*IN_TS*CNN2-24*EX_TS*CNN2	0,122*	0,009	0	0,08	0,165	-0,667*	0,015	0	-0,736	-0,599
24*IN_TS*SWIN-24*EX_TS*SWIN	0,31*	0,004	0	0,29	0,329	-5,764*	0,007	0	-5,796	-5,732
25*IN_TS*GAN-25*EX_TS*GAN	0,65*	0,004	0	0,631	0,668	-3,503*	0,009	0	-3,542	-3,464
25*IN_TS*CNN1-25*EX_TS*CNN1	0,313*	0,003	0	0,3	0,326	-2,405*	0,004	0	-2,422	-2,388
25*IN_TS*CNN2-25*EX_TS*CNN2	0,023*	0,004	0	0,004	0,042	-0,034*	0,006	0	-0,059	-0,009

25*IN_TS*SWIN-25*EX_TS*SWIN	0,074*	0,003	0	0,061	0,086	-0,28*	0,005	0	-0,303	-0,258
26*IN_TS*GAN-26*EX_TS*GAN	0,638*	0,14	0,013	0,04	1,235	-4,27	2,063	1	-10,67	2,130
26*IN_TS*CNN1-26*EX_TS*CNN1	0,095*	0,019	0,002	0,013	0,177	-0,191*	0,034	0	-0,34	-0,041
26*IN_TS*CNN2-26*EX_TS*CNN2	0,209*	0,002	0	0,201	0,217	-1,334*	0,074	0	-1,665	-1,003
26*IN_TS*SWIN-26*EX_TS*SWIN	0,29*	0,021	0	0,197	0,382	-1,667*	0,081	0	-2,032	-1,302
27*IN_TS*GAN-27*EX_TS*GAN	0,774*	0,051	0	0,547	1,001	-10,308	2,54	0,203	-21,421	0,804
27*IN_TS*CNN1-27*EX_TS*CNN1	0,076*	0,004	0	0,059	0,094	-0,153*	0,021	0	-0,246	-0,059
27*IN_TS*CNN2-27*EX_TS*CNN2	0,059*	0,007	0	0,028	0,09	-0,118	0,039	1	-0,268	0,032
27*IN_TS*SWIN-27*EX_TS*SWIN	0,155*	0,029	0	0,032	0,278	-0,413	0,15	1	-0,956	0,131
28*IN_TS*GAN-28*EX_TS*GAN	0,757*	0,038	0	0,587	0,926	-6,022*	0,563	0	-8,546	-3,498
28*IN_TS*CNN1-28*EX_TS*CNN1	0,002	0,005	1	-0,008	0,012	-0,009	0,009	1	-0,031	0,013
28*IN_TS*CNN2-28*EX_TS*CNN2	-0,046*	0,008	0	-0,078	-0,01	-0,054	0,088	1	-0,248	0,141
28*IN_TS*SWIN-28*EX_TS*SWIN	0,235*	0,03	0	0,102	0,367	-3,725*	0,5	0	-5,948	-1,501
29*IN_TS*GAN-29*EX_TS*GAN	0,793*	0,048	0	0,582	1,004	-8,891*	1,603	0	-15,967	-1,815
29*IN_TS*CNN1-29*EX_TS*CNN1	0,864*	0,024	0	0,759	0,97	-45,871	11,525	0,278	-96,24	4,498
29*IN_TS*CNN2-29*EX_TS*CNN2	0,317*	0,012	0	0,264	0,369	-1,646*	0,088	0	-2,042	-1,250
29*IN_TS*SWIN-29*EX_TS*SWIN	0,03	0,013	1	-0,013	0,073	-0,347	0,107	1	-0,777	0,083
30*IN_TS*GAN-30*EX_TS*GAN	0,605*	0,057	0	0,353	0,858	-9,543	6,29	1	-26,674	7,587
30*IN_TS*CNN1-30*EX_TS*CNN1	0,098*	0,007	0	0,068	0,127	-0,224*	0,031	0	-0,363	-0,085
30*IN_TS*CNN2-30*EX_TS*CNN2	0,026*	0,003	0	0,013	0,039	-0,037*	0,005	0	-0,057	-0,017
30*IN_TS*SWIN-30*EX_TS*SWIN	0,075*	0,009	0	0,034	0,115	-0,125*	0,02	0	-0,214	-0,037
31*IN_TS*GAN-31*EX_TS*GAN	0,78*	0,073	0	0,458	1,102	-9,489	2,982	1	-21,316	2,338
31*IN_TS*CNN1-31*EX_TS*CNN1	0,311*	0,006	0	0,285	0,337	-1,721*	0,102	0	-2,176	-1,266
31*IN_TS*CNN2-31*EX_TS*CNN2	0,014*	0,003	0,005	0,001	0,027	-0,011	0,005	1	-0,028	0,005
31*IN_TS*SWIN-31*EX_TS*SWIN	0,19*	0,012	0	0,139	0,242	-0,591*	0,081	0	-0,952	-0,229
32*IN_TS*GAN-32*EX_TS*GAN	0,847*	0,033	0	0,702	0,991	-8,572	2,218	0,425	-18,239	1,095
32*IN_TS*CNN1-32*EX_TS*CNN1	0,206*	0,006	0	0,182	0,23	-0,698*	0,013	0	-0,756	-0,640
32*IN_TS*CNN2-32*EX_TS*CNN2	0,062*	0,002	0	0,052	0,072	-0,092*	0,003	0	-0,105	-0,079
32*IN_TS*SWIN-32*EX_TS*SWIN	0,124*	0,023	0	0,025	0,222	-1,099	0,324	1	-2,441	0,243
33*IN_TS*GAN-33*EX_TS*GAN	0,029*	0,005	0	0,008	0,05	-0,093*	0,011	0	-0,144	-0,043
33*IN_TS*CNN1-33*EX_TS*CNN1	0,918*	0,009	0	0,879	0,958	-90,05	33,098	1	-209,002	28,902
33*IN_TS*CNN2-33*EX_TS*CNN2	0,175*	0,018	0	0,098	0,253	-0,458*	0,073	0	-0,781	-0,135
33*IN_TS*SWIN-33*EX_TS*SWIN	0,178*	0,013	0	0,119	0,236	-0,641*	0,095	0	-1,061	-0,222
34*IN_TS*GAN-34*EX_TS*GAN	0,69*	0,098	0	0,263	1,117	-88,438	69,815	1	-267,644	90,768
34*IN_TS*CNN1-34*EX_TS*CNN1	0,7*	0,071	0	0,387	1,014	-25,762	13,914	1	-66,772	15,249
34*IN_TS*CNN2-34*EX_TS*CNN2	0,075*	0,004	0	0,059	0,092	-0,298*	0,059	0,002	-0,558	-0,038
34*IN_TS*SWIN-34*EX_TS*SWIN	0,194	0,07	1	-0,06	0,448	-1,576	0,809	1	-4,015	0,863
35*IN_TS*GAN-35*EX_TS*GAN	0,911*	0,003	0	0,897	0,925	-61,028	37,323	1	-165,555	43,498
35*IN_TS*CNN1-35*EX_TS*CNN1	0,212*	0,028	0	0,088	0,336	-0,916*	0,17	0	-1,664	-0,167
35*IN_TS*CNN2-35*EX_TS*CNN2	0,005	0,004	1	-0,004	0,015	-0,603	0,054	0	-0,845	-0,361
35*IN_TS*SWIN-35*EX_TS*SWIN	0,104*	0,013	0	0,046	0,162	-0,854*	0,172	0,003	-1,61	-0,099
36*IN_TS*GAN-36*EX_TS*GAN	0,852*	0,018	0	0,775	0,929	-29,258	18,39	1	-80,224	21,708
36*IN_TS*CNN1-36*EX_TS*CNN1	0,346*	0,032	0	0,207	0,485	-4,185*	0,504	0	-6,434	-1,936
36*IN_TS*CNN2-36*EX_TS*CNN2	-0,001	0,004	1	-0,008	0,007	-0,744*	0,069	0	-1,05	-0,437
36*IN_TS*SWIN-36*EX_TS*SWIN	0,005	0,003	1	-0,003	0,013	-0,577*	0,104	0	-1,035	-0,119
37*IN_TS*GAN-37*EX_TS*GAN	0,785*	0,055	0	0,543	1,027	-31,729	22,652	1	-91,741	28,283
37*IN_TS*CNN1-37*EX_TS*CNN1	0,004	0,011	1	-0,019	0,027	-0,028	0,029	1	-0,097	0,041
37*IN_TS*CNN2-37*EX_TS*CNN2	-0,033*	0,005	0	-0,056	-0,01	0,049*	0,01	0,002	0,006	0,092

37*IN_TS*SWIN-37*EX_TS*SWIN	0,343*	0,012	0	0,291	0,396	-2,876*	0,043	0	-3,067	-2,685
-----------------------------	--------	-------	---	-------	-------	---------	-------	---	--------	--------

Note.—Posthoc test comparisons are reported in terms of Mean Difference (Mean Diff.), Standard Error (Std. Error), sequential Bonferroni correction (Seq. Bonf. Sig.) and 95% confidence intervals (CIs) for differences, with lower and upper limits. DSC = Dice similarity coefficient, BAHD = Balanced average Hausdorff distance, UNT = UNETR, SWIN = Swin-UNETR, *P* = Pretraining, S = Scratch, IN = Internal Dataset, EX = External Dataset, IN_TS = Internal test set, EX_TS = External test set.

RSNA

Optimizing Performance of Transformer-based Models for Fetal Brain MRI Segmentation

Key Results

Swin-UNETR with pre-training weights, dataset size and input size optimization showed comparable performance over state-of-the-art models in resting-state functional MRI (rs-fMRI) fetal brain segmentation, with performance differences between mid- and late-fetal periods.

Patients:

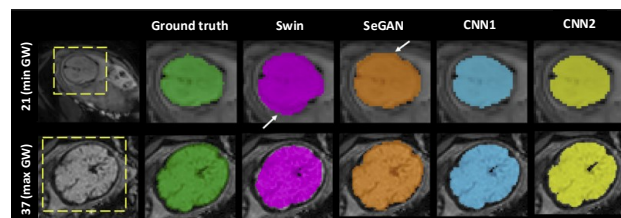
- Internal (172 fetuses; 519 images) and External (131 fetuses; 561 images) rs-fMRI fetal brain datasets

Methods:

- Study the influence of dataset size, pre-training approaches and image input size on Swin-UNETR and UNETR models.
- The internal and external datasets were used to cross-validate and assess the generalization capability of the best model against state-of-the-art models on scanner-types and gestational weeks (GW).

Results:

- Swin-UNTER showed flexibility in dealing with smaller datasets, regardless of pretraining approaches.
- Swin-UNETR performed best on mean image voxel size (Dice, 0.92).
- Swin-UNTER showed comparable performance with the reference standard on late-fetal period images but lower performance on images during early GWs.



Arrows = model segmentation errors

Pecco N et al. Published Online: June 26, 2024
<https://doi.org/10.1148/ryai.230229>

Radiology:Artificial Intelligence