BMC Medical Imaging

RESEARCH

Open Access

Automatic lateral ventricle and choroid plexus segmentation method in infant brain MR images

Junghwa Kang[1], Hyun Gi Kim[2,3*], Na-Young Shin[4,5] and Yoonho Nam[1*]

Abstract

Introduction Accurate segmentation of the lateral ventricles (LV) and choroid plexus (CP) in infant brain MRI is essential for understanding cerebrospinal fluid dynamics and early neurodevelopment. However, segmentation methods recently introduced for adult populations often underperform on infant data because of rapid anatomical changes, low tissue contrast, and motion artifacts, and they frequently misclassify tissue boundaries.

Method To address these challenges, we propose a fully automated deep learning method for joint LV and CP segmentation using T1-weighted MRI (Baby Connectome Project (BCP) dataset total $n = 154$; in-house retrospective dataset $n = 52$). Our approach integrates an anatomy-aware loss function that explicitly enforces the topological constraint of CP containment within the LV. The method was validated on two independent datasets to demonstrate clinical adaptability using Dice score, 95% Hausdorff distance (HD95), and the average symmetric surface distance (ASSD).

Results The method achieved Dice scores of 0.818 ± 0.075 for LV and 0.827 ± 0.084 for CP in the BCP dataset, with HD95 of 7.487 ± 7.351 mm and 4.925 ± 2.897 mm, and ASSD of 1.175 ± 0.524 mm and 0.818 ± 0.239 mm, respectively. In the in-house dataset, the method achieved Dice scores of 0.964 ± 0.060 for LV and 0.932 ± 0.059 for CP, with HD95 of 0.310 ± 0.542 mm and 4.148 ± 3.726 mm, and ASSD of 0.088 ± 0.151 mm and 0.280 ± 0.239 mm, respectively.

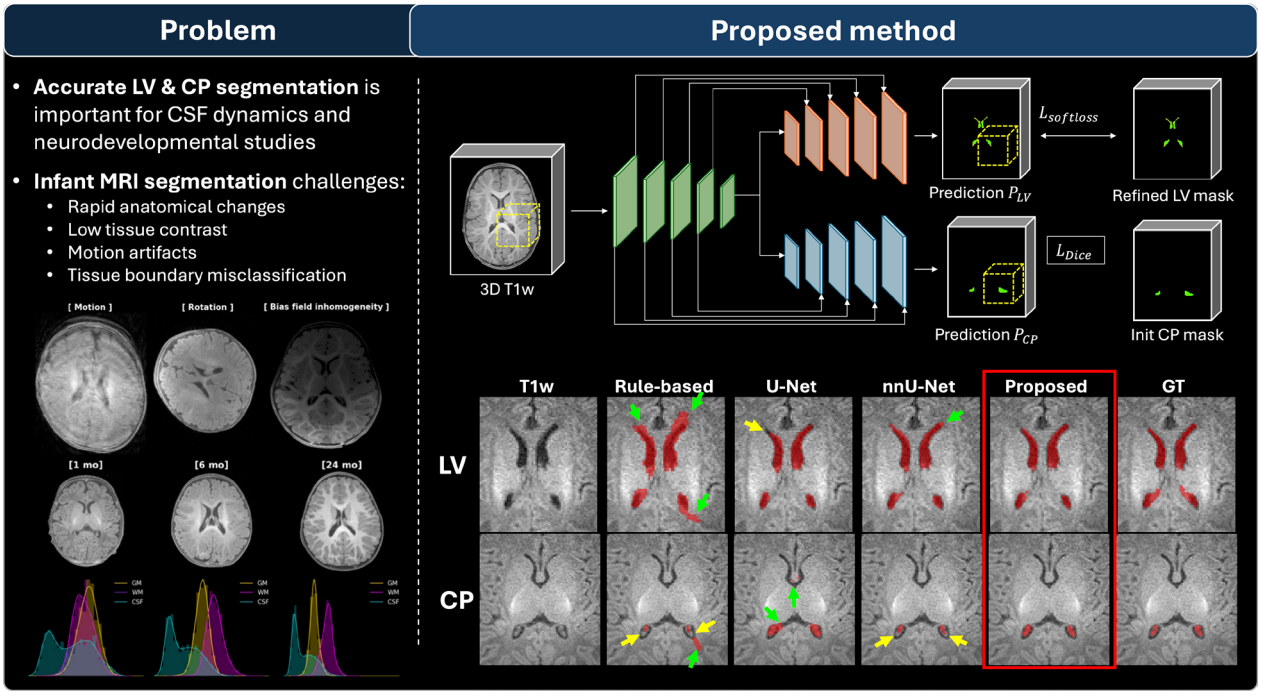
Conclusion This method addresses limitations of prior methods by ensuring anatomical consistency without manual annotation. The approach has the potential to support large-scale studies investigating CP morphology and its relevance to early neurodevelopment.

Keywords Choroid plexus, Lateral ventricle, Segmentation, Infant Brain

*Correspondence:
Hyun Gi Kim
catharina@amc.seoul.kr
Yoonho Nam
yoonhonam@hufs.ac.kr

Full list of author information is available at the end of the article

Graphical Abstract



Introduction

The choroid plexus (CP), a highly vascularized structure within the brain ventricles, is essential for cerebrospinal fluid (CSF) production and regulation [1, 2]. Through CSF circulation, the CP plays an important role in the glymphatic system, facilitating the clearance of metabolic waste from the brain [3]. Alterations in CP morphology are associated with aging, development, and several neurological disorders, underscoring the importance of accurate CP quantification [4–7]. Because the lateral ventricles (LV) contain the largest CP volume, quantitative studies typically focus on CP measurements within the LV region.

Several studies have analyzed CP volume using manual delineation [6, 8], but this approach is labor-intensive and exhibits substantial inter-observer variability due to irregular CP morphology and low contrast on T1-weighted (T1w) MRI, where the CP lies adjacent to the lateral ventricular wall. These challenges limit scalability and reproducibility in large-scale studies, highlighting the need for automated segmentation methods. Automatic methods widely applied in previous studies, such as Gaussian mixture models (GMM), rely on accurate prior LV segmentation [9–11]. Inaccuracies in LV segmentation can directly propagate and lead to errors in downstream CP segmentation. While these approaches can show improved accuracy over atlas-based methods like FreeSurfer [12] in adult populations, they still require careful initialization and are sensitive to image quality.

More recently, deep learning approaches have been introduced for LV and CP segmentation. Contrast-agnostic methods like SynthSeg [13] and FetalSynthSeg [14] enable robust LV segmentation across MRI contrasts using synthetic data generation, though CP-specific adaptation remains challenging. A two-step 3D U-Net architecture has been proposed for CP segmentation in adults with neurological disorders [15]. ASCHOPLEX et al. (2024) applied ensemble learning to improve model generalization on adult data [16], and UX-Net incorporated attention mechanisms for this task [17]. Some studies perform joint segmentation of LV and CP [18]. However, these supervised deep learning methods typically require either extensive manual correction or large volumes of high-quality annotated training data, limiting their clinical application where well-curated datasets are unavailable.

These challenges are particularly severe in infant populations, where labeled training data are scarce [19, 20]. Unlike adults, infant brains show rapid developmental changes that alter both anatomy and MRI signal

characteristics (Fig. 1d). In infants, CP segmentation is especially challenging because the CP is close to the ventricular wall (Fig. 1a) and motion artifacts are common (Fig. 1b). Consequently, segmentation methods designed for adults, such as FreeSurfer, often fail on infant data. Atlas-based approaches and some deep learning methods have also shown limitations in handling individual anatomical variability and indistinct boundaries, particularly in infants during early developmental stages [21, 22]. Even an infant-specific tool like Infant FreeSurfer (InfantFS) [23] often misclassifies adjacent structures, particularly in the presence of suboptimal image quality, producing LV false positives (FP) that affect CP segmentation (Fig. 1c). Recent synthetic-data driven deep learning methods like BabySeg [24] may provide improved LV segmentation across infant MRI contrasts but do not output CP, highlighting the need for joint LV-CP segmentation.

In fetal and neonatal brain imaging, T2-weighted (T2w) MRI is widely used because of its superior tissue contrast in this period [25–27]. However, practical considerations remain important for quantitative volumetric studies. In routine clinical practice and many existing datasets, three-dimensional (3D) T1w acquisitions are more commonly available than 3D T2w images. In addition, 3D imaging provides advantages for quantitative analysis of thin and complex structures, such as CP, which can be substantially influenced by slice thickness and partial-volume effects in two-dimensional (2D) acquisitions. Although several methods have been proposed to reconstruct 3D volumes from 2D slices [28–30], validation for quantitative assessment of fine structures such as the choroid plexus remains limited.

To address these challenges in the infant population, we propose an automated method for joint segmentation of LV and CP using only 3D T1w brain MR images. The proposed method employs a single network architecture with an anatomy-aware loss designed to encourage anatomically plausible spatial relationships, specifically enforcing containment of the CP within the LV. To reduce reliance on manual annotations, the model is trained using automatically generated pseudo-labels. The proposed method is evaluated on two independent datasets to examine its applicability across different data sources. In addition, the proposed method is applied to a large infant cohort to investigate age-related trends in LV and CP volumes, demonstrating its potential utility for developmental studies.

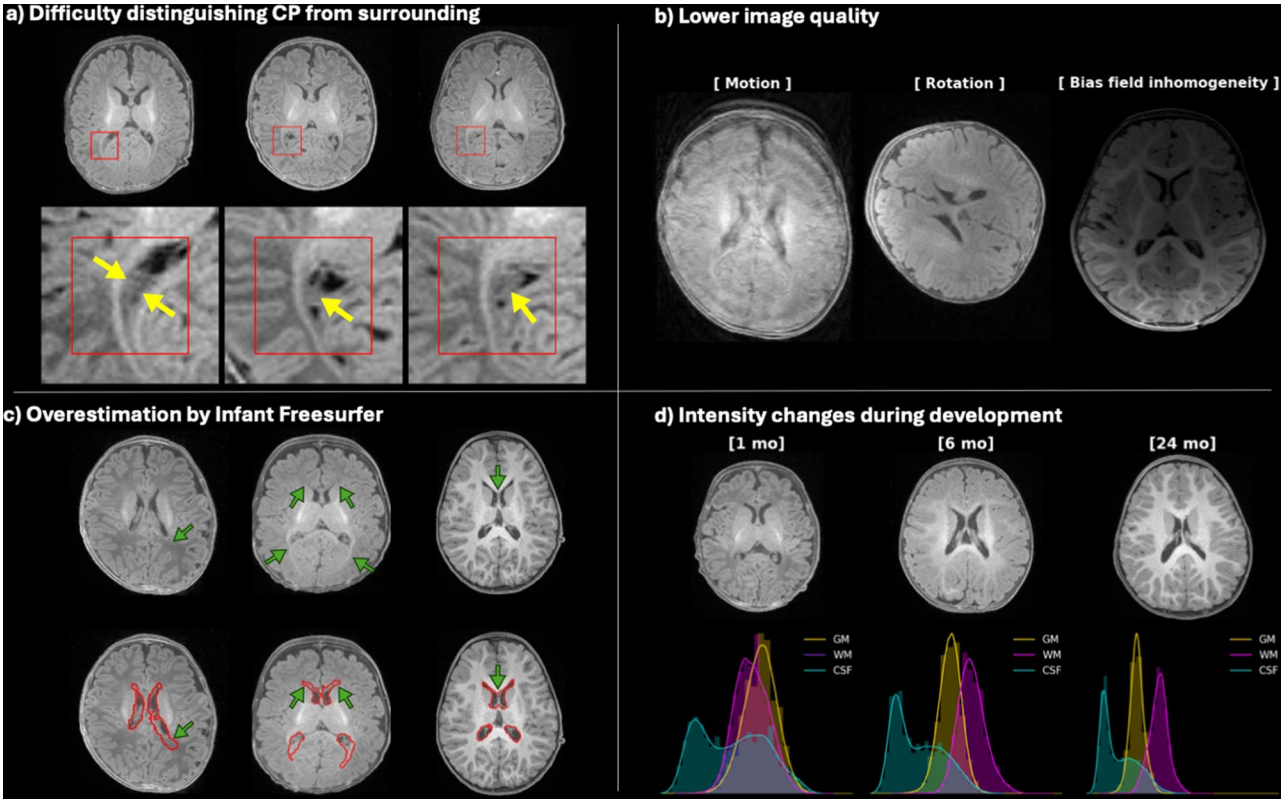


Fig. 1 Common challenges in LV and CP segmentation for infant brain MRI. (**a**) demonstrates the difficulty of distinguishing the CP from surrounding structures (yellow arrows indicate the CP). (**b**) presents representative examples of lower image quality in infant scans, which impacts segmentation accuracy due to reliance on image quality. (**c**) shows that the LV label generated by InfantFS frequently produces false positives (green arrows). (**d**) shows the challenges in segmenting brain tissues due to variations in tissue contrasts as the infant brain develops

Method

Dataset

BCP dataset

We used a cohort of 154 healthy infants from the UNC/UMN Baby Connectome Project (BCP) [31]. The cohort included infants with postnatal age (PA) up to 24 months. We randomly selected 52 infants using age-stratified sampling for model development and evaluation, with 13 infants per age group (0–6 months, 7–12 months, 13–18 months, and 19–24 months). We split these 52 infants into training (n = 32), validation (n = 10), and test (n = 10) sets. The imaging protocol included T1w. Data preprocessing and coregistration were performed according to the procedures summarized in Table 1. The remaining 102 subjects with longitudinal scans (278 scans) were used in a separate study to analyze age-related trends and comparison, and were not included in development. All procedures were approved by the Institutional Review Boards of the University of North Carolina at Chapel Hill and the University of Minnesota.

In-house dataset

To demonstrate the adaptability of the proposed method to real-world clinical datasets, as well as its methodological robustness, we applied the entire segmentation pipeline to an independent retrospective clinical cohort. This cohort consisted of 52 infants (gestational age (GA) at birth 24–41 weeks) collected from our institution, including cases with MRI-visible hypoxic-ischemic injury (HIE). The imaging data comprised T1w MRI scans, with acquisition details summarized in Table 1. The dataset was partitioned into training (n = 32), validation (n = 10), and test (n = 10) sets using age-stratified random sampling. This study was approved by our Institutional Review Board with a waiver of informed consent.

Table 1 Summary of MRI acquisition protocols and demographic characteristics for two datasets (BCP and in-house datasets)

	BCP	In-house
Scanner	3T Siemens Prisma	3T Siemens Magnetom Vida
Coil	32 channel head coil	64 channel head coil
Sequence	T1-weighted MPRAGE	T1-weighted MPRAGE
TR/TE (ms)	2400/2.24	2400/2.37
TI (ms)	1060	1000
Resolution (mm^3)	0.8 × 0.8 × 0.8	0.8 × 0.8 × 0.8
Total scan time	6:38	3:49
Number of subjects (training/valid/test/total)	32/10/10/154	32/10/10/52
Age (mean (SD))	10.27 (PNA in months)	32.25 (GA in weeks)
Sex (subjects, male/female)	89/91	24/28

LV and CP segmentation method

The proposed method for LV and CP segmentation, consisting of pseudo-label generation and subsequent segmentation model training, is described in Fig. 2.

Generation of training labels

The labels used for training were derived from InfantFS automated segmentations (Fig. 2a). To reduce the boundary over-segmentation often present in atlas-based InfantFS outputs, we adopted the region-informed GMM strategy using LV and CP masks [10, 11]. This method enhances segmentation accuracy by restricting intensity modeling to the LV voxels, including the CP, thereby effectively resolving the intensity ambiguity between the CP and surrounding isointense tissues. The procedure involved three specific steps. First, LV masks (including CP) were extracted from the InfantFS (aseg.mgz, version 4.2) [23] to define the region of interest. Second, spherical erosion (radius = 1 voxel) was applied to the masks to remove partial volume effects at the boundaries. Third, a three-component GMM (n_components = 3; full covariance, Dirichlet-process prior, default initialization; scikit-learn v1.4.2) was fitted to the T1w intensities exclusively within these eroded LV masks to classify the CP and CSF components, resulting in the final refined CP labels. For GMM fitting, the means, standard deviations, and component proportions were initialized as [0.30, 0.58, 0.81], [0.09, 0.08, 0.07], and [0.32, 0.24, 0.44], respectively.

Network architecture

For automatic segmentation, we employed a deep learning model based on a 3D U-Net architecture [32] with a shared encoder and two separate decoders for simultaneous segmentation of the LV and CP. This dual-decoder structure addresses the volumetric imbalance and distinct textural characteristics of the two structures. By separating the decoding pathways, the model allows for independent optimization, preventing the larger LV from dominating the learning of the smaller CP.

The shared encoder consists of four stages with increasing numbers of channels, followed by a bottleneck layer. The convolutional blocks use instance normalization and non-linear activation functions. The two independent decoder branches perform upsampling via transposed convolutions and skip connections. This configuration allows each branch to reconstruct task-specific features, specifically boundary delineation for the LV and textural differentiation for the CP, while minimizing feature interference.

The two decoder branches use distinct output configurations and loss functions to address structure-specific optimization requirements. The LV decoder outputs a single–channel probability map through a sigmoid activation, formulating the task as a boundary regression

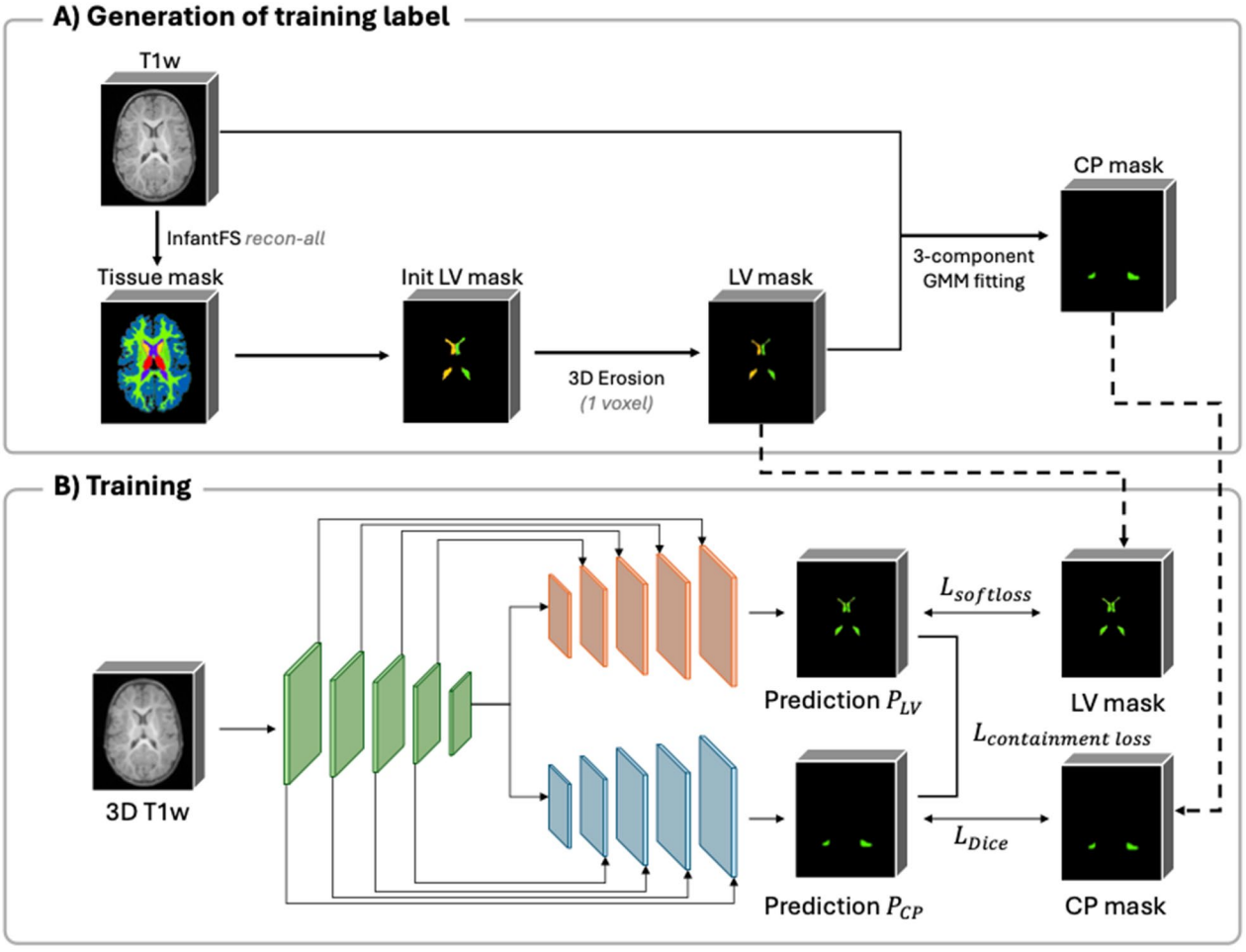


Fig. 2 Pipeline of proposed training strategy for automatic LV and CP segmentation in infant MRI

problem. This branch employs a boundary-aware soft loss [33] for precise boundary delineation. The loss utilizes a pixel-wise spatial weight map (W), calculated as:

$$W = 0.7 \cdot W_{edge} + 0.3 \cdot Y_{LV}$$

where Y_{LV} denotes the ground truth (GT) mask and W_{edge} represents the morphological gradient calculated by subtracting the eroded mask from the dilated mask using a $3 \times 3 \times 3$ kernel. This applies larger weights to the boundary regions.

The CP decoder generates a two-channel probability map (background and CP) using a softmax activation. This branch utilizes a hybrid loss function combining Dice loss and weighted Cross-Entropy loss. Class weights were empirically set to 0.3 for the background and 1.7 for the CP to mitigate class imbalance and enhance sensitivity for the smaller CP structure.

Based on ventricular anatomy, we assume the CP should be contained within the LV. We first convert the predicted LV probability map $P_{LV}(x)$ into a binary mask by thresholding at 0.5 and then dilate this mask by $r = 3$ voxels to tolerate small boundary errors. This yields a relaxed LV region $\widetilde{M}_{LV}(x)$. We penalize CP probability predicted outside this relaxed LV region using the containment loss [34]:

$$\widetilde{M}_{\mathrm{LV}}(x) = \mathrm{dilate}\left(\mathbb{I}[P_{\mathrm{LV}}(x) > 0.5], r\right)$$

$$L_{cont} = \frac{\lambda}{|\Omega|} \sum_{x \in \Omega} P_{CP}(x) \cdot [1 - \widetilde{M}_{\mathrm{LV}}(x)]$$

Here, P_{CP} and P_{LV} are the predicted probability maps for CP and LV, $\mathbb{I}[\cdot]$ is the indicator function, and $|\Omega|$ is the number of voxels in the loss domain. The dilation radius ($r = 3$ voxels) compensates for boundary uncertainty inherent in infant MRI. The regularization weight $\lambda = 0.3$ balances segmentation performance with anatomical plausibility. During optimization, we stop gradients through P_{LV} for this term so that L_{cont} affects only the CP decoder.

Model training

Models were implemented in PyTorch 2.4.1 and trained on an NVIDIA Quadro RTX 6000 GPU using AdamW optimizer [35] with learning rate 1×10^{-4}, weight decay 1×10^{-5}, and gradient clipping at norm 3.0. The model was trained using the CP and LV pseudo-labels generated according to the procedure described in section "Dataset". Data augmentation included random isotropic resampling, intensity scaling, Gaussian smoothing, and bias-field simulation. We sampled $64 \times 64 \times 64$ patches with random rotations and flips during training. Additional implementation details can be found at [https://github.com/jhkang0526/Infant_LVCPseg].

Evaluation

Manual segmentation

GT labels for the test set were generated semi-automatically. These GT labels were used exclusively for evaluation and were not employed in the model training. For the LV, we first obtained automated segmentations from InfantFS on T1w and then performed manual corrections. For the CP, we generated an initial mask within the manually corrected LV using a Gaussian mixture model and then performed manual corrections for both FP and false negative (FN) voxels. A trained researcher conducted this process using ITK-SNAP (version 3.8), and a board-certified neuroradiologist independently reviewed and adjudicated any discrepancies, ensuring high accuracy and consistency for reliable GT labels.

Comparison method

To assess the effectiveness of the proposed method, we compared the proposed method with four alternative approaches for LV and CP segmentation (Fig. 3):

- InfantFS and InfantFS + GMM - an atlas-assisted tool (InfantFS) that first produces LV masks, and CP masks are then obtained by fitting a GMM to intensities restricted to the LV region.
- BabySeg and BabySeg + GMM - Domain-randomized deep learning framework trained on synthetic images augmented with real infant MRI data, using group convolutions for robust multi-contrast brain segmentation. However, automated CP-specific segmentation is not currently supported. Then, CP masks are then obtained by fitting a GMM to intensities restricted to the LV region.
- 3D U-Net - Four resolution levels with two convolutional blocks per level with InstanceNorm and residual connections. Base width 32 that doubles at each downsampling step. A combined segmentation loss $L_{\text{seg}} = \alpha\, L_{\text{Dice}} + (1-\alpha)\, L_{\text{CE}}$ with $\alpha = 0.5$, where L_{Dice} and L_{CE} denote the Dice and cross-entropy losses.
- nnU-Net [36] - Automated architecture and training.
- Proposed method trained with T1w.

All models utilized the same single-channel T1w inputs and per-volume intensity normalization. For three deep learning models, training was performed using a consistent set of GT masks for the LV and CP across all experiments. For the proposed method and U-Net, training hyperparameters were standardized to ensure a fair comparison (section "Model training"). In contrast, nnU-Net was trained using its default self-configuring pipeline (https://github.com/MIC-DKFZ/nnUNet).

Accuracy evaluation

Segmentation performance on the test sets was evaluated using the Dice similarity coefficient (DSC), the

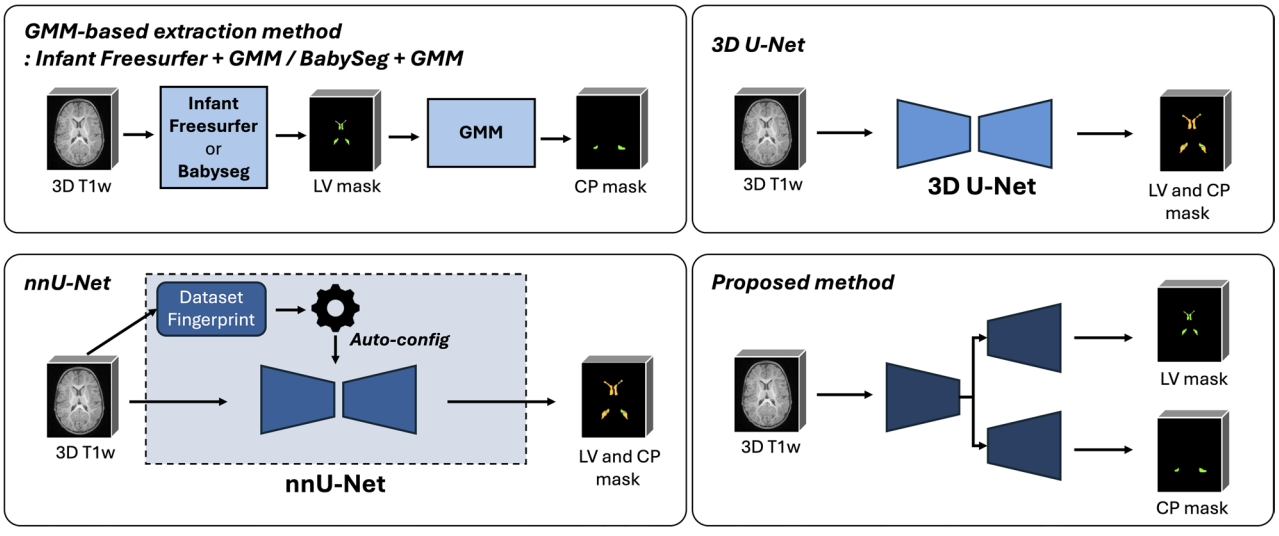


Fig. 3 Schematic overview of the comparison methods: InfantFS + GMM, BabySeg + GMM, 3D U-Net, nnU-Net, and the proposed methods

95th-percentile Hausdorff distance (HD95), and the average symmetric surface distance (ASSD) with manually generated GT. DSC measures spatial overlap and higher values mean more overlap. HD95 is the 95th percentile of the bidirectional surface distance between the segmentation and the reference and lower values mean closer boundaries. ASSD is the mean of the bidirectional surface distances in millimeters, with lower values indicating more accurate boundary alignment.

In addition to these segmentation performance metrics, we evaluated the volumetric agreement between the automated segmentation and the GT. The Pearson correlation coefficient (r) and the Mean Absolute Error (MAE) were calculated to assess the linear relationship and absolute magnitude of errors.

Statistical analysis

Performance differences between the proposed method and alternative approaches were evaluated using two-sided Wilcoxon signed-rank tests on paired per-subject metrics. The same statistical test was employed to compare the automated segmentations against the manual reference. Statistical significance was defined as $p < 0.05$. All statistical computations were conducted in Python (v3.12.4) utilizing the SciPy (v1.11.4) and statsmodels (v0.14.2) libraries.

Because manual GT labeling for the entire scans (102 subjects; 279 scans) would be prohibitively time-consuming, we assessed age-dependent, method-specific differences in the estimated CP volume using linear mixed-effects models (LMM). Analyses were conducted separately for absolute CP volume (mm^3) and normalized CP fraction (CP/ total intracranial volume (ICV)). Total ICV was derived using a separate U-Net model trained on manually annotated ICV masks. For each outcome (y), all observations from 0 to 25 months were modeled with a cubic B-spline for age (months) and a method-by-age interaction:

$$\begin{aligned} y_{\{ij\}} = \beta_0 + \ (Model_j) + \ f\,(month_i) \\ + \ (Model_j \times \ \ f\,(month_i)) \\ + \ b_j + \ u_{\{ij\}} + \ \epsilon_{\,\{ij\}} \end{aligned}$$

where $model_j$ denotes the categorical fixed effect of segmentation method (U-Net, nnU-Net, InfantFS(+GMM), and proposed), $f\,(month_i)$ is a cubic B-spline basis of month, b_j is a subject-level random intercept, and $u_{\{ij\}}$ is a visit-level variance component capturing within-visit correlation across methods. We tested the global method-by-age interaction using a Wald χ^2test and performed planned contrasts at 1 and 3 months (Δ = Proposed – U-Net/nnU-Net/InfantFS) with 95% confidence intervals. The p-values across the six contrasts (3 comparison methods × 2 ages) were adjusted using Holm correction.

Result

Segmentation performance on the BCP dataset

The quantitative segmentation performance on the BCP dataset is summarized in Table 2. For LV segmentation, the proposed method shows the highest spatial overlap, with better DSC than InfantFS, BabySeg, U-Net, and nnU-Net. Boundary distance metrics (HD95 and ASSD) also indicate competitive or improved performance compared with the other methods. For CP segmentation, the proposed method achieves the best overall accuracy across DSC and HD95, while nnU-Net attains a slightly lower ASSD. These trends suggest that the proposed method provides more robust CP boundaries than InfantFS + GMM, BabySeg + GMM, and U-Net baselines, with performance comparable to or better than nnU-Net.

In volumetric analysis, the proposed method showed the highest agreement with the GT among the compared methods. For LV, it achieved the highest Pearson correlation ($r = 0.982$) and the lowest MAE (218.60 mm^3). For CP, the proposed method also resulted in a high correlation ($r = 0.703$), surpassing the InfantFS + GMM ($r = 0.272$), BabySeg + GMM ($r = 0.368$), and U-Net ($r = 0.574$) baselines. While nnU-Net showed a slightly lower MAE for CP compared with the proposed method (332.01 mm^3 vs. 351.50 mm^3), the proposed method achieved the highest correlation coefficient among all methods.

Figures 4 and 5 present a qualitative comparison of segmentation results using subjects selected from the 24-, 3-, and 1-month age groups. In LV segmentation, while InfantFS showed a tendency toward boundary expansion relative to the GT, the proposed method demonstrated reduced FP and FN, as indicated by the yellow and green arrows in Fig. 4. For CP segmentation, the proposed method demonstrated the closest agreement with the GT compared with other methods. In contrast, InfantFS and U-Net failed to segment properly and exhibited excessive over-segmentation and under-segmentation in the 24-, 3-, and 1-month images, as highlighted by the yellow and green arrows in Fig. 5.

Figure 6 shows the robustness of the proposed method in cases affected by imaging artifacts. In the presence of artifact and suboptimal bias field correction, the proposed method provides more accurate segmentation than the InfantFS, as indicated by the green arrows. Likewise, in cases where InfantFS overestimates structures, the proposed method shows improved performance in both LV and CP segmentation.

Table 2 Performance comparisons for LV and CP segmentation in BCP dataset

	LV					CP				
	InfantFS	**BabySeg**	**U-Net**	**nnU-Net**	**Proposed**	**InfantFS + GMM**	**BabySeg + GMM**	**U-Net**	**nnU-Net**	**Proposed**
DSC ↑	0.680 ± 0.157	0.784 ± 0.304	0.781 ± 0.069	0.782 ± 0.059	0.818 ± 0.075	0.569 ± 0.167	0.492 ± 0.211	0.589 ± 0.117	0.723 ± 0.116	0.827 ± 0.084
HD95 ↓	7.264 ± 7.826	6.348 ± 5.519	6.878 ± 5.759	9.535 ± 9.070	7.487 ± 7.351	15.027 ± 9.775	27.223 ± 11.820	37.283 ± 7.124	5.184 ± 2.563	4.925 ± 2.897
ASSD ↓	1.538 ± 1.502	2.753 ± 3.888	1.574 ± 0.621	0.972 ± 0.404	1.175 ± 0.524	2.214 ± 2.753	4.248 ± 5.839	2.934 ± 0.816	0.774 ± 0.309	0.818 ± 0.239
Pearson r	0.871	0.962	0.931	0.968	0.982	0.272	0.386	0.574	0.608	0.703
MAE (mm^3)	1481.11	1401.959	1589.05	1265.528	1218.6	1471.953	319.14	734.304	343.490	351.329

Segmentation performance on clinical dataset (in-house dataset)

The quantitative segmentation performance on the clinical dataset is summarized in Table 3. For LV segmentation, the proposed method shows the highest DSC and the lowest boundary errors among all methods, achieving better performance than the InfantFS, BabySeg, U-Net, and nnU-Net. For CP segmentation, the proposed method also provides the best performance across DSC, HD95, and ASSD, with clear improvements over the other approaches.

In the volumetric analysis, the proposed method demonstrated robust agreement with the GT. For LV, it achieved the highest Pearson correlation ($r = 0.853$) and the lowest MAE (191.131 mm^3) when compared with the GT. For CP, the proposed method also resulted in the highest correlation ($r = 0.763$) and the lowest MAE (118.5 mm^3), surpassing other comparison methods.

Figures 7 and 8 present representative LV and CP segmentation results from three randomly selected subjects at GA of 37–39 weeks. The proposed method demonstrates more consistent visual agreement with the GT, whereas the comparison methods show more FP and FN, including over-segmentation, under-segmentation, and isolated outlier regions.

Age-related trends in segmentation volumes (BCP dataset)

Figure 9 summarizes age-binned distributions of LV and CP volumes and their ICV-normalized ratios in the 0–24-month BCP subset. For each age bin, boxplots show the empirical spread across subjects, while the overlaid mixed-effects model estimates depict the population-level trajectory with 95% confidence intervals for each method.

For LV volume, the mixed-effects model indicated an age-dependent model effect (model × age: $\chi^2(12) = 305.9$, $p < 0.001$). InfantFS produced the largest LV volumes, whereas the proposed model yielded more conservative estimates, with the largest separations at 1–3 months (See Additional_file1 - Supplementary Table 1). The proposed model was also lower than nnU-Net at 1 and 3 months, while the difference from U-Net was not significant at 1 month but became significant at 3 months. For LV/ICV, the age-dependent model effect remained pronounced ($\chi^2(12) = 1505.5$, $p < 0.001$), and the proposed model showed consistently lower ratios than InfantFS at 1–3 months (See Additional_file2 - Supplementary Table 2).

For CP volume, a clear age-dependent model effect was likewise observed ($\chi^2(12) = 147.6$, $p < 0.001$). At both 1 and 3 months, the proposed model produced lower CP volumes than GMM and U-Net but higher volumes than nnU-Net (See Additional_file3 - Supplementary Table 3). For CP/ICV, the age-dependent model effect persisted

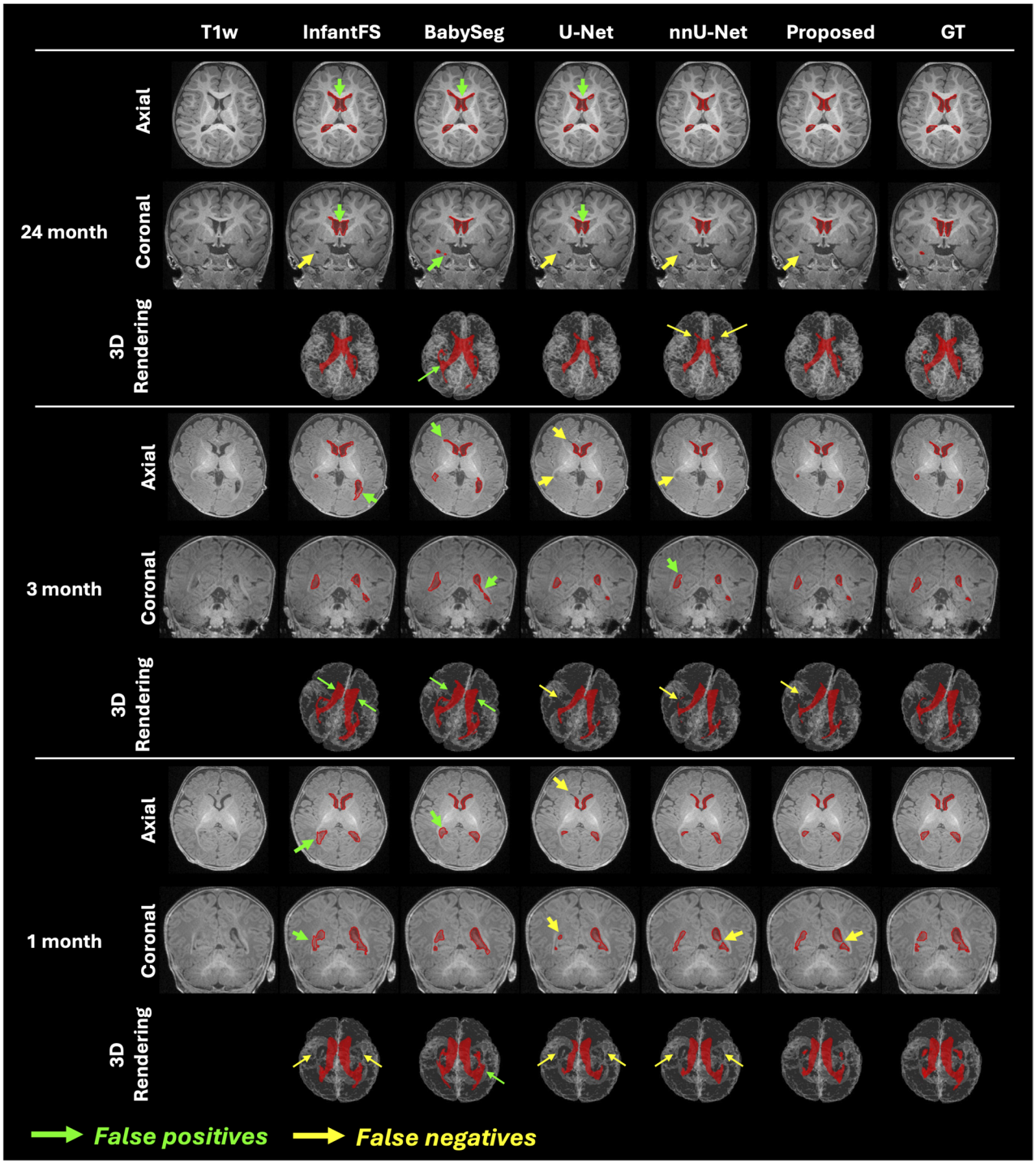


Fig. 4 Visual comparison of the segmentation results of the LV on 24 month, 3 month, and 1 month T1w from BCP dataset, obtained from manual delineation, InfantFS, BabySeg, and three automatic methods. Yellow arrow indicated false negatives, and green arrow indicated false positives

($\chi^2(12) = 214.2$, $p < 0.001$), with the proposed model showing lower ratios than GMM and U-Net yet higher ratios than nnU-Net at 1 and 3 months (See Additional_file4 - Supplementary Table 4).

Discussion

This study developed and validated a deep learning method for automated joint segmentation of the LV and CP in infant MRI from T1w images. This single-modality approach avoids potential errors from multi-modal co-registration and allows for application even when 3D

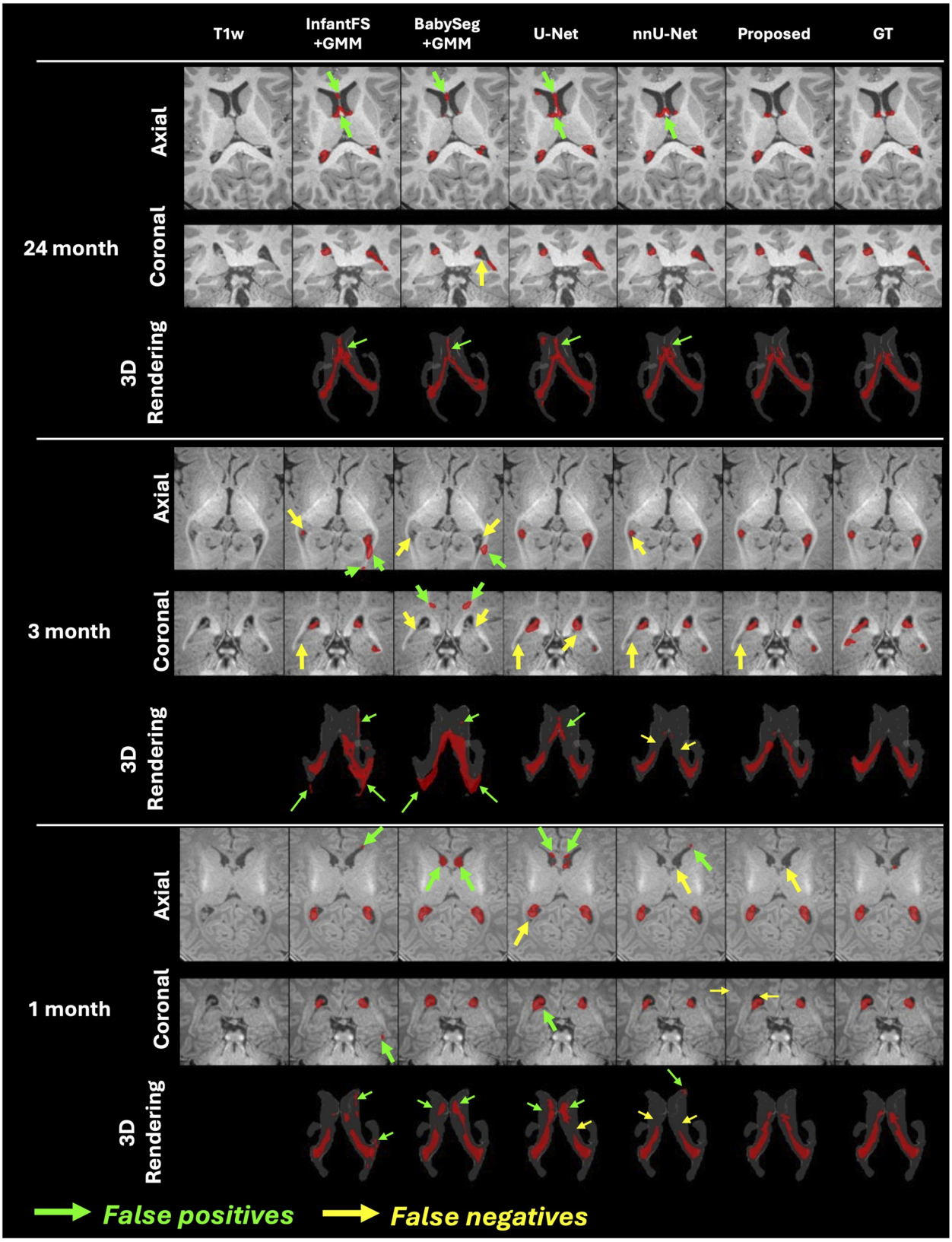


Fig. 5 Visual comparison of the segmentation results of the CP on both 24 month, 3 month, and 1 month T1w images from BCP dataset, obtained from manual delineation, InfantFS + GMM, BabySeg + GMM, and three automatic methods. Yellow arrow indicated false negatives, and green arrow indicated false positives

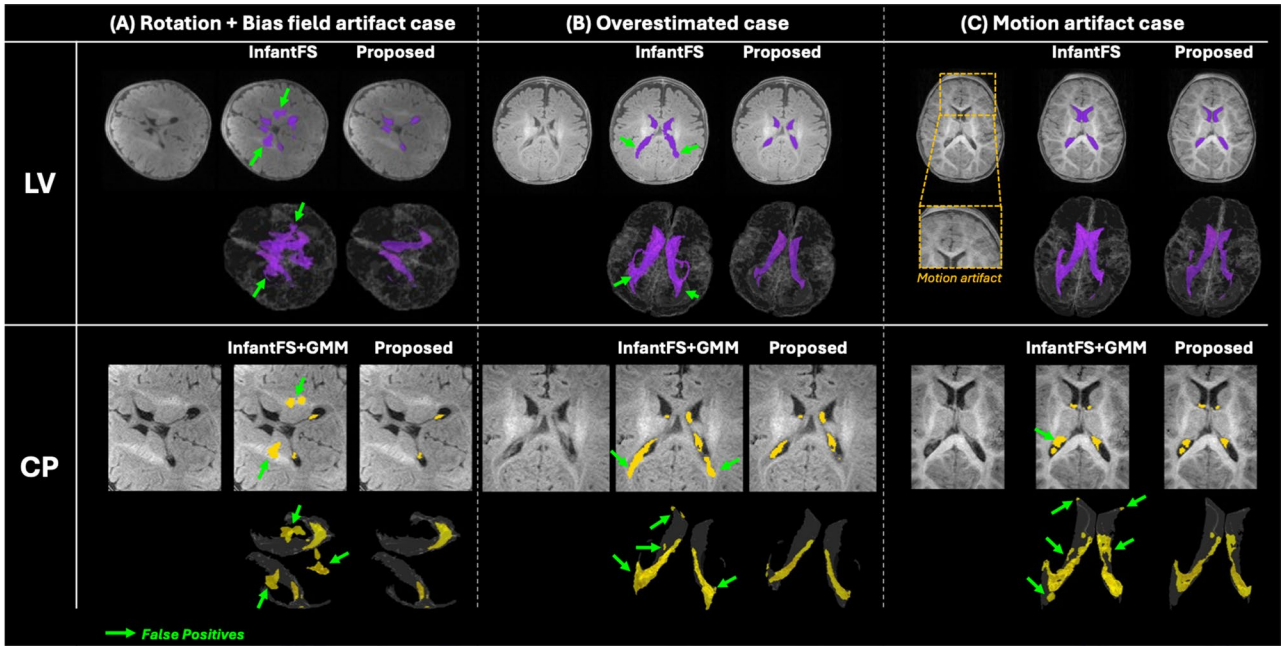


Fig. 6 Comparison of InfantFS and proposed method on cases with (**A**) rotation and bias field artifact case, (**B**) overestimation case, and (**C**) motion artifact case. Yellow arrow indicated false negatives, and green arrow indicated false positives

T2w images are unavailable. Although 3D T2w acquisition is increasingly recommended in neonatal brain MRI protocols [25], it is not as commonly obtained as 3D T1w images [26, 27]. To enable application across a broader age range, the utilization of 3D T1w images instead of 3D T2w images was a reasonable approach. By integrating an anatomy-aware loss and utilizing automatically refined labels, our method addresses key limitations of prior approaches. It shows improved performance across multiple age-specific subgroups below 24 months. The adaptability of the proposed method was further verified through its application to an independent clinical dataset with a different imaging protocol. Compared with the InfantFS-based method and deep learning-based segmentation models, the proposed method provided more precise and consistent delineation of the boundary between the LV and brain parenchyma. A boundary-sensitive loss and an anatomical containment constraint reduced mislabeling of CP voxels outside the LV. In addition, the method demonstrated robust performance in the presence of common artifacts, such as subject motion and intensity inhomogeneity.

The improvements in Dice, lower HD95, and lower ASSD compared to InfantFS and GMM-based methods show the efficacy of the proposed method. In contrast to GMM approaches that rely on voxel intensities, our model uses a dual-decoder architecture to capture both spatial context and boundary features. In addition, the anatomical containment loss promotes anatomically plausible containment relationships and reduces topologically inconsistent predictions. These components contribute to boundary refinement in challenging cases with artifacts and weak tissue contrast. From an architectural perspective, the shared encoder supports feature sharing across objectives. Dedicated decoder branches allow task-specific specialization, which may reduce interference between heterogeneous feature representations and further support consistent delineation.

Analysis of age-dependent LV and CP volumes revealed significant model-by-age interactions [5], indicating that developmental trajectories differed across methods. The proposed model consistently produced more conservative LV and CP volumes than InfantFS and U-Net, while showing slightly higher values than nnU-Net, particularly at early ages (1–3 months), resulting in balanced LV/ICV and CP/ICV ratios that closely followed physiologically plausible developmental patterns. These quantitative trends align with the reductions in FP and FN observed in Figs. 4 and 5. Decreased FP near tissue boundaries explains the lower volumes relative to InfantFS and U-Net, whereas reduced FN accounts for the higher estimates than nnU-Net. This correspondence between voxel-level segmentation accuracy and volumetric outcomes supports the internal validity of the mixed-effects results. It suggests that the proposed approach may provide biologically consistent and less biased quantification across infancy, although direct validation against GT annotations would be required to confirm this. The isointense phase (approximately 6–8 months) is known to be particularly challenging for infant brain MRI analysis due to reduced tissue contrast. Despite this intrinsic difficulty, the proposed model showed stable performance

Table 3 Performance comparisons for LV and CP segmentation in in-house dataset

	LV					CP				
	InfantFS	**BabySeg**	**U-Net**	**nnU-Net**	**Proposed**	**InfantFS + GMM**	**BabySeg + GMM**	**U-Net**	**nnU-Net**	**Proposed**
DSC ↑	0.627 ± 0.078	0.650 ± 0.046	0.692 ± 0.077	0.803 ± 0.046	0.964 ± 0.060	0.501 ± 0.140	0.557 ± 0.086	0.563 ± 0.113	0.561 ± 0.116	0.932 ± 0.059
HD95 ↓	4.534 ± 1.372	13.797 ± 9.362	15.57 ± 7.613	1.749 ± 0.365	0.310 ± 0.542	11.310 ± 7.005	14.819 ± 8.009	12.671 ± 7.744	6.723 ± 3.140	4.148 ± 3.726
ASSD ↓	0.959 ± 0.200	1.410 ± 0.483	1.447 ± 0.503	0.436 ± 0.081	0.088 ± 0.151	1.502 ± 0.769	2.10 ± 00.822	1.746 ± 1.658	1.152 ± 0.450	0.280 ± 0.239
Pearson r	0.415	0.848	0.757	0.846	0.853	0.601	0.711	0.306	0.600	0.763
MAE (mm^3)	2389.196	2651.570	837.562	731.13	191.131	502.555	264.98	453.26	460.52	118.5

trends across age groups, suggesting potential robustness even during this challenging period. To further examine this aspect, representative cases within the isointense age range were additionally evaluated, and the corresponding qualitative and quantitative results are provided in the Supplementary Tables 5, 6, and Supplementary Fig. 1 (See Additional_file 5, 6, and 7). These preliminary observations suggest that the proposed framework may maintain reasonable performance during the isointense phase. Further validation in larger cohorts is needed.

This study has several limitations. First, quantitative evaluation was performed using relatively small test sets, with 10 subjects from each dataset, due to the challenges associated with generating high quality GT labels. Nevertheless, we conducted an age-related analysis by performing inference on 279 MRI scans (not included in model training) spanning a wide age range up to 24 months and qualitatively reviewing the results to assess the practical applicability of the proposed method. Representative results from the public dataset are available at the provided link [https://github.com/jhkang0526/Infant_LVCPseg]. Second, as both two datasets were acquired using similar protocols (0.8 mm isotropic 3D resolution at 3T), further validation is required to establish the generalizability to other vendors, field strengths, or acquisition protocols. Third, our analysis focused on the CP within the LV, as the CP in the third and fourth ventricles is smaller and requires different acquisition protocols optimized for posterior fossa structures. Future work could extend the method to these regions. Additionally, SynthSeg-based synthetic data generation approach [13] may be explored to support fine-grained CP labeling and reduce annotation burden, but this requires careful handling of the CP's small size, complex shape, and imperfect labels. Fourth, this study did not include neonates with major pathological conditions, such as ventricular dilatation. Such abnormalities are clinically important and may be associated with altered CSF dynamics and disrupted neurodevelopment. In this work, we focused on structurally normal neonates to establish a foundational segmentation model. As a qualitative reference, we additionally included inference results from cases with relatively large LV volumes in Supplementary Fig. 2 (See Additional_file 8). These examples suggest that the proposed method can still provide plausible LV and CP segmentations in such cases. However, further validation in pathological cohorts will be necessary to assess the generalizability of the method.

In this study, we presented and evaluated an automated method for joint segmentation of the LV and CP in infant MRI. The proposed method showed improved segmentation performance compared with other approaches and showed a reliable LV boundary delineation in the presence of motion artifacts and subtle tissue contrast at the

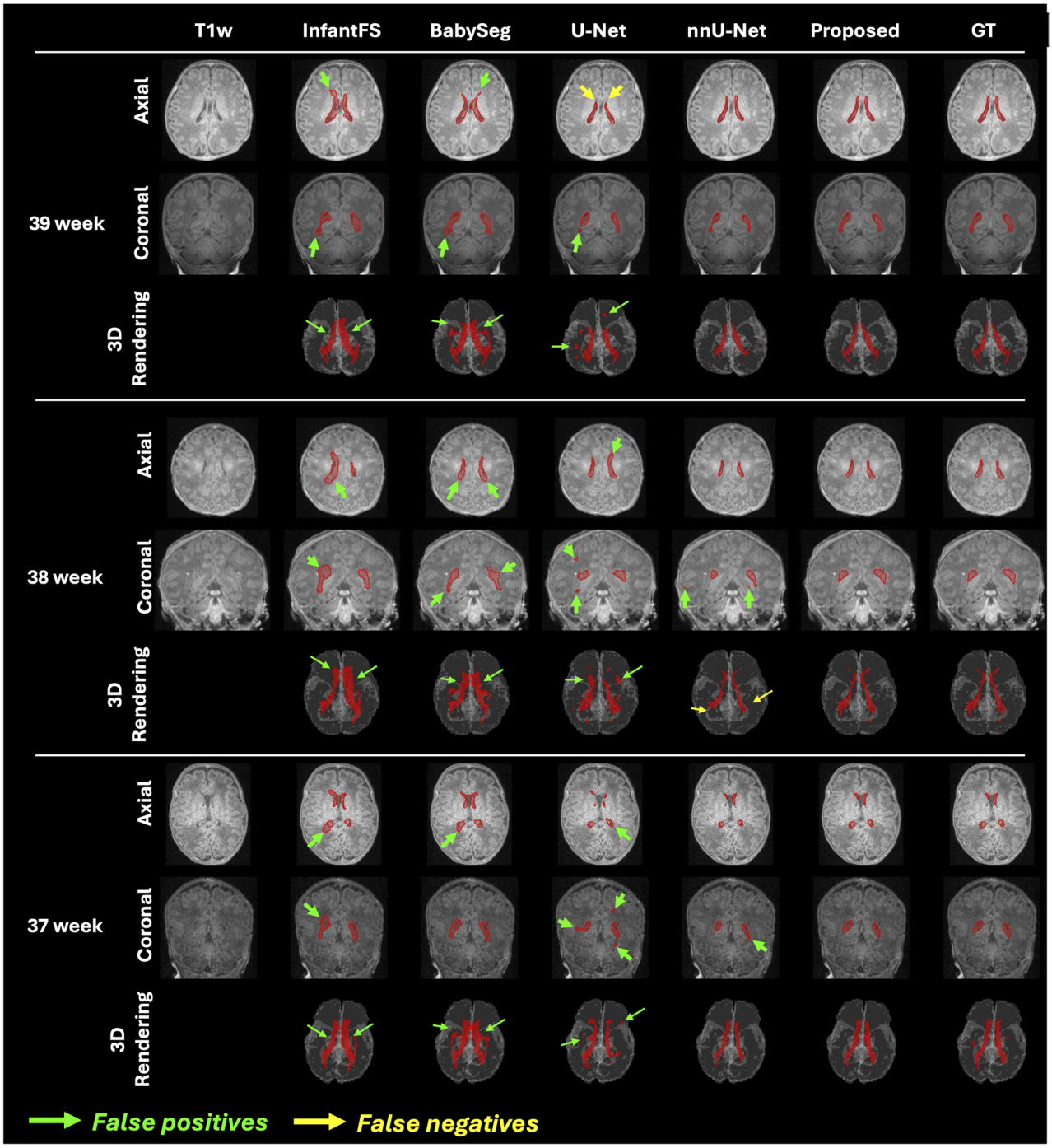


Fig. 7 Visual comparison of the segmentation results of the LV on both 37-, 38-, and 39-week T1w images from in-house dataset, obtained from InfantFS, BabySeg, U-Net, nnU-Net, proposed method, and manual annotation

boundaries. Although further validation across a broader range of imaging protocols and field strengths is warranted, the proposed approach may help reduce manual annotation burden and support large-scale neuroimaging studies. The proposed method could contribute to future investigations of CP-related neurodevelopmental processes and may have potential utility in exploratory clinical risk stratification.

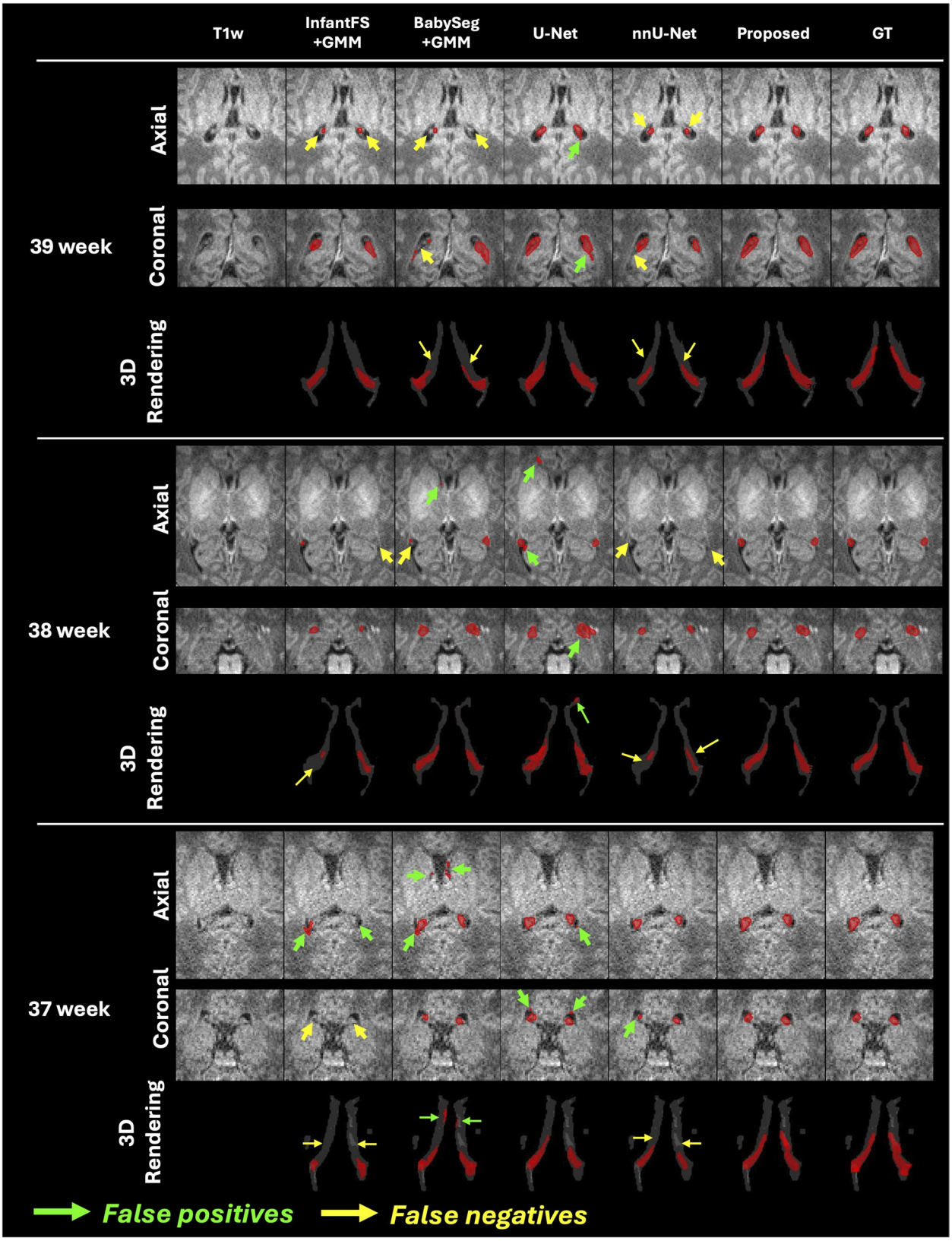


Fig. 8 Visual comparison of the segmentation results of the CP on both 37-, 38-, and 39-week T1w images from in-house dataset, obtained from InfantFS + GMM, BabySeg + GMM, U-Net, nnU-Net, proposed method, and manual annotation

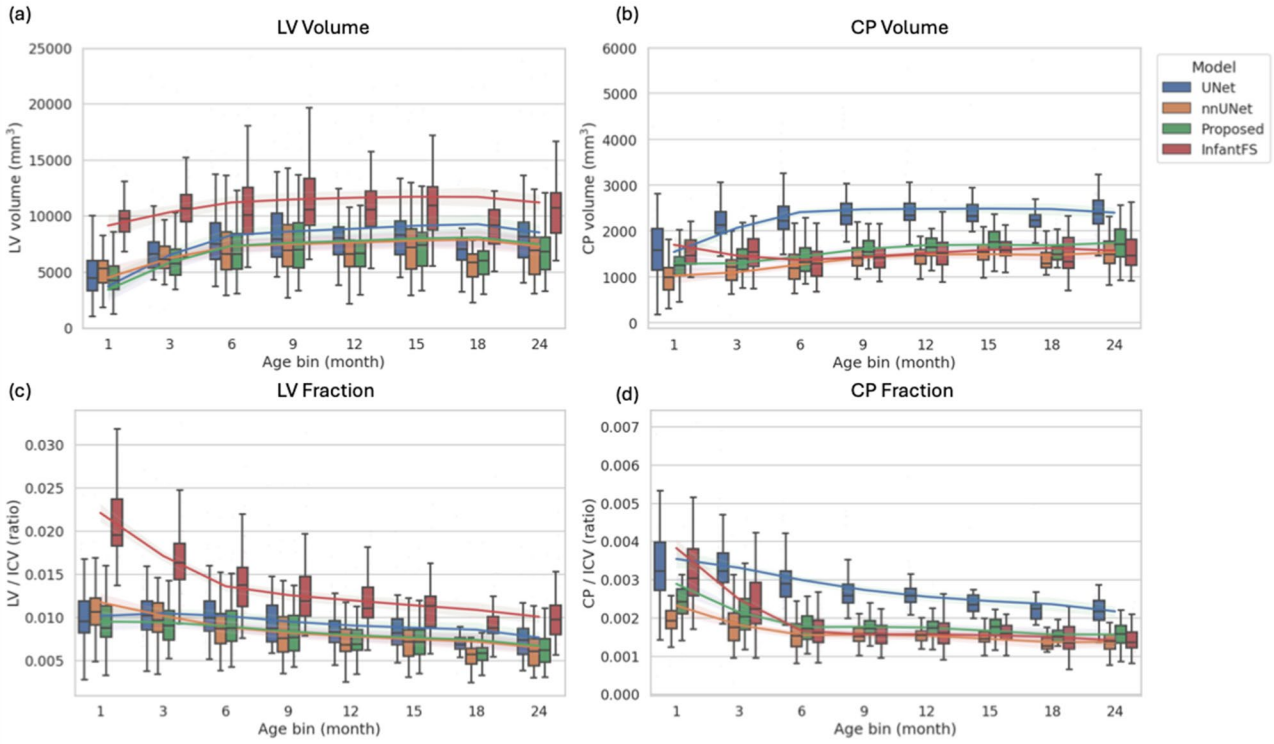


Fig. 9 Boxplots show age-binned distributions (median/IQR; whiskers 1.5×IQR), with individual observations overlaid. Lines and shaded ribbons denote LMM estimated marginal means and 95% CIs (0–25 months). Colors indicate segmentation pipelines

Abbreviations

CP	Choroid plexus
LV	Lateral ventricle
CSF	Cerebrospinal fluid
LMM	Linear mixed-effects models
T1w	T1-weighted
T2w	T2-weighted
GMM	Gaussian mixture models
InfantFS	Infant freesurfer
FP	False positives
FN	False negatives
PA	Postnatal age
BCP	Baby connectome project
GA	Gestational age
DSC	Dice similarity coefficient
HC95	95th-percentile Hausdorff distance
ASSD	Average symmetric surface distance
MAE	Mean absolute error

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-026-02335-x.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

Supplementary Material 9

Acknowledgements

Data utilized in this study were obtained from the Baby Connectome Project, supported by the NIH Blueprint for Neuroscience Research (Grant no. 1U01MH110274).

Author contributions

Junghwa Kang: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. Na-Young Shin: Validation, Investigation, Writing – review & editing. Hyun Gi Kim: Resources, Investigation, Data curation. Yoonho Nam: Supervision, Methodology, Project administration, Writing – review & editing, Funding acquisition, Conceptualization.

Funding

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00248100).

Data availability

Baby Connectome Project Dataset:- The dataset used validation and test in this study was derived from the publicly available BCP dataset (Howell et al., 2019), which is fully anonymized and shared under a data use agreement. https://www.humanconnectome.org/study/lifespan-baby-connectome-project - In-house Dataset (Traveling Dataset):- The dataset used validation and test was obtained from Eunpyeong St. Mary's Hostpital, The Catholic of Korea. Due to privacy and ethical restrictions, this dataset is not publicly available, subject to Institutional Review Board (IRB) approval - The source code will be made publicly available upon publication at (https://github.com/jhkang0526/Infant_LVCPseg). The labels and pretrained model weights derived from public data are available from the corresponding author ([yoonhonam@hufs.ac.kr] (mailto: yoonhonam@hufs.ac.kr)) upon reasonable request.

Declarations

Ethical approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of the Eunpyeong St. Mary's Hospital (approval no. PC24RISI0104). The board waived the requirement for informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Biomedical Engineering, Hankuk University of Foreign Studies, Yongin-si, Gyeonggi-do 17035, Korea
[2]Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul 05505, Republic of Korea
[3]Department of Radiology, Eunpyeong St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Eunpyeong-gu, Seoul, Korea
[4]Department of Radiology and Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Yonsei University College of Medicine, Seoul, Republic of Korea
[5]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Republic of Korea

Received: 20 January 2026 / Accepted: 1 April 2026
Published online: 03 April 2026

References

1. Wolburg H, Paulus W. Choroid plexus: biology and pathology. Acta Neuropathol. 2010;119(1):75–88. https://doi.org/10.1007/s00401-009-0627-8.
2. Redzic ZB, Segal MB. The structure of the choroid plexus and the physiology of the choroid plexus epithelium. Adv Drug Deliv Rev. 2004;56(12):1695–716. https://doi.org/10.1016/j.addr.2004.07.005.
3. Jessen NA, Munk ASF, Lundgaard I, Nedergaard M. The glymphatic system: a beginner's guide. Neurochem Res. 2015;40(12):2583–99. https://doi.org/10.1007/s11064-015-1581-6.
4. Jeong SH, Park CJ, Cha J, Kim SY, Lee SK, Kim YJ, et al. Choroid plexus volume, amyloid burden, and cognition in the Alzheimer's disease continuum. Aging Dis. 2024;16(1):552. https://doi.org/10.14336/AD.2024.0118.
5. Li J, Gao Y, Xu Y, Dai W, Hu Y, Feng X, et al. Morphological changes of the choroid plexus in the lateral ventricle across the lifespan: 5551 subjects from fetus to elderly. NeuroImage. 2025;121392. https://doi.org/10.1016/j.neuroimage.2025.121392.
6. Choi JD, Moon Y, Kim HJ, Yim Y, Lee S, Moon WJ. Choroid plexus volume and permeability at brain MRI within the Alzheimer disease clinical spectrum. Radiology. 2022;304(3):635–45. https://doi.org/10.1148/radiol.212400.
7. Solár P, Zamani A, Kubíčková L, Dubový P, Joukal M. Choroid plexus and the blood–cerebrospinal fluid barrier in disease. Fluids Barriers CNS. 2020;17(1):35. https://doi.org/10.1186/s12987-020-00196-2.
8. Visani V, Pizzini FB, Natale V, Tamanti A, Anglani M, Bertoldo A, et al. Choroid plexus volume in multiple sclerosis can be estimated on structural MRI avoiding contrast injection. Eur Radiol Exp. 2024;8(1):33. https://doi.org/10.1186/s41747-024-00421-9.
9. Storelli L, Pagani E, Rubin M, Margoni M, Filippi M, Rocca MA. A fully automatic method to segment choroid plexuses in multiple sclerosis using conventional MRI sequences. J Magn Reson Imaging. 2024;59(5):1643–52. https://doi.org/10.1002/jmri.28937.
10. Zhou L, Butler TA, Wang XH, Keil SA, Hojjati SH, Hu TW, et al. Region-informed machine learning model for choroid plexus segmentation in Alzheimer's disease. Front Aging Neurosci. 2025;17:1613320. https://doi.org/10.3389/fnagi.2025.1613320.
11. Tadayon E, Moret B, Sprugnoli G, Monti L, Pascual-Leone A, Santarnecchi E, et al. Improving choroid plexus segmentation in the healthy and diseased brain: relevance for tau-PET imaging in dementia. J Alzheimers Dis. 2020;74(4):1057–68. https://doi.org/10.3233/JAD-190706.
12. Fischl B, FreeSurfer. NeuroImage. 2012;62(2):774–81. https://doi.org/10.1016/j.neuroimage.2012.01.021.
13. Billot B, Greve DN, Puonti O, Thielscher A, Van Leemput K, Fischl B, et al. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. Med Image Anal. 2023;86:102789. https://doi.org/10.1016/j.media.2023.102789.
14. Zalevskyi V, Sanchez T, Roulet M, Aviles Verdera J, Hutter J, Kebiri H, et al. Improving cross-domain brain tissue segmentation in fetal MRI with synthetic data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention; 2024. pp. 437 – 47. https://doi.org/10.1007/978-3-031-72089-5_44.
15. Zhao L, Feng X, Meyer CH, Alsop DC. Choroid plexus segmentation using optimized 3D U-Net. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). 2020. pp. 381–384. https://doi.org/10.1109/ISBI45749.2020.9098443.
16. Visani V, Veronese M, Pizzini FB, Colombi A, Natale V, Marjin C, et al. ASCHOPLEX: A generalizable approach for the automatic segmentation of choroid plexus. Comput Biol Med. 2024;182:109164. https://doi.org/10.1016/j.compbiomed.2024.109164.
17. Wang X, Wang X, Yan Z, Yin F, Li Y, Liu X, et al. Enhanced choroid plexus segmentation with 3D UX-Net and its association with disease progression in multiple sclerosis. Mult Scler Relat Disord. 2024;88:105750. https://doi.org/10.1016/j.msard.2024.105750.
18. Eisma JJ, McKnight CD, Hett K, Elenberger J, Han CJ, Song AK, et al. Deep learning segmentation of the choroid plexus from structural magnetic resonance imaging (MRI): validation and normative ranges across the adult lifespan. Fluids Barriers CNS. 2024;21(1):21. https://doi.org/10.1186/s12987-024-00525-9.
19. Makropoulos A, Gousias IS, Ledig C, Aljabar P, Serag A, Hajnal JV, et al. Automatic whole brain MRI segmentation of the developing neonatal brain. IEEE Trans Med Imaging. 2014;33(9):1818–31. https://doi.org/10.1109/TMI.2014.2322280.
20. Wang L, Nie D, Li G, Puybareau E, Dolz J, Zhang Q, et al. Benchmark on automatic six-month-old infant brain segmentation algorithms: the iSeg-2017 challenge. IEEE Trans Med Imaging. 2019;38(9):2219–30. https://doi.org/10.1109/TMI.2019.2901712.
21. Habas PA, Kim K, Rousseau F, Glenn OA, Barkovich AJ, Studholme C. Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses. Hum Brain Mapp. 2010;31(9):1348–58. https://doi.org/10.1002/hbm.20935.
22. Wang Y, Haghpanah FS, Zhang X, Santamaria K, da Costa Aguiar Alves GK, Bruno E, et al. ID-Seg: an infant deep learning-based segmentation framework to improve limbic structure estimates. Brain Inf. 2022;9(1):12. https://doi.org/10.1186/s40708-022-00161-9.
23. Zöllei L, Iglesias JE, Ou Y, Grant PE, Fischl B. Infant FreeSurfer: An automated segmentation and surface extraction pipeline for T1-weighted neuroimaging data of infants 0–2 years. NeuroImage. 2020;218:116946. https://doi.org/10.1016/j.neuroimage.2020.116946.
24. Hoffmann M, Zöllei L, Dalca AV. Deep infant brain segmentation from multi-contrast MRI. arXiv preprint arXiv:2512.05114. 2025. https://doi.org/10.48550/arXiv.2512.05114.
25. Barkovich MJ, Williams C, Barkovich AJ. Technical and practical tips for performing brain magnetic resonance imaging in premature neonates. Semin Perinatol. 2021;45(7):151468. https://doi.org/10.1016/j.semperi.2021.151468.
26. Barkovich AJ. Pediatric Neuroimaging. 6th ed. Philadelphia: Lippincott Williams & Wilkins; 2018.
27. Martinez-Biarge M, Arnaez J, Arca G, Valverde E, Llorens-Salvador R, García-Alix A, et al. Recommendations for the use of brain MRI in the neonatal period. Pediatr (Engl Ed). 2025;103(3):503–935. https://doi.org/10.1016/j.anpede.2025.503935.
28. Kuklisova-Murgasova M, Quaglierini C, Claret N, Lacerda HM, Mattmann J, Martin L, et al. Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. Med Image Anal. 2012;16(8):1550–64. https://doi.org/10.1016/j.media.2012.08.001.
29. Xu J, Moyer D, Gagoski B, Iglesias JE, Grant PE, Golland P, et al. NeSVoR: implicit neural representation for slice-to-volume reconstruction in MRI. IEEE Trans Med Imaging. 2023;42(6):1707–19. https://doi.org/10.1109/TMI.2023.3239165.
30. Prayssac A, Giraud R, Boiger C, Gori A, Lombaert H, Dardenne E, et al. NiftyMIC - A PyTorch implementation of a state-of-the-art slice-to-volume registration and super-resolution reconstruction method. GitHub Repository. 2020. Available from: https://github.com/gift-surg/NiftyMIC.

31. Howell BR, Styner MA, Gao W, Yap PT, Wang L, Baluyot K, et al. The UNC/UMN Baby Connectome Project (BCP): An overview of the study design and protocol development. NeuroImage. 2019;185:891–905. https://doi.org/10.1016/j.neuroimage.2018.03.049.
32. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2016. pp. 424–432. https://doi.org/10.48550/arXiv.1606.06650.
33. Gros C, Lemay A, Cohen-Adad J, SoftSeg. Advantages of soft versus binary training for image segmentation. Med Image Anal. 2021;71:102038. https://doi.org/10.1016/j.media.2021.102038.
34. Zhang C, Guan W, Xing X, Yang G. Preserving cardiac integrity: a topology-infused approach to whole heart segmentation. MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images. Springer; 2024. pp. 24–33. https://doi.org/10.48550/arXiv.2410.10551.
35. Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv Preprint. 2017. https://doi.org/10.48550/arXiv.1711.05101. arXiv:1711.05101.
36. 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):203–11. https://doi.org/10.1038/s41592-020-01008-z.

Publisher's note