A Modular Multi-Model Framework for Multi-Structure Segmentation of Ultra-Low-Field Pediatric Brain MRI

Chun Cao[1] and Lyuyang Tong[1*✉] and Bo Du[1*✉]

School of Computer Science, National Engineering Research Center for Multimedia Software, Institute of Artificial Intelligence, and Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University, Wuhan, China.
[*]Corresponding authors: Lyuyangtong@whu.edu.cn, dubo@whu.edu.cn

Abstract. Accurately characterizing structural changes in the developing human brain during early life is essential for understanding healthy development and identifying neurodevelopmental disorders. However, existing deep learning segmentation tools designed for brain MRI do not generalize well to pediatric brain segmentation, and the problem is further amplified in low-field MRI because of its lower image quality. To address these challenges, we propose a modular multi-model framework for multi-structure segmentation of ultra-low-field pediatric brain MRI, which consists of a heterogeneous segmentation ensemble (HSE) with four complementary multi-class segmentation models, an anatomical prior refinement (APR) module with four auxiliary binary segmentation models, and an anatomical consistency post-processing (ACP) module for hard-label refinement. Based on out-of-fold segmentation results from five-fold cross-validation on 79 labeled cases, the HSE achieved a mean Dice score of 0.793788 across the 11 structures. Incorporating APR increased the mean Dice score to 0.796778, while the final ACP module further improved it to 0.807148.

Keywords: Ultra-low-field MRI · Pediatric brain segmentation · nnU-Net · Multi-model fusion · Multi-structure segmentation

1 Introduction

Portable low-field MRI systems reduce infrastructure requirements and can extend neuroimaging to bedside and resource-constrained settings [14,13,1]. The accessibility is particularly relevant to pediatric neuroimaging, where patient transport, scanner availability, and acquisition burden may limit the use of conventional high-field MRI systems [4]. Automated anatomical segmentation can support reproducible volumetric analysis and downstream clinical or developmental studies. However, the image characteristics of low-field MRI differ substantially from those of standard clinical scans, making direct application of existing segmentation tools challenging.

LISA 2026 Task 2 focuses on automatic multi-structure segmentation of early-childhood brain MRI acquired using ultra-low-field 0.064 T T2-weighted imaging [10]. The task requires segmentation of 11 anatomical structures, including the bilateral hippocampi, ventricles, caudate nuclei, lentiform nuclei, thalami, and the corpus callosum. Under the ultra-low-field acquisition setting, reduced signal-to-noise ratio, limited spatial resolution, weak tissue contrast, and acquisition artifacts make structural boundaries difficult to identify [8,2]. The challenge is particularly pronounced for small and bilateral structures, where accurate delineation and correct left–right semantic assignment are both required. Related brain MRI studies have likewise highlighted the difficulty of segmentation under rapidly changing anatomy and low or ambiguous tissue contrast [20,5].

Encoder–decoder architectures such as U-Net and 3D U-Net have become standard approaches for volumetric medical image segmentation [17,3]. nnU-Net further improves practical applicability by automatically adapting preprocessing, network configuration, and training strategies to the target dataset [6,7]. Nevertheless, a single multi-class model may remain sensitive to ambiguous boundaries and low-confidence predictions in ultra-low-field MRI. Deep ensembles have been shown to improve calibration and reduce occasional severe failures in medical image segmentation [15,16], and ensemble-based strategies have also been investigated for infant brain MRI segmentation [5]. These observations motivate combining complementary architectures with structure-specific anatomical priors.

To address these challenges, we propose a modular multi-model framework consisting of three sequential modules: heterogeneous segmentation ensemble (HSE), anatomical prior refinement (APR), and anatomical consistency post-processing (ACP). HSE combines four complementary multi-class segmentation models: the default nnU-Net v2 3D full-resolution PlainConvUNet (PConv), a large-patch PlainConvUNet (LP-PConv), a medium ResidualEncoderUNet (ResEnc-M), and the Transformer-centric Primus architecture [19]. The probability maps are fused to obtain the primary 12-class prediction for the background and 11 anatomical structures. The architectural and contextual diversity among these models provides complementary representations for the subsequent refinement stages. APR further calibrates the HSE probability tensor using four auxiliary binary anatomical priors. Foreground prior modulation (FPM) uses an all-structure foreground prediction to reduce excessive background probability. Hippocampal-union prior refinement (HUPR) and ventricular-union prior refinement (VUPR) adjust the total probability masses of the bilateral hippocampi and ventricles while retaining their left–right proportions. Callosal prior fusion (CPF) incorporates an auxiliary corpus-callosum probability into the corresponding multi-class channel. All APR operations are performed at the probability level before hard-label generation. ACP operates on the resulting hard-label prediction and consists of left–right consistency correction (LRCC) and relative connected-component filtering (RCCF). LRCC enforces consistent side identities for the five bilateral anatomical pairs according to the NIfTI left–right orientation, while RCCF removes relatively small disconnected components

using a label-wise relative size criterion. Overall, the operations provide deterministic correction of severe anatomical inconsistencies and isolated false-positive regions. Our contributions are summarized as follows:

- We propose a modular HSE–APR–ACP framework for multi-structure segmentation of ultra-low-field pediatric brain MRI, improving the mean Dice score from 0.788484 to 0.807148 over all 11 structures.
- We combine heterogeneous multi-class models with structure-specific anatomical priors to improve probability-level segmentation accuracy and robustness.
- We introduce deterministic anatomical consistency post-processing to correct left–right semantic errors and remove small disconnected components.

2 Method

As illustrated in Fig. 1, the proposed framework contains three sequential modules. The first module is a heterogeneous segmentation ensemble (HSE), which combines four complementary multi-class segmentation models to generate the primary probability tensor with 12 output classes. The second module is anatomical prior refinement (APR), which uses four auxiliary binary anatomical priors to refine the ensemble probabilities. The third module is anatomical consistency post-processing (ACP), which applies deterministic hard-label refinement to reduce left–right semantic inconsistencies and small isolated false-positive components.

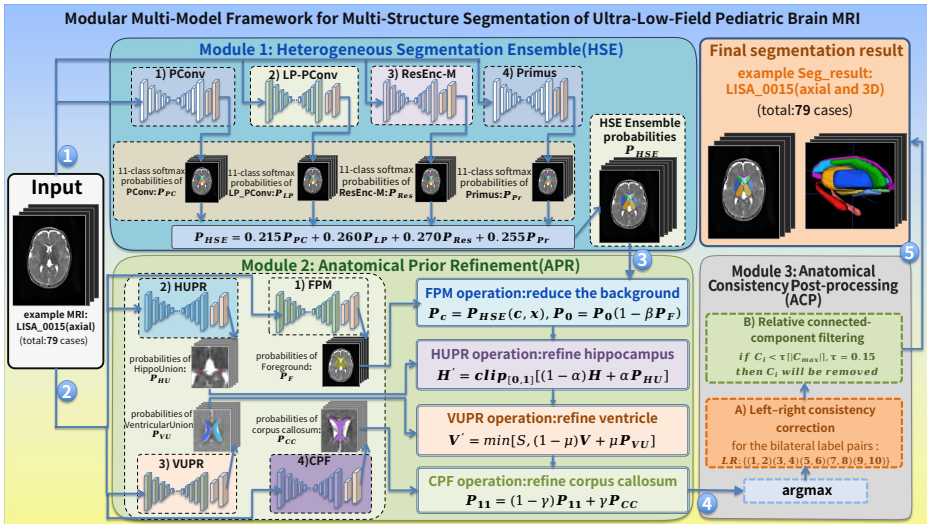

Fig. 1. Overview of the proposed modular multi-model segmentation framework

2.1 Module 1: Heterogeneous Segmentation Ensemble

Table 1. Model configurations used in HSE.

Model	Configuration	Patch size	Checkpoint selection
PConv	nnU-Net v2 3D full-resolution PlainConvUNet with the default plan	$112 \times 160 \times 128$	final (F0–F4)
LP-PConv	PlainConvUNet with an enlarged patch size in the 3D full-resolution plan	$160 \times 160 \times 128$	best (F2); final (others)
ResEnc-M	nnU-Net v2 medium ResidualEncoderUNet plan [7]	auto-configured	final (F0–F4)
Primus	PrimusV3S with the nnU-Net v2 3D full-resolution pipeline [19]	$112 \times 160 \times 128$	final (F0–F4)

Table 1 summarizes the architecture, patch size, and checkpoint selection of each multi-class model used in HSE. HSE combines the spatially aligned 12-channel softmax probability tensors produced by the four multi-class models:

$$P_{\mathrm{HSE}}(c,\mathbf{x}) = w_1 P_{\mathrm{PC}}(c,\mathbf{x}) + w_2 P_{\mathrm{LP}}(c,\mathbf{x}) + w_3 P_{\mathrm{Res}}(c,\mathbf{x}) + w_4 P_{\mathrm{Pr}}(c,\mathbf{x}), \\ (w_1, w_2, w_3, w_4) = (0.215,\ 0.260,\ 0.270,\ 0.255). \tag{1}$$

where $\mathbf{x}$ represents a voxel, $c \in \{0, 1, \ldots, 11\}$ denotes the background or one of the 11 anatomical classes, and P_{PC}, P_{LP}, P_{Res}, and P_{Pr} are the corresponding model probabilities. The fusion weights were selected using five-fold out-of-fold evaluation over all 11 foreground structures and were fixed for all subsequent experiments. The resulting probability tensor P_{HSE} is passed to APR.

2.2 Module 2: Anatomical Prior Refinement

The anatomical prior refinement (APR) module sequentially calibrates the HSE probability tensor using four auxiliary binary nnU-Net v2 PlainConvUNet models trained with structure-specific targets. FPM reduces excessive background competition using class-agnostic foreground evidence. HUPR and VUPR refine the bilateral hippocampal and ventricular probability masses while preserving left–right identity; these structures were among the more challenging targets in the out-of-fold analysis, with the hippocampi showing the lowest accuracy. CPF further refines the relatively difficult corpus callosum. Algorithm 1 summarizes the complete APR procedure.

Algorithm 1 Anatomical Prior Refinement

Require: HSE probabilities P_{HSE}; auxiliary probabilities P_F, P_{HU}, P_{VU}, and P_{CC}

Ensure: Refined probability tensor P_{APR}

$P_c \leftarrow P_{\mathrm{HSE}}(c, \mathbf{x}), \quad c = 0, \ldots, 11$

FPM: Foreground prior modulation

$P_0 \leftarrow P_0(1 - \beta P_F), \qquad \beta = 0.85$

HUPR: Hippocampal-union prior refinement

$H \leftarrow P_1 + P_2$

$H' \leftarrow \mathrm{clip}_{[0,1]}((1-\alpha)H + \alpha P_{\mathrm{HU}}), \qquad \alpha = 0.52$

$r_L \leftarrow \dfrac{P_1}{\max(H, \varepsilon)}, \qquad r_R \leftarrow 1 - r_L$

If $H \le \varepsilon$, set $r_L = r_R = 0.5$.

$P_c \leftarrow P_c \dfrac{1 - H'}{\max(1 - H, \varepsilon)}, \qquad c \notin \{1, 2\}$

$(P_1, P_2) \leftarrow (r_L H',\, r_R H')$

VUPR: Ventricular-union prior refinement

$V \leftarrow P_3 + P_4, \qquad S \leftarrow \sum_{k=0}^{11} P_k$

$V' \leftarrow \min(S, (1-\mu)V + \mu P_{\mathrm{VU}}), \qquad \mu = 0.17$

$r_L^V \leftarrow \dfrac{P_3}{\max(V, \varepsilon)}, \qquad r_R^V \leftarrow 1 - r_L^V$

If $V \le \varepsilon$, set $r_L^V = r_R^V = 0.5$.

$P_c \leftarrow P_c \dfrac{S - V'}{\max(S - V, \varepsilon)}, \qquad c \notin \{3, 4\}$

$(P_3, P_4) \leftarrow (r_L^V V',\, r_R^V V')$

CPF: Callosal prior refinement

$P_{11} \leftarrow (1-\gamma)P_{11} + \gamma P_{\mathrm{CC}}, \qquad \gamma = 0.34$

$P_{\mathrm{APR}}(c, \mathbf{x}) \leftarrow P_c, \qquad c = 0, \ldots, 11$

where P_c denotes the working class probability initialized from P_{HSE} and sequentially updated by APR. P_F, P_{HU}, P_{VU}, and P_{CC} are the auxiliary probabilities for the all-structure foreground, hippocampal union, ventricular union, and corpus callosum, respectively. H and V denote the original hippocampal- and ventricular-union probability masses, while H' and V' denote their refined values; S is the total probability mass before VUPR. r_L, r_R, r_L^V, and r_R^V preserve the original left–right proportions of the hippocampi and ventricles, and $\varepsilon = 10^{-8}$ is used for numerical stability. The resulting P_{APR} is passed to Module 3 for hard-label generation and anatomical consistency post-processing.

2.3 Module 3: Anatomical Consistency Post-processing

The anatomical consistency post-processing (ACP) module converts the refined APR probabilities into a hard-label segmentation and subsequently applies two deterministic operations. LRCC enforces consistent side identities for the five bilateral anatomical pairs, while RCCF removes relatively small disconnected components. Algorithm 2 summarizes the complete ACP procedure.

Algorithm 2 Anatomical Consistency Post-processing

Require: Refined probability tensor P_{APR}

Ensure: Final segmentation $\hat{Y}_{\mathrm{final}}$

$\hat{Y}_{\mathrm{int}}(\mathbf{x}) \leftarrow \arg\max_{c\in\{0,\dots,11\}} P_{\mathrm{APR}}(c,\mathbf{x})$

$\hat{Y} \leftarrow \mathrm{TransposeToNIfTI}(\hat{Y}_{\mathrm{int}})$

LRCC: Left–right consistency correction

$\mathcal{LR} \leftarrow \{(1,2),(3,4),(5,6),(7,8),(9,10)\}$

$m_x \leftarrow \dfrac{N_x-1}{2}$

for all $(l,r) \in \mathcal{LR}$ **do**

  if $\hat{Y}(\mathbf{x}) \in \{l,r\}, \quad \hat{Y}(\mathbf{x}) \leftarrow \begin{cases} l, & x < m_x, \\ r, & x \geq m_x \end{cases}$

end for

RCCF: Relative connected-component filtering

for $c = 1,\dots,11$ **do**

  $\{C_i\} \leftarrow \mathrm{CC}_{26}(\hat{Y}=c)$

  if $\{C_i\} \neq \emptyset$ **then**

    $C_{\max} \leftarrow \arg\max_{C_i} |C_i|$

    remove each C_i **if** $|C_i| < \lceil \tau |C_{\max}| \rceil, \quad \tau = 0.15$

  end if

end for

$\hat{Y}_{\mathrm{final}} \leftarrow \hat{Y}$

where $\hat{Y}_{\mathrm{int}}$ represents the hard-label prediction in the internal inference axis order. $\hat{Y}$ denotes the prediction after conversion to the NIfTI axis order, and $\mathcal{LR}$ contains the five bilateral label pairs. N_x is the image size along the NIfTI x-axis and m_x is its midpoint. C_i denotes a 3D connected component obtained using 26-connectivity, $C_{\max}$ is the largest component of the same label, and $|C_i|$ denotes its size in voxels. Components smaller than $\lceil \tau |C_{\max}| \rceil$ are assigned to the background.

3 Experiments and Results

3.1 Dataset Description

Our experiments use the LISA 2026 Task 2 dataset for multi-structure segmentation of pediatric ultra-low-field brain MRI [9,10]. Each input is a combined isotropic (CISO) T2-weighted volume derived from orthogonal 0.064 T Hyperfine Swoop acquisitions and registered to the corresponding subject-matched high-field scan. The high-field images are not released; the reference masks were delineated on the matched high-field volumes and reviewed by an expert medical image evaluator. The dataset is divided as follows:

– **Cross-validation dataset:** 79 CISO volumes with 11-label reference segmentations. A fixed five-fold split was shared by all model configurations.

Each subject contributed one out-of-fold segmentation result from models that excluded that subject during training.
– **Validation set:** 12 CISO volumes without public reference labels, evaluated through the official challenge server.

The label definitions are background (0), left and right hippocampi (1, 2), left and right ventricles (3, 4), left and right caudate nuclei (5, 6), left and right lentiform nuclei (7, 8), left and right thalami (9, 10), and corpus callosum (11). The submitted masks are required to preserve the input geometry and contain all 11 foreground labels [11]. In this study, all analyses and ablation experiments are reported over the 11 foreground structures.

3.2 Experimental Setup

All models were implemented in Python using PyTorch and trained on NVIDIA GeForce RTX 4090 GPUs. The nnU-Net-based models followed the preprocessing, data augmentation, optimization, deep-supervision, and sliding-window inference procedures of nnU-Net v2 [6,7]. All segmentation networks used in the final framework were trained for 1000 epochs.

A fixed five-fold cross-validation split was used for the 79 labeled cases. Folds 0–3 contained 16 validation cases each and fold 4 contained 15, with the remaining cases used for training. The validation cases in each fold are listed in Table 2.

Table 2. Five-fold cross-validation split of the 79 labeled cases. All case IDs are prefixed with "LISA_".

Fold	Validation case IDs
F0	0003, 0005, 0007, 0014, 0021, 0029, 0031, 0038, 0041, 0043, 0045, 0052, 0054, 0063, 1009, 1010
F1	0010, 0019, 0022, 0023, 0026, 0027, 0036, 0040, 0049, 0051, 0058, 0061, 0062, 1001, 1004, 1005
F2	0009, 0013, 0017, 0020, 0025, 0033, 0046, 0047, 0050, 0053, 0055, 0059, 1003, 1006, 1012, 1013
F3	0001, 0004, 0006, 0016, 0018, 0028, 0032, 0034, 0048, 0056, 0057, 1002, 1007, 1008, 1011, 1014
F4	0002, 0008, 0011, 0012, 0015, 0024, 0030, 0035, 0037, 0039, 0042, 0044, 0060, 1015, 1016

The same fold assignment was used across the multi-class and auxiliary models, and all reported results on the labeled dataset were obtained from five-fold out-of-fold predictions. The evaluation metrics included the Dice similarity coefficient (DSC), Hausdorff distance (HD), 95th-percentile Hausdorff distance (HD95), average symmetric surface distance (ASSD), and absolute relative volume error (|RVE|) [18,12]. The primary endpoint was the mean DSC across all 11 foreground structures.

3.3 Experimental Results

We conducted a factorial ablation study on the 79 labeled cases using five-fold out-of-fold predictions. The evaluated factors were HSE, APR, and ACP. To evaluate the contribution of HSE, PConv, the best-performing individual multi-class model, was used as the single-model baseline. Table 3 presents the complete $2 \times 2 \times 2$ ablation results.

Table 3. Factorial ablation study on the 79 labeled cases using five-fold out-of-fold predictions. HD, HD95, and ASSD are reported in millimeters. NF denotes the number of cases containing at least one non-finite surface-distance measurement due to an empty structure prediction.

Probability predictor	APR	ACP	DSC ↑	HD ↓	HD95 ↓	ASSD ↓	|RVE| ↓	NF ↓
Single model (PConv)	Off	Off	0.788484	∞	∞	∞	0.145554	1
Single model (PConv)	**On**	Off	0.792564	∞	∞	∞	0.149951	1
Single model (PConv)	Off	**On**	0.799568	3.863144	1.829182	0.766669	0.131351	0
Single model (PConv)	**On**	**On**	0.803908	3.829010	1.795209	0.755401	0.134133	0
HSE	Off	Off	0.793788	∞	∞	∞	0.134447	1
HSE	**On**	Off	0.796778	5.542447	2.489122	0.903125	0.138869	0
HSE	Off	**On**	0.803884	3.806996	1.787335	0.749596	**0.126467**	0
HSE	**On**	**On**	**0.807148**	**3.784127**	**1.775122**	**0.742050**	0.130703	0

The complete HSE–APR–ACP framework achieved the best overall performance, improving the mean DSC from 0.788484 for the best single model to 0.807148. The final configuration also obtained the lowest HD, HD95, and ASSD among all evaluated variants.

Both HSE and APR provided consistent improvements. Under the complete processing setting, HSE improved the mean DSC by 0.003240, while APR improved it by 0.003264. ACP produced a larger average gain of 0.010370, mainly by correcting a small number of severe left–right anatomical inconsistencies.

Non-finite surface-distance values occurred in one case for several configurations without ACP, due to empty structure predictions. Overall, the results demonstrate that HSE and APR improve segmentation accuracy, while ACP enhances robustness against severe anatomical errors.

Table 4. Per-structure DSC of the final HSE–APR–ACP framework on the 79 labeled cases using five-fold out-of-fold predictions.

Structure	DSC	Structure	DSC
Hippocampus L	0.668741	Hippocampus R	0.684491
Ventricle L	0.805547	Ventricle R	0.790113
Caudate L	0.833272	Caudate R	0.827619
Lentiform L	0.855735	Lentiform R	0.852928
Thalamus L	0.898463	Thalamus R	0.900293
Corpus callosum	0.761429	**Mean**	**0.807148**

Table 4 summarizes the per-structure DSC of the final HSE–APR–ACP framework on the 79 labeled cases using five-fold out-of-fold predictions. The

final model achieved the highest DSCs for the bilateral thalami, whereas the bilateral hippocampi remained the most challenging structures.

Figure 2 shows the qualitative segmentation result for case LISA_0015 in axial, coronal, sagittal, and 3D views.

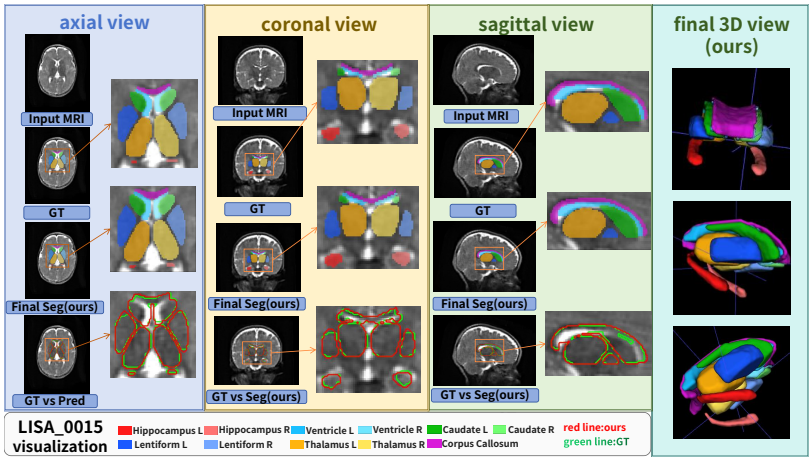

Fig. 2. Qualitative visualization of the final segmentation results.

4 Discussion

The proposed framework achieved improved segmentation performance by combining complementary multi-class models with structure-specific anatomical refinement. The ablation study showed that both HSE and APR consistently improved the mean Dice score, while the complete HSE–APR–ACP pipeline increased the mean Dice from 0.788484 for the best single-model baseline to 0.807148. These results support the use of heterogeneous model fusion and anatomical priors for reducing uncertainty in ultra-low-field MRI segmentation, where image contrast and structural boundaries can be ambiguous. The benefit of HSE is also consistent with previous observations that deep ensembles can improve robustness in medical image segmentation [15,16].

The three modules serve distinct but complementary roles. HSE strengthens the primary probability representation by exploiting architectural diversity among the multi-class models. APR performs probability-level refinement using foreground and structure-specific anatomical priors while preserving the multi-class prediction framework. ACP provides hard-label anatomical correction for infrequent but severe errors, particularly left–right semantic inconsistencies, and conservatively removes small disconnected components. The improvement in the mean score therefore arises mainly from the correction of a limited number of failure cases, whereas anatomically consistent predictions remain largely unchanged.

Several limitations remain. First, the available labeled dataset contains only 79 cases, and the current parameter selection was therefore based on five-fold

out-of-fold evaluation within this dataset. Second, the multi-model framework increases inference cost compared with a single segmentation network. In addition, as shown in Table 4, the hippocampi remain the most challenging structures in the final results, suggesting that further improvements may require stronger structure-specific modeling. Future work could investigate more efficient model fusion and improved anatomical priors for these difficult structures while maintaining the robustness of the current framework.

Acknowledgements This work was supported in part by the National Key Research and Development Program of China under grants (2023YFC2705705 and 2023YFC2705701), the National Natural Science Foundation of China under grants 62306217 and 62225113, the Innovative Research Group Project of Hubei Province under grants 2024AFA017. This work has been supported by the New Cornerstone Science Foundation through the XPLORER PRIZE. This work was supported by WHU-Kingsoft Joint Lab. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Disclosure of Interests. The authors declare no competing interests relevant to the content of the article.

References

1. Arnold, T.C., Freeman, C.W., Litt, B., Stein, J.M.: Low-field MRI: Clinical promise and challenges. Journal of Magnetic Resonance Imaging **57**(1), 25–44 (2023). https://doi.org/10.1002/jmri.28408
2. de Leeuw den Bouter, M.L., Ippolito, G., O'Reilly, T.P.A., Remis, R.F., van Gijzen, M.B., Webb, A.G.: Deep learning-based single image super-resolution for low-field MR brain images. Scientific Reports **12**, 6362 (2022). https://doi.org/10.1038/s41598-022-10298-6
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Lecture Notes in Computer Science, vol. 9901, pp. 424–432. Springer (2016). https://doi.org/10.1007/978-3-319-46723-8_49
4. Deoni, S.C.L., Bruchhage, M.M.K., Beauchemin, J., Volpe, A., D'Sa, V., Huentelman, M., Williams, S.C.R.: Accessible pediatric neuroimaging using a low field strength MRI scanner. NeuroImage **238**, 118273 (2021). https://doi.org/10.1016/j.neuroimage.2021.118273
5. Dolz, J., Desrosiers, C., Wang, L., Yuan, J., Shen, D., Ben Ayed, I.: Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. Computerized Medical Imaging and Graphics **79**, 101660 (2020). https://doi.org/10.1016/j.compmedimag.2019.101660
6. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z

7. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jäger, P.F.: nnU-Net Revisited: A call for rigorous validation in 3d medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Lecture Notes in Computer Science, vol. 15009, pp. 488–498. Springer (2024). https://doi.org/10.1007/978-3-031-72114-4_47
8. Koonjoo, N., Zhu, B., Bagnall, G.C., Bhutto, D., Rosen, M.S.: Boosting the signal-to-noise of low-field MRI with deep learning image reconstruction. Scientific Reports **11**, 8248 (2021). https://doi.org/10.1038/s41598-021-87482-7
9. LISA 2026 Organizing Committee: LISA 2026 data summary (2026), https://www.synapse.org/Synapse:syn72118611/wiki/640472, accessed 1 July 2026
10. LISA 2026 Organizing Committee: LISA 2026 Task 2: Segmentation (2026), https://www.synapse.org/Synapse:syn72118611/wiki/637244, accessed 1 July 2026
11. LISA 2026 Organizing Committee: LISA 2026 Task 2 submission formatting (2026), https://www.synapse.org/Synapse:syn72118611/wiki/640942, accessed 1 July 2026
12. LISA 2026 Organizing Committee: LISA 2026 task evaluation metrics (2026), https://www.synapse.org/Synapse:syn72118611/wiki/637246, accessed 1 July 2026
13. Liu, Y., Leong, A.T.L., Zhao, Y., et al.: A low-cost and shielding-free ultra-low-field brain MRI scanner. Nature Communications **12**, 7238 (2021). https://doi.org/10.1038/s41467-021-27317-1
14. Mazurek, M.H., Cahn, B.A., Yuen, M.M., et al.: Portable, bedside, low-field magnetic resonance imaging for evaluation of intracerebral hemorrhage. Nature Communications **12**, 5119 (2021). https://doi.org/10.1038/s41467-021-25441-6
15. Mehrtash, A., Wells, W.M., Tempany, C.M., Abolmaesumi, P., Kapur, T.: Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. IEEE Transactions on Medical Imaging **39**(12), 3868–3878 (2020). https://doi.org/10.1109/TMI.2020.3006437
16. Petrov, Y., Malik, B., Fredrickson, J., Jemaa, S., Carano, R.A.D.: Deep ensembles are robust to occasional catastrophic failures of individual DNNs for organs segmentations in CT images. Journal of Digital Imaging **36**, 2060–2074 (2023). https://doi.org/10.1007/s10278-023-00857-2
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Taha, A.A., Hanbury, A.: Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Medical Imaging **15**, 29 (2015). https://doi.org/10.1186/s12880-015-0068-x
19. Wald, T., Roy, S., Isensee, F., Ulrich, C., Ziegler, S., Trofimova, D., Stock, R., Baumgartner, M., Köhler, G., Maier-Hein, K.: Primus: Enforcing attention usage for 3d medical image segmentation. Transactions on Machine Learning Research (2026), https://openreview.net/forum?id=x4vZE4PDEu
20. Wang, L., Wu, Z., Chen, L., Sun, Y., Lin, W., Li, G.: iBEAT V2.0: A multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction. Nature Protocols **18**(5), 1488–1509 (2023). https://doi.org/10.1038/s41596-023-00806-x