Spatial Prior-Guided Boundary and Region-Aware 2D Lesion Segmentation in Neonatal Hypoxic Ischemic Encephalopathy

Amog Rao[1⋆[0009−0000−4877−6834]], Ananya Shukla[1⋆[0009−0006−4484−8224]], Jia Bhargava[1], Yangming Ou[2], and Rina Bao[2†]

[1] Plaksha University, India
{amog.rao,ananya.shukla,jia.bhargava}@plaksha.edu.in
[2] Boston Children's Hospital and Harvard Medical School, Boston, USA
{yangming.ou,rina.bao}@childrens.harvard.edu

Abstract. Segmenting acute and hyper-acute brain lesions in neonatal hypoxic ischemic encephalopathy (HIE) from diffusion-weighted MRI (DWI) is critical for prognosis and treatment planning but remains challenging due to severe class imbalance and lesion variability. We propose a computationally efficient 2D segmentation framework leveraging ADC and Z_{ADC} maps as a three-channel input to UNet++ with an Inception-v4 encoder and scSE attention for enhanced spatial-channel recalibration. To address class critical imbalance and lack of volumetric context in 2D methods, we introduce a novel boundary-and-region-aware weighted loss integrating Tversky, Log-Hausdorff, and Focal losses. Our method surpasses state-of-the-art 2D approaches and achieves competitive performance against computationally intensive 3D architectures, securing a DSC of 0.6060, MASD of 2.6484, and NSD of 0.7477. These results establish a new benchmark for neonatal HIE lesion segmentation, demonstrating superior detection of both acute and hyper-acute lesions while mitigating the challenge of loss collapse. The code is available at https://github.com/BONBID-HIE/Neonatal-HIE-SPARSeg.

Keywords: Neonatal Hypoxic Ischemic Encephalopathy (HIE) · Diffusion Weighted MRI (DWI) · 2D Lesion Segmentation.

1 Introduction

Hypoxic ischemic encephalopathy (HIE) is a type of brain dysfunction (lesion injury) that occurs when the infant's brain experiences a sudden decrease in oxygen or blood flow during the prenatal, intrapartum, or postnatal period [10]. HIE affects around 1 to 5 per 1000 term-born infants, leading to significant long-term neurocognitive deficits such as developmental delays, cognitive impairment, cerebral palsy, epilepsy, or death in about 30%-50% of cases despite receiving therapeutic hypothermia treatment [2]. Accurate identification of brain lesions in

⋆ These authors contributed equally to this work. † Corresponding author.

neonatal MRIs is critical for prognosis, treatment evaluation, and understanding disease progression with the help of reliable biomarkers and detection methods. However, HIE lesions are often hyper-acute and multi-focal, posing challenges for algorithms that perform well on larger, focal lesions like brain tumors.

Apparent Diffusion Coefficient (ADC) maps help address this challenge by measuring the magnitude of water diffusion within a voxel of brain tissue and are calculated using Diffusion-Weighted MRI (DWI) in clinical settings [8]. The degree of reduction in ADC is correlated with the severity of the injury, indicating areas of decreased water diffusion. However, the normal range of ADC varies in space (different brain regions) and in time (as the brain develops rapidly during infancy) [4,11], making expert interpretation error-prone. For example, an ADC value of $800\,(\times 10^{-6}\,\mathrm{mm}^2/\mathrm{s})$ may be considered normal for cortical gray matter, where ADC values typically range from $780 - 1090\,(\times 10^{-6}\,\mathrm{mm}^2/\mathrm{s})$ [5]. However, the same ADC value in white matter, where normal ADC ranges from $620 - 790\,(\times 10^{-6}\,\mathrm{mm}^2/\mathrm{s})$, could be considered lesioned. This regional variability underscores the challenge of using fixed ADC thresholds for lesion detection [2].

Therefore, Z-Score ADC ($\mathrm{Z_{ADC}}$) was proposed to quantify deviations in a patient's ADC values relative to the normative brain-specific range [2,19]. Normative ADC atlases, built from healthy neonates, provide voxel-wise mean and standard deviation of ADC values. The patient's ADC map is deformably registered to the atlas space, ensuring anatomical correspondence across voxels. The $\mathrm{Z_{ADC}}$ at each voxel $\mathbf{x}$ is computed as $Z_{\mathrm{ADC}}(\mathbf{x}) = \frac{I(\mathbf{x}) - \mu_{\phi(\mathbf{x})}}{\sigma_{\phi(\mathbf{x})}}$, where $I(\mathbf{x})$ is the ADC value at voxel $\mathbf{x}$, and $\phi(\mathbf{x})$ denotes its atlas-mapped location via deformation ϕ. This enables voxel-wise normalization and highlights regions of abnormal diffusion relative to normative variability. Comparisons are made by thresholding them at values of -1.5, -2, and -2.5, $Z_{ADC} \in [-10, 10]$. Evaluating the accuracy of predicted masks using Dice Similarity Coefficient (DSC), sensitivity, and specificity, Bao et al. [2] demonstrated the highest DSC (0.54 ± 0.28) at a threshold of -2. This threshold provided the best balance of accuracy and consistency in lesion detection across subjects, further highlighting the effectiveness of $\mathrm{Z_{ADC}}$ maps in lesion detection.

Given the current challenges of spatial variability and multi-focal nature of HIE lesions, Toubal et al. [3] integrated Swin-UNETR, a vision transformer-based segmentation model, with random forest classifier to enhance local feature discrimination. Their approach processes ADC and $\mathrm{Z_{ADC}}$ maps to generate a voxel-wise lesion probability map, segmented into overlapping 5×5 2D patches and classified using random forest. While effective for small datasets, the method is computationally intensive due to quadratic self-attention, 3D voxel-wise distance transforms, and sliding-window patch classification overhead. Building on CNN architectures, Koirala et al. [7] proposed an ensemble of six 3D UNet variants, including dual-branch and attention-based architectures, with a hybrid loss combining BCE, MS-SSIM, and Jaccard Loss. Stratified 5-fold cross-validation, gradient accumulation, and sigmoid-averaged ensembling improved robustness and mitigated over-fitting. However, this approach incurs a high computational

overhead arising due to multiple volumetric inferences and multi-scale loss computations, making training memory intensive.

In this work, we present a computationally efficient 2D segmentation framework for critically imbalanced, acute and hyper-acute lesion detection in neonatal HIE from sparsely sampled volumetric ADC. To achieve this we propose,

(a) A three-channel input representation of ADC, Z_{ADC}, and Z_{ADC} thresholded at -2 ($Z_{ADC}<-2$) improving lesion detectability through enhanced diffusion-based feature representation.
(b) A 2D framework, based on UNet++ with InceptionV4-based encoder and scSE attention, leveraging both spatial and channel recalibration for improved feature extraction.
(c) Tversky-Log-Hausdorff-Focal (TLHF) loss, a boundary and region-aware weighted loss designed to address loss collapse, where weak supervision suppresses gradient updates and leads to near-zero lesion predictions. TLHF improves spatial context, boundary delineation, and stabilizes gradient flow in highly imbalanced lesion segmentation.

Extensive results validate the effectiveness of our 2D framework, demonstrating competitive performance against computationally intensive 3D architectures.

2 Dataset

The BOston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE) [2] is the first public HIE dataset comprising of skull-stripped Apparent Diffusion Coefficient (ss-ADC) and Z_{ADC} maps with manually annotated lesion masks for 133 neonates. Patients had volumetric scan resolutions ranging from $128 \times 128 \times D$ to $256 \times 256 \times D$, where $D \in [16, 64]$. The train set consisted of about 89 patients, whereas the test set included 44 patients. All subjects were term-born neonates with clinically diagnosed HIE; scans were acquired at a mean age of 3.9 ± 2.7 days. Lesion volume analysis stratified patients into hyperacute ($< 1\%$, N=74), acute (1–5%, N=26), and severely acute ($> 5\%$, N=33) categories based on the percentage of DWI-visible abnormalities relative to brain volume. The large and uneven inter-slice gaps (2–6 mm), which are common in real-world neonatal MRI substantially reduce slice-to-slice correlation, limiting the effectiveness of 3D convolutions, which tend to over-smooth distant anatomy or introduce spurious continuity. Consequently, a 2D slice-based framework was better suited for this dataset's anisotropic resolution.

3 Methodology

3.1 Inputs and Pre-Processing

As shown in Figure 1, ADC, Z_{ADC}, and a binary thresholded mask ($Z_{ADC}<-2$) $\in \{0, 1\}$ were stacked slice by slice along the axial axis. The thresholded mask was generated by applying a threshold of -2 to Z_{ADC}, where pixels <-2 were

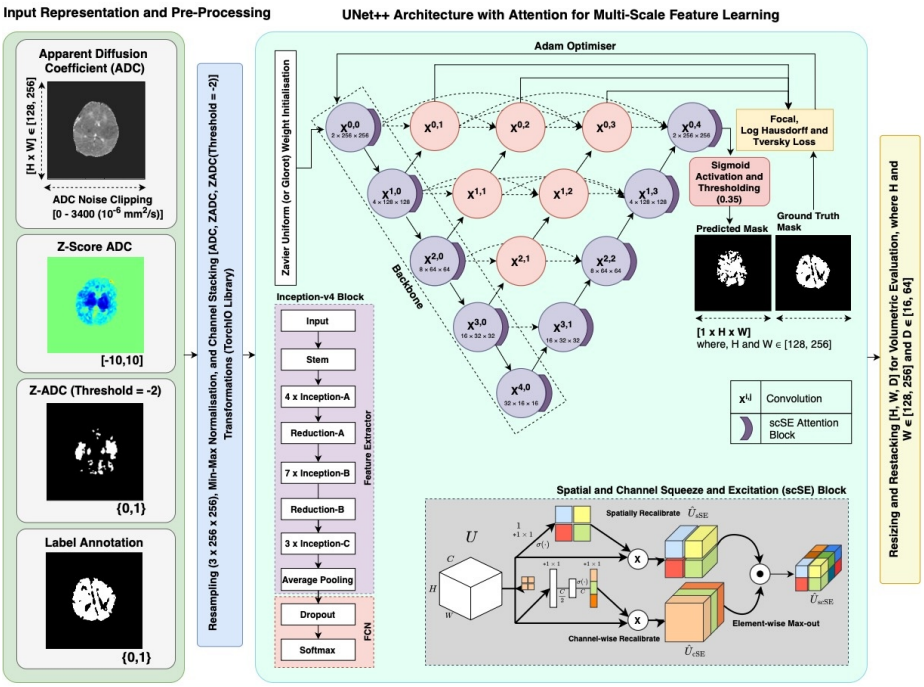

Fig. 1. Overview of our 2D segmentation pipeline using the thresholded Z_{ADC} as a spatial prior channel. The addition of scSE attention enhances small lesion detection, with training guided by a novel region- and boundary-aware TLHF loss.

given a positive label, while the remaining were assigned a negative label, creating a binary representation. These stacked three channel images were then resampled to a resolution of (256×256) using nearest-neighbors interpolation to prevent data distortion or addition of artifacts. To remove noise, ADC maps were clipped to eliminate negative pixel values captured during scanning. Next, min-max normalization was applied such that ADC $[0, 3400]$ and Z_{ADC} $[-10, 10]$ were scaled to $[0, 1]$, ensuring uniform pixel intensity. To introduce diversity in input representations, non-aggressive dynamic augmentations such as random flips, gamma correction, anisotropic transformations and blurring were applied.

3.2 Network Architecture

UNet++ [21] extends UNet [13] by introducing dense nested skip connections which progressively bridge the semantic gap between encoder and decoder feature maps. This enhances multi-scale feature aggregation and information flow. Additionally, deep supervision at multiple decoder stages ensures better gradient propagation, enabling the model to capture both global context and fine-grained features, which is critical for precise medical image segmentation. For 2D brain lesion segmentation, a three-channel input ADC - Z_{ADC} - $Z_{ADC}<-2$ is stacked,

normalized, and then fed into the Inception-v4 encoder blocks [16], constituting an early fusion strategy that enables joint reasoning over continuous (raw, standardized) and binary (thresholded) diffusion cues. Spatial and Channel Squeeze-and-Excitation (scSE) blocks [14] are applied after each encoder-decoder block to enhance feature representation via dynamic intra-feature recalibration, serving as a lightweight fusion mechanism across spatial and channel dimensions. The Channel Squeeze-and-Excitation (cSE) branch applies global average pooling to compute channel-wise attention while, the Spatial Squeeze-and-Excitation (sSE) branch uses 1×1 convolution and sigmoid activation to generate a spatial attention map, emphasizing pixel-level regions with higher importance. Concurrent recalibration by the cSE and sSE blocks, combined via element-wise addition, enhances lesion boundary delineation: ($\hat{U}_{scSE} = \hat{U}_{cSE} + \hat{U}_{sSE}$).

3.3 Loss Function

Tversky Loss [15] is a generalized Dice Loss (F-score equivalent) with tunable parameters α and β to control false positive and false negative penalties, improving performance on imbalanced datasets,

$$L_{Tversky}(p,q) = 1 - \frac{\sum_{v\in V} p(v)q(v)}{\sum_{v\in V}[p(v)q(v) + \alpha(1-q(v))p(v) + \beta(1-p(v))q(v)]}. \quad (1)$$

Focal Loss [9] is a variant of Cross-Entropy loss that focuses more on hard misclassified examples using a focusing parameter γ,

$$L_{Focal}(p,q) = -\frac{1}{|V|}\sum_{v\in V} \alpha \cdot q(v) \cdot (1-p(v))^{\gamma} \log(p(v)), \quad (2)$$

where, the weighting factor α balances the importance of positive and negative samples, while $\gamma \geq 0$ reduces the relative loss for easy-to-classify cases.

Log-Hausdorff Loss [6] captures boundary discrepancies by incorporating distance transforms ($d_p(v)$ and $d_q(v)$), with a logarithmic term enhancing sensitivity to small errors,

$$L_{LH}(p,q) = \log\left(1 + \frac{1}{|V|}\sum_{v\in V}\big(p(v)-q(v)\big)^2 \cdot \big(d_p(v)^{\alpha} + d_q(v)^{\alpha}\big)\right). \quad (3)$$

Tversky-Log-Hausdorff-Focal Loss — our customized boundary-and-region-aware loss, is a weighted combination of the above mentioned losses,

$$L_{TLHF}(p,q) = \gamma_T \cdot L_{Tversky} + \gamma_{LH} \cdot L_{LH} + \gamma_L \cdot L_{Focal}. \quad (4)$$

3.4 Hyperparameters, Training Strategy, and Evaluation

Training was performed on an NVIDIA RTX A5000 (24GB) for 75–100 epochs with a batch size of 16, using early stopping with a patience of 15–20 epochs. Model weights were initialized using Xavier uniform distribution, and optimized using Adam with an initial learning rate of 0.0001, adjusted using ReduceLROnPlateau scheduler. TLHF Loss is a combination of Tversky ($\alpha = 0.3$, $\beta = 0.7$, $\gamma_T = 1.5$), Log-Hausdorff ($\gamma_{LH} = 2$), and Focal ($\gamma = 3$, $\gamma_F \in \{2, 4\}$) losses. In a lightweight $\pm 20\%$ grid sweep around these γ values, we observed minor performance variation ($\pm 0.29\%$ DSC, $\pm 0.42\%$ MASD, $\pm 0.37\%$ NSD), confirming robustness to local weight changes. γ_F was dynamically adjusted, initialized at 2, and linearly ramped to 4 in a curriculum-style schedule to prioritize region-wise learning early and emphasize small, hard lesions later. A sigmoid threshold of 0.35 was empirically found to yield optimal binary segmentation performance.

Evaluation Metrics. Segmentation accuracy is assessed using Dice Similarity Coefficient (DSC), Mean Average Surface Distance (MASD), and Normalized Surface Distance (NSD) [12], and evaluated on the BONBID-HIE challenge[3].

(a) **Dice Similarity Coefficient (DSC)** quantifies global overlap between predicted (p) and ground truth (q) masks: $\mathrm{DSC}(p, q) = \frac{2\times|p\cap q|}{|p|+|q|}$.

(b) **Mean Average Surface Distance (MASD)** measures average boundary distance between p and q: $\mathrm{MASD}(p, q) = \frac{1}{2}\left(\frac{\sum d(\delta(q),\delta(p))}{|\delta(q)|} + \frac{\sum d(\delta(p),\delta(q))}{|\delta(p)|}\right)$.

(c) **Normalized Surface Distance (NSD)** captures surface overlap at boundaries within a dilated tolerance τ: $\mathrm{NSD}(p, q) = \frac{|\delta(q)\cap\gamma_\tau(p)|+|\gamma_\tau(q)\cap\delta(p)|}{|\delta(q)|+|\delta(p)|}$.

4 Results and Inference

We structure our results to follow a modular, incremental flow across four key design dimensions: (i) we evaluate input encodings across UNet++ backbones, (ii) we fix the best-performing encoder and input design to compare loss functions and introduce TLHF, (iii) we apply attention (scSE) to the TLHF setup and analyze performance by lesion size, and (iv) we benchmark our final model against top 3D BONBID-HIE submissions. This progression highlights the combined contributions of input priors, loss design, attention, and architecture choices.

Effectiveness of Input Channel Stacking. Table 1 evaluated the impact of input channel stacking on segmentation performance using a UNet++ architecture with DenseNet-161 and Inception-v4 encoders, trained with Dice-Focal loss. Models trained on a 3-channel input (ADC, $\mathrm{Z_{ADC}}$, and $\mathrm{Z_{ADC}}<-2$) achieved the highest scores across all three metrics, significantly outperforming models trained on their 2-channel (ADC, $\mathrm{Z_{ADC}}$) counterparts, with DSC improvements of up to 5.98% for DenseNet-161 and 5.83% for Inception-v4. These results suggest that

[3] https://bonbid-hie2024.grand-challenge.org/

Table 1. Effect of input channel stacking using UNet++ with different encoders.

Input	DenseNet-161			Inception-v4		
	DSC(↑)	MASD(↓)	NSD(↑)	DSC(↑)	MASD(↓)	NSD(↑)
ADC	0.2977	5.9840	0.3863	0.3239	5.1795	0.4361
Z_{ADC}	0.5128	3.5919	0.6689	0.5221	4.1288	0.6641
[ADC, Z_{ADC}]	0.5438	3.2168	0.6655	0.5540	3.1597	0.6800
[ADC, Z_{ADC}, $Z_{ADC}<-2$]	0.5763	3.1023	0.6751	**0.5863**	**3.0711**	**0.6941**

$Z_{ADC}<-2$ masks act as non-exclusive spatial lesion priors, highlighting extreme diffusion deviations without masking borderline voxels, which remain visible via ADC and Z_{ADC} channels, thereby enhancing regions with higher likelihood of lesion detection. Furthermore, Inception-v4 demonstrated superior performance across all metrics, due to its multi-scale feature extraction capability. Therefore, Inception-v4 was used as the encoder for our whole pipeline.

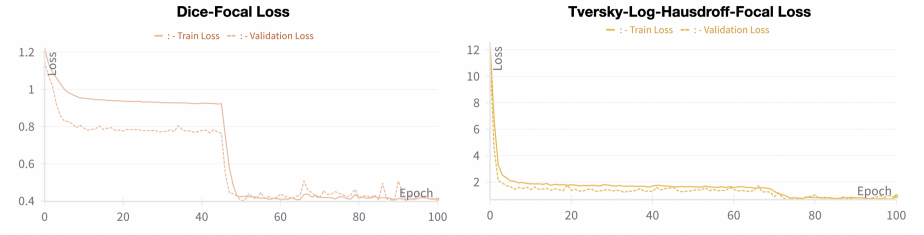

Fig. 2. (a) Loss Collapse with Dice-Focal Loss (b) Mitigating Loss Collapse by using Tversky-Log-Hausdorff-Focal Loss

Effectiveness of Loss and Attention. Due to the nature of HIE lesions, which are typically small and diffusely distributed, over half of the patients in the cohort have less than 1% of their brain volume injured [2]. Directly training using Dice-Focal loss leads to loss collapse, where lesion predictions shrink to near-zero, resulting in minimal gradients and trapping the model in a local minimum, as illustrated in Figure 2(a). Dice loss struggles with vanishing gradients, while Focal Loss, designed to address class imbalance amplifies the issue by suppressing false positives. Recovery from collapse occurs through stochastic perturbations, which restores gradients and stabilizes training. Large lesions (>5%) recover effectively due to stronger gradients, whereas smaller lesions (<1% and 1–5%) remain underrepresented, limiting effective supervision and refinement.

To address loss collapse, we introduced a hybrid TLHF loss function combining boundary-based Log-Hausdorff loss, with region-specific Tversky loss, and class-balancing Focal loss. This approach enhanced training during early epochs, as indicated in Figure 2(b), ensuring stable gradient flow and improved hyperacute lesion segmentation. Performance of TLHF loss is evidenced in Tables 2

Table 2. Ablation of loss functions (Dice-Focal vs. TLHF) and attention (± scSE).

Loss	Attention	DSC (↑)	MASD (↓)	NSD (↑)
Dice-Focal Loss	None	0.5863	3.0711	0.6941
TLHF Loss	None	**0.6060**	**2.6484**	0.7153
TLHF Loss	**scSE**	0.5986	2.7986	**0.7477**

and 3, where it consistently outperforms Dice-Focal loss across all metrics. TLHF demonstrates improved lesion localization as indicated by MASD (2.6484), reduced boundary-based pixel-level misclassification through NSD (0.7477), resulting in a stronger global overlap captured by DSC (0.6060). Table 3 further con-

Table 3. Performance stratified by lesion volume groups using TLHF ± scSE.

Design	<1%			1-5%			>5%		
	DSC	MASD	NSD	DSC	MASD	NSD	DSC	MASD	NSD
TLHF	0.4740	**4.1310**	0.5997	**0.6753**	**1.6719**	**0.8011**	**0.8132**	**0.8550**	**0.8665**
TLHF+scSE	**0.4776**	4.2458	**0.6777**	0.6449	1.8769	0.7802	0.8090	0.8653	0.8626

firms that the proposed TLHF loss, combined with scSE attention, is particularly effective for smaller lesions (<1% of brain volume), boosting NSD relatively by 13% and enhancing the model's ability to capture fine-grained representations.

Table 4. Performance evaluation across BONBID-HIE 2023 and 2024 submissions.

Participant/Team	Method	Architecture	DSC (↑)	MASD (↓)	NSD (↑)
Sejong_AI_team	3D	UNet	0.6262	2.5017	0.7316
Toubal et al. [3]	3D	Swin-UNETR w/ RF	0.6215	2.2556	0.7678
Ours	**2D**	**UNet++**	**0.6060**	**2.6484**	0.7153
Ours	**2D**	**UNet++ w/ scSE**	0.5986	2.7986	**0.7477**
Koirala et al. [7]	3D	UNet Ensemble	0.5800	2.5993	0.7379
Wodzinski et al. [20]	3D	ResUNet	0.5770	2.9419	0.7379
Tahmasebi et al. [17]	3D	nnUNet	0.4998	5.3327	0.6449
Aydın et al. [1]	3D	Attention UNet	0.4826	3.4626	0.6176

Comparisons with 3D Networks. Our parameter-efficient 2D models, using three channel stacked inputs, enhanced with scSE attention, and trained on our boundary-and region-aware TLHF loss, achieve competitive performance in volumetric segmentation tasks despite the extensive parameters and contextual depth of 3D methods. Notably, our lightweight 2D approach matches or outperforms several 3D benchmarks, securing the second-highest NSD score and ranking in the top three for DSC, as shown in Table 4, demonstrating strong performance across existing studies on neonatal HIE lesion segmentation.

5 Conclusion

In this work, we present a computationally efficient 2D segmentation framework for neonatal HIE lesions, leveraging a novel three-channel input including (ADC, Z_{ADC}, and $Z_{ADC}<-2$) as an anatomical prior to mitigate spatial and temporal uncertainties in ADC by encoding region-specific diffusion deviations. We introduce a hybrid region-and boundary-aware loss function, Tversky-Log-Hausdorff-Focal loss, that effectively addresses lack of spatial context in 2D methods and loss collapse in critically imbalanced datasets. Through our extensive ablation study, we demonstrate that spatial and channel attention mechanisms significantly enhance segmentation, particularly for hyper-acute (<1%) lesions. Our findings establish that our parametrically efficient 2D framework, outperforms resource-intensive 3D architectures, setting a new benchmark for neonatal HIE lesion segmentation. Future work can explore ConvLSTM-based approaches [18] to model 3D data sequentially for improved spatial coherence along with mid-level or modality-aware fusion strategies, such as dual-stream encoders or cross-attention, to further enrich multi-contrast feature integration.

Acknowledgments. The first authors thank Plaksha University for seed financial support and Dr. Anupam Sobti for early mentorship in the Deep Learning (AI3011) course. The authors also thank the BONBID-HIE challenges and workshops of 2023 and 2024 for providing the dataset, insights and evaluation platform.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Aydın, M.A., Abdinli, E., Unal, G.: SegResNet Based Reciprocal Transformation for BONBID-HIE Lesion Segmentation. In: Bao, R., Grant, E., Kirkpatrick, A., Wachs, J., Ou, Y. (eds.) AI for Brain Lesion Detection and Trauma Video Action Recognition, pp. 39–44. Springer Nature Switzerland (2025)
2. Bao, R., Song, Y., Bates, S.V., Weiss, R.J., Foster, A.N., Jaimes, C., Sotardi, S., Zhang, Y., Hirschtick, R.L., Grant, P.E., Ou, Y.: BOston Neonatal Brain Injury Data for Hypoxic Ischemic Encephalopathy (BONBID-HIE): I. MRI and Lesion Labeling. Scientific Data **12**(1) (2025)
3. Eddine Toubal, I., Soltani Kazemi, E., Rahmon, G., Kucukpinar, T., Almansour, M., Ho, M.L., Palaniappan, K.: Fusion of deep and local features using random forests for neonatal HIE segmentation. In: Bao, R., Grant, E., Kirkpatrick, A., Wachs, J., Ou, Y. (eds.) AI for Brain Lesion Detection and Trauma Video Action Recognition. pp. 3–13. Springer Nature Switzerland (2025)
4. Goergen, S., Ang, H., Wong, F., Carse, E., Charlton, M., Evans, R., Whiteley, G., Clark, J., Shipp, D., Jolley, D., Paul, E., Cheong, J.: Early MRI in term infants with perinatal hypoxic–ischaemic brain injury: Interobserver agreement and MRI predictors of outcome at 2 years. Clinical Radiology **69**(1), 72–81 (2014)
5. Helenius, J., Soinne, L., Perkiö, J., Salonen, O., Kangasmäki, A., Kaste, M., Carano, R.A., Aronen, H.J., Tatlisumak, T.: Diffusion-weighted MR imaging in normal human brains in various age groups. American Journal of Neuroradiology **23**(2), 194–199 (2002)

6. Karimi, D., Salcudean, S.: Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. IEEE Transactions on Medical Imaging **PP**, 1–1 (07 2019)
7. Koirala, C.P., Mohapatra, S., Schlaug, G.: An ensemble approach for segmentation of neonatal HIE lesions. In: Bao, R., Grant, E., Kirkpatrick, A., Wachs, J., Ou, Y. (eds.) AI for Brain Lesion Detection and Trauma Video Action Recognition. pp. 23–27. Springer Nature Switzerland (2025)
8. Liauw, L., van Wezel-Meijler, G., Veen, S., Van Buchem, M., van der Grond, J.: Do apparent diffusion coefficient measurements predict outcome in children with neonatal hypoxic-ischemic encephalopathy? American Journal of Neuroradiology **30**(2), 264–270 (2009)
9. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988 (2017)
10. Long, M., Brandon, D.H.: Induced hypothermia for neonates with hypoxic–ischemic encephalopathy. Journal of Obstetric, Gynecologic & Neonatal Nursing **36**(3), 293–298 (2007)
11. Ozturk, A., Sasson, A., Farrell, J., Landman, B., da Motta, A., Aralasmak, A., Yousem, D.: Regional differences in diffusion tensor imaging measurements: Assessment of intrarater and interrater variability. American Journal of Neuroradiology **29**(6), 1124–1127 (2008)
12. Reinke, A., Eisenmann, M., Tizabi, M., Sudre, C., Radsch, T., Antonelli, M., Arbel, T., Bakas, S., Cardoso, M., Cheplygina, V., Farahani, K., Glocker, B., Heckmann-Notzel, D., Isensee, F., Jannin, P., Kahn, C., Kleesiek, J., Kurç, T., Kozubek, M., Landman, B., Litjens, G., Maier-Hein, K., Menze, B., Muller, H., Petersen, J., Reyes, M., Rieke, N., Stieltjes, B., Summers, R., Tsaftaris, S., van Ginneken, B., Kopp-Schneider, A., Jager, P., Maier-Hein, L.: Common limitations of image processing metrics: A picture story. arXiv (2021)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer (2015)
14. Roy, A.G., Navab, N., Wachinger, C.: Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. IEEE Transactions on Medical Imaging **38**(2), 540–549 (2019)
15. Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: International Workshop on Machine Learning in Medical Imaging. pp. 379–387. Springer (2017)
16. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 4278–4284. AAAI Press (2017)
17. Tahmasebi, N., Punithakumar, K.: A deep neural network approach for the lesion segmentation from neonatal brain magnetic resonance imaging. In: Bao, R., Grant, E., Kirkpatrick, A., Wachs, J., Ou, Y. (eds.) AI for Brain Lesion Detection and Trauma Video Action Recognition. pp. 34–38. Springer Nature Switzerland (2025)
18. Wang, Z., Sun, D., Zeng, X., Wu, R., Wang, Y.: Contextual embedding learning to enhance 2D networks for volumetric image segmentation. Expert Systems with Applications **253**, 124279 (2024)

19. Weiss, R.J., Bates, S.V., Song, Y., Zhang, Y., Herzberg, E.M., Chen, Y.C., Gong, M., Chien, I., Zhang, L., Murphy, S.N., Gollub, R.L., Grant, P.E., Ou, Y.: Mining multi-site clinical data to develop machine learning MRI biomarkers: application to neonatal hypoxic ischemic encephalopathy. Journal of Translational Medicine **17**(1), 385 (2019)
20. Wodzinski, M., Müller, H.: Improving segmentation of hypoxic ischemic encephalopathy lesions by heavy data augmentation: Contribution to the BONBID challenge. In: Bao, R., Grant, E., Kirkpatrick, A., Wachs, J., Ou, Y. (eds.) AI for Brain Lesion Detection and Trauma Video Action Recognition. pp. 28–33. Springer Nature Switzerland (2025)
21. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: UNet++: A nested U-Net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Lecture Notes in Computer Science, vol. 11045, pp. 3–11. Springer (2018)