SUPPLEMENT: DESCRIPTIONS OF 14 SUBMITTED ALGORITHMS

S1. CIVALAB - DEEP & LOCAL FEATURE FUSION WITH RANDOM FOREST

A. Pre-processing

In the initial stages of our approach, two preprocessing steps on the two-channel dataset [57] were applied.

Intensity regularization. We employed the z-transform method. By standardizing the intensity values, we provide a consistent scale per channel, which aids the network in interpreting the data more robustly.

Space Normalization. We ensure uniformity in image resolution by interpolating all images into a fixed resolution. We utilized trilinear interpolation to standardize all images to a consistent 1mm x 1mm x 1mm resolution.

B. Swin-UNETR

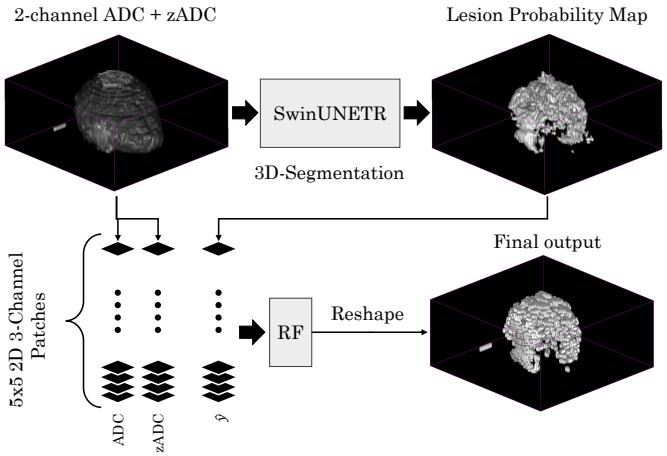

Fig. 9. The proposed HIE brain image segmentation system. The diagram illustrates the integration of Swin-UNETR with a random forest classifier, taking space-normalized 2-channel input (ADC and zADC) to produce a lesion probability map. This map, combined with the input channels, is divided into 5x5, 3-channel patches, which are subsequently fed into the random forest (RF) classifier. Following individual patch classification, the system reconstructs the final segmentation volume, providing a comprehensive visualization of the detected lesions.

Swin-UNETR [68] represents a significant advancement in the field of medical image segmentation, particularly tailored for intricate tasks like lesion prediction in brain images. While traditional Fully Convolutional Neural Networks (FCNNs) [71]–[74] have been the cornerstone of 3D medical image segmentation, they often falter in capturing long-range information due to the limited effective receptive field [75]. This limitation becomes especially pronounced when dealing with lesions that exhibit considerable variability in size and shape. Swin-UNETR addresses this challenge head-on. For the specific task of 3D HIE brain image segmentation [57], we trained the network on the data by taking a space-normalized 2-channel input (ADC and zADC) and training it to predict lesions. The architecture reformulates the segmentation task as a sequence-to-sequence prediction challenge. Here, the multi-modal input data is projected into a 1D sequence of embeddings, which then serves as the input to a hierarchical Swin transformer [76], [77] acting as the encoder. This encoder, leveraging shifted windows for self-attention computations, extracts features at multiple resolutions. These features are subsequently integrated with an CNN-based decoder at each resolution through skip connections, ensuring a comprehensive and detailed representation of the data, conducive to accurate lesion prediction.

C. Loss function

In addition to a standard DICE Loss [78], we employ a novel loss inspired by the Hausdorff distance measure [79], [80], a metric renowned for its efficacy in quantifying the spatial discrepancy between two point sets. Recognizing the inherent limitations of the traditional Hausdorff distance loss [81] in certain segmentation scenarios, we enhance its robustness by incorporating a logarithmic term. This log term ensures that the loss is more sensitive to finer discrepancies, allowing for a more nuanced optimization process. Our log Hausdorff loss $\mathcal{L}_{LH} : (p, q) \rightarrow \mathbb{R}$ is given as:

$$\mathcal{L}_{LH}(p,q) = \log\left(1 + \frac{1}{|V|}\sum_{v\in V}\Big((p(v) - q(v))^2 \cdot \Big(d_p(v)^\alpha + d_q(v)^\alpha\Big)\Big)\right) \quad (4)$$

Where d_p and d_q represent the distance transforms for p and q respectively. α determines how strongly larger errors are penalized. In our experiments, we set $\alpha = 2.0$. Additionally, we define the DICE loss $\mathcal{L}_{DICE} : (p, q) \rightarrow \mathbb{R}$ as:

$$\mathcal{L}_{DICE}(p,q) = 1 - \frac{2 \times \sum p(v)q(v) + \epsilon}{\sum p(v)^2 + \sum q(v)^2 + \epsilon} \quad (5)$$

Where ϵ is a small constant (e.g., 10^{-5}) added for numerical stability to prevent division by zero. Finally, we define our final loss function to minimize both DICE loss and Log Hausdorff loss as the following:

$$\mathcal{L}(p,q) = \beta_1 \cdot \mathcal{L}_{DICE} + \beta_2 \cdot \mathcal{L}_{LH} \quad (6)$$

Where β_1 and β_2 are coefficient weights used to adjust the numerical values, ensuring that each term has a roughly equal contribution.

D. Random forest fusion

Ensemble approaches are known to be effective in creating a robust deep-learning system [82]–[84]. The integration of a random forest classifier (inspired by [85]) into the segmentation process serves as a refinement step to enhance the accuracy of lesion detection. After Swin-UNETR processes the 3D HIE brain images, producing a lesion probability map, a moving window approach is employed to dissect this map and the two input channels (ADC and zADC) into smaller 5×5 2D windows. Each of these windows, now represented

as a 3-channel image, is fed into the random forest classifier. The classifier, trained on numerous such windows, predicts the lesion probability for the central pixel of each window as shown in Figure. 9. By leveraging the ensemble learning capabilities of the random forest, we sidestep the need to train multiple resource-intensive 3D segmentation models. Our method uniquely combines the long-range information derived from Swin-UNETR with the local features captured in the 5×5 windows, ensuring a balanced integration of both global and local features. This approach allows the model to effectively capture the intricate patterns of HIE lesions. As a result of these combined strategies, we produce a lesion segmentation that stands out in both quality and robustness, setting a new standard in HIE lesion detection and segmentation.

S2. XLERATORXLERATOR9 - LABEL AWARE DENOISING PRETRAINING

Algorithm and Data: The goal of the proposed framework is for the model to develop stronger representations of regions of interest, based on the denoising pre-training procedure of Decoder Denoising Pretraining (DDP) [86] In broad terms, the method equates to adding noise to pixels sampled from $\mathcal{N}(0,\sigma_a)$ for regions of interest and $\mathcal{N}(0,\sigma_b)$ elsewhere. Given the challenge setting, we take the ROIs to be the annotated areas with brain lesions including areas adjacent to lesion contours. Following CarveMix [87], given an annotation $y \in \mathbb{R}^{w\times h}$ we define an indicator map for the region of interest for an image as

$$\mathcal{C}(y) = \begin{cases} 1 & D(\cdot|y) \leq \lambda \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $D(\cdot|y)$ is the distance between a pixel and the contour of the annotated area. Notably, $D(\cdot|y)$ is negative for annotated pixels within a lesion. λ is sampled from

$$\lambda \sim \frac{1}{2}U\Big(-\frac{1}{2}\Big|D(\cdot|y)_{min}\Big|,0\Big) + \frac{1}{2}U\Big(0,\Big|D(\cdot|y)_{min}\Big|\Big)$$

where $D(\cdot|y)_{min} = \min_x D(x|y)$ is an indicator of lesion size for a given annotation y. Hence, the distribution of λ will ensure the sampled ROI is proportionate to the size of the lesions. In practice, we first obtain a random matrix $\mathbf{W}^y_{ij} \in \mathbb{R}^{2\times h\times w}$, defined as

$$\mathbf{W}^y_{kij} \sim \begin{cases} \mathcal{N}(0,\sigma_a^2) & \text{if } \mathcal{C}(y)_{ij} = 1 \\ \mathcal{N}(0,\sigma_b^2) & \text{otherwise,} \end{cases} \tag{8}$$

before merging it with the image. Using this construction, we can add in the noise pixel-wise so that given an image $\mathbf{x} \in \mathbb{R}^{2\times h\times w}$ we can apply the noising transform

$$\mathbf{x}' = \sqrt{\gamma}\mathbf{x} + \sqrt{1-\gamma}\mathbf{W}^y \tag{9}$$

where γ is set to 0.95. Consistent with DDP, the model is trained to predict the noise $\mathbf{W}^y$ using the L2 loss. Altogether, given a segmentation model f_θ with learnable parameters θ, we pre-train the model using

$$L_{\text{pt}}(\theta) = \mathbb{E}_x\mathbb{E}_{\mathbf{W}^y}\Big\|f_\theta(\sqrt{\gamma}\mathbf{x} + \sqrt{\mathbf{1-\gamma}}\mathbf{W}^{\mathbf{y}}) - \mathbf{W}^y\Big\|_2^2. \tag{10}$$

The hyperparameters σ_a and σ_a are empirically chosen where here we set $\sigma_a = 0.8$ and $\sigma_b = 1.5$ based on cross-validation experiments.

To create the inputs to the model, the skull-stripped ADC and ZADC maps are normalized to have zero mean and unitary variance when computed across the training dataset. The resulting two maps are concatenated and resized to have a width and height of 256 pixels.

Training and Testing: The model is trained using the denoising procedure as previously outlined then, the model is subsequently fine-tuned to predict segmentation labels. The model is subsequently fine-tuned to predict segmentation labels using the following weighted segmentation loss:

$$L_{\text{ft}}(p,y) = L_{\text{BCE}}(p,y) + L_{\text{Dice}}(p,y) + 3L_{\text{Focal}}(p,y)$$

where $L_{\text{BCE}}, L_{\text{Dice}}$, and L_{Focal} are the binary cross-entropy, Dice [78], and Focal [88] losses respectively. Each model is trained until saturation on a validation set, where we use the Dice [78] to measure performance.

During both pretraining and fine-tuning we apply a simple augmentation where images are randomly flipped horizontally and vertically. Other augmentation strategies typically reduce performance. For both training phases, the model is optimized using the Adam optimizer with a learning rate of 0.0001 for 30 epochs with a batch size of 16. Due to time constraints, the final algorithm used for inference is a voting ensemble of 8 models, however, better results can be achieved with even bigger ensembles. Using these methods, models can focus resources to learn stronger representations only in a few key areas that are the most relevant to the downstream lesion segmentation task.

S3. FRIMPZ - ENHANCING LESION SEGMENTATION IN THE BONBID-HIE CHALLENGE: AN ENSEMBLE STRATEGY

Algorithm and Data: Swin-UNETR [68] signifies a major breakthrough in the filed of medical image segmentation, specifically designed to address complex challenges such as predicting lesions in brain images. Although conventional Fully Convolutional Neural Networks (FCNNs) [71], [72] have been fundemental in 3D image segmentation, they frequently struggle to capture long-range information because of the restricted effective receptive filed [75] of their convolution layers. On the other hand, Swin-UNETR, drawing from the advancements in vision transformers, directly overcomes such challenges. In the particular context of 3D HIE brain image segmentation [57], we effectivelly fitted the model to the data by utilizing a space-normalized 2-channel input (ADC and zADC) and training it to predict lesions. More specifically, we formulate the image segmentation task as a mapping from a sequence of inputs to a sequence of outputs, then leveraging sequence-to-sequence models to solve it. Here, the multi-modal input data is transformed into a 1D sequence of embeddings, which subsequently serves as the input to a hierarchical Swin transformer [76], [77] functioning as the encoder. This encoder, utilizing shifted windows for self-attention computations, captures features across multiple resolutions. These features are then combined with a decoder

Embedded Dimension	768
Feature size	24
Number of Heads	12
Number of Parameters	15.7M

TABLE VI
THE BEST CONFIGURATIONS OF THE SWIN UNETR MODEL.

based on fully convolutional neural networks (FCNN) at each resolution using skip connections, ensuring a detailed representation of the data that facilitates accurate lesion prediction.

The model is trained using the 1st Boston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE), which consists of 85 cases in the training set. The model's performance is evaluated on separate validation and test sets, each containing 4 and 3 cases, respectively. Ground truth segmentation labels for the validation and test sets are unavailable. Thus, model evaluation is performed exclusively through the official BONBID-HIE challenge server.

To achieve maximum performance, we perform a four fold cross-validation on the dataset, using 60 cases for training and 24 cases for validation in each run. It must be noted that our training solely relies on the data provided by the challenge. The final results are achieved by utilizing an ensemble of four Swin-UNETR models, resulting in an improved performance and more robust predictions in comparison to each individual split.

Training and Testing: The Swin-UNETR model [68] is developed using PyTorch and MONAI libraries. The model is trained on a HPC server containing NVIDIA A100 GPUs. Table VI provides information about the Swin UNETR architecture, including its configurations, and the total parameters of the model. The learning rate is adjusted to 0.0001. In the training phase, random patches with dimensions of 128×128×128 are taken from 3D image volumes. We apply a data augmentation of random rotation with a probability of 0.5 for z axis. Likewise, with a probability of 0.5, we randomly flip input image channels along the x axis. Additionally, we perform an intensity normalization preprocessing step to guarantee that the values of individual voxels within 3D volume images have consistent and standardized ranges. We also establish the size of the batch to be one. The model is trained within 1500 epochs by setting the weight decay value to 0.00001. Furthermore, the loss function employed for optimizing the model's parameters is the Dice loss [89]. During the inference phase, we employ a sliding window technique with a size of 128 × 128 × 128 and a 0 overlap for adjacent voxels.

S4. SCHAUGLAB - INTEGRATED ENSEMBLE UNET-ONET MODEL WITH AUGMENTATION

Algorithm and Data: In this work, we used an ensemble approach for segmenting neonatal hypoxic-ischemic encephalopathy (HIE) lesions using deep learning techniques applied to high-resolution diffusion-based MRI data from the Boston Neonatal Brain Injury Dataset (BONBID-HIE). The methodology integrated preprocessing steps like min-max normalization, central patch extraction, and channel stacking of ADC and Z-score ADC maps, creating an information-rich input for model training. Six variants of the UNet model, including a basic 3D UNet [90] and dual-branch ONet [91], were used to leverage their unique segmentation capabilities. Hybrid loss functions comprising Binary Cross-Entropy, MS-SSIM, and Jaccard index were utilized to refine segmentation across hierarchical levels. The models underwent stratified k-fold cross-validation to ensure robust hyperparameter optimization. Gradient accumulation was employed during training to simulate larger batch sizes on memory-constrained GPUs, with models trained for 400 epochs. The final ensemble output was generated by averaging the sigmoid activations of all models, followed by a 0.5 thresholding step for mask creation, and was further refined using postprocessing techniques to remove noise.

This ensemble approach demonstrated a significant improvement in segmentation performance, achieving a Dice score of 0.5800 ± 0.2557, with NSD and MASD scores indicating promising precision in lesion delineation. However, post-processing did not significantly enhance the unseen test dataset results. The ensemble's robustness highlights the advantages of combining diverse architectures and hybrid loss functions for precise neonatal HIE lesion segmentation. Future work aims to incorporate larger datasets, such as dHCP or BCP, to allow pre-training and transfer learning, potentially improving segmentation accuracy. This study provides a foundation for the development of advanced tools in neonatal neuroimaging and the improvement of clinical diagnostic workflows.

S5. RAJROY

This variation of the frimpz approach utilizes Swin-UNETR for 3D brain image segmentation, specifically to predict lesions in the BONBID-HIE dataset. Unlike traditional FCNNs, Swin-UNETR benefits from vision transformers, capturing long-range dependencies via shifted windows in the self-attention mechanism. We use a space-normalized 2-channel input (ADC and zADC) and apply a sequence-to-sequence framework for segmentation. The model is trained with five-fold cross-validation on 84 training and validation cases, with an ensemble of five Swin-UNETR models enhancing performance. Training includes data augmentation (random rotation, flipping), intensity normalization, and Dice loss optimization. Inference is performed using a sliding window technique. All training and evaluation were done on an HPC server with NVIDIA A100 GPUs.

S6. MEDGIFT (LWM) - HEAVY AUGMENTED MULTIMODAL RUNET

Algorithm and Data: The method is based on a neural network similar to the 3-D ResUNet architecture [92]. The input volumes are firstly concatenated and resampled to 256x256x128. Then they are normalized channel-wise to [0-1] range with min/max window for the ADC map and [-10, 10] window for the ZADC map. The network takes as input Bx2xHxWxD volumes and outputs segmentations with the Bx1xHxWxD. The details related to the network architecture are available in the associated repository [93]. Only the data

provided by the challenge organizers was used during training. No external data was used. No pretrained networks were used. The proposed method has several limitations that could be addressed. First, the use of patch-based approach could be potentially beneficial since some of the lesions are extremely small. Secondly, the boundaries of the segmentations masks are usually quite sharp which is not desired for CNNs. Therefore, initial smoothing of the segmentation masks could improve the network training and convergence. Finally, the use of a transformer-based architecture could potentially further improve the results by providing larger receptive field capturing long-range spatial relationships.

Training and Testing: The network was trained fully supervisedly using the provided dense annotations. The AdamW was used as the optimizer, with exponentially decaying learning rate scheduler. The network was trained until convergence on the validation set. Additional experiments with 5-fold validations were performed. No ensembles were employed, the best final model was used for the final Docker container. The objective function is a linear combination of two loss terms: (i) Soft Dice Loss, (ii) Focal Loss, with the same weight for both the loss terms. During training, the data is augmented by: (i) random affine transformation, (ii) random intensity transformation, (iii) random Gaussian noise, (iv) random flipping (all axes), (v) random motion artifacts, (vi) random anisotropy transformation, (vii) random Gaussian blurring. The transformations are applied with random order, each with probability equal to 0.5. The intensity augmentations are defined and applied separately to each modality. Moreover, prior to training the data was offline augmented by elastic transformation to generate 10000 cases. The reason for performing the elastic augmentation prior to training was related to the computational complexity of this transformation and the resulting CPU bottleneck. The source code and all network, training, augmentation, and inference hyperparameters are reported in the repository [93]. The pretrained model is available upon request, as well as the access to the Grand-Challenge algorithm [94].

S7. IMAD.TOUBAL

This method uses Swin-UNETR, a deep learning model for 3D medical image segmentation. It combines the Swin Transformer backbone, which captures long-range dependencies in the data through efficient self-attention, with the U-Net architecture, known for its effective segmentation capabilities. The Swin Transformer allows the model to handle complex 3D structures, such as lesions or tumors, by learning both local and global image context. The model is trained with techniques like data augmentation and Dice loss optimization to improve segmentation accuracy. Swin-UNETR has shown superior performance in segmenting small, irregular structures, making it particularly useful for medical imaging tasks like brain lesion detection.

S8. UNEIMAGE - VOXEL-SPECIFIC LOGISTIC REGRESSION

Algorithm and Data: Expert segmentation of hypoxic ischemic encephalopathy focuses on abnormally low ADC values. However, what constitutes an abnormally low ADC value varies from region to region [57]. Therefore, we made a model that would segment each voxel independently depending on its location in the brain and its level of ADC abnormality, that is z-score value. To allow for the alignment of voxels between subjects, we created a custom template with the provided challenge ADC images using iterative registrations with ANTs [95]. We then aligned the z-score images provided by the challenge to the custom template.

The logistic regression model was designed such that z-score was the only predictor variable, and separate weights were assigned to each voxel in the template space.

Training and Testing: We trained the logistic regression model twice to with minimization of dice similarity score loss as the objective function. First, we trained the model for 1,000 iterations with all of the model weights shared between voxels. This ensured that the initial values for each of the voxel-specific parameters was appropriate. Next, we trained the model for 10,000 iterations while updating the voxel-specific parameters independently of each other.

The model achieved a Dice similarity score of 0.56, a mean absolute surface distance of 2.93, and a normalized surface distance of 0.73, ranking 8th out of 14 entries on the testing leaderboard. We visualized decision boundaries of the model, revealing that voxels adjacent to cerebrospinal fluid and the brainstem required more extreme z scores for lesion prediction. While the model finished in the middle of the leaderboard, we found that the voxel-specific logistic regression model provided interpretability and simplicity. Future work could explore expanding the logistic regression model to be region-specific, rather than voxel-specific, to allow for better fine-tuning of the parameters.

S9. NGZVH

This is a variant of frimpz, civalab apply Swin-UNETR for lesion segmentation in 3D brain images from the BONBID-HIE dataset. Swin-UNETR, which integrates vision transformers, addresses the challenges of capturing long-range dependencies that are common in traditional FCNNs by using a shifted window self-attention mechanism. The model takes a space-normalized 2-channel input (ADC and zADC) and formulates the segmentation task as a sequence-to-sequence problem. To improve generalization, we perform 10-fold cross-validation, training the model on 75 cases and validating on 9 cases per fold. An ensemble of 10 trained models is used to further boost performance. During training, we apply data augmentation techniques like random rotations and flips and perform intensity normalization. The model is optimized using Dice loss, and inference is carried out with a sliding window approach. All training runs are executed on a high-performance computing server equipped with NVIDIA A100 GPUs.

S10. ASHWIN_DHAKAL & CIVA - UNET-3D FOR ACCURATE LESION SEGMENTATION

Algorithm and Data: Hypoxic ischemic encephalopathy (HIE) is a frequently observed brain injury in newborns,

and current methods for delineating HIE brain lesions have demonstrated limited efficacy. Our research introduces the utilization of a UNet-3D model for the segmentation of HIE brain lesions. Our deep neural network processes three-dimensional HIE images, employing two channels (ADC and zADC), to generate predictions regarding the likelihood of lesion occurrence. We propose an innovative hybrid loss function based on the log Hausdorff distance, which is explained in detail in *Section 5.3*. This function aims to regularize the 3D anatomical shape of hypoxic regions and implicitly optimize surface distance metrics (MASD and NSD), while incorporating the DICE loss to optimize the intersection over union metric. Two preprocessing steps were undertaken on the two-channel dataset: space normalization, ensuring uniformity in image resolution, and intensity regularization, achieved through the application of the z-transform method.

UNet-3D [96] is a three-dimensional extension of the UNet [72] architecture. UNet is a convolutional neural network architecture that was originally developed for biomedical image segmentation, particularly for tasks involving 2D images, such as cell segmentation in microscopy images. UNet and UNet-3D have proven to be highly effective tools for medical image segmentation [74], particularly in the context of medical image segmentation. This deep learning architecture excels in its ability to accurately and precisely delineate lesions within 3D medical images, making it a valuable asset in clinical applications. When applied to brain image segmentation tasks, UNet-3D demonstrates its prowess by producing detailed and reliable segmentation of lesions associated with conditions like HIE. The benefits of utilizing UNet-3D in such applications are manifold. First and foremost, it leverages the power of three-dimensional information, capturing spatial context and fine details crucial for accurate segmentation. Furthermore, its adaptability and capacity for feature learning allow it to excel in diverse medical imaging scenarios.

For the particular task of segmenting 3D HIE brain images as described in the reference [57], we skillfully configured the neural network to suit the data. This involved using a space-normalized input comprising two channels (ADC and zADC) and training the network to make predictions about lesions. UNet-3D employs an encoder-decoder architecture enriched with skip connections, enabling it to capture both high-level and low-level features present in the input image. These features are then seamlessly integrated with a decoder based on Fully Convolutional Neural Networks (FCNN) at each resolution level through skip connections. This integration process ensures the creation of a comprehensive and detailed representation of the data, which is conducive to accurate lesion predictions.

Training and Testing: The network underwent training from the beginning to execute 3D segmentation. The input dimensions for the network were set at 128x128x128 (mm). The training dataset supplied by the challenge was divided into 40 segments for training and 37 for validation. In order to assess the impact of the proposed log Hausdorff distance on the network's accuracy, two separate training sessions were conducted. The first network exclusively employed the DICE loss, while the second network utilized a combination of log Hausdorff distance and DICE loss.

The network utilizing only the DICE loss, denoted as *civa* in TABLE III, attained a Dice score of 0.4999, an MASD of 3.4620, and an NSD of 0.6961, securing the 11th position on the testing leaderboard. On the other hand, the network employing a combination of log Hausdorff distance and DICE loss, labeled as *ashwin_dhakal* in TABLE III, achieved a Dice score of 0.4914, an MASD of 3.3839, and an NSD of 0.7018, ranking 10th in the test leaderboard. This highlights that the inclusion of Hausdorff distance as a loss does not notably impact the DICE score. However, it does contribute to the enhancement of MASD and NSD scores.

In the context of the BONBID-HIE challenge, our approach yields competitive results when compared to other methods across all evaluation metrics, ranking within the top 10 methods. This demonstrates the potential of deep learning as an effective approach for segmenting HIE brain lesions.

S11. CIVA

A variation of the ashwin_dhakal approach, utilizing the UNET-3D model for segmenting HIE brain lesions. This deep neural network processes three-dimensional HIE images using two input channels, ADC and zADC, to predict the likelihood of lesion occurrence. The network was trained from scratch for 3D segmentation, with input dimensions set to 128x128x128 (mm). The training dataset provided by the challenge was split into 40 samples for training and 37 for validation. During training, the Dice loss function was employed. The model, referred to as Civa in Table 3, achieved a Dice score of 0.4999, an MASD of 3.4620, and an NSD of 0.6961, securing 11th place on the testing leaderboard.

S12. ARDA.AYDN - SEGRESNET-BASED RECIPROCAL TRANSFORMATION

Algorithm and Data: Segmenting HIE lesions is a challenging task due to their nature (i.e., more diffuse and smaller). Therefore, direct application of some state-of-the-art algorithms such as U-Net [72] gives unsatisfactory results. In this method, we try to solve this problem by proposing a new transform called "Reciprocal Transformation" and combining this transform with SegResNet [97](without the variational auto-encoder (VAE) part).

Neuroradiologists state that low ADC values in the brain are indicative of acute HIE, and a simple threshold segmentation method based on Z_{ADC} maps can get good results, as shown in [57]. For this reason, we tried to penalize higher ADC values and emphasize lower ADC values based on voxel-wise Z_{ADC} maps of images. We developed a new formula to create a new image with more emphasis on lower ADC values using ADC and Z_{ADC} maps.

$$result = \frac{ADC}{(1 + min(|Z_{\text{ADC}}|) + Z_{\text{ADC}})^n} \tag{11}$$

where ADC and Z_{ADC} stand for ADC and Z_{ADC} maps of 2D images (without depth) respectively. These maps have the shape of (C, H, W) where C is the number of channels (in this case C=1), and H and W are the spatial sizes of the

image. We calculate the resulting tensor for each depth (D) value and obtain the resulting tensor has the shape of (C, H, W, D). n ($n > 0$) is another hyper-parameter of our model, and it represents the degree of the transformation. During the experiments, we found that $n = 1.5$ gives the best result for our case.

Since we used the absolute value of the Z-scores to calculate the term $min(|Z_{\text{ADC}}|)$, the result will be equal to the ADC value for the smallest Z-score value. On the other hand, the value of the resulting tensor gradually decreases as the Z-score value increases. This feature of the transformation penalizes higher ADC values and emphasizes lower ADC values.

Training and Testing: Since each magnetic resonance image has different depth values, to make the dataset's shape compatible with network architecture, we padded the depth values of all data to multiples of 8. Furthermore, the addition of extra features to the dataset by concatenating Z_{ADC} and ADC maps with the result of reciprocal transformation significantly improved accuracy scores during the evaluation phase. We normalized the intensity to scale pixel values to a specific range for all data in the dataset and to gain numerical stability. To minimize overfitting during training, random augmentation techniques such as random zoom, random flip, and random rotation with a probability of 0.1, were implemented.

We implemented the SegResNet [97]model (without the VAE part) and random data augmentations using the MONAI (https://monai.io/) framework. We used the Adam optimizer [98] with a learning rate of 1e-5. Since HIE lesions are smaller than the background, we used Tversky loss [99] to handle the class imbalance. For normalization, we used Group Normalization [100] as described in the original paper [97]. The proposed method achieved a mean dice score of 48.26 in the test set of the challenge.

S13. SVCC — nnU-Net approach with reorientation of the data

A nnU-Net [101] neural network was trained in its default configuration using pairs of ss-ADC and z-ADC datasets. The nnU-Net approach is a self-configuring framework with U-Net [72] as its base neural network architecture. The image datasets and the corresponding annotations are oriented with the coordinate axes, and the image spacing was set to 1 mm in the x, y, and z-coordinate directions before using them as the training sets for the nnU-Net approach. In addition, the datasets were converted from MHA format to NIFTI format to be compatible with version 1 of the nnU-Net approach using the *SimpleITK* python module.

A. Neural net training and evaluation

The training data consists of pairs of ss-ADC and z-ADC image volumes acquired from 85 patients, along with their ground truth annotations by experts. The image volume sizes on the training dataset varied from $(128 \times 128 \times 23)$ to $(256 \times 256 \times 60)$. The image resolutions varied from $(0.7031\,mm \times 0.7031\,mm \times 2.2\,mm)$ to $(2.0\,mm \times 2.0\,mm \times 3.0\,mm)$. The nnU-Net framework was trained using *3d_fullres* and *2d* options, which correspond to 3D and 2D versions of the U-Net, respectively. Although nnU-Net framework consists of a 3D U-Net cascade option where a 3D U-Net processes a low-resolution image first and then a high-resolution to refine the predictions, it was not executed based on the suggestion by the nnU-Net framework. Once trained, the nnU-Net utility to find the best configuration was executed, which suggested using only the 3D full-resolution option for the final neural network model.

The neural network yielded average Dice scores of 67.35% and 49.98% for the validation and test sets, respectively.

S14. CIHUAI

A. Data process

We performed our data preprocessing in three stages. First, we resized all scans using a Nearest Neighbor interpolator with a spacing of (0.625, 0.625, 2). Next, we standardized the data using Z-normalization. To overcome the memory limitations of our 24GB GPU when working with cerebral images, we adopted a patch-based training method based on NIFTYNET [102]. We employed Grid Sampling for both training and test data, using patch-size of 32x224x224 with an overlap size of 16x112x112. The augmentation methods encompassed random affine, flipping, random gamma, random blurring, and random bias transformations.

B. Training pipeline

As our base model, we employed SwinUNETR [68], which had a base feature size of 48. Both ssADC and zADC scans were simultaneously trained, with the model input channel set to 1. To optimize our model, we utilized AdamW with a beta of (0.9, 0.99), and employed CrossEntropy as the loss function.

C. Inference pipeline

The final prediction was only based on zADC scans. According to our internal validation results, the Dice score of standalone ssADC and the voxel average of ssADC and zADC were both lower compared to the zADC predictions.