An Ensemble Approach to 3D Deep Learning for High-Resolution Diffusion MRI Segmentation in Neonatal Hypoxic-Ischemic Encephalopathy

Chiranjeewee Prasad Koirala[1*], Sovesh Mohapatra[1,2*], Advait Gosai[1*], Gottfried Schlaug[2,3,4#]

[1]Manning College of Information and Computer Sciences, University of Massachusetts Amherst, MA
[2]Institite for Applied Life Sciences, University of Massachusetts Amherst, MA
[3]Department of Biomedical Engineering, University of Massachusetts Amherst, MA
[4]Department of Neurology, Baystate Medical Center, and UMass Chan Medical School - Baystate Campus, Springfield, MA.

[*]contributed equally
[#]corresponding author

Corresponding author: E-mail: gschlaug@umass.edu;

Abstract

Neonatal hypoxic-ischemic encephalopathy (HIE) has been a major concern around the globe. This paper explores the applications of 3D based deep learning models on high-resolution diffusion based MRI images on the Boston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE). We propose an ensemble method along with a hybrid loss function that consists of combining six different variants of UNet architecture and three different loss functions respectively. Our findings reveal that the ensemble approach, combining different architectures, outperforms single models, leading to improved evaluation metrics. Specifically, in this case, 0.5800 ± 0.2557 dice score was achieved using the ensemble approach. These results show the importance of tailored deep learning techniques in precisely segmenting HIE lesions and lay groundwork for future work to fine-tune models.

1. Introduction

Neonatal hypoxic-ischemic encephalopathy (HIE) poses a major concern in neonatology and pediatrics, leading to significant deficits in oxygen and blood circulation within the infant brain[1,2]. With an incidence rate ranging between 1 to 5 per 1,000 term births, this condition affects a global estimate of approximately 200,000 term-born neonates annually[3]. HIE often causes irreversible neural damage, long-term cognitive deficits, and increased mortality rates[4]. Early diagnosis and precise quantification of the affected neural areas are essential for initiation of the neuroprotective measures and therapeutic hypothermia, which have shown effectiveness in decreasing the extent of the neural tissue damage[1,4,5].

With the increasing use of deep learning (DL) and advancement in the magnetic resonance imaging (MRI) techniques, the precision of quantification in medical imaging has been significantly improved and fastened. However, due to less segmented data in the neonatal HIE lesion domain, the DL methods often suffer from limitations such as sensitivity to hyperparameter tuning, susceptibility to overfitting and variations in imaging and scanning protocols, which makes it different for building robust and generalized models which would applicable over a large and diverse dataset[6].

In this paper, we are focusing on the Boston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE). Here, we present a unique ensemble method using six different variants of UNet architecture, optimizing the individual strengths while mitigating their shortcomings.

2. Methodology

2.1. Preprocessing

We devised three different techniques for preprocessing the data before adding it for the model training.

- Image normalization: All the acquired images were normalized using min-max normalization.
- Central patch extraction: From the normalized image, a central patch of dimensions $64 \times 224 \times 224$ was extracted which was further used for model training.
- Channel stacking: Both the apparent diffusion coefficient (ADC) and Z-score ADC (zADC) maps were combined by stacking them together which resulted in a final image with dimensions of $2 \times 64 \times 224 \times 224$.

2.2. Model Architecture

Our architecture combines six different variants of the UNet keeping in mind each model's individual strengths and weaknesses.

- UNet: A standard 3D UNet model that starts with a base channel depth of 8, which doubles with each encoding step and is progressively halved during the decoding process. There are four encoding and four decoding layers. The architecture incorporates double convolution blocks with group normalization[7].
- ONet: A dual-branch 3D U-Net model. Both branches process data in parallel through four encoding and decoding layers. Outputs from both branches are combined and passed through an output convolution to produce the segmentation mask. Three instances of this model are trained, and two of its variants, one with double convolution and another with GELU activation are trained[8].

2.3. Loss Functions

We utilize the Basnet hybrid loss, a composite loss function combining three key components:

- Binary cross-entropy Loss: This loss function penalizes deviations between the model's probabilistic predictions and the actual binary labels.

$$binary\_cross\_entropy\_loss \; = \; -\frac{1}{N} \sum_{i=1}^{N} \left[y_i log(p_i) \; + \; (1 - y_i) log(1 - p_i) \right]$$

- Multi-Scale Structural Similarity Loss (MS-SSIM): A perceptual loss evaluating the structural similarity between predicted and true images.

$$ms\_ssim\_loss \; = \; 1 - \prod_{j=1}^{M} SSIM(x_j, \; y_j)^{w_j}$$

where:

- $SSIM(x_j, \; y_j)$ is the structural similarity index (SSIM) between images $x_j \; and \; y_j$
- M is the number of scales
- $x_j \, and \, y_j$ are the images x and y downsampled to the jth scale respectively
- w_j are the weights for each scale

- Jaccard Index Loss: This loss function measures the overlap between predicted and true binary masks, emphasizing penalization for inaccurate segmentation areas.

$$jaccard\_loss \; = \; 1 - \frac{\sum prediction\_probabilities * targets}{\sum prediction\_probabilities + \sum targets - \sum prediction\_probabilities * targets}$$

The final loss (basnet hybrid loss) is an aggregation:

$$Basnetloss \; = \; binary\_cross\_entropy\_loss + \; ms\_ssim\_loss \; + \; jaccard\_loss$$

2.4. Training

We employed the Adam optimizer with an initial learning rate of 1×10^{-3}. The learning rate was scheduled to decrease to 75% of its value every 50 epochs. Even though we used a batch size of one, we simulated the effects of larger batch sizes on memory-restricted GPUs by utilizing gradient accumulation. In this approach, the gradients are updated every 4 batches instead of after each batch. Leveraging this method, the models were trained for a total of 400 epochs.

2.5. Ensembling

The sigmoids of the outputs of all models are averaged to yield the final probability mask for every image. A threshold of 0.5 is then applied to this mask to generate the final prediction. Pixels with probabilities greater than or equal to the thresholded value are the final segmented region.

$$prediction = (\frac{1}{M}\sum_{i=1}^{M} sigmoid(model_i)) >= 0.5$$

2.6. Post-Processing

Following ensemble prediction, the output is refined in the post-processing step. Connected regions in the mask that have a pixel size of less than the threshold of 2 are considered noise and discarded.

3. Results

Table 1: Comparison of effects of the post processing on the unseen testing dataset

Method	Dice	NSD	MASD
Ensemble	**0.5800 ± 0.2557**	**0.7379 ± 0.2557**	**2.5993 ± 3.1263**
Ensemble + Post-Processing	0.5767 ± 0.2620	0.7179 ± 0.2620	2.9171 ± 3.7313

In our assessment of various techniques, we developed and submitted the results to the portal for validating the model's prediction on the unseen testing data. The findings revealed that the applied post processing methods don't improve the results on the unseen data as shown in Table -1. Along with that, the developing stage has shown us that the ensemble training approach with the proposed hybrid loss function effectively segments the HIE lesions.

4. Discussion and Conclusion

In this paper, we showcase the importance of using an ensemble approach along with the hybrid loss function. We have trained six different models on the provided dataset and ensemble the results. Due to time constraints, the transfer learning approach was not being used. Hence, the future directions will be focused towards using larger datasets such as dHCP or BCP for first training the models and then performing segmentation which may result in significant improvement in the results from the ensembled models.

References

1. Cowan, F. M., Pennock, J. M., Hanrahan, J. D., Manji, K. P. & Edwards, A. D. Early detection of cerebral infarction and hypoxic ischemic encephalopathy in neonates using diffusion-weighted magnetic resonance imaging. *Neuropediatrics* **25**, (1994).

2. Simbruner, G. *et al.* Systemic Hypothermia after Neonatal Encephalopathy: Outcomes of neo.nEURO.network RCT. *Pediatrics* **126**, (2010).

3. Bao, R. *et al.* BOston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE): Part I. MRI 2 and Manual Lesion Annotation 3. doi:10.1101/2023.06.30.546841.

4. Chock, V. Y., Cho, S. H. & Frymoyer, A. Aminophylline for renal protection in neonatal hypoxic–ischemic encephalopathy in the era of therapeutic hypothermia. *Pediatr Res* **89**, (2021).

5. Guan, B. *et al.* Early diagnosis and outcome prediction of neonatal hypoxic-ischemic encephalopathy with color Doppler ultrasound. *Diagn Interv Imaging* **98**, (2017).

6. Mohapatra, S. *et al.* Meta-Analysis of Transfer Learning for Segmentation of Brain Lesions. (2023).

7. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 9351 (2015).

8. Tan, B. *50.039 Deep Learning Project Report Brain Tumor Segmentation in 3D Pre-Operative MRI Scans*. (2023).