A SwinUNETR-Based Approach with Window-Level Transformation for Segmenting Hypoxic Ischemic Encephalopathy Lesions

A. Senih Yıldırım[1] and Yusuf H. Şahin[2]

[1] Gazi University, Ankara, Turkey
ahmedsenih.yildirim@gazi.edu.tr
[2] Istanbul Technical University, Istanbul, Turkey
sahinyu@itu.edu.tr

Abstract. Hypoxic Ischemic Encephalopathy (HIE) is a critical neonatal brain condition requiring precise diagnosis. Segmenting HIE lesions in brain MRI is challenging due to their diffuse and subtle nature. We propose a SwinUNETR-based method for 3D segmentation, incorporating rigorous preprocessing, advanced data augmentation, and the Focal Tversky Loss to handle class imbalance. Although we initially experimented with a voting-based ensemble, we shifted to a single model after observing inconsistent results on the competition platform. Our approach achieved a Dice score of 0.35 on the test set, and we share our code to aid reproducibility and future research. For more information, visit our GitHub repository: Unet_WindowLevel_Segmentation

Keywords: 3D brain image segmentation · SwinUNETR · Magnetic Resonance Imaging (MRI)

1 Introduction

Deep learning has revolutionized medical imaging, offering effective solutions for tasks like segmentation, with notable success in delineating tumors and organs. However, applying these models to complex conditions like Hypoxic Ischemic Encephalopathy (HIE) remains challenging. HIE, a severe neonatal brain injury, requires precise and timely diagnosis to improve outcomes. The diffuse and subtle nature of HIE lesions complicates segmentation, even for advanced models, while intrapartum hypoxia-ischemia presents further challenges due to its complex pathophysiology and variability [2, 3].

Brain tumors also pose significant health concerns, where early detection is crucial. Variations in tumor location, shape, and size hinder precise segmentation [1]. To address these challenges, we initially tested a SwinUNETR model with window-level transformations and a voting-based ensemble. Although this performed well on local validation, it showed inconsistencies on the competition platform. Thus, we opted for a single model approach to improve generalization. Our method aims to boost segmentation accuracy on the Boston Neonatal Brain Injury Dataset (BONBID-HIE), supporting more accurate diagnostics and treatment planning.

2 Dataset

The Boston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE) is a unique MRI dataset dedicated to the segmentation of HIE lesions. It features Apparent Diffusion Coefficient (ADC) maps alongside voxel-wise Z-scores (ZADC) of these values. The dataset comprises 85 training cases and 4 validation case. The labeled MRI data, sourced from Massachusetts General Hospital (MGH) and Boston Children's Hospital (BCH), captures a wide range of HIE lesion characteristics, providing a comprehensive foundation for training and evaluating segmentation models. [4]

3 Method

3.1 Focal Tversky Loss

The Focal Tversky Loss effectively addresses class imbalance in medical image segmentation, particularly for small lesions. It builds upon the Tversky Index, an extension of the Dice coefficient, incorporating adjustable weights for false positives (FP) and false negatives (FN).

Tversky Index: The Tversky Index is calculated using true positives (TP), false negatives (FN), and false positives (FP):

$$\text{Tversky Index} = \frac{\text{TP}}{\text{TP} + \alpha \cdot \text{FN} + \beta \cdot \text{FP}}$$

where α and β are parameters that balance the importance of FN and FP.

Focal Tversky Loss: The Focal Tversky Loss incorporates a focusing parameter γ to emphasize difficult-to-segment regions:

$$\text{Focal Tversky Loss} = (1 - \text{Tversky Index})^{\gamma}$$

Adjusting γ helps the model focus on challenging areas, enhancing segmentation performance under class imbalance [5].

3.2 Window Level Transformation

Window Leveling is a technique used to enhance the visibility of specific intensity ranges in CT images by adjusting the grayscale values. The window level (WL) sets the midpoint of the intensity range, and the window width (WW) defines the range around this midpoint. This method enhances contrast and preserves critical details within a defined Hounsfield unit range, aiding in better segmentation outcomes [6].

In our project, we applied this transformation to improve the segmentation of Hypoxic Ischemic Encephalopathy (HIE) lesions in brain MRI. Using the formula:

$$\text{Transformed Value} = \frac{\text{Original Value} - \left(\text{Window Level} - \frac{\text{Window Width}}{2}\right)}{\text{Window Width}}$$

we experimented with different WL and WW settings to maximize lesion visibility. This adjustment was crucial for reducing background noise and enhancing subtle lesion details, significantly improving our model's performance.

3.3 Model

Our segmentation model is based on the SwinUNETR architecture, which leverages the strengths of transformer-based networks for 3D medical image analysis. The model is designed to efficiently capture both local and global contextual information, making it suitable for complex segmentation tasks like Hypoxic Ischemic Encephalopathy (HIE) lesions. We use an input size of $128 \times 128 \times 128$, with two input channels (the transformed image and the Z-map) and a single output channel for lesion segmentation. The feature size is set to 48, and we enable checkpointing to optimize memory usage during training.

The model is trained using the AdamW optimizer with a learning rate of 1×10^{-4} and a cosine annealing learning rate scheduler to dynamically adjust the learning rate. To address class imbalance in the lesion segmentation task, we use the Focal Tversky Loss, which emphasizes harder-to-segment regions. The Dice Metric is used for evaluation, measuring the overlap between the predicted and ground truth masks.

3.4 Voting Mechanism

To enhance segmentation robustness and accuracy, we employ a voting mechanism that utilizes multiple models trained with different window level settings. Specifically, we use the top three models, each trained with distinct window level and width parameters, to better account for image contrast variations and improve generalization.

During inference, each model applies its own window level transformation, producing three separate segmentation predictions. We then use majority voting: a voxel is labeled as a lesion if at least two out of three models agree; otherwise, it is labeled as background. This approach reduces the impact of individual model biases, ensuring more reliable segmentation and consistent performance across varying image intensities.

4 Experimental Results

Initially, we experimented with a voting system for our segmentation task. Using this system on our local dataset, we achieved a promising 0.7170 Dice score. However, upon uploading our approach to the competition's platform, we observed a significant drop in performance. Consequently, to ensure more consistent results, we decided to submit our code using a single model.

During the Docker Validation phase, our single model approach achieved a 0.55 Dice score. Similarly, in the Test Phase, the single model yielded a Dice score of 0.35. These results indicate that the model's performance on the platform was considerably lower than what we observed locally.

5 Conclusion

In this study, we developed a deep learning-based model for lesion segmentation in brain MRI images. Although we initially experimented with a voting system, we had to switch to a single model approach due to the lower-than-expected performance on the competition platform. Despite the discrepancy between the success achieved in our local environment and the performance on the platform, these experiments provided valuable insights into the strengths and weaknesses of our model.

The results indicate that the model performed better on local data but exhibited a performance drop on the platform. In the future, we plan to work on improving the model's generalizability and achieving higher accuracy by focusing on data diversity and model optimization.

References

1. Amin, J., Sharif, M., Raza, M., Saba, T., & Anjum, M. A.: Brain Tumor Detection Using Deep Learning: A Comprehensive Review. *Journal of Medical Imaging and Health Informatics* **12**(3), 1–15 (2022). doi:10.1166/jmihi.2022.3502
2. Lee, A. C. C., Kozuki, N., Blencowe, H., Vos, T., Bahalim, A., Darmstadt, G. L., Niermeyer, S., Ellis, M., Robertson, N. J., Cousens, S., & Lawn, J. E.: Intrapartum-related neonatal encephalopathy incidence and impairment at regional and global levels for 2010 with trends from 1990. *Pediatric Research* **74**(S1), 50–72 (2013). `https://pubmed.ncbi.nlm.nih.gov/24366463/`
3. Graham, E. M., Ruis, K. A., Hartman, A. L., Northington, F. J., & Fox, H. E.: A systematic review of the role of intrapartum hypoxia-ischemia in the causation of neonatal encephalopathy. *American Journal of Obstetrics & Gynecology* **199**(6), 587–595 (2008). `https://pubmed.ncbi.nlm.nih.gov/19084096/`
4. Bao, R., Song, Y., Bates, S. V., Weiss, R. J., Foster, A. N., Jaimes Cobos, C., Sotardi, S., Zhang, Y., Gollub, R. L., Grant, P. E., Ou, Y.: BOston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE): Part I. MRI and Manual Lesion Annotation. *bioRxiv* (2023). `https://doi.org/10.1101/2023.06.30.546841`
5. Submitted on 9 Oct 2020. *Convolutional Recurrent Residual U-Net Embedded with Attention Mechanism and Focal Tversky Loss Function for Cancerous Nuclei Detection.*
6. Apivanichkul, K., Phasukkit, P., Dankulchai, P.: Performance Comparison of Deep Learning Approach for Automatic CT Image Segmentation by Using Window Leveling. In: Proceedings of the 2021 Biomedical Engineering International Conference (BMEiCON-2021), pp. 1–5. IEEE (2021). `https://doi.org/10.1109/BMEiCON53485.2021.9745252`