LESION SEGMENTATION FOR HYPOXIC ISCHEMIC ENCEPHALOPATHY

E. Almazán Sánchez[1], G. Ots Rodríguez [2], J. Villoldo Fernández [3]

[1] Biomedical Engineering, Rey Juan Carlos University, Madrid, Spain, e.almazan.2020@alumnos.urjc.es

[2] Biomedical Engineering, Rey Juan Carlos University, Madrid, Spain, g.ots.2020@alumnos.urjc.es

[3] Biomedical Engineering, Rey Juan Carlos University, Madrid, Spain, ja.villoldo.2020@alumnos.urjc.es

Abstract

Hypoxic Ischemic Encephalopathy (HIE) remains a substantial global concern, causing severe brain injury in neonates despite therapeutic hypothermia. High mortality and morbidity rates underscore the urgency for advancements in clinical care. This study focuses on the segmentation of HIE lesions in 2D brain images obtained from 3D MRI images, employing a comprehensive three-pronged approach.

Firstly, a U-Net architecture is deployed, aiming to segment annotated expert masks. Secondly, a computationally efficient convolutional neural network (CNN) that only contains convolutional layers extracts features from the input images, serving as input for a Random Forest model. This innovative methodology significantly reduces computational costs and prediction times. Finally, a Transformer-based CNN emphasizes key regions, prioritizing areas deemed more clinically relevant.

Challenges in HIE lesion segmentation arise from the diffuse and subtle nature of abnormalities in MRI scans. Current methods yield suboptimal accuracy (Dice overlap ~0.5), necessitating refinement for enhanced prognostic accuracy. In order to tackle this issue, the BONBID-HIE lesion segmentation challenge leverages publicly available 3D MRI data, aiming to elevate diagnostic precision for HIE. [1, 2]

Our proposed methodologies not only encompass traditional deep learning approaches but also introduce a novel fusion of CNNs and Random Forests, optimizing computational efficiency. The Transformer-based CNN further refines segmentation by assigning varying weights to different regions.

1. Introduction

Hypoxic Ischemic Encephalopathy (HIE) stands as a formidable global health challenge, imposing severe brain injury on neonates despite therapeutic interventions such as hypothermia. This condition is marked by:

- Elevated mortality rate.
- Elevated morbidity rate.

These issues emphasize the critical need for advancements in clinical care strategies. To address the complexities of HIE, accurate segmentation of the associated lesions becomes a pivotal step in enhancing patient outcomes.

The imperative for lesion segmentation within the clinical context of HIE is underscored by its multifaceted impact on prognosis, neurological symptom understanding, and the timely prediction of therapeutic responses. Firstly, a more accurate estimation of prognosis is essential for tailoring individualized treatment plans and informing families about potential outcomes. Secondly, a nuanced understanding of neurological symptoms, intricately linked to lesion distribution, is vital for comprehensive patient care.

Despite the potential benefits, HIE lesion segmentation in Magnetic Resonance Imaging (MRI) remains a formidable challenge. The lesions often manifest as diffuse and small abnormalities, contributing to the intricacies of accurate segmentation. Present segmentation methodologies fall short, yielding suboptimal accuracy, with Dice overlap metrics hovering around 0.5, emphasizing the need for refinement to achieve precise prognostic insights.

Addressing this critical gap, the BONBID-HIE Lesion Segmentation Challenge has been initiated, providing publicly available 3D MRI data. This collaborative effort aims to propel advancements in diagnostic precision for HIE, fostering innovation within the scientific community.

In our pursuit of improved segmentation methodologies, this study introduces a three-methodology approach:

- A U-Net architecture segment lesions outlined by expert annotations.
- A computationally efficient convolutional neural network (CNN) used to extract features from the input images that serve as input for a Random Forest model. This novel strategy significantly reduces computational costs and prediction times, addressing practical challenges in real-world clinical applications.
- A Transformer-based CNN that prioritizes clinically relevant regions, offering a nuanced approach to lesion segmentation.

These methodologies, while incorporating a traditional deep learning model, also present an innovative integration of CNNs and Random Forests, showcasing versatility and efficiency. By contributing to the refinement of HIE lesion segmentation, this research not only addresses an immediate clinical need but also contributes to the broader landscape of medical image analysis.

2. Materials and methods

Regarding the materials, we used the 1st Boston neonatal brain injury dataset for hypoxic ischemic encephalopathy, named BONBID-HIE Lesion Segmentation database.

This dataset provides diverse and annotated images crucial for the development and evaluation of mask segmentation models. It contains a wide range of scenarios, capturing varying lighting conditions, perspectives, and mask types.

The data was obtained from Massachusetts General Hospital. It includes MRIs from different scanners (Siemens 3T and GE 1.5T), different MRI protocols, and from patients of different races/ethnicities and ages (0-14 days postnatal age). Inclusion criteria were: (1) term-born (at physician discretion) (2) clinical diagnosis of HIE; (3) initially treated at MGH between 2001 and 2018; (4) no comorbidities such as hydrocephalus or congenital syndromes; and (5) high-quality MRI acquired in Day 0-14 after birth (visually checked by RW, AF, YO). Exclusion criteria were: (1) excessive motion artifacts or missing images; (2) secondary HIE diagnosis to a primary perinatal stroke.

The dataset is already split into training, validation and test subsets. However, as we had to ask for the images, they only sent us the training images, which are 85, so we decided to work with that amount as if no training-validation-test split had been done before.

It is composed of 3D images from 85 patients. Each group contains three 3D images:

- 1ADCss: Skull stripped Apparent Diffusion Coefficient (ADC) map.
- 2ZADC: ZADC map. These images were developed from the ADC map to normalize and make ADC values comparable across brain voxel locations.
- 3LABEL: Expert lesion annotations. HIE lesions were manually annotated (by physicians >3 years of experience) as a binary mask on the 3D ADC maps in the patient's raw image space, using the MRICroN software. The annotations started from the axial slice and were subsequently modified in the coronal and sagittal planes for the 3D integrity of lesion regions. This image is what we want to predict with our models (we will talk about them as 'masks').

Moving on to methods, first we loaded the images with Skimage library. Then, preprocessing techniques are applied to optimize the quality and homogeneity of brain images. Subsequently, data augmentation is employed to expand the dataset, enhancing its robustness and diversity.

Following the preparatory steps, we focused on three distinct models that constitute the core of our methodology. Firstly, a U-Net architecture is deployed for segmentation, followed by a Random Forest-based model where an efficient set of convolutional layers acts as the feature generator. Lastly, a tailored Transformer is introduced to highlight clinically relevant areas during the segmentation process.

2.1. Preprocessing techniques

In order to prepare the dataset for using it at our segmentation models, we performed several preprocessing techniques:

- Splitting the whole dataset into training (80% of all the 3D images), validation (20% of the training subset) and test (the remaining 20% of all the 3D images) subsets.
- Normalization of the image's intensities.
- Resizing of the images, in which we convert all the slices of the 3D images into (128,128) shape since otherwise we had slices with different shapes.
- Getting 2D slices from the 3D images.

Due to the lack of memory and computational resources, the use of the volumetric (3D) data for our objective was not possible. Hence, their 2D slices were obtained instead. As a drawback, we lost some information in the conversion from 3D to 2D images. However, we expanded the size of our database since we included almost all of the slices of each 3D image.

2.2. Data Augmentation

The UNET model will be further fitted with data augmented images. Data augmentation allows to overcome limitations on the learning process of the model due to insufficient data, by increasing the number of samples and adding more variability to the dataset. This is achieved by applying random transformations to the original images (color, texture, intensity level, geometry, etc). Therefore, data augmentation becomes a useful technique to prevent overfitting and improve the robustness of the model.

To create the augmentation pipeline, first the types and values of the transformations that will be applied to our original images are defined. Transformations ought to be realistic, mirroring scenarios that might be found in clinical cases, to ensure the trained models are reliable and safe for clinical use. The pipeline is created considering that the same random transformations on a random set of images also need to be applied to the corresponding set of masks.

Then, instances of the *ImageDataGenerator* class from *Keras* are created. They are configured with the augmentation parameters specified earlier. Next, training images and masks are fitted to the data generators and a seed parameter is given to ensure reproducibility from images to masks.

Finally, batches of augmented data for training are generated with the *flow* method. The same procedure is applied to validation and testing sets only that in these cases data augmentation is not applied, simply have training and validation images and masks in the same format when fitting the UNET. Evaluation of the model is performed on the original test set.

2.3. Segmentation models

U-NET

[3] The first method we used is a U-Net model. This convolutional neural network (CNN) architecture is designed to discern intricate spatial patterns and capture hierarchical features.

The architecture begins with an input layer (128x128x1), followed by a series of convolutional layers interspersed with rectified linear unit (ReLU) activation functions. Consecutive convolutional layers (Conv2D) with kernel sizes of 3x3 and 'same' padding allow the network to progressively extract and refine features. MaxPooling2D layers down sample the spatial dimensions, aiding in feature abstraction.

The contracting path of the U-Net culminates in a bottleneck layer, represented by Conv2D layers with an expansive number of filters (1024). Subsequently, the expansive path involves up sampling (UpSampling2D) and concatenation operations, facilitating the recovery of spatial information while maintaining contextual details.

This U-Net model concludes with a Conv2D layer using a 1x1 kernel and a sigmoid activation function, producing a probability map for lesion presence. The model is trained using the Adam optimizer and binary cross entropy loss, with the evaluation metric being the Dice coefficient, that measures the degree of overlap of the prediction with the mask.

The model training process is governed by early stopping, monitoring the validation Dice coefficient to ensure optimal convergence while preventing overfitting.

Random Forest

[4] Moreover, the second method we used is the combination of a CNN that only has convolutional layers with a random forest to predict the segmentation. This methodology seamlessly integrates the feature extraction capabilities of a CNN with the robust classification power of a Random Forest.

Initially, a CNN with two convolutional layers is employed as a feature extractor. The trained feature extractor is then used to predict features from the training dataset. Subsequently, these features are reshaped and used alongside the corresponding annotated masks to train a Random Forest classifier. This process is iteratively applied to the multiple training images, ensuring the model learns the intricacies of HIE lesion features.

The resulting Random Forest model is saved and later loaded for predicting lesion masks on unseen test data. The feature extractor, a pre-trained segment of the CNN, is utilized to extract features from the test images. These features are then input into the previously trained Random Forest model, producing predictions for lesion presence.

Evaluation of the ensemble model is conducted using the Dice coefficient, a metric assessing the overlap between predicted and ground truth masks. The calculated Dice coefficients from all the images are averaged, providing a comprehensive measure of the model's segmentation accuracy across the entire test dataset.

U-NETR

[5] Lastly, the third method defines a neural network architecture known as UNETR, which combines elements of a U-Net and a Transformer for 3D image segmentation.

A Transformer is a type of neural network architecture introduced in the paper "Attention is All You Need" by Vaswani et al. Its key components include self-attention mechanisms and feed-forward neural networks. It's widely used in natural language processing and has been adapted for computer vision tasks.

The self-attention mechanism takes the input hidden states, applies linear transformations to obtain query, key, and value tensors, computes attention scores, and produces an attention output. It allows the model to weigh different parts of the input sequence differently during processing.

Also, a simple multi-layer perceptron (MLP) with a linear layer, activation function (GELU), and dropout, is introduced, as well as a feed-forward neural network that applies linear transformations to the input features.

Additionally, it includes patch and position embeddings, which are obtained using 3D convolutional layers, in order to handle the input data. It includes a layer normalization for both the attention and the MLP parts.

Then, the `Transformer` module, which consists of multiple blocks stacked on top of each other and an embedding layer, with the forward function iteratively applies it to the input hidden states and extracts intermediate outputs at specified layers.

Finally, the `U-NETR` combines the Transformer and U-Net architectures. It uses a Transformer encoder to extract features from the input, and then a U-Net decoder to generate the final segmentation output.

The overall workflow of the algorithm is as follows:

1. The input data is passed through the Transformer encoder (`Transformer` module).

2. Intermediate outputs are extracted at specified layers during the Transformer processing.

3. The U-Net decoder takes these intermediate outputs, upsamples and combines them to generate the final segmentation output.

Some important parameters are highlighted:

- img_shape: Shape of the input 3D image.
- input_dim: Number of input channels.
- output_dim: Number of output channels.
- embed_dim: Dimensionality of the embeddings.
- patch_size`: Size of the patches used in the Transformer.
- num_heads: Number of attention heads in the Transformer.

- dropout: Dropout rate.
- num_layers: Number of layers in the Transformer.
- ext_layers: List of layers from which intermediate outputs are extracted.

Thus, the strengths of both, Transformer and U-Net architecture for 3D image segmentation. The model learns hierarchical representations through the Transformer encoder and refines the segmentation output using the U-Net decoder. [6]

3. Results

With the implementation of a tailored benchmark UNET model, results from the state-of-the-art described in the BONBID-HIE article [2], were successfully replicated. HIE lesion segmentation involves challenges due to the diffuse and small nature of the lesions. Our model achieved a Dice Coefficient of 0.41 training with the original images and improved to 0.43 after further fitting with the data augmentation. These results were obtained by using ZADC maps, that address anatomical variations in ADC values, and demonstrated superior segmentation accuracy compared to traditional ADC maps. Said finding also accords with the article.

The graphs below depict the learning process of the UNET over successive epochs, portraying its good generalization capability.

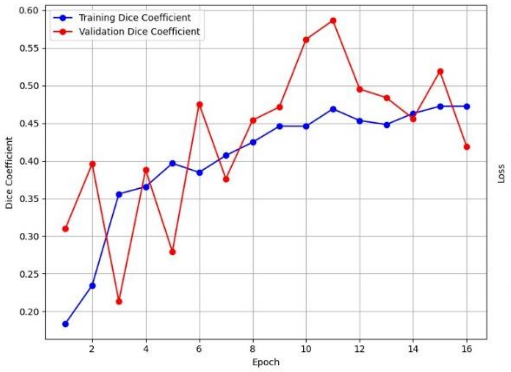

Figure 1. *Training and Validation Dice Coefficient in UNET with original data*

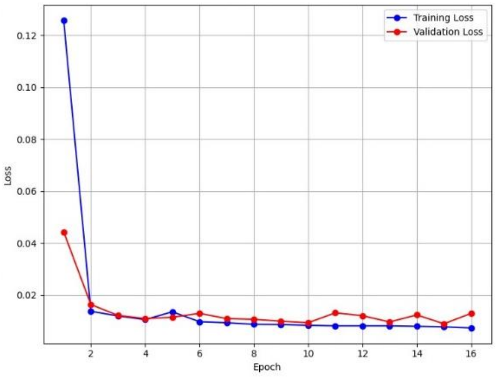

Figure 2. *Training and Validation Binary Cross-Entropy Loss in UNET with original data*

With the goal of improving these results we attended to other methods of segmentation. For the random forest model, which uses a simple convolution architecture to extract features and classify the fixed points, the results obtained were worse, with a Dice Coefficient of 0.29. However, computation time must also be considered. For the UNET the training time was approximately 4 hours whereas for the Random Forest the training time was in the order of minutes. So, there is a major time/performance trade-off.

Below a boxplot is shown visually comparing the performance of the two segmentation methods. This was obtained by computing the Dice Coefficient of the ground truth with predicted masks from both models. The distribution of coefficient scores shows that the UNET has higher median value than Random Forest but also higher variability of results.

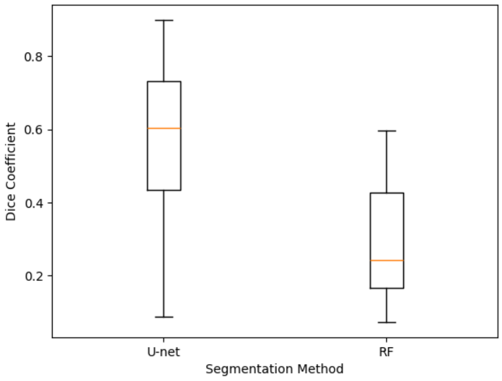

Figure 3. *Dice Coefficient Comparison Boxplots between UNET and RF Models*

The final attempt to improving the benchmark for the field was developing a UNETR architecture. It is believed that this model, which combines the spatial feature capturing strength of U-Net with the Transformer's attention mechanisms for understanding relationships between distant pixels, would adapt much better to the HIE segmentation task. Unfortunately, these assumptions are based on theoretical concepts since due to computational limitations, testing of this approach was not fulfilled.

To bring this section to an end a pair of predicted versus true masks is showed for each model.

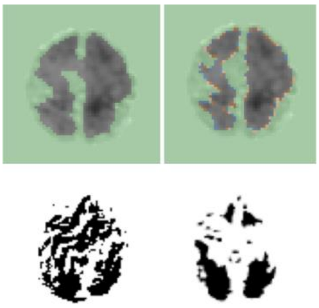

Figure 4. *Example of Predicted and Ground Truth Masks. Predicted Masks are on the left and True masks are on the right. On top is the UNET and below the Random Forest.*

4. Discussion

At the beginning of the research, the challenges posed by limited computational resources in implementing and evaluating advanced models, particularly the UNETR, for the segmentation of hypoxic ischemic encephalopathy (HIE) lesions from 3D MRI data, appeared.

Due to resource constraints, a practical approach was adopted, involving the conversion of 3D images into 2D slices. Thus, both models, including UNET and Random Forest, were implemented in 2D due to their computational efficiency.

Despite efforts to create a UNETR model, it was not practically implemented on the images due to persistent computational errors related to memory and time constraints. The inability to conduct a real-world application of UNETR underscores the challenges posed by resource limitations.

Even though the BONBID-HIE lesion segmentation challenge, where advanced models like UNETR have shown promising results, there is an acknowledgment that theoretical understanding alone is insufficient. The potential benefits of UNETR in providing superior segmentation results are acknowledged, but the lack of practical implementation on the available images highlights the need for additional computational resources.

5. Conclusion

The critical role of accurate segmentation in addressing the clinical challenges associated with HIE should be emphasized. The resource constraints led to the adoption of 2D models, but the inability to practically implement UNETR indicates the necessity for enhanced computational capabilities.

The observed limitations of existing state-of-the-art models, such as UNET, and the suboptimal performance of alternative approaches like Random Forest combined with neural networks underscore the need for further advancements.

Thus, the development and implementation of sophisticated algorithms, like UNETR, is imperative, encouraging and advocating for further research and investment in computational infrastructure.

This will harness the potential of advanced algorithms, enabling the exploration of cutting-edge models for accurate medical image segmentation, ultimately leading to improved diagnostic precision, enhanced patient care and prognosis, and boost clinical outcomes in conditions or diseases such as HIE.

Acknowledgments

The authors extend their heartful gratitude to Professors Norberto Malpica, Borja Rodríguez Vila, and Elena Aparicio Esteve for their invaluable guidance and expertise in the 'Medical Image Analysis' course. Their dedication has significantly enriched our knowledge.

Special thanks are also extended to the Rey Juan Carlos University for fostering an environment conducive to learning and exploration.

References

[1] M. P. Ellen Grant, «Grand Challenge,» [En línea]. Available: https://bonbid-hie2023.grand-challenge.org/bonbid-hie2023/.

[2] R. Bao, «BOston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE): Part I. MRI and Manual Lesion Annotation,» *PubMed Central,* 2023.

[3] «U-Net: Convolutional Networks for Biomedical Image Segmentation,» [En línea]. Available: https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/.

[4] DigitalScreeni, «Convolutional filters + Random Forest for image segmentation.,» [En línea]. Available: https://www.youtube.com/watch?v=5ct8Yqkiioo&ab_channel=DigitalSreeni.

[5] K. He, «ScienceDirect,» [En línea]. Available: https://www.sciencedirect.com/science/article/pii/S2667102622000717?via%3Dihub.

[6] A. Hatamizadeh, «Cornell University,» 2021. [En línea]. Available: https://arxiv.org/abs/2103.10504.