Skull-Base Tumor Segmentation and Classification

Harshit[1] u1471783@umail.utah.edu
[1] *School of Computing, University of Utah*
Tyler Richards[2] tyler.richards@hsc.utah.edu
[2] *Department of Radiology, University of Utah School of Medicine*
Tolga Tasdizen[3] tolga@sci.utah.edu
[3] *Scientific Computing and Imaging Institute, University of Utah*

Editors: Under Review for MIDL 2025

Abstract

Skull-base tumor segmentation is a critical yet challenging task due to the rarity of such tumors and the complexity of the anatomical region. This work focuses on developing an automated segmentation and classification framework for central skull base tumors using multimodal 3D imaging data, including CT and MRI scans. By leveraging state-of-the-art deep learning architectures tailored for 3D medical imaging, we aim to accurately delineate tumor boundaries while preserving anatomical context. Our approach incorporates advanced data augmentation techniques and loss functions optimized for handling imbalanced datasets. Initial experiments demonstrate the framework's capability to achieve precise tumor segmentation, highlighting its potential to assist clinicians in diagnosis and treatment planning. Future work will focus on expanding the dataset and integrating interpretability mechanisms to further enhance clinical adoption.

Keywords: Skull-base tumors, Medical Imaging, UNets.

1. Introduction

Skull base tumors, located at the interface of critical neurovascular and bony structures, are among the most challenging types of tumors to diagnose and treat. Their rarity, combined with the high variability in shape, size, and location, necessitates precise imaging and computational tools for effective management. Traditional manual segmentation is time-consuming, error-prone, and dependent on expert radiologists, highlighting the need for automated solutions.

Deep learning has emerged as a transformative tool in medical imaging, particularly in segmentation tasks. Convolutional Neural Networks (CNNs) (Cun et al., 1990) and their derivatives have shown remarkable performance in processing 3D volumetric data, making them well-suited for medical image analysis. However, the intricate anatomy of the skull base and the limited availability of annotated datasets present significant challenges. In this work, we aim to address these challenges through a tailored approach leveraging multi-channel imaging modalities, data augmentation, and transfer learning.

2. Methodology

2.1. Dataset and Preprocessing

We developed our own dataset from scratch, comprising MRI and CT scans of skull base tumors. This dataset includes T1-weighted, T2-weighted, T1-contrast-enhanced, and FLAIR sequences, all meticulously annotated by expert radiologists. To ensure consistency and quality, we standardized the data through intensity normalization, resampling to a uniform voxel size, and spatial alignment using rigid and affine transformations. Additionally, data augmentation techniques, such as rotation, flipping, and elastic deformations, were applied to improve model generalization and robustness.

2.2. Model Architecture

The segmentation model is based on the 3D UNet architecture (Ronneberger et al., 2015), which has been adapted to the specific requirements of skull base tumor segmentation. The model accepts both single-channel and multi-channel input configurations (e.g., T1-T2, T1-T1FS, and T2-T1-T1FS combinations) and outputs a binary segmentation mask. Transfer learning was implemented by initializing the model with pre-trained weights from a BRATS (Menze and et al, 2015; Bakas et al., 2017; Bakas and et al, 2019) brain tumor segmentation model, ensuring a better starting point for fine-tuning. Key architectural modifications include:

- Incorporating DiceCELoss, a composite loss function combining Dice loss and Cross-Entropy loss, optimized for class imbalance.
- Employing the sliding window inference method to handle large 3D volumes effectively.

2.3. Training and Optimization

The training pipeline involved the AdamW optimizer (Loshchilov and Hutter, 2019) with a learning rate scheduler based on cosine annealing. The model was trained for 5000 iterations with periodic validation. Dice similarity coefficient (DSC) was used as the primary evaluation metric. The training process was monitored and logged using WandB for detailed analysis and visualization. To ensure effective learning, we adopted a phased approach:

1. Initial fine-tuning of the pre-trained model on the skull base dataset.
2. Enabling gradient updates for all layers to refine the network for the target domain.

2.4. Classification Pipeline

In addition to segmentation, we developed a classification pipeline to determine tumor subtypes as primary or secondary tumors using features extracted from the fine-tuned 3D UNet model. Following the training of the segmentation model, the same UNet was repurposed to extract deep, case-specific feature representations. These extracted features served as input for a linear classifier, specifically designed to predict tumor subtypes. By leveraging the spatial and contextual information captured during segmentation, the classifier was trained using a cross-entropy loss function, enabling it to effectively complement the segmentation model and enhance overall diagnostic utility.

3. Results

We experimented with various input configurations, including single-channel inputs (T1, T2, T1FS, and FLAIR) and multi-channel inputs (T1-T2, T1-T1FS, T2-T1-T1FS). Table 1 presents the results of these configurations, showcasing the segmentation performance across different modalities. Despite leveraging advanced deep learning techniques and multi-channel input configurations, the model's performance, as indicated by the Dice scores, was relatively modest. The limited amount of training data, consisting of less than 90 annotated cases, is likely a significant factor contributing to these results.

The highest performance was observed for the T2 input configuration, achieving a Dice score of 0.56, while FLAIR had the lowest performance at 0.47. Multi-channel configurations, such as T1-T2 and T2-T1-T1FS, did not significantly improve performance compared to single-channel inputs, with Dice scores of 0.48 and 0.49, respectively. These results underscore the challenges posed by the limited dataset size and the complex nature of skull base tumors.

We also evaluated the classification pipeline using different input configurations, reporting validation performance metrics such as F1-score and accuracy for predicting tumor subtypes as primary or secondary. Among the single-channel inputs, the T2 configuration demonstrated the best performance, achieving a validation F1-score of 0.89 and accuracy of 0.80, highlighting its superior discriminatory capability. The T1FS input yielded a validation F1-score of 0.71 and accuracy of 0.55, while T1 resulted in a validation F1-score of 0.63 and accuracy of 0.45. These results indicate that T2 imaging provides the most reliable features for tumor subtype classification, significantly outperforming the other modalities.

Input Configuration	**Dice Score**
T1	0.53
T2	0.56
T1FS	0.52
FLAIR	0.47
T1-T2	0.48
T1-T1FS	0.54
T2-T1-T1FS	0.49

Table 1: Performance comparison of single-channel and multi-channel input configurations.

4. Conclusion

Automated segmentation of skull base tumors is a crucial step toward improving diagnostic accuracy and treatment planning. This study demonstrates the feasibility of leveraging deep learning for this task, offering a robust and scalable solution. With further advancements in model design and dataset availability, this approach has the potential to become a standard tool in neuro-oncology.

References

Spyridon Bakas and et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge, 2019. URL https://arxiv.org/abs/1811.02629.

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific Data*, 4, 09 2017. doi: 10.1038/sdata.2017.117.

Y. Le Cun, B. Boser, J. S. Denker, R. E. Howard, W. Habbard, L. D. Jackel, and D. Henderson. *Handwritten digit recognition with a back-propagation network*, page 396–404. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990. ISBN 1558601007.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019. URL https://arxiv.org/abs/1711.05101.

Bjoern H. Menze and et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015. doi: 10.1109/TMI.2014.2377694.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. URL https://arxiv.org/abs/1505.04597.