Brain MRI Segmentation

P21: Tanya Dubey, Aman Waoo, Girish Wangikar, Sainath Gorige
Department of Computer Science, North Carolina State University
Raleigh, NC 27695
{tdubey, awaoo, gwangik, sgorige}@ncsu.edu

1 Background and Introduction

Lower-grade gliomas (LGG) are a group of brain tumors with diverse molecular subtypes and clinical behaviors. Accurate prediction of patient outcomes based on histopathological data is challenging due to inter-observer variability. To address this issue, recent research has explored the identification of genomic subtypes within LGG tumors and their association with specific shape features. In this study, we explore a novel method for quantifying tumor imaging characteristics using deep learning-based segmentation and examine whether these features can predict the genomic subtypes of LGG.

Lower-grade gliomas, which include WHO grade II and grade III brain tumors, are characterized by their infiltrative nature and potential for recurrence and progression to higher-grade lesions. Accurate prediction of patient outcomes for LGG based solely on histopathological data is problematic due to the inherent variability in interpretations. To address this challenge, researchers have explored the molecular classification of LGG based on DNA methylation, gene expression, DNA copy number, and microRNA expression, in addition to established markers like IDH mutation and 1p/19q co-deletion. These molecular subtypes have been shown to correlate with disease course and overall survival.

Radiogenomics, a burgeoning field in cancer research, aims to investigate the relationship between tumor genomic characteristics and medical imaging. Recent studies have identified associations between MRI-derived tumor shape features and genomic subtypes, providing a potential avenue for non-invasive prediction of tumor behavior. However, many of these studies relied on manual tumor segmentation, which is labor-intensive, time-consuming, and prone to inter-observer variability.

Significant advances in the field of Deep learning have demonstrated the potential to transform the field of radiogenomics by enabling high-quality, automated tumor segmentation in LGG. Such automated segmentation could facilitate the identification of genomic subtypes, offering a faster, cost-effective, and consistent approach compared to manual segmentation. In this study, we leverage deep learning techniques to explore an automated algorithm for quantifying tumor shape features and assess whether these features can predict the molecular subtypes of LGG. The ability to derive imaging-based biomarkers for tumor genomics could provide valuable information to clinicians non-invasively and improve the stratification of tumors, particularly in cases where surgical resection is not feasible. Brief introductions to the considered data sets are given in the subsequent sub-sections.

In the context of our project on the association of genomic subtypes of lower-grade gliomas with shape features extracted by deep learning algorithms, the article "Brain Tumor Segmentation Based on Refined Fully Convolutional Neural Networks with A Hierarchical Dice Loss" provides valuable insights. This study presents significant advancements in brain tumor segmentation, focusing on the refinement of fully convolutional neural networks (FCNNs) and the introduction of a novel hierarchical dice loss function. The modifications to classic FCNN architectures, including the integration of residual structures and batch normalization layers, demonstrate improved performance in key segmentation metrics. This is particularly pertinent for detecting enhancing tumors, a crucial aspect in the characterization of gliomas.

The relevance of this research to our project lies in its exploration of complex brain tumor segmentation challenges. Brain tumors, particularly gliomas, present a significant medical challenge due to their variable shapes, locations, and modalities. Prior research has underlined the difficulties in man-

ual segmentation, which is often time-consuming and requires expert knowledge. The advancements in neural network architectures and loss functions proposed in this research could directly inform the development of deep learning models for extracting shape features from lower-grade gliomas, aiding in the correlation of these features with genomic subtypes.

2 Method

Initially, we employed the VGG16 architecture as our baseline model. VGG16, standing for Very Deep Convolutional Networks, is a renowned convolutional neural network (CNN) with 16 layers. The choice of VGG16 was motivated by its well-established performance in image recognition tasks, including its remarkable 92.7% top-5 test accuracy on the ImageNet dataset. We used the VGG16 model from Keras with ImageNet weights and excluding the dense layers. Therefore, we can utilize the pre-trained convolutional layers of VGG16 for feature extraction while excluding the classification part.

In tackling the problem at hand, the Unet architecture was selected over VGG because the Unet architecture, characterized by its encoder-decoder design with skip connections, is adept at preserving spatial information during the downsampling process, crucial for maintaining the fidelity of segmented structures. This characteristic is especially advantageous in scenarios where precise localization is paramount, such as in medical imaging or object detection.

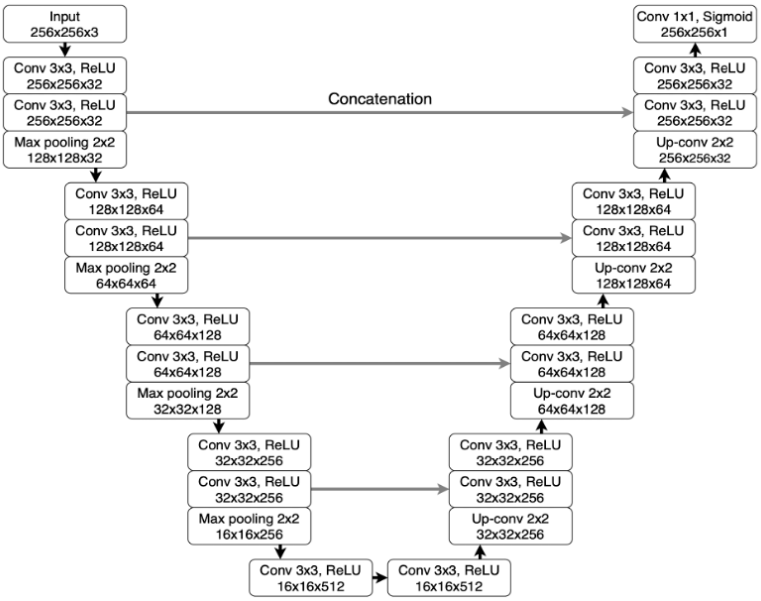


Figure 1: UNet Architecture

Now for statistical analysis we experimented with ANOVA and Kruskal-Wallis, because the choice of ANOVA and Kruskal-Wallis tests.

ANOVA is selected for its suitability in assessing variance across multiple groups simultaneously. In our case, where we are exploring the impact of different factors on imaging features, ANOVA allows us to discern whether there are statistically significant differences among various groups.

The Kruskal-Wallis test is a non-parametric alternative to ANOVA, chosen to accommodate scenarios where assumptions of normality might be violated. This robust statistical method allows us to evaluate group differences without relying on the assumption of a normal distribution.

3 Plan and Experiment

3.1 Dataset

The primary dataset for our research project is sourced from The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA). This dataset comprises preoperative imaging and genomic data from 110 patients diagnosed with lower-grade gliomas. The imaging data includes fluid-attenuated inversion recovery (FLAIR) sequences, and the genomic data involves various molecular classifications such as IDH mutation, 1p/19q co-deletion, RNASeq clusters, DNA methylation clusters, DNA copy number clusters, microRNA expression clusters, and a cluster of clusters.

Patient Population Data

The data set was collected from The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA). Our research included 110 patients with lower-grade gliomas from five different institutions, each contributing to the TCGA LGG collection.

3.1.1 Imaging Data

Imaging data was obtained from TCIA and included various MRI modalities, with fluid-attenuated inversion recovery (FLAIR) sequences as the primary focus. The dataset comprised patients with all sequences available, as well as those with missing post-contrast or pre-contrast sequences. To maintain the original tumor growth pattern, only preoperative data was analyzed. Manual annotation of FLAIR images was performed by a researcher to create training data for the segmentation algorithm.

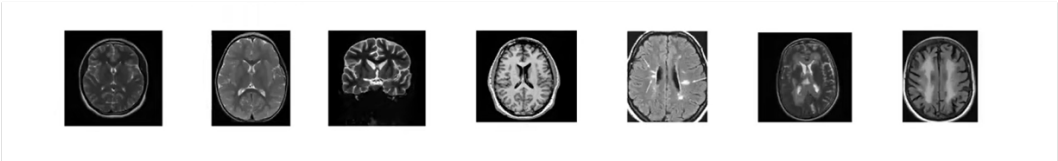


Figure 2: No tumor (2556)

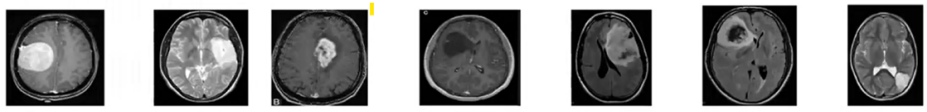


Figure 3: With Tumor (1373)

There is an imbalance among tumor and non tumor images.

3.1.2 Genomic Data

The genomic data is categorized as follows:

1. DNA methylation data (5 clusters)
2. Gene expression data (RNASeq, 4 clusters)
3. DNA copy number data (3 clusters)
4. microRNA expression data (4 clusters)

This high-dimensional genomic data is used to classify tumors into different molecular subtypes and clusters. The RNASeq data is divided into 4 clusters with sizes of 27, 23, 31, and 29 samples, while the methylation data is more evenly distributed, with 5 clusters of sizes ranging from 21 to 23 samples each.

3.2 Hypotheses

Association between Imaging Features and Genomic Subtypes:

Hypothesis: There exists a significant association between automatically extracted imaging features and genomic subtypes of lower-grade gliomas.

Rationale: Previous studies, such as the research conducted by Buda et al., have highlighted robust connections between manually assessed tumor shapes and the distinct molecular classifications present in lower-grade gliomas. Differences in the bounding ellipsoid volume ratio among RNASeq clusters were noted, along with the ability of margin fluctuation to distinguish specific methylation-based groups.

However, manual segmentation has inherent drawbacks—it's time-consuming, prone to variability among different observers, and can be quite labor-intensive. Deep learning presents an alternative by automating and standardizing tumor delineation rapidly, potentially enhancing subsequent analyses linking radiology and genomics.

3.3 Experimental Design

3.3.1 Automatic segmentation

The algorithm to automatically obtain the segmentation mask consists of three steps: first, the image is preprocessed; next, it's segmented. After getting these segmentation masks, we pulled out shape characteristics that could predict molecular types. The next parts will delve into each of these procedures.

3.3.2 Preprocessing

Images differed in size among patients. In the preprocessing phase, the images were:

1. Adjusted to a standard size.
2. The skull was removed to solely analyze the brain.
3. Image brightness and contrast were standardized.
4. Data Augmentation: Oversampling with data augmentation to imbalance between tumor and non-tumor pixels.

3.3.3 Segmentation

The segmentation process utilized a U-Net architecture, a type of neural network. This network has multiple layers, including convolutional layers with specific functions, and uses skip connections to enhance training. Manual segmentations were the reference for training this automatic segmentation tool. To counteract the imbalance between tumor and non-tumor data, we increased the number of tumor images by tweaking their rotation and scale, and removed images without any relevant data after skull stripping. This ensured effective network training.

Data Splitting(Holdout Strategy): We randomly split the dataset into 75:25 ratio for training and testing.

U-Net convolutional neural network architecture is trained on manually annotated tumor masks obtained from the MRI scans to segment tumors automatically. Extensive data augmentation and class re-balancing are employed during training to enhance segmentation performance.

Evaluation Metric We use Dice coefficient as an evaluation metric for the segmentation task which quantifies the similarity between predicted and ground truth masks. Ranging from 0 to 1, where 1 indicates perfect overlap, the Dice coefficient measures the ratio of twice the intersection area to the sum of areas of the predicted and reference masks.

$$Dice\ Coefficient = \frac{2 \times |A \cap B|}{|A| + |B|}$$

where A is the predicted set of pixels and B is the ground truth.

3.3.4 Extraction of Shape Features

Three shape features are extracted from the automatic tumor segmentations:

1. Bounding ellipsoid volume ratio: The ratio of the volume of the tumor to the volume of the bounding ellipsoid.
2. Margin fluctuation: The average distance from the tumor surface to the nearest point on the bounding ellipsoid.
3. Angular standard deviation: The standard deviation of the angles between the tumor surface normal vectors and the major axis of the bounding ellipsoid.

The relationship between these imaging features and molecular subtypes will be analyzed using statistical tests. The code was implemented in Python using Keras and other deep learning libraries.

3.3.5 Statistical Methods

We will employ ANOVA to assess the variance in imaging features across different genomic subtypes. Additionally, the Kruskal-Wallis test will be utilized as a non-parametric alternative to account for potential deviations from normality in the data.

We are essentially testing the following with the two tests:

Null Hypothesis (H0): There is no significant difference in imaging features among different genomic subtypes.

Alternative Hypothesis (H1): At least one genomic subtype significantly differs in imaging features compared to others.

We will use a significance level of 0.05 to determine statistical significance.

We will interpret the results by examining effect sizes, and visual representations, providing a comprehensive understanding of the associations between imaging features and genomic subtypes.

4 Results

We employed a U-Net convolutional neural network architecture to perform automated segmentation of lower-grade gliomas within brain MRI scans. Preprocessing steps, including rescaling, and contrast normalization, ensured standardized and focused analysis on brain regions. The model's performance was evaluated on a test set, achieving a mean Dice similarity coefficient of 0.9466, signifying substantial agreement between the algorithmic and manual segmentations. Our baseline model(VGG) achieved a Dice similarity coefficient of 0.86.

This precise delineation enabled the extraction of three vital shape features from the segmented tumor regions:

- Bounding ellipsoid volume ratio
- Margin fluctuation
- Angular standard deviation

The angular standard deviation marks irregularities in tumor boundaries, while the volume ratio measures tumor convexity, and margin fluctuation quantifies surface smoothness.

In our exploration beyond segmentation, our analysis aimed to uncover correlations between these shape features and the diverse genomic subtypes present in lower-grade gliomas. Employing ANOVA and Kruskal-Wallis tests across a cohort of 110 patients, we drew robust statistical inferences.

The comparison against RNASeq clusters unveiled a notable discovery:

- The RNASeq-R2 subtype exhibited significantly higher irregularity in tumor shape, as shown by BEVR (p-value = 0.0000642 via ANOVA), suggesting distinct geometric variations linked to different RNASeq subtypes.

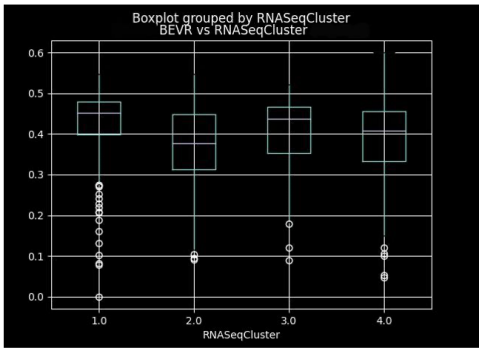
Associations	p-value for ANOVA	p-value for Kruskal-Wallis
BEVR - RNASeqCluster	0.0000642	0.00001
MF - RNASeqCluster	0.073	0.018
ASD - MethylationCluster	0.044	0.019

Figure 4: Results of p-values for two different statistical test

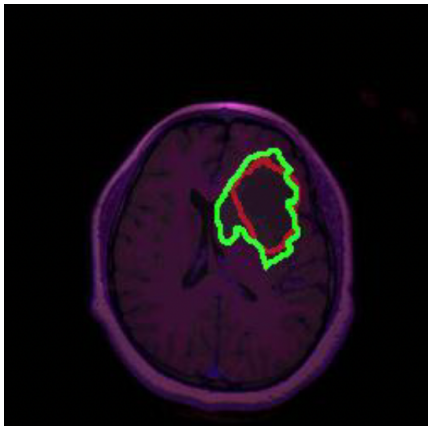
- Further investigation into the relationship of margin fluctuation with RNASeq clusters highlighted a discernible difference in boundary smoothness across genomic subtypes, indicated by a p-value of 0.018. This is illustrated in Figure 4.

Additionally, exploring angular standard deviation within DNA methylation clusters revealed substantial variations in margin patterns among subtypes (p-value = 0.019, Kruskal-Wallis test).

The associations of features are visually represented in Figure 5 (a), which demonstrates the Box plot for association between BEVR and RNASeq Cluster.



(a) BEVR vs RNASeq



(b) Ground Truth vs Automatic Segmentation Output

Figure 5

Furthermore, Figure 5 (b) displays the actual (green) and predicted (red) outlines of tumors in brain sections, providing a visual comparison between the algorithmic predictions and the real outlines of the tumors.

The combined findings underscore the significance of both the segmentation methodology's accuracy and the extracted shape features in characterizing lower-grade gliomas. These quantitative descriptors not only capture the complex geometries but also reveal nuanced differences among genomic subtypes. Understanding these associations could pave the way for tailored treatments and prognostic insights, potentially enabling more targeted and effective interventions for patients diagnosed with lower-grade gliomas.

5 Conclusion

In the conclusion of our project, the following points are highlighted:

Demonstration of MRI Features Association with Tumor Subtypes: The project successfully demonstrated that features extracted from MRI scans using deep learning algorithms were associated with tumor molecular subtypes in lower-grade gliomas. This was determined using genomic assays.

Potential for Non-Invasive Imaging-Based Surrogates: The findings of this project indicate the potential for developing reproducible, non-invasive imaging-based surrogates of tumor genomics in brain cancer. This could be a significant advancement in the field, providing a less invasive method for tumor analysis and characterization.

Insights About the Problem and Approaches: This project has provided valuable insights into the application of deep learning algorithms for analyzing MRI features. One key lesson learned is the effectiveness of these algorithms in associating MRI features with tumor molecular subtypes in lower-grade gliomas. This demonstrates a significant potential in the field of medical imaging and cancer diagnosis.

Experiments Conducted: The experiments conducted have shown the promise of using non-invasive imaging-based methods as surrogates for more invasive genomic assays. This approach could revolutionize the process of tumor genomics analysis in brain cancer, making it more accessible and less burdensome for patients.

Ideas for Future Exploration: Although not fully explored in this project due to time constraints, there is a suggestion for future exploration in enhancing the accuracy and specificity of the deep learning models used. Further research could involve integrating additional data sources or employing more advanced algorithmic techniques to refine the predictive capabilities of these models.

6 References

- M. Buda, A. Saha, M. A. Mazurowski (2019). Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm.
- Olaf R., Philipp F., Thomas B. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation.
- https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation/data
- Jiachi Zhang, Xiaolei Shen, Tianqi Zhuo, and Hong Zhou. Brain tumor segmentation based on refined fully convolutional neural networks with a hierarchical dice loss. arXiv preprint arXiv:1712.09093, 2017.
- Lucas Fidon, Wenqi Li, Luis C Garcia-Peraza-Herrera, Jinendra Ekanayake, Neil Kitchen,Sebastien Ourselin, and Tom Vercauteren. Generalised wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks. In International MICCAI Brainlesion Workshop, pages 6476. Springer, 2017.
- Simon Jegou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 1119, 2017.