Final Report for Syslab

GlioGrade: Using 3D Convolutional Neural Networks to Type and Grade Gliomas within Human MRI Scans

Anurag Perakalapudi

Date: 5/27/25

Yilmaz - Period 1

Table of Contents

Abstract

Gliomas are the most common malignant brain tumors, making up the vast majority of deaths related to brain tumors. Accurate diagnosis requires identifying both the glioma type and grade. Most machine learning models use 2D Convolutional Neural Networks (CNNs), which lack the spatial awareness needed to process full 3D MRI volumes. This project proposes "GlioGrade," a novel, locally hosted tool that uses 3D CNNs to classify glioma types and grades using the UCSF-PDGM dataset. Our models achieved 84.57% and 83.84% accuracy for typing and grading, respectively. The project integrated preprocessing techniques like skull stripping and N4 bias correction and evaluated multiple CNN architectures, including geodesic and pretrained models. The system was deployed via a user-friendly web interface for clinical consideration.

I. Introduction

Gliomas are the most common form of cancerous tumor with 6 cases per 100,000 people. There are 4 major types of glioma: Astrocytoma (IDH-Wildtype), Astrocytoma (IDH-Mutant), Oligodendroglioma, and Glioblastoma. The types are determined by genetic and histological characteristics. There are three different grades of glioma: 2, 3, and 4, which describe the severity of the tumor. These two metrics help doctors diagnose and treat gliomas accurately.

There have been previous attempts at grading and typing gliomas, however these have only involved 2D images which doesn't take the whole brain into consideration as a doctor would. Our tool would be the first to type and grade using 3D Convolutional Neural Networks (CNN) Models. The World Health Organization (WHO) reclassified the glioma identifications in 2021 and our tool is the first tool to be trained for the new classification.

II. Background

Previous models often relied on manual MRI analysis or 2D CNNs. Gutta et al. (2021) demonstrated that CNNs outperform traditional methods for glioma grading, which further motivated our 3D approach. These methods are limited by human error or lack of 3D spatial context. Many models also used outdated classification systems prior to the 2021 WHO revision, which redefined glioma categories and diagnostic guidelines. To remain clinically relevant, this tool was developed and trained using the updated WHO 2021 glioma classification.

For this project, the UCSF-PDGM dataset (495 patients) was used for training. It includes skull-stripped and bias-corrected MRI scans preprocessed for machine learning applications. The Erasmus Glioma Dataset (774 patients) was used for validation and contained raw MRI scans, offering a contrast in preprocessing protocols and challenging the model's generalization capability. The UCSF-PDGM dataset was used for training and includes 495 preprocessed patient scans classified under WHO 2021 (Calabrese et al., 2022; TCIA, n.d.). For testing, we used the Erasmus Glioma Database containing raw MRI scans (Voort et al., 2021).

Preprocessing included standard methods recommended in MRI-based brain tumor classification literature:

1. Skull stripping, which removes the bright skull tissue surrounding the brain to reduce size and focus the model on relevant data. We applied skull stripping using a method described in Analytics Vidhya's tutorial (2021).
2. N4 bias correction, which mitigates low-frequency intensity nonuniformities often seen in MRI scans due to magnetic field inhomogeneities. We applied N4ITK bias field correction following techniques explained by Ramr777 (2021).
3. Segmentation (optional), where models such as MONAI were explored to isolate enhancing tumor, tumor core, and whole tumor masks, though these did not significantly improve model performance.

III. Applications:

This tool has direct applications in clinical decision-making. It can be deployed in hospitals to help radiologists quickly determine glioma type and grade non-invasively from MRI scans. This reduces reliance on biopsy and manual labeling, which are time-consuming and subjective.

Furthermore, the locally hosted web interface ensures that all medical images remain on-site, protecting patient privacy while allowing for powerful machine learning diagnosis. This is especially relevant in healthcare settings where HIPAA compliance is an issue.

IV. Methods:

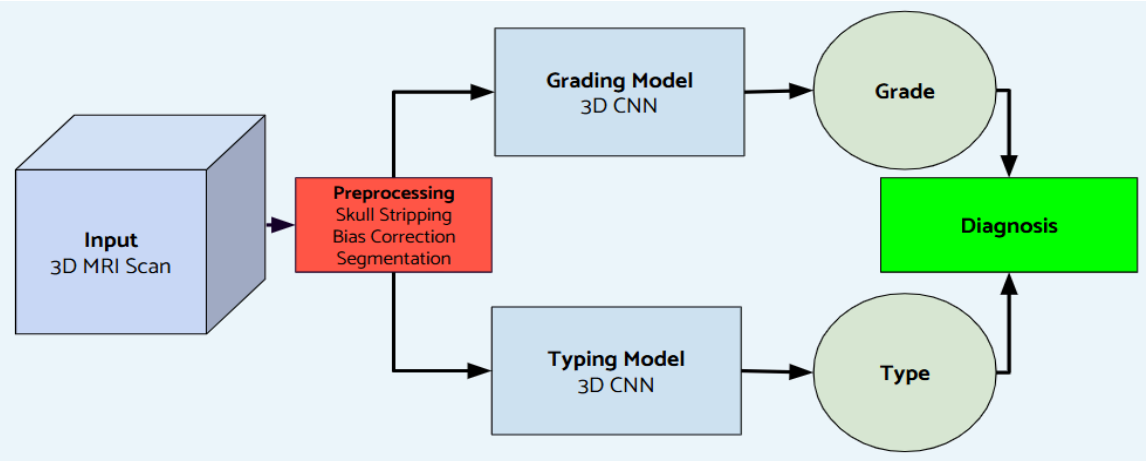

Figure 1: Systems Architecture Diagram

Systems Diagram Overview:

Figure 1 illustrates the end-to-end process in the GlioGrade system. The 3D MRI scan undergoes preprocessing steps (bias correction and skull stripping) before being passed into two sequential CNN models: one for typing and one for grading gliomas. The final result is displayed to the user via a user-friendly local interface.

Input/Output:
MRI inputs were processed in NIfTI format after skull stripping and bias correction. The outputs were two categorical labels: glioma type (Astrocytoma IDH-Mutant/Wildtype, Glioblastoma, Oligodendroglioma) and glioma grade (2, 3, 4). The model was designed to handle multi-class classification using softmax outputs and categorical cross entropy loss.

Architecure:

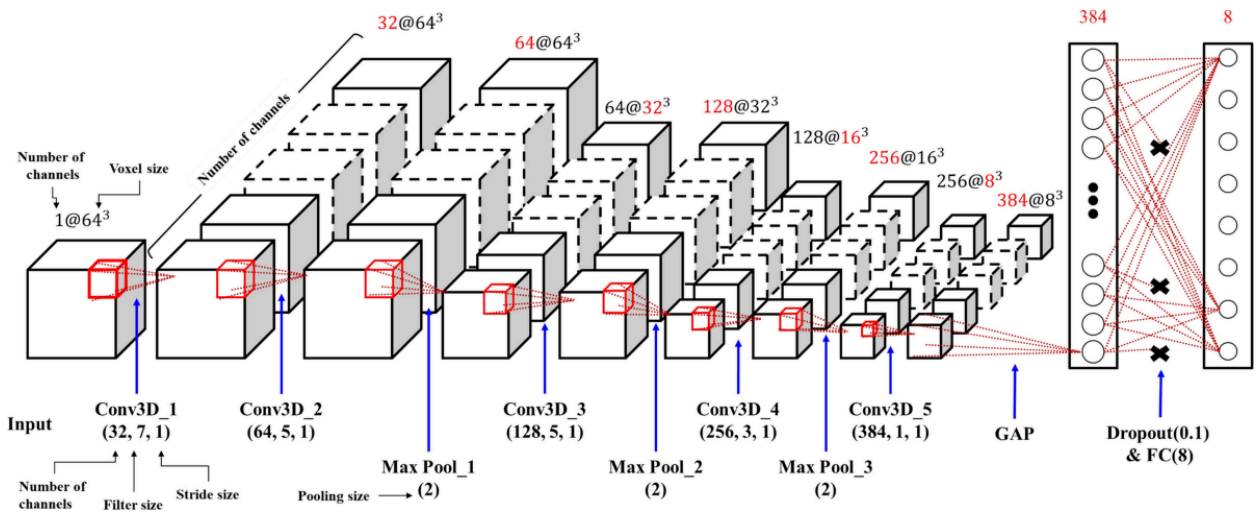

Figure 2: 3D CNN Architecture (Inspired by VGG-16):

Figure 2 shows the architecture of a deep 3D CNN adapted from a model published in Scientific Reports by Kim, J., et al. (2022). It includes multiple convolutional layers with increasing channel depths and decreasing spatial dimensions, illustrating the network's capability to extract increasingly abstract features from volumetric brain MRI data. The model utilizes various kernel sizes, pooling strategies, and a final fully connected layer with dropout to prevent overfitting. This structure inspired our own CNN design for glioma classification.

We modeled parts of our architecture on designs used in deep multi-scale 3D CNNs, as shown in Mzoughi et al. (2020). The decision to use a VGG-inspired model was motivated by its architectural simplicity and prior success on similar medical imaging tasks. It provided a good balance between complexity and computational efficiency. While more complex alternatives such as geodesic convolutions and segmentation-based isolation were explored, they proved either too unstable or ineffective within our training constraints.

The architecture follows a series of Conv3D and MaxPool3D layers:

- Input: (1, 64, 64, 64)
- Conv3D layers: (1 → 32), (32 → 64), (64 → 128), (128 → 256)
- Every convolution is followed by BatchNorm3D and ReLU activation
- MaxPooling reduces spatial dimensions by half
- Fully connected layers: 16,384 → 512 → Softmax Output

The input for our models consisted of 3D T2-weighted FLAIR MRI scans stored in NIfTI format, which were preprocessed using skull stripping and N4 bias correction. The goal was to classify each scan into one of four glioma types and subsequently grade gliomas into one of three severity levels if applicable.

Each MRI scan was normalized and passed through a custom-built 3D CNN model. This model applied a series of convolutional layers to extract spatial features, followed by pooling operations to reduce dimensionality. These features were flattened and fed into fully connected layers for classification. The final output was generated using a softmax activation layer.

Training:

Models were trained on Google Colab using the UCSF dataset. A 70/20/10 split ensured that training, validation, and test sets were distinct. Loss and accuracy metrics were monitored using TensorBoard. Data was balanced using weighted loss functions to offset class imbalances in oligodendroglioma and Grade 2 cases.

Model evaluation was based on categorical accuracy and confusion matrices. Internal metrics such as F1 scores per class were also tracked during training. The final architecture featured four 3D convolutional blocks, each followed by Batch Normalization and Max Pooling layers, and two fully connected dense layers. Input volumes were shaped as (1, 64, 64, 64), with output dimensions of either 4 (for glioma types) or 3 (for grades).

Other model architectures and preprocessing configurations were explored:

1. A Geodesic CNN that incorporates surface-based learning did not outperform the standard model.
2. Using only segmented tumor regions (ET/TC/WT) reduced model robustness to real-world scans.
3. MedicalNet, a pretrained 3D CNN, was difficult to adapt to our dataset structure.

V. Results:

The final models performed well in classification tasks. The typing model achieved an accuracy of 84.57% in distinguishing between glioma subtypes including astrocytoma (IDH-mutant/wildtype), glioblastoma, and oligodendroglioma. The grading model achieved an accuracy of 83.84% in classifying tumor grades 2, 3, and 4. These models were trained using the UCSF-PDGM dataset, evaluated on a held-out validation set, then externally verified on the Erasmus dataset to assess robustness across various preprocessing and acquisition standards.

The confusion matrices showed that the bulk of errors occurred between adjacent classes. For example, due to overlapping imaging characteristics, the model occasionally misclassified

glioblastoma as astrocytoma IDH-wildtype, and Grade 3 gliomas as Grade 4. These misclassifications are consistent with issues experienced in real-world clinical settings, supporting the model's performance realism. While not perfect, these models strike a reasonable compromise between accuracy, efficiency, and generalizability, making them valuable clinical decision-support tools.

In addition to the classification models, we developed a locally hosted website that allows users to upload their MRI scans and view predictions in real time. This interface was designed for ease of use by clinicians and researchers and supports DICOM and NIfTI formats. The tool includes dynamic image slicing and class probability visualization, offering interpretability and accessibility. Its deployment ensures patient data privacy and compatibility with hospital systems, reinforcing the practical clinical value of the overall solution.

VI. Limitations:

There were several limitations that affected model performance and generalizability. One major challenge was the class imbalance present in the UCSF dataset, particularly the underrepresentation of oligodendroglioma cases. This made it difficult for the model to learn adequate representations for this subtype. Another limitation was the mismatch in preprocessing steps and imaging protocols between the UCSF training dataset and the Erasmus test dataset.

This difference occasionally hindered generalization and highlights the importance of preprocessing standardization. Additionally, the 3D CNN architecture, while effective, demanded substantial computational resources, limiting the number of experiments and hyperparameter tuning we could perform. Lastly, efforts to incorporate segmentation or pretrained architectures such as MedicalNet and geodesic CNNs did not yield improved results and sometimes degraded performance, pointing to the complexity of domain-specific model generalization.

VII. Conclusion:

We successfully built and validated GlioGrade, a novel tool using 3D CNNs to type and grade gliomas. It uses cutting-edge WHO 2021 standards and provides promising accuracy. The tool has potential as a clinical assistant for MRI-based diagnosis.

VIII. Future Work:

If given the opportunity to begin again, we would have emphasized preprocessing harmonization across datasets and applied stratified K-fold validation to better capture variability across glioma subtypes. With additional time, we would have prioritized collecting more samples for underrepresented classes such as oligodendrogliomas. We would have also adopted cross-dataset

validation techniques and explored advanced regularization strategies like dropout scheduling and learning rate warm-ups.

IX. Materials:

The materials used for this project included the UCSF-PDGM and Erasmus Glioma datasets (see References). Tools and packages utilized included Python 3.8, TensorFlow/Keras for model development, NiBabel for reading .nii.gz files, MONAI for segmentation experimentation, and visualization libraries such as matplotlib and seaborn. All code was developed and run using Google Colab as the computing platform.

References:

1. Kim, J., et al. (2022). A deep 3D convolutional neural network architecture with a fully connected layer for classification of glioma grades. Scientific Reports, 12, 15294. https://doi.org/10.1038/s41598-022-19212-6

2. Gutta, S., Acharya, J., Shiroishi, M.S., Hwang, D., Nayak, K.S. (2021). Improved Glioma Grading Using Deep Convolutional Neural Networks. AJNR Am J Neuroradiol, 42(2), 233–239. https://doi.org/10.3174/ajnr.A6882

3. Mzoughi, H., et al. (2020). Deep Multi-Scale 3D Convolutional Neural Network (CNN) for MRI Gliomas Brain Tumor Classification. J Digit Imaging, 33(4), 903–915. https://doi.org/10.1007/s10278-020-00335-4

4. UCSF Dataset: https://www.cancerimagingarchive.net/collection/ucsf-pdgm/

5. Erasmus Dataset: https://www.sciencedirect.com/science/article/pii/S2352340921004753

6. Bias correction: https://medium.com/@alexandro.ramr777/how-to-do-bias-field-correction-with-python-156b9d51dd79

7. Skull stripping: https://www.analyticsvidhya.com/blog/2021/06/introduction-to-skull-stripping-image-segmentation-on-3d-mri-images/

8. Demo video: https://www.youtube.com/watch?v=J6PYP13rp0o

APPENDIX:

I. **CODE:**
https://github.com/2025aperakal/SyslabFinal