A multimodal 3D deep learning framework for glioma segmentation, classification and precision therapy

Study objective

The objective of this study was to design and evaluate a biologically inspired 3D deep learning architecture that integrates spatial attention and multi-branch convolutional pathways to enhance glioma segmentation, tumor type and grade classification, molecular marker prediction, and survival estimation. In the final step, the outputs of the model are used to generate automated radiology reports, which are further processed by a retrieval-augmented generation (RAG) system to suggest individualized treatment plans.

Rationale

Management of glioma patients faces several clinical challenges:

- Molecular testing and tumor classification require surgical biopsy, which carries procedural risks and potential delays.
- Biopsies from multiple tumor sites are often needed due to inter- and intra-tumoral heterogeneity.
- Manual delineation and interpretation of imaging are time-consuming, costly, and prone to inter-radiologist variability.

To address these challenges, our approach emphasizes boundary-sensitive learning and multimodal feature fusion while maintaining computational efficiency suitable for large-scale clinical use. We hypothesize that incorporating biologically inspired mechanisms and attention-driven modules will enable the model to generate clinically relevant, interpretable predictions, bridging the gap between radiological imaging and precision oncology.

Materials and methods

Datasets

For pretraining the model, we used BraTS 2025 data that includes 2873 adult glioma patients and for training and testing the model we used UCSF-PDGM dataset (495 adult glioma patients) .The 10% and 20% of UCSF-PDGM dataset were used as external validation and internal validation data respectivily . Since gliomas are heterogeneous, no single sequence captures all tumor components we choose two sequence for each patient including T1-contrast (T1c) which highlights regions with a disrupted blood–brain barrier, showing the active, enhancing tumor core and T2-FLAIR (T2f) which suppress CSF and makes edema and infiltrative tumor spread visible . The original voxel spacing and sizes were (1 mm, 182×218×182) for BraTS, and (1 mm, 240×240×155) for UCSF-PDGM . Tumor masks contained three classes aside from background (class 0): enhancing tumor (class 3), non-enhancing/necrotic tumor (class

1), edema (class 2) . Tabular patient data included sex, age, tumor type, tumor grade, IDH and MGMT status, and survival information.

Preprocessing for 3D Segmentation

For preprocessing of the data of 3D segmentation model first we exclude patient that have corrupted or missing data of images or masks. Then for unifying images and mask sizes across all patients we use cropping and zero padding to covert size of images to (224,224,160) then use voxel spacing of 2mm to resize them to (112,112,80). To maintain the original structure of images and masks during the resizing, we use interpolation of bilinear and nearest neighbor for images and masks respectively. Then we stack the T2f and T1c images of each patient to create a image with shape of (112,112,80 ,2). The mask classes converted to one hot encoded classes and masks get shape of (112,112,80,4).

Preprocessing for Classification

For preprocessing of the data of classification model first we exclude patient that have corrupted or missing tabular data or images . The same preprocessing step like 3D segmentation model were applied to images and we create images with shape of (112,112,80,2) . In addition to images the clinical data including age and sex used as inputs of the classification model . The labels of classification model includes tumor type , tumor grade , IDH and MGMT status and overall survival The Age and overall survival underwent z-score normalization and min–max normalization respectively .In our dataset, 16% of patients were missing MGMT status labels. To address this issue, we employed data imputation using CatBoost, which leverages the available clinical and demographic features to predict the most likely MGMT status for these patients. Next, tumor type, tumor grade, IDH status, and MGMT status were one-hot encoded before being fed into the model. For both segmentation and classification tasks, 20% of the dataset was held out as an internal validation set. Validation data underwent the same preprocessing steps as the training set but were never used for model weight updates, serving solely to monitor model performance and avoid overfitting.

Data augmentation

To improve model generalization, address limited data diversity and class imbalance, we applied on-the-fly 3D data augmentation during training. For the segmentation model, augmentations included random flips along the sagittal , coronal and transverse planes and Random Gaussian noise . For the classification model, the same image augmentations were applied to the MRI inputs while tabular clinical features remained unchanged. Augmentation was applied probabilistically per training sample, ensuring that each epoch saw a diverse set of variations without altering the ground truth labels.

3D Segmentation Model Architecture

We developed a biologically inspired 3D convolutional neural network for volumetric glioma segmentation, implemented in TensorFlow v2.18 using the Keras API. The model receives multimodal MRI input of shape

88×104×88×2 and employs parallel encoder branches inspired by the human visual pathway. Two Sensory Neuron branches process the input with 3D convolutions: the nasal branch uses 32 and 64 filters with a 3×3×3 kernel, while the temporal branch uses 32 and 64 filters with a 5×5×5 kernel, both incorporating dilation rates of (1,1,1). Each branch includes residual Nucleus blocks with multi-scale convolutions and batch normalization, followed by a Spatial Attention mechanism to emphasize salient tumor regions.

Outputs from the nasal and temporal branches are concatenated in a crossover module and processed through additional Sensory Neuron blocks with 128 filters and multi-scale dilations. The resulting feature maps are then fed into three parallel Lateral Geniculate Body blocks with 240 filters and kernel sizes 1×1×1, 3×3×3, and 5×5×5, each followed by residual Nucleus blocks and 2×2×2 max pooling for hierarchical feature extraction. The outputs of these blocks are concatenated into the optic radiation feature map.

The decoder employs four sequential Motor Neuron blocks with transposed convolutions for upsampling (filter sizes: 240, 128, 64), skip connections from corresponding encoder outputs, and batch normalization. The final stages include additional 3D convolutions with 64 and 32 filters, producing a volumetric feature map that is passed through a 1×1×1 convolution with softmax activation to generate a four-class segmentation output.

This architecture combines multi-scale convolutions, dilations, residual connections, attention mechanisms, and biologically inspired branching to capture heterogeneous tumor structures and peritumoral regions effectively, providing accurate volumetric segmentation from multimodal MRI data.

Classification Model Architecture

The classification model uses encoder part of segmentation model plus a tabular input head with shape of (2,) .The high level features of encoder bottleneck after global max pooling and global average pooling is concatenated with the clinical data features . Then the features feds to three dense layers(relu activation function) with 512 , 256 , 128 neurons respectively . Finaly the classification model have five output heads including tumor type , tumor grade , IDH status , MGMT status , overall survival .

Model Training

The model was trained on a NVIDIA P100 GPU setup with 16 GB VRAM, 30 GB of system RAM and 300 GB HDD, using a batch size of 2 for segmentation model and 3 for classification model to balance GPU memory constraints and stable gradient updates. A two-stage training protocol was adopted, consisting of 200 epochs of pretraining followed by 160 epochs for segmentation model training and 80 epochs for classification model training. The Adam optimizer with a learning rate of ranging from 1×10^{-4} to 1×10^{-5} was employed to ensure efficient convergence. For segmentation task a custom combined loss function, integrating weighted Dice loss and categorical focal cross-entropy, was designed to enhance boundary sensitivity and address class imbalance. Model performance was monitored using Intersection-over-Union (IoU) across the non-background classes as the primary evaluation metric. For classification model the Huber loss and a mean absolute error (MAE) was used as loss and metric respectively for overall survival task and for the remaining tasks, the weighted categorical focal cross-entropy were used to address severe class imbalance and accuracy used as metric.

Interpretability

To enhance model transparency and support clinical decision-making, we applied interpretability methods for both the segmentation and classification models. For the segmentation model, Gradient-weighted Class Activation Mapping (Grad-CAM) was used to visualize salient regions influencing voxel-wise predictions, highlighting tumor subregions detected by the network. For the classification model, SHapley Additive exPlanations (SHAP) were employed to quantify the contribution of each clinical feature and imaging-derived representation to the predicted tumor type, grade, and molecular markers. These techniques provided insight into which image regions and patient features drove model outputs, allowing evaluation of model reasoning and identification of potential biases.

Radiology Report Generation

After training of segmentation and classification models , their outputs is utilized to generate a radiology report .Frist the predicted mask is given to our inbuilt python library that first registers brain atlas according to patient MRI image and then according to registered atlas and predicted mask, gives us the size of each tumor subregion along with their locations they involving in brain. Then this data integrate with classification model outputs to generated a radiology report .

Retrieval-Augmented Generation (RAG) for Treatment Planning

At final stage we designed a classic RAG system that suggest us a treatment plan according to generated radiology report. RAG combines a retrieval system with a large language model (LLM) to improve responses. It first retrieves relevant documents or passages from a large knowledge base based on the input query. The language model then conditions its generation on the retrieved content to produce more accurate and informative answers. We used OpenAI API to implement GPT-4o-mini as our LLM for RAG , all-MiniLM-L6-v2 as embedding model and Chroma DB as vector store . We used latest guidelines in glioma management as relevant documents and then chunk them into subdocuments with 1500 characters with overlap of 200 characters . Then our subdocuments save as embedding in our vector store . The radiology report is utilized as RAG query and relevant document are retrieved based on the similarity score threshold (score threshold: 0.5). In the final stage the LLM give us specific personalized treatment plan according to our prompt .

Results

3D segmentation model

The proposed model was evaluated on a dataset of 495 glioma patients, including T1c and T2f MRI sequences, that underwent automated segmentation using an ensemble model consisting of prior BraTS challenge winning segmentation algorithms. Images were then manually corrected by trained radiologists and approved by 2 expert reviewers. Data were split into training, validation, and test sets in an 70:20:10 ratio.

During **training**, the model achieved a mean **IoU of 0.8066** with a corresponding loss of **0.3498**, while validation performance yielded an IoU of **0.7298** and loss of **0.4059**.

On the independent **test set**, the model achieved a **mean IoU of 0.7142** with an **overall loss of 0.6448**. Subregion-specific IoU scores demonstrated consistent performance across tumor compartments, with **0.6819** for non-enhancing/necrotic tumor, **0.7481** for peritumoral edema, and **0.7126** for enhancing tumor regions.

Classification model

The proposed model was evaluated on a dataset of 495 glioma patients, including T1c and T2f MRI sequences plus tabular data . Data were split into training, validation, and test sets in an 70:20:10 ratio.

On the **training set**, the model achieved high predictive performance across outputs. IDH mutation status was predicted with an accuracy of **0.9463** and a loss of **0.0186**, while MGMT promoter methylation reached an accuracy of **0.7573** with a loss of **0.0371**. Tumor grade classification yielded an accuracy of **0.9272** (loss = **0.0295**), and tumor type classification reached **0.9347** (loss = **0.0278**). For survival prediction, the model achieved a MAE of **0.0997** with a corresponding survival loss of **0.0106**. The overall training loss was **0.2200**.

On the **validation set**, performance was slightly lower, as expected. IDH prediction achieved **0.8090** accuracy (loss = **0.0925**), and MGMT classification reached **0.6742** accuracy (loss = **0.0699**). Tumor grade and tumor type prediction achieved accuracies of **0.7640** and **0.7753**, with corresponding losses of **0.0920** and **0.0653**, respectively. The survival prediction task resulted in an MAE of **0.1719** (loss = **0.0171**), while the total validation loss was **0.5947**.

On the **test set**, the model achieved an **IDH mutation prediction accuracy of 0.8722** with a loss of **0.0562**, and **MGMT methylation accuracy of 0.7675** (loss = **0.0854**). Tumor grade classification reached **0.8400 accuracy** (loss = **0.0385**), while tumor type classification achieved the highest accuracy of **0.8926** (loss = **0.0438**). For survival prediction, the model obtained a MAE of **0.2792** with a loss of **0.0706**. The total test loss was **0.4567**.