Explainable Radiogenomics: Non-Invasive Prediction of IDH Mutation in Gliomas on Low-Memory GPUs

Researchers:

Sayna Sarvar, Arman Sajjadi

Supervisor:

Dr. Tara Nezhad (Data scientist at Acosta)

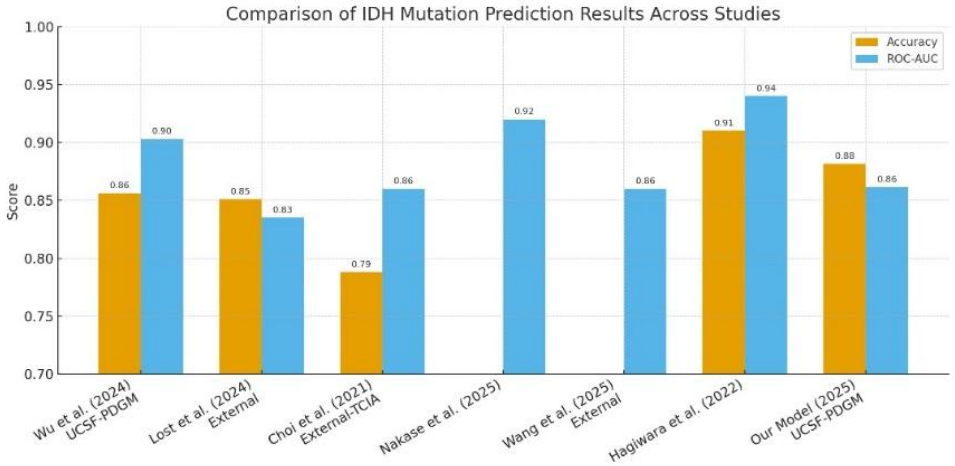
Abstract

The Isocitrate Dehydrogenase (IDH) mutation status is a critical biomarker for the prognosis and treatment planning of gliomas. In this study we developed a 3D deep learning model using the public **UCSF-PDGM dataset**. For each of the 501 patient, four pre-operative MRI sequences **(FLAIR, T1, T1c, and T2)** were used as input. The core of our pipeline is a **3D EfficientNet-B7** architecture, chosen for its high performance and computational efficiency for working on Low-Memory GPUs. The model was trained using extensive data augmentation and regularization techniques to ensure robustness. For model interpretability, we employed **Gradient-weighted Class Activation Mapping (Grad-CAM)** to generate heatmaps that identify the most salient predictive regions within the brain tumor. This makes the AI's predictions easier for doctors to trust.

On a held-out test set, our model achieved excellent performance, reaching an **F1-Score of 0.88**, an accuracy of 88.2%, a precision of 0.88, a recall of 0.88, and an ROC-AUC of 0.86. The Grad-CAM visualizations successfully highlighted specific intra-tumoral areas that were highly correlated with the model's predictions.

1. Related Work

Figure below compares the performance of our model against prior studies on IDH mutation prediction, demonstrating that our model achieves accuracy and ROC-AUC scores competitive with state-of-the-art methods reported in the literature.

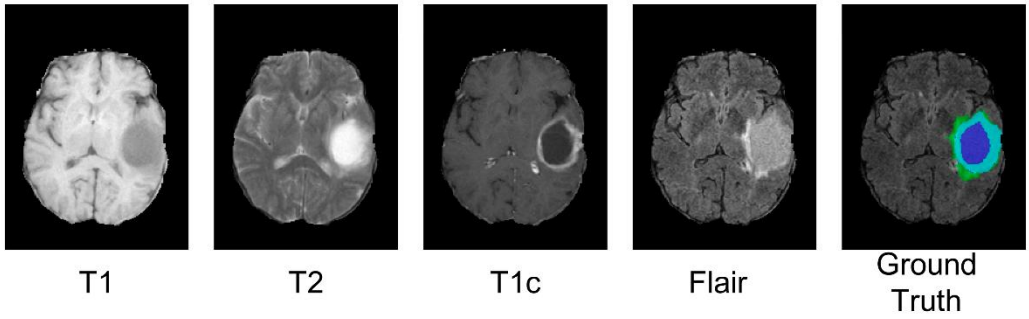


2. Materials and Methods

2.1 Dataset and Preprocessing

This study utilized the UCSF-PDGM dataset, which contains multi-parametric pre-operative MRI scans with confirmed IDH mutation status for 501 patients. For each patient, four MRI sequences (T1, T1c, T2, and FLAIR) were employed to capture complementary tumor characteristics. The dataset also includes ground truth tumor segmentations, which were used in this study exclusively for validating the model's focus in the final explainability analysis.

Prior to model training, all images were resampled to a uniform size of 128×128×128 voxels and normalized. For model development, the dataset was partitioned into training (70%), validation (15%), and testing (15%) sets.



2.2 Model Architecture and Fine-Tuning Strategy

For this study, we employed a **3D EfficientNet-B7** architecture, leveraging transfer learning by initializing the network with weights pre-trained on ImageNet. The model was specifically adapted for our multi-modal task: the input layer was reconfigured to accept four channels corresponding to the stacked MRI sequences (T1, T1c, T2, and FLAIR), and the final classifier layer was replaced to output two classes for the binary prediction of IDH mutation status.

A key component of our methodology was a resource-efficient fine-tuning strategy designed for low-memory GPUs. To achieve this, the majority of the pre-trained network's layers were frozen, making them non-trainable. Only the parameters in the final two convolutional blocks and the new classifier head were unfrozen, allowing them to be specifically adapted to the radiogenomic features of gliomas. This approach dramatically reduces the computational and memory footprint while preserving the rich feature hierarchy learned from the original dataset.

2.3 Training Procedure

The model was trained with a robust procedure designed to handle the specific challenges of the dataset and to promote generalization.

Class Imbalance Mitigation The dataset exhibited a significant class imbalance, with a majority of wildtype samples (Class 0: 398) compared to mutant samples (Class 1: 103). A two-pronged strategy was implemented to counteract this:

1. **Data-Level Sampling:** A WeightedRandomSampler was integrated into the training data loader. This technique oversamples the minority class (mutant) to ensure that each training batch contains a balanced distribution of both classes, preventing the model from developing a bias towards the majority class.

2. **Loss Function:** Focal Loss was selected as the objective function. Unlike standard cross-entropy, Focal Loss dynamically reduces the contribution of easy-to-classify examples (typically the majority class), compelling the model to focus more of its learning effort on challenging examples from the minority class.

Optimization and Regularization The model's weights were updated using the **AdamW** optimizer. The learning rate was dynamically managed by a **OneCycleLR** scheduler, which is known to improve training stability and speed up convergence. To prevent overfitting, a comprehensive set of 3D-specific data augmentations were applied to the training data at runtime.

2.4. Evaluation Metrics

The model's final performance was evaluated on the independent test set using a suite of metrics: **Accuracy, Precision, Recall, F1-Score, and ROC-AUC**. Due to the dataset's class imbalance, the F1-Score was considered a primary metric for its robustness. A confusion matrix was also generated to provide a detailed per-class visualization of the results.

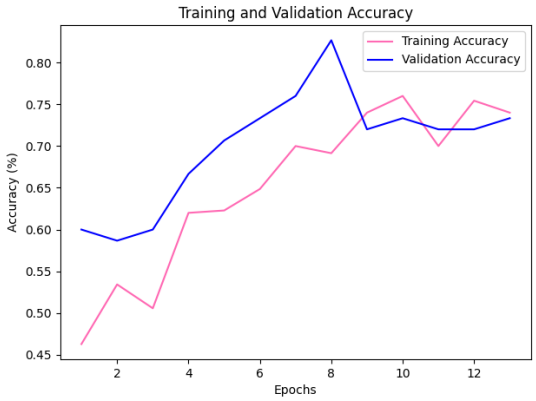
3. Results

3.1. Quantitative Performance

The trained 3D EfficientNet-B7 model was evaluated on the held-out test set, which comprised 15% of the UCSF-PDGM dataset. The comprehensive evaluation metrics are summarized in the table below.

Table: Performance Metrics on the Test Set

Metric	score
F1-score	0.87
Accuracy	0.88
Precision	0.87
Recal	0.88
ROC-AUC	0.86



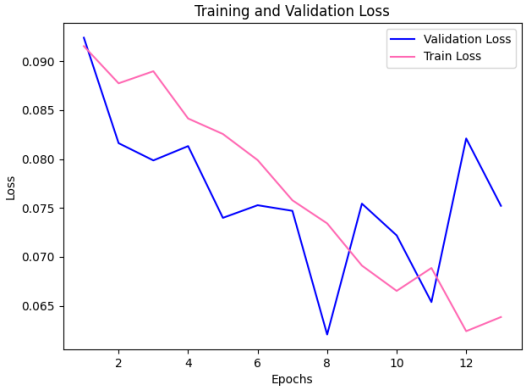


Figure: Model Training and Validation Curves.

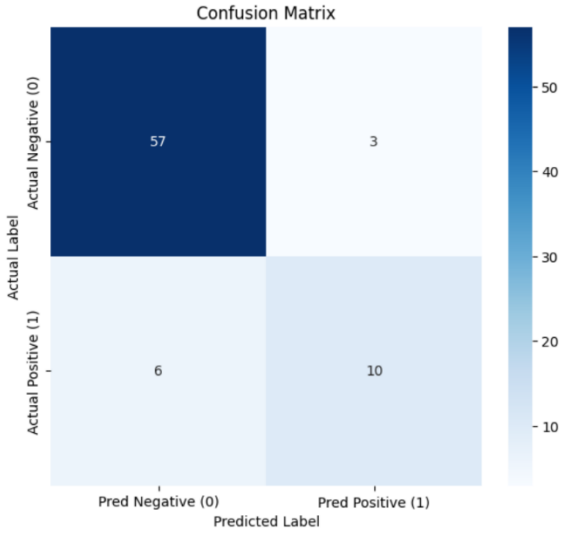


Figure: Confusion Matrix on the Test Set

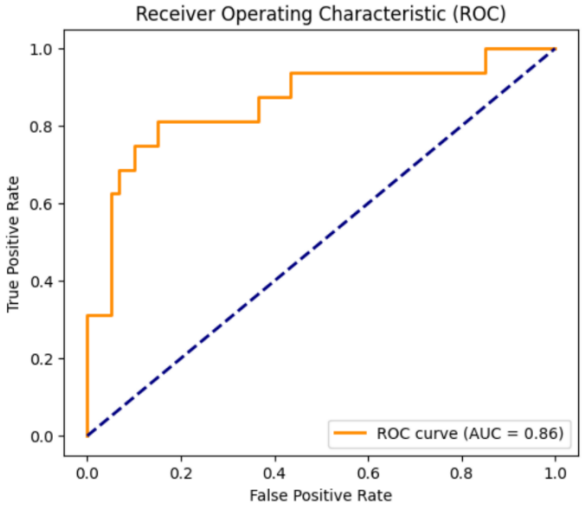


Figure: Receiver Operating Characteristic

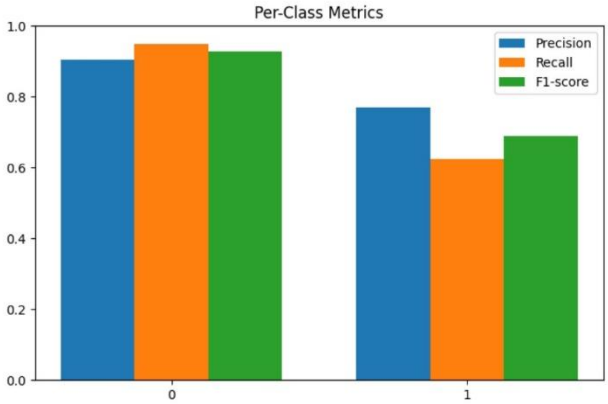


Figure 5 : pre-class Metrics

3.2. Qualitative Results: Visualization of Model Explainability

To provide insight into the model's decision-making process, we generated Grad-CAM heatmaps for cases in the test set. These heatmaps were overlaid onto the FLAIR MRI sequence and the corresponding tumor segmentation mask.

As shown in the figures below, the model consistently focused on specific intra-tumoral regions. The areas highlighted in red and yellow, indicating high importance.

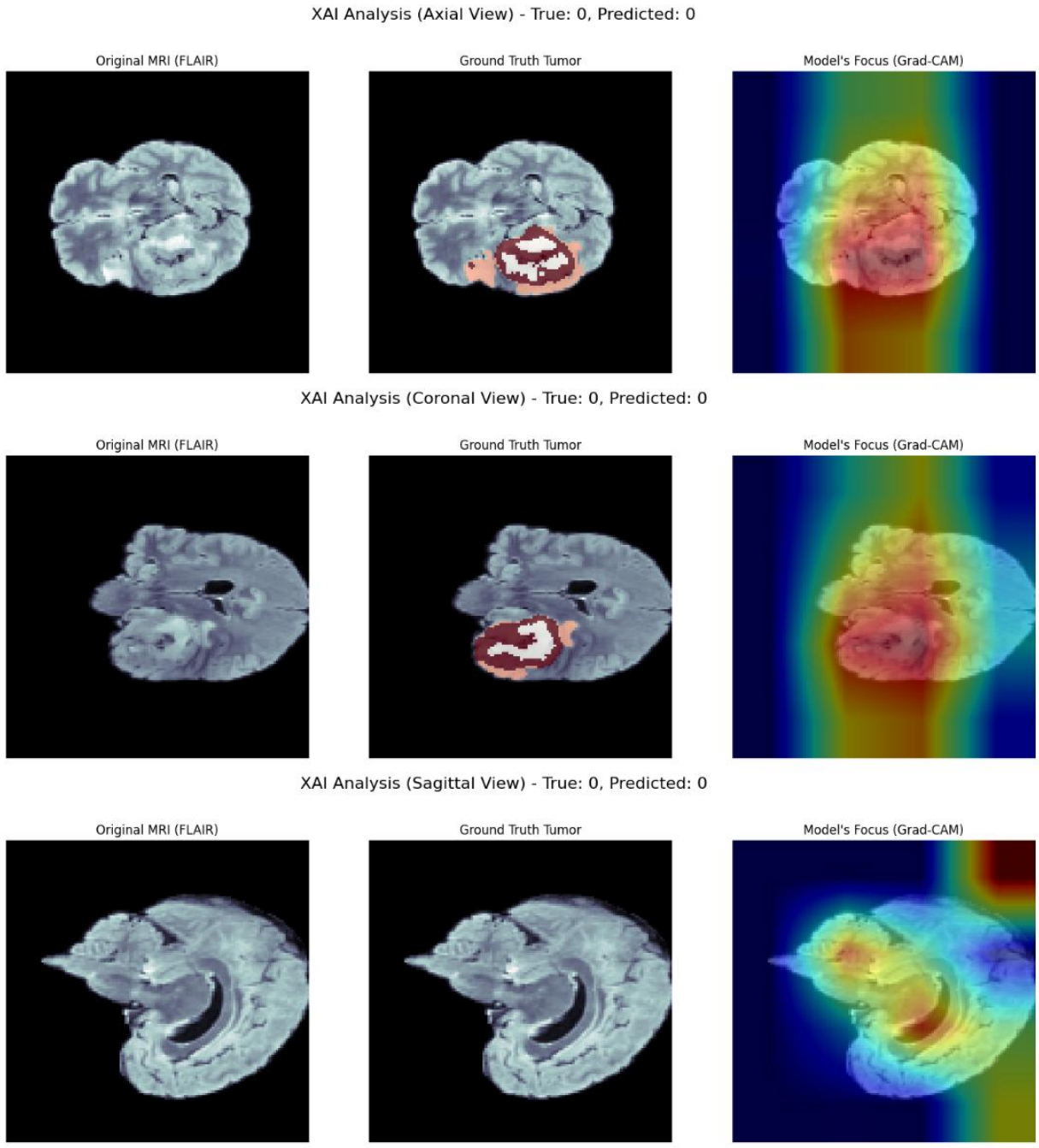


Figure: Grad-CAM visualization for a correctly classified IDH-mutant patient.