Non-Invasive Isocitrate Dehydrogenase Detection in Glioma from Multiparametric MRI using Artificial Intelligence: A Fully Automated, Explainable, Multicenter Study

Somayeh Farahani (PhD)[1,2,3], Marjaneh Hejazi (PhD)[1,4*], Antonio Di Ieva (MD, PhD)[3], Emad Fatemizadeh (PhD)[5], Ahmad Pour-Rashidi (MD)[1], Sidong Liu (PhD)[2,3]

[1]Department of Medical Physics and Biomedical Engineering, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran
[2]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW, Australia
[3]Computational NeuroSurgery (CNS) Lab, Faculty of Medicine, Health and Human Sciences, Macquarie Medical School, Macquarie University, Sydney, NSW, Australia
[4]Research Center for Molecular and Cellular Imaging, Bio-Optical Imaging Group, Tehran University of Medical Sciences, Tehran, Iran
[5]Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

ABSTRACT

Background: Glioma management depends on molecular profiling, particularly Isocitrate Dehydrogenase (IDH) status, but requires invasive biopsy. Radiomics applied to Magnetic Resonance Imaging (MRI) offers a non-invasive alternative, though translation is hindered by variability in segmentation, limited external validation, and lack of interpretability.

Objective: We aimed to develop and validate an interpretable radiomics pipeline for non-invasive IDH classification across MRI data.

Material and Methods: In this retrospective study, we assembled 1117 preoperative MRI scans from three public cohorts: The Cancer Genome Atlas (TCGA), the University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM), and the Erasmus Glioma Database (EGD). Tumor subregions were segmented using BrainSegFounder model. From the segmented regions, we extracted approximately 1900 PyRadiomics features per patient. Boruta selection was applied, followed by classification with k-nearest neighbors (kNN), Light Gradient Boosting Machine (LightGBM), and Logistic Regression (LR). Model interpretability was evaluated using Shapley Additive Explanations (SHAP).

Results: All classifiers achieved strong internal performance. On the external EGD test set, kNN provided the most balanced generalization (area under the receiver operating characteristic curve [AUC]=0.94, Matthews correlation coefficient [MCC]=0.74), whereas LR maximized sensitivity (0.93). On the TCGA test set, LR provided the best overall balance (AUC=0.91, MCC=0.68), and LightGBM achieved the highest specificity (0.94). SHAP analysis highlighted texture features derived from post-contrast T1-weighted images as key predictors.

Conclusion: Our pipeline achieves high IDH classification performance with full automation, cross-site generalizability, and transparency. These findings support its potential for clinical application in non-invasive glioma genotyping.

Keywords

Glioma; Machine Learning; Image Segmentation; Isocitrate Dehydrogenase; Radiomics

*Corresponding author:
Marjaneh Hejazi
Department of Medical Physics and Biomedical Engineering, School of Medicine, Tehran University of Medical Sciences, Tehran
E-mail:
mhejazi@sina.tums.ac.ir

Received: 3 October 2025
Accepted: 6 December 2025

Introduction

Gliomas are highly aggressive and heterogeneous primary brain tumors. Accurate molecular and histopathological profiling, particularly determination of Isocitrate Dehydrogenase (IDH) mutation status, is central to patient management [1]. IDH status carries critical prognostic and therapeutic significance. For instance, IDH-mutant gliomas generally have slower progression and better outcomes compared to IDH–wildtype tumors [2,3]. Invasive biopsy and genetic testing remain the reference standard, yet these procedures are invasive, vulnerable to sampling error, and carry procedural risk. Such limitations have motivated growing interest in non-invasive, imaging-based approaches [4].

Radiomics provides one such avenue by converting routine Magnetic Resonance Imaging (MRI) scans into high-dimensional quantitative features [5]. These features can capture subtle variations in tumor intensity, texture, and morphology that correlate with molecular status [6]. While previous studies have applied radiomics to glioma genotyping, findings remain inconsistent [7,8]. Radiomics workflows comprise lesion segmentation, feature extraction and selection, and predictive modeling [9]. Among these steps, segmentation is a major bottleneck. Most prior studies have relied on manual or semi-automatic tumor delineation, which is labor-intensive and subject to considerable inter-observer variability [10-12]. Differences in how tumors are contoured can significantly influence extracted features and subsequent model outputs, thereby undermining the reproducibility of radiomic biomarkers [13]. Automated segmentation methods can mitigate these issues by improving consistency and reducing the burden of manual delineation [10]. However, prior works did not implement robust automated approaches across heterogeneous, multicenter MRI scans, thereby limiting the generalizability of their findings [11,12,14].

Limited external validation represents another important issue. Many studies achieve strong performance on internal cross-validation but fail to generalize to unseen cohorts [7,15-18]. A recent systematic review reported that fewer than 30 percent of radiomics studies included an independent external validation cohort [7], and another review found that only 2 of 11 studies did so [8]. Without external testing, results risk overfitting to site-specific imaging characteristics. Furthermore, interpretability remains a persistent challenge. Both handcrafted radiomics and deep learning (DL) models are often criticized because their complex feature representations are not easily understandable to clinicians. This opacity can reduce trust in model outputs and ultimately hinder their integration into clinical practice [18,19].

To address these challenges, we developed an explainabe radiomics pipeline for non-invasive IDH classification on heterogeneous MRI data. We employed BrainSegFounder, a SWIN-UNETR–based architecture pretrained on more than 41000 brain MRIs, for automated tumor segmentation [20]. This approach provided consistent, reproducible segmentations and stabilized subsequent radiomics analyses across multi-institutional datasets. We then compared three widely used Machine Learning (ML) classifiers, k-nearest neighbor (kNN) [21], Light Gradient-Boosting Machine (LightGBM) [22], and logistic regression (LR) [23]. These methods span complementary families: ensemble learners (LightGBM), Linear Models (LR), and distance-based approaches (kNN). For interpretability, we applied SHapley Additive exPlanations (SHAP) to provide global and instance-level insights into model decisions [24]. Finally, we curated a multi-institutional cohort of 1117 preoperative glioma MRIs, of which 628 were used for external validation. Together, this framework integrates robust automatic segmentation, interpretable machine learning, and multicenter external validation, advancing radiomics toward clinically reliable glioma genotyping.

Material and Methods

Patient Population

This was a retrospective, multicenter observational diagnostic-modeling study with external validation. Because we used only publicly available, fully anonymized data, no institutional review board approval was required. We studied preoperative MRI scans from 1736 glioma patients (WHO grades 2–4) across three public cohorts: The Cancer Genome Atlas (TCGA; including the TCGA-LGG and TCGA-GBM collections; ~2006–2013) [25], the University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM; 2015–2021) [26], and the Erasmus Glioma Database (EGD; 2008–2018) dataset [27]. Patients were included if IDH status was known and all four MRI modalities, T1-weighted (T1), post-contrast T1-weighted (T1C), T2-weighted (T2), and Fluid-Attenuated Inversion Recovery (FLAIR), were available. We evaluated two experimental setups (Table 1). In scenario A, TCGA and UCSF-PDGM were used for training and internal validation, with EGD as the external test set. In scenario B, EGD and UCSF-PDGM were used for training and internal validation, with TCGA as the external test set.

Image Preprocessing and Tumor Segmentation

Raw MR images were only available for the TCGA dataset, which we processed using the Integrative Imaging Informatics for Cancer Research: Workflow Automation for Neuro-oncology (I3CR-WANO) framework [28]. For the other cohorts, we relied on the preprocessed data. All datasets, either preprocessed by I3CR-WANO or supplied preprocessed, were standardized through a common pipeline: registration to a shared anatomical space with 1×1×1 mm^3 voxels, bias-field correction, and skull stripping. As this was a retrospective study, acquisition protocols varied across centers. To maintain real-world variability, we did not exclude cases based on protocol differences or image quality. We further applied Z-score normalization of image intensities (centering to zero mean, unit variance) to mitigate inter-scanner and acquisition-protocol variation. MRI signal amplitudes differ substantially across scanners and protocols, which can introduce ‘batch effects’ and obscure the underlying biology. Z-score standardization ensures that intensity-based features reflect consistent, biologically relevant variation rather than variation introduced by scanner-specific scaling, thereby improving feature robustness and model reliability [29,30].

Expert tumor masks were not consistently available (Table 1). We therefore used the BrainSegFounder-Tiny model (SWIN-UNETR backbone) [20], originally self-supervised on UK Biobank data (41400 healthy brain MRIs) [31] and BraTS 2021 [32], and fine-tuned for tumor segmentation (stage 3). For this work, we further fine-tuned the stage 3 weights on UCSF-PDGM, restricting training to cases with all four preoperative modalities, expert masks, and no overlap with the BraTS 2021 training set. Model fine-tuning was carried out on a single A100 GPU (32 GB memory) with a batch size of 2 and a learning rate of 1×10^{-4}. To prevent overfitting and enhance training stability, early stopping with a patience of 15 epochs was employed to optimize the Dice loss. The model outputs three nested subregions: Enhancing Tumor (ET), peritumoral Edema (ED), and the Necrotic/Non-Enhancing Core (NCR/NET).

Radiomics Pipeline

Figure 1 summarizes our radiomics workflow. Radiomic features were extracted from each segmented subregion with PyRadiomics [33]. Feature families included first-order statistics, shape descriptors, and texture metrics derived from Gray-Level Run-Length Matrix (GLRLM), Gray-Level Co-Occurrence Matrix (GLCM), gray-Level Size Zone Matrix (GLSZM), Neighboring Gray-Tone

Table 1: Patient characteristics, availability of conventional MRI sequences, expert tumor segmentation labels, and exclusion criteria across TCGA, UCSF-PDGM, and EGD datasets. Class distributions for molecular and histological grades are reported as counts and percentages.

Datasets		TCGA (n=461)	UCSF-PDGM (n=501)	EGD (n=470)
Grade	2	47 (22%)	56 (11%)	119 (16%)
	3	59 (28%)	43 (9%)	78 (11%)
	4	107 (50%)	390 (80%)	474 (66%)
	Unknown	0	0	48 (7%)
IDH Status	Mutated	89 (42%)	103 (21%)	139 (20%)
	Wildtype	124 (58%)	386 (79%)	276 (38%)
	Unknown	0	0	304 (42%)
MRI Modalities		(T1, T1C, T2, FLAIR)	(T1, T1C, T2, FLAIR)	(T1, T1C, T2, FLAIR)
Expert Segmentation masks		Not available	All labels available	Only WT region available
Exclusion Criteria		•Missing preoperative MRI sequences (n=221) •Missing molecular pathological markers (n=14) •Failed in preprocessing (n=13)	•Follow-up images of the same patient (n=6) •Overlap with the TCGA dataset (n=6)	•Missing molecular pathological markers (n=3) •Failed in preprocessing (n=52)
Dataset Utilization	Scenario A	Training + Internal Validation(n=213)	Training + Internal Validation (n=489)	External Testing (n=415)
	Scenario B	External Testing (n=213)	Training + Internal Validation (n=489)	Training + Internal Validation (n=415)

TCGA: The Cancer Genome Atlas, MRI: Magnetic Resonance Imaging, UCSF-PDGM: University of California San Francisco Preoperative Diffuse Glioma MRI, EGD: Erasmus Glioma Database, T1: T1-weighted, T1C: Post-contrast T1-weighted, T2: T2-weighted, FLAIR: Fluid-Attenuated Inversion Recovery, IDH, Isocitrate Dehydrogenase, WT: Whole Tumor

Difference Matrix (NGTDM), and Gray-Level Dependence Matrix (GLDM). Per patient, this yielded around 1900 features. We assessed features from individual subregions, combined subregions and across sequences. We applied Boruta feature selection [34] to identify robust predictors. The algorithm introduces random "shadow" features and retains only real features that consistently outperform them. We then applied SHAP [24], a model-agnostic interpretability method, to estimate the contribution of each feature. SHAP assigns values that represent how much each feature influences the model's predictions, leading to the overall output to be interpreted as the sum of these contributions.

Evaluation and Statistical Analysis

No class weighting or resampling was used. All models were trained on the observed class frequencies. Model performance was assessed using balanced Accuracy (ACC), area under the receiver-operating characteristic Curve (AUC), sensitivity (SEN), specificity (SPE), F1-score (F1), and Matthews correlation

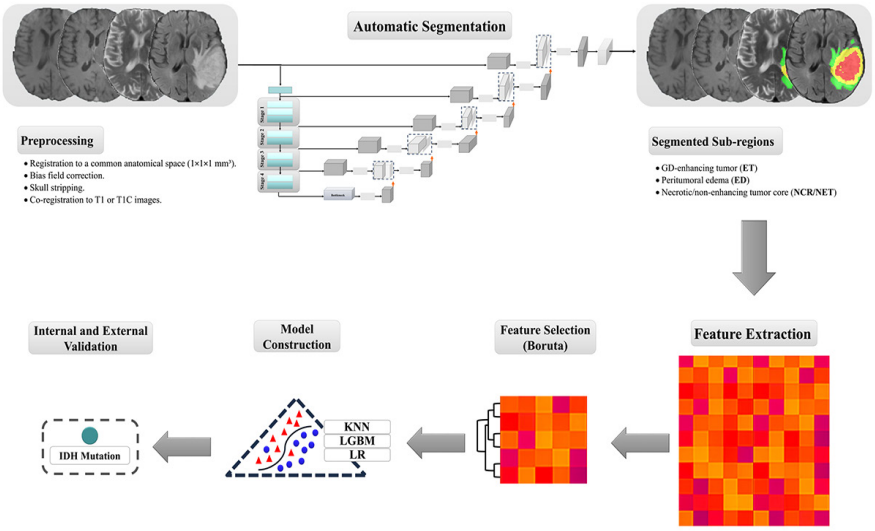

Figure 1: Overview of the proposed radiomics workflow.
(kNN: k-Nearest Neighbor, LightGBM: Light Gradient-Boosting Machine, LR: Logistic Regression)

coefficient (MCC). ACC reflects the proportion of correctly classified samples across classes, while AUC measures the ability of the model to discriminate between positive and negative cases across thresholds. F1 provides a harmonic balance between precision and recall, whereas MCC, ranging from −1 to +1, offers a more robust measure under class imbalance.

Normality was assessed with the Shapiro–Wilk test. Normally distributed measures were compared using one-way ANOVA with post-hoc pairwise tests (P-value<0.05), and non-normal data were analyzed with the Kruskal–Wallis test followed by pairwise Wilcoxon comparisons. All analyses were performed in R (v4.4.1) using the rstatix package.

Results

Following data curation, cases with missing preoperative scans (n=221), incomplete pathology records (n=17), duplicate follow-up scans from the same patients (n=6), cohort overlap (n=6), and failed preprocessing (n=65) were excluded. The final cohort comprised 1117 patients, including 331 with IDH-mutant and 786 with IDH-wildtype gliomas (Table 1).

In the first scenario, model performance was evaluated with five-fold cross-validation on the TCGA and UCSF-PDGM cohorts and independently tested on the EGD cohort (Table 2). In internal validation, all models demonstrated high AUC values (>0.90). LightGBM delivered the strongest overall performance, reaching an AUC of 0.94±0.05 and MCC of 0.72±0.11, while LR achieved the highest sensitivity (0.83±0.09). On the external EGD test set, kNN achieved the top AUC (0.94±0.01), with LightGBM close behind (0.92±0.01). LR again maximized sensitivity (0.93±0.02) but at the cost of specificity (0.76±0.05). Among all models, kNN demonstrated the best overall balance, achieving the highest F1-score (0.83±0.01) and MCC (0.74±0.02). The confusion matrices for the EGD cohort are shown in Figure 2A.

In scenario B, UCSF-PDGM and EGD were used for training and validation, with TCGA serving as the independent test set (Table 3). In internal validation, LightGBM achieved the highest specificity (0.97±0.02), while LR demonstrated the highest sensitivity (0.86±0.05). On the TCGA test set, LR showed the most balanced generalization performance

Table 2: Mean±standard deviation of evaluation metrics for prediction models across 5-fold cross-validation and external test cohort. The datasets include TCGA (The Cancer Genome Atlas), UCSF-PDGM (University of California San Francisco Preoperative Diffuse Glioma MRI), and EGD. The TCGA and UCSF-PDGM were used for training and cross-validation, and EGD served as external validation set (Scenario A). The best-performing mean values are highlighted in bold. Symbols denote statistically significant differences (P<0.05). Asterisks (*) indicate a significant difference from kNN, (#) indicate a significant difference from LightGBM, and daggers (†) denote a significant difference from LR.

Dataset	Metrics	Model		
		kNN	LightGBM	LR
Internal Validation	ACC	0.84±0.04	0.84±0.04	0.87±0.03
	SEN	0.72±0.15	0.71±0.13	0.83±0.09
	SPE	0.95±0.02	0.96±0.01#	0.90±0.04
	F1	0.77±0.10	0.78±0.10	0.78±0.05
	MCC	0.70±0.12	0.72±0.11	0.70±0.07
	AUC	0.92±0.05	0.94±0.05	0.93±0.03
EGD	ACC	0.88±0.01	0.84±0.01	0.85±0.03
	SEN	0.87±0.02	0.79±0.03†	0.93±0.02
	SPE	0.89±0.02	0.89±0.01†	0.76±0.05*
	F1	0.83±0.01	0.79±0.03	0.77±0.04
	MCC	0.74±0.02	0.69±0.04	0.65±0.06
	AUC	0.94±0.01	0.92±0.01	0.90±0.02

EGD: Erasmus Glioma Database, ACC: Accuracy, SEN: Sensitivity, SPE: Specificity, MCC: Matthews Correlation Coefficient, AUC: Area Under The Curve, kNN: k-Nearest Neighbor, LightGBM: Light Gradient-Boosting Machine, LR: Logistic Regression

(ACC: 0.85±0.01; MCC: 0.68±0.03; AUC: 0.91±0.01). kNN remained competitive (ACC: 0.82±0.01; MCC: 0.67±0.01), and LightGBM maintained the highest specificity (0.94±0.01) but with lower sensitivity (0.69±0.03). The corresponding confusion matrices for the TCGA cohort are shown in Figure 2B.

To probe feature contributions, we performed SHAP analyses. In the EGD cohort (Figure 3A), the highest-impact predictors were T1C-derived texture descriptors, GLRLM and GLSZM, together with T1C first-order total energy. Additional contributors included T1C GLSZM, T1C first-order Skewness, FLAIR NGTDM coarseness, and whole-tumor T2 GLCM. In the TCGA cohort (Figure 3B), SHAP again highlighted the predominance of T1C features, GLSZM, GLCM, and first-order statistics (median/mean and total energy), with complementary effects from whole-tumor shape sphericity, T1 root mean squared, T2 10th-percentile intensity, and FLAIR coarseness.

Instance-level SHAP force plots (Figure 3C and 3D) on the EGD cohort further illustrate how individual features influenced predictions. In Figure 3C, a correctly classified IDH-mutant case shows the score rising sharply above the base value (from 0.23 to 0.87), driven by strong positive contributions from T1C-derived texture and intensity descriptors, such as GLSZM, GLRLM, and first-order total energy and skewness. Conversely, Figure 3D depicts an IDH wild-type case, where dominant blue contributions lower the score to 0.03, with T1C texture features

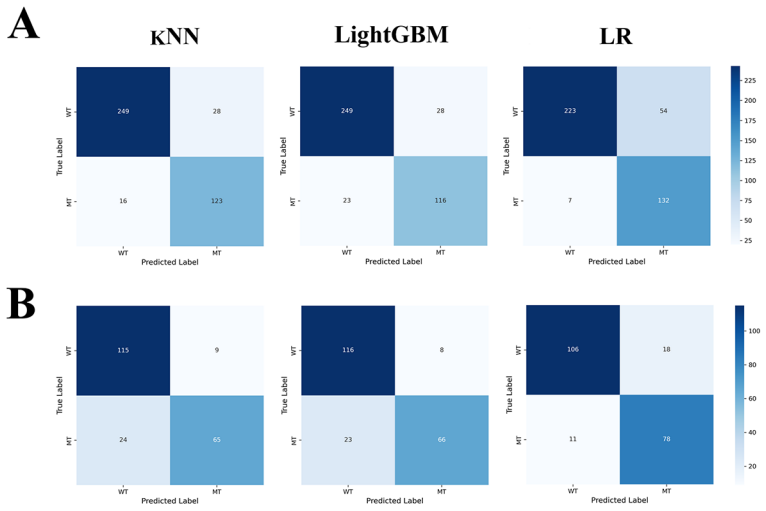

Figure 2: Confusion matrices of the six classifiers on external testing: (**A**) EGD cohort and (**B**) TCGA cohort.
(EGD: Erasmus Glioma Database, TCGA: The Cancer Genome Atlas, kNN: k-Nearest Neighbor, LightGBM: Light Gradient-Boosting Machine, LR: Logistic Regression, WT: Wildtype, MT: Mutant)

(GLRLM short-run high gray-level emphasis, GLSZM low-gray-level zone emphasis, and GLSZM gray-level non-uniformity) and shape sphericity collectively suppressing the probability despite minor positive effects from first-order Total Energy.

Discussion

Our results demonstrate that an integrated, explainable radiomics pipeline can non-invasively predict IDH mutation status in gliomas with high accuracy, even across heterogeneous, multicenter MRI data. Using a robust automated segmentation and a multimodal feature approach, we achieved high performance (AUC ~0.90–0.94), which is comparable to the best-reported values in the radiomics literature [7,8,19]. All three ML classifiers showed strong internal validation results and, importantly, maintained high performance on independent external datasets. In particular, kNN provided the most balanced generalization on the EGD cohort, while LR model maximized performance on the TCGA test set. Our external test AUC (~0.94) surpasses that of some prior multicenter radiomics studies (AUC ~0.82–0.86), highlighting the strength of our pipeline in capturing the IDH-specific imaging phenotype. This level of performance is also comparable to recent multi-task or foundation-based DL studies for IDH prediction [35-39]. Thus, with careful design, conventional radiomics coupled with modern segmentation can achieve accuracy rivaling more complex DL models, while offering greater transparency.

Our findings both confirm and extend observations from earlier radiomics studies. The recent systematic review and meta-analysis of DL-based radiomics reported pooled AUC of 0.89 for IDH prediction [7]. This analysis identified segmentation strategy and DL integration depth as major sources of heterogeneity and over-optimistic results when external validation was limited. Our framework was designed to address these issues. First, instead of relying on variable manual masks, we used BrainSegFounder [20], a state-of-the-art SWIN-UNETR model pretrained on more than 41000 MRI cases, to automatically

Table 3: Mean±standard deviation of evaluation metrics for prediction models across 5-fold cross-validation and external test cohort. The datasets include TCGA, UCSF-PDGM (University of California San Francisco Preoperative Diffuse Glioma MRI), and EGD (Erasmus Glioma Database). The EGD and UCSF-PDGM were used for training and cross-validation (Scenario B). The best-performing mean values are highlighted in bold. Symbols represent statistically significant differences (P<0.05). Asterisks (*) indicate a significant difference from kNN , (#) indicate a significant difference from LightGBM, and daggers (†) denote a significant difference from LR.

Dataset	Metrics	Model		
		kNN	LightGBM	LR
Internal Validation	ACC	0.85±0.02	0.86±0.02	0.89±0.02
	SEN	0.75±0.09	0.74±0.08	0.86±0.05
	SPE	0.94±0.02	0.97±0.02	0.91±0.02#
	F1	0.78±0.04	0.80±0.05	0.82±0.03
	MCC	0.71±0.05	0.75±0.07	0.75±0.04
	AUC	0.92±0.02	0.94±0.02	0.94±0.02
TCGA	ACC	0.82±0.01	0.82±0.02	0.85±0.01
	SEN	0.72±0.02	0.69±0.03†	0.82±0.04*
	SPE	0.92±0.01	0.94±0.01	0.87±0.02#
	F1	0.79±0.01	0.77±0.02	0.81±0.02
	MCC	0.67±0.01	0.66±0.03	0.68±0.03
	AUC	0.88±0.01	0.90±0.01	0.91±0.01

TCGA: The Cancer Genome Atlas, ACC: Accuracy, SEN: Sensitivity, SPE: Specificity, MCC: Matthews Correlation Coefficient, AUC: Area Under The Curve, kNN: k-Nearest Neighbor, LightGBM: Light Gradient-Boosting Machine, LR: Logistic Regression

segment tumor subregions. This approach not only reduced manual effort but also generated consistent tumor masks, improving the stability of downstream radiomic features. Second, unlike previous studies that trained and tested models on the same institutional data [7, 8], we intentionally built training and test splits across different centers and scanners. This cross-site validation is essential to prevent models from exploiting cohort-specific artifacts and ensures stronger generalizability.

Beyond raw performance, a major contribution of this work is the integration of explainability via SHAP analysis. ML and DL models have often been criticized as "black boxes," which hinders clinical trust and adoption [18,19]. In our study, SHAP enabled us to examine the model's decision process and identify the features that most strongly influenced IDH predictions. Texture features from T1C, particularly GLRLM and GLSZM, consistently emerged as the most important contributors. This aligns with clinical radiology experience, IDH-wildtype gliomas (generally higher grade) may exhibit more irregular, heterogeneous enhancement, whereas IDH-mutant tumors (often lower grade) tend to show more uniform or minimal enhancement [29]. These explainable insights are valuable to both clinicians and researchers. They provide biological plausibility to the model's behavior and could guide further hypothesis generation.

We acknowledge several limitations in our study. First, as a retrospective study spanning multiple scanners and protocols, acquisition heterogeneity may still affect feature distributions despite our harmonization. While we chose not to exclude cases with lower quality

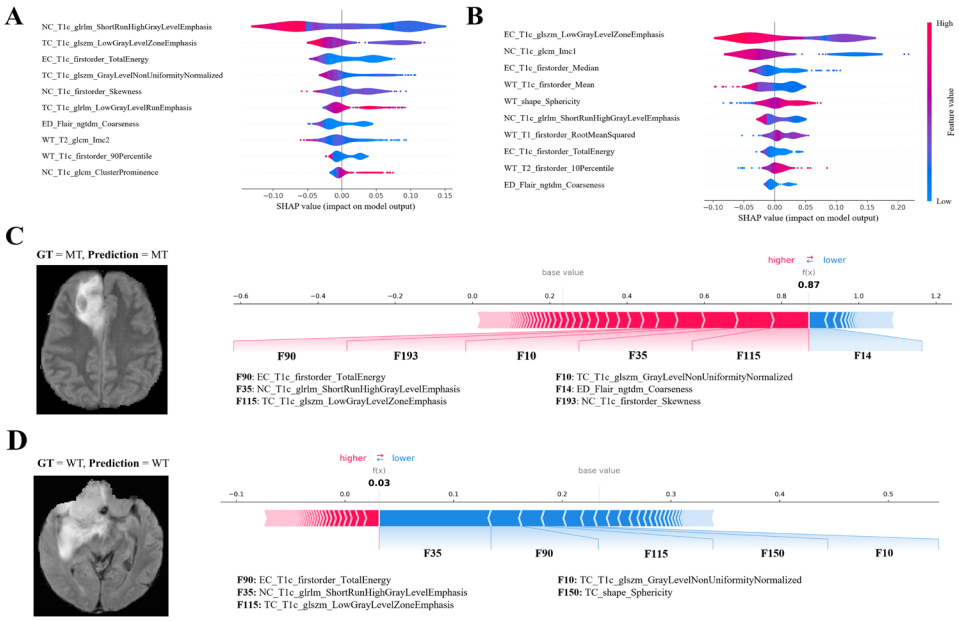

Figure 3: SHAP explanations for IDH prediction. (**A**) SHAP summary plot of the top ten features influencing model predictions in the EGD cohort and (**B**) in the TCGA cohort. (**C**) SHAP force plot illustrating a case correctly classified as IDH-mutant, and (**D**) a case correctly classified as IDH–wildtype.
(SHAP: Shapley Additive Explanations, IDH: isocitrate dehydrogenase, EGD: Erasmus Glioma Database, TCGA: The Cancer Genome Atlas, NC: Necrotic Core, TC: Tumor Core, EC: Enhancing Tumor Core, ED: Peritumoral Edema, WT: Wildtype, GT: Ground Truth, MT: Mutant)

or differing protocols (to maintain real-world heterogeneity), this could introduce noise. Second, our radiomic features were purely imaging-based. We did not integrate clinical factors or other biomarkers. In practice, a physician would consider imaging alongside patient data, and recent evidence indicates that combining radiomics with clinical features can improve predictive performance [8]. Additionally, we focused solely on IDH mutation classification. Downstream decisions, however, also depend on 1p/19q co-deletion status and tumor grade (in IDH-mutants). An ideal non-invasive tool would provide a full molecular profile. Lastly, while we did perform external validation, prospective clinical validation is still needed [18]. A prospective study would evaluate how our model performs on new patients in real-time and whether its predictions can truly influence or improve clinical decision-making. This step will be crucial for regulatory approval and clinical uptake of any artificial intelligence-driven diagnostic tool.

Conclusion

In this retrospective study, we developed and validated a fully automated, explainable radiomics pipeline for non-invasive prediction of IDH mutation status in gliomas using routine multiparametric MRI. Using a foundation-based SWIN-UNETR model for tumor segmentation, handcrafted radiomic features, and three complementary ML classifiers, we achieved consistently high performance across internal datasets. The models also maintained strong performance on independent test sets from different institutions and scanners, highlighting the generalizability of the proposed approach.

Beyond predictive accuracy, our framework

provides transparent insights into model decision-making. SHAP analyses consistently highlighted texture and first-order features derived from T1C images as dominant contributors to IDH classification. This interpretability can facilitate clinician trust, enable quality control of predictions, and support hypothesis generation regarding imaging correlates of glioma biology.

Our study addresses key barriers to the clinical translation of radiomics in neuro-oncology, including dependence on manual segmentation, limited external validation, and lack of model transparency. Important steps remain before clinical deployment. Future work should assess the value of integrating clinical, histopathologic, and additional molecular markers, including 1p/19q co-deletion and grade within IDH-mutant tumors. With prospective evaluation, this approach has the potential to support preoperative decision-making, reduce reliance on invasive tissue sampling, and contribute to more personalized management of patients with glioma.

Acknowledgment

During the preparation of this work, the first author used OpenAI ChatGPT-4o to assist with identifying and correcting grammatical and typographical errors. All content was reviewed and revised by the authors as needed, and the authors take full responsibility for the final manuscript.

Authors' Contribution

S. Farahani conceived and designed the study, gathered and curated the datasets, developed the methodology, performed formal analysis and validation, generated all visualizations, and wrote the original draft. M. Hejazi supervised the work and critically reviewed and edited the manuscript. A. Di Ieva reviewed and edited the manuscript. E. Fatemizadeh reviewed and edited the manuscript. A. Pour-Rashidi reviewed and edited the manuscript. S. Liu provided project administration and supervision, and reviewed and edited the manuscript. All authors read, revised, and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Ethical Approval

This retrospective analysis used only publicly available, de-identified data: TCGA (NCI policies and consent framework), UCSF-PDGM (hosted on TCIA as a public, de-identified collection), and the Erasmus Glioma Database (EGD) (released with scans defaced and registered to a common atlas). Per ICMJE recommendations and institutional policy on secondary analysis of public de-identified data, this work is not human subjects research and IRB review/informed consent were not required. All use complied with the datasets' terms, and the study followed the Declaration of Helsinki.

Funding

This work was supported by the Research Chancellor of Tehran University of Medical Sciences (TUMS), Tehran, Iran (Grant No. 1402-4-410-69326).

Conflict of Interest

None

Data Availability Statement

The datasets are available at the following links:

TCGA-LGG [25]: https://www.cancerimagingarchive.net/collection/tcga-lgg

TCGA-GBM [25]: https://www.cancerimagingarchive.net/collection/tcga-gbm

UCSF-PDGM [26]: https://www.cancerimagingarchive.net/collection/ucsf-pdgm

EGD [27]: https://xnat.health-ri.nl/REST/projects/egd.

References

1. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO Classification

of Tumors of the Central Nervous System: a summary. *Neuro Oncol.* 2021;**23**(8):1231-51. doi: 10.1093/neuonc/noab106. PubMed PMID: 34185076. PubMed PMCID: PMC8328013.

2. Cai J, Zhu P, Zhang C, Li Q, Wang Z, Li G, et al. Detection of ATRX and IDH1-R132H immunohistochemistry in the progression of 211 paired gliomas. *Oncotarget.* 2016;**7**(13):16384-95. doi: 10.18632/oncotarget.7650. PubMed PMID: 26918938. PubMed PMCID: PMC4941322.
3. Núñez FJ, Mendez FM, Kadiyala P, Alghamri MS, Savelieff MG, Garcia-Fabiani MB, et al. IDH1-R132H acts as a tumor suppressor in glioma via epigenetic up-regulation of the DNA damage response. *Sci Transl Med.* 2019;**11**(479):eaaq1427. doi: 10.1126/scitranslmed.aaq1427. PubMed PMID: 30760578. PubMed PMCID: PMC6400220.
4. Han S, Liu Y, Cai SJ, Qian M, Ding J, Larion M, et al. IDH mutation in glioma: molecular mechanisms and potential therapeutic targets. *Br J Cancer.* 2020;**122**(11):1580-9. doi: 10.1038/s41416-020-0814-x. PubMed PMID: 32291392. PubMed PMCID: PMC7250901.
5. Jang K, Russo C, Di Ieva A. Radiomics in gliomas: clinical implications of computational modeling and fractal-based analysis. *Neuroradiology.* 2020;**62**(7):771-90. doi: 10.1007/s00234-020-02403-1. PubMed PMID: 32249351.
6. Di Salle G, Tumminello L, Laino ME, Shalaby S, Aghakhanyan G, Fanni SC, et al. Accuracy of Radiomics in Predicting IDH Mutation Status in Diffuse Gliomas: A Bivariate Meta-Analysis. *Radiol Artif Intell.* 2024;**6**(1):e220257. doi: 10.1148/ryai.220257. PubMed PMID: 38231039. PubMed PMCID: PMC10831518.
7. Farahani S, Hejazi M, Tabassum M, Di Ieva A, Mahdavifar N, Liu S. Diagnostic performance of deep learning for predicting glioma isocitrate dehydrogenase and 1p/19q co-deletion in MRI: a systematic review and meta-analysis. *Eur Radiol.* 2025:1-30. doi: 10.1007/s00330-025-11898-2. PubMed PMID: 40817944.
8. Chung CYC, Pigott LE. Predicting IDH and ATRX mutations in gliomas from radiomic features with machine learning: a systematic review and meta-analysis. *Front Radiol.* 2024;**4**:1493824. doi: 10.3389/fradi.2024.1493824. PubMed PMID: 39544481. PubMed PMCID: PMC11560782.
9. Scalco E, Rizzo G, Mastropietro A. The stability of oncologic MRI radiomic features and the potential role of deep learning: a review. *Phys Med Biol.* 2022;**67**(9). doi: 10.1088/1361-6560/ac60b9. PubMed PMID: 35325881.
10. Yu Y, Li GF, Tan WX, Qu XY, Zhang T, Hou XY, et al. Towards automatical tumor segmentation in radiomics: a comparative analysis of various methods and radiologists for both region extraction and downstream diagnosis. *BMC Med Imaging.* 2025;**25**(1):63. doi: 10.1186/s12880-025-01596-2. PubMed PMID: 40000987. PubMed PMCID: PMC11863488.
11. Ren Y, Zhang X, Rui W, Pang H, Qiu T, Wang J, Xie Q, et al. Noninvasive Prediction of IDH1 Mutation and ATRX Expression Loss in Low-Grade Gliomas Using Multiparametric MR Radiomic Features. *J Magn Reson Imaging.* 2019;**49**(3):808-17. doi: 10.1002/jmri.26240. PubMed PMID: 30194745.
12. Shboul ZA, Chen J, M Iftekharuddin K. Prediction of Molecular Mutations in Diffuse Low-Grade Gliomas using MR Imaging Features. *Sci Rep.* 2020;**10**(1):3711. doi: 10.1038/s41598-020-60550-0. PubMed PMID: 32111869. PubMed PMCID: PMC7048831.
13. Pati S, Verma R, Akbari H, Bilello M, Hill VB, Sako C, et al. Reproducibility analysis of multi-institutional paired expert annotations and radiomic features of the Ivy Glioblastoma Atlas Project (Ivy GAP) dataset. *Med Phys.* 2020;**47**(12):6039-52. doi: 10.1002/mp.14556. PubMed PMID: 33118182. PubMed PMCID: PMC8382093.
14. Haubold J, Hosch R, Parmar V, Glas M, Guberina N, Catalano OA, et al. Fully Automated MR Based Virtual Biopsy of Cerebral Gliomas. *Cancers (Basel).* 2021;**13**(24):6186. doi: 10.3390/cancers13246186. PubMed PMID: 34944806. PubMed PMCID: PMC8699054.
15. Kim JY, Park JE, Jo Y, Shim WH, Nam SJ, Kim JH, et al. Incorporating diffusion- and perfusion-weighted MRI into a radiomics model improves diagnostic performance for pseudoprogression in glioblastoma patients. *Neuro Oncol.* 2019;**21**(3):404-14. doi: 10.1093/neuonc/noy133. PubMed PMID: 30107606. PubMed PMCID: PMC6380413.
16. Casale R, Lavrova E, Sanduleanu S, Woodruff HC, Lambin P. Development and external validation of a non-invasive molecular status predictor of chromosome 1p/19q co-deletion based on MRI radiomics analysis of Low Grade Glioma patients. *Eur J Radiol.* 2021;**139**:109678. doi: 10.1016/j.ejrad.2021.109678. PubMed PMID: 33848780.
17. Doniselli FM, Pascuzzo R, Mazzi F, Padelli F, Moscatelli M, Akinci D'Antonoli T, et al. Quality assessment of the MRI-radiomics studies for MGMT promoter methylation prediction in glioma: a systematic review and meta-analysis. *Eur Radiol.* 2024;**34**(9):5802-15. doi: 10.1007/s00330-024-10594-x. PubMed PMID: 38308012. PubMed PMCID: PMC11364578.
18. Farahani S, Hejazi M, Moradizeyveh S, Di Ieva A, Fatemizadeh E, Liu S. Diagnostic Accuracy of Deep Learning Models in Predicting Glioma Molecular Markers: A Systematic Review and Meta-Analysis. *Diagnostics (Basel).* 2025;**15**(7):797. doi: 10.3390/diagnostics15070797. PubMed PMID: 40218147. PubMed PMCID: PMC11988998.
19. Bhandari AP, Liong R, Koppen J, Murthy SV, Lasocki A. Noninvasive Determination of IDH and 1p19q Status of Lower-grade Gliomas Using MRI Radiomics:

A Systematic Review. *AJNR Am J Neuroradiol.* 2021;**42**(1):94-101. doi: 10.3174/ajnr.A6875. PubMed PMID: 33243896. PubMed PMCID: PMC7814803.

20. Cox J, Liu P, Stolte SE, Yang Y, Liu K, See KB, et al. BrainSegFounder: Towards 3D foundation models for neuroimage segmentation. *Med Image Anal.* 2024;**97**:103301. doi: 10.1016/j.media.2024.103301. PubMed PMID: 39146701. PubMed PMCID: PMC11382327.
21. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory*. 1967;**13**(1):21-7. doi: 10.1109/TIT.1967.1053964.
22. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. 30th International Conference on Advances in Neural Information Processing Systems; NIPS; 2017. p. 3149-57.
23. Cox DR. The regression analysis of binary sequences. *J R Stat Soc Series B Stat Methodol.* 1958;**20**(2):215-32. doi: 10.1111/j.2517-6161.1958.tb00292.x.
24. Mangalathu S, Hwang SH, Jeon JS. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures.* 2020;**219**:110927. doi: 10.1016/j.engstruct.2020.110927.
25. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data.* 2017;**4**:170117. doi: 10.1038/sdata.2017.117. PubMed PMID: 28872634. PubMed PMCID: PMC5685212.
26. Calabrese E, Villanueva-Meyer JE, Rudie JD, Rauschecker AM, Baid U, Bakas S, et al. The University of California San Francisco Preoperative Diffuse Glioma MRI Dataset. *Radiol Artif Intell.* 2022;**4**(6):e220058. doi: 10.1148/ryai.220058. PubMed PMID: 36523646. PubMed PMCID: PMC9748624.
27. Van Der Voort SR, Incekara F, Wijnenga MMJ, Kapsas G, Gahrmann R, Schouten JW, et al. The Erasmus Glioma Database (EGD): Structural MRI scans, WHO 2016 subtypes, and segmentations of 774 patients with glioma. *Data Brief.* 2021;**37**:107191. doi: 10.1016/j.dib.2021.107191. PubMed PMID: 34159239. PubMed PMCID: PMC8203723.
28. Chakrabarty S, Abidi SA, Mousa M, Mokkarala M, Hren I, Yadav D, et al. Integrative Imaging Informatics for Cancer Research: Workflow Automation for Neuro-Oncology (I3CR-WANO). *JCO Clin Cancer Inform.* 2023;**7**:e2200177. doi: 10.1200/CCI.22.00177. PubMed PMID: 37146265. PubMed PMCID: PMC10281444.
29. Carré A, Klausner G, Edjlali M, Lerousseau M, Briend-Diop J, Sun R, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Sci Rep.* 2020;**10**(1):12340. doi: 10.1038/s41598-020-69298-z. PubMed PMID: 32704007. PubMed PMCID: PMC7378556.
30. Panic J, Defeudis A, Balestra G, Giannini V, Rosati S. Normalization Strategies in Multi-Center Radiomics Abdominal MRI: Systematic Review and Meta-Analyses. *IEEE Open J Eng Med Biol.* 2023;**4**:67-76. doi: 10.1109/OJEMB.2023.3271455. PubMed PMID: 37283773. PubMed PMCID: PMC10241248.
31. Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat Commun.* 2020;**11**(1):2624. doi: 10.1038/s41467-020-15948-9. PubMed PMID: 32457287. PubMed PMCID: PMC7250878.
32. Baid U, Ghodasara S, Mohan S, Bilello M, Calabrese E, Colak E, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification [Internet]. arXiv [Preprint]. 2021 [cited 2021 Jun 5]. Available from: https://arxiv.org/abs/2107.02314.
33. Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017;**77**(21):e104-7. doi: 10.1158/0008-5472.CAN-17-0339. PubMed PMID: 29092951. PubMed PMCID: PMC5672828.
34. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw.* 2010;**36**:1-3. doi: 10.18637/jss.v036.i11.
35. Wu X, Zhang S, Zhang Z, He Z, Xu Z, Wang W, et al. Biologically interpretable multi-task deep learning pipeline predicts molecular alterations, grade, and prognosis in glioma patients. *NPJ Precis Oncol.* 2024;**8**(1):181. doi: 10.1038/s41698-024-00670-2. PubMed PMID: 39152182. PubMed PMCID: PMC11329669.
36. Chen Q, Wang L, Deng Z, Wang R, Wang L, Jian C, Zhu YM. Cooperative multi-task learning and interpretable image biomarkers for glioma grading and molecular subtyping. *Med Image Anal.* 2025;**101**:103435. doi: 10.1016/j.media.2024.103435. PubMed PMID: 39778265.
37. Zhang J, Cao J, Tang F, Xie T, Feng Q, Huang M. Multi-level Feature Exploration and Fusion Network for Prediction of IDH Status in Gliomas from MRI. *IEEE J Biomed Health Inform.* 2023;**28**(1):42-53. doi: 10.1109/JBHI.2023.3279433. PubMed PMID: 37247314.
38. Farahani S, Hejazi M, Di Ieva A, Liu S. FoundBioNet: A Foundation-Based Model for IDH Genotyping of Glioma from Multi-parametric MRI. In International Conference on Medical Image Computing and Computer-Assisted Intervention; Switzerland: Springer, Cham; 2025. p. 259-70.
39. Chen M, Zhang M, Yin L, Ma L, Ding R, Zheng T, Yue Q, Lui S, Sun H. Medical image foundation models in assisting diagnosis of brain tumors: a pilot study. *Eur Radiol.* 2024;**34**(10):6667-79. doi: 10.1007/s00330-024-10728-1. PubMed PMID: 38627290.