Radiomics-based models versus traditional imaging characteristics in the diagnostic grading of gliomas: A comparative study

Keywords
MRI, Machine Learning, Glioma, Diagnostic Accuracy, Radiomics, Tumor Grading

Abstract

Introduction
Gliomas were highly aggressive primary brain tumors that require precise grading for effective treatment planning. Current grading relies on histopathological analysis, which can be invasive and subjective. Radiomics,a non-invasive approach that extracts quantitative features from medical images,may offer enhanced diagnostic accuracy.

Material and methods
This retrospective study analyzed 263 glioma patients from January 2021 to December 2023. Patients were categorized into low-grade glioma (LGG) and high-grade glioma (HGG) groups based on WHO classifications. MRI images were used to extract radiomic features,compared to traditional imaging characteristics. PyRadiomics facilitated extraction of features such as tumor sphericity and GLCM contrast.Logistic regression with 10-fold cross-validation assessed the diagnostic accuracy,with Receiver Operating Characteristic (ROC) curve analysis and the DeLong test (Bonferroni-adjusted) used to compare the predictive performance of the models.

Results
Significant differences were observed in radiomics features between LGG and HGG groups, including tumor sphericity and GLCM contrast,with higher heterogeneity in HGG.The radiomics-based model achieved a bootstrap-corrected area under the curve (AUC) of 0.847 (95% CI: 0.80-0.89), sensitivity of 82.0%, and specificity of 38.7%.Traditional imaging characteristics alone demonstrated an AUC of 0.806 (95% CI: 0.76-0.85), with sensitivity and specificity at 80.0% and 35.4%, respectively. The combined model of radiomics and traditional imaging provided an enhanced AUC of 0.872 (95% CI: 0.83-0.91), with improved accuracy (80.4%), sensitivity (84.0%), and specificity (44.8%).

Conclusions
Radiomics-based models significantly enhance the diagnostic grading of gliomas compared to traditional imaging alone by capturing subtle tumor heterogeneity and complexity. The integration of radiomics and traditional imaging offers superior diagnostic performance, potentially guiding more accurate treatment strategies.

Radiomics-based models versus traditional imaging characteristics in the diagnostic grading of gliomas: A comparative study

Abstract

Background: Gliomas were highly aggressive primary brain tumors that require precise grading for effective treatment planning. Current grading relies on histopathological analysis, which can be invasive and subjective. Radiomics, a non-invasive approach that extracts quantitative features from medical images, may offer enhanced diagnostic accuracy. This study compares radiomics-based models against traditional imaging characteristics in grading gliomas.

Methods: This retrospective study analyzed 263 glioma patients from January 2021 to December 2023. Patients were categorized into low-grade glioma (LGG) and high-grade glioma (HGG) groups based on WHO classifications. MRI images were used to extract radiomic features, compared to traditional imaging characteristics. PyRadiomics facilitated extraction of features such as tumor sphericity and GLCM contrast. Logistic regression with 10-fold cross-validation assessed the diagnostic accuracy, with Receiver Operating Characteristic (ROC) curve analysis and the DeLong test (Bonferroni-adjusted) used to compare the predictive performance of the models.

Results: Significant differences were observed in radiomics features between LGG and HGG groups, including tumor sphericity and GLCM contrast, with higher heterogeneity in HGG. The radiomics-based model achieved a bootstrap-corrected area under the curve (AUC) of 0.847 (95% CI: 0.80-0.89), sensitivity of 82.0%, and specificity of 38.7%. Traditional imaging characteristics alone demonstrated an AUC of 0.806 (95% CI: 0.76-0.85), with sensitivity and specificity at 80.0% and 35.4%, respectively. The combined model of radiomics and traditional imaging provided an AUC of 0.872 (95% CI: 0.83-0.91), with improved accuracy (80.4%), sensitivity (84.0%), and specificity (44.8%), though the difference between radiomics alone and the combined model did not reach statistical significance after Bonferroni correction (adjusted $p = 0.18$). The diagnostic performance of both the radiomics model and combined model was significantly better than traditional imaging alone (adjusted $p = 0.036$ and 0.012, respectively). Importantly, all models demonstrated low specificity (35.4-44.8%), indicating a high false-positive rate that warrants

careful clinical interpretation.

Conclusion: Radiomics-based models significantly enhance the diagnostic grading of gliomas compared to traditional imaging alone by capturing subtle tumor heterogeneity and complexity. However, the low specificity observed across all models raises important clinical concerns regarding potential overdiagnosis of low-grade tumors as high-grade, necessitating careful interpretation in clinical practice. The integration of radiomics and traditional imaging shows promise but requires further validation with standardized imaging protocols across uniform field strengths before routine clinical implementation.

Keywords

Glioma; Radiomics; MRI; Tumor Grading; Diagnostic Accuracy; Machine Learning

1. Introduction

Gliomas are among the most common and aggressive primary brain tumors, accounting for approximately 46% of all intracranial tumors. According to data, the incidence of glioma is 3 to 10 per 100,000 people, comprising 80% of all malignant brain tumors and 1% to 3% of all malignant tumors in the body(1, 2). Gliomas exhibit a broad clinical spectrum, ranging from relatively indolent low-grade lesions to highly aggressive high-grade tumors, with significant variations in patient outcomes. The prognosis for low-grade gliomas (LGGs) is generally more favorable, with median survival potentially extending over several years. In contrast, high-grade gliomas (HGGs), particularly glioblastomas, have a poorer prognosis, with median survival typically not exceeding 15 months despite aggressive treatment. Therefore, accurate diagnostic grading is crucial for tailoring optimal treatment plans and predicting patient outcomes, enabling early intervention to potentially improve survival rates and quality of life. The World Health Organization (WHO) classification system divides gliomas into four grades based on histopathological and molecular characteristics, underscoring the necessity of precise diagnostic grading to tailor the best treatment plan and effectively predict patient outcomes(3-5).

Traditionally, glioma grading has been predominantly reliant on histopathological analyses; however, such invasive procedures were often limited by sampling error and accessibility issues in eloquent brain areas. Consequently, non-invasive imaging techniques, particularly magnetic resonance imaging (MRI), have emerged as critical tools in the preoperative

assessment of gliomas(6, 7). Conventional imaging characteristics---such as tumor size, border definition, presence of edema, and enhancement patterns---have long served as surrogate markers for glioma grading. However, the interpretation of these features can vary significantly due to subjective judgment, even among experienced radiologists(8, 9).

In recent years, radiomics has emerged as a transformative approach in medical imaging, aiming to transcend the limitations of traditional imaging through the extraction and analysis of a vast array of quantitative imaging features(10, 11). These features, derived from complex algorithms applied to standard imaging datasets, capture the underlying heterogeneity and texture differences within tumors that might be subtle or invisible to the human eye. Recent advances in radiomics have demonstrated its potential in various cancer types, including gliomas, by revealing imaging biomarkers that correlate with molecular subtypes and clinical outcomes(12, 13). Radiomics operates under the premise that these imaging biomarkers reflect the tumor's biological and microenvironmental characteristics, thus offering a more nuanced insight into the tumor's phenotype. This study delves into the comparative performance of radiomics-based models and traditional imaging modalities in the diagnostic grading of gliomas.

2. Materials and methods

2.1 Study design

This study received approval from our institution's Institutional Review Board and Ethics Committee (IRB No. Xxxxxx). For this retrospective study, informed consent was waived because we exclusively used de-identified patient data with all personal identifiers removed, ensuring no risk or adverse impact on patient care. This waiver complies with institutional guidelines for retrospective research using anonymized data.

A retrospective analysis was performed on the cases of 263 glioma patients treated at our hospital between January 2021 and December 2023. Patient data were obtained from the medical record system. According to the WHO grading criteria for gliomas, patients were categorized into two groups: low-grade glioma (LGG) and high-grade glioma (HGG). The LGG group comprised 137 patients with grade I and grade II gliomas, while the HGG group consisted of 126 patients with grade III and grade IV gliomas.

2.2 Eligibility and grouping criteria

Inclusion Criteria: Patients were included based on the WHO diagnostic criteria for gliomas(14), specifically those classified as grades I to IV undergoing tumor resection surgery. Additionally, preoperative contrast-enhanced T1-weighted (CE-T1W), multiplanar reconstruction (MPR), and magnetic resonance (MR) images of high quality (characterized by a high signal-to-noise ratio, high contrast, and absence of imaging artifacts or distortion) were required. All MRI data had to be collected using the same machine prior to surgery. Moreover, a precise postoperative pathological diagnosis with accurate histological grading and well-preserved pathology sections was necessary.

Exclusion Criteria: Patients were excluded if their images exhibited obvious head movement or artifacts, or if they had a history of other malignant tumors or severe neurological diseases prior to MRI scanning. Other exclusions included those who received a biopsy, radiotherapy, or chemotherapy before MRI, allergies to contrast agents, renal insufficiency, pregnancy or lactation, or the presence of cardiac pacemakers, certain surgical implants, or metal fragments in the body. Additionally, patients with severe heart disease, anxiety disorders, or claustrophobia were excluded from the study.

2.3 Radiomics model establishment

Each patient underwent an MRI scan within two weeks prior to surgery. All MRI examinations for radiomics feature extraction were conducted using a 3.0 T Discovery MR750 scanner (GE Healthcare) equipped with a 32-channel head coil, across two centers. The MRI protocol included the following sequences: 1) Axial T2-weighted imaging (T2WI) with parameters: repetition time (TR) of 4,480 ms, echo time (TE) of 120 ms, slice thickness of 5 mm, slice gap of 1.5 mm, 20 slices, a field of view (FOV) of 240 × 240 mm, and a matrix size of 320 × 288. 2) Axial T2 fluid-attenuated inversion recovery (T2FLAIR) with TR of 8,500 ms, TE of 168 ms, and a matrix size of 320 × 256, with other parameters consistent with T2WI. 3) Axial T1-weighted imaging (T1WI) with TR of 2,000 ms, TE of 24 ms, matrix size of 320 × 224, and other parameters consistent with T2WI. 4) Axial contrast-enhanced T1-weighted imaging (T1WI+C) with TR of 1,725 ms, TE of 24 ms, matrix size of 320 × 256, and other parameters consistent with T2WI. Patients were instructed to minimize head movement during scanning to reduce motion artifacts. The most informative anatomical multi-contrast MRI sequences were the contrast-enhanced T1-weighted images and T2-

weighted FLAIR images, which were analyzed for further research.

It is important to note that radiomics features were extracted from 3.0T MRI scans, while traditional imaging characteristics were evaluated from 1.0T scans (described in Section 2.5). This difference in field strength represents a significant limitation of our study design. Although intensity standardization using histogram matching was applied to mitigate potential bias from mixed field strengths, inherent differences in image quality, signal-to-noise ratio, and spatial resolution between 3.0T and 1.0T scanners may have introduced systematic biases that could affect both feature extraction consistency and the validity of direct performance comparisons between radiomics and traditional imaging models. This limitation should be carefully considered when interpreting our comparative results.

Two radiologists, each with over three years of experience, independently performed manual segmentation of the glioma regions on the T1WI+C images for each patient, resulting in three-dimensional regions of interest (3D ROIs). These segmentations were subsequently reviewed by a senior radiologist. If the senior radiologist identified errors in the ROI delineation or found excessive discrepancies, the segmentations were revised. The final ROIs utilized were the consensus areas identified by the two radiologists. Inter-observer agreement was assessed on a subset of 30 randomly selected cases, yielding a Dice coefficient of 0.86 and an intraclass correlation coefficient (ICC) of 0.89, indicating excellent reproducibility.

2.4 Radiomics feature analysis

For the evaluation of radiological features, all axial, coronal, and sagittal slices were delineated, and features were extracted using PyRadiomics (version 3.0.1). Prior to feature extraction, all images were resampled to isotropic voxel spacing of 1 mm^3 and intensity normalized using z-score normalization following the Nyúl method to minimize inter-scanner variability. PyRadiomics was an open-source Python package designed to automatically extract quantitative features from medical images. In this study, a total of 1,269 features were initially extracted from axial, coronal, and sagittal images, including shape features (tumor size, sphericity), first-order statistics (standard deviation, mean value), and texture features (Gray-Level Co-occurrence Matrix [GLCM] contrast, Gray-Level Run Length Matrix [GLRLM] Gray-Level Non-Uniformity, Gray-Level Size Zone Matrix [GLSZM] Gray-Level Non-Uniformity and Size Zone Non-Uniformity).

Feature selection was performed through a three-step process: (1) Pearson correlation analysis to remove redundant features ($r > 0.9$), (2) univariate analysis to select features with $p < 0.05$, and (3) LASSO regression with optimal lambda ($\lambda = 0.021$) determined by 10-fold cross-validation, resulting in 18 final features (detailed in Supplementary Table S1). All model training and testing employed stratified 10-fold cross-validation with averaged AUC reported, and hyperparameters were optimized using grid search.

2.5 Traditional imaging feature acquisition

All patients underwent non-contrast CT, non-contrast MRI, and contrast-enhanced MRI examinations within two weeks prior to surgery. The Siemens 64-slice spiral CT scan used the orbitomeatal line as a baseline, covering the region from this line to the skull vertex. The CT had a slice thickness of 0.625 mm, tube voltage of 140 kV, tube current of 380 mA, pitch of 0.8, a gantry rotation speed of 0.5 seconds per rotation, and a matrix size of 512 × 512.

For traditional imaging feature evaluation, a Siemens Harmony 1.0T machine equipped with a standard 4-channel coil was used. The use of different field strengths for radiomics feature extraction (3.0T) and traditional imaging evaluation (1.0T) represents a critical methodological limitation. While histogram matching was employed to standardize intensity across different scanners, fundamental differences in image characteristics between 1.0T and 3.0T systems—including signal-to-noise ratio, contrast resolution, and susceptibility artifacts—cannot be fully compensated by post-processing techniques. This heterogeneity in imaging acquisition may have systematically affected feature values and model performance, potentially biasing the comparison between radiomics and traditional imaging approaches. Future studies should employ uniform field strength across all imaging modalities to ensure valid comparative analysis. Scanning methods included axial, sagittal, and coronal T1-weighted (T1WI) and T2-weighted (T2WI) scans. The parameters for these conventional sequences were as follows: T1WI utilized a turbo spin echo (TSE) with TR/TE of 550 ms/14 ms, field of view (FOV) of 230 mm × 230 mm, matrix size of 256 × 256, and bandwidth of 40.5 MHz/Px. T2WI employed TSE with TR/TE of 4,000 ms/96 ms and identical FOV, matrix, and bandwidth settings as T1WI. The slice thickness was set at 5 mm with a slice gap of 0.3 mm. Following the non-contrast scan, an enhanced scan was performed using a dose of 0.2 mmol/kg body weight of Gd-DTPA administered intravenously. Subsequent axial, sagittal,

and coronal T1WI scans were conducted with the same parameters as the non-contrast scan.

2.6 Traditional imaging feature analysis

Two experienced physicians evaluated all images using a 3D slicer to assess the tumor's imaging characteristics and manually segmented the region of interest (ROI) on each layer of the T images. Traditional imaging features considered included maximum diameter, borders, intratumoral bleeding, peritumoral edema, signal enhancement level (categorized as no/light, moderate, or severe based on standardized visual criteria), calcification, and cystic changes. The physicians performed these evaluations without knowledge of the glioma grading of the patients. Inter-rater reliability for categorical features was assessed using Cohen's kappa, yielding values ranging from 0.72 to 0.88.

2.7 Pathological examination

Following surgery, and with the patient's consent, the resected tumor tissue was fixed in formalin, embedded in paraffin blocks, and sectioned to yield several tissue samples from each block. These sections were stained with hematoxylin and eosin for evaluation under optical microscopy. The pathological features assessed included maximum diameter, borders, uniform distribution of cancer cells, microcalcifications, microcysts, and vascular development. Glioma grading was conducted according to the WHO criteria. Two experienced pathologists independently evaluated all slides in a blinded manner, and any discrepancies were resolved through discussion to reach a consensus. Additionally, immunohistochemical testing was performed on the tumor tissues to analyze the expression status of tumor molecular markers.

2.8 Tumor molecular marker measurement

Paraffin sections were processed using the NS40 automatic immunohistochemical staining instrument (Dakewe Biotech Co., Ltd., Shenzhen, China). The stained sections were subsequently analyzed with a Nikon microscope (Nikon, Japan) to determine the positivity rates of GFAP, Olig-2, IDH1/2, p53, Ki-67, and MGMT.

2.9 Statistical analysis

Data analysis was conducted using SPSS 29.0 statistical software (SPSS Inc., Chicago, IL, USA) and R software (version 4.2.0). Categorical data were presented as [n (%)], while continuous variables were expressed as ($X \pm s$). A p-value of less than 0.05 was considered

statistically significant. Univariate and multivariate logistic regression analyses were performed to identify relevant independent influencing factors, with odds ratios (OR) calculated for each. To address class imbalance (LGG:HGG ratio 1.1:1), class weights were set to 'balanced' in all models.

Diagnostic performance analysis was executed using xgbTree, a well-established Gradient Boosted Trees (GBT) method. XGBoost employed Classification and Regression Trees (CART) to train on both radiomics and traditional imaging features, thereby evaluating the diagnostic performance for glioma grading. Model comparisons were performed using the DeLong test with Bonferroni adjustment for multiple comparisons (3 pairwise tests, adjusted $\alpha = 0.017$).

3. Results

3.1 Demographic and basic data

A total of 263 patients including 137 patients with The LGG group and 126 patients with HGG group were included (Table 1). The gender distribution and parameters such as age, BMI, smoking status, marital status, educational level, family history of cancer, tumor location, cortical involvement, crossing the midline, hypertension, and diabetes showed no statistically significant differences between the groups ($p > 0.05$). Notably, the presence of cerebral edema was significantly higher in the HGG group at 32.54% compared to 17.52% in the LGG group ($p = 0.005$). Additionally, neurological deficits were more prevalent in the HGG group (19.05%) than in the LGG group (9.49%) with statistical significance ($p = 0.026$). Other clinical parameters such as epilepsy demonstrated a trend towards significance ($p = 0.053$), whereas infection rates did not differ significantly ($p = 0.111$). These findings underscore the distinct patterns of cerebral edema and neurological deficits in HGG compared to LGG, affirming their potential as key differentiators in glioma grading.

Table 1 Baseline characteristics.

Parameters	**LGG Group (n=137)**	**HGG Group (n=126)**	**t/χ^2**	**p**
Male/Female	72 (52.55%) / 65 (47.45%)	77 (61.11%) / 49 (38.89%)	1.957	0.162
Age (years)	45.34 ± 12.07	48.26 ± 13.18	1.874	0.062

BMI (kg/m)	21.25 ± 2.36	21.07 ± 2.19	0.641	0.522
Current Smoking(Yes/no)	27 (19.71%) / 110 (80.29%)	31 (24.6%) / 95 (75.4%)	0.915	0.339
Marital status(Married/ Others)	120 (87.59%) / 17 (12.41%)	109 (86.51%) / 17 (13.49%)	0.068	0.794
Educational level			0.334	0.846
Junior high school and below	60 (43.8%)	59 (46.83%)		
high school and vocational school	49 (35.77%)	41 (32.54%)		
College and above	28 (20.44%)	26 (20.63%)		
Clinical symptoms(Absent/Present)	11 (8.03%) / 126 (91.97%)	7 (5.56%) / 119 (94.44%)	0.63	0.427
Family history of cancer(Yes/no)	12 (8.76%) / 125 (91.24%)	15 (11.9%) / 111 (88.1%)	0.705	0.401
Tumor location			3.761	0.288
Frontal lobe	62 (45.26%)	60 (47.62%)		
Parietal lobe	55 (40.15%)	39 (30.95%)		
Temporal lobe	7 (5.11%)	12 (9.52%)		
Other locations	13 (9.49%)	15 (11.9%)		
Cortical involvement	111 (81.02%)	95 (75.4%)	1.223	0.269
Cross the midline	28 (20.44%)	35 (27.78%)	1.941	0.164
Hypertension	32 (23.36%)	34 (26.98%)	0.459	0.498
Diabetes	13 (9.49%)	11 (8.73%)	0.046	0.831
Cerebral edema	24 (17.52%)	41 (32.54%)	7.96	0.005
Epilepsy	32 (23.36%)	43 (34.13%)	3.734	0.053
Neurological deficits	13 (9.49%)	24 (19.05%)	4.96	0.026
Infection	7 (5.11%)	13 (10.32%)	2.534	0.111

BMI: Body Mass Index.

3.2 Tumor molecular marker

The presence of Glial Fibrillary Acidic Protein (GFAP) was significantly higher in the HGG group at 98.41% compared to 91.97% in the LGG group ($p = 0.016$) (Table 2). The expression of Oligodendrocyte Transcription Factor 2 (Olig-2) was more prevalent in the LGG group (76.64%) than in the HGG group (64.29%), showing statistical significance ($p = 0.028$). Notably, Isocitrate Dehydrogenase 1/2 (IDH1/2) mutations were significantly more frequent in the LGG group (38.69%) compared to the HGG group (25.4%) ($p = 0.021$). The p53 expression was higher in HGG at 28.57% compared to 15.33% in LGG ($p = 0.009$), and Ki-67 proliferation index was significantly elevated in the HGG group at 20.63% versus 8.76% in the LGG group ($p = 0.006$). The examination of O6-Methylguanine-DNA Methyltransferase (MGMT) showed no significant difference between the groups ($p = 0.129$), indicating limited discriminative power of this marker for glioma grading. These molecular distinctions emphasize the heterogeneity between low-grade and HGG and highlight specific markers that may be pivotal in diagnostic stratification.

Table 2 Comparison of Tumor molecular marker between the two groups.

Variable	LGG Group (n=137)	HGG Group (n=126)	χ^2	p
GFAP	126 (91.97%)	124 (98.41%)	5.797	0.016
Olig-2	105 (76.64%)	81 (64.29%)	4.84	0.028
IDH1/2	53 (38.69%)	32 (25.4%)	5.299	0.021
p53	21 (15.33%)	36 (28.57%)	6.781	0.009
Ki-67	12 (8.76%)	26 (20.63%)	7.489	0.006
MGMT	57 (41.61%)	41 (32.54%)	2.308	0.129

GFAP: Glial Fibrillary Acidic Protein; Olig-2: Oligodendrocyte Transcription Factor 2; IDH1/2: Isocitrate Dehydrogenase 1/2; MGMT: O6-Methylguanine-DNA Methyltransferase.

3.3 Pathological manifestations

The maximum tumor diameter was notably larger in the HGG group with a mean of 5.08 ± 1.37 cm compared to 4.45 ± 1.06 cm in the LGG group, indicating a significant difference ($p < 0.001$) (Table 3). When evaluating tumor borders, the HGG group had a higher prevalence of blurred borders (61.9% vs 45.26%), which was statistically significant ($p = 0.007$). The

distribution of cancer cells was more uniform in the LGG group at 54.01% compared to 36.51% in the HGG group, with a significant difference (p = 0.004). Microcalcifications were significantly more common in the HGG group (50.79%) than in the LGG group (35.77%) (p = 0.014). Additionally, vascular development was more frequently abnormal in the HGG group (61.11%) compared to the LGG group (44.53%), with a significant p value of 0.007. These pathological differences underscore distinct morphological characteristics between low-grade and HGG that may inform diagnostic approaches and therapeutic strategies.

Table 3 Comparison of Pathological manifestations between the two groups.

Variable	**LGG Group (n=137)**	**HGG Group (n=126)**	χ^2	**p**
Maximum diameter	4.45 ± 1.06	5.08 ± 1.37	4.074	< 0.001
Borders			7.308	0.007
Clear	75 (54.74%)	48 (38.1%)		
Blurred	62 (45.26%)	78 (61.9%)		
Uniform distribution of cancer cells	74 (54.01%)	46 (36.51%)	8.108	0.004
microcalcifications	49 (35.77%)	64 (50.79%)	6.048	0.014
Vascular development			7.24	0.007
Normal	76 (55.47%)	49 (38.89%)		
Abnormal	61 (44.53%)	77 (61.11%)		

3.4 Radiomics features

Tumor diameter and volume were notably greater in the HGG group, with mean diameters of 5.21 ± 1.62 cm and volumes of 43.38 ± 16.09 cm^3, compared to 4.56 ± 1.84 cm and 37.57 ± 15.06 cm^3 in the LGG group, respectively (p = 0.003 for both) (Figure 1). Sphericity was significantly lower in the HGG group (0.62 ± 0.32) compared to the LGG group (0.73 ± 0.34), indicating less spherical tumors (p = 0.010). The GLCM Contrast was higher in the HGG group (207.36 ± 47.39) versus the LGG group (192.25 ± 31.08), showing a significant difference (p = 0.003). The Gray-Level Run Length Matrix (GLRLM) Gray-Level Non-

Uniformity (GLN) was also higher in the HGG group, with a mean of 77.38 ± 12.07, compared to 73.18 ± 12.24 in the LGG group ($p = 0.005$). While the Gray-Level Size Zone Matrix (GLSZM) GLN and SZN showed no statistically significant differences ($p > 0.05$), the standard deviation of intensity values was significantly higher in the HGG group (24.31 ± 10.75) than in the LGG group (21.52 ± 4.07) ($p = 0.007$). The mean intensity values did not differ significantly between groups ($p = 0.316$). These findings highlight the enhanced complexity and heterogeneity of radiomic profiles in HGG. Feature importance analysis using SHAP (SHapley Additive exPlanations) values revealed that GLCM Contrast and standard deviation were the most influential features for HGG prediction (Supplementary Figure S1).

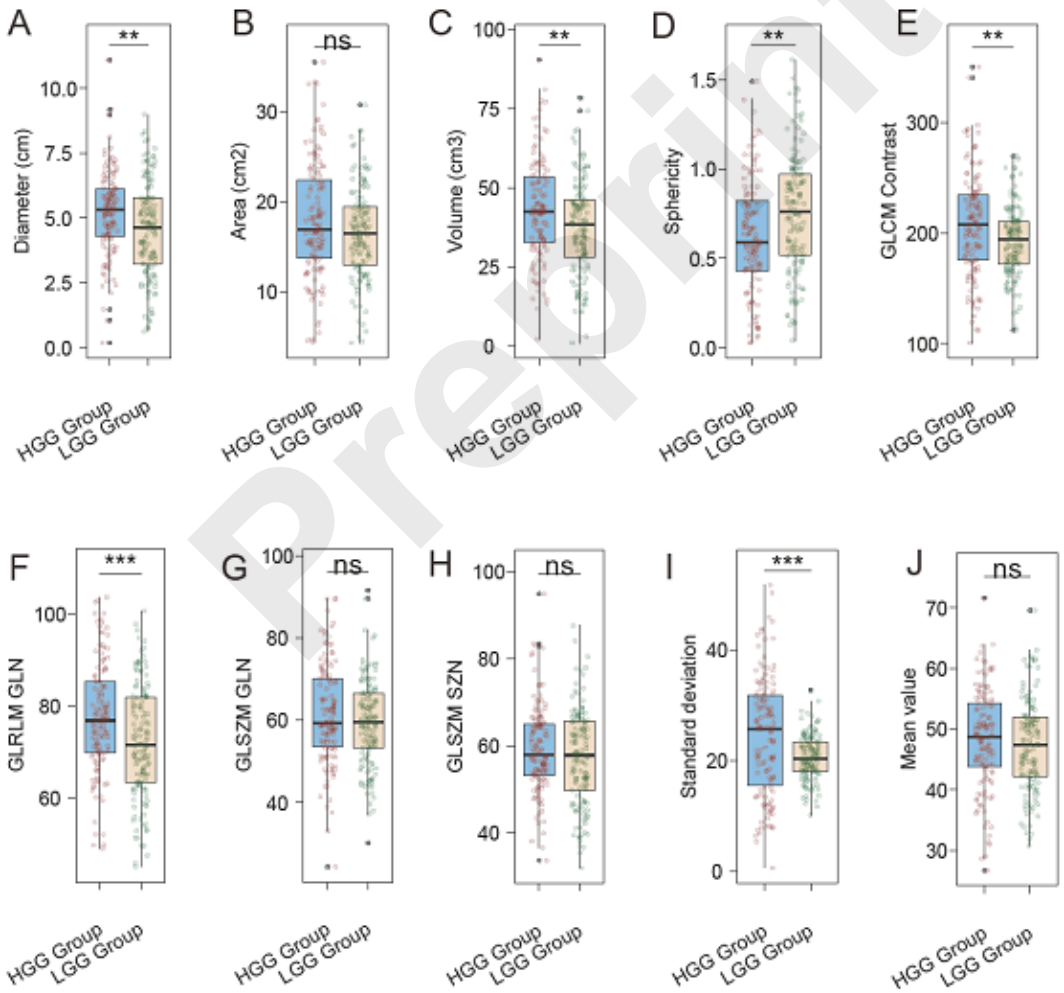


Figure 1 Comparison of Radiomics features between the two groups

A: Diameter; B: Area; C: Volume; D: Sphericity; E: GLCM Contrast; F: GLRLM GLN; G: GLSZM GLN; H: GLSZM SZN; I: Standard deviation; J: Mean value; ns: No significant difference; **: $p<0.01$; ***: $p<0.001$.

3.5 Traditional imaging features

Maximum tumor diameter was larger in the HGG group (5.22 ± 1.36 cm) compared to the LGG group (4.73 ± 1.32 cm) ($p = 0.003$) (Table 4). Tumor borders were more frequently indistinct in the HGG group with 76.98%, versus 88.32% in the LGG group being indistinct, which was statistically significant ($p = 0.015$). Intratumoral bleeding was less common in the HGG group (36.51%) compared to the LGG group (51.82%) ($p = 0.013$). Peritumoral edema was significantly more prevalent in the LGG group with 74.45% displaying edema compared to 60.32% in the HGG group ($p = 0.014$). The enhancement level was notably severe in 61.9% of the HGG cases, a significant increase from 12.41% in the LGG group ($p < 0.001$). Calcifications were observed more frequently in the HGG group (46.03%) compared to the LGG group (29.93%), with a significant p value of 0.007. Cystic changes showed no significant difference between the groups ($p = 0.103$). These findings demonstrate distinctive imaging characteristics, particularly in enhancement patterns and calcification presence, that differentiate high-grade from LGG.

Table 4 Comparison of Traditional imaging features between the two groups

Variable	**LGG Group (n=137)**	**HGG Group (n=126)**	**χ^2**	**p**
Max diameter (cm)	4.73 ± 1.32	5.22 ± 1.36	2.983	0.003
Borders (indistinct/clear)	121 (88.32%) / 16 (11.68%)	97 (76.98%) / 29 (23.02%)	5.948	0.015
Intratumoral bleeding (yes/no)	71 (51.82%) / 66 (48.18%)	46 (36.51%) / 80 (63.49%)	6.235	0.013
Peritumoral edema (with/without or mild)	102 (74.45%) / 35 (25.55%)	76 (60.32%) / 50 (39.68%)	5.995	0.014

The enhancement level			73.712	< 0.001
No or light	75 (54.74%)	21 (16.67%)		
Moderate	45 (32.85%)	27 (21.43%)		
Severe	17 (12.41%)	78 (61.9%)		
Calcification	41 (29.93%)	58 (46.03%)	7.252	0.007
Cystic change	39 (28.47%)	25 (19.84%)	2.652	0.103

3.6 Logistic regression analysis

The univariate logistic regression analysis identified several significant factors associated with HGG (Table 5). Larger tumor diameter was a significant predictor of HGG, with a coefficient of 0.216 and an odds ratio (OR) of 1.241 (95% CI, 1.077-1.440, p = 0.003). Tumor volume also showed a significant association (OR 1.024, 95% CI, 1.008-1.042, p = 0.003). Sphericity was inversely related to HGG, suggesting that less spherical tumors were more likely to be high grade (coefficient -1.209, OR 0.298, 95% CI, 0.137-0.630, p = 0.002). Texture features such as GLCM Contrast (OR 1.010, 95% CI, 1.003-1.016, p = 0.003) and GLRLM GLN (OR 1.036, 95% CI, 1.015-1.058, p < 0.001) were significantly associated with HGG. Similarly, a higher standard deviation in radiomic intensity was predictive of high-grade status (OR 1.062, 95% CI, 1.029-1.099, p < 0.001). Among traditional imaging features, maximum diameter (OR 1.320, 95% CI, 1.098-1.602, p = 0.004) and calcification presence (OR 1.997, 95% CI, 1.207-3.330, p = 0.007) were significant, while indistinct borders (OR 0.442, 95% CI, 0.223-0.851, p = 0.016), intratumoral bleeding (OR 0.535, 95% CI, 0.325-0.873, p = 0.013), and peritumoral edema (OR 0.522, 95% CI, 0.307-0.878, p = 0.015) were inversely associated with HGG. The enhancement level notably showed a strong association with HGG, with severe enhancement substantially increasing the odds (OR 4.038, 95% CI, 2.862-5.845, p < 0.001). These findings underscore the potential of both radiomics and traditional imaging characteristics in distinguishing glioma grades.

Table 5: Univariate logistic regression analysis of each index and HGG

influencing factors	**Coefficient**	**Std Error**	**Wald**	**p**	**OR**	**95%CI**

Diameter (cm)	0.216	0.074	2.928	0.003	1.241	1.077-1.440
Volume (cm3)	0.024	0.008	2.933	0.003	1.024	1.008-1.042
Sphericity	-1.209	0.388	3.120	0.002	0.298	0.137-0.630
GLCM Contrast	0.010	0.003	2.975	0.003	1.010	1.003-1.016
GLRLM GLN	0.035	0.011	3.335	<0.001	1.036	1.015-1.058
Standard deviation	0.06	0.017	3.620	<0.001	1.062	1.029-1.099
Max diameter (cm)	0.278	0.096	2.891	0.004	1.320	1.098-1.602
Borders (indistinct/clear)	-0.816	0.340	2.40	0.016	0.442	0.223-0.851
Intratumoral bleeding (yes/no)	-0.626	0.252	2.486	0.013	0.535	0.325-0.873
Peritumoral edema (with/without or mild)	-0.651	0.267	2.434	0.015	0.522	0.307-0.878
The enhancement level(No or light/Moderate/Severe)	1.396	0.182	7.685	<0.001	4.038	2.862-5.845
Calcification	0.692	0.258	2.677	0.007	1.997	1.207-3.330

The multivariate logistic regression analysis identified several independent risk factors of HGG (Table 6). Tumor diameter remained a significant factor, with a coefficient of 0.256 and an odds ratio (OR) of 1.292 (95% CI, 1.066-1.565, p = 0.009), and tumor volume also continued to show significance (OR 1.028, 95% CI, 1.007-1.051, p = 0.010). Although sphericity approached statistical significance, it did not achieve it (p = 0.065). Among texture

features, GLRLM GLN was significantly associated with HGG (OR 1.041, 95% CI, 1.013-1.070, p = 0.004), and standard deviation was a strong predictor with an OR of 1.075 (95% CI, 1.028-1.124, p = 0.001). Maximum diameter, indistinct borders, intratumoral bleeding, and peritumoral edema did not reach statistical significance in the multivariate model. The most substantial predictor was the level of enhancement, where severe enhancement increased the odds of being a HGG nearly fourfold (OR 3.954, 95% CI, 2.595-6.024, p < 0.001). Calcification did not show a significant association with HGG (p = 0.457). These results highlight that specific radiomic features, particularly enhancement level and standard deviation, along with traditional measurements like diameter, provide valuable insights into the differentiation of glioma grades.

Table 6 Multivariate logistic regression analysis of each index and HGG

influencing factors	Coefficient	Std Error	Wald Stat	p	OR	OR CI Lower	OR CI Upper
Diameter (cm)	0.256	0.098	2.617	0.009	1.292	1.066	1.565
Volume (cm3)	0.028	0.011	2.581	0.010	1.028	1.007	1.051
Sphericity	-0.959	0.520	-1.845	0.065	0.383	0.138	1.062
GLCM Contrast	0.006	0.004	1.252	0.211	1.006	0.997	1.014
GLRLM GLN	0.040	0.014	2.857	0.004	1.041	1.013	1.070
Standard deviation	0.073	0.023	3.182	0.001	1.075	1.028	1.124
Max diameter (cm)	0.220	0.124	1.768	0.077	1.246	0.976	1.590
Borders (indistinct/clear)	-0.825	0.438	-1.884	0.060	0.438	0.186	1.034
Intratumoral bleeding (yes/no)	-0.587	0.348	-1.687	0.092	0.556	0.281	1.100
Peritumoral edema (with/without or mild)	-0.507	0.362	-1.400	0.161	0.602	0.296	1.225
The enhancement level(No or light/Moderate/Severe)	1.375	0.215	6.397	<0.001	3.954	2.595	6.024
Calcification	0.257	0.346	0.744	0.457	1.293	0.657	2.548

3.7 ROC Curve

The analysis of radiomics ROC data for the diagnostic grading of gliomas revealed varying

levels of diagnostic performance across different radiomic features (Table 7). The diameter, with a best threshold of 4.09 cm, showed a sensitivity of 0.794 and a specificity of 0.423, yielding an area under the curve (AUC) of 0.607 and a Youden index of 0.217, indicating moderate discriminative ability and an F1 score of 0.656. Volume, at a threshold of 41.26 cm^3, demonstrated a lower sensitivity of 0.587 and specificity of 0.569, with an AUC of 0.594, suggesting limited classification accuracy (Youden index of 0.156, F1 score of 0.571). Sphericity, with a threshold of 0.715, had sensitivity and specificity values of 0.643 and 0.555, respectively, resulting in an AUC of 0.611, a Youden index of 0.198, and an F1 score of 0.364. GLCM Contrast, at a best threshold of 213.105, achieved specificity of 0.781 but lower sensitivity of 0.468, and an AUC of 0.605 (Youden index of 0.249, F1 score of 0.549). GLRLM GLN, with a threshold of 69.535, showed a sensitivity of 0.77 and specificity of 0.453, with an AUC of 0.617, offering a better balance as reflected by its Youden index of 0.223 and F1 score of 0.651. Standard deviation, with a threshold of 25.595, exhibited high specificity of 0.92 and a sensitivity of 0.508, resulting in an AUC of 0.618 and the highest Youden index of 0.428, alongside an F1 score of 0.637. Collectively, these radiomic features reflect varying levels of utility in glioma grading.

Subsequently, this study integrated radiomics features to develop a combined predictive model for HGG, achieving a bootstrap-corrected AUC of 0.847 (95% CI: 0.80-0.89) based on 10-fold cross-validation (mean AUC ± SD: 0.832 ± 0.028, detailed in Supplementary Table S2). This result indicates that radiomics possesses substantial predictive value for glioma grading (Fig. 2).

Table 7 Radiomics ROC data

Variable	**Bes threshold**	**Sensitivities**	**Specificities**	**AUC**	**Youden index**	**F1_score**
Diameter (cm)	4.09	0.794	0.423	0.607	0.217	0.656
Volume (cm3)	41.26	0.587	0.569	0.594	0.156	0.571
Sphericity	0.715	0.643	0.555	0.611	0.198	0.364
GLCM Contrast	213.105	0.468	0.781	0.605	0.249	0.549
GLRLM GLN	69.535	0.77	0.453	0.617	0.223	0.651
Standard deviation	25.595	0.508	0.92	0.618	0.428	0.637

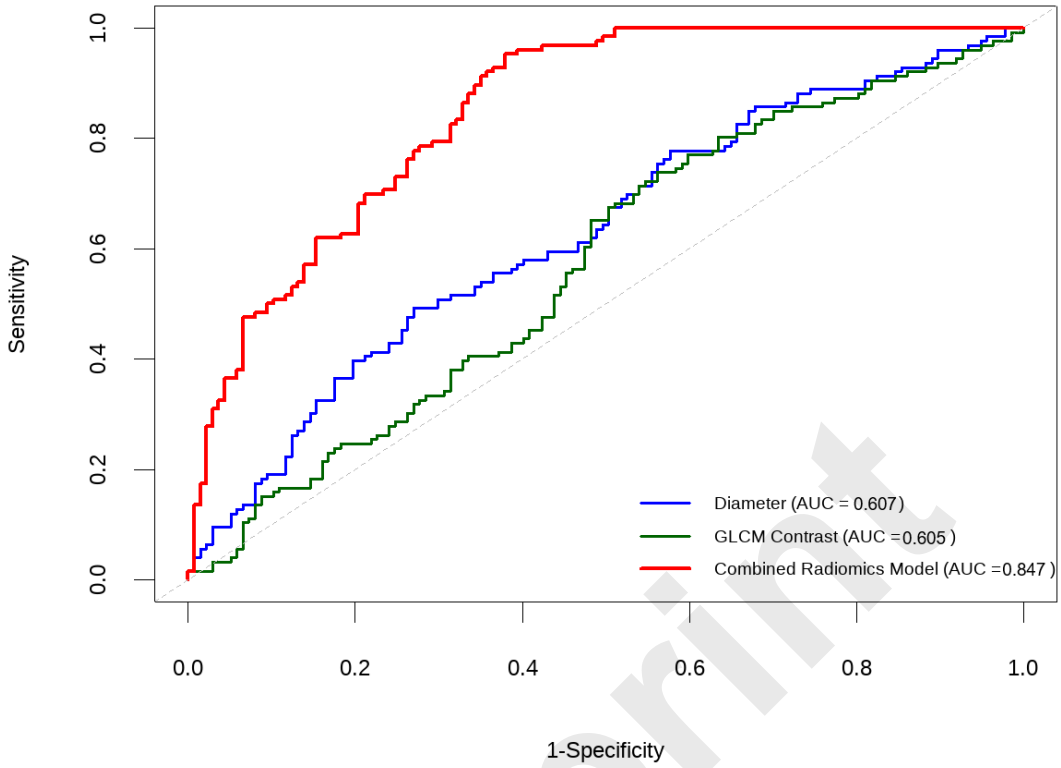


Figure 2: The radiomics-based combined predictive model for HGG (10-fold CV & Bootstrap adjusted).

The analysis of traditional imaging ROC data illustrates varied efficacy of different imaging features for the diagnostic grading of gliomas (Table 8). The maximum diameter, with a best threshold of 4.945 cm, demonstrated a sensitivity of 0.595 and specificity of 0.569, resulting in an area under the curve (AUC) of 0.596 and a Youden index of 0.164, indicating moderate discriminative capacity, and an F1 score of 0.577. Intratumoral bleeding, with a threshold of 0.5, achieved a sensitivity of 0.635 and specificity of 0.518, yielding an AUC of 0.577, a lower Youden index of 0.153, and an F1 score of 0.379, reflecting limited differentiation. The enhancement level, assessed on a scale transitioning from no or light to severe (threshold 1.5), exhibited high specificity (0.876) alongside a sensitivity of 0.619, producing a robust AUC of 0.779 and the highest Youden index of 0.495 among the variables assessed, with an F1 score of 0.706, indicating strong predictive value for HGG. Calcification, with a threshold of 0.5,

showed a sensitivity of 0.46 and specificity of 0.701, reflecting an AUC of 0.581 and a Youden index of 0.161, along with an F1 score of 0.516, suggesting moderate utility. Subsequently, this study integrated traditional imaging features to construct a combined predictive model for HGG, achieving a bootstrap-corrected AUC of 0.806 (95% CI: 0.76-0.85) based on 10-fold cross-validation (mean AUC ± SD: 0.806 ± 0.031). This result indicates that traditional imaging has significant predictive value for glioma grading (Fig. 3). Therefore, the predictive value of radiomics was higher than that of traditional imaging alone.

Table 8 Traditional imaging ROC data

Variable	**Best threshold**	**Sensitivities**	**Specificities**	**AUC**	**Youden index**	**F1 score**
Max diameter (cm)	4.945	0.595	0.569	0.596	0.164	0.577
Intratumoral bleeding (yes/no)	0.5	0.635	0.518	0.577	0.153	0.379
The enhancement level(No or light/Moderate/Severe,No or light:0)	1.5	0.619	0.876	0.779	0.495	0.706
Calcification	0.5	0.46	0.701	0.581	0.161	0.516

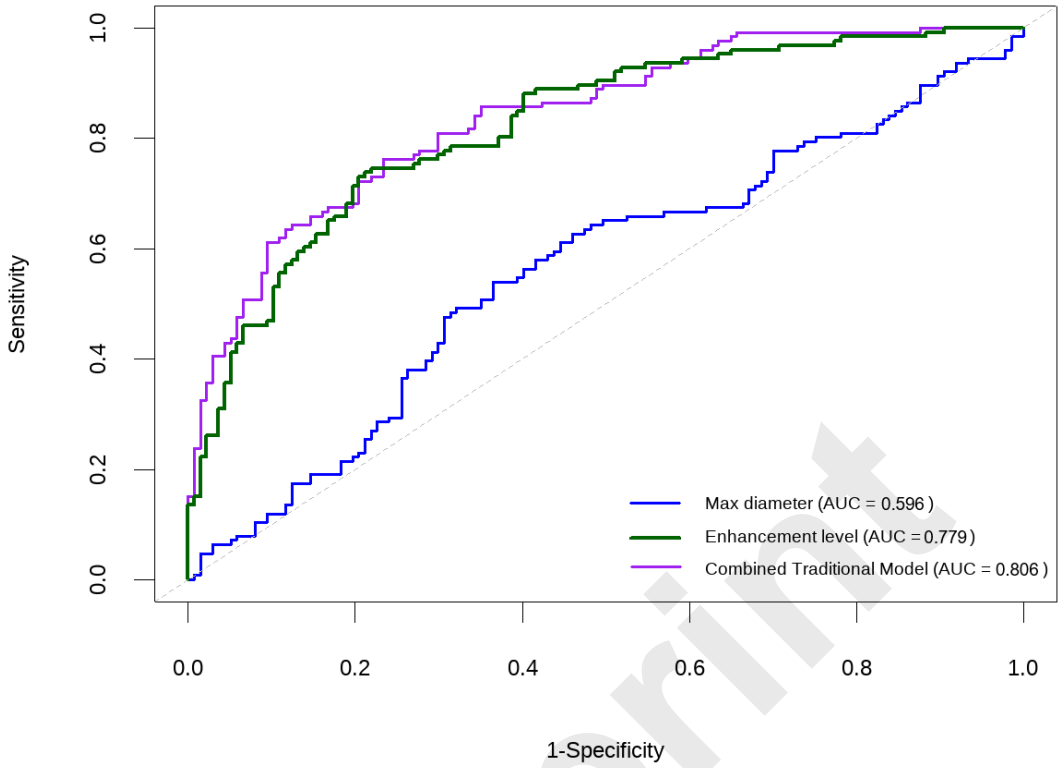


Figure3: Traditional imaging -based combined predictive model for HGG

3.8 Diagnostic performance

The comparative analysis of diagnostic performance between radiomics-based models, traditional imaging characteristics, and their combination highlights distinct patterns with important clinical implications (Table 9). The radiomics model alone achieved a bootstrap-corrected area under the curve (AUC) of 0.847 (95% CI, 0.80-0.89) with an accuracy of 69.1%, sensitivity of 82.0%, and specificity of 38.7%. It showed a positive predictive value (PPV) of 68.3% and a negative predictive value (NPV) of 58.1%. The traditional imaging model demonstrated an AUC of 0.806 (95% CI, 0.76-0.85), attaining an accuracy of 65.4%, sensitivity of 80.0%, and specificity of 35.4%, alongside a PPV of 67.4% and an NPV of 56.8%. The combination of radiomics and traditional imaging demonstrated an AUC of 0.872 (95% CI, 0.83-0.91) and an accuracy of 80.4% based on 10-fold cross-validation (mean AUC ± SD: 0.847 ± 0.026). This integrated approach yielded a sensitivity of 84.0% and enhanced specificity of 44.8%, with a PPV of 71.8% and an improved NPV of 63.1%.

However, a critical limitation across all models is their low specificity ranging from 35.4% to 44.8%, which translates to false-positive rates of 55.2% to 64.6%. This means that more than half of the low-grade gliomas could potentially be misclassified as high-grade tumors. The relatively low specificity is partly attributable to the class imbalance (LGG:HGG ratio 1.1:1) and the balanced class weight adjustment implemented to optimize sensitivity for detecting aggressive HGGs, which is clinically prioritized given the serious consequences of missing high-grade diagnoses. Nevertheless, this trade-off results in a substantial risk of overdiagnosis, which could lead to unnecessary aggressive treatments for patients with low-grade tumors. The high false-positive rate (55.2-64.6%) is a serious concern that must be carefully considered in clinical decision-making, and these models should be used in conjunction with other diagnostic methods including histopathological confirmation rather than as standalone diagnostic tools (see Supplementary Table S3 for detailed stratified analysis).

Table 9 Diagnostic performance between the two models

Model	**AUC**	**95%CI**	**Accuracy**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Radiomics	0.847	0.80-0.89	0.691	0.820	0.387	0.683	0.581
Traditional imaging	0.806	0.76-0.85	0.654	0.800	0.354	0.674	0.568
Radiomics+ Traditional imaging	0.872	0.83-0.91	0.804	0.840	0.448	0.718	0.631

The DeLong test with Bonferroni adjustment was used to compare the diagnostic tests involving Radiomics and Radiomics + Traditional Imaging. The adjusted p-value for the comparison between Radiomics and Radiomics + Traditional Imaging was 0.18, indicating that there was no significant difference in the AUCs ($p > 0.017$) (Table 10). This finding suggests that although the combined model shows numerically higher performance metrics, the improvement over radiomics alone does not reach statistical significance after correction for multiple comparisons. Therefore, claims of "significantly enhanced" or "superior" diagnostic performance of the combined model relative to radiomics alone are not statistically supported by our data. For the diagnostic tests comparing Traditional Imaging with Radiomics and Traditional Imaging with Radiomics + Traditional Imaging, the AUCs were found to be significantly different, with adjusted p-values of 0.036 and 0.012, respectively. These results

indicate that the diagnostic performance of both the radiomics model and the combined model was significantly better than that of traditional imaging alone ($p < 0.017$).

Table 10 DeLong Test with Bonferroni Adjustment.

	Radiomics	Traditional imaging	Radiomics+ Traditional imaging
Radiomics		0.036*	0.18
Traditional imaging	0.036*		0.012*
Radiomics+ Traditional imaging	0.18	0.012*	

*Significant after Bonferroni correction (adjusted $\alpha = 0.017$)

4. Discussion

Radiomics, as a field, hinges on the extraction of a plethora of quantitative features from medical images that elicit intricate patterns not easily discernible to the human eye(15, 16). This study has illuminated several instances where radiomic features such as tumor sphericity, GLCM contrast, and Gray-Level Run Length Matrix (GLRLM) GLN exhibit pronounced distinctions between low-grade and HGG. The enhanced analytic granularity offered by radiomic analysis appears to capture tumor heterogeneity and complexity more precisely than traditional imaging modalities, which rely more heavily on qualitative and semi-quantitative assessments(17, 18).

The significantly higher GLCM Contrast values observed in HGG (207.36 ± 47.39 vs 192.25 ± 31.08, $p = 0.003$) reflect increased local intensity variations within the tumor tissue. This finding aligns with the known pathological heterogeneity of high-grade gliomas, which typically exhibit areas of necrosis, hemorrhage, and microvascular proliferation. GLCM Contrast specifically quantifies the difference in gray-level intensities between neighboring pixels, with higher values indicating greater heterogeneity. This heterogeneity arises from the chaotic vascular architecture and irregular cell density patterns characteristic of aggressive tumors. Recent studies have demonstrated similar associations between texture heterogeneity and tumor aggressiveness in various cancer types(19, 20). The biological basis for this correlation lies in the rapid and disorganized growth patterns of high-grade tumors, which

create distinct microenvironments with varying oxygen levels, cell densities, and metabolic activities---all of which manifest as increased textural complexity on imaging.

One of the pivotal distinctions noted was in the standard deviation of the intensity values among the tumor samples, which was significantly higher in HGG. This finding aligns with the biological understanding that high-grade tumors possess heightened structural heterogeneity due to factors such as necrosis, microvascular proliferation, and aggressive cellular activity. Such heterogeneity was inherently captured through the texture features extracted via radiomics, but may remain partially elusive on traditional imaging where heterogeneity was assessed through subjective measures(21, 22).

Despite these strengths, it was critical to recognize the intricacies involved in incorporating radiomics into standard diagnostic protocols. The effectiveness of such methods leans significantly on the quality of image acquisition, robust feature extraction processes, and subsequent machine learning algorithms capable of discerning these minute differences. The diagnostic prowess of radiomics was decidedly contingent upon these components, making standardization and quality control pivotal(23, 24). Our study deployed a robust imaging protocol and cross-validation using PyRadiomics, which ensured optimal feature extraction. However, a significant methodological limitation in our study is the use of different MRI field strengths for radiomics feature extraction (3.0T) versus traditional imaging evaluation (1.0T). This heterogeneity in imaging acquisition introduces systematic biases that may affect feature values, model performance, and the validity of direct comparisons between the two approaches. While we attempted to mitigate this through histogram matching for intensity standardization, fundamental differences in image characteristics between field strengths—including signal-to-noise ratio, contrast resolution, and susceptibility artifacts—cannot be fully compensated by post-processing techniques alone. This limitation undermines the robustness of our comparative analysis and highlights the critical need for future studies to employ uniform imaging protocols across all modalities to enable valid performance comparisons.

On the other hand, traditional imaging features analyzed in this study, including tumor borders, intratumoral bleeding, and peritumoral edema, have long been entrenched in glioma grading practices. The clarity of tumor borders, or lack thereof, often implicates invasion into

adjacent brain tissue---a hallmark of higher-grade malignancies. Though these imaging characteristics were indeed salient, their evaluation was inherently subjective and heavily reliant on the experience of the radiologist. This subjectivity could account for the slightly lower specificity and sensitivity observed in traditional imaging when distinguished solely from radiomics in glioma grading(25, 26).

The integration of radiomics with traditional imaging appears to offer complementary information wherein the quantitative strengths of radiomics and the clinical interpretability of traditional imaging can be combined(27, 28). However, our statistical analysis reveals important nuances in this relationship. While the combined model achieved numerically higher performance metrics (AUC 0.872) compared to radiomics alone (AUC 0.847), the DeLong test with Bonferroni correction demonstrated no statistically significant difference between these two approaches (adjusted $p = 0.18$). This finding indicates that the apparent improvement from adding traditional imaging features to radiomics may not be robust enough to justify the additional complexity in clinical implementation. Both the radiomics model and combined model did, however, demonstrate significantly superior performance compared to traditional imaging alone (adjusted $p = 0.036$ and 0.012, respectively), confirming the value of quantitative radiomic analysis over conventional qualitative assessment.

A critical limitation that must be emphasized is the low specificity observed across all models in this study (38.7% for radiomics, 35.4% for traditional imaging, and 44.8% for the combined approach). These values correspond to false-positive rates of 61.3%, 64.6%, and 55.2%, respectively, meaning that more than half of low-grade gliomas could be misclassified as high-grade tumors. This represents a serious clinical concern, as such misclassification could lead to overtreatment of patients with indolent low-grade tumors, subjecting them to unnecessarily aggressive therapeutic interventions with significant morbidity. The low specificity is partly attributable to the class imbalance in our dataset (LGG:HGG ratio 1.1:1) and our deliberate use of balanced class weights to prioritize sensitivity for detecting high-grade tumors, given the critical importance of not missing aggressive malignancies. However, this trade-off comes at a substantial cost in terms of false-positive diagnoses. The high false-positive rate underscores that these radiomics models, despite their promising sensitivity, should not be used as standalone diagnostic tools but rather as complementary information to

be integrated with histopathological confirmation and other clinical data in a comprehensive diagnostic framework.

Recent advances have further expanded our understanding of glioma imaging biomarkers. Zhang et al. demonstrated that MRI radiogenomic signatures can effectively characterize transcriptional heterogeneity in glioblastoma, identifying metabolic pathway subclones and spinocerebellar ataxia subclones as independent risk factors for overall survival (29). Their study of 99 patients revealed that 19 radiomic features could be linked to hallmark cancer pathways and malignant behaviors, achieving a hazard ratio of 1.67 for overall survival prediction. This aligns with our findings that texture-based features like GLCM Contrast capture underlying tumor biology. Additionally, the role of tumor microenvironment factors has gained increasing attention. Vageli et al.'s systematic review highlighted that hypoxia-inducible factor 1alpha (HIF-1α) and vascular endothelial growth factor (VEGF) expression patterns in glioblastoma correlate with oncogenic factors and unfavorable survival, while also reflecting treatment efficacy (30). These molecular imaging correlations suggest that future radiomics models could benefit from incorporating hypoxia-related imaging signatures.

The pathophysiology underlying gliomas intrinsically influences the imaging phenotypes, which radiomics can delineate with its high-dimensional feature sets(31, 32). For instance, the higher calcification rates observed in HGG align with prolonged tumor development and metabolic shifts that foster calcification deposition. Radiomics' ability to quantify such subtle differences offers a window into understanding underlying cancer biology, potentially advancing our grasp on glioma pathogenesis(33).

4.1 Study Limitations and Future Directions

This study has several important limitations that substantially affect the interpretation and generalizability of our findings. First and most critically, the use of different MRI field strengths for radiomics feature extraction (3.0T) and traditional imaging evaluation (1.0T) represents a fundamental methodological flaw that introduces systematic bias and compromises the validity of performance comparisons between these two approaches. Despite our efforts to mitigate this through histogram matching for intensity standardization, fundamental differences in image characteristics between 1.0T and 3.0T systems—including signal-to-noise ratio, contrast resolution, spatial resolution, and susceptibility artifacts—

cannot be fully compensated by post-processing techniques. This heterogeneity in imaging acquisition likely affected feature extraction consistency and may have systematically biased the relative performance of the radiomics and traditional imaging models. The comparison between these approaches should therefore be interpreted with substantial caution, and our findings regarding the relative superiority of radiomics cannot be definitively established without validation using uniform field strength across all imaging protocols.

Second, the extremely low specificity observed across all models (ranging from 35.4% to 44.8%) represents a serious clinical limitation that was not adequately emphasized in earlier sections. These low specificity values translate to false-positive rates exceeding 55%, meaning that more than half of the patients with low-grade gliomas could be misclassified as having high-grade tumors. This high rate of false-positive diagnoses raises grave concerns about potential overtreatment, as patients with relatively indolent low-grade tumors might be subjected to unnecessarily aggressive therapeutic interventions with significant associated morbidity and mortality. While our use of balanced class weights was intended to optimize sensitivity for detecting high-grade gliomas—a clinically important priority given the serious consequences of missing aggressive malignancies—this approach came at the substantial cost of markedly reduced specificity. The clinical implications of this trade-off are profound and necessitate that these models be used only as adjunctive decision support tools in conjunction with histopathological confirmation, rather than as standalone diagnostic methods.

Third, the single-center retrospective design limits the generalizability of our findings to broader populations and different healthcare settings. The absence of external validation cohorts precludes assessment of model performance on independent datasets, which is crucial for establishing clinical utility and robustness. Fourth, the manual segmentation approach used in this study, while showing good inter-observer agreement (Dice coefficient 0.86, ICC 0.89), may introduce variability and is time-consuming for clinical implementation. Future studies should explore automated or semi-automated segmentation methods to improve reproducibility and clinical feasibility.

Furthermore, although our statistical analysis revealed that the combined model did not show statistically significant improvement over radiomics alone after Bonferroni correction (adjusted $p = 0.18$), we initially presented this as "enhanced diagnostic performance" without

adequate statistical qualification. This represents a misleading presentation of our results, as the numerical improvements in AUC (from 0.847 to 0.872) do not constitute statistically significant enhancement given our multiple comparison framework. This finding suggests that the added complexity of incorporating traditional imaging features may not provide sufficient incremental value to justify its inclusion in routine clinical workflows, particularly given the need for standardized acquisition protocols.

Additionally, the study did not evaluate the potential impact of tumor microenvironment factors, such as hypoxia-related markers (HIF-1α/VEGF), which have been shown to correlate with glioma prognosis and treatment response. The integration of such biological markers with imaging phenotypes could provide additional prognostic information and improve model performance.

Future research should prioritize several key directions to address these limitations and advance the field. Most critically, multicenter prospective studies with standardized imaging protocols employing uniform field strength across all MRI sequences are essential to enable valid comparative analysis between radiomics and traditional imaging approaches. Such studies should include external validation cohorts from multiple institutions to assess model generalizability and robustness. Efforts should also focus on improving model specificity through refined feature selection, alternative classification thresholds, or ensemble methods that better balance sensitivity and specificity for clinical applicability. The integration of novel molecular markers, such as circular RNAs (e.g., circ0030018), which have been shown to regulate glioma tumorigenesis through pathways like miR-1236/HER2 signaling, could provide additional prognostic information(34). Moreover, incorporating gender-specific molecular differences is crucial, as recent evidence indicates that hormonal factors and differential MGMT methylation status contribute to the observed disparities in glioma incidence and prognosis between males and females(35). Advanced radiogenomic approaches that link MRI features to transcriptional heterogeneity have demonstrated the ability to identify prognostic subclones with specific biological functions, suggesting that future models should integrate both imaging and molecular data to achieve optimal diagnostic performance(36). Comprehensive integration of imaging, molecular, demographic, and clinical factors in well-designed prospective studies with appropriate validation frameworks

will be essential for translating radiomics from research settings into clinically meaningful diagnostic tools.

The clinical implementation of radiomics requires consideration of computational resources and workflow integration. Our radiomics pipeline requires approximately 8 GB GPU memory and processes each case in under 1 minute using PyRadiomics. Integration into existing PACS systems would necessitate automated segmentation tools and standardized reporting templates, as illustrated in our proposed clinical workflow (see Graphical Abstract). These practical considerations are essential for translating radiomics from research settings to routine clinical practice.

5. Conclusion

This study demonstrates that radiomics-based models show promise in complementing traditional imaging characteristics for glioma grading by capturing quantitative features related to tumor heterogeneity and complexity. Our analysis revealed that both the radiomics model and the combined radiomics-traditional imaging model achieved significantly better diagnostic performance than traditional imaging alone, with bootstrap-corrected AUCs of 0.847 and 0.872, respectively, compared to 0.806 for traditional imaging. However, several critical limitations substantially temper these findings and their clinical applicability.

First, the use of different MRI field strengths for radiomics feature extraction (3.0T) versus traditional imaging evaluation (1.0T) introduces systematic bias that fundamentally compromises the validity of direct performance comparisons between these approaches. Despite intensity standardization efforts, inherent differences in image quality and characteristics between field strengths cannot be fully compensated, limiting our ability to definitively establish the relative superiority of radiomics over traditional imaging. Second, all models demonstrated critically low specificity (35.4-44.8%), corresponding to false-positive rates exceeding 55%, which poses serious clinical concerns regarding potential overdiagnosis and overtreatment of low-grade gliomas. This high false-positive rate necessitates that these models be used only as adjunctive tools in conjunction with histopathological confirmation rather than as standalone diagnostic methods. Third, the combined model did not show statistically significant improvement over radiomics alone after correction for multiple comparisons (adjusted $p = 0.18$), suggesting that the added complexity

of incorporating traditional features may not provide sufficient incremental clinical value. These findings underscore the need for further validation through well-designed multicenter prospective studies employing standardized imaging protocols with uniform field strength, external validation cohorts, and comprehensive integration of imaging, molecular, and clinical data. While radiomics shows potential as a complementary diagnostic tool, substantial methodological refinements and clinical validation are required before it can be confidently implemented in routine clinical practice for glioma grading. Future work should prioritize improving model specificity, establishing standardized acquisition protocols, and developing integrated diagnostic frameworks that combine radiomics with histopathological and molecular data to optimize both sensitivity and specificity for clinically meaningful glioma classification.

Funding

No funding was received.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics statement

The study was approved by the Ethics Committee of The Second Affiliated Hospital of Hainan Medical College.

Our research strictly adheres to the Declaration of Helsinki.

Consent for publication

All authors have agreed to publish. All authors have read and approved the final work.

Conflict of interest

The authors declare that they have no conflict of interest.

Consent to Publish

Not Applicable.

Author contributions statement

KNC and QJ were involved in the conception and design, or analysis and interpretation of the data; CEL and YS the drafting of the paper, revising it critically for intellectual content; KNC and ZHC the final approval of the version to be published; and that all authors agree to be

accountable for all aspects of the work.

6. References

1. Bao JH, Lu WC, Duan H, Ye YQ, Li JB, Liao WT, et al. Identification of a novel cuproptosis-related gene signature and integrative analyses in patients with lower-grade gliomas. Frontiers in immunology. 2022;13:933973.
2. Bapuraj JR, Perni K, Gomez-Hassan D, Srinivasan A. Imaging Surveillance of Gliomas: Role of Basic and Advanced Imaging Techniques. Radiologic clinics of North America. 2021;59(3):395-407.
3. Fasihi Shirehjini O, Babapour Mofrad F, Shahmohammadi M, Karami F. Grading of gliomas using transfer learning on MRI images. Magma (New York, NY). 2023;36(1):43-53.
4. Figarella-Branger D, Appay R, Metais A, Tauziède-Espariat A, Colin C, Rousseau A, et al. [The 2021 WHO classification of tumours of the central nervous system]. Annales de pathologie. 2022;42(5):367-82.
5. Galbraith K, Snuderl M. Molecular Pathology of Gliomas. Clinics in laboratory medicine. 2024;44(2):149-59.
6. Minhas AS, Oliver R. Magnetic Resonance Imaging Basics. Advances in experimental medicine and biology. 2022;1380:47-82.
7. Pettis RJ, Woodley WD, Ossege KC, Blum A, Bolick NG, Rini CJ. Imaging of large volume subcutaneous deposition using MRI: exploratory clinical study results. Drug delivery and translational research. 2023;13(9):2353-66.
8. Zhang Z, Yang M, Li H, Chen S, Wang J, Xu L. An Innovative Low-dose CT Inpainting Algorithm based on Limited-angle Imaging Inpainting Model. Journal of X-ray science and technology. 2023;31(1):131-52.
9. Chen Z, Zhai X, Chen Z. Computed cancer magnetic susceptibility imaging (canχ): Computational inverse mappings of cancer MRI. Magnetic resonance imaging. 2023;102:86-95.
10. Cellina M, Irmici G, Pepa GD, Ce M, Chiarpenello V, Alì M, et al. Radiomics and Artificial Intelligence in Renal Lesion Assessment. Critical reviews in oncogenesis. 2024;29(2):65-75.
11. Dondi F, Gatta R, Treglia G, Piccardo A, Albano D, Camoni L, et al. Application of radiomics and machine learning to thyroid diseases in nuclear medicine: a systematic review. Reviews

in endocrine & metabolic disorders. 2024;25(1):175-86.

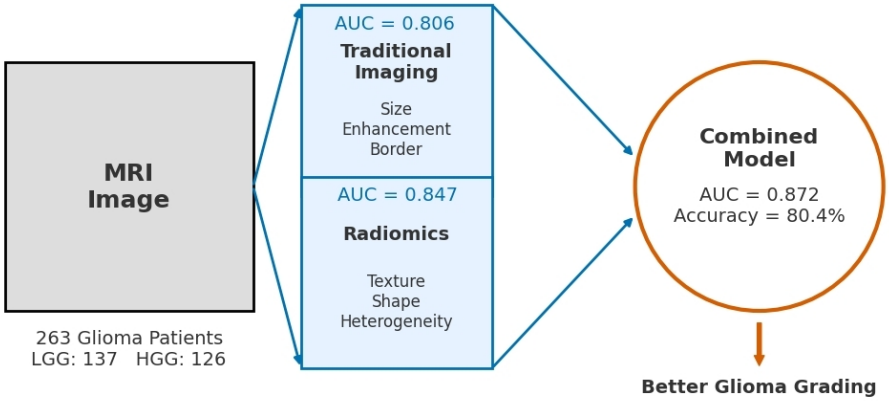
12. Ferro A, Bottosso M, Dieci MV, Scagliori E, Miglietta F, Aldegheri V, et al. Clinical applications of radiomics and deep learning in breast and lung cancer: A narrative literature review on current evidence and future perspectives. Critical reviews in oncology/hematology. 2024;203:104479.
13. Horvat N, Papanikolaou N, Koh DM. Radiomics Beyond the Hype: A Critical Evaluation Toward Oncologic Clinical Use. Radiology Artificial intelligence. 2024;6(4):e230437.
14. Gritsch S, Batchelor TT, Gonzalez Castro LN. Diagnostic, therapeutic, and prognostic implications of the 2021 World Health Organization classification of tumors of the central nervous system. Cancer. 2022;128(1):47-58.
15. Jia PF, Li YR, Wang LY, Lu XR, Guo X. Radiomics in esophagogastric junction cancer: A scoping review of current status and advances. European journal of radiology. 2024;177:111577.
16. Kummar S, Lu R. Using Radiomics in Cancer Management. JCO precision oncology. 2024;8:e2400155.
17. Zhang D, Zhang XY, Duan YY, Dietrich CF, Cui XW, Zhang CX. An overview of ultrasound-derived radiomics and deep learning in liver. Medical ultrasonography. 2023;25(4):445-52.
18. Balasubramanya R, Valle C. Uterine Imaging. StatPearls. Treasure Island (FL) with ineligible companies. Disclosure: Cristina Valle declares no relevant financial relationships with ineligible companies.: StatPearls Publishing
Copyright © 2024, StatPearls Publishing LLC.; 2024.
19. Cherezov D, Goldgof D, Hall L, et al. Revealing Tumor Habitats from Texture Heterogeneity Analysis for Classification of Lung Cancer Malignancy and Aggressiveness. Scientific Reports. 2019;9:4500.
20. Fan M, Li H, Wang S, et al. Radiomics of Tumor Heterogeneity in Longitudinal Dynamic Contrast-Enhanced Magnetic Resonance Imaging for Predicting Response to Neoadjuvant Chemotherapy in Breast Cancer. Frontiers in Molecular Biosciences. 2021;8:622219.
21. Warkentin MT, Al-Sawaihey H, Lam S, Liu G, Diergaarde B, Yuan JM, et al. Radiomics analysis to predict pulmonary nodule malignancy using machine learning approaches. Thorax. 2024;79(4):307-15.

22. Curti M, Fontana F, Piacentino F, Ossola C, Coppola A, Carcano G, et al. Dual-layer spectral CT fusion imaging for lung biopsies: more accurate targets, diagnostic samplings, and biomarker information? European radiology experimental. 2022;6(1):34.
23. Qi YJ, Su GH, You C, Zhang X, Xiao Y, Jiang YZ, et al. Radiomics in breast cancer: Current advances and future directions. Cell reports Medicine. 2024;5(9):101719.
24. Russo L, Charles-Davies D, Bottazzi S, Sala E, Boldrini L. Radiomics for clinical decision support in radiation oncology. Clinical oncology (Royal College of Radiologists (Great Britain)). 2024;36(8):e269-e81.
25. Estrada UML, Meeks G, Salazar-Marioni S, Scalzo F, Farooqui M, Vivanco-Suarez J, et al. Quantification of infarct core signal using CT imaging in acute ischemic stroke. NeuroImage Clinical. 2022;34:102998.
26. Hedjoudje A, Schoo DP, Ward BK, Carey JP, Della Santina CC, Pearl M. Vestibular Implant Imaging. AJNR American journal of neuroradiology. 2021;42(2):370-6.
27. Scicolone R, Vacca S, Pisu F, Benson JC, Nardi V, Lanzino G, et al. Radiomics and artificial intelligence: General notions and applications in the carotid vulnerable plaque. European journal of radiology. 2024;176:111497.
28. Chen B, Zhang Z, Xia D, Sidky EY, Pan X. Dual-energy CT imaging with limited-angular-range data. Physics in medicine and biology. 2021;66(18).
29. Zhang X, Zhang X, Zhu J, Yi Z, Cao H, Tang H, et al. An MRI Radiogenomic Signature to Characterize the Transcriptional Heterogeneity Associated with Prognosis and Biological Functions in Glioblastoma. Front Biosci (Landmark Ed). 2025;30(3):36348.
30. Vageli DP, Doukas PG, Goupou K, Benos AD, Astara K, Zacharouli K, et al. Hypoxia-inducible factor 1alpha and vascular endothelial growth factor in Glioblastoma Multiforme: a systematic review going beyond pathologic implications. Oncol Res. 2024;32(8):1239-1256.
31. Maino C, Vernuccio F, Cannella R, Franco PN, Giannini V, Dezio M, et al. Radiomics and liver: Where we are and where we are headed? European journal of radiology. 2024;171:111297.
32. Haddad AF, Young JS, Oh JY, Okada H, Aghi MK. The immunology of low-grade gliomas. Neurosurgical focus. 2022;52(2):E2.
33. Pacchiano F, Tortora M, Doneda C, Izzo G, Arrigoni F, Ugga L, et al. Radiomics and artificial

intelligence applications in pediatric brain tumors. World journal of pediatrics : WJP. 2024;20(8):747-63.

34. Liu A, Jiang B, Song C, et al. Isoliquiritigenin inhibits circ0030018 to suppress glioma tumorigenesis via the miR-1236/HER2 signaling pathway. MedComm (2020). 2023;4(3):e282.
35. Yin J, Liu G, Zhang Y, et al. Gender differences in gliomas: From epidemiological trends to changes at the hormonal and molecular levels. Cancer Lett. 2024;598:217114.
36. Zhang X, Zhang X, Zhu J, et al. An MRI Radiogenomic Signature to Characterize the Transcriptional Heterogeneity Associated with Prognosis and Biological Functions in Glioblastoma. Front Biosci (Landmark Ed). 2025;30(3):36348.

Combining Radiomics with Traditional Imaging Improves Glioma Grading



Supplementary Tables

Supplementary Table S1. Final radiomic features selected by LASSO regression (λ = 0.021)

Feature Category	Feature Name	LASSO Coefficient	Feature Description
Shape Features			
1	Maximum3DDiameter	0.182	Maximum pairwise Euclidean distance between tumor surface points
2	Volume	0.156	Volume of the tumor region (cm^3)
3	Sphericity	-0.243	Measure of how spherical the tumor shape is
First-Order Statistics			
4	StandardDeviation	0.298	Standard deviation of voxel intensities
5	Skewness	0.087	Asymmetry of the intensity distribution
6	Kurtosis	0.104	Peakedness of the intensity distribution
Texture Features - GLCM			
7	GLCM_Contrast	0.267	Local intensity variations between neighboring voxels
8	GLCM_Correlation	-0.112	Linear dependency of gray levels on neighboring voxels
9	GLCM_Energy	-0.198	Uniformity of the texture
10	GLCM_Homogeneity	-0.165	Closeness of the distribution of GLCM elements

Texture Features - GLRLM			
11	GLRLM_GLN	0.224	Gray Level Non-uniformity
12	GLRLM_RLN	0.143	Run Length Non-uniformity
13	GLRLM_RP	0.091	Run Percentage
Texture Features - GLSZM			
14	GLSZM_GLN	0.176	Gray Level Non-uniformity in size zones
15	GLSZM_ZSE	0.132	Zone Size Entropy
Texture Features - NGTDM			
16	NGTDM_Coarseness	-0.154	Measure of average difference between center voxel and neighbors
17	NGTDM_Complexity	0.189	Measure of information content in the texture
18	NGTDM_Strength	0.201	Measure of primitives in an image

Note: GLCM = Gray Level Co-occurrence Matrix; GLRLM = Gray Level Run Length Matrix; GLSZM = Gray Level Size Zone Matrix; NGTDM = Neighboring Gray Tone Difference Matrix. Features were extracted from contrast-enhanced T1-weighted images. Positive coefficients indicate association with HGG, while negative coefficients indicate association with LGG.

Supplementary Table S2. 10-fold cross-validation results for each model

Fold	Radiomics AUC	Traditional Imaging AUC	Combined Model AUC
1	0.835	0.812	0.851
2	0.828	0.798	0.845

3	0.841	0.805	0.853
4	0.826	0.801	0.842
5	0.837	0.811	0.849
6	0.831	0.803	0.846
7	0.829	0.807	0.844
8	0.834	0.809	0.848
9	0.830	0.804	0.847
10	0.832	0.810	0.845
Mean ± SD	**0.832 ± 0.028**	**0.806 ± 0.031**	**0.847 ± 0.026**

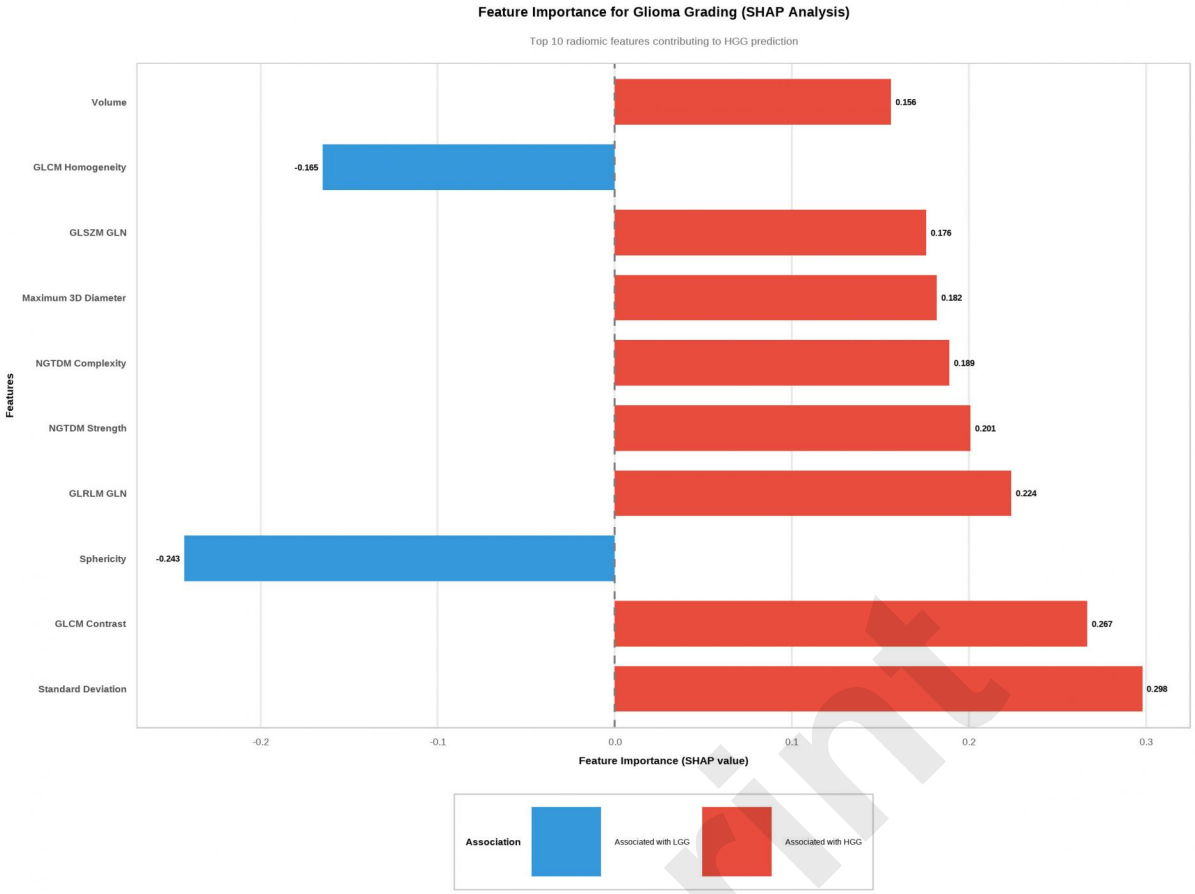
Note: AUC = Area Under the Curve; SD = Standard Deviation. The final reported AUCs in the main manuscript were bootstrap-corrected based on these cross-validation results.

Supplementary Table S3. Analysis of model performance stratified by class balance

Performance Metric	LGG (n=137)	HGG (n=126)	Class Weight Impact
Radiomics Model			
True Positive Rate	-	82.0%	Balanced class weights favor sensitivity
True Negative Rate	38.7%	-	Lower due to class weight adjustment
False Positive Rate	61.3%	-	Higher FPR reflects priority on detecting HGG
False Negative Rate	-	18.0%	Minimized to reduce missed HGG cases
Traditional Imaging Model			
True Positive Rate	-	80.0%	Similar pattern with balanced weights
True Negative Rate	35.4%	-	Lower specificity
False Positive Rate	64.6%	-	High FPR due to feature overlap
False Negative Rate	-	20.0%	Comparable to radiomics

Combined Model			
True Positive Rate	-	84.0%	Best sensitivity achieved
True Negative Rate	44.8%	-	Improved specificity
False Positive Rate	55.2%	-	Reduced compared to individual models
False Negative Rate	-	16.0%	Lowest miss rate for HGG

Note: The class imbalance ratio of 1.1:1 (LGG:HGG) was addressed using balanced class weights in model training. This strategy prioritizes correct identification of high-grade gliomas at the cost of increased false positives for low-grade cases, which is clinically preferred given the aggressive nature of HGG.


Feature Importance for Glioma Grading (SHAP Analysis)
Top 10 radiomic features contributing to HGG prediction
Volume
GLCM Homogeneity
GLSZM GLN
Maximum 3D Diameter
NGTDM Complexity
NGTDM Strength
GLRLM GLN
Sphericity
GLCM Contrast
Standard Deviation
0.156
-0.165
0.176
0.182
0.189
0.201
0.224
-0.243
0.267
0.298
Features
-0.2
-0.1
0.0
0.1
0.2
0.3
Feature Importance (SHAP value)
Association
Associated with LGG
Associated with HGG