

Development and External Validation of a Multimodal Habitat Radiomics Model to Support Preoperative Differentiation of Glioblastoma and Solitary Brain Metastasis

Mengyang He

Zhengzhou University

Xingyu Liu

Zhengzhou University

Yufei Gao

Zhengzhou University

Eryuan Gao

First Affiliated Hospital of Zhengzhou University

Peng Wang

Affiliated Hospital of Inner Mongolia Medical University

Guang Yang

East China Normal University

Chengxiu Zhang

East China Normal University

Mengzhu Wang

Siemens (China)

Yang Song

Siemens (China)

Xiaoyue Ma

First Affiliated Hospital of Zhengzhou University

Jie Bai

First Affiliated Hospital of Zhengzhou University

Jing Yan

First Affiliated Hospital of Zhengzhou University

Yong Zhang

First Affiliated Hospital of Zhengzhou University

Yang Gao

Affiliated Hospital of Inner Mongolia Medical University

Guohua Zhao

ghzhao@ha.edu.cn

First Affiliated Hospital of Zhengzhou University

Research Article

Keywords:

Posted Date: August 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10302559/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Preoperative differentiation between glioblastoma (Gb) and solitary brain metastasis (SBM) is clinically important because treatment planning and prognosis differ substantially. This study aimed to develop and externally test habitat-based radiomics models using conventional morphologic MRI and multiple diffusion MRI techniques, and to assess whether multimodal fusion improved diagnostic performance.

Methods

This retrospective two-center study included 253 patients with histologically confirmed Gb or SBM. Patients from Center A were divided into a training set ($n = 142$) and an internal test set ($n = 60$), while patients from Center B formed an independent external test set ($n = 51$). Radiomics features were extracted from tumor burden volume, peritumoral edema, and abnormal burden volume (ABV). Imaging sources included conventional morphologic MRI, DWI, DTI, DKI, NODDI, and MAP-MRI. Single-modality and multimodal fusion models were developed for each habitat. All data-dependent preprocessing, feature selection, imbalance correction, classifier selection, and hyperparameter tuning were restricted to the training set. Discrimination was assessed using the area under the receiver operating characteristic curve (AUC). External performance was primarily evaluated in the unharmonized external test set, with ComBat harmonization performed separately as a domain-adaptation sensitivity analysis. Calibration, Brier score, and decision curve analysis were assessed in the internal test set.

Results

Single-modality model performance varied across habitats, imaging sources, and test sets. Multimodal fusion models showed higher observed internal-test AUCs than the corresponding best single-modality models across the three habitats. Among the fusion models, the ABV fusion model achieved the highest observed AUC, reaching 0.940 (95% CI, 0.884–0.996) in the internal test set and 0.839 (95% CI, 0.708–0.969) in the unharmonized external test set. In the ComBat-harmonized domain-adaptation sensitivity analysis, the external-test AUC increased to 0.917 (95% CI, 0.829–1.000). The ABV fusion model also had the lowest Brier score among the evaluated ABV-based models and showed favorable net benefit on decision curve analysis.

Conclusions

Habitat-based multimodal radiomics integrating conventional morphologic MRI and diffusion MRI showed potential for differentiating Gb from SBM. In this dataset, the ABV fusion model showed the highest observed performance and may provide adjunctive imaging-based support for preoperative diagnostic decision-making.

Introduction

Preoperative differentiation between glioblastoma (Gb) and solitary brain metastasis (SBM) remains a common and clinically important challenge in neuro-oncology [1–3]. On conventional MRI, both entities may appear as a solitary contrast-enhancing brain lesion with surrounding edema, yet their diagnostic work-up, surgical planning, adjuvant treatment strategies, and prognosis differ substantially [4, 5]. Although histopathologic assessment remains the reference standard, reliable preoperative imaging-based risk stratification may support multidisciplinary decision-making, particularly when the primary malignancy is unknown, biopsy is delayed, or invasive procedures are considered high risk [6, 7]. These considerations have motivated the development of noninvasive imaging models intended to complement, rather than replace, clinical and pathological evaluation.

Advanced diffusion MRI provides quantitative information on tissue microstructure that is not fully captured by conventional morphologic MRI and has shown potential for differentiating Gb from SBM [8–10]. Diffusion-weighted imaging (DWI), diffusion tensor imaging (DTI), diffusion kurtosis imaging (DKI), neurite orientation dispersion and density imaging (NODDI), and mean

apparent propagator MRI (MAP-MRI) characterize complementary aspects of water diffusion, including cellularity, white matter integrity, non-Gaussian diffusion behavior, and microstructural complexity [11–13]. Radiomics can further translate these imaging characteristics into high-dimensional quantitative features for prediction model development [14, 15]. However, previous studies have differed in imaging protocols, analyzed regions of interest, feature extraction strategies, and modeling approaches. Consequently, the relative contribution of different diffusion-derived radiomics sources remains uncertain, particularly when different tumor habitats are considered.

The spatial habitat from which radiomics features are extracted is particularly relevant for distinguishing Gb from SBM. Features extracted from the enhancing or necrotic tumor burden may reflect intratumoral heterogeneity, whereas features from peritumoral edema may capture differences in tumor infiltration, vasogenic edema, or adjacent white matter disruption. A combined abnormal burden region may provide a more comprehensive representation of both the lesion and surrounding abnormal tissue [16, 17]. Because Gb and SBM may differ not only within the tumor core but also in their effects on the surrounding brain parenchyma, comparing habitat-specific models may help clarify which spatial representation is most informative for imaging-based differential diagnosis [18, 19].

In this retrospective two-center study, we developed and externally tested habitat-based radiomics models for differentiating Gb from SBM using morphologic MRI (MMRI) and five diffusion MRI techniques. We compared single-modality models with multimodal fusion models across three tumor habitats: tumor burden volume, peritumoral edema, and abnormal burden volume. We also used Shapley Additive Explanations to summarize the contribution of selected features and improve model interpretability. The aim of this study was to assess the relative contribution of multiple diffusion MRI-derived radiomics sources within a unified machine-learning framework and to identify a compact, externally tested model that could provide adjunctive support for preoperative diagnostic decision-making.

Materials and Methods

This retrospective study was approved by the institutional review boards of both participating institutions, and the requirement for written informed consent was waived. The study was conducted in accordance with the Declaration of Helsinki.

Patients

Patients with histopathologically confirmed Gb or SBM from two medical centers were retrospectively included. Patients were identified from Center A between December 2015 and December 2024 and from Center B between June 2018 and December 2024. Because the study period spanned the transition from the 2016 to the 2021 WHO classification of central nervous system tumors, a prespecified harmonization protocol was used to recategorize eligible adult IDH-wildtype glioblastoma cases according to the 2021 WHO criteria. Cases without sufficient histological or molecular evidence were excluded [20]. The inclusion and exclusion criteria are shown in Fig. 1, and the mapping protocol is provided in Supplementary Material E1.

A total of 202 eligible patients from Center A were randomly divided at the patient level into a training set and an internal test set at a 7:3 ratio. An additional 51 eligible patients from Center B were assigned to the external test set. The internal and external test sets were used only for model evaluation and were not involved in model training, feature selection, classifier selection, or hyperparameter tuning. Demographic and clinical characteristics are summarized in Table 1.

Table 1
Clinical characteristics of patients in the training, internal test, and external test sets

Characteristic	Training set (<i>n</i> = 142)			Internal test set (<i>n</i> = 60)			External test set (<i>n</i> = 51)		
	Gb (<i>n</i> = 101)	SBM (<i>n</i> = 41)	<i>P</i> value	Gb (<i>n</i> = 43)	SBM (<i>n</i> = 17)	<i>P</i> value	Gb (<i>n</i> = 36)	SBM (<i>n</i> = 15)	<i>P</i> value
Age, years									
Mean ± SD	52.5 ± 11.6	55.7 ± 12.4	0.156	57.0 ± 10.6	58.8 ± 10.6	0.566	55.3 ± 11.4	62.1 ± 10.9	0.060
Sex, <i>n</i>									
Male (%)	58 (57.4)	23 (56.1)	0.982	23 (53.5)	10 (58.8)	0.931	17 (47.2)	9 (60.0)	0.600
Female (%)	43 (42.6)	18 (43.9)		20 (46.5)	7 (41.2)		19 (52.8)	6 (40.0)	
Primary origin of SBM, <i>n</i>									
Lung, <i>n</i>									
Adenocarcinoma (%)		31 (75.6)			12 (68.8)			9 (60.0)	
Squamous cell carcinoma (%)		1 (2.4)							
Neuroendocrine carcinoma (%)		1 (2.4)			2 (12.6)			1 (6.7)	
Small cell lung carcinoma (%)		1 (2.4)						1 (6.7)	
Poorly differentiated carcinoma (%)		1 (2.4)						1 (6.7)	
Stomach, <i>n</i>									
Adenocarcinoma (%)		1 (2.4)			0			1 (6.7)	
Kidney, <i>n</i>									
Clear cell carcinoma (%)		3 (7.3)			1 (6.2)			2 (13.2)	
Uterus, <i>n</i>									
Endometrial carcinoma (%)		1 (2.4)			1 (6.2)				
Unknown site, <i>n</i> (%)		1 (2.4)			1 (6.2)				
Percentages may not sum to 100 due to rounding. Gb, glioblastoma; SBM, solitary brain metastasis; SD, standard deviation									

MRI Acquisition and Image Preprocessing

MRI examinations were performed on 3.0-T scanners at both centers. Center A used a MAGNETOM Prisma scanner, and Center B used a MAGNETOM Skyra scanner (Siemens Healthineers, Erlangen, Germany). The MRI protocol included morphologic MRI (MMRI) and diffusion MRI. MMRI sequences included T1-weighted imaging, T2-weighted imaging, fluid-attenuated inversion recovery, and contrast-enhanced T1-weighted imaging. Center A used three-dimensional contrast-enhanced T1-weighted

magnetization-prepared rapid gradient-echo imaging, whereas Center B used a three-dimensional CE-T1-weighted sequence. Detailed acquisition parameters are provided in Supplementary Material E2 and Supplementary Tables S1 and S2.

Diffusion MRI data were processed using in-house software, NeuDiLab, built on the open-source DIPY toolbox, to generate the following parametric maps: DWI metrics, apparent diffusion coefficient (ADC); DTI metrics, including fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD); DKI metrics, including mean kurtosis (MK), axial kurtosis (AK), radial kurtosis (RK), FA, MD, AD, and RD; NODDI metrics, including intracellular volume fraction (ICVF), isotropic volume fraction (ISOVF), and orientation dispersion index (ODI); and MAP-MRI metrics, including return-to-plane probability (RTPP), return-to-origin probability (RTOP), return-to-axis probability (RTAP), mean squared displacement (MSD), q-space inverse variance (QIV), non-Gaussianity (NG), axial non-Gaussianity (NGAx), and radial non-Gaussianity (NGRad) [17, 18]. The study workflow is shown in Fig. 2. All MMRI sequences and diffusion-derived parametric maps were coregistered to FLAIR images using ITK-SNAP version 3.8.0.

Tumor Habitat Segmentation

Three volumes of interest (VOIs) were defined to represent different tumor habitats: tumor burden volume (TBV; VOI 1), which included necrosis and solid tumor; peritumoral edema (VOI 2); and abnormal burden volume (ABV; VOI 3), defined as the union of VOI 1 and VOI 2.

Initial masks were generated automatically using an nnU-Net model trained on the BraTS 2020 dataset [21]. The segmentation results were reviewed and manually refined by two radiologists with 5 and 10 years of experience. Disagreements were resolved by consensus, and the final consensus masks were used for radiomics feature extraction.

Radiomics Feature Extraction and Model Development

Radiomics feature extraction was performed using Feature Explorer version 0.6.0 [22]. First-order and texture features were extracted from each habitat for each MRI sequence or parametric map. Texture features included gray-level co-occurrence matrix, gray-level run-length matrix, gray-level size zone matrix, gray-level dependence matrix, and neighboring gray-tone difference matrix features. Eight wavelet-transformed images were also generated. For each VOI and MRI sequence or parametric map, 837 radiomics features were extracted. The feature categories are listed in Supplementary Table S3.

The primary modeling task was binary classification of Gb versus SBM. Model development was restricted to the training set. For each tumor habitat, baseline models were developed separately for six imaging sources: MMRI, DWI, DTI, DKI, NODDI, and MAP-MRI. Multimodal fusion models were then constructed by combining features from two to six imaging sources. Each model was limited to four features; the rationale for this feature limit relative to sample size is provided in Supplementary Material E3.

Each candidate modeling pipeline consisted of class-imbalance correction, feature normalization, redundancy reduction, feature selection, and classifier training [23]. Class imbalance correction was performed only within the training data. Normalization methods included min-max scaling, z-score standardization, and mean normalization. Redundant features were removed using Pearson correlation analysis. Candidate feature selection methods included analysis of variance, the Kruskal-Wallis test, recursive feature elimination, and Relief. Candidate classifiers included support vector machine, logistic regression, and L1-regularized logistic regression.

To reduce the risk of data leakage, all data-dependent preprocessing and model-development steps were fitted within the training data only. During five-fold cross-validation, class imbalance correction, normalization, redundancy reduction, feature selection, and classifier training were repeated within each training fold. The held-out fold was used only for performance estimation. After the optimal pipeline had been selected based on cross-validation performance in the training set, the final model was refitted on the full training set and applied without further modification to the internal and external test sets.

Inter-Center Harmonization and Domain-Adaptation Analysis

Because the two centers used different scanners and diffusion acquisition schemes, inter-center variability was considered a predefined methodological issue [24]. Each final model was first evaluated in the external test set without harmonization. As a sensitivity analysis, ComBat harmonization was then applied to align the external cohort with the internal reference distribution, and model evaluation was repeated.

ComBat harmonization was performed using in-house software, nnFAE, based on scikit-learn version 1.4 and PyRadiomics version 3.0.1. Because this procedure used cohort-level information from the external site, the ComBat-harmonized analysis was interpreted as a domain-adaptation sensitivity analysis rather than as fully independent external validation. Both unharmonized and ComBat-harmonized external test results were reported.

Model Evaluation and Statistical Analysis

The final locked models were evaluated in the internal and external test sets. Discrimination was assessed using the area under the receiver operating characteristic curve, sensitivity, specificity, accuracy. Ninety-five percent confidence intervals were calculated for AUCs. Differences in AUCs between models were compared using the DeLong test. Statistical analyses were performed using MedCalc version 20.0.15. Two-sided P values less than 0.05 were considered statistically significant.

Calibration curves were used to compare predicted probabilities with observed outcomes [25]. The Brier score was calculated as the mean squared difference between the predicted probability and the binary outcome, with lower values indicating better probabilistic accuracy [26]. Calibration was assessed in the internal test set for the ABV fusion model and selected ABV-based single-modality models. These comparators included the ABV MMRI model, representing conventional morphologic imaging, and the ABV NODDI and ABV MAP-MRI models, representing advanced diffusion MRI models with distinct microstructural information. This analysis was intended to evaluate the probabilistic performance of the fusion model relative to clinically and technically relevant single-modality alternatives.

Clinical utility was assessed in the internal test set using decision curve analysis [27]. Net benefit was calculated across a range of threshold probabilities. The ABV fusion model was compared with the same ABV-based single-modality comparators and with two default strategies: treating all patients as Gb and treating no patients as Gb. This analysis was used to assess whether the fusion model provided greater net benefit than selected imaging-model comparators across plausible decision thresholds.

Model Interpretation

Shapley Additive Explanations were used to summarize the contribution of selected radiomics features to model predictions [28]. SHAP feature importance bar plots and summary plots were generated for the final selected fusion model. These analyses were used to improve model transparency and to describe the relative contribution and direction of effect of imaging-derived features in model predictions.

Results

Patient Characteristics

A total of 253 patients were included, comprising 180 patients with Gb and 73 patients with SBM. The Center A cohort was divided into a training set of 142 patients and an internal test set of 60 patients. The independent Center B cohort included 51 patients and served as the external test set. The training set included 101 patients with Gb and 41 with SBM, the internal test set included 43 patients with Gb and 17 with SBM, and the external test set included 36 patients with Gb and 15 with SBM. Within each dataset, age and sex did not differ significantly between patients with Gb and those with SBM (all $P > 0.05$; Table 1).

Performance of Single-Modality Baseline Models

The performance of the single-modality baseline models varied across tumor habitats and imaging sources (Table 2). In the tumor burden volume (TBV), internal-test AUCs ranged from 0.722 to 0.907. The NODDI model had the highest internal-test AUC in this habitat (AUC, 0.907; 95% CI, 0.833–0.981), whereas the highest external-test AUC among the TBV baseline models was

observed for the DTI model (AUC, 0.822; 95% CI, 0.673–0.971). In the peritumoral edema habitat, internal-test AUCs ranged from 0.679 to 0.816. The DTI model had the highest internal-test AUC (AUC, 0.816; 95% CI, 0.691–0.941), and the MAP-MRI model had the highest external-test AUC (AUC, 0.780; 95% CI, 0.622–0.938). In the abnormal burden volume (ABV), internal-test AUCs ranged from 0.757 to 0.847. The NODDI model had the highest internal-test AUC (AUC, 0.847; 95% CI, 0.748–0.945), whereas the MAP-MRI model had the highest external-test AUC (AUC, 0.880; 95% CI, 0.779–0.980).

Table 2
Performance of single-modality baseline models in the internal and external test sets

VOI	Model	Internal test set				External test set			
		Sen	Spe	Acc	AUC (95% CI)	Sen	Spe	Acc	AUC (95% CI)
TBV	MMRI	0.767	0.824	0.783	0.773 (0.617–0.928)	0.806	0.733	0.784	0.767 (0.624–0.908)
	DWI	0.744	0.824	0.767	0.762 (0.618–0.906)	0.639	0.800	0.686	0.691 (0.529–0.852)
	DTI	0.674	0.882	0.733	0.792 (0.655–0.929)	0.917	0.667	0.843	0.822 (0.673–0.971)
	DKI	0.698	0.824	0.733	0.791 (0.667–0.914)	0.722	0.800	0.745	0.811 (0.686–0.935)
	NODDI	0.791	0.941	0.833	0.907 (0.833–0.981)	0.639	0.867	0.706	0.809 (0.686–0.932)
	MAP-MRI	0.442	0.941	0.583	0.722 (0.582–0.863)	0.611	0.800	0.667	0.732 (0.570–0.893)
Peritumoral edema	MMRI	0.744	0.765	0.750	0.755 (0.603–0.908)	0.889	0.600	0.804	0.707 (0.518–0.896)
	DWI	0.763	0.706	0.747	0.679 (0.524–0.833)	0.750	0.667	0.725	0.662 (0.504–0.820)
	DTI	0.721	0.882	0.767	0.816 (0.691–0.941)	0.600	0.838	0.784	0.772 (0.637–0.907)
	DKI	0.698	0.765	0.718	0.740 (0.607–0.873)	0.889	0.533	0.784	0.728 (0.569–0.887)
	NODDI	0.744	0.882	0.783	0.803 (0.677–0.930)	0.833	0.600	0.765	0.732 (0.560–0.904)
	MAP-MRI	0.861	0.765	0.833	0.799 (0.652–0.946)	0.917	0.667	0.843	0.780 (0.622–0.938)
ABV	MMRI	0.884	0.706	0.833	0.836 (0.727–0.945)	0.778	0.800	0.784	0.850 (0.740–0.960)
	DWI	0.861	0.647	0.800	0.757 (0.611–0.901)	0.417	1.000	0.588	0.741 (0.598–0.884)
	DTI	0.767	0.882	0.800	0.837 (0.737–0.937)	0.694	0.933	0.765	0.824 (0.709–0.939)
	DKI	0.721	0.941	0.783	0.833 (0.733–0.933)	0.722	0.933	0.784	0.818 (0.702–0.935)
	NODDI	0.651	0.941	0.733	0.847 (0.748–0.945)	0.917	0.667	0.843	0.815 (0.677–0.953)
	MAP-MRI	0.791	0.882	0.817	0.839 (0.715–0.962)	0.889	0.800	0.863	0.880 (0.779–0.980)

Sen, Sensitivity; Spe, specificity; Acc, accuracy; AUC, area under the receiver operating characteristic curve; CI, confidence interval; TBV, tumor burden volume; ABV, abnormal burden volume.

Overall, no single imaging source consistently achieved the highest AUC across all habitats and test sets. Key steps in the radiomics workflow and the four selected features for the baseline models are provided in Supplementary Table S4.

Performance of Multimodal Fusion Models

The selected multimodal fusion models for each habitat are summarized in Table 3, and the corresponding ROC curves are shown in Fig. 3.

Table 3
Performance of the multimodal fusion models

Habitat	Fusion model	Evaluation set	AUC (95% CI)	Sensitivity	Specificity	Accuracy
TBV	MMRI + DWI + DKI + NODDI + MAP-MRI	Internal test	0.912 (0.813-1.000)	0.907	0.941	0.917
TBV	MMRI + DWI + DKI + NODDI + MAP-MRI	External test, unharmonized	0.818 (0.683–0.952)	0.833	0.733	0.804
TBV	MMRI + DWI + DKI + NODDI + MAP-MRI	External test, ComBat-harmonized	0.874 (0.750–0.998)	0.917	0.800	0.882
Peritumoral edema	MMRI + DWI + MAP-MRI	Internal test	0.907 (0.794-1.000)	0.861	0.941	0.883
Peritumoral edema	MMRI + DWI + MAP-MRI	External test, unharmonized	0.781 (0.640–0.923)	0.694	0.733	0.705
Peritumoral edema	MMRI + DWI + MAP-MRI	External test, ComBat-harmonized	0.865 (0.753–0.976)	0.833	0.800	0.824
ABV	MMRI + DWI + NODDI + MAP-MRI	Internal test	0.940 (0.884–0.996)	0.930	0.824	0.900
ABV	MMRI + DWI + NODDI + MAP-MRI	External test, unharmonized	0.839 (0.708–0.969)	0.916	0.733	0.862
ABV	MMRI + DWI + NODDI + MAP-MRI	External test, ComBat-harmonized	0.917 (0.829-1.000)	0.889	0.867	0.882

In the TBV, the selected fusion model combined MMRI, DWI, DKI, NODDI, and MAP-MRI features. This model achieved an internal-test AUC of 0.912 (95% CI, 0.813–1.000), with sensitivity of 0.907, specificity of 0.941, and accuracy of 0.917. In the external test set before harmonization, the AUC was 0.818 (95% CI, 0.683–0.952). In the ComBat-harmonized sensitivity analysis, the AUC was 0.874 (95% CI, 0.750–0.998).

For the peritumoral edema habitat, the selected fusion model combined MMRI, DWI, and MAP-MRI features. The internal-test AUC was 0.907 (95% CI, 0.794–1.000), with sensitivity of 0.861, specificity of 0.941, and accuracy of 0.883. In the external test set before harmonization, the AUC was 0.781 (95% CI, 0.640–0.923). In the ComBat-harmonized sensitivity analysis, the AUC was 0.865 (95% CI, 0.753–0.976).

For the ABV, the selected fusion model combined MMRI, DWI, NODDI, and MAP-MRI features. This model had the highest numerical AUC among the three habitat-based fusion models in the internal test set, with an AUC of 0.940 (95% CI, 0.884–0.996), sensitivity of 0.930, specificity of 0.824, and accuracy of 0.900. In the external test set before harmonization, the AUC was 0.839 (95% CI, 0.708–0.969). In the ComBat-harmonized sensitivity analysis, the AUC was 0.917 (95% CI, 0.829–1.000).

In the internal test set, DeLong testing showed no statistically significant difference between the ABV fusion model and either the TBV fusion model ($P = 0.12$) or the peritumoral edema fusion model ($P = 0.080$), although the ABV fusion model had the highest

numerical AUC (Supplementary Table S5). Detailed fusion-model results and selected features are provided in Supplementary Tables S6–S8.

Calibration, Clinical Utility, and Model Interpretation

Calibration and clinical utility were evaluated in the internal test set (Fig. 4). Among the ABV-based models evaluated, the ABV fusion model had the lowest Brier score, indicating better probabilistic accuracy than the selected single-modality comparators (Fig. 4A). Calibration curves showed some deviation from the ideal line for all models; however, the ABV fusion model generally showed closer agreement between predicted probabilities and observed outcomes than the comparator models (Fig. 4B).

Decision curve analysis showed that the ABV fusion model provided greater net benefit than the treat-none strategy across the evaluated threshold probabilities and generally outperformed the selected ABV-based single-modality comparators over most of the threshold range (Fig. 4C). The difference was most apparent at intermediate-to-high threshold probabilities, whereas the net benefit of different strategies was similar at very low thresholds.

SHAP analysis was performed for the final selected fusion model, which corresponded to the ABV fusion model (Fig. 5). The most influential features were ODI_wavelet-HLL_firstorder_InterquartileRange, Flair_original_glszm_ZoneEntropy, CET1_original_glcM_DifferenceEntropy, and NGAx_wavelet-LLL_glszm_GrayLevelNonUniformityNormalized (Fig. 5A). In the SHAP summary plot, lower values of the ODI and FLAIR features tended to increase the model output for Gb, whereas higher values of the CET1 and NGAx features tended to increase the model output for Gb (Fig. 5B). These findings indicate that the final selected fusion model used information from both diffusion-derived and morphology-derived habitat features.

Discussion

In this retrospective multicenter study of 253 patients, we developed habitat-based multimodal radiomics models to differentiate glioblastoma (Gb) from solitary brain metastasis (SBM) using morphologic MRI and multiple diffusion MRI techniques. The performance of the single-modality models varied across tumor habitats, imaging sources, and test sets, suggesting that no single imaging source provided consistently optimal discrimination. By contrast, the multimodal fusion models showed stronger overall performance. Among them, the ABV fusion model achieved the highest numerical AUC, with an AUC of 0.940 in the internal test set and 0.839 in the external test set before harmonization. In the ComBat-harmonized sensitivity analysis, the external AUC increased to 0.917. The ABV fusion model also showed favorable calibration and decision-curve findings in the internal test set. These results suggest that habitat-based multimodal MRI radiomics may provide useful imaging-derived information for the preoperative differentiation of Gb and SBM.

The comparison across tumor habitats was a key component of this study. Gb and SBM may show substantial overlap on conventional MRI, particularly when both present as solitary enhancing masses with necrosis and surrounding edema [29, 30]. However, the enhancing tumor, peritumoral region, and combined abnormal burden may reflect different aspects of tumor-related imaging heterogeneity [6, 31]. By evaluating TBV, peritumoral edema, and ABV in parallel, this study examined whether radiomics performance varied according to the spatial definition of the lesion. The good performance observed across the habitat-based fusion models suggests that different lesion definitions can capture diagnostically relevant information. The ABV fusion model may have benefited from incorporating information from both the enhancing lesion and the surrounding abnormal signal, which could have contributed to its higher numerical performance in this dataset.

The multimodal fusion strategy was another important aspect of the proposed approach. Morphologic MRI provides information on macroscopic tumor appearance, contrast enhancement, necrosis, and edema, whereas diffusion MRI reflects water diffusion and tissue microstructural characteristics [9, 32]. Across habitats, the single-modality model with the highest AUC was not consistent between the internal and external test sets. This finding suggests that individual imaging sources may capture only part of the discriminative information needed to differentiate Gb from SBM. By combining selected features from morphologic MRI and diffusion-derived parameter maps, the fusion models may have incorporated complementary imaging information that was not fully represented by any single modality alone [33, 34].

External testing showed lower AUCs before harmonization than in the ComBat-harmonized sensitivity analysis, which is consistent with the known sensitivity of radiomics features to scanner-, acquisition-, and center-related variation [24, 35]. In this study, the unharmonized external test set was used as the primary assessment of external model performance, whereas the ComBat-harmonized analysis was used as a sensitivity analysis to explore the potential influence of center-related feature variation. The improved external AUCs after ComBat harmonization suggest that part of the performance decrease in the independent cohort may have been related to technical heterogeneity across centers.

Calibration and clinical utility analyses provided additional information beyond discrimination. In the internal test set, the ABV fusion model had the lowest Brier score among the evaluated ABV-based models, indicating better probabilistic accuracy than the selected single-modality comparators. Calibration curves showed some deviation from the ideal line for all models, but the ABV fusion model generally showed closer agreement between predicted probabilities and observed outcomes. Decision curve analysis showed that the ABV fusion model provided greater net benefit than the treat-none strategy and generally outperformed the selected ABV-based single-modality comparators across much of the evaluated threshold range.

SHAP analysis was used to interpret the final selected fusion model, which corresponded to the ABV fusion model. The most influential features were derived from both morphologic MRI and diffusion MRI, including features from ODI, FLAIR, contrast-enhanced T1-weighted imaging, and NGAx maps. This finding indicates that the model used information from multiple imaging sources when generating predictions. The SHAP summary plot further showed that the direction and magnitude of feature contributions varied across patients, reflecting the heterogeneous imaging presentation of Gb and SBM. These results should be interpreted as model-level explanations rather than direct evidence of underlying biological mechanisms [36].

This study has several limitations. First, the retrospective design may have introduced selection bias. Second, although an independent external test set was included, the external cohort was modest in size, which limited the precision of external performance estimates. Third, radiomics features are sensitive to image acquisition, preprocessing, segmentation, and feature-extraction settings. Although a standardized pipeline was used and ComBat harmonization was explored in a sensitivity analysis, residual technical variability may remain. Fourth, calibration, decision curve analysis, and SHAP interpretation were performed in the internal test set; further evaluation in larger external cohorts is warranted.

In conclusion, habitat-based multimodal radiomics using morphologic MRI and diffusion MRI provided useful information for differentiating Gb from SBM. The ABV fusion model showed the highest numerical performance among the evaluated habitat-based fusion models and demonstrated favorable calibration and decision-curve findings in the internal test set. These findings support the potential value of ABV-based multimodal radiomics as an adjunctive approach for preoperative diagnostic decision support, although further validation in larger multicenter cohorts is warranted.

Abbreviations

ABV = abnormal burden volume, AUC = area under the receiver operating characteristic curve, DKI = diffusion kurtosis imaging, DTI = diffusion tensor imaging, DWI = diffusion-weighted imaging, Gb = glioblastoma, MAP-MRI = mean apparent propagator MRI, NODDI = neurite orientation dispersion and density imaging, SBM = solitary brain metastasis, TBV = tumor burden volume

Declarations

Ethics approval and consent to participate

This retrospective study was approved by the institutional review boards (IRBs) of the two participating institutions [First Affiliated Hospital of Zhengzhou University (2019-KY-231, 2019) and Affiliated Hospital of Inner Mongolia Medical University (KY2023064, 2023)], both of which waived the requirement for written informed consent from patients. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials

The data could be requested from the corresponding author upon reasonable request.

Competing interests

The authors declared no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 82402395 and 82502477)

Authors' contributions

M.H. and X.L. drafted the main manuscript text and prepared the figures and tables. Y.G. and E.G. prepared the supplementary material. P.W. curated the external-center data. G.Y. and C.Z. provided the image preprocessing programs. M.W. and Y.S. provided technical guidance and maintenance support for MRI acquisition. X.M. and J.B. reviewed and refined the imaging annotations. J.Y. and Y.Z. organized data collection and curation at the primary center. Y.G. organized data collection and curation at the external center. G.Z. supervised the study and provided funding support. All authors reviewed and approved the final manuscript.

Acknowledgements

We thank all the authors for their efforts in this paper. Large language models were not utilized for the creation of this manuscript.

References

1. Sun Q, Chen Y, Liang C, et al. Biologic pathways underlying prognostic radiomics phenotypes from paired MRI and RNA sequencing in glioblastoma. *Radiology*. 2021;301(3):654–63.
2. Cagney DN, Martin AM, Catalano PJ, et al. Incidence and prognosis of patients with brain metastases at diagnosis of systemic malignancy: a population-based study. *Neuro Oncol*. 2017;19(11):1511–21.
3. Schaff LR, Mellinghoff IK. Glioblastoma and other primary brain malignancies in adults: a review. *JAMA*. 2023;329(7):574–87.
4. Pouyan A, Ghorbanlo M, Eslami M, et al. Glioblastoma multiforme: insights into pathogenesis, key signaling pathways, and therapeutic strategies. *Mol Cancer*. 2025;24(1):58.
5. Mansouri A, Ozair A, Bhanja D, et al. Stereotactic radiosurgery for patients with brain metastases: current principles, expanding indications and opportunities for multidisciplinary care. *Nat Rev Clin Oncol*. 2025;22(5):327–47.
6. Chen Y, Lin H, Sun J, Pu R, Zhou Y, Sun B. Texture feature differentiation of glioblastoma and solitary brain metastases based on tumor and tumor-brain interface. *Acad Radiol*. 2025;32(1):400–10.
7. Dong F, Li Q, Jiang B, et al. Differentiation of supratentorial single brain metastasis and glioblastoma by using peri-enhancing oedema region-derived radiomic features and multiple classifiers. *Eur Radiol*. 2020;30(5):3015–22.
8. Fordham AJ, Hacherl CC, Patel N, et al. Differentiating glioblastomas from solitary brain metastases: an update on the current literature of advanced imaging modalities. *Cancers (Basel)*. 2021;13(12):2960.
9. Mao J, Zeng W, Zhang Q, et al. Differentiation between high-grade gliomas and solitary brain metastases: a comparison of five diffusion-weighted MRI models. *BMC Med Imaging*. 2020;20(1):124.
10. Byrnes TJ, Barrick TR, Bell BA, Clark CA. Diffusion tensor imaging discriminates between glioblastoma and cerebral metastases in vivo. *NMR Biomed*. 2011;24(1):54–60.
11. Gao E, Wang P, Bai J, et al. Radiomics analysis of diffusion kurtosis imaging: distinguishing between glioblastoma and single brain metastasis. *Acad Radiol*. 2024;31(3):1036–43.
12. Kadota Y, Hirai T, Azuma M, et al. Differentiation between glioblastoma and solitary brain metastasis using neurite orientation dispersion and density imaging. *J Neuroradiol*. 2020;47(3):197–202.

13. Wang P, Weng L, Xie S, et al. Primary application of mean apparent propagator-MRI diffusion model in the grading of diffuse glioma. *Eur J Radiol.* 2021;138:109622.
14. Bai J, He M, Gao E, et al. Radiomic texture analysis based on neurite orientation dispersion and density imaging to differentiate glioblastoma from solitary brain metastasis. *BMC Cancer.* 2023;23(1):1231.
15. Artzi M, Bressler I, Ben Bashat D. Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. *J Magn Reson Imaging.* 2019;50(2):519–28.
16. Moon HH, Park JE, Kim N, et al. Prospective longitudinal analysis of physiologic MRI-based tumor habitat predicts short-term patient outcomes in IDH-wildtype glioblastoma. *Neuro Oncol.* 2025;27(3):841–53.
17. Park JE, Oh JY, Park DH, et al. Mapping tumor habitats in isocitrate dehydrogenase-wild type glioblastoma: integrating MRI, pathologic, and RNA data from the Ivy Glioblastoma Atlas Project. *Neuro Oncol.* 2025;27(1):291–301.
18. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5:4006.
19. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology.* 2020;295(2):328–38.
20. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol.* 2021;23(8):1231–51.
21. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* 2021;18(2):203–11.
22. Song Y, Zhang J, Zhang YD, et al. FeAture Explorer (FAE): a tool for developing and comparing radiomics models. *PLoS ONE.* 2020;15(8):e0237587.
23. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep.* 2015;5:13087.
24. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. *Radiology.* 2019;291(1):53–9.
25. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med.* 2019;17(1):230.
26. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128–38.
27. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak.* 2006;26(6):565–74.
28. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems 30 (NIPS 2017).* Curran Associates, Inc.; 2017. pp. 4765–74.
29. Law M, Cha S, Knopp EA, Johnson G, Arnett J, Litt AW. High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. *Radiology.* 2002;222(3):715–21.
30. Lee EJ, terBrugge K, Mikulis D, et al. Diagnostic value of peritumoral minimum apparent diffusion coefficient for differentiation of glioblastoma multiforme from solitary metastatic lesions. *AJR Am J Roentgenol.* 2011;196(1):71–6.
31. Chen Y, Guo W, Li Y, et al. Differentiation of glioblastoma and solitary brain metastasis using brain-tumor interface radiomics features based on MR images: a multicenter study. *Acad Radiol.* 2025;32(7):4164–76.
32. Bai J, He M, Gao E, et al. High-performance presurgical differentiation of glioblastoma and metastasis by means of multiparametric neurite orientation dispersion and density imaging (NODDI) radiomics. *Eur Radiol.* 2024;34(10):6616–28.
33. Han Y, Zhang L, Niu S, et al. Differentiation between glioblastoma multiforme and metastasis from the lungs and other sites using combined clinical/routine MRI radiomics. *Front Cell Dev Biol.* 2021;9:710461.
34. Zhang Y, Zhang H, Zhang H, et al. Glioblastoma and solitary brain metastasis: differentiation by integrating demographic-MRI and deep-learning radiomics signatures. *J Magn Reson Imaging.* 2024;60(3):909–20.
35. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007;8(1):118–27.

36. Xia X, Wu W, Tan Q, Gou Q. Interpretable machine learning models for differentiating glioblastoma from solitary brain metastasis using radiomics. Acad Radiol. 2025;32(9):5388–400.

Figures

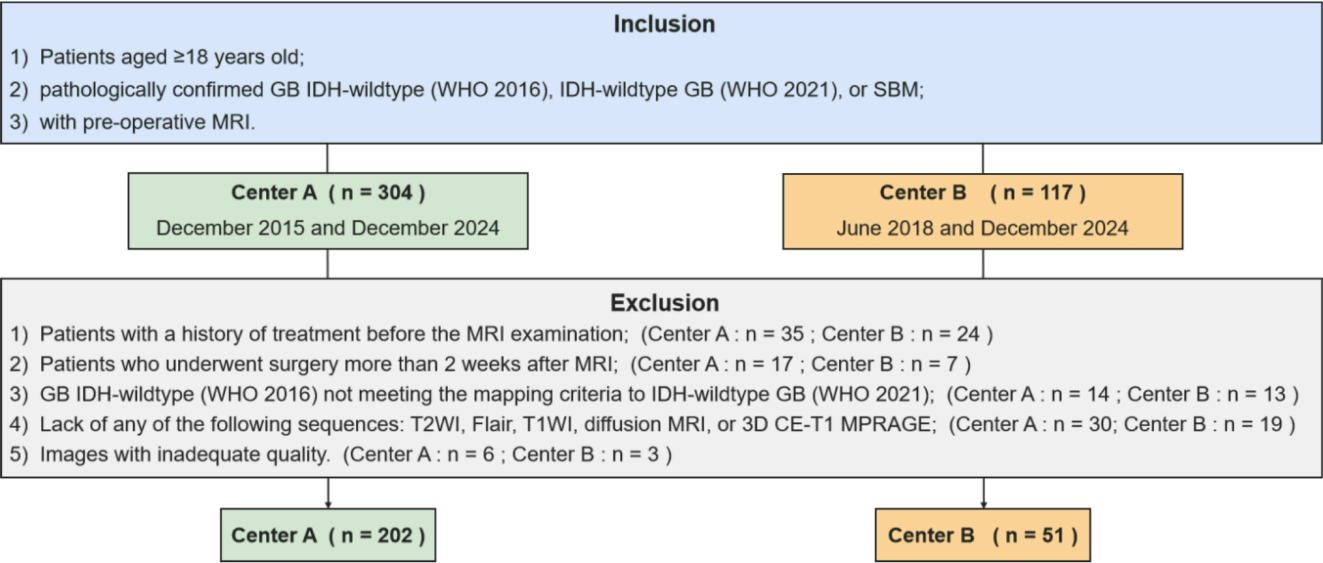


Figure 1

Flow diagram of patient selection and cohort allocation.

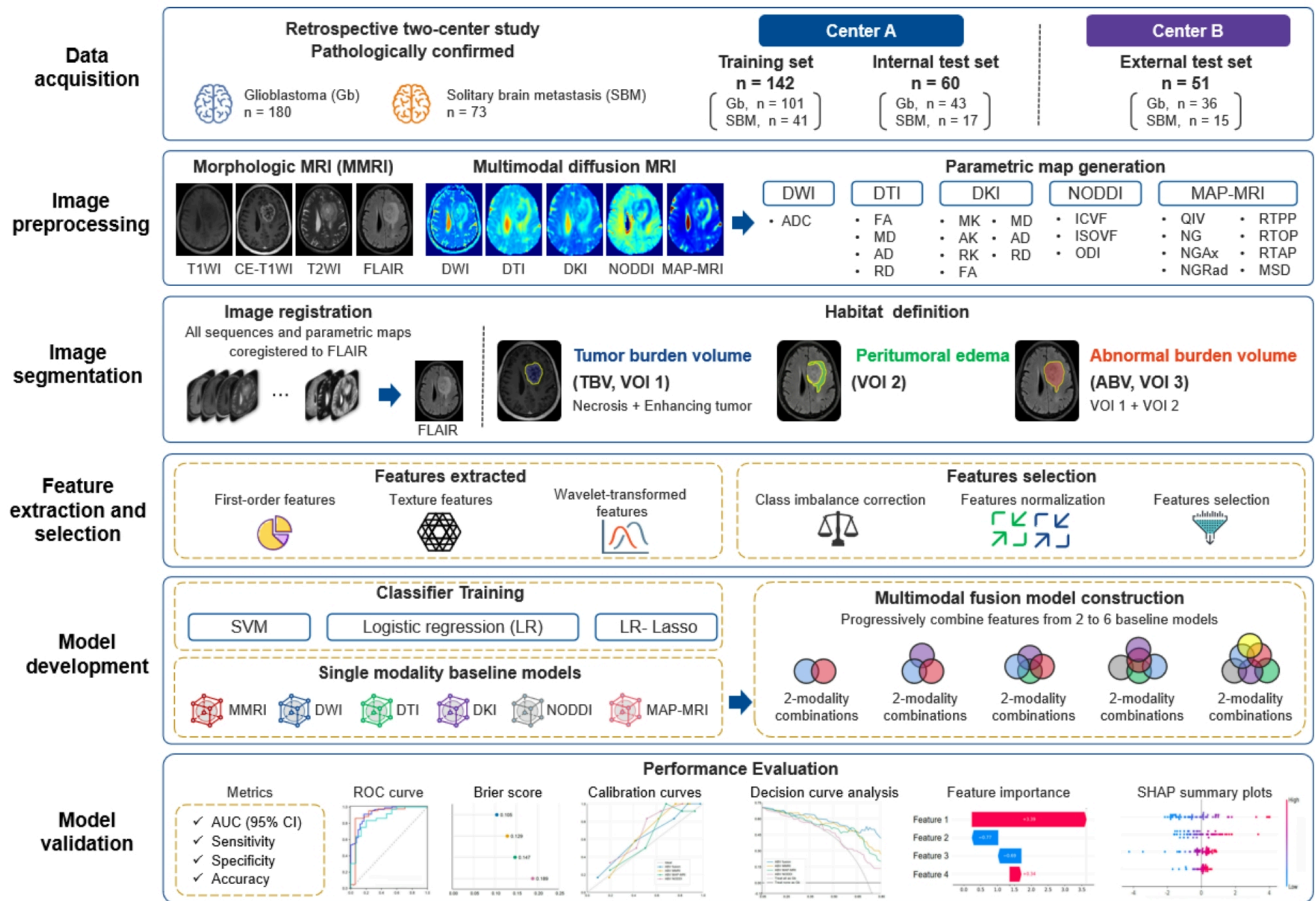


Figure 2

Overview of the habitat-based multimodal radiomics workflow.

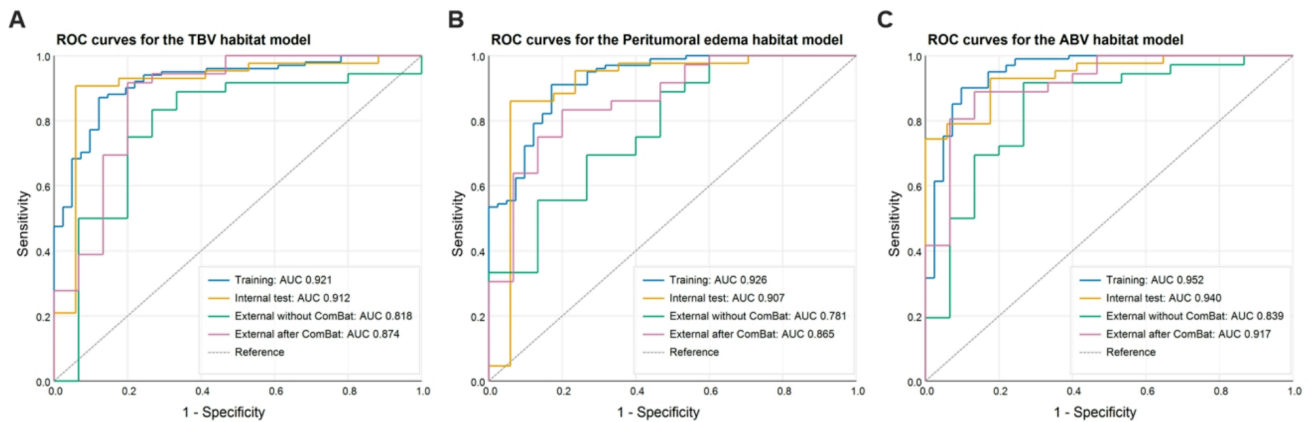


Figure 3

ROC analysis of the habitat-based multimodal fusion models. ROC curves are shown for the (A) TBV model, (B) peritumoral edema model, and (C) ABV model. For each habitat, model performance is shown in the training set, internal test set, external test set before ComBat harmonization, and external test set after ComBat harmonization.

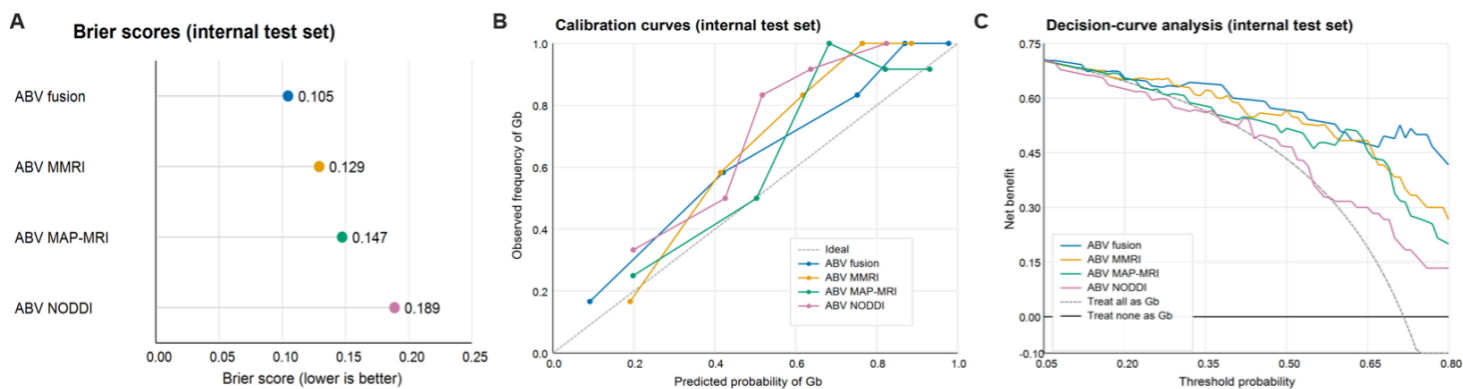


Figure 4

Calibration, Brier score, and decision-curve analysis of ABV-based models in the internal test set. (A) Brier scores for the ABV fusion model and the ABV MMRI, ABV MAP-MRI, and ABV NODDI models. Lower Brier scores indicate better overall probabilistic performance. (B) Calibration curves comparing predicted probabilities with observed diagnostic outcomes. The dashed diagonal line represents ideal calibration. (C) Decision-curve analysis showing net benefit across threshold probabilities for the ABV-based models, with treat-all-as-Gb and treat-none strategies included as references.

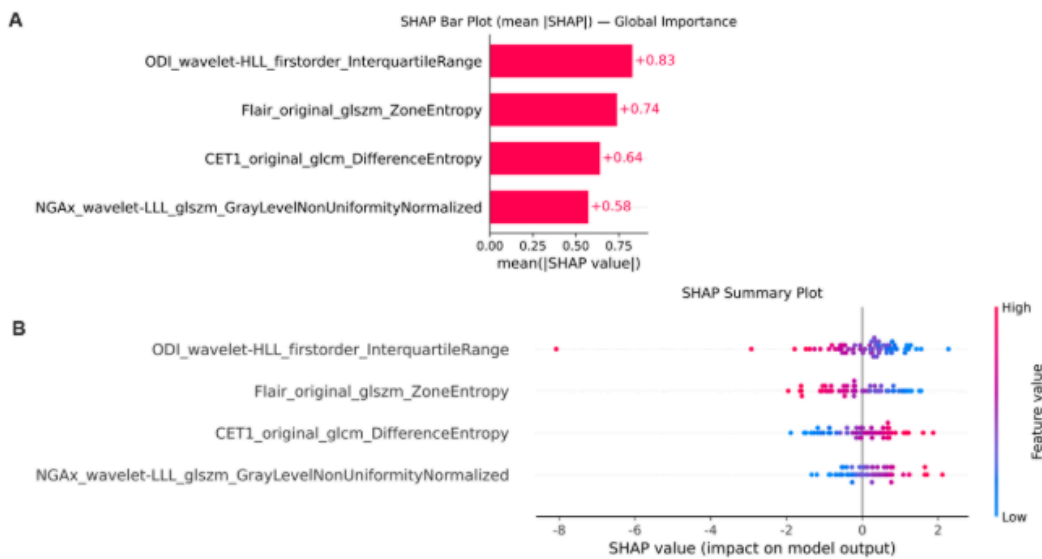


Figure 5

SHAP-based interpretation of the ABV fusion model. (A) SHAP bar plot showing the global importance of the selected radiomics features in the ABV fusion model, ranked by mean absolute SHAP value. (B) SHAP summary plot showing the distribution and direction of each feature's contribution to model output across patients. Positive SHAP values indicate that a feature increases the model output toward glioblastoma, whereas negative SHAP values indicate that a feature decreases the model output and shifts the prediction toward solitary brain metastasis. Feature color indicates the original feature value, ranging from low to high.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial.docx](#)