Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Spatially Weighted Patient Similarity Graph Diffusion Radiomics Framework for Preoperative Prediction of Short- and Long-Term Survival in Glioblastoma

Zeyuan Li
Shandong Second Medical University
Mengqi Wang
Shandong Second Medical University
Xun Sun
Shandong Second Medical University
Yushi Zhang
Shandong Second Medical University
Kun Han
Shandong Second Medical University
Yue Huang
Shandong Second Medical University
Na Sun
Shandong Second Medical University
Xiao Qi
Shandong Second Medical University
Suzhen Wang
wangsz@sdsmu.edu.cn

Shandong Second Medical University

Research Article

Keywords: glioblastoma, radiomics, graph diffusion, patient similarity graph, survival risk stratification

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11496684/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Spatially Weighted Patient Similarity Graph Diffusion Radiomics Framework for Preoperative Prediction of Short- and Long-Term Survival in Glioblastoma

Ze-yuan Li[1], Meng-qi Wang[1], Xun Sun[1], Yu-shi Zhang[1], Kun Han[1], Yue Huang[1], Na Sun[1], Xiao Qi[1]*, Su-zhen Wang[1]*

[1]Department of Health Statistics, School of Public Health, Shandong Second Medical University, No. 7166 Baotong West Street, Weicheng, Weifang 261053, Shandong, China

*Correspondence: Su-zhen Wang (e-mail: wangsz@sdsmu.edu.cn) , Xiao Qi(e-mail: qixiao@sdsmu.edu.cn)

Abstract

Background: This study aimed to develop and validate a spatially weighted patient similarity graph diffusion radiomics framework integrated with age for prediction of short-term versus long-term survival in glioblastoma (GBM).

Methods: This study included 234 GBM patients from BraTS 2020 (116 short-term and 118 long-term survivors, defined by a 365-day threshold) and 146 patients from UPenn-GBM for external validation. Radiomic and spatial features were extracted from whole-tumor regions on four MRI sequences (T1, T1ce, T2, and FLAIR). Following unsupervised feature filtering, a spatially weighted patient similarity graph was constructed using a radial basis function (RBF)kernel. Single-step, nonparametric graph diffusion was applied to incorporate inter-patient relationships into imaging feature representations. The top 35 imaging features ranked by Elastic Net were combined with age and entered into a logistic regression classifier for short-term survival prediction. Predicted probabilities of short-term survival were evaluated using receiver operating characteristic analysis, Kaplan–Meier survival analysis, decision curve analysis, and SHAP interpretation.

Results: The proposed model achieved an AUC of 0.736 (95% CI: 0.670–0.800, $P < 0.001$) in the internal cross-validation and 0.778 (95% CI: 0.699– 0.851, $P < 0.001$) in the external validation, demonstrating significant risk stratification capabilities in both cohorts. The model achieved a Harrell's C-index of 0.687 in the external validation. SHAP analysis showed that age was the most influential predictor, followed by the

FLAIR-derived original GLRLM feature RunLengthNonUniformity and wavelet-based texture features from multiple sequences, with high-ranked imaging predictors being derived from all four MRI sequences and mainly related to texture.

Conclusions: The proposed framework integrates multimodal MRI radiomics, spatially weighted patient similarity, and age to support preoperative survival risk stratification in GBM. Its lightweight, nonparametric graph diffusion approach offers an interpretable alternative to deep graph neural networks, particularly for limited-sample medical imaging datasets.

Key words: glioblastoma; radiomics; graph diffusion; patient similarity graph; survival risk stratification

1. Introduction

Glioblastoma (GBM) is the most common primary malignant brain tumor in adults, accounting for approximately 70% of all newly diagnosed gliomas[1]. Despite advances in medical technology, the overall median survival (OS) is only about 15 months, and individual prognosis varies significantly [2-4]. Therefore, accurate preoperative identification of high-risk patients is essential for individualized treatment planning and for improving patients' quality of life. However, diagnosis for patients in preoperative is usually invasive and traumatic. On the other hand, Some important prognostic factors, particularly extent of resection, are only determined after surgery. Although KPS can be assessed preoperatively, it is not always available at the time of imaging-based decision-making [5].

Magnetic resonance imaging (MRI) is the standard, non-invasive diagnostic tool for patients with suspected brain tumors, providing comprehensive structural and complementary anatomical and tissue-contrast information into tumor location, morphology, and adjacent tissues[6]. However, conventional MRI interpretation relies on subjective visual assessment and may overlook subtle microstructural heterogeneity that carries important prognostic information. This leads to limited diagnostic accuracy, especially in early lesion detection, differential diagnosis of benign and malignant lesions, and prediction of molecular phenotypes. Radiomics is a high-throughput quantitative image analysis method, which can help to extract a large number of image features from medical scanning images, often identify subtle tumor features that cannot be recognized by human eyes[7-9], and can transform digitally encrypted medical images containing information related to tumor pathophysiology into mining high-dimensional data[10]. Prior research has indicated that these radiomic features may quantitatively describe the microscopic histopathological characteristics of

glioblastoma multiform from a macroscopic imaging perspective, and suggest a significant correlation with overall survival of patients[11-13].

Existing prognostic models based on radiomics mainly treat patients as independent samples and ignore the potential relationships between them[7,14,15]. From a clinical perspective, however, physicians often assess a patient's prognosis by referring to similar cases encountered in their prior clinical experience. For example, two patients sharing similar imaging phenotypes and tumor locations may exhibit similar biological behaviors and clinical trajectories. In addition, Graph Convolutional Networks (GCNs) have been initially applied to medical prognostic prediction by integrating imaging and clinical features, demonstrating the capability of graph-based patient population modeling[16]. However, deep graph learning architecture introduces a large number of trainable parameters, which easily leads to the risk of overfitting of small samples and high-dimensional data that often exists in medical radiomics research[17]. Consequently, how to effectively utilize inter-patient relational structure without relying on heavily parameterized architecture remains an urgent and unresolved challenge in the field.

We hypothesized that patient similarity, explicitly modeled by a spatially weighted patient similarity graph structure and a nonparametric graph diffusion without trainable diffusion parameters, could reveal potential prognostic information beyond individual radiomic features[18]. Based on this, we developed a population-aware graph diffusion radiomics framework for preoperative survival risk stratification in GBM patients. The model was internally validated on the BraTS 2020 dataset[19-21] using 5-fold cross-validation and externally validated on the independent UPenn-GBM cohort[22]. This study is expected to help identify GBM patients at high risk of early death, potentially facilitating individualized treatment planning and optimized follow-up management.

2. Materials and methods

2.1 Data sources and preprocessing

This study used two publicly available GBM MRI datasets:

The model development cohort was sourced from the training set of the MICCAI BraTS 2020 multimodal brain tumor dataset, which originally comprised 369 patients, among whom 236 patients had publicly available overall survival labels and comprised the official survival prediction subset, 2 patients were excluded because of incomplete feature extraction or unmatched identifiers. Therefore, 234 patients were finally included in the model development cohort, possessing complete multimodal MRI scans

(including four modalities: T1, T2, FLAIR, and T1ce), tumor segmentation labels, and survival records in days. Because explicit censoring information was not available in the dataset, we assumed that the recorded survival times were representative of observed events. According to the official survival time labels, a 365-day threshold was used to dichotomize the cohort into a short-term survival group (overall survival <365 days, n=116) and a long-term survival group (overall survival ≥365 days, n=118) to construct a logistic regression model.

The external validation dataset was derived from the UPenn-GBM public dataset, comprising 146 patients with complete four-modality MRI, tumor segmentation labels, and survival information. All patients had documented survival status, allowing deaths to be distinguished from censored observations. The same 365-day threshold was applied to define the long/short-term survival labels.

For all MRI images, the enhancing tumor (label 4), necrotic/non-enhancing tumor core (label 1), and peritumoral edema (label 2) were combined to construct the whole tumor (WT) region of interest (ROI) based on the official BraTS segmentation labels. Subsequently, the voxel intensities within the WT mask for each modality were standardized using Z-score normalization to eliminate intensity variations caused by different scanners and acquisition parameters. The study flowchart is shown in Figure 1.

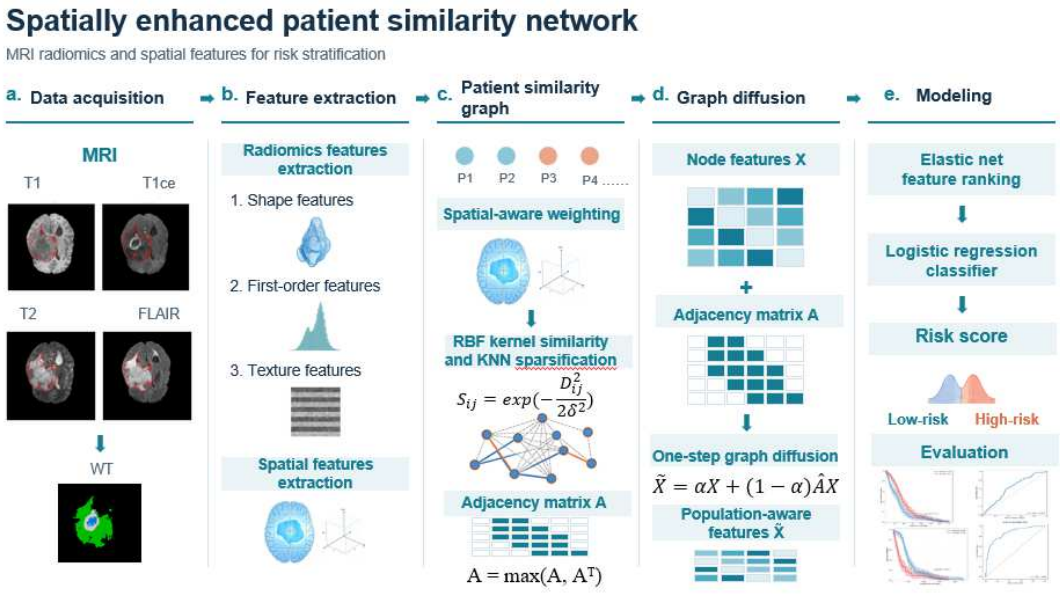


Figure 1. Workflow of study development and evaluation.

(a) Data acquisition: Multimodal MRI and tumor segmentation masks from the BraTS and UPenn-GBM datasets. (b) Feature extraction: Radiomic and spatial features

extracted from the whole-tumor region. (c) Patient similarity graph: A spatially weighted RBF -KNN graph was constructed ($k = 10$, $\lambda = 0.7$). (d) Graph diffusion: Single-step nonparametric graph diffusion without trainable diffusion parameters ($\alpha = 0.6$) generated population-aware feature representations. (e) Modeling: Elastic Net feature selection, Logistic Regression classification, and internal 5-fold cross-validation with external validation on UPenn-GBM.

2.2 Feature extraction

Multimodal radiomics feature extraction: The PyRadiomics package (v.3.1.0) was employed to extract radiomic features from the WT regions across the four MRI modalities. Feature categories included shape, first-order statistics, gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), gray-level dependence matrix (GLDM), and neighboring gray tone difference matrix (NGTDM). Approximately 800–1000 features were extracted per modality, yielding a total of about 3400 radiomic features.

Spatial feature extraction: All MRI images were preprocessed using the standard BraTS pipeline and co-registered to the same anatomical template. In this study, the voxel coordinates of the tumor segmentation masks were transformed into the physical coordinate system of the common atlas space using the NIfTI affine transformation to calculate the spatial location and morphological features of each patient. These included: tumor centroid coordinates (x, y, z; in mm), physical range along each axis (range_x, range_y, range_z), tumor volume (computed based on voxel volume), hemispheric laterality (left/right/bilateral), midline crossing status (binary), and a deep location score. A total of 15 spatial features were extracted. As spatial features have been proven effective for survival prediction tasks, they were concatenated with the radiomic features to form the initial feature.

2.3 Unsupervised feature pre-filtering

The initial feature matrix underwent unsupervised pre-filtering without any exposure to the survival labels. The procedures were as follows: (a) features with a missing rate >30% were excluded, and the remaining missing values were imputed using the median; (b) features with a variance <0.01 were eliminated; (c) Pearson correlation coefficients were computed for all pairs of features. For any pair with an absolute correlation coefficient exceeding 0.95, the feature with the higher sum of absolute correlations with all other features was removed, as it could be more effectively represented by the remaining feature set. Following pre-filtering, the feature

dimension was reduced to approximately 671. It should be noted that age is not included in unsupervised filtering, graph construction, or feature ranking. It was normalized and added to the selected 35 image features only at the final classifier input.

2.4 Spatially weighted patient similarity graph construction

Each patient was treated as a node in an undirected graph G=(V,ε), The edges of the graph were constructed based on the similarity of integrated multimodal radiomic and spatial feature vectors, explicitly modeling patient-to-patient phenotypic relationships.

Spatially weighted mechanism: Before distance calculation, the spatial features were multiplied by a weighted factor $\lambda_s = 0.7$, This weighted factor is used to adjust the relative contribution of spatial location information and image texture information in the similarity calculation, so as to avoid any kind of features from overly dominating the construction of the patient similarity map due to the scale difference, so as to improve the robustness of the similarity measure. The value of λ_s was selected among $\{0.5, 0.7, 1.0, 1.5\}$ by cross validation, and 0.7 was finally determined as the optimal value.

Similarity kernel selection: A radial basis function (RBF) kernel was adopted to calculate nonlinear patient similarity. For feature vectors x_i, x_j, the similarity was defined as:

$$S_{ij} = \exp(-\frac{D_{ij}^2}{2\delta^2})$$

where $D_{ij} = ||x_i - x_j||_2$ is the Euclidean distance, with the kernel bandwidth δ set to the median of all non-zero elements in the distance matrix.

Adjacency matrix construction: For each node, the k=10 nearest neighbors with the highest similarities were retained, forming a sparse weighted adjacency matrix A, if patient j was among the selected neighbors of patient i, is $A_{ij} = S_{ij}$,or $A_{ij} = 0$ .The matrix was then symmetrized via:

$$A \leftarrow max(A, A^T)$$

to ensure an undirected graph structure. This weighted construction preserved the continuous strength of inter-patient similarity, allowing for more subtle information dissemination during graph diffusion.

2.5 Graph diffusion and population-aware radiomics

Graph diffusion represents the core methodological innovation of this study, aiming to fuse the phenotypic expressions of individual patients with information from similar populations through neighbor-to-neighbor feature propagation, thereby achieving population-aware risk stratification. Graph construction was based only on imaging and spatial features. Age was not included in graph building or graph diffusion and was added only at the final classifier input stage. It is added only at the final classification input stage.

On the training set, given the weighted adjacency matrix A and the node feature matrix X, a row-normalized transition matrix $\hat{A} = D^{-1}A$ was first constructed, where $D_{ii} = \Sigma_j A_{ij}$. The single-step graph diffusion was defined as:

$$\tilde{X} = \alpha X + (1-\alpha)\hat{A}X, \alpha \in [0,1]$$

where $\alpha = 0.6$ controls the trade-off between the original features and the neighbors' features. This diffusion process involves no learnable parameters, inherently avoiding the overfitting risks associated with deep graph neural networks on small-sample datasets.

For patients in the test set, an inductive residual propagation strategy was applied. For a test node u, its diffused feature $\tilde{x}_u$ and the aggregated diffused features of its selected training-set neighbors:

$$\tilde{x}_u = \beta x_u + (1-\beta)\sum_{j \in N_u} w_{uj}\tilde{x}_j$$

where N_u represents the k nearest neighbors selected from the training set based on RBF similarity, w_{uj} denotes the normalized similarity weight, and $\beta = 0.5$. This design ensures that test samples also acquire population context while completely remaining blinded to the test set labels or other test samples, strictly preventing information leakage.

2.6 Elastic Net feature ranking and Logistic Regression prognostic classification

Following graph diffusion, each patient obtained a population-aware feature vector $\tilde{X}$. To mitigate the overfitting risks posed by high dimensionality and small sample sizes, an Elastic Net logistic regression was employed only for feature ranking within the internal training set. The Elastic Net objective function was defined as:

$$\min_{w,b} \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{y}_i) + \lambda\left[\rho||w||_1 + (1-\rho)\frac{1}{2}||w||_2^2\right]$$

Where $L(y_i, \hat{y}_i)$ denotes the logistic (cross-entropy) loss, $\rho = 0.5$ to balance the L1 and L2 penalties, and α is the regularization strength. In the implementation, $\alpha = 1/C$, and the regularization parameter was set to $C = 0.5(l1_ratio = 0.5)$. After model fitting, the absolute regression coefficients were used to rank all diffusion-based imaging features, and the top *N*=35 features were retained for subsequent modeling.

The selected imaging features and the per-fold standardized age were concatenated as the final model input. A logistic regression classifier (L2 regularization, C=0.15, solver=lbfgs) was then used to predict the probability of short-term survival. No additional class-imbalance correction was applied, since the short-term and long-term survival groups were approximately balanced in both cohorts.

The final model aimed to predict the probability of short-term survival (overall survival, < 365 days) for each patient. This continuous predicted probability was then used directly for survival stratification.

2.7 Model evaluation and statistical analysis

All supervised modeling steps were performed within each cross-validation fold; unsupervised feature filtering was performed on the complete dataset before cross-validation without using any survival labels. The final model established on the BraTS training set, including the feature scaler, age scaler, feature ranking indices, and the trained logistic regression classifier, was directly applied to the UPenn-GBM dataset without any retraining or parameter fine-tuning. Age values were standardized using the age scaler derived from the BraTS training set and then concatenated with the selected imaging features before model prediction.

Classification performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, and F1-score. $\bar{X} \pm S$ of the five-fold AUC was reported to assess model stability, and bootstrap resampling (2000 iterations with replacement) was used to estimate 95% confidence intervals for the AUC in the external validation cohort.

The predicted probability of short-term survival (<365 days) was used as the model-derived prognostic probability for survival stratification. Patients were stratified into high-risk and low-risk groups based on the median of this predicted probability, and Kaplan-Meier survival curves were plotted. The log-rank test was used to assess the significance of survival differences between the two groups. Because the BraTS dataset lacks censoring information, we assumed that all patients had experienced the

death event at the last follow-up.

In the UPenn cohort, the locked model was applied without any retraining or parameter fitting. Model discrimination was evaluated using the area under the receiver operating characteristic curve (AUC) and Harrell's C-index. Decision curve analysis was additionally performed to assess the potential clinical utility of the predicted short-term survival probability.

We calculated SHAP values based on the selected graph-diffused radiomic features to evaluate their contribution to model predictions, and the retention frequency of each imaging feature across five folds was calculated to assess feature stability. A two-sided $P < 0.05$ was considered statistically significant.

3.Result

3.1 Patient Baseline Characteristics

Baseline characteristics of the study population in the training and external validation test datasets are summarized in Table 1. In the BraTS training set, 236 patients possessed survival labels. After radiomic feature extraction and data cleaning, 234 patients were finally included in the model for training and internal cross-validation with a mean age of 61.2 ± 11.9 years. The median overall survival (OS) was 369 days, comprising 116 patients (49.6%) in the short-term survival group and 118 patients (50.4%) in the long-term survival group. The UPenn external validation set included 146 patients with a mean age of 62.3 ± 12.0 years and a median OS of 385 days, of which 68 patients (46.6%) were in the short-term survival group and 78 patients (53.4%) were in the long-term survival group. The distribution of survival labels was largely comparable between the two cohorts.

Table 1

Baseline characteristics of the BraTS training cohort and the UPenn external validation set.

Variable	BraTS training set (n=234)	UPenn external validation set (n=146)
Age (years)	61.2 ± 11.9	62.3 ± 12.0
Gender, n (%)		
Male	N/A	88 (60.3%)
Female	N/A	58 (39.7%)
Median overall survival (days)	369	385

Variable	BraTS training set (n=234)	UPenn external validation set (n=146)
Survival label, n (%)		
Short-term survival (<365 days)	116 (49.6%)	68 (46.6%)
Long-term survival (≥365 days)	118 (50.4%)	78 (53.4%)

Data are presented as $\bar{X} \pm S$ deviation or number of patients (percentage), unless otherwise indicated.

Note: The BraTS 2020 dataset contained 236 patients with survival labels; 2 patients were excluded due to missing data during the merging process, resulting in a final analytical sample of 234 patients. The BraTS dataset does not provide explicit censoring information; all patients were assumed to have experienced the event of death at the recorded overall survival time for the purpose of Kaplan–Meier analysis.

3.2 Model Classification Performance and Internal Cross-Validation

In the 5-fold cross-validation of the 234 BraTS patients, our final model achieved a Mean Fold AUC of 0.740 ± 0.029, an accuracy of 0.658, a sensitivity of 0.664, a specificity of 0.652, and an F1-score of 0.658 (Table 2). Kaplan-Meier survival analysis demonstrated that the model exhibited significant risk stratification capabilities (log-rank P < 0.001, Figure. 2B). The corresponding receiver operating characteristic (ROC) curve is shown in Figure. 2A.

Table 2

Model performance in internal cross-validation and external validation.

Performance metric	Internal CV (BraTS, n=234)	External validation (UPenn, n=146)
Mean Fold AUC	0.740 ± 0.029	—
AUC	0.736(0.670-0.800)	0.778(0.699-0.851)
Accuracy	0.658	0.691
Sensitivity	0.664	0.691
Specificity	0.652	0.692
F1-score	0.658	0.676
log-rank P	<0.001	<0.001
Harrell's C-index	—	0.687

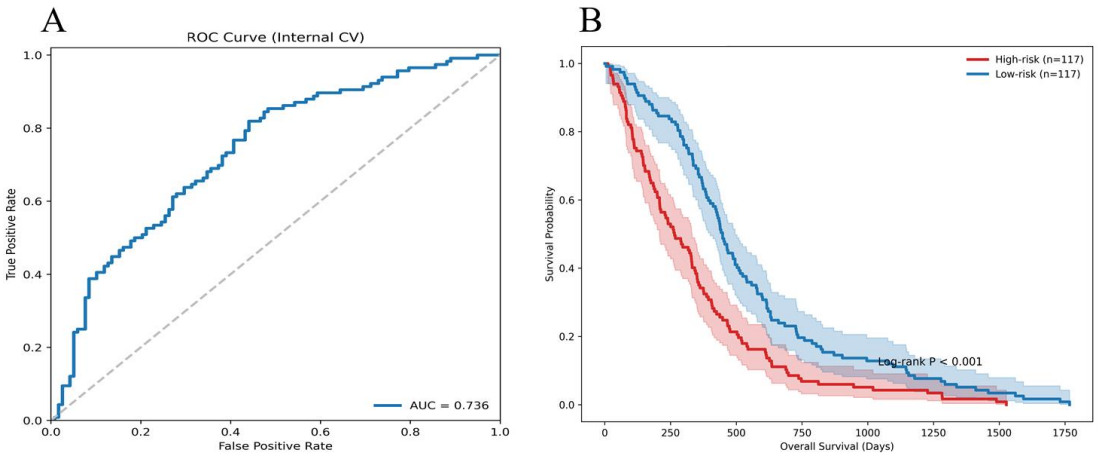


Figure 2. Model performance in the internal cross-validation (BraTS, n=234).

(A) Receiver operating characteristic curve of the proposed model. The shaded area represents the 95% confidence interval. (B) Kaplan-Meier survival curves of patients stratified into high-risk and low-risk groups using the median model-predicted short-term survival probability. The log-rank test was used to compare survival differences between the two groups.

3.3 Ablation Study

Table 3 summarizes the ablation results. Compared with the radiomics-only baseline (M0), adding age (M1) increased the AUC from 0.674 to 0.706 and improved the risk stratification significance (log-rank P from 0.008 to 0.001), indicating that age provided additional prognostic information at the individual level. Graph module alone (M2) had a slightly higher AUC (0.687) than M0, while maintaining significant survival stratification (log-rank P = 0.003).

The final model (M3), which combined graph-diffused radiomic features and age, achieved the best performance (AUC = 0.736; mean fold AUC = 0.740 ± 0.029; log-rank P < 0.001). These results suggest that graph diffusion and age were complementary and that the most robust risk stratification can be achieved by integrating them.

Table 3

Results of ablation experiments in the internal 5-fold cross-validation on the BraTS training cohort (n = 234)

Model	M0	M1	M2	M3
Spatially Weighted	×	×	√	√

Graph Construction	×	×	√	√
Graph Diffusion	×	×	√	√
Age	×	√	×	√
AUC	0.674	0.706	0.686	0.736
Mean Fold AUC	0.678 ± 0.016	0.710 ± 0.025	0.689 ± 0.045	0.740 ± 0.029
log-rank P	0.008	0.001	0.003	< 0.001

Point estimates from 5-fold cross-validation. √ = included, × = excluded. P values from log-rank test.

Abbreviations: AUC, area under the receiver operating characteristic curve.

3.4 External Validation

The final model locked on the BraTS training set was directly applied to the UPenn external validation cohort and yielded an AUC of 0.778, an accuracy of 0.691, a sensitivity of 0.691, and a specificity of 0.692 (Table 2). Bootstrap resampling (2000 iterations) yielded a 95% CI of 0.699-0.851 for the external AUC, indicating highly stable predictive performance on external set. Kaplan-Meier analysis demonstrated that the model could significantly stratify UPenn patients into high-risk and low-risk groups (log-rank $P < 0.001$, Figure. 3).

Decision curve analysis showed that across threshold probabilities from 0.20 to 0.60, the model yielded a higher net benefit than the treat-all strategy (Table 4, Figure 4). At a threshold of 0.20, the model achieved a net benefit of 0.339, compared with 0.332 for treating all patients. The advantage became more evident at higher thresholds: at a threshold of 0.50, the net benefit of the model was 0.158, whereas the treat-all strategy produced a negative net benefit of −0.069. These results suggest that the model-derived short-term survival probability could help identify patients at sufficiently high risk to warrant more intensive clinical management, while avoiding unnecessary treatment in patients with lower predicted risk.

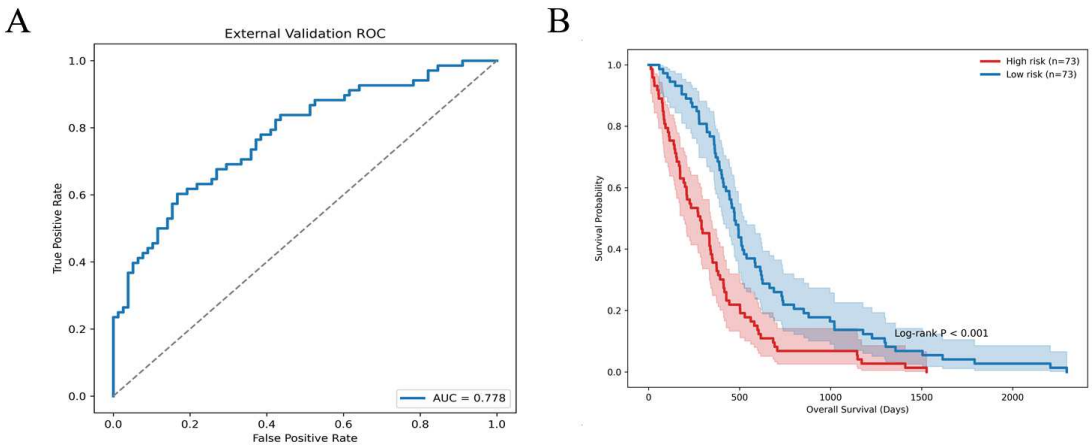


Figure 3. Model performance in the external validation (UPenn, n=146).

(A) Receiver operating characteristic curve of the proposed model. The shaded area represents the 95% confidence interval obtained via bootstrap resampling (2000 iterations). (B) Kaplan-Meier survival curves of patients stratified into high-risk and low-risk groups using the median model-predicted short-term survival probability. The log-rank test was used to compare survival differences between the two groups.

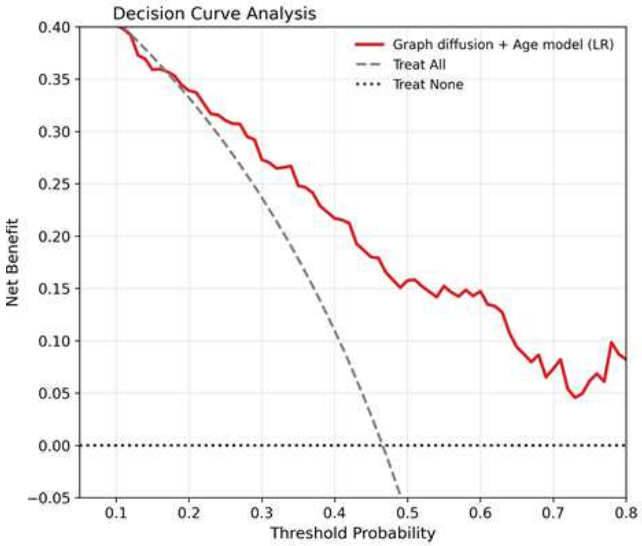


Figure 4. Decision curve analysis of the proposed model in the external validation (UPenn, n=146).

The net benefit is plotted against a range of threshold probabilities. The horizontal line at zero represents the strategy of treating no patients, and the gray curve represents the strategy of treating all patients. The model provides a higher net benefit compared with these alternative strategies across a wide range of clinically relevant thresholds, indicating its potential clinical utility.

Table 4

Decision curve analysis-based net benefits at selected threshold probabilities.

Threshold probability	Model net benefit	Treat-all net benefit	Treat-none net benefit	Advantage (Model – Treat-all)
0.10	0.402	0.406	0.000	-0.004
0.20	0.339	0.332	0.000	0.007
0.30	0.273	0.237	0.000	0.036
0.40	0.217	0.110	0.000	0.107
0.50	0.158	-0.069	0.000	0.227
0.60	0.143	-0.303	0.000	0.446

Net benefit values were derived from decision curve analysis in the UPenn external validation cohort (n=146). The model-derived short-term survival probability was used as the risk estimate. The treat-none strategy has a net benefit of 0 at all thresholds. Abbreviation: NB, net benefit.

3.5 SHAP Feature Importance and Feature Stability

The final model input consisted of 35 graph-diffused imaging features selected by Elastic Net ranking and 1 age variable. SHAP analysis showed that age was the most influential predictor, followed by the FLAIR-derived original GLRLM feature RunLengthNonUniformity, the shape feature Maximum2DDiameterColumn, and several wavelet-based texture features (Figure 5B). Highly ranked imaging predictors were derived from all four MRI sequences, including FLAIR, T2, T1, and T1ce. Wavelet-transformed texture features, particularly those based on GLSZM, GLRLM, and GLDM, accounted for a substantial proportion of the highly ranked imaging predictors, indicating that the model captured heterogeneous spatial and gray-level patterns within the tumor. Feature stability analysis demonstrated that most of the displayed features were consistently selected across the five cross-validation folds, with several features retained in all five folds and others retained in four or three folds (Figure 5A), supporting the stability of the selected feature set. The feature heatmap further illustrated heterogeneous patterns of the selected graph-diffused features across patients ordered by the predicted short-term survival probability (Figure 5C), suggesting that the final model integrated complementary information from multiple MRI sequences together with age.

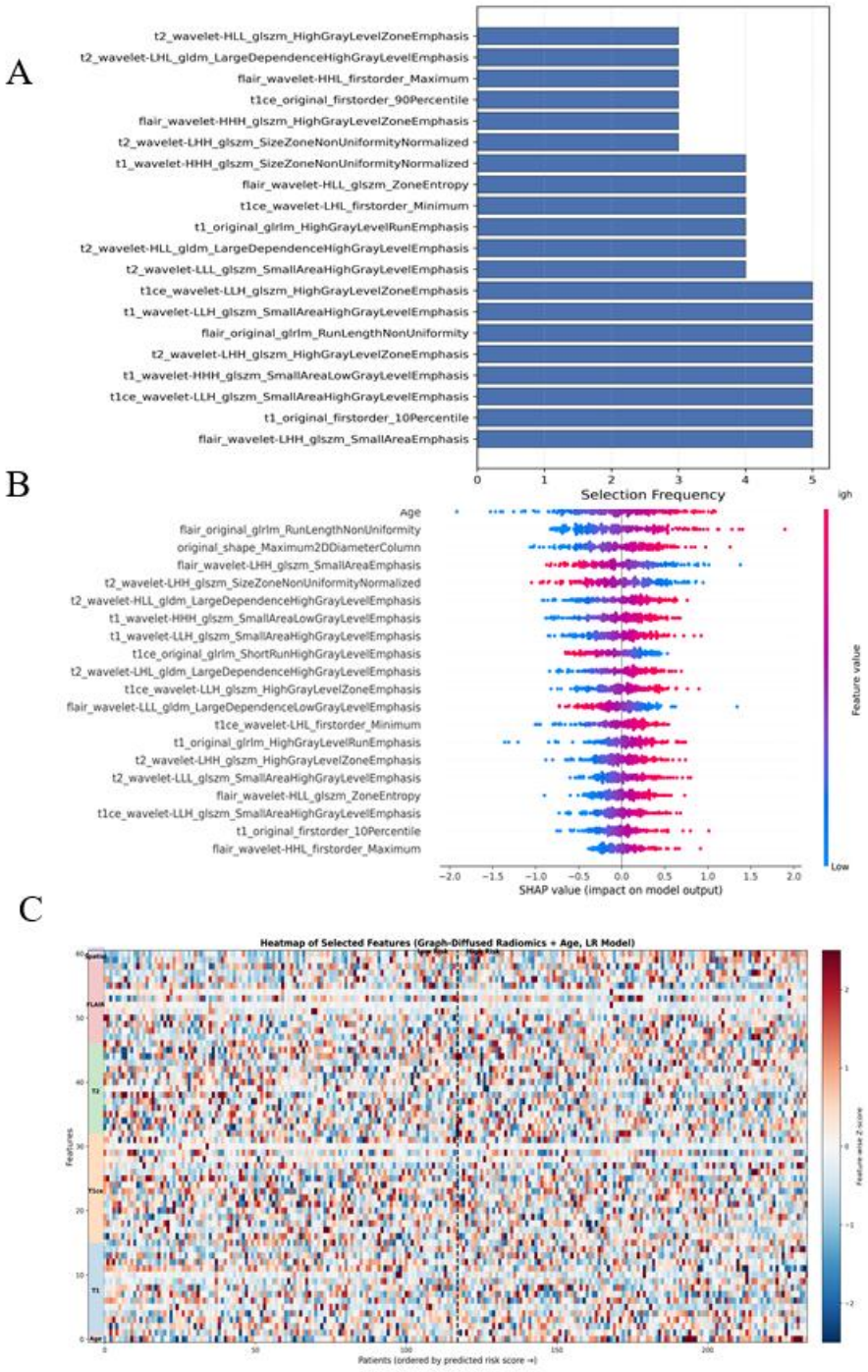


Figure 5. **Feature selection stability and SHAP-based interpretation of graph-diffused radiomic features.**

(A) Feature selection stability across the five cross-validation folds. The bar chart displays the frequency with which each feature was retained by Elastic Net regularization, with a maximum score of 5 indicating consistent selection in all folds. (B) SHAP beeswarm summary plot illustrating the distribution of SHAP values across all patients for each feature. Each point represents a single patient; red indicates high

feature values and blue indicates low feature values. Positive SHAP values drive the prediction toward short-term survival (<365 days). (C) Heatmap of the selected graph-diffused features across patients ordered according to their predicted short-term survival probability. Features are grouped according to their imaging modality and spatial origin.

4.Discussion

Radiomics has emerged as a promising prognosis approach for GBM because quantitative MRI features can capture intratumoral heterogeneity that may not be fully appreciated through traditional visual assessment[23,24]. Previous radiomics-based studies of glioblastoma have mainly focused on extracting high-dimensional imaging features and combining them with molecular or clinical variables, where each patient is usually treated as an independent sample regardless of the potential relationships between patients[25,26]. A recent meta-analysis of 17 studies reported a pooled C-index of 0.65(95% CI, 0.58-0.72) for MR-based radiomics in predicting overall survival [13]. In comparison, the proposed framework explicitly models patient-to-patient phenotypic similarity through a weighted graph and propagates imaging information across similar patients through a non-parametric graph diffusion mechanism. Using only conventional structural MRI and age, the model achieved an internal AUC of 0.736 and an external Harrell's C-index of 0.687 in an independent UPenn cohort. These results are comparable to the pooled radiomics value, but our model was developed using only conventional structural MRI and age, without requiring postoperative molecular or detailed clinical information. The independent external validation further supports the potential generalizability of the proposed framework. This supports the potential value of population-aware graph-based representation in improving the generalizability of radiomics models.

Ablation experiments revealed two complementary contributions. Compared with the radiomics baseline alone, adding age alone increased the AUC from 0.674 to 0.706 and was associated with more significant survival stratification (log-rank P from 0.008 to 0.001), confirming the prognostic importance of this established clinical factor. Graph module further improved the AUC to 0.686 while maintaining significant risk stratification (log-rank P = 0.003), suggesting that population-level information provides complementary value beyond individual imaging features[18]. The final model combining graph diffusion and age achieved the best overall performance (AUC = 0.736; log-rank P < 0.001), indicating that graph-derived population information and individual clinical priors are indeed complementary. These findings support the notion that inter-patient relational modeling, rather than simply increasing model complexity,

can enhance preoperative risk stratification in glioblastoma.

SHAP analysis further revealed the contribution of clinical and imaging features to model predictions, with age being the most influential predictor, consistent with its established prognostic role in glioblastoma[27]. Among the imaging features, the FLAIR-derived original GLRLM feature RunLengthNonUniformity and the shape feature Maximum2DDiameterColumn ranked highly, along with a number of wavelet-transformed texture features based on GLSZM, GLRLM, and GLDM. Highly ranked imaging predictors were derived from all four MRI sequences, including FLAIR, T2, T1, and T1ce. These features characterize spatial and gray-level patterns within the tumor and may capture aspects of intratumoral heterogeneity that are difficult to appreciate through conventional visual assessment. The observed contribution of heterogeneous radiomic features further supports the potential value of MRI radiomics for preoperative survival risk stratification in glioblastoma[28-30].

External validation remains a major challenge for medical artificial intelligence due to differences in patient populations and imaging acquisition protocols at different institutions can greatly affect model performance [31,32]. Despite these variations, our framework maintained significant prognostic stratification capability in the independent UPenn cohort. Interestingly, the external validation performance was numerically higher than the internal cross-validation estimate, which may reflect differences in cohort composition, sample size, and imaging heterogeneity rather than improved model performance alone. Nevertheless, the consistent separation of high- and low-risk groups across datasets supports the potential generalizability of the proposed approach.

Several limitations should be acknowledged. First, because explicit censoring information was not available in the BraTS dataset, we performed survival analyses under the assumption that recorded survival times were representative of observed events; Second, both cohorts were derived primarily from Western populations, and further validation in more diverse populations is required; Third, the current framework relies derived features, and the integration of molecular biomarkers such as IDH mutations and MGMT promoter methylation status may further improve prognostic accuracy [33,34]. Fourth, graph construction parameters, including spatially weighted and diffusion coefficients, were empirically determined and require systematic optimization in future studies[35]. Finally, prospective multicenter validation using standardized imaging protocols and clinical workflows is required before clinical translation.

In conclusion, this study demonstrates that modeling patient similarity relationships through a spatially weighted graph diffusion framework can improve population-level survival stratification in GBM. The combination of image-derived imaging findings and age provides a lightweight, interpretable, and preliminary external generalizability method that may support preoperative risk assessment and individualized treatment planning.

5.Conclusions

This study presents a spatially weighted patient similarity graph diffusion radiomics framework integrating age as a clinical covariate for preoperative survival risk stratification in GBM. By constructing patient similarity maps and performing non-parametric graph diffusion, the framework extends traditional radiomics from the individual level to the population-aware representation. This lightweight graph diffusion strategy does not require additional trainable parameters, and is especially suitable for the analysis of small samples and high-dimensional medical image data. In the future, larger cohorts and multicenter prospective data are needed to further verify its clinical translation value.

List of abbreviations

GBM	Glioblastoma
MRI	Magnetic resonance imaging
OS	Overall survival
WT	Whole tumor
ROI	Region of interest
RBF	Radial basis function
KNN	K-nearest neighbors
AUC	Area under the receiver operating characteristic curve
SHAP	SHapley Additive exPlanations
DCA	Decision curve analysis
GLRLM	Gray-level run length matrix

GLCM	Gray-level co-occurrence matrix
GLSZM	Gray-level size zone matrix
GLDM	Gray-level dependence matrix
NGTDM	Neighboring gray tone difference matrix

Declarations

Ethics approval and consent to participate

This study was a analysis of existing, de-identified neuroimaging and clinical data from the BraTS 2020 and UPenn-GBM datasets. No new patient recruitment, direct patient contact, or additional data collection was involved. The original UPenn-GBM study was approved by the Institutional Review Board of the University of Pennsylvania Health System, and informed consent was obtained from the participants, as reported in the original dataset publication. The BraTS 2020 dataset was previously approved by the institutional review boards of the contributing institutions and all patient identifiers were removed prior to public release. Consent to participate was obtained by the original data collectors at the contributing institutions.

Consent for publication

Not applicable. This study uses publicly available anonymized data with no identifiable patient information.

Availability of data and materials

All data used in this study are publicly available and were obtained from two open-access datasets.

The Brain Tumor Segmentation (BraTS) 2020 dataset used in this study is publicly available through registration at the CBICA Image Processing Portal (https://ipp.cbica.upenn.edu). The dataset can be accessed following the data request procedure described at https://www.med.upenn.edu/cbica/brats2020/registration.html.

The UPENN-GBM dataset can be accessed via The Cancer Imaging Archive (TCIA) at https://www.cancerimagingarchive.net/collections/. No proprietary or patient-identifiable data were used. All analyses were conducted on de-identified data and no additional restrictions were applied to data access.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China [grant number 81872719 and T2600463] and the Natural Science Foundation of Shandong Province, China [grant numbers ZR2019MH034].

Author contributions

All authors contributed to the conception and design of the study. Z.L contributed to data collection and preprocessing, statistical analysis, and drafting of the original manuscript. M.W, X.S, Y.Z, K.H, Y.H, N.S contributed to interpretation of findings and critical revision of the manuscript. S.W and X.Q supervised the study, contributed to the study conception , provided methodological guidance, and critically reviewed and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Ostrom QT, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014-2018. Neuro Oncol. 2021;23(12 Suppl 2):iii1-iii105.
2. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med. 2005;352(10):987-96.
3. Tykocki T, Eltayeb M. Ten-year survival in glioblastoma. A systematic review. J Clin Neurosci. 2018;54:7-13.
4. Yang J, Cai H, Zhang L, Alafate W, Xi S, Du J, et al. A nomogram for predicting the diagnosis of central malignant tumors based on preoperative clinical characteristics and laboratory indicators: a diagnostic study. Int J Surg. 2025;111(11):7583-93.

5. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. J Neurosurg. 2001;95(2):190-8.

6. Ellingson BM, Wen PY, Cloughesy TF. Modified Criteria for Radiographic Response Assessment in Glioblastoma Clinical Trials. Neurotherapeutics. 2017;14(2):307-20.

7. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48(4):441-6.

8. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017;77(21):e104-e7.

9. Zwanenburg A, Vallières M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;295(2):328-38.

10. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-62.

11. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

12. Kickingereder P, Burth S, Wick A, Götz M, Eidel O, Schlemmer HP, et al. Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models. Radiology. 2016;280(3):880-9.

13. Choi Y, Jang J, Kim BS, Ahn KJ. Pretreatment MR-based radiomics in patients with glioblastoma: A systematic review and meta-analysis of prognostic endpoints. Eur J Radiol. 2023;168:111130.

14. Pai S, Bader GD. Patient Similarity Networks for Precision Medicine. J Mol Biol. 2018;430(18 Pt A):2924-38.

15. Chaddad A, Kucharczyk MJ, Daniel P, Sabri S, Jean-Claude BJ, Niazi T, et al. Radiomics in Glioblastoma: Current Status and Challenges Facing Clinical

Implementation. Front Oncol. 2019;9:374.

16. Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. International Conference on Learning Representations (ICLR). 2017.

17. Zhang S, Tong H, Xu J, Maciejewski R. Graph convolutional networks: a comprehensive review. Comput Soc Netw. 2019;6(1):11.

18. Zhou D, Bousquet O, Lal TN, Weston J, Schölkopf B. Learning with local and global consistency. Advances in Neural Information Processing Systems (NeurIPS). 2004;16:321–328.

19. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694.

20. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117.

21. S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al., "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge", arXiv preprint arXiv:1811.02629 (2018).

22. Bakas, S., Sako, C., Akbari, H., Bilello, M., Sotiras, A., Shukla, G., Rudie, J. D., Flores Santamaria, N., Fathi Kazerooni, A., Pati, S., Rathore, S., Mamourian, E., Ha, S. M., Parker, W., Doshi, J., Baid, U., Bergman, M., Binder, Z. A., Verma, R., ⋯ Davatzikos, C. (2021). Multi-parametric magnetic resonance imaging (mpMRI) scans for de novo Glioblastoma (GBM) patients from the University of Pennsylvania Health System (UPENN-GBM) (Version 2) [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/TCIA.709X-DN49.

23. Choi Y, Nam Y, Jang J, Shin NY, Lee YS, Ahn KJ, et al. Radiomics may increase the prognostic value for survival in glioblastoma patients when combined with conventional clinical and genetic prognostic models. Eur

Radiol. 2021;31(4):2084-93.

24. Delgado-López PD, Corrales-García EM. Survival in glioblastoma: a review on the impact of treatment modalities. Clin Transl Oncol. 2016;18(11):1062-71.

25. Bae S, Choi YS, Ahn SS, Chang JH, Kang SG, Kim EH, et al. Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction. Radiology. 2018;289(3):797-806.

26. Sun Q, Chen Y, Liang C, Zhao Y, Lv X, Zou Y, et al. Biologic Pathways Underlying Prognostic Radiomics Phenotypes from Paired MRI and RNA Sequencing in Glioblastoma. Radiology. 2021;301(3):654-63.

27. Miller KD, Ostrom QT, Kruchko C, Patil N, Tihan T, Cioffi G, et al. Brain and other central nervous system tumor statistics, 2021. CA Cancer J Clin. 2021;71(5):381-406.

28. Ashraf M, Abdelsadg M, Grivas A. Relationship between molecular characteristics of glioblastoma multiforme and the subventricular zone. Br J Neurosurg. 2024;38(5):1100-7.

29. Li ZC, Bai H, Sun Q, Li Q, Liu L, Zou Y, et al. Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: A multicentre study. Eur Radiol. 2018;28(9):3640-50.

30. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563-77.

31. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195.

32. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. Radiology. 2018;286(3):800-9.

33. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med. 2005;352(10):997-1003.

34. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1

and IDH2 mutations in gliomas. N Engl J Med. 2009;360(8):765-73.

35. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.