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.

Imaging Biomarkers and Machine Learning for Preoperative Prediction of Progression in Glioblastoma

Mateo Moreno Gómez

mateo.moreno@utp.edu.co

Technological University of Pereira

Adriana Milena Páez Rodríguez

Technological University of Pereira

Iván Mauricio Herrera Mora

Technological University of Pereira

Research Article

Keywords: Glioblastoma, Imaging biomarkers, Radiomics, Machine learning, Magnetic resonance imaging, Tumor progression

Posted Date: March 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8936121/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

Glioblastoma is characterized by marked biological heterogeneity and early clinical deterioration, making progression assessment and risk stratification challenging in routine practice. Preoperative MRI radiomics may provide quantitative imaging biomarkers to support individualized prognostic modeling. We investigated whether a standardized radiomics–machine learning workflow can predict progression in histologically confirmed glioblastoma.

Methods

We retrospectively analyzed adults with CNS WHO grade 4 glioblastoma treated between 2019 and 2024. Progression was defined using RANO criteria with clinical-radiological follow-up to reduce misclassification in the early post-treatment window. From preoperative contrast-enhanced T1-weighted MRI, 159 radiomic features were extracted using an IBSI-oriented pipeline after semi-automatic 3D segmentation. After preprocessing and correlation filtering, feature selection was performed within a nested leave-one-out cross-validation framework. Nine classifiers were benchmarked, and uncertainty was quantified by bootstrap resampling (1,000 iterations). Performance was assessed by balanced accuracy, AUROC, AUPRC, Brier score, and calibration.

Results

Eleven patients met eligibility criteria (progression, n = 5; no progression, n = 6). Three GLRLM texture features were consistently selected across validation folds (ShortRunHighGrayLevelEmphasis, ShortRunLowGrayLevelEmphasis, LowGrayLevelRunEmphasis). The Extra Trees classifier achieved the best overall performance (balanced accuracy 0.817, AUROC 0.833, AUPRC 0.710, Brier score 0.188), outperforming alternative models in this cohort, with favorable calibration.

Conclusions

A compact preoperative MRI radiomics signature combined with machine learning showed promising performance for progression prediction in glioblastoma. This study supports the feasibility of a rigorously standardized, calibration-aware predictive framework and provides a basis for external multicenter validation.

Introduction

Glioblastoma (GBM), classified by the World Health Organization as a diffuse grade IV glioma, is the most aggressive primary brain tumor in adults [22]. It is characterized by highly infiltrative growth, rapid

proliferation, and marked genetic and histological heterogeneity, features that complicate management and are associated with poor outcomes [2, 37].

Despite standard multimodal treatment including maximal safe resection followed by radiotherapy with concomitant and adjuvant temozolomide according to the Stupp protocol [34], prognosis remains unfavorable. Median overall survival is approximately 12–15 months, and more than 90% of patients experience recurrence within two years of diagnosis [34, 26]. This limited therapeutic durability is largely driven by intratumoral heterogeneity, which promotes differential treatment sensitivity across cellular subclones within a dynamic, therapy-resistant microenvironment [2, 37].

During follow-up of patients with this condition, one of the major clinical challenges is distinguishing true tumor progression from pseudoprogression, a transient treatment-related inflammatory effect that can mimic recurrence on post-treatment imaging [1, 20, 24]. This distinction is especially important to guide treatment and avoid inappropriate therapeutic decisions. The RANO criteria for high-grade gliomas provide a standardized framework, defining progression as a ≥ 25% increase in bidimensional area or ≥ 40% increase in measurable tumor volume, appearance of new lesions, or unequivocal tumor-related clinical deterioration, with specific caution during the first 12 weeks after chemoradiotherapy, when persistence on subsequent assessment is required [38]. Even with these criteria, however, imaging-based assessment may remain difficult in borderline scenarios, particularly when radiographic patterns incompletely reflect the biological heterogeneity of tumor behavior.

Within this context, radiomics has emerged as a non-invasive strategy to quantify intratumoral heterogeneity through automated or semi-automated extraction of high-dimensional texture, shape, and intensity descriptors from medical imaging [17, 6]. These quantitative features have been associated with biological processes linked to glioblastoma aggressiveness, including proliferation, necrosis, and angiogenesis [7, 8, 29], supporting the role of radiomics as a bridge between conventional imaging and precision medicine [18].

Consistent with this rationale, prior studies have shown that combining radiomic and clinical variables with machine-learning methods can improve prediction of survival, progression, and treatment response in high-grade gliomas [14, 19, 28]. At the same time, clinical use has advanced slowly because methods often differ across studies, particularly in image processing and reporting, which limits reproducibility and cross-study comparison [19, 20]. In addition, many neuro-oncology datasets are small and single-center, so performance estimates can be unstable if validation is not carefully designed [16, 27].

In this setting, the present study was designed to develop and internally validate a model for GBM progression by combining preoperative contrast-enhanced T1 MRI radiomic features with clinical variables. The workflow followed IBSI recommendations to standardize image preprocessing and feature extraction [39], with the aim of improving consistency and reproducibility of quantitative imaging variables. Model development and reporting were aligned with TRIPOD-AI guidance to make predictor selection, validation procedures, and performance reporting more transparent, together with a rigorous

internal validation strategy [9]. The underlying hypothesis was that this standardized approach could provide reproducible progression-risk estimates and support future external multicenter validation.

Methods

Reporting, ethics, and data governance

The study was conducted in accordance with the Declaration of Helsinki and applicable national regulations. Approval was obtained from institutional Research Ethics Committee. Clinical and imaging data were de-identified prior to analysis by removing direct identifiers and re-identification keys. Data was stored in a secure cloud environment (with encryption, role-based access control, and audit logging. The study was retrospective and did not modify clinical management.

Study design and setting

An observational, analytical, retrospective predictive study was performed to develop a machine-learning model for estimating glioblastoma progression in routine clinical conditions. Adults with histopathologically confirmed glioblastoma treated between 2019 and 2024 at Hospital Universitario San Jorge (Pereira, Colombia) were screened. To reduce technical heterogeneity, inclusion required a technically adequate preoperative contrast-enhanced T1-weighted MRI (T1CE) suitable for segmentation and radiomics processing.

Participant selection and clinical data

Eligible participants were adults with CNS WHO grade 4 glioblastoma and available preoperative T1CE MRI in DICOM format. Clinical variables were extracted in a structured manner from electronic medical records. Exclusion criteria included insufficient image quality for segmentation/feature extraction and missing key clinical or follow-up information required for outcome adjudication.

Outcome definition: tumor progression

The primary endpoint was true tumor progression, defined according to RANO criteria for contrast-enhancing high-grade gliomas [38]. Progression was defined by ≥ 25% increase in bidimensional enhancing tumor area or ≥ 40% increase in measurable enhancing tumor volume, appearance of new lesions, or unequivocal tumor-related clinical deterioration under stable or increasing corticosteroid doses [38]. To improve traceability of the clinical component, neurological status was operationalized into three clinically assessable variables: headache, seizures, and focal neurological deficit. Events occurring within 12 weeks after chemoradiotherapy were classified as progression only when persistence or worsening was documented on subsequent clinical–radiological assessment, consistent

with the RANO early post-treatment caution window. Outside this interval, progression status was assigned based on the available clinical and imaging evidence.

Imaging and segmentation

Preoperative MRI were retrieved in DICOM format. Enhancing tumor volumes were segmented on T1CE images using 3D Slicer (v5.6) with semi-automatic tools. Segmentations underwent independent expert review and consensus reconciliation to mitigate inter-observer variability, resulting in a single finalized 3D tumor mask for each case (Fig. 1). Final masks were additionally used to generate 3D surface reconstructions for qualitative visualization and morphological documentation (Fig. 2).

Coronal (A, D), axial (B, E), and sagittal (C, F) views are shown. The lower row (D–F) displays the final tumor segmentation mask (cyan) overlaid on the corresponding MRI slices.

Multiplanar views with tumor mask overlay (A–C) and the corresponding three-dimensional surface reconstructions (D–F).

Radiomics preprocessing and feature extraction

Preprocessing and feature computation followed an IBSI-oriented workflow to support reproducibility [39]. Images were resampled to isotropic spacing (1×1×1 mm) using B-spline interpolation. Intensity normalization was performed using z-score normalization with a scaling parameter (normalizeScale = 100). Texture discretization used a fixed bin width of 31; GLCM matrices were computed using a symmetric configuration. Radiomic features were extracted with PyRadiomics (v3.1.0) [11] using a single configuration (.yaml) file to ensure parameter traceability across the pipeline. Extracted features included first-order intensity metrics, shape descriptors (e.g., volume, surface area, sphericity, elongation, compactness), and texture features derived from standard matrices including GLCM, GLRLM, GLSZM, and NGTDM.

Data handling and preprocessing

Radiomic features were integrated with clinical variables into a unified analytic dataset. Variables were reviewed to identify outliers and distributional asymmetry. Numeric variables were standardized using z-scores; categorical variables were encoded using label encoding. Identifiers, technical meta-fields, and postoperative variables without direct clinical interpretation were removed. Missing values were imputed using the median within the analytic dataset. To reduce redundancy, correlated features were filtered prior to selection using a correlation threshold ($|r| > 0.85$), reflecting collinearity patterns commonly observed within radiomic feature families.

Feature selection

Feature selection aimed to identify a compact predictor set suitable for the sample modeling. Three approaches were evaluated: ANOVA-based K-best ranking, LASSO, and Elastic Net. Feature selection was implemented within the internal validation framework to reduce information leakage and to avoid optimistic performance estimates.

Machine-learning development and internal validation

Modeling was implemented in Python (v3.10) using Kaggle Notebooks and standard scientific libraries (pandas, numpy, scikit-learn, matplotlib, seaborn). Model development and reporting were aligned with TRIPOD-AI/TRIPOD + AI guidance [9]. Given the sample size, nested leave-one-out cross-validation was used. Within this scheme, preprocessing and feature selection were performed in the inner loop, while performance was estimated in the outer loop to maintain separation between model selection and evaluation. Nine classifiers were benchmarked: Logistic Regression, Linear Discriminant Analysis, Support Vector Machine (linear kernel), Gaussian Naïve Bayes, Ridge Classifier, Stochastic Gradient Descent classifier, Random Forest, Gradient Boosting, and Extra Trees.

Performance metrics, calibration, and uncertainty

Balanced accuracy was prespecified as the primary metric to account for potential class imbalance. Discrimination was summarized using AUROC, and precision–recall performance using AUPRC. Calibration was assessed using the Brier score and calibration diagnostics (intercept/slope and calibration curves). Uncertainty was quantified using bootstrap resampling (1,000 iterations) to derive 95% confidence intervals for performance estimates. A schematic overview of the complete analytical workflow is provided in Fig. 3.

Figure 3 **Overview of the study workflow from cohort identification and RANO-based outcome adjudication to preoperative T1CE segmentation, IBSI-oriented radiomics feature extraction, feature filtering/selection, machine-learning benchmarking, and internal validation using nested leave-one-out cross-validation with bootstrap-based uncertainty estimation**

Results

Approximately 30 patients with a diagnosis of glioblastoma were initially identified. After eligibility assessment, including image-quality review, verification of completeness of clinical records, and data cleaning to resolve inconsistencies, 11 patients met inclusion criteria and were retained for analysis. The primary outcome was tumor progression adjudicated using RANO criteria. Event distribution comprised 5 progression cases (45.5%) and 6 patients (54.5%) remaining progression-free during clinical and radiological follow-up. Imaging data were processed under an IBSI-oriented workflow, ensuring isotropic 1-mm spacing and spatial consistency between images and segmentation masks.

The initial radiomics set comprised 159 candidate features and was reduced to 125 after removal of administrative fields, technical metadata, and postoperative variables not suitable for preoperative

prediction. Missing values were imputed using the median, and predictors were standardized using z-score within each cross-validation iteration to limit information leakage. Exploratory correlation analysis demonstrated substantial collinearity among morphometric and texture-derived descriptors within the same feature families. Multiple near-duplicate pairs showed Pearson correlations approaching unity (Table 1), including volume-related measures (e.g., VoxelNum.1–VoxelVolume; MeshVolume–VoxelNum.1) and first-order intensity metrics (e.g., Energy–TotalEnergy; JointAverage–SumAverage). Consistent with this finding, a correlation filter ($|r| \leq 0.85$) was applied prior to final feature selection. The overall correlation structure among the top-ranked preoperative predictors is summarized in Fig. 3, showing clustering patterns consistent with partial redundancy across texture families after preprocessing.

Table 1
Highest correlations among preoperative radiomic features

Feature 1	Feature 2	Pearson r
Energy	TotalEnergy	1.000
JointAverage	SumAverage	1.000
Minimum.2	Minimum.3	1.000
Maximum.2	Maximum.3	1.000
VoxelNum.1	VoxelVolume	1.000
Mean.2	Mean.3	1.000
MeshVolume	VoxelNum.1	1.000
LargeAreaEmphasis	ZoneVariance	1.000
GrayLevelVariance.2	Variance	0.999

Feature selection was constrained to three predictors in keeping with TRIPOD-AI recommendations for small-sample model development. ANOVA-based K-best ranking, LASSO, and Elastic Net were evaluated within the leave-one-out cross-validation framework in combination with the candidate classifiers. Across all 11 LOOCV iterations, three GLRLM textural features were consistently retained: ShortRunHighGrayLevelEmphasis, ShortRunLowGrayLevelEmphasis, and LowGrayLevelRunEmphasis. These features showed comparable contributions to outcome discrimination, with mean normalized importance values of 0.634, 0.627, and 0.597, respectively, indicating stable selection patterns across folds.

The Extra Trees classifier achieved the highest overall performance, with balanced accuracy (BA) of 0.817, AUROC of 0.833, and AUPRC of 0.710. Performance was higher or comparable to logistic regression (BA 0.817; AUROC 0.800) and random forest (BA 0.733; AUROC 0.800). Results for all nine

benchmarked classifiers are summarized in Table 2, and the corresponding metric comparison is shown in Fig. 4.

Table 2
Performance of classification models

Model	Balanced accuracy	AUROC	AUPRC	F1-macro	Brier score	Calibration error
Extra Trees Classifier	0.817	0.833	0.710	0.817	0.188	0.361
Random Forest	0.733	0.800	0.686	0.727	0.170	0.162
Logistic Regression	0.817	0.800	0.686	0.817	0.212	0.375
Linear Discriminant Analysis	0.733	0.767	0.660	0.727	0.215	0.333
Gradient Boosting	0.633	0.767	0.726	0.633	0.269	0.386
Gaussian Naive Bayes	0.717	0.767	0.660	0.718	0.263	0.231
SVM (Linear Kernel)	0.733	0.667	0.563	0.727	0.262	0.408
Ridge Classifier	0.733	0.667	0.563	0.727	0.213	0.283
Stochastic Gradient Descent	0.633	0.533	0.520	0.633	0.365	0.355

Bar plot showing balanced accuracy, AUROC, and AUPRC for the nine machine-learning models evaluated under leave-one-out cross-validation.

This model also showed the most favorable calibration, with a Brier score of 0.188. Inspection of the ROC, precision–recall, and calibration curves **(**Fig. 5**)** suggested good agreement between predicted and observed risks, with calibration slope close to 1 and intercept near 0. Bootstrap-derived confidence intervals were wide, as expected for a limited sample size, but centered on the point estimates, supporting consistency of model performance within the available cohort.

Receiver operating characteristic (ROC) curve (left) and precision–recall curve (center) summarize discrimination, while the calibration plot (right) shows agreement between predicted probabilities and observed outcomes; the Brier score is reported on the calibration panel.

Across the 11 LOOCV iterations, the three retained GLRLM features—ShortRunHighGrayLevelEmphasis, ShortRunLowGrayLevelEmphasis, and LowGrayLevelRunEmphasis—showed moderate inter-feature dependence (r = 0.71–0.76), suggesting partially shared texture information without complete redundancy after correlation filtering. This pattern is consistent with the broader clustering observed among top-ranked preoperative radiomic predictors in the correlation map **(**Fig. 3**).** `

Univariable screening further supported the relevance of these predictors. Point-biserial correlation and F-score ranking identified the same GLRLM descriptors as the strongest individual associates of progression (Table 3), led by ShortRunHighGrayLevelEmphasis (|r| = 0.634; p = 0.036; F = 6.06) and ShortRunLowGrayLevelEmphasis (|r| = 0.627; p = 0.039; F = 5.84), followed by LowGrayLevelRunEmphasis (|r| = 0.597; p = 0.052; F = 4.98). The complete top 10 ranking is illustrated in Fig. 6.

Table 3
Correlation and discriminative strength of the 10 predictors most associated with the outcome

Feature	Point-biserial r	p-value	F-score
ShortRunHighGrayLevelEmphasis	0.634	0.036	6.06
ShortRunLowGrayLevelEmphasis	0.627	0.039	5.84
LowGrayLevelRunEmphasis	0.597	0.052	4.98
HighGrayLevelRunEmphasis	0.571	0.067	4.34
SmallDependenceLowGrayLevelEmphasis	0.539	0.087	3.69
SumEntropy	0.535	0.090	3.61
LowGrayLevelZoneEmphasis	0.522	0.100	3.37
HighGrayLevelEmphasis	0.516	0.104	3.27
LowGrayLevelEmphasis	0.509	0.110	3.15
SmallAreaLowGrayLevelEmphasis	0.491	0.125	2.86

Horizontal bar plot showing the absolute point-biserial correlation (|r|) between each candidate feature and outcome. The ranking highlights a predominance of GLRLM texture descriptors, with ShortRunHighGrayLevelEmphasis, ShortRunLowGrayLevelEmphasis, and LowGrayLevelRunEmphasis among the strongest individual associations.

Finally, comparison using a radar plot (Fig. 7) showed a more uniform performance profile for the Extra Trees model when balanced accuracy, AUPRC, and F1-macro were considered jointly. A similar pattern was observed for other ensemble-based approaches, whereas penalized linear classifiers and the stochastic gradient descent model showed less consistent profiles across metrics.

Radar plot summarizing balanced accuracy, AUROC, AUPRC, and F1-macro across the evaluated classifiers under LOOCV to facilitate joint assessment of discrimination and overall performance.

Discussion

The present study suggests that a set of clinical and preoperative MRI radiomic features, processed under IBSI-oriented settings and developed and reported according to TRIPOD-AI principles, can support progression-risk modeling in CNS WHO grade 4 glioblastoma within a real cohort of patients. These findings align with the rationale for radiomics as a quantitative approach for individualized risk stratification in biologically heterogeneous brain tumors [22, 6, 8].

Methodological standardization is essential for reproducible radiomics and remains a barrier to clinical translation. In neuro-oncology Radiomics Quality Score (RQS) systematic assessments have consistently documented low overall methodological quality and substantial heterogeneity in acquisition, segmentation, feature handling, and reporting, often reflecting an average RQS on the order of ~ 19%, well below thresholds expected for clinically deployable models [27, 36]. Accordingly, the present workflow was designed to address recurrent sources of bias by adopting an IBSI-oriented preprocessing, enforcing fold-wise imputation and scaling to limit information leakage, and using leave-one-out cross-validation with bootstrap-derived uncertainty intervals to reduce optimism in performance estimation [3, 32]. Consistent with TRIPOD-AI, reporting emphasizes not only discrimination but also calibration, an aspect frequently omitted in neuro-oncology radiomics despite being critical for interpreting predicted risks and supporting clinical decision-making

Among the evaluated classifiers, the Extra Trees model achieved AUROC 0.833 with balanced accuracy 0.817 and favorable calibration in this cohort. These estimates fall within the range commonly reported for glioblastoma progression or survival prediction models, where AUC values frequently span 0.70–0.85 depending on cohort composition, endpoints, and validation design [14, 3, 25]. Tree-based ensemble methods may be advantageous in small datasets because they can capture non-linear interactions and reduce variance through aggregation without requiring large training samples [4]. Consistent with recent reviews, well-calibrated classical machine-learning ensembles may achieve performance comparable to more complex architectures when data are limited, while offering practical advantages in reproducibility and implementation [5, 8].

The three retained predictors, ShortRunHighGrayLevelEmphasis, ShortRunLowGrayLevelEmphasis, and LowGrayLevelRunEmphasis, belong to the GLRLM family and quantify fine-scale textural heterogeneity within enhancing tumor regions. Such texture descriptors may reflect microstructural complexity associated with key biological processes in glioblastoma, including cellular proliferation, necrosis, and neovascularization [7]. Prior work has linked irregular enhancement patterns and textural complexity to clinically relevant outcomes, including survival and treatment response, independent of tumor size [3]. These observations support the plausibility that GLRLM-based descriptors capture prognostically meaningful aspects of intratumoral heterogeneity, although direct biological interpretation remains inferential without radiogenomic or histopathologic correlation.

Progression was operationalized using RANO criteria [38], which provide standardized thresholds for response assessment in contrast-enhancing high-grade gliomas. Although RANO also addresses the challenge of distinguishing true progression from pseudoprogression, the present endpoint focused on

true progression to reduce ambiguity related to early post-treatment imaging changes and to maintain a clinically reproducible definition. The use of a consensus framework supports comparability with contemporary neuro-oncology literature and situates the modeling approach within established response assessment standards [32].

While deep learning methods have expanded rapidly in neuro oncology imaging, evidence of consistent superiority over classical machine-learning approaches remains mixed, particularly in settings with limited data. Recent systematic reviews and comparative studies report that many deep learning models for glioblastoma outcome prediction achieve performance similar to conventional approaches (often AUC ~ 0.80–0.85), with substantial variability across study designs and validation strategies [5, 13]. In this context, Extra Trees offers a pragmatic alternative with lower computational demands and simpler reproducibility in small cohorts [30]. Future work may assess whether integrating radiomic signatures with molecular, genomic, or transcriptomic information improves prognostic performance and supports radiogenomic modeling in glioblastoma [30, 10].

Limitations and future directions

This study has important limitations. The most relevant is the sample size, which reduces statistical precision, increases the risk of model instability, and limits generalizability. In glioblastoma radiomics, small development cohorts have been used in early-phase model building; however, those studies are primarily pilot studies and require careful control of overfitting and transparent reporting to be interpretable [15, 21, 33, 35]. Therefore, the present findings should be viewed as preliminary and not as evidence of clinical readiness.

Multiple design choices were made to reduce though not eliminate optimism in performance estimates. Image preprocessing followed an IBSI-oriented workflow, predictor handling and reporting were aligned with TRIPOD-AI principles, and model evaluation used leave-one-out cross-validation with bootstrap-based uncertainty quantification. These steps help limit information leakage and provide a clearer view of uncertainty but they cannot substitute for independent validation. A second limitation is the single-center nature of the dataset and the likely heterogeneity in MRI acquisition parameters, which can alter radiomic feature distributions and affect transportability across scanners and institutions. This limitation is broadly recognized in neuro-oncology radiomics and remains a major barrier to translation [27]. External validation in independent, multicenter cohorts is therefore the next step; using harmonized public datasets such as BraTS may facilitate benchmarking and improve reproducibility across sites [23]. Finally, future work should test whether the selected GLRLM texture features add incremental value beyond established clinical and molecular predictors and whether multimodal integration improves calibration and clinical utility [10]. Incorporating explainable AI approaches would also be useful to characterize model behavior, identify failure modes, and support responsible implementation in clinical decision-making [12, 31].

Conclusions

In this study, a preoperative MRI radiomics pipeline developed under IBSI-oriented preprocessing and TRIPOD-AI–aligned reporting yielded an internally validated model for predicting RANO-defined progression in glioblastoma. Using a parsimonious predictor set constrained to three GLRLM texture features, the best-performing classifier (Extra Trees) achieved an AUROC of 0.833 with balanced accuracy of 0.817 and acceptable calibration in leave-one-out cross-validation.

These results support the hypothesis that fine-scale textural heterogeneity encodes prognostically relevant information that can be quantified non-invasively from routine imaging. Given the sample size and single-cohort design, the findings should be interpreted as preliminary and require confirmation in independent, preferably multicenter datasets. Future work should evaluate clinical utility (e.g., decision-curve analysis), external calibration, and the incremental value of radiomics beyond established clinical predictors, as well as explore multimodal integration with molecular and histopathologic markers to advance precision risk stratification in neuro-oncology.

Abbreviations

AUPRC: area under the precision–recall curve

AUROC: area under the receiver operating characteristic curve

BA: balanced accuracy

BraTS: Brain Tumor Segmentation Benchmark

CBTRUS: Central Brain Tumor Registry of the United States

CNS: central nervous system

DICOM: Digital Imaging and Communications in Medicine

GBM: glioblastoma

GLCM: gray level co-occurrence matrix

GLRLM: gray level run length matrix

GLSZM: gray level size zone matrix

IBSI: Image Biomarker Standardisation Initiative

LOOCV: leave-one-out cross-validation

MRI: magnetic resonance imaging

NGTDM: neighbouring gray tone difference matrix

RANO: Response Assessment in Neuro-Oncology

RQS: Radiomics Quality Score

SVM: support vector machine

T1CE: contrast-enhanced T1-weighted MRI

TRIPOD-AI / TRIPOD+AI: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis – Artificial Intelligence extension

WHO: World Health Organization

Declarations

Funding: the authors did not receive support from any organization for the submitted work**.**

Competing Interests: the authors have no relevant financial or non-financial interests to disclose.

Human Ethics and Consent to Participate declarations: this study involved human participant data. Ethical approval was granted by the Institutional Review Board of Universidad Tecnológica de Pereira (Protocol No. 59 - 260525). Given the retrospective design and the exclusive use of fully de-identified data, the requirement for informed consent to participate was waived by the ethics committee.

Consent to Participate: informed consent to participate was waived by the Institutional Review Board of Universidad Tecnológica de Pereira (Protocol No. 67-040725) due to the retrospective nature of the study and the use of fully anonymized, de-identified data.

Consent for Publication: not applicable.

Data Availability: the datasets and models generated and/or analyzed during the current study are available upon request and subject to institutional approval. Data will be provided in de-identified form.

Author Contributions

MMG: Mateo Moreno Gómez

IMHM: Iván Mauricio Herrera Mora

AMPR: Adriana Milena Páez Rodríguez

The contributions of each author to this study are as follows: conceptualization of the research question and study design were performed by AMPR and MMG. Methodology and analytic workflow, including patient selection and data preprocessing strategy, were developed by MMG, AMPR, and IMHM. Data curation, encompassing clinical data collection, verification of eligibility criteria, and database cleaning,

was carried out by AMPR and MMG. Quality control of imaging-derived data and oversight of the radiomics workflow were conducted By AMPR and IMHM. Radiomic feature extraction, computational analysis, and machine-learning model development were performed by MMG and AMPR. Statistical analysis and internal validation procedures were executed by MMG. Interpretation of results and contribution to the discussion were provided by all authors. The original manuscript draft was written by MMG AND AMPR, and all authors contributed to reviewing and editing the manuscript for important intellectual content. Supervision and project oversight were provided by IMHM. All authors have read and approved the final version of the manuscript. All authors approved the final version of the manuscript and any subsequent revisions.

References

1. Abbasi AW, Westerlaan HE, Holtman GA, Aden KM, van Laar PJ, van der Hoorn A (2018) Incidence of Tumour Progression and Pseudoprogression in High-Grade Gliomas: a Systematic Review and Meta-Analysis. Clin Neuroradiol 28(3):401–411
2. Alexander BM, Cloughesy TF (2017) Adult Glioblastoma. J Clin Oncol 35(21):2402–2409
3. Bae S, Choi YS, Ahn SS, Chang JH, Kang S-G, Kim EH, Kim SH, Lee S-K (2018) Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction. Radiology 289(3):797–806
4. Biau G, Scornet E (2016) A random forest guided tour. TEST 25(2):197–227
5. Booth TC, Williams M, Luis A, Cardoso J, Ashkan K, Shuaib H (2019) Machine learning and glioma imaging biomarkers. Clin Radiol 75(1):20–32
6. Chaddad A, Kucharczyk MJ, Daniel P, Sabri S, Jean-Claude BJ, Niazi T, Abdulkarim B (2019) Radiomics in Glioblastoma: Current Status and Challenges Facing Clinical Implementation. Front Oncol. 10.3389/fonc.2019.00374
7. Choi Y, Jang J, Kim B-S, Ahn K-J (2023) Pretreatment MR-based radiomics in patients with glioblastoma: A systematic review and meta-analysis of prognostic endpoints. Eur J Radiol 168:111130
8. Christodoulou RC, Pitsillos R, Papageorgiou PS, Petrou V, Vamvouras G, Rivera L, Papageorgiou SG, Solomou EE, Georgiou MF (2025) Artificial Intelligence in Glioma Diagnosis: A Narrative Review of Radiomics and Deep Learning for Tumor Classification and Molecular Profiling Across Positron Emission Tomography and Magnetic Resonance Imaging. Eng 6(10):262
9. Collins GS, Moons KGM, Dhiman P et al (2024) TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 385:e078378
10. Fathi Kazerooni A, Bakas S, Saligheh Rad H, Davatzikos C (2020) Imaging signatures of glioblastoma molecular characteristics: A radiogenomics review. J Magn Reson Imaging 52(1):54–69
11. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin J-C, Pieper S, Aerts HJWL (2017) Computational Radiomics System to Decode the

Radiographic Phenotype. Cancer Res 77(21):e104–e107
12. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H (2019) Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 9(4):e1312
13. Khan MKH, Guo W, Liu J, Dong F, Li Z, Patterson TA, Hong H (2023) Machine learning and deep learning for brain tumor MRI image segmentation. Exp Biol Med (Maywood) 248(21):1974–1992
14. Kickingereder P, Burth S, Wick A et al (2016) Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models. Radiology 280(3):880–889
15. Kihira S, Tsankova NM, Bauer A et al (2021) Multiparametric MRI texture analysis in prediction of glioma biomarker status: added value of MR diffusion. Neurooncol Adv 3(1):vdab051
16. Kotrotsou A, Zinn PO, Colen RR (2016) Radiomics in Brain Tumors: An Emerging Technique for Characterization of Tumor Environment. Magn Reson Imaging Clin N Am 24(4):719–729
17. Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. Magn Reson Imaging 30(9):1234–1248
18. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14(12):749–762
19. Lao J, Chen Y, Li Z-C, Li Q, Zhang J, Liu J, Zhai G (2017) A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. Sci Rep 7(1):10353
20. Li M, Ren X, Dong G et al (2021) Distinguishing Pseudoprogression From True Early Progression in Isocitrate Dehydrogenase Wild-Type Glioblastoma by Interrogating Clinical, Radiological, and Molecular Features. Front Oncol 11:627325
21. Long H, Zhang P, Bi Y, Yang C, Wu M, He D, Huang S, Yang K, Qi S, Wang J (2023) MRI radiomic features of peritumoral edema may predict the recurrence sites of glioblastoma multiforme. Front Oncol 12:1042498
22. Louis DN, Perry A, Wesseling P et al (2021) The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol 23(8):1231–1251
23. Menze BH, Jakab A, Bauer S et al (2015) The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging 34(10):1993–2024
24. Moassefi M, Faghani S, Conte GM et al (2022) A deep learning model for discriminating true progression from pseudoprogression in glioblastoma patients. J Neurooncol 159(2):447–455
25. Nakhate V, Gonzalez Castro LN (2023) Artificial intelligence in neuro-oncology. Front Neurosci 17:1217629
26. Ostrom QT, Price M, Neff C, Cioffi G, Waite KA, Kruchko C, Barnholtz-Sloan JS (2022) CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015–2019. Neuro Oncol 24(Suppl 5):v1–v95
27. Park JE, Kim HS, Kim D, Park SY, Kim JY, Cho SJ, Kim JH (2020) A systematic review reporting quality of radiomics research in neuro-oncology: toward clinical utility and quality improvement

using high-dimensional imaging features. BMC Cancer 20(1):29

28. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL (2015) Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep 5(1):13087
29. Pérez-Beteta J, Molina-García D, Ortiz-Alhambra JA et al (2018) Tumor Surface Regularity at MR Imaging Predicts Survival and Response to Surgery in Patients with Glioblastoma. Radiology 288(1):218–225
30. Rathore S, Akbari H, Doshi J, Shukla G, Rozycki M, Bilello M, Lustig R, Davatzikos C (2018) Radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma: implications for personalized radiotherapy planning. J Med Imaging (Bellingham) 5(2):021219
31. Reyes M, Meier R, Pereira S, Silva CA, Dahlweid F-M, von Tengg-Kobligk H, Summers RM, Wiest R (2020) On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. Radiology: Artificial Intelligence 2(3):e190043
32. Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman MLD, Smith TR, Arnaout O (2018) Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review. World Neurosurg 109:476–486e1
33. Shim KY, Chung SW, Jeong JH et al (2021) Radiomics-based neural network predicts recurrence patterns in glioblastoma using dynamic susceptibility contrast-enhanced MRI. Sci Rep 11(1):9974
34. Stupp R, Mason WP, van den Bent MJ et al (2005) Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 352(10):987–996
35. Styliara EI, Astrakas LG, Alexiou G, Xydis VG, Zikou A, Kafritsas G, Voulgaris S, Argyropoulou MI (2024) Survival Outcome Prediction in Glioblastoma: Insights from MRI Radiomics. Curr Oncol 31(4):2233–2243
36. Ugga L, Perillo T, Cuocolo R, Stanzione A, Romeo V, Green R, Cantoni V, Brunetti A (2021) Meningioma MRI radiomics and machine learning: systematic review, quality score assessment, and meta-analysis. Neuroradiology 63(8):1293–1304
37. Weller M, Wick W, Aldape K et al (2015) Glioma. Nat Rev Dis Primers 1:15017
38. Wen PY, Macdonald DR, Reardon DA et al (2010) Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. J Clin Oncol 28(11):1963–1972
39. Zwanenburg A, Vallières M, Abdalah MA et al (2020) The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology 295(2):328–338

Figures

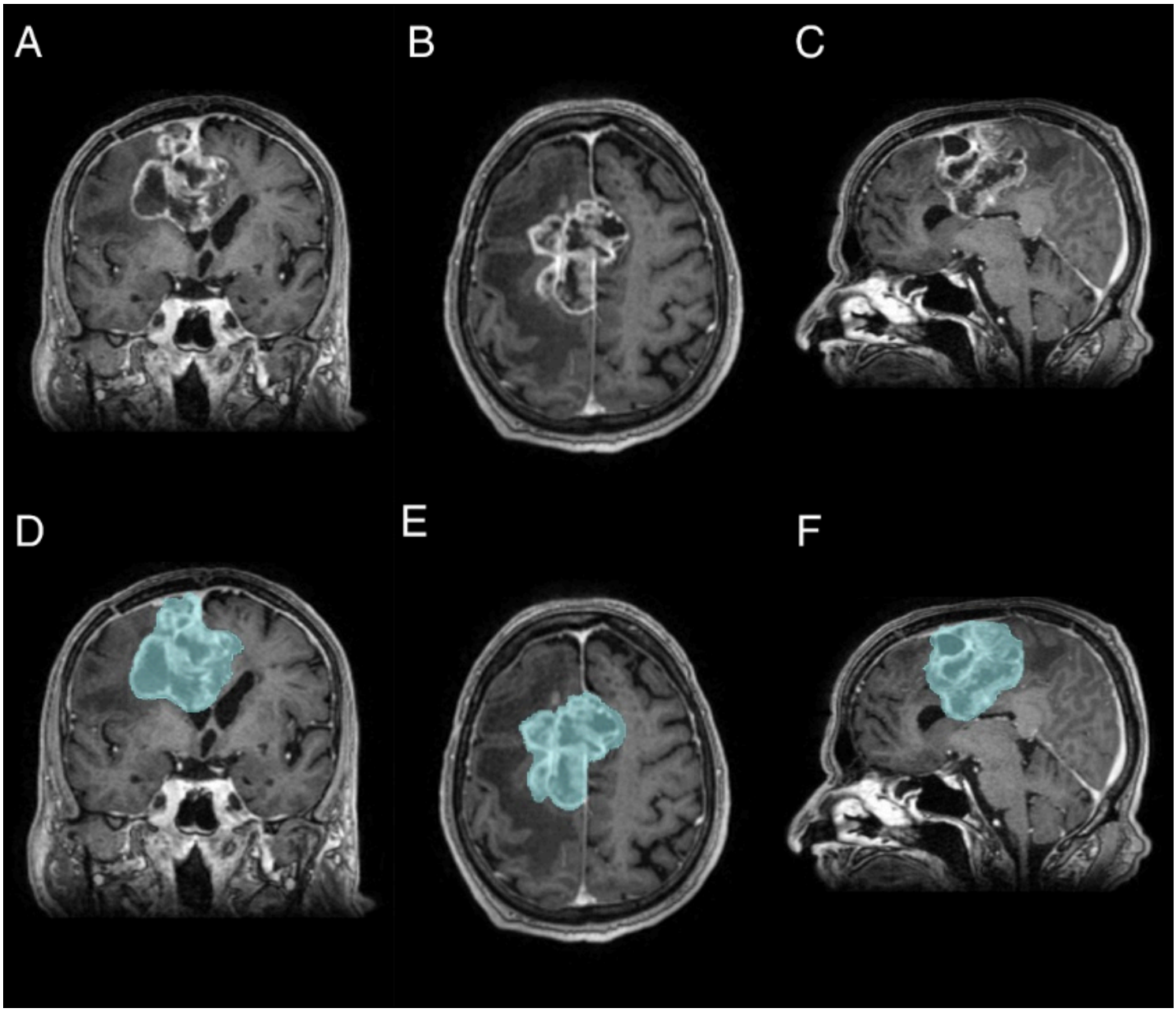


Figure 1

Volumetric glioblastoma segmentation on preoperative contrast-enhanced T1-weighted MRI (T1CE). Coronal (A, D), axial (B, E), and sagittal (C, F) views are shown. The lower row (D–F) displays the final tumor segmentation mask (cyan) overlaid on the corresponding MRI slices.

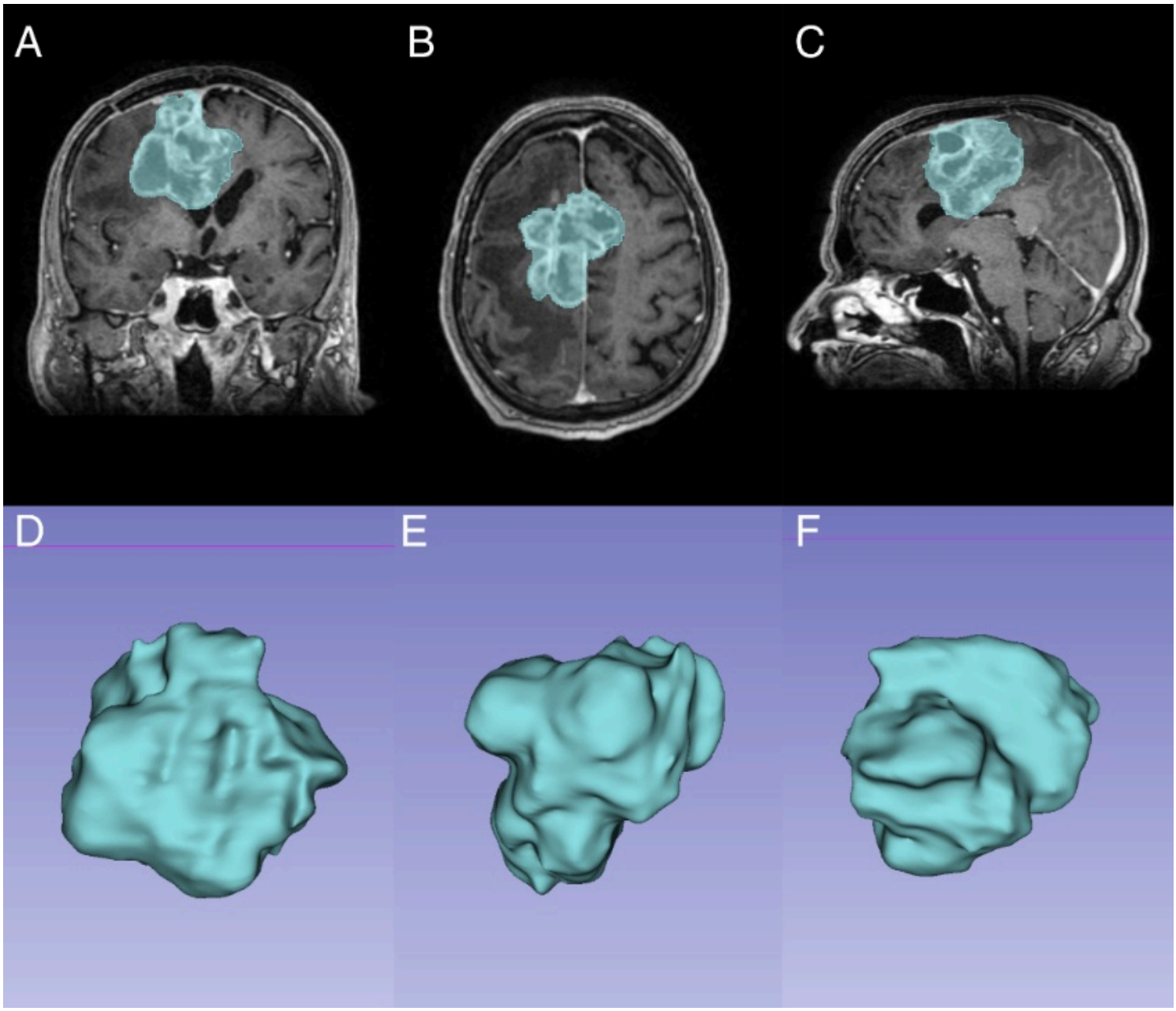


Figure 2

Three-dimensional reconstruction of the segmented glioblastoma volume.
Multiplanar views with tumor mask overlay (A–C) and the corresponding three-dimensional surface reconstructions (D–F).

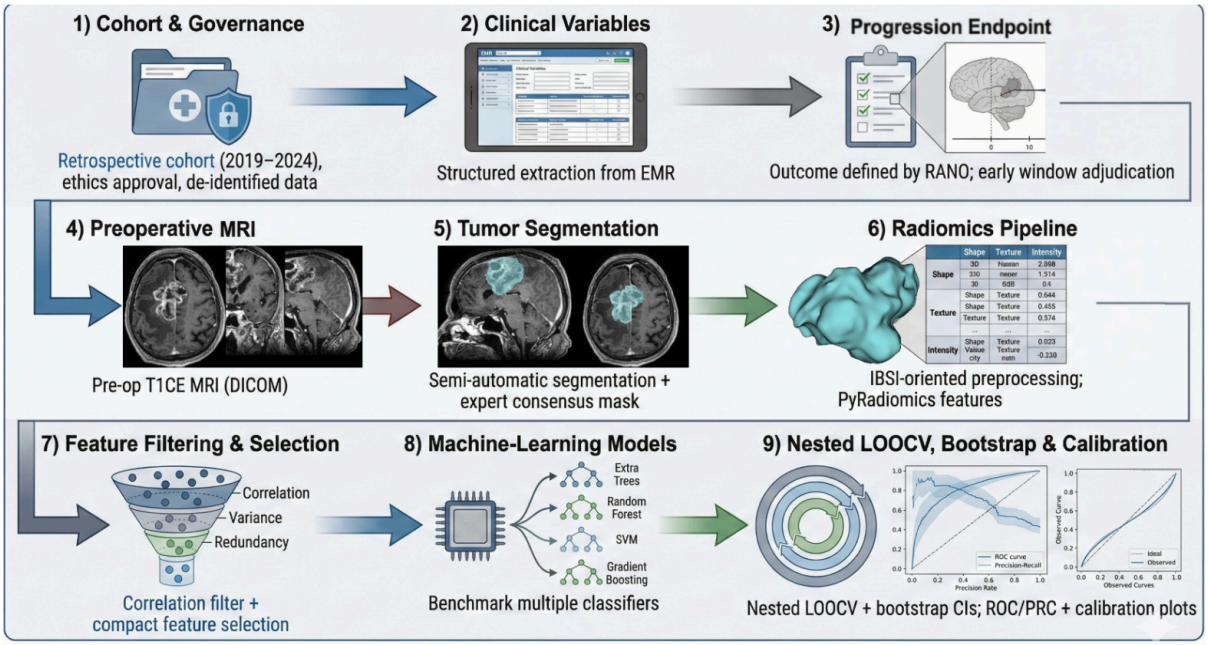


Figure 3

Overview of the study workflow from cohort identification and RANO-based outcome adjudication to preoperative T1CE segmentation, IBSI-oriented radiomics feature extraction, feature filtering/selection, machine-learning benchmarking, and internal validation using nested leave-one-out cross-validation with bootstrap-based uncertainty estimation

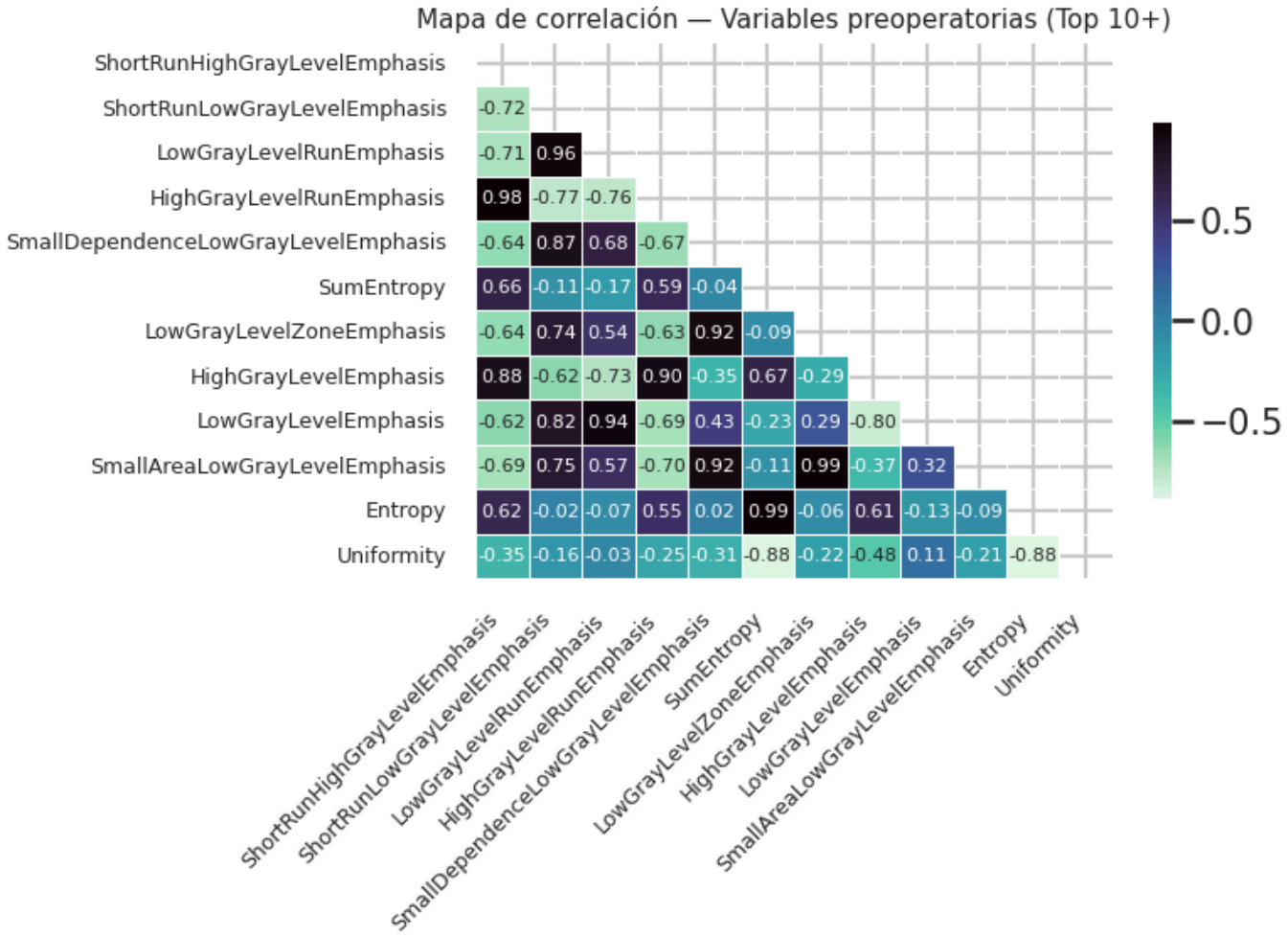


Figure 4

Fig. 3 Correlation structure among top-ranked preoperative radiomic predictors.

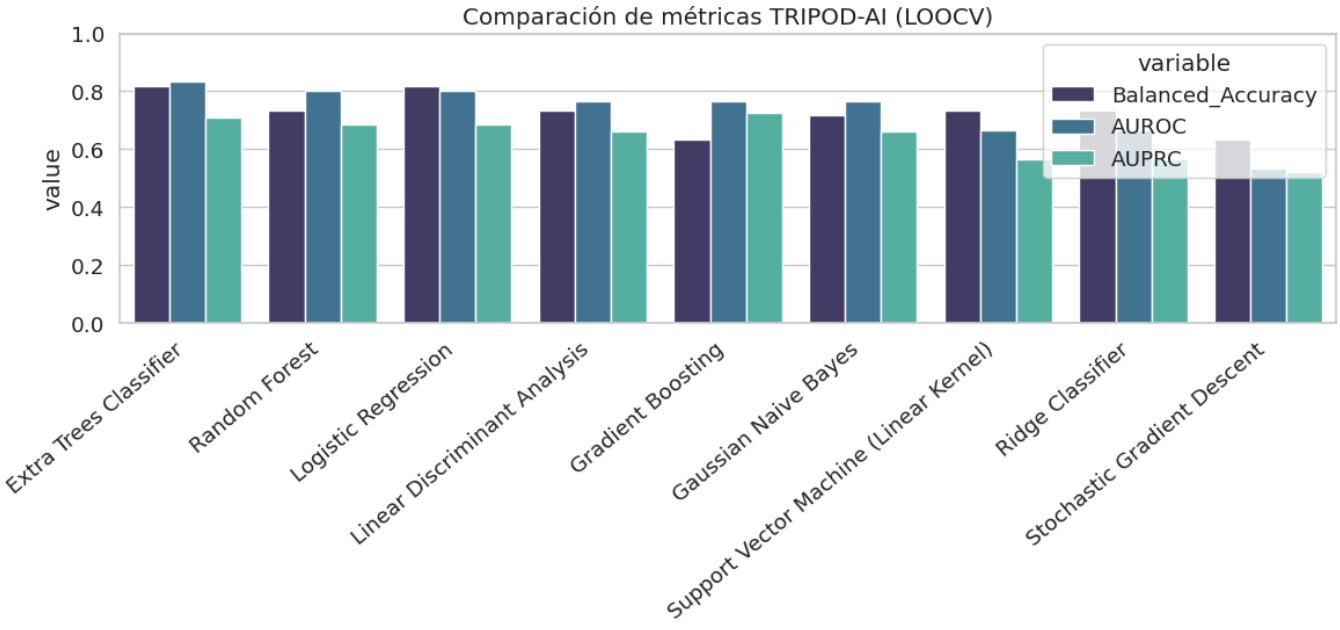


Figure 5

Fig. 4 Performance metric comparison across benchmarked classifiers.
Bar plot showing balanced accuracy, AUROC, and AUPRC for the nine machine-learning models evaluated under leave-one-out cross-validation.

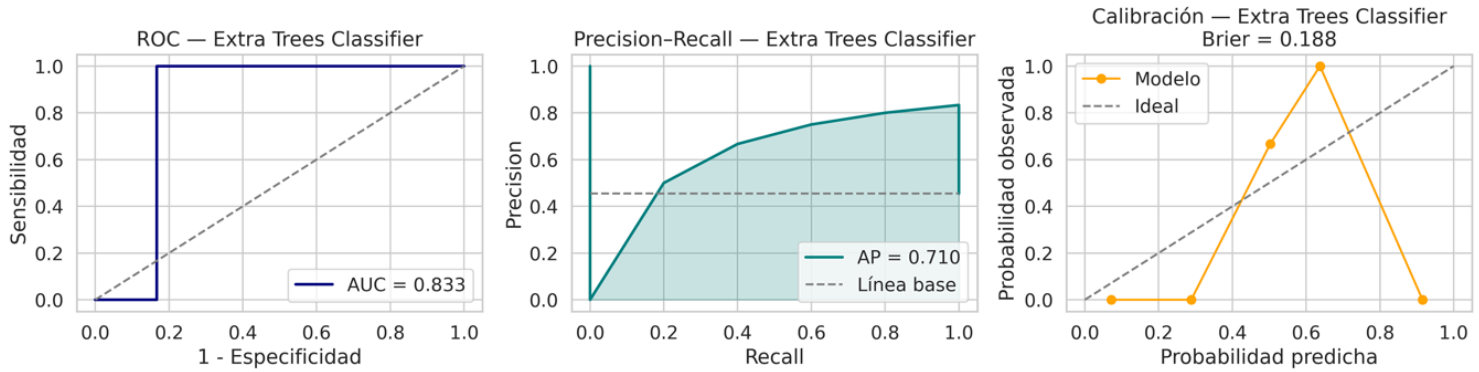


Figure 6

Fig. 5 Discrimination and calibration performance of the best-performing model (Extra Trees). Receiver operating characteristic (ROC) curve (left) and precision–recall curve (center) summarize discrimination, while the calibration plot (right) shows agreement between predicted probabilities and observed outcomes; the Brier score is reported on the calibration panel.

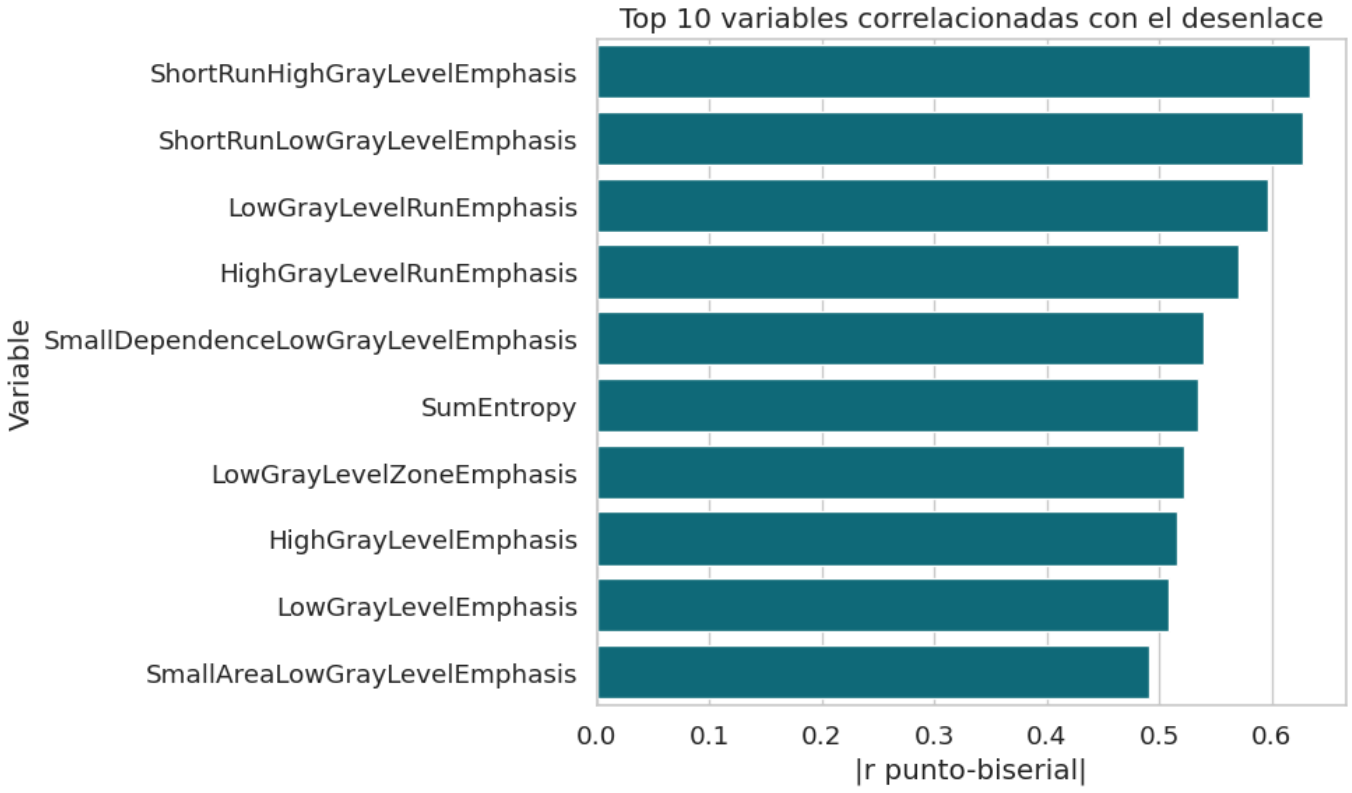


Figure 7

Fig 6. Univariable ranking of the top 10 preoperative radiomic predictors associated with progression. Horizontal bar plot showing the absolute point-biserial correlation (|r|) between each candidate feature and outcome. The ranking highlights a predominance of GLRLM texture descriptors, with ShortRunHighGrayLevelEmphasis, ShortRunLowGrayLevelEmphasis, and LowGrayLevelRunEmphasis among the strongest individual associations.

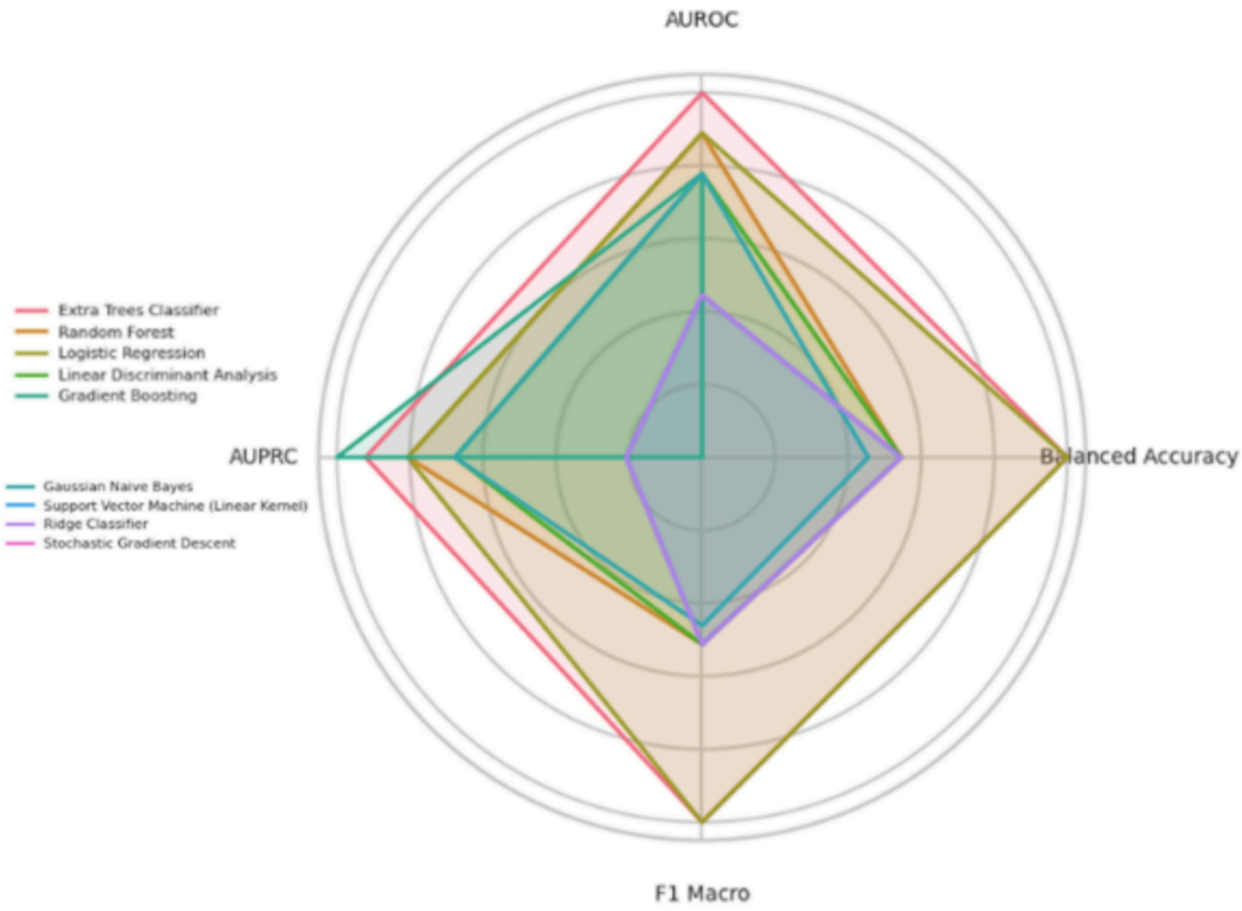


Figure 8

Fig.7 Multimetric radar comparison of classifier performance.
Radar plot summarizing balanced accuracy, AUROC, AUPRC, and F1-macro across the evaluated classifiers under LOOCV to facilitate joint assessment of discrimination and overall performance.