NEUROSURGICAL FOCUS

Neurosurg Focus 61(3):E16, 2026

Multicenter machine learning model using clinical and radiomic features for prediction of postoperative residual status in meningioma

Nafiye Sanlier, MD,[1,3] Murat Yuce, MD, MS,[2] Umid Sulaimanov, MD,[1] Gular Ismayilova, MD,[3] Cagin Huryol, MD,[4] Huseyin Bozkurt, MD,[5] Oyku Ozturk, MD,[1,6] Fikret Ozturk, MD,[7] Irem Altun, MD,[8] Taha Yasin Altunkaya, MD,[3] Simon G. Ammanuel, MD,[1] Behman Demir, MD,[1] Yerkebulan Serikkanov, MD,[1] Ahmed Rasim Bayramoglu, MD,[1,7] Huseyin Demir, MD,[9] Ariorad Moniri,[10] Yesim Karagoz, MD,[5] Burak Karaaslan, MD,[6] Ilhan Aydin, MD,[7] Mehmet Sabri Gurbuz, MD,[8] Okan Turk, MD,[3] Abdullah Keles, MD,[1] Ufuk Erginoglu, MD,[1] and Mustafa K. Baskaya, MD[1]

[1]Department of Neurological Surgery, University of Wisconsin School of Medicine & Public Health, Madison, Wisconsin; [2]Icahn School of Medicine at Mount Sinai, Biomedical Engineering and Imaging Institute, New York, New York; [3]Department of Neurosurgery, University of Health Sciences, Istanbul Training and Research Hospital, Istanbul, Turkey; [4]Department of Radiology, Gazimagusa State Hospital, Gazimagusa, Cyprus; [5]Department of Radiology, University of Health Sciences, Istanbul Training and Research Hospital, Istanbul; [6]Department of Neurosurgery, Gazi University Faculty of Medicine, Ankara; [7]Department of Neurosurgery, University of Health Sciences, Bakırkoy Prof. Dr. Mazhar Osman Training and Research Hospital for Psychiatry, Neurology, and Neurosurgery, Istanbul; [8]Department of Neurosurgery, Istanbul Medeniyet University, Prof. Dr. Suleyman Yalcin City Hospital, Istanbul; [9]Department of Neurosurgery, Istanbul Health and Technology University, Istanbul; and [10]Acıbadem University School of Medicine, Istanbul, Turkey

OBJECTIVE Gross-total resection is the primary surgical objective in meningioma management; however, predicting resectability preoperatively remains challenging, particularly for meningiomas located in the skull base that exhibit complex anatomical relationships. There is a lack of validated reproducible models that integrate known anatomical factors for surgery; similarly, the value of radiomics as a stand-alone predictor is still unclear. The aim of this study was to develop a machine learning model that combines clinical and radiomic features to estimate early postoperative residual meningioma.

METHODS This retrospective multicenter study included 369 patients who underwent meningioma resection from 2020 to 2024 and had available preoperative contrast-enhanced T1-weighted MRI. Patient data from 3 centers (n = 307) were used to develop the model using leave-one-center-out cross-validation, and an independent cohort (n = 62) was used for external validation. Radiomic features were derived from manually segmented meningiomas, filtered for reproducibility, and integrated with 6 predefined clinical variables. The clinical-only, radiomics-only, and combined models were trained using 4 machine learning classifiers. Model performance was assessed through cross-validation, independent external validation, and receiver operating characteristic analysis. Formal incremental benefit analyses were performed on the common overlap external subset, with predictions available for all compared models.

RESULTS Residual meningioma occurred in 23.5% of patients in the development cohort and 14.5% in the external cohort. Lesion location and venous sinus involvement were significantly associated with residual status. Following the feature selection, 2 stable radiomic texture features were retained. In external validation, the combined radiomics-clinical k-nearest neighbors model achieved the highest area under the curve (0.821), with sensitivity of 0.889, specificity of 0.706, and accuracy of 0.733. Clinical variables provided most of the predictive value, whereas radiomic features pro-

ABBREVIATIONS AUC = area under the curve; BRSHH = Bakırkoy Prof. Dr. Mazhar Osman Psychiatry, Neurology, and Neurosurgery Training and Research Hospital; DCA = decision curve analysis; EOR = extent of resection; GAZI = Gazi University Faculty of Medicine Hospital; GEAH = Goztepe Prof. Dr. Suleyman Yalcin City Hospital; GTR = gross-total resection; ICC = intraclass correlation coefficient; IEAH = Istanbul Training and Research Hospital; KNN = k-nearest neighbors; LOCO = leave one center out; LR = logistic regression; ML = machine learning; RF = random forest; ROC = receiver operating characteristic; SHAP = Shapley Additive Explanations; XGBoost = Extreme Gradient Boosting.

SUBMITTED March 15, 2026. **ACCEPTED** June 11, 2026.
INCLUDE WHEN CITING DOI: 10.3171/2026.6.FOCUS26266.

vided only limited incremental value when added to the clinical model. Decision curve analysis revealed net benefit for the combined model within a narrow range of low thresholds.

CONCLUSIONS In this multicenter study, which included external validation, clinical variables provided most of the predictive value for postoperative residual meningioma. Radiomic features alone showed limited discrimination and only modest added value beyond clinical predictors. These combined models could serve as decision-support tools for preoperative risk assessment in meningioma surgery but are not yet suitable for routine stand-alone clinical use.

https://thejns.org/doi/abs/10.3171/2026.6.FOCUS26266

KEYWORDS machine learning; meningioma; MRI; prediction; radiomics; surgical resectability

Meningioma is the most common primary intracranial tumor, and extent of resection (EOR) remains a key determinant of recurrence and prognosis, traditionally evaluated using the Simpson grading classification.[1,2] However, gross-total resection (GTR) is often limited by a skull base location and venous sinus or neurovascular involvement combined with the need to preserve neurological function.[3] Therefore, preoperative prediction of MRI-defined residual status is vital for surgical strategy.[4,5] Therefore, balancing oncological radicality with functional preservation requires more precise preoperative tools for estimating resectability.[6,7]

Several clinical and radiological factors, including location, bone invasion, and venous sinus involvement, have been associated with incomplete resection.[7,8] However, these predictors are commonly evaluated individually, and reproducible models integrating clinical and imaging features for preoperative risk estimation remain limited.[7,9,10] Although traditional regression models are commonly accepted, machine learning (ML) algorithms often provide superior discriminative power.[11] ML might be particularly useful in meningioma studies for which sample sizes are often constrained.[12,13]

Radiomics and ML are increasingly implemented in meningioma research. These methods extract quantitative imaging features beyond visual evaluation and have shown promise in predicting tumor grade and biological behavior, especially when combined with clinical data.[14–17] However, the majority of published studies are retrospective, single-center, and focused on histological grading or recurrence rather than surgical resectability, and external validation remains insufficient.[10,18,19] Moreover, few studies have examined whether radiomic features offer additional value beyond well-known anatomical predictors of resection outcome.[19] Many models depend on surgeon-reported EOR rather than objective postoperative imaging findings, whereas imaging-based analyses provide a more objective endpoint, particularly in multicenter settings.[20] Current radiomic models remain limited by imaging heterogeneity and insufficient integration of clinical, anatomical, and radiomic data.[10,21]

In this multicenter study, we aimed to develop and externally validate an ML model for predicting early radiological residual meningioma by combining clinical variables with MRI-based radiomics. Particular attention was given to whether known clinical and anatomical predictors could be incorporated into a reproducible model for objective preoperative risk assessment, and whether radiomics provides measurable incremental value beyond these predictors.

Methods

Study Design and Patient Population

Patients who underwent meningioma resection from January 2020 to December 2024 were retrospectively identified from four tertiary neurosurgical centers: Istanbul Training and Research Hospital (IEAH); Bakırkoy Prof. Dr. Mazhar Osman Psychiatry, Neurology, and Neurosurgery Training and Research Hospital (BRSHH); Gazi University Faculty of Medicine Hospital (GAZI); and Goztepe Prof. Dr. Suleyman Yalcin City Hospital (GEAH). Three centers (IEAH, BRSHH, and GAZI) constituted the development cohort, and GEAH was reserved as an independent external cohort. Patients with recurrent meningioma or prior treatment for the index lesion were not included. The study protocol was approved by the local ethics committee of IEAH, and informed consent was waived based on the retrospective nature of the study.

The primary outcome was the presence of residual meningioma within 24–72 hours on postoperative contrast-enhanced T1-weighted MRI, defined as any focal enhancement distinct from surgical changes. No minimum volume threshold was applied. A binary endpoint was selected over a graded stratification because the limited event would have rendered per-group sample sizes insufficient for stable model development, and standardized volumetric cutoffs for residual status have not been established.[22] Postoperative residual volume was calculated only as a secondary descriptive analysis. Residual status was determined by a radiologist who was blinded to the intraoperative resection assessment.

Imaging Acquisition, Segmentation, and Radiomic Feature Extraction

Preoperative contrast-enhanced T1-weighted MRI was acquired at all centers using Siemens Healthineers scanners at 1.5T (IEAH, BRSHH, GEAH) or 3T (GAZI) field strengths. Both 2D axial T1-weighted spin-echo and 3D T1-weighted gradient-echo sequences were used after gadolinium-based contrast administration. Acquisition parameters varied by center and sequence type with the following ranges: repetition time 580–2720 msec, echo time 2.9–15 msec, slice thickness 1–5 mm, and in-plane resolution from approximately 0.5 × 0.5 mm to 1.0 × 1.0 mm. Acquisition parameter heterogeneity across centers was harmonized through standardized preprocessing prior to feature extraction. Tumor segmentation was performed manually on preoperative contrast-enhanced T1-weighted MRI using 3D Slicer (version 5.10.0) by two board-certified radiologists. Segmentation masks were

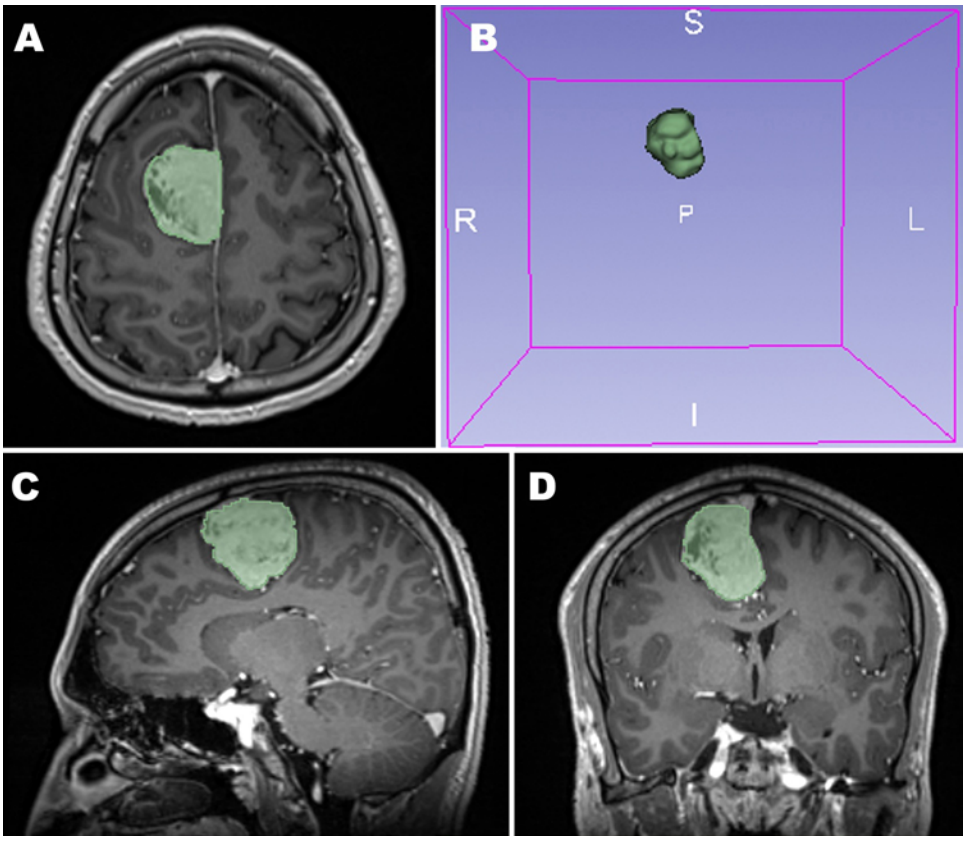


FIG. 1. Representative example of manual segmentation used for radiomic analysis. Contrast-enhanced T1-weighted MR images showing the segmented meningioma in axial (**A**), sagittal (**C**), and coronal (**D**) planes, together with the corresponding 3D reconstruction (**B**). The *green mask* indicates the segmentation used for radiomic feature extraction.

stored in NIfTI (Neuroimaging Informatics Technology Initiative) format. An example of manual segmentation and 3D reconstruction applied for radiomic feature extraction is presented in Fig. 1.

Radiomic features were extracted using PyRadiomics (version 3.1.0). Images were resampled to 1 × 1 × 1 mm^3 isotropic voxel spacing using B-spline interpolation for images and nearest-neighbor interpolation for segmentation masks. Intensity normalization was performed with a normalizeScale parameter of 100, which centered each image according to its own mean and scaled intensity values to a standard deviation of 100. Gray-level discretization was performed using a fixed bin width of 25. Extracted features consisted of first-order statistics, with the texture features of GLCM (gray-level co-occurrence matrix), gray-level run-length matrix, gray-level size-zone matrix, gray-level dependence matrix, and neighboring gray-tone difference matrix, as well as 3D shape descriptors. Features were analyzed from original, Laplacian-of-Gaussian, and wavelet-filtered images. This pipeline yielded a high-dimensional radiomic feature set, with all features prefixed with "t1c_" to indicate their contrast-enhanced T1-weighted origin.

Feature reproducibility was assessed with the intraclass correlation coefficient (ICC) based on dual-reader segmentations in a subset of cases. ICC(3,1), a 2-way mixed-effects single-measure consistency model, was implemented manually to evaluate segmentation-dependent feature stability. Radiomic features with ICC > 0.80 were classified as reproducible and retained for subsequent analysis, whereas features with lower ICC, zero variance, or inadequate dual-reader data were excluded. Thus, only segmentation-stable radiomic features were integrated into the ML pipeline.

Clinical Variables

The following 6 clinical variables were identified as potential predictors: patient age at surgery, sex, primary location, peritumoral edema, bone invasion, and involvement of venous sinus. These parameters were selected a priori based on clinical relevance and availability of data across centers. WHO grade was recorded for cohort characterization but excluded from the modeling; only the 6 preoperative variables were included. Resection status and WHO grade serve solely for cohort description.

Although the surgeon's intraoperative assessment of resection (GTR vs subtotal resection) was recorded, it was considered an outcome-adjacent variable rather than a preoperative baseline factor. The agreement between surgeon assessment and MRI-confirmed residual status was evaluated using Cohen's kappa (κ) across cohorts to support objective imaging. There were no missing data for the variables included in the final analysis.

ML Pipeline

The three-center dataset was used to develop the model, while the GEAH dataset was used for independent external validation. In the development cohort, model training and internal validation were performed with the leave-one-center-out (LOCO) cross-validation. In each LOCO fold, patients from one development center were held out as the validation set, while patients from the other two centers were used for training, allowing evaluation of intercenter generalizability. All preprocessing, feature selection, class balancing, and model training procedures were conducted only on the training subsets of each cross-validation fold to prevent data leakage. Figure 2 summarizes the workflow for radiomic feature extraction, model construction, and validation.

Feature selection was executed through a 5-step dimensionality reduction pipeline. Univariable logistic regression (LR) was used to calculate probability values, areas under the curve (AUCs), and odds ratios, followed by Sure Independence Screening ($p > 0.5$). Redundant features were eliminated using Spearman correlation (> 0.8) and variance inflation factor (> 4). Features with $p < 0.2$ or an AUC > 0.55 were kept for further analysis. Then, LASSO (Least Absolute Shrinkage and Selection Operator, $\alpha = 0.005$) was used for final selection. This procedure was repeated independently within each LOCO training fold, and only features selected across folds were retained. In the clinical-only model, all 6 clinical features were integrated without feature selection.

SMOTE (Synthetic Minority Over-Sampling Technique) was applied to the training partitions within each cross-validation fold but never applied to the validation or external sets. Synthetic samples were generated using k-nearest neighbor interpolation, with the number of neighbors defined as k = min(5, n_minority-1), whereby n_minority is the number of minority-class samples in the training fold. For each data modality, 4 ML classifiers were evaluated: Extreme Gradient Boosting (XGBoost), random forest (RF), k-nearest neighbors (KNN), and LR. Thus, 12 models were evaluated across the 3 modalities (clinical-only, radiomics-only, and combined). Within each data modality, all 4 classifiers were trained using the same feature set.

After hyperparameter optimization, the best hyperparameter configuration with the highest mean validation AUC across LOCO folds was used to train models within each fold. The model was trained on the corresponding training partition in each fold, resulting in 3 trained models. The probabilities from all fold-specific models were averaged to produce a single final probability estimate per patient to predict new data.

For models sensitive to feature scale (LR and KNN), feature standardization was performed with z-score normalization. StandardScaler was fitted exclusively to the training data of the development set, and the same transformation was applied to both the 3 validation folds and the external set to prevent data leakage. Tree-based models (XGBoost and RF) were trained with the original feature scales.

Three model configurations were evaluated: 1) A clinical-only model, which served as a baseline, used preoperative characteristics (age, sex, location, edema, bone invasion, and involvement of venous sinus). 2) A radiomics-only model used stable contrast-enhanced T1-weighted features. 3) A combined model assessed the incremental predictive value of radiomics by integrating clinical characteristics with radiomic features.

Statistical Analysis

Model performance was evaluated using LOCO cross-validation within the development cohort and independent external validation. The performance measures included AUC, accuracy, sensitivity, specificity, F1-score, and predictive values. Receiver operating characteristic (ROC) curves were generated from predicted probabilities, and the optimal threshold was determined using the Youden index from cross-validation predictions. Corresponding 95% confidence intervals were estimated using 2000 bootstrap resamples of the external cohort. Statistical significance was set at $p < 0.05$. The best combined and best clinical-only models were compared on the external cohort using the AUC (DeLong's test), Brier score, and Hosmer-Lemeshow test. Calibration was assessed with intercept and slope, while continuous NRI (Net Reclassification Improvement) and IDI (Integrated Discrimination Improvement) were calculated to evaluate reclassification. Within each data modality, the best nonlinear ML classifier was compared with LR using DeLong's test for paired ROC curves.

Decision curve analysis (DCA) was performed to evaluate the net clinical benefit of the best-performing model from each modality across threshold probabilities ranging from 0.01 to 0.99. Net benefit was compared with treat-all and treat-none strategies. DCA was performed in the external cohort in which predictions were available for all compared modalities.

To improve model interpretability, Shapley Additive Explanations (SHAP) were performed for the best-performing model of each modality. TreeExplainer was used for tree-based models (XGBoost and RF), and KernelExplainer was used for KNN and LR. SHAP values were calculated on the external dataset to interpret the final model predictions, and global feature importance was determined as the mean absolute SHAP value across the entire external cohort.

Statistical analyses were conducted in Python (version 3.13). PyRadiomics (version 3.1.0), scikit-learn, XGBoost, and SHAP were used for feature extraction and model interpretation. A fixed random seed of 42 was used to ensure reproducibility.

Results

Patient Characteristics and Surgeon-MRI Agreement

Overall, 369 eligible patients with preoperative contrast-enhanced T1-weighted MRI who underwent meningioma resection were included in this analysis (Table 1). The development cohort comprised 307 patients from 3 centers (IEAH, n = 117; BRSHH, n = 100; and GAZI, n = 90), and the external validation cohort comprised 62 patients. Residual meningioma was observed in 23.5% (72/307) of patients in the development cohort and 14.5% (9/62) in the

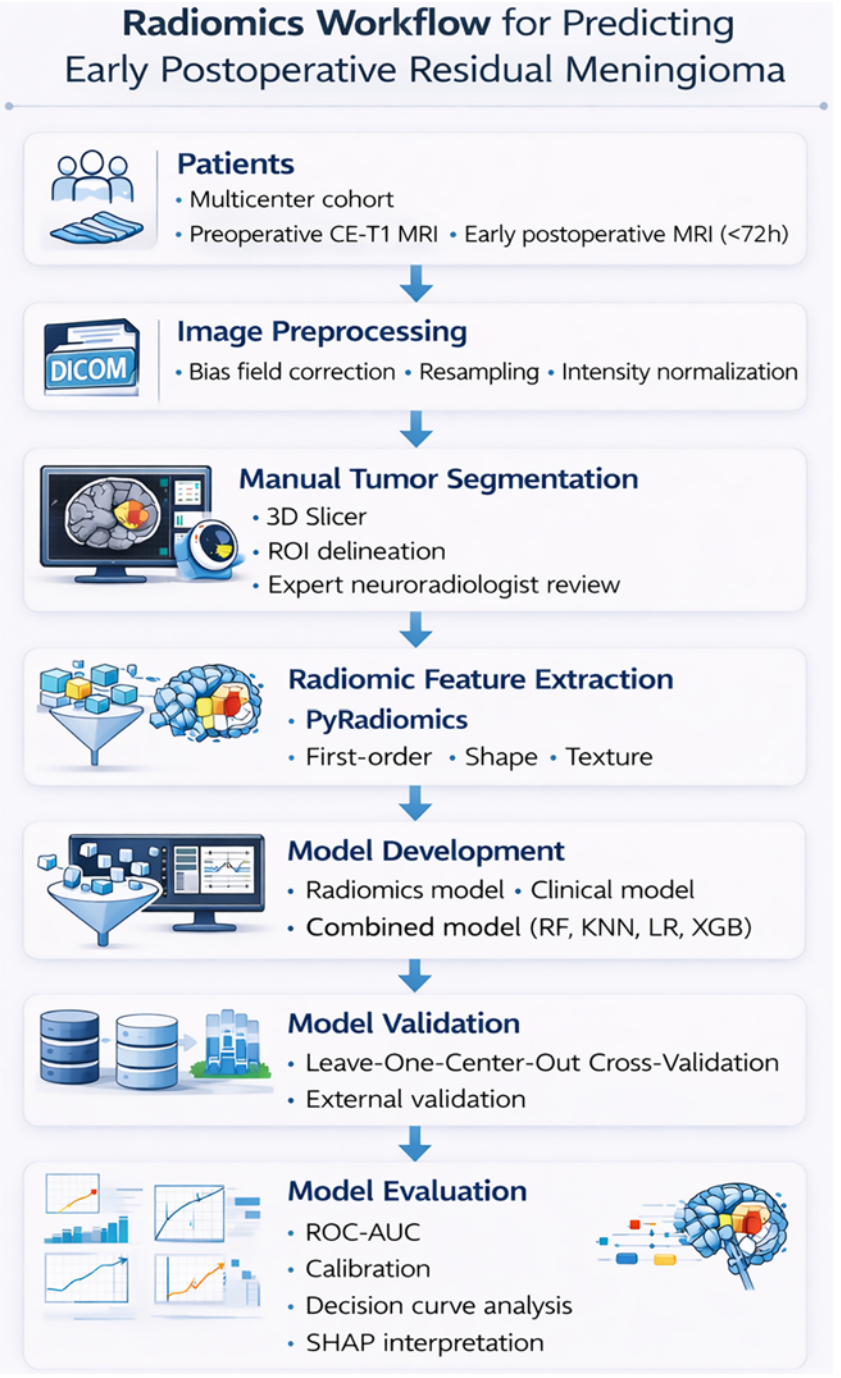


FIG. 2. Schematic showing the multicenter ML pipeline for prediction of early postoperative radiological residual in meningioma surgery. Preoperative contrast-enhanced T1-weighted (CE-T1) MRI underwent standardized preprocessing followed by manual segmentation performed by two board-certified radiologists with expert quality control. Radiomic features and structured anatomical variables were integrated to develop radiomics-only, clinical-only, and combined ML models evaluated across multiple classifiers. Model performance was evaluated using LOCO cross-validation and independent external validation, with AUC, calibration, DCA, and SHAP-based interpretability. h = hours; ROI = region of interest. Created in BioRender. Sanlier, N. (2026) https://BioRender.com/1inakc9.

external cohort. Patients with residual tumor (n = 81) more frequently had skull base meningiomas (43.2% vs 21.9%) and posterior fossa lesions (19.8% vs 12.2%) compared with patients without residual tumor (n = 288). Histological WHO grade distribution is summarized in Table 1. Lesion location ($p < 0.001$) and venous sinus involvement ($p < 0.001$) were significantly associated with residual status. Other baseline characteristics did not differ significantly

TABLE 1. Baseline characteristics of patients who underwent meningioma resection stratified by postoperative residual meningioma status

Characteristic	No Residual	Residual	Total	p Value
No. of patients	288	81	369	
Mean age ± SD, yrs	55.7 ± 14.0	54.5 ± 13.0	55.4 ± 13.8	0.482
Male sex	80 (27.8)	33 (40.7)	113 (30.6)	**0.029**
Lesion location				**<0.001**
Convexity	107 (37.2)	12 (14.8)	119 (32.2)	
Parasagittal/falx	79 (27.4)	17 (21.0)	96 (26.0)	
Skull base	63 (21.9)	35 (43.2)	98 (26.6)	
Posterior fossa	35 (12.2)	16 (19.8)	51 (13.8)	
Other	4 (1.4)	1 (1.2)	5 (1.4)	
Surgeon-reported EOR				
GTR	280 (97.2)	21 (25.9)	301 (81.6)	**<0.001**
STR	8 (2.8)	60 (74.1)	68 (18.4)	
WHO grade				0.611
1	230 (79.9)	61 (75.3)	291 (78.9)	
2	50 (17.4)	18 (22.2)	68 (18.4)	
3	4 (1.4)	1 (1.2)	5 (1.4)	
Missing	4 (1.4)	1 (1.2)	5 (1.4)	
Peritumoral edema				
None	128 (44.4)	30 (37.0)	158 (42.8)	
Present	160 (55.6)	51 (63.0)	211 (57.2)	0.254
Bone invasion	36 (12.5)	16 (19.8)	52 (14.1)	0.105
Venous sinus involvement				**<0.001**
No contact	198 (68.8)	41 (50.6)	239 (64.8)	
Contact	40 (13.9)	27 (33.3)	67 (18.2)	
Suspected invasion	50 (17.4)	13 (16.0)	63 (17.1)	

STR = subtotal resection.
Values are presented as the number of patients (%), unless indicated otherwise. Patients with missing data were excluded from between-group comparisons. Boldface type indicates statistical significance ($p < 0.05$). For comparisons, the Mann-Whitney U-test was used for continuous variables, Fisher's exact test for binary variables, and chi-square test for categorical variables with ≥ 3 groups.

between the two groups. Among patients with residual tumor, 61.7% (50/81) had residual volume < 5 cm^3 and 61.7% (50/81) had EOR > 75% (Supplementary Table 1). Surgeon-MRI agreement was substantial, with an agreement rate of 92.0% and κ value of 0.75 in the overall cohort. The κ value was 0.73 (agreement 91.0%) in the development cohort and 0.86 (agreement 96.8%) in the external cohort.

Radiomic Feature Extraction and Selection

Radiomic feature extraction was successfully completed for all 369 patients, with 1595 radiomic features extracted per patient from the original LoG-filtered wavelet-transformed images. Dual-reader segmentations in 30 patients enabled reproducibility evaluation. ICC analysis determined 1049 features (65.8%) with ICC(3,1) > 0.80, which were retained for model development. Subsequent feature selection retained 2 wavelet-derived texture features (t1c_wavelet-LHH_glcm_Correlation, and t1c_wavelet-LHL_glcm_Correlation) for the radiomics-only model. The final combined model incorporated these 2 radiomic features together with 6 predefined clinical variables, yielding a total of 8 predictors.

Model Performance and Interpretation

Internal validation was performed with LOCO cross-validation within the development cohort. The combined XGBoost model achieved the highest mean cross-validated AUC (0.765), followed by the combined KNN model (AUC = 0.757) and clinical XGBoost model (AUC = 0.756). Clinical-only models exhibited moderate discrimination among classifiers (AUC range 0.730–0.756), while radiomics-only models demonstrated inferior discrimination, nearing chance-level performance (AUC range 0.544–0.634).

External validation was conducted using the independent GEAH cohort. The combined radiomics–clinical KNN model achieved the highest AUC (0.821, 95% CI 0.664–0.942), with sensitivity of 0.889, specificity of 0.706, precision of 0.348, and overall accuracy of 0.733. The best-performing clinical-only model (RF) demon-

FIG. 3. Model performance comparison across data modalities. Heat map showing performance metrics for the 12 ML models developed using combined, clinical-only, and radiomics-only feature sets in the independent external validation cohort. Metrics include the validation (Val) AUC from LOCO cross-validation (CV), test AUC, F1-score, recall, and accuracy. Cell coloration represents relative performance, with *blue* indicating better performance and *orange* indicating worse performance. The best-performing classifier for each metric within its modality group is marked with a *star*. Combined models consistently outperformed single-modality models. The KNN model showed the highest discriminative power when tested on external data. T1C = contrast-enhanced T1-weighted.

strated moderate discrimination (AUC = 0.741, 95% CI 0.579–0.882), whereas the best radiomics-only model (XGBoost) had restricted predictive capability (AUC = 0.522, 95% CI 0.304–0.746) (Fig. 3). Among all 12 models, AUCs ranged from 0.590 to 0.821 (combined), from 0.632 to 0.741 (clinical-only), and from 0.414 to 0.522 (radiomics-only) (Fig. 3).

Incremental Value of Radiomics and ML

The best nonlinear classifier significantly outperformed LR in the clinical-only (ΔAUC +0.100, p = 0.027) and combined (ΔAUC +0.231, p < 0.001) modalities, but not in the radiomics-only modality (ΔAUC +0.103, p = 0.261). Brier scores changed most in the clinical-only and radiomics-only models (ΔBrier +0.076 and +0.028, respectively); whereas the gain in the combined model was minimal (+0.010) (Table 2).

In the head-to-head external comparison, the combined KNN model achieved a higher AUC than the clinical-only RF model (0.821 vs 0.741) but this difference was not significant (ΔAUC +0.080; paired DeLong's test, p = 0.22). Calibration was inferior for the combined KNN model, with a higher Brier score (0.198 [95% CI 0.160–0.240] vs 0.132 [95% CI 0.090–0.176]), steeper calibration slope (1.53 vs 0.97), and significant Hosmer-Lemeshow test (p < 0.001 vs p = 0.07) (Supplementary Fig. 1 and Supplementary Table 2). Despite this, reclassification modestly improved (continuous NRI +0.218 and IDI +0.072). Among the combined models, KNN performance (AUC = 0.821) exceeded XGBoost (AUC = 0.723; ΔAUC +0.098, p = 0.109) and significantly outperformed RF (ΔAUC +0.161, p = 0.040). Among the clinical-only models, RF exceeded LR (ΔAUC = 0.109, p = 0.019).

DCA demonstrated that the combined KNN model provided net benefit within a narrow low range, from approximately 0.10 to 0.19, remaining above both the treat-all and treat-none strategies (Fig. 4A). The clinical-only RF model had a more restricted beneficial window, while the radiomics-only XGBoost model provided limited net benefit across the evaluated threshold range. The corresponding confusion matrix is presented in Fig. 4B.

SHAP analysis was used to interpret model predictions. In the clinical RF model, the most influential variables

TABLE 2. Incremental gain of the top ML classifier over baseline logistic regression across modalities in the external validation cohort

Modality	Best ML Classifier	AUC, ML	AUC, LR	ΔAUC	p Value*	Brier, ML	Brier, LR	ΔBrier
Clinical	RF	0.741	0.641	+0.100	0.027	0.132	0.208	+0.076
Radiomics	XGBoost	0.522	0.418	+0.103	0.261	0.221	0.249	+0.028
Combined	KNN	0.821	0.590	+0.231	<0.001	0.198	0.208	+0.010

ΔAUC = AUC of best ML classifier minus AUC of LR model trained on the same LASSO-selected feature set; ΔBrier = Brier score of LR minus Brier score of ML (positive values favor ML).
Within each modality, classifiers used identical preprocessing pipelines and feature sets; therefore, differences reflect the contribution of the algorithm itself.
* DeLong's test for paired ROC curves.

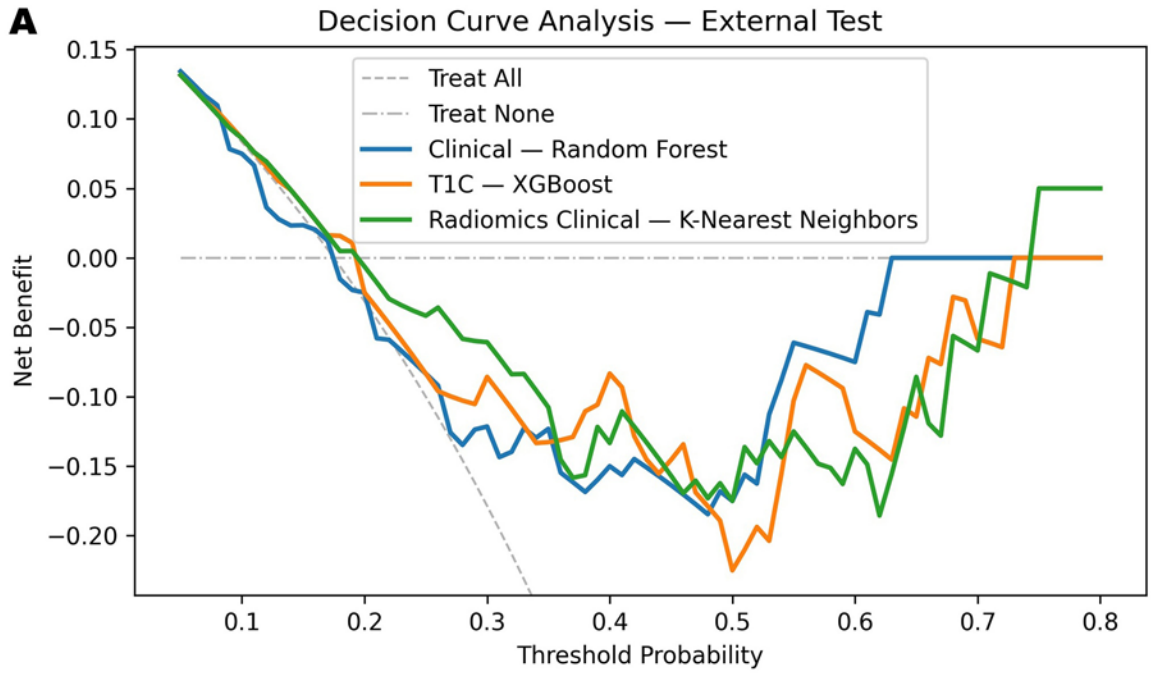
were location, sex, age, and venous sinus involvement (Fig. 5A). Preoperative edema and bone invasion contributed to a lesser extent. Higher SHAP values for skull base location were associated with increased probability of residual tumor. In the combined model, t1c_wavelet-LHH_glcm_Correlation ranked among the top predictors alongside tumor location, sex, and venous sinus involvement, while the t1c_wavelet-LHL_glcm_Correlation feature contributed less (Fig. 5B). ROC analyses for the external validation cohort are shown in Fig. 6, including the best-performing combined KNN model, and that model compared with the best-performing clinical-only RF model.

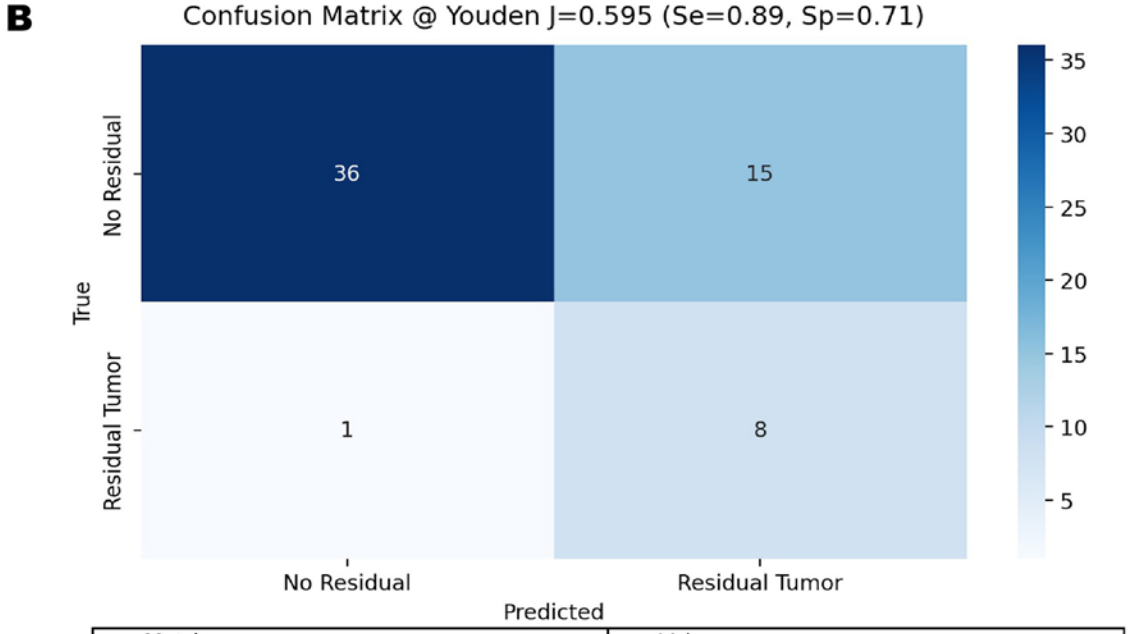
Discussion

GTR remains the primary surgical objective in meningioma management, yet preoperative prediction of resectability remains challenging, particularly for meningiomas exhibiting complex anatomical relationships.[7,9] In our study, location was the main predictor of residual meningioma, consistent with evidence that a skull base location and venous sinus involvement limit complete resection.[5] This confirms that resectability is anatomically driven, emphasizing the importance of objective imaging-based assessment. Preoperative prediction of residual meningioma likelihood might facilitate personalized surgical planning and more accurate patient counseling.[23]

In this multicenter study, anatomical-clinical variables accounted for most of the predictive value for residual meningioma, whereas radiomic features provided only limited incremental value when combined with clinical predictors. Although prior radiomics studies have addressed either single-center resection prediction,[19] multicenter deep learning models,[24] or multicenter combined models for recurrence,[16] less work has combined these elements specifically for MRI-defined residual disease with formal incremental-benefit metrics. The contribution of the current study is the integration of known predictors into a standardized framework and quantification of the incremental value of radiomics, rather than identifying new predictors. The combined model achieved the highest AUC (0.821), but its improvement over the best clinical-only model was modest and was not significant. Calibration of the combined model was also inferior, suggesting that radiomics added value primarily for ranking patients by risk rather than for generating well-calibrated absolute risk estimates. The reclassification pattern aligns with this interpretation, as the improvement was primarily driven by better event classification, at the cost of shifting more nonevents toward higher predicted risk. These findings suggest that radiomic features might provide complementary information without replacing established anatomical predictors, consistent with prior studies demonstrating that combined clinical-radiomic models might numerically outperform single-modality approaches.[16,24] However, broader clinical implementation remains limited by challenges related to calibration, reproducibility, and external validation.[10,21] The two retained features (wavelet-domain GLCM correlation indices on contrast-enhanced T1-weighted MRI) reflect the spatial linear dependence of voxel intensities. These features have been linked to intratumoral heterogeneity across tumor types[25,26] and to brain and sinus invasion in meningioma specifically.[18] Although histopathological correlation is needed, their LASSO retention suggests prediction-relevant texture information beyond global tumor characteristics.

Clinical models demonstrated moderate performance, while radiomics-only models yielded near-chance discrimination, consistent with prior reports on the limited standalone performance of ML models in meningioma.[11,16,17] These findings support the concept that radiomic features alone are insufficient but can provide complementary value when integrated with clinical variables. Building on recent multialgorithm benchmarking efforts in meningioma radiomics,[27,28] we compared multiple classifiers to select the best-performing model.[29,30] Although deep learning studies report high accuracy, combined clinical-radiomic models might provide better interpretability and clinical utility.[24] Our study extends these findings by using a multicenter design with external validation and MRI-confirmed data. The observed discordance between surgeon assessment and postoperative imaging further supports the need for objective imaging-based prediction tools, emphasizing the potential value of combined models. The near-chance performance of radiomics-only models is consistent with the observation that conventional intratumoral features might not capture peritumoral or spatially extended anatomical relationships that are relevant to surgical resectability.[31] Manual segmentation variability and limited standardization further constrain radiomics-only approaches.[32] Deep learning approaches using 3D convolutional networks might better capture peritumoral and vascular features that conventional radiomics cannot capture.[24]





Metric	Value
Sensitivity	0.889
Specificity	0.706
Precision	0.348
F1-score	0.500
Accuracy	0.733
TP / TN / FP / FN	8/36/15/1

FIG. 4. Clinical utility and classification predictive performance of the models. **A:** Decision curves showing the net clinical benefit in the independent external cohort. The combined radiomics-clinical KNN model (*green*) exhibits a greater net benefit at lower decision thresholds compared with the clinical-only (*blue*), radiomics-only (*orange*), and treat-all (*dashed gray line*) strategies. **B:** Confusion matrix (*upper*) and chart (*lower*) showing performance metrics of the most effective radiomics+clinical KNN model, which was evaluated at the optimal Youden threshold with sensitivity (Se) of 0.889 and specificity (Sp) of 0.706, in the external test set. FN = false negative; FP = false positive; TN = true negative; TP = true positive.

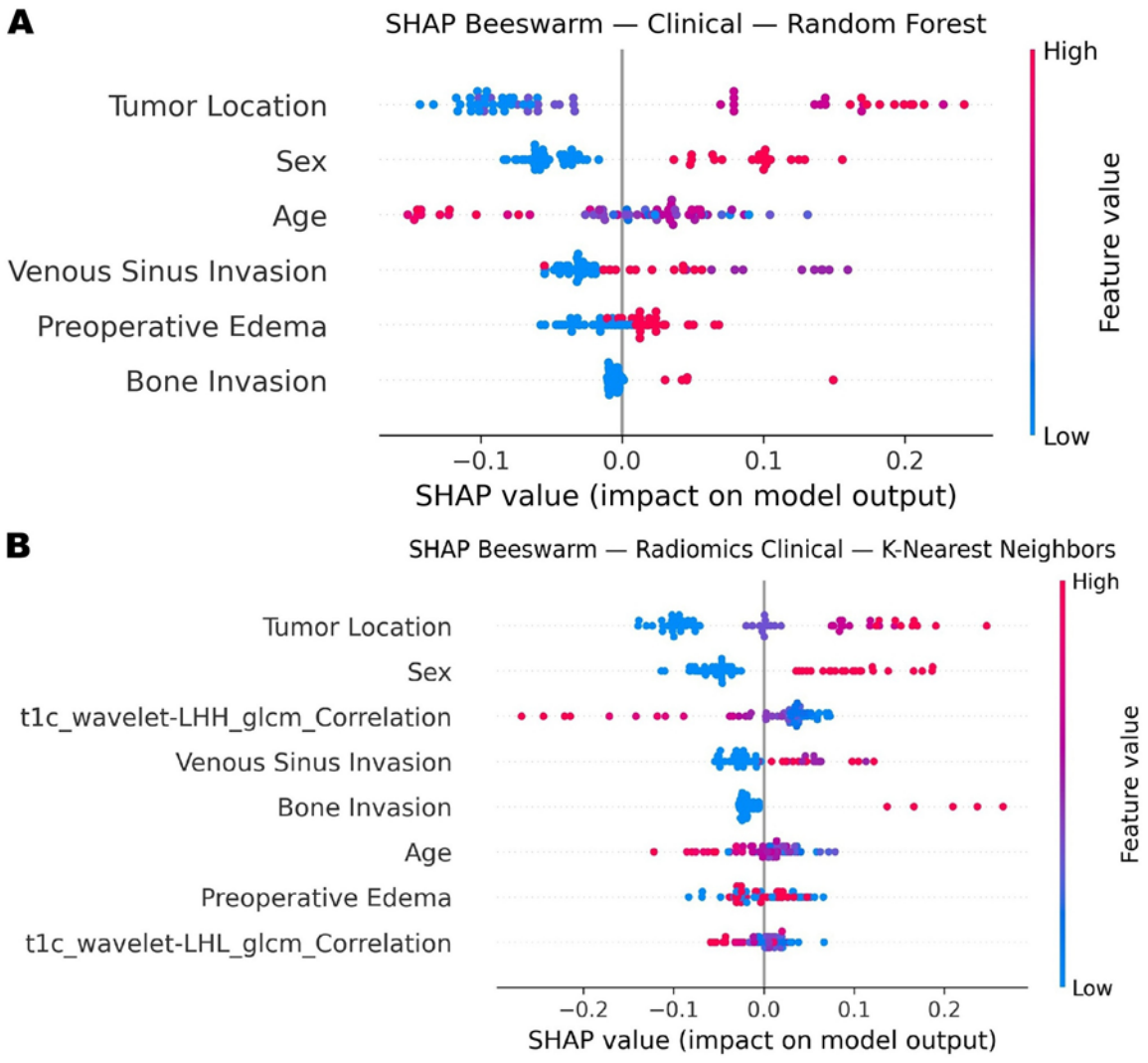


FIG. 5. Model interpretability using SHAP analysis. Beeswarm plots showing feature importance for the clinical RF (**A**) and combined KNN (**B**) models. Each *point* represents an individual patient, with *red* indicating high and *blue* indicating low feature values. Positive SHAP values indicate increased predicted probability of residual meningioma, whereas negative values indicate lower predicted probability. Meningioma location and venous sinus involvement are the most influential predictors, while radiomic texture features provide additional discriminative weight in the combined configuration.

DCA showed modest net benefit only across a limited threshold range, suggesting a possible supportive role rather than as a stand-alone tool. Given the low external event prevalence and associated false-positive predictions, the operating threshold is better interpreted as a triage-oriented threshold than confirmatory. With 15 false-positive predictions for every 8 true-positive predictions in the external cohort, most model-flagged high-risk cases lacked actual residual disease. Clinically, the combined model might function better as a high-sensitivity preoperative risk-stratification tool to identify patients who might benefit from closer postoperative follow-up, rather than as a model for precise absolute risk estimation. The low precision reflects a sensitivity-precision tradeoff, which is clinically rational when missing the identification of high-risk patients carries greater consequences than over-triage. Consequently, the combined model might prove most beneficial in anatomically complex cases, such as skull base meningiomas with unclear sinus involvement, serving to complement rather than replace expert opinion. Further validation in larger and more diverse cohorts is needed before broader clinical implementation. Future studies should also evaluate residual tumor burden, automated segmentation, and deep learning pipelines to automate imaging-based clinical variable assessment.

This study had several limitations. First, reliance on a single MRI sequence (contrast-enhanced T1-weighted) and the small external cohort with a low event rate might limit model performance, as well as statistical power and precision, while manual segmentation might affect scalability. Second, residual status was analyzed as a binary endpoint rather than stratified by residual burden as re-

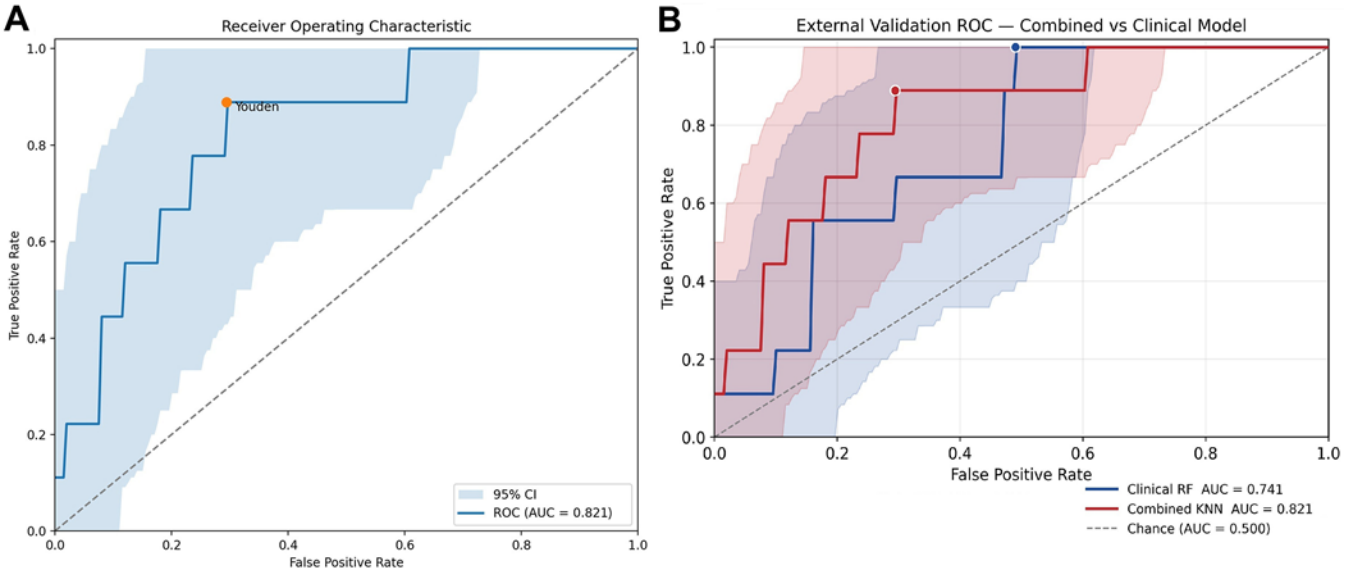


FIG. 6. ROC analysis in the independent external validation cohort. **A:** ROC curve showing the discriminative ability of the best-performing combined KNN model (AUC = 0.821, 95% CI 0.664–0.942). *Shading* represents the bootstrap-derived confidence interval, and the *orange dot* indicates the optimal classification threshold as determined by the Youden index. **B:** ROC curves showing the relative discriminative ability of the combined KNN and clinical-only RF models. The combined KNN model achieved a higher AUC than the RF model (AUC = 0.741, 95% CI 0.579–0.882). *Dots* indicate optimal thresholds and the *dashed line* represents chance performance (AUC = 0.50).

sidual events were limited, particularly in the external cohort. Therefore, postoperative residual volume was reported only as a descriptive secondary analysis, and future work in larger cohorts should re-evaluate residual burden as a graded outcome. Third, several potentially important determinants of resectability, including surgical approach, intraoperative anatomical findings, and surgeon-specific factors, were not incorporated. Finally, calibration of the combined model was imperfect in the external cohort, so recalibration would be required before any clinical use of predicted probabilities.

Conclusions

Clinical variables provided most of the predictive value for residual meningioma, while radiomics showed limited stand-alone discrimination and only modest added value beyond clinical predictors. The combined model achieved the highest numerical performance, suggesting that radiomic features might complement established anatomical predictors. This study demonstrates that known risk factors might support externally validated and objective preoperative risk estimation. These models could serve as decision-support tools for selected risk-stratification scenarios, but they are not yet sufficient for stand-alone clinical use.

References

1. Dincer A, Morales-Valero SF, Robert SM, et al. Surgical strategies for intracranial meningioma in the molecular era. *J Neurooncol*. 2023;162(2):253-265. doi:10.1007/s11060-023-04272-z
2. Goldbrunner R, Minniti G, Preusser M, et al. EANO guidelines for the diagnosis and treatment of meningiomas. *Lancet Oncol*. 2016;17(9):e383-e391. doi:10.1016/S1470-2045(16)30321-7
3. Anthofer J, Seidel-Schulz R, Proescholdt M, Brawanski A, Schebesch KM. Meningiomas adjacent to major venous sinuses-clinical outcome and recurrence. *World Neurosurg*. 2017;104:560-566. doi:10.1016/j.wneu.2017.05.025
4. Schwartz TH, McDermott MW. The Simpson grade: abandon the scale but preserve the message. *J Neurosurg*. 2021;135(2):488-495. doi:10.3171/2020.6.JNS201904
5. Wang YC, Chuang CC, Wei KC, et al. Skull base atypical meningioma: long term surgical outcome and prognostic factors. *Clin Neurol Neurosurg*. 2015;128:112-116. doi:10.1016/j.clineuro.2014.11.009
6. Corniola MV, Lemée JM, Da Broi M, et al. Posterior fossa meningiomas: perioperative predictors of extent of resection, overall survival and progression-free survival. *Acta Neurochir (Wien)*. 2019;161(5):1003-1011. doi:10.1007/s00701-019-03862-z
7. Lemée JM, Corniola MV, Da Broi M, et al. Extent of resection in meningioma: predictive factors and clinical implications. *Sci Rep*. 2019;9(1):5944. doi:10.1038/s41598-019-42451-z
8. Takase H, Yamamoto T. Bone invasive meningioma: recent advances and therapeutic perspectives. *Front Oncol*. 2022;12:895374. doi:10.3389/fonc.2022.895374
9. Aizer AA, Bi WL, Kandola MS, et al. Extent of resection and overall survival for patients with atypical and malignant meningioma. *Cancer*. 2015;121(24):4376-4381. doi:10.1002/cncr.29639
10. Song D, Cai R, Lou Y, et al. Advancements in the application of MRI radiomics in meningioma. *Radiat Oncol*. 2025;20(1):105. doi:10.1186/s13014-025-02679-8
11. Abualnaja SY, Morris JS, Rashid H, Cook WH, Helmy AE. Machine learning for predicting post-operative outcomes in meningiomas: a systematic review and meta-analysis. *Acta*

Neurochir (Wien). 2024;166(1):505. doi:10.1007/s00701-024-06344-z
12. An C, Park YW, Ahn SS, Han K, Kim H, Lee SK. Radiomics machine learning study with a small sample size: single random training-test set split may lead to unreliable results. *PLoS One*. 2021;16(8):e0256152. doi:10.1371/journal.pone.0256152
13. Hale AT, Stonko DP, Wang L, Strother MK, Chambless LB. Machine learning analyses can differentiate meningioma grade by features on magnetic resonance imaging. *Neurosurg Focus*. 2018;45(5):E4. doi:10.3171/2018.8.FOCUS18191
14. Xu N, Ma W, Dong W, Yin B. A scoping review of artificial intelligence applications in meningioma from image analysis to prognostic prediction. *Discov Oncol*. 2026;17(1):319. doi:10.1007/s12672-026-04476-5
15. Hamerla G, Meyer HJ, Schob S, et al. Comparison of machine learning classifiers for differentiation of grade 1 from higher gradings in meningioma: a multicenter radiomics study. *Magn Reson Imaging*. 2019;63:244-249. doi:10.1016/j.mri.2019.08.011
16. Ren L, Chen J, Deng J, et al. The development of a combined clinico-radiomics model for predicting post-operative recurrence in atypical meningiomas: a multicenter study. *J Neurooncol*. 2024;166(1):59-71. doi:10.1007/s11060-023-04511-3
17. Saroh S, Pendem S, Prakashini K, et al. Machine learning based radiomics approach for outcome prediction of meningioma - a systematic review. *F1000Res*. 2025;14:330. doi:10.12688/f1000research.162306.1
18. Gui Y, Hu W, Ren J, et al. Preoperative diagnosis of meningioma sinus invasion based on MRI radiomics and deep learning: a multicenter study. *Cancer Imaging*. 2025;25(1):20. doi:10.1186/s40644-025-00845-5
19. Musigmann M, Akkurt BH, Krähling H, et al. Assessing preoperative risk of STR in skull meningiomas using MR radiomics and machine learning. *Sci Rep*. 2022;12(1):14043. doi:10.1038/s41598-022-18458-4
20. Brokinkel B, Spille DC, Brokinkel C, et al. The Simpson grading: defining the optimal threshold for gross total resection in meningioma surgery. *Neurosurg Rev*. 2021;44(3):1713-1720. doi:10.1007/s10143-020-01369-1
21. Bakas S, Vollmuth P, Galldiks N, et al. Artificial Intelligence for Response Assessment in Neuro Oncology (AI-RANO), part 2: recommendations for standardisation, validation, and good clinical practice. *Lancet Oncol*. 2024;25(11):e589-e601. doi:10.1016/S1470-2045(24)00315-2
22. Gillespie CS, Taweel BA, Richardson GE, et al. Volumetric growth of residual meningioma - a systematic review. *J Clin Neurosci*. 2021;91:110-117. doi:10.1016/j.jocn.2021.06.033
23. Chotai S, Schwartz TH. The Simpson grading: is it still valid? *Cancers (Basel)*. 2022;14(8):2007. doi:10.3390/cancers14082007
24. Akkurt BH, Wanderer S, Schwyzer L, et al. Predicting meningioma resection status: use of deep learning. *Acad Radiol*. 2023;30(7):1232-1237. doi:10.1016/j.acra.2022.10.004
25. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006. doi:10.1038/ncomms5006
26. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: the promise of radiomics. *Phys Med*. 2017;38:122-139. doi:10.1016/j.ejmp.2017.05.071
27. Karabacak M, Patil S, Feng R, Shrivastava RK, Margetis K. A large scale multi institutional study for radiomics driven machine learning for meningioma grading. *Sci Rep*. 2024;14(1):26191. doi:10.1038/s41598-024-78311-8
28. Zhang S, Richter J, Veale J, et al. Development of hybrid radiomic machine learning models for preoperative prediction of meningioma grade on multiparametric MRI. *J Clin Neurosci*. 2025;135:111118. doi:10.1016/j.jocn.2025.111118
29. Stulp F, Sigaud O. Many regression algorithms, one unified model: a review. *Neural Netw*. 2015;69:60-79. doi:10.1016/j.neunet.2015.05.005
30. Zhang S, Li X, Zong M, Zhu X, Wang R. Efficient kNN classification with different numbers of nearest neighbors. *IEEE Trans Neural Netw Learn Syst*. 2018;29(5):1774-1785. doi:10.1109/TNNLS.2017.2673241
31. Tabassum M, Suman AA, Suero Molina E, Pan E, Di Ieva A, Liu S. Radiomics and machine learning in brain tumors and their habitat: a systematic review. *Cancers (Basel)*. 2023;15(15):3845. doi:10.3390/cancers15153845
32. Lohmann P, Bousabarah K, Hoevels M, Treuer H. Radiomics in radiation oncology—basics, methods, and limitations. *Strahlenther Onkol*. 2020;196(10):848-855. doi:10.1007/s00066-020-01663-3

Disclosures

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author Contributions

Conception and design: Baskaya, Sanlier, Yuce, Sulaimanov, Ismayilova, Huryol, Bozkurt, O Ozturk, B Demir, Serikkanov, Bayramoglu, H Demir, Karagoz, Karaaslan, Turk, Erginoglu. Acquisition of data: Sanlier, Sulaimanov, Ismayilova, Huryol, Bozkurt, O Ozturk, F Ozturk, Altun, Altunkaya, B Demir, Serikkanov, Bayramoglu, Karagoz, Karaaslan, Aydin, Gurbuz, Turk. Analysis and interpretation of data: Sanlier, Yuce, Sulaimanov, Ismayilova, Huryol, Bozkurt, O Ozturk, F Ozturk, Altun, Altunkaya, B Demir, Serikkanov, H Demir, Moniri, Karagoz, Karaaslan, Gurbuz, Turk, Erginoglu. Drafting the article: Sanlier, Yuce, Sulaimanov, Huryol, O Ozturk, Ammanuel, B Demir, Serikkanov, Bayramoglu, H Demir, Moniri, Karagoz, Karaaslan, Turk. Critically revising the article: Baskaya, Yuce, Sulaimanov, Ismayilova, Bozkurt, Altun, Altunkaya, Ammanuel, Serikkanov, Karagoz, Gurbuz, Keles, Erginoglu. Reviewed submitted version of manuscript: Baskaya, Sanlier, Yuce, Sulaimanov, Bozkurt, O Ozturk, Ammanuel, Serikkanov, H Demir, Karagoz, Karaaslan, Turk, Keles. Approved the final version of the manuscript on behalf of all authors: Baskaya. Statistical analysis: Sanlier, Yuce, O Ozturk, Serikkanov, Moniri, Karagoz, Karaaslan, Turk. Administrative/technical/material support: Altun, Altunkaya, Bayramoglu, Karagoz, Gurbuz, Keles. Study supervision: Baskaya, Karagoz, Keles.

Supplemental Information

Online-Only Content

Supplemental material is available online.

Supplementary Figure and Tables. https://thejns.org/doi/suppl/10.3171/2026.6.FOCUS26266.

Correspondence

Mustafa K. Baskaya: School of Medicine and Public Health, University of Wisconsin, Madison, WI. baskaya@neurosurgery.wisc.edu.