

Radiomic analysis of magnetic resonance imaging predicts brain metastases velocity and clinical outcome after upfront radiosurgery

Che-Yu Hsu, Furen Xiao, Kao-Lang Liu, Ting-Li Chen, Yueh-Chou Lee, and Weichung Wang

Division of Radiation Oncology, Department of Oncology, National Taiwan University Hospital, Taipei, Taiwan (C-Y.H.); National Taiwan University Cancer Center, Taipei, Taiwan (C-Y.H.); Graduate Program of Data Science, National Taiwan University and Academia Sinica, Taipei, Taiwan (C-Y.H., T-L.C., W.W.); Department of Neurosurgery, National Taiwan University Hospital, Taipei, Taiwan (F.X.); Department of Medical Imaging, National Taiwan University Cancer Center, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan (K-L.L.); Department of Mathematics, National Taiwan University, Taipei, Taiwan (Y-C.L.); Institute of Applied Mathematical Sciences, National Taiwan University, Taipei, Taiwan (W.W.); Institute of Statistical Science, Academia Sinica, Taipei, Taiwan (T-L.C.)

Corresponding Authors: Weichung Wang, PhD, Institute of Applied Mathematical Sciences, National Taiwan University, No.1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan (wwang@ntu.edu.tw); Che-Yu Hsu, MD, Division of Radiation Oncology, Department of Oncology, National Taiwan University Hospital, No. 7, Chung-Shan South Rd, Taipei 100, Taiwan (codiwa@gmail.com).

Abstract

Background. Brain metastasis velocity (BMV) predicts outcomes after initial distant brain failure (DBF) following upfront stereotactic radiosurgery (SRS). We developed an integrated model of clinical predictors and pre-SRS MRI-derived radiomic scores (R-scores) to identify high-BMV (BMV-H) patients upon initial identification of brain metastases (BMs).

Methods. In total, 256 patients with BMs treated with upfront SRS alone were retrospectively included. R-scores were built from 1246 radiomic features in 2 target volumes by using the Extreme Gradient Boosting algorithm to predict BMV-H groups, as defined by BMV at least 4 or leptomeningeal disease at first DBF. Two R-scores and 3 clinical predictors were integrated into a predictive clinico-radiomic (CR) model.

Results. The related R-scores showed significant differences between BMV-H and low BMV (BMV-L), as defined by BMV less than 4 or no DBF ($P < .001$). Regression analysis identified BMs number, perilesional edema, and extra-cranial progression as significant predictors. The CR model using these 5 predictors achieved a bootstrapping corrected *C*-index of 0.842 and 0.832 in the discovery and test sets, respectively. Overall survival (OS) after first DBF was significantly different between the CR-predicted BMV-L and BMV-H groups (median OS: 26.7 vs 13.0 months, $P = .016$). Among patients with a diagnosis-specific graded prognostic assessment of 1.5–2 or 2.5–4, the median OS after initial SRS was 33.8 and 67.8 months for CR-predicted BMV-L, compared to 13.5 and 31.0 months for CR-predicted BMV-H ($P < .001$ and $< .001$), respectively.

Conclusion. Our CR model provides a novel approach showing good performance to predict BMV and clinical outcomes.

Key Points

- A model of clinical predictors and MRI-derived radiomic scores was created.
- The model offers a good prediction of brain metastasis velocity and patient survival.
- The model may inform brain metastases monitoring and radiation treatment strategies.

Importance of the Study

Brain metastasis velocity (BMV) is a reliable measurement of the kinetics of distant brain failure (DBF) and predicts outcomes after initial DBF following upfront stereotactic radiosurgery (SRS). Based on pre-SRS information and magnetic resonance imaging, we developed a machine learning-based clinico-radiomic (CR) model to predict BMV and, in turn, clinical outcomes. Particularly, the CR-predicted BMV risk category not only

offers good BMV prediction and survival after the first DBF, but also represents a good survival prediction factor among patients with a diagnosis-specific graded prognostic assessment of 1.5–2 or 2.5–4 upon initial identification of brain metastases. Our proposed CR model may help customize optimal radiation treatment strategies and DBF monitoring frequencies for brain metastases patients in this era of precision medicine.

Approximately 20–40% of patients diagnosed with cancer develop brain metastases (BMs), whose incidence has been rising due to more efficacious systemic therapies and improved survival.^{1,2} In the past decades, whole brain radiotherapy (WBRT) was the treatment of choice for multiple BMs. However, recently, evidence from randomized clinical trials has supported stereotactic radiosurgery (SRS) alone as the preferred treatment modality in patients with 1–4 BMs.^{3,4} Thus, SRS provides better cognitive and quality of life outcomes with no detrimental effect on survival, compared with WBRT.^{5,6} However, designing an SRS plan is more time- and effort-consuming than WBRT. Moreover, in patients with poor life expectancy or with rapidly developing new distant BMs, the use of SRS alone, to avoid WBRT neurocognitive sequelae, may be less beneficial.^{4,7,8} Selecting those BM patients most likely to achieve better intracranial control and survival outcome is essential to maximize the effectiveness of SRS alone.

However, distant brain failure (DBF) is a common occurrence following SRS alone. Brain metastasis velocity (BMV) is a novel prognostic metric that measures the number of new BMs appearing between initial SRS and first DBF.⁹ BMV has emerged as a reliable predictor of overall survival (OS) after the first DBF following initial SRS. As validated in several retrospective analyses from multi-institutional cohorts, the median OS after first DBF is around 12, 8, and 4 months for patient groups showing a BMV less than 4, between 4–13, and greater than 13 new lesions per year, respectively.^{9–11} Nevertheless, BMV can only be used following the first DBF, which narrows the target population for clinical application. Difficulties in obtaining BMV before SRS also hinder the integration of BMV with other prognostic indices such as the diagnosis-specific graded prognostic assessment (DS-GPA) and some nomograms^{12,13} to predict DBF and OS outcomes at the time of initial BMs identification.

Radiomics, a newly emerging field of image analysis, provides radiographic information by extracting high-throughput imaging phenotypes and selecting features via statistical data analysis or machine learning (ML) algorithms. The development of radiomic models has revealed promising routes for prognosis prediction in various cancer types. Recent works include the development of a radiomics-based nomogram to predict lymph node metastasis in colorectal cancer,¹⁴ the utilization of a radiomic

signature (RS) to evaluate the pathological complete response to neoadjuvant chemoradiation in locally advanced rectal cancer,¹⁵ and the application of large-scale radiomic features to stratify anti-angiogenic treatment response in recurrent glioblastoma.¹⁶ As for BMs, MRI radiomic approaches are used to distinguish true progression from radionecrosis after SRS,^{17,18} to find associations between radiographic features and prognosis after treatment with immune checkpoint inhibitors,¹⁹ and to predict local tumor control following SRS.²⁰ However, to the best of our knowledge, there is no study optimally assessing the prognostic potential of MRI-based radiomics in capturing the kinetics of developing DBF after initial SRS. This study investigated whether the integration of large-scale MRI radiomic features and clinical profiles could predict high BMV (BMV-H) in patients treated with SRS alone. Additionally, we explored the impact of our clinico-radiomic (CR) model on survival outcomes after initial SRS and first DBF.

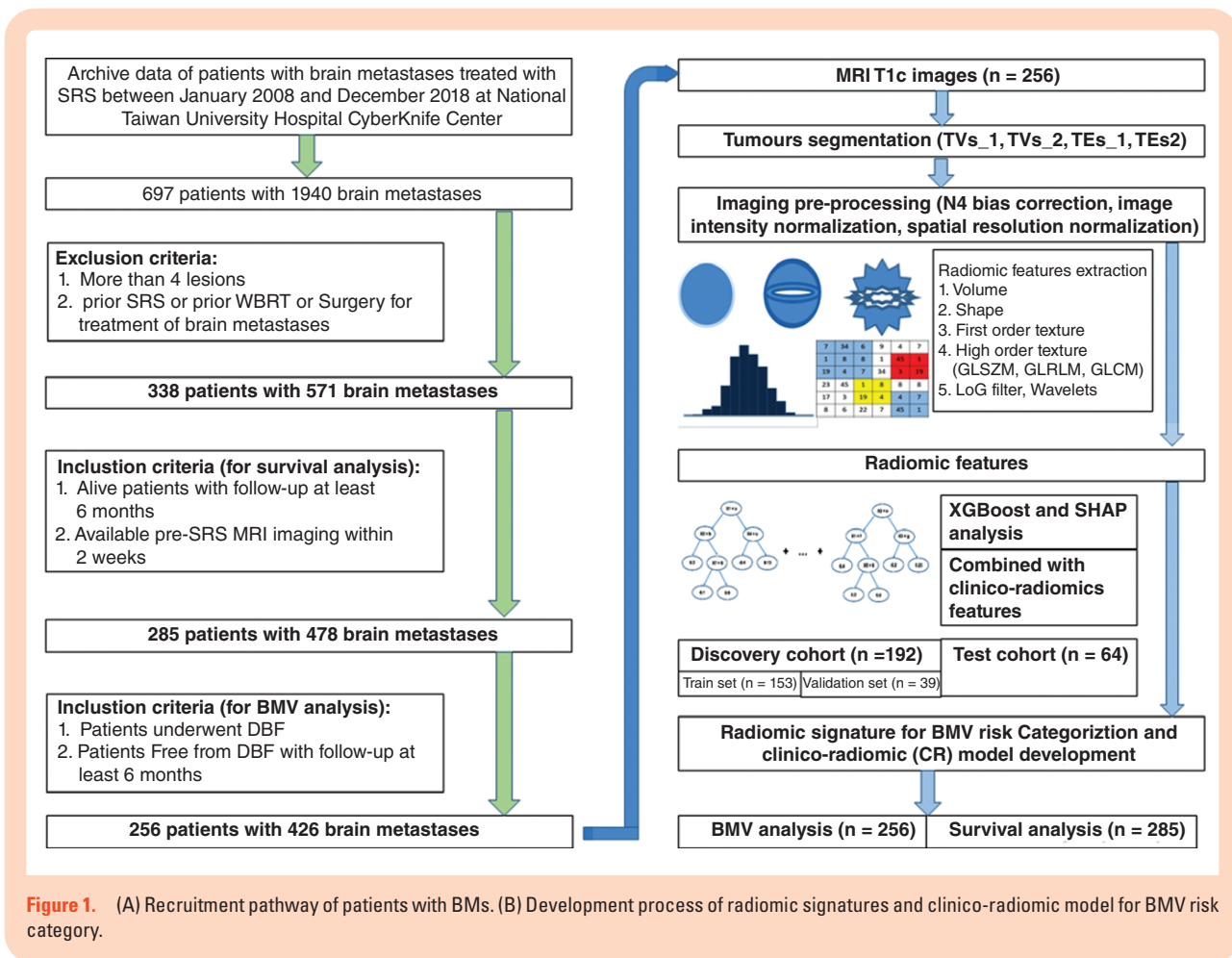
Materials and Methods

Patients and Database Acquisition

This retrospective study was approved by the National Taiwan University Hospital (NTUH) Institutional Review Board. The dataset was derived from the NTUH SRS database and consisted of all patients with BMs treated between January 2008 and January 2018 using a CyberKnife G4 image-guided robotic SRS system. The exclusion criteria were 5 or more BMs; prior SRS, WBRT, or surgery; and no available pre-SRS MRI images. The detailed recruitment pathway is presented in [Figure 1A](#).

We reviewed electronic medical records to determine the clinical outcome and infer putative clinical factors for DBF, including patient age; sex; histology of primary malignancy; number of initial BMs; DS-GPA; lowest SRS margin dose; maximum tumor volume (TV), defined as the volume of the maximum sized tumor among all metastases; extracranial disease burden; and extracranial progression status (EP).

Patients underwent MRI and clinical examination follow-ups approximately 1 month following SRS and every 3–4 months thereafter. DBF was identified as the apparition



of new lesions on follow-up imaging outside of the 95% prescribed isodose line. The number of lesions and timing of the first DBF were estimated for subsequent calculation of BMV. Leptomeningeal disease (LMD) was diagnosed either through cerebrospinal fluid cytology or neuroimaging. Regarding neuroimaging, LMD was defined as the presence of multifocal enhancing subarachnoid nodules on T1-weighted contrast-enhanced sequences (T1c) brain MRI or contrast-enhanced computed tomography (CT). Putative prognostic radiographic features, including perilesional edema (PE) and tumor peripheral enhancement, were recorded according to radiology reports and reconfirmed by radiologist K-L.L., with 20 years of radiology experience, according to edematous change on FLAIR and enhanced tumor signal on T1c imaging sequence.

Risk Group Categorization

We sorted patients into 2 groups according to their estimated BMV. The low-BMV (BMV-L) group gathered patients who underwent no DBF events during a follow-up period of at least 6 months or had a BMV of 1–3 BMs per year. BMV-H patients had an estimated BMV of 4 BMs per year or more or were diagnosed as having LMD at first DBF.

Image Preprocessing and Quantitative Image Analysis Workflow for RS

For each clinical case, pre-SRS MRI T1c was obtained after gadolinium-DTPA injection by using a 1.5-T unit MRI system (GE, Signa Excite HDxt 1.5T). [Supplementary Table S1](#) displays the imaging parameters used for the MRI sequences.

TV segmentation₁ (TVs₁) was obtained by retrieving gross TV (GTV) from Digital Imaging and Communications in Medicine-RT structure sets of each patient. TVs₂ was delineated by an independent radiation oncologist (C-Y.H., with 10 years of brain tumor radiotherapy experience) using a 3D Slicer software (3D Slicer, version 4.7.0-1) to assess inter-observer variations. To reveal the radiographic features at tumor edges, we also created tumor_{edge} segmentation₁ (TEs₁) and TEs₂ through subtraction tumor boundary via an isotropic 2 mm contraction from isotropic 3 mm expansion around the TVs₁ and TVs₂ surface, respectively ([Supplementary Figure S1](#)).

As for MRI image preprocessing, the N4 bias correction algorithm was used to remove unwanted low-frequency intensity non-uniformity by implementing the Insight Toolkit.²¹ Next, image intensity normalization was performed to transform arbitrary MR imaging signal intensity

values within 5 mm from the TV surface into standardized intensity ranges from 0 to 1024 Gy. Finally, linear interpolation was applied in the T1c images to make the voxel size isotropic ($1 \times 1 \times 1 \text{ mm}^3$; Figure 1B).

Radiomic features were extracted from T1c images encompassed within TVs and TEs by using the Pyradiomics package (version 2.0.0).²² A total of 1246 radiomic features were extracted from the preprocessed original and derived T1c images (5 derived images from Laplacian of Gaussian filter of 5 sigma levels and 8 derived images from Wavelet decompositions), including (1) first-order statistics features; (2) shape-based features; and (3) texture features derived from gray-level co-occurrence matrix (GLCM), gray-level run-length matrix, and gray-level size zone matrix (Supplementary Table S4; Supplementary Note S1). We calculated the intraclass correlation coefficient (ICC) for the radiomic features extracted from the segmentation of TVs₁/TVs₂ and TEs₁/TEs₂. Only radiomic features from TVs₁ and TEs₁ and with ICC values higher than 0.9 were selected for subsequent analysis to ensure good reliability.

Extreme Gradient Boosting (XGBoost) algorithm, based on a gradient boosting decision tree method,²³ was used to build mathematical models that predict BMV-H based on radiomic features. The XGBoost function was as follows:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (\text{Equation 1})$$

$$L(\phi) = \sum_i l_{\text{loss}}(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (\text{Equation 2})$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|\omega_i\|^2 \quad (\text{Equation 3})$$

In Equation 1, the predictive value y_i is generated from the tree ensemble model f_k (k -additive functions) by using the independent input variable x_i . The goal of the training process is to minimize objective function (Equation 2), where l_{loss} is a loss function and Ω used for regularization via penalizing the complexity of the model. The Ω comprised T , which is the number of leaves, and ω_i is the complexity score of the i th leaf.

Hyper-parameters accounting for fitness (including learning_rate, scale_pos_weight, and base_score) and regularity (consisting of min_child_weight, max_depth, gamma, subsample, reg_alpha, and colsample_bytree) were tuned via a grid search approach to find an optimal hyper-parameters set (Supplementary Table S2). For each grid search trial, we performed 10-fold cross-validation and searched for the optimal hyper-parameters giving the best cross-validation results within a total of 1000 grid search trials on the training set. Then, we fine-tuned the chosen hyper-parameters by the early stopping of training iterations according to the loss of objective function on the validation set to improve regularity. The output of the XGBoost model was finally converted into a probability score, namely the radiomic score (R-score), indicating the probability for the patient to belong to the BMV-H group.

Model Interpretation

We used the SHapley Additive exPlanations (SHAP) algorithm,²⁴ which is based on game theory and local explanations,²⁵ to interpret the output of our XGBoost model.²⁶ SHAP can estimate the contribution of each input feature to the model output based on their marginal contribution. The mean absolute SHAP values of radiomic features can represent their impact on the R-score. For these radiomic features with the top 5 SHAP values, we further used student t -test and analysis of variance to evaluate their correlation with putative clinical factors, where a P value of less than .01 was considered statistically significant.

Development and Performance Evaluation of the Predictive Model

Univariate analysis was performed to determine significant clinical predictors for BMV-H, and the significant predictors were further combined to develop the clinical (C) model or integrated with R-scores into the CR model via logistic regression by using IBM SPSS Statistics 20.

Harrell's C -index and the corrected C -index, derived from the bootstrapping validation with 1000 resamples, were calculated by using the "Hmisc" package in the R software in order to evaluate the discriminative ability of our clinical predictors, C model, R-scores, and CR model between BMV-H and BMV-L. We also performed decision curve analysis (DCA) to evaluate the net benefits of the R-score added to the C model. The net benefit of the CR model was estimated by using the decision curve with the difference between the true-positive and false-positive rates, weighted by the odds of the selected threshold probability of risk.²⁷

Follow-Up and Statistical Analysis of Clinical Outcomes

Median follow-up and time-to-event outcomes were assessed from the beginning of SRS to the most recent follow-up or event of interest. Time-to-event outcomes were summarized using the Kaplan–Meier and log-rank test. A P value of less than .05 was considered statistically significant.

Results

Patient Characteristics

Between January 2008 and February 2018, a total of 256 patients with 426 newly diagnosed BMs treated with SRS as a single modality were included. The entire cohort was randomly split into the discovery (training and validation) and test cohorts with the ratio of 3:1 to establish and validate the prediction of the R-score for BMV-H. The demographic characteristics of the discovery and test cohorts are listed in Supplementary Table S3. The median follow-up time of the entire cohort was 20.2 months (95% CI: 18.6–22.3). There were no significant differences between the

2 cohorts in the distribution of elderly patients ($P = .885$), gender ($P = .640$), primary malignancy ($P = .182$), BMs number ($P = .971$), DS-GPA ($P = .100$), lowest SRS margin dose ($P = .348$), total TV ($P = .744$), median follow-up time ($P = .559$), and first DBF status ($P = .814$).

Clinical Predictors

As given in Table 1, only the number of BMs, presence of PE, and status of EP were significantly different between BMV-H and BMV-L in the discovery cohort. Other clinical factors, including age, gender, primary malignancy, systemic disease burden, DS-GPA, tumor peripheral enhancement, lowest SRS margin dose, TV, and epidermal growth factor receptor (EGFR) mutant status, were distributed similarly across groups. Risk coefficients estimated by univariate analysis are summarized in Table 1.

Radiomic Analysis and Predictive CR Model

Radiomic features extracted from T1c images in the discovery cohort were selected and quantitatively integrated into 2 R-scores, namely Radi_Tumor based on TVs_1 and Radi_Edge based on TEs_1, using XGBoost. The selected XGBoost model hyper-parameters are listed in Supplementary Table S2. The corrected C-indexes of Radi_Tumor and Radi_Edge were 0.811 and 0.794 in the discovery cohort and 0.713 and 0.764 in the test cohort, respectively (Table 2).

We further combined 3 significant clinical predictors without and with 2 R-scores to develop the C model and CR model by using logistic regressions, as described:

$$Y = -2.50 + 0.53 \times \text{BM_numbers} + 0.79 \times \text{PE} + 1.08 \times \text{EP (C model)}$$

$$Y = -4.8 + 0.02 \times \text{BM_numbers} + 0.31 \times \text{PE} + 1.21 \times \text{EP} + 8.78 \times \text{Radi_Tumor} + 4.00 \times \text{Radi_Edge (CR model, Figure 2A)}$$

Compared with the C model only, the resulting CR model demonstrated better discriminative ability in estimating the probability of BMV-H, with a C-index of 0.858 (95% CI: 0.800–0.917) and corrected C-index of 0.842 in the discovery cohort, and a C-index of 0.833 (95% CI: 0.698–0.967) and corrected C-index of 0.832 in the test cohort (Table 2 and Figure 2B). The CR model's calibration curve demonstrated good agreement between the predicted and observed BMV groups in both the discovery and test cohorts (Figure 2C and D). The Hosmer–Lemeshow test gave a χ^2 of 10.719 ($P = .218$) and 6.901 ($P = .547$) for the discovery and the test cohorts, respectively, indicating that the CR model was appropriate for both datasets. Furthermore, we also tested the incremental net benefit of the CR model with respect to the C model for prediction of BMV-H using DCA, as described in Supplementary Figure S2. DCA showed that adding 2 R-scores resulted in a net benefit compared with the use of the C model only.

By using a cutoff value of predictors and models for assuring both a specificity greater than 75% and maximizing

the Youden index of the receiver operating characteristic curve analysis from the discovery cohort, we dichotomized patients into binary risk groups. As given in Table 2, we further evaluated the predictive power of these binary classifiers using metrics of sensitivity, specificity, positive predictive value, negative predictive value, and accuracy (Table 2). The CR model-predicted BMV-H group (CR-predicted BMV-H) and BMV-L group (CR-predicted BMV-L), using a cutoff value of 0.215, yielded the predictive accuracy of 0.781 and 0.797 in the discovery and test cohorts, respectively (Table 2). The results also outperformed the accuracy of the binary classifier based on the C model (0.736 and 0.703 in the discovery and test cohorts, respectively). As for BMs from various origins, the C-index of the CR model was around 0.791–0.886 (Supplementary Figure S3A; Supplementary Note S2A). For BMs of lung origin, administration of tyrosine kinase inhibitors or not did not change the prediction performance of the CR model (C-indexes: 0.837 and 0.859, respectively; Supplementary Figure S3B; Supplementary Note S2A).

Survival Predictions

As of December 2018, 285 patients completed the OS follow-up since the initial SRS, and 121 patients completed the OS follow-up since the first DBF. The overall DBF rate was 42.5% (121/285), and the overall death rate was 47.7% (136/285). The median OS after first DBF for the entire cohort was 18.5 months (95% CI: 12.0–24.9), more specifically 13.0 months (95% CI: 7.7–18.4) for BMV-H and 25.6 months (95% CI: 15.9–35.2) for BMV-L patients (log-rank test, $P = .043$; Figure 3A). Moreover, the OS after first DBF was 13.0 months (95% CI: 4.8–21.3) for CR-predicted BMV-H and 26.7 months (95% CI: 8.1–45.4) for CR-predicted BMV-L patients (log-rank test, $P = .016$; Figure 3B). As for patients receiving salvage systemic treatment only, both BMV-L and CR-predicted BMV-L patients would have longer OS after the first DBF, which is less prominent in the subgroup treated with added salvage WBRT or SRS (Supplementary Figure S4; Supplementary Note S2B). In the BMV-L subgroup and the CR-predicted BMV-L patients, salvage SRS was associated with better OS after DBF compared to that of WBRT ($P = .001$ and $P < .001$, respectively; Supplementary Figure S5; Supplementary Note S2B).

The median OS after initial SRS was 32.4 months (95% CI: 26.4–38.5); more specifically 16.9 months (95% CI: 12.5–21.3) for CR-predicted BMV-H and 52.7 months (95% CI: 30.3–75.0) for CR-predicted BMV-L (log-rank test, $P < .001$; Figure 3C). Among patients with a DS-GPA of 1.5–2 or 2.5–4, the median OS after initial SRS was 13.5 and 31.0 months for CR-predicted BMV-H patients, respectively. The OS was significantly worse than for CR-predicted BMV-L, with a survival of 30.8 and 67.8 months (log-rank test, $P < .001$, and $< .001$), respectively (Figure 3D and E). Nevertheless, for patients with a DS-GPA of 0–1, no survival difference was observed between the CR-predicted BMV-H and CR-predicted BMV-L groups (13.3 and 11.2 months, respectively, $P = .531$; Figure 3F).

Table 1. Distribution of Clinical Factors Between BMV-H and BMV-L in Discovery Set

	BMV-L	BMV-H	Odds Ratio	P
Age, mean±SD, years				
≤60	69 (47.9)	25 (52.1)	1	.617
>60	75 (52.1)	23 (47.9)	0.846 (0.440–1.628)	
Gender, n (%)				
Male	61 (42.4)	18 (37.5)	1	.554
Female	83 (57.6)	30 (62.5)	1.225 (0.626–2.397)	
Primary malignancy, n (%)				
Lung	111 (77.1)	38 (79.2)	1	
Breast	16 (11.1)	4 (8.3)	0.730 (0.230–2.320)	.594
GI	11 (7.6)	1 (2.1)	0.266 (0.33–2.126)	.212
Others	6 (4.2)	5 (10.4)	2.434 (0.703–8.434)	.161
Number of brain metastases, n (%)				
1	92 (63.9)	16 (33.3)	1	
2	33 (22.9)	20 (41.7)	3.485 (1.616–7.514)	.001
3–4	19 (13.2)	12 (25.0)	3.632 (1.481–8.903)	.005
Systemic disease burden, n (%)				
None	72 (50.0)	22 (45.8)	1	
Oligometastatic (1–5 extracranial metastases)	49 (34.0)	16 (33.3)	1.069 (0.510–2.238)	.860
Widespread (>5 extracranial metastases)	23 (16.0)	10 (20.8)	1.423 (0.589–3.440)	.434
Extracranial progression				
No	122 (84.7)	30 (62.5)	1	
Yes	22 (15.3)	18 (37.5)	3.327 (1.588–6.974)	.001
DS-GPA (%)				
0–1	18 (12.5)	10 (20.8)	1	
1.5–2	43 (29.9)	20 (41.7)	0.837 (0.328–2.138)	.710
2.5–3	55 (38.2)	13 (27.1)	0.425 (0.159–1.135)	.088
3.5–4	28 (19.4)	5 (10.4)	0.321 (0.094–1.095)	.070
Central necrosis				
No	75 (52.1)	24 (50)	1	.803
Yes	69 (47.9)	24 (50)	1.087 (0.565–2.089)	
Perilesional edema				
No	119 (61.0)	22 (36.1)	1	.011
Yes	76 (39.0)	93 (63.9)	2.401 (1.226–4.702)	
Peripheral enhancement				
No	68 (47.2)	16 (33.3)	1	.095
Yes	76 (52.8)	32 (66.7)	1.789 (0.903–3.545)	
EGFR mutant status (lung primary)				
Mutant	59 (41)	18 (37.5)	1	
Wild type	29 (38.9)	9 (18.7)	1.017 (0.407–2.541)	.971
Not available or non-lung primary	56 (20.1)	21 (43.8)	1.229 (0.594–2.546)	.579
Maximum tumor volume, mean±SD, mm ³				
≤2500	83 (57.6)	24 (50.0)	1	.357
>2500	61 (42.4)	24 (50.0)	1.361 (0.707–2.620)	
Lowest SRS margin dose, mean±SD, cGy				
	21.00 (20.00–22.00)	20.00 (20.00–22.00)		.798
Redi_Tumor, mean±SD				
	0.239±0.091	0.364±0.119		<.001
Redi_Edge, mean±SD				
	0.154±0.150	0.333±0.192		<.001

GI, gastrointestinal system; DS-GPA, diagnosis-specific grade prognostic assessment; BMV, brain metastases velocity; SD, standard deviation.

Table 2. Discrimination Performance Evaluation of Models on Discovery and Test Cohorts

	Discovery Cohort					Test Cohort				
	BM's	PE	EP	C Model	CR Model	BM's	PE	EP	C Model	CR Model
C-index	0.653	0.608	0.611	0.736	0.858	0.517	0.679	0.604	0.721	0.833
95% CI	0.562–0.741	0.516–0.700	0.514–0.708	0.658–0.815	0.800–0.917	0.335–0.698	0.515–0.844	0.421–0.787	0.581–0.861	0.698–0.967
Corrected C-index	0.651	0.604	0.610	0.718	0.842	0.453	0.676	0.581	0.7192	0.832
Cutoff value	1.5	0.5	0.5	0.275	0.215	1.5	0.5	0.5	0.275	0.215
Sensitivity	66.7%	62.5%	37.5%	58.3%	85.4%	46.2%	69.2%	38.5%	69.2%	92.3%
Specificity	63.9%	59.0%	84.7%	76.4%	75.7%	54.9%	66.7%	82.4%	70.6%	76.5%
PPV	38.1%	33.7%	45.0%	45.2%	53.9%	20.7%	34.6%	35.7%	37.5%	50.0%
NPV	85.2%	82.5%	80.3%	84.6%	94.0%	80.0%	89.5%	84.0%	90.0%	97.5%
Accuracy	0.646	0.599	0.729	0.736	0.781	0.531	0.672	0.734	0.703	0.797

C-index, concordance index; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; BM's, numbers of brain metastases; PE, presence of perilesional edema; EP, presence of extracranial progression.

Feature Analysis

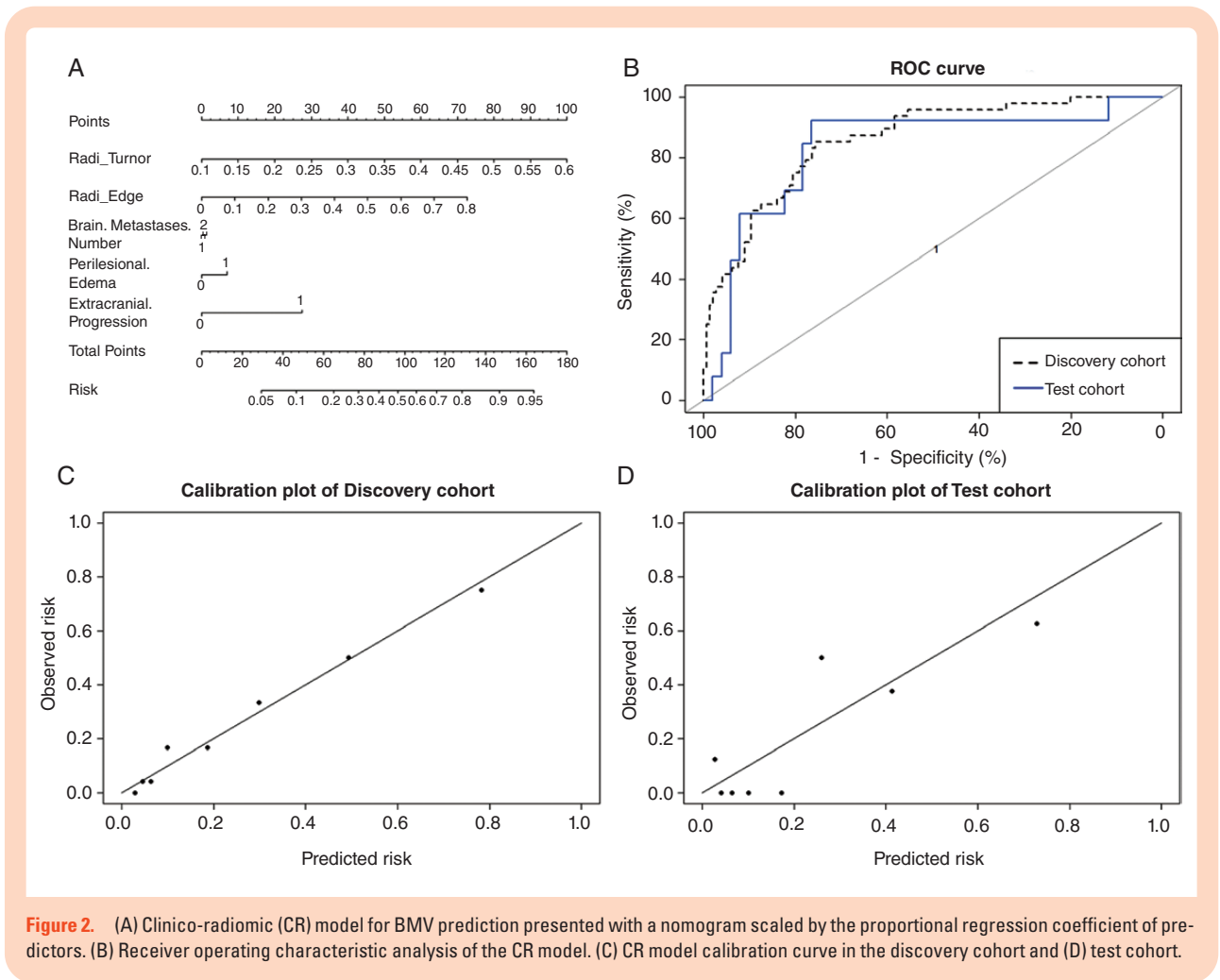
There were 16 and 9 radiomic features with non-zero SHAP values based on the prediction of BMV-H using Radi_Tumor and Radi_Edge, respectively. Out of the top 5 ranked features of Radi_Tumor, log-sigma-2-5-mm-3D_firstorder_Maximum had the highest SHAP value of 0.153, followed by 2 shape-based features (shape_MajorAxisLength and shape_Elongation) and 2 high-order textural features (wavelet-LHH_glcmln and log-sigma-3-0-mm-3D_glcmln_Correlation; [Figure 4A](#)). As for Radi_Edge, the top 4 ranked features were all high-order textural features, namely, log-sigma-1-5-mm-3D_glcmln, log-sigma-2-0-mm-3D_glcmln_ClusterShade, log-sigma-3-0-mm-3D_glcmln_MCC, and wavelet-HLL_glcmln_JointEntropy, followed by the shape-based feature shape_MajorAxisLength ([Figure 4B](#)).

Further correlation analysis between the top 5 features of Radi_Tumor and Radi_Edge with clinical predictors indicated that shape-based features, including MajorAxisLength and Elongation, were significantly associated with the BMs number ($P < .001$). The distribution of the Radi_Tumor first-order statistics features and Radi_Tumor and Radi_Edge high-order textural features varied across groups with different BMs numbers ([Figure 4B and C](#)). On the other hand, 1 Radi_Tumor first-order statistics feature (log-sigma-2-5-mm-3D_firstorder_Maximum) and 2 Radi_Tumor (wavelet-LHH_glcmln and log-sigma-3-0-mm-3D_glcmln_Correlation) and 3 Radi_Edge high-order textural features (log-sigma-2-0-mm-3D_glcmln_ClusterShade, log-sigma-3-0-mm-3D_glcmln_MCC, and wavelet-HLL_glcmln_JointEntropy) also correlated with PE. Additionally, for the subgroup of tumors with a lung origin, Radi_Tumor wavelet-LHH_glcmln was related to the EGFR mutant status ($P = .024$).

Discussion

This study investigated the prognostic potential of an ML-assisted radiomic model derived from pre-SRS MR imaging in predicting DBF kinetics and outcomes in a cohort of BMs patients treated with upfront SRS. MRI-based radiomic features, converted into the quantitative R-scores Radi_Tumor and Radi_Edge, emerged as independent BMV-H predictors. Our risk model integrating clinical factors and R-scores displayed good discriminative power for BMV-H and BMV-L, with a corrected C-index of 0.842 and 0.832 in the discovery and test cohorts, respectively. Using the cutoff value of 0.215 allowed us to identify 85.4% of BMV-H cases from the discovery cohort with a specificity of 75.7% and 92.3% of BMV-H cases from the test cohort with a specificity of 76.5%.

Additionally, the CR model predicted that the 2 BMV subgroups independently correlated with survival after both initial SRS and first DBF. Especially for patients with an initial DS-GPA of 1.5–2 or 2.5–4, our dichotomy category could help select CR-predicted BMV-L patients, who have a better median OS of 30.8–67.8 months after initial SRS. Our findings may help assist future clinical decisions for newly diagnosed BMs patients.



Radiomics has been emerging as an important novel imaging methodology in oncology and the source of quantitative biomarkers for cancer treatment. However, technical challenges including feature reliability and overfitting problems with high-dimensional data need to be overcome. All MRI images in our study were obtained from unique scan equipment to increase reproducibility by minimizing imaging variations related to scanner characteristics. Furthermore, in addition to using the N4 bias correction algorithm to remove unwanted low-frequency intensity non-uniformity, we performed regional voxel intensity normalization within 5 mm from the lesion surface to enhance the relationship between voxel intensities instead of that between raw numerical intensity values, which is unstable due to the lack of fixed tissue-specific numeric values from MRI scans. Next, we adopted XGBoost to train the prediction model as it allows the implementation of gradient-boosted decision trees and applies a higher regularized approach to control overfitting. Due to its scalability and speed,²³ XGBoost has also emerged as a widely used ML method for high-dimensional biomedical problems.^{28,29} Another important step for improving regularity is to adopt an ML ensemble method approach. In addition to radiomic features extracted from entire tumors such as Radi_Tumor, which reflect intratumoral heterogeneities, we used a set

of radiomic features obtained from the tumor edge zone to build up Radi_Edge. The corrected C-indexes of Radi_Tumor and Radi_Edge for the test cohort were initially 0.709 and 0.770, respectively. After making an ensemble of the 2 R-scores and clinical predictors, the CR model provided a corrected C-index value of 0.832, indicating better regularity.

One weakness of ML-derived predictive models is that they function as “a black box,” offering little information on how exactly the model makes predictions. For model interpretation, we took advantage of the SHAP algorithm, which provided insightful measures on the importance of features in an ML model. Thus, the SHAP value of radiomic features determined their impact on the XGBoost model.

For Radi_Tumor and Radi_Edge, shape-based features accounted for two-fifths and one-fifth of the top 5 important radiomic features, respectively. Upon correlation analysis, both MajorAxisLength and shape_Elongation were associated with multiple BMs status. Compared with single BM cases, the tumor-enclosing ellipsoid of multiple BMs will likely have a greater largest axis length. Intracranial BMs number is widely regarded as an important prognostic factor for either DBF or survival,^{12,30,31} with a similar role found for DBF in our study. MajorAxisLength not only quantifies the number of BMs but also their spatial

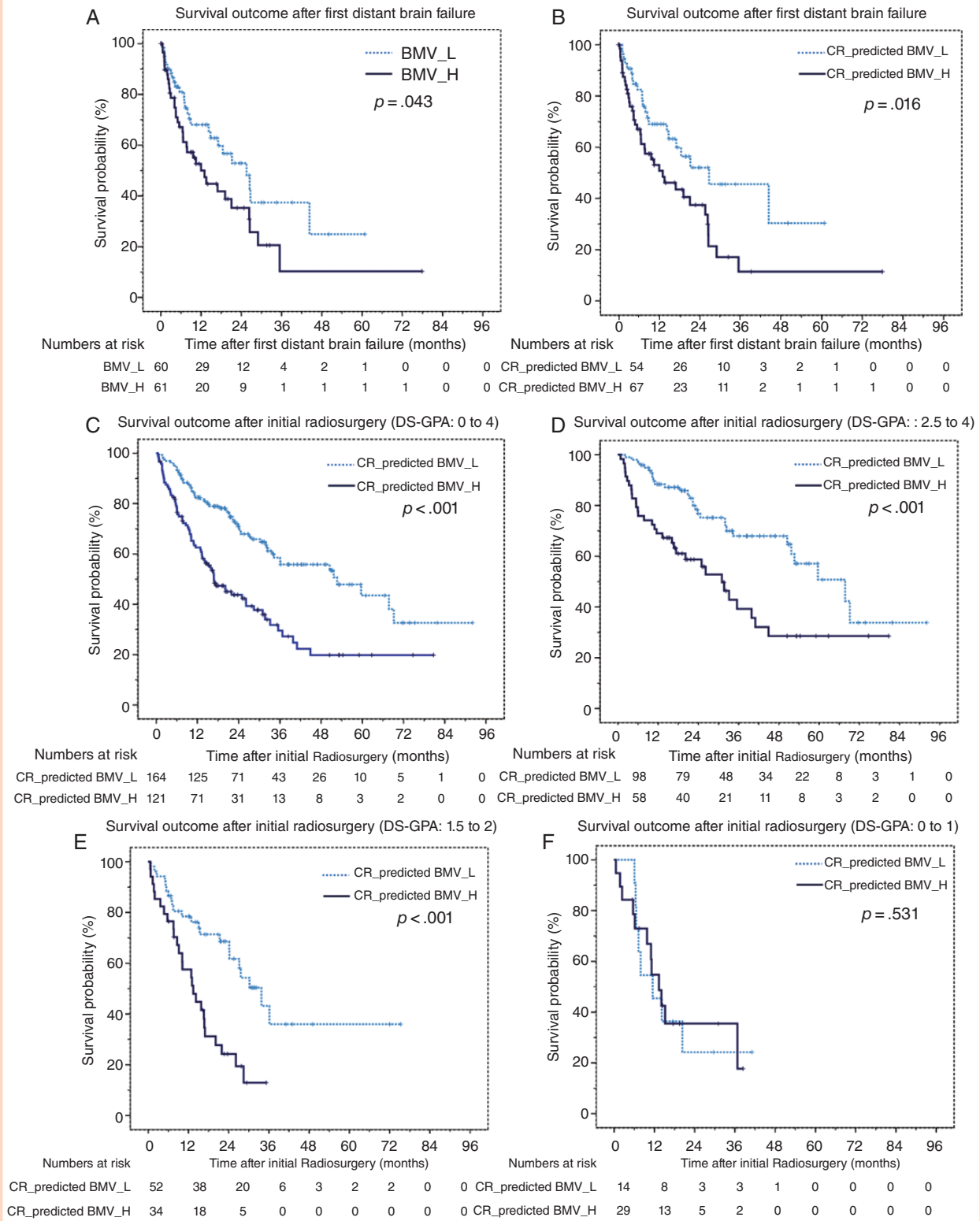


Figure 3. (A and B) Overall survival (OS) curves since the first DBF scaled by (A) BMV status and by (B) CR-predicted BMV status with Kaplan–Meier analysis. (C, D, E, and F) OS curves since initial SRS scaled by CR-predicted BMV status with Kaplan–Meier analysis for (C) all patients and (D) patients with a diagnosis-specific graded prognostic assessment (DS-GPA) of 1.5–2, (E) 2.5–4, or (F) 0–1.

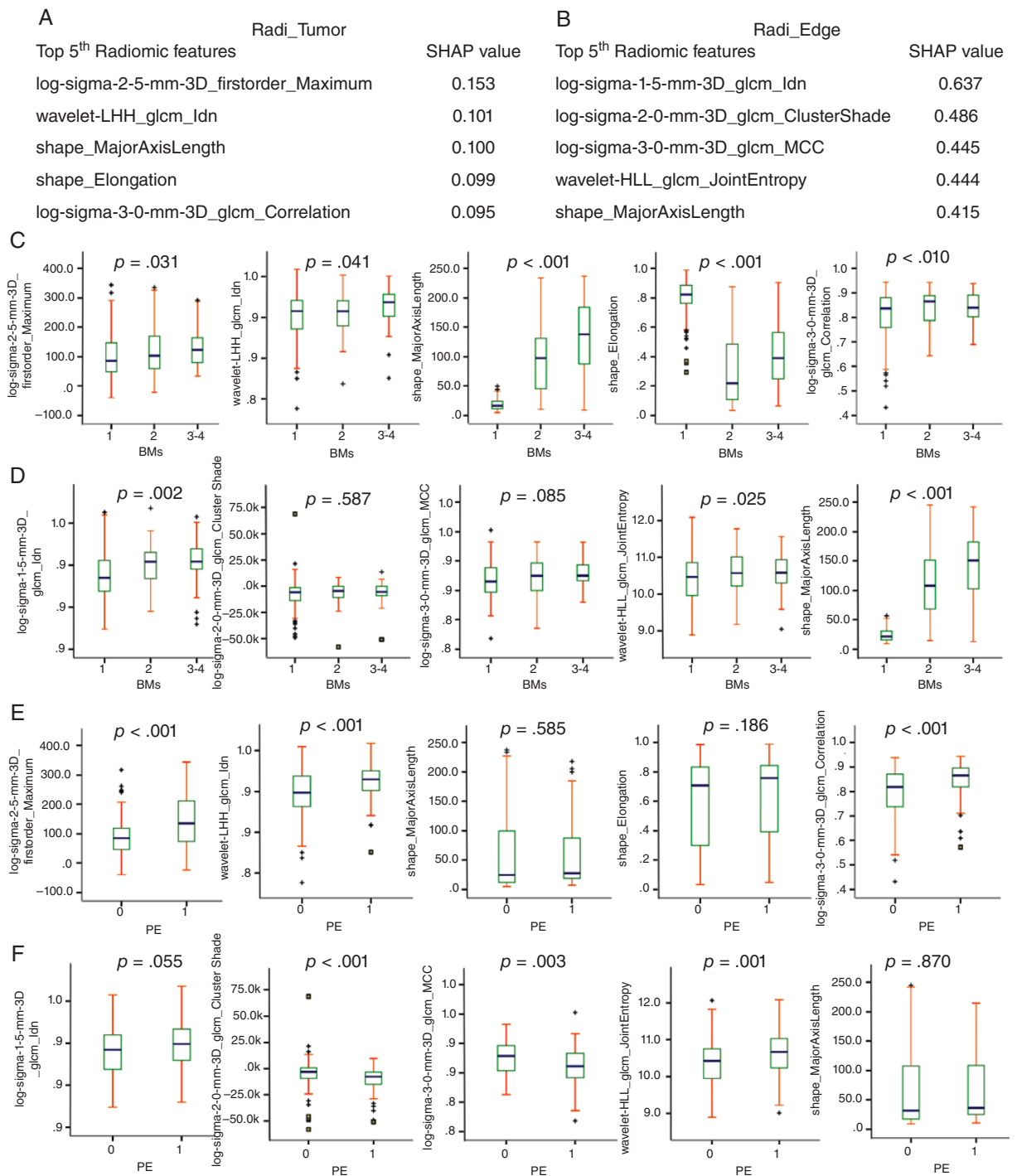


Figure 4. (A and B) Top 5 important radiomic features of (A) Radi_Tumor and (B) Radi_Edge according to SHAP value. (C and D) Box plot illustrations of the correlation between metastases number and top 5 important radiomic features of (C) Radi_Tumor or (D) Radi_Edge. (E and F) Box plot illustrations of the correlation between perilesional edema and top 5 important radiomic features of (E) Radi_Tumor and (F) Radi_Edge.

distribution; thus, the more widely BMs spread, the larger the MajorAxisLength value. In addition, shape_Elongation is regarded as an invasive phenotype giving rise to variations at the border between the tumor and brain tissue and associated with local control of BMs treated with SRS.³² Altogether, the properties of shape features and their

correlations with BMs number may provide one possible explanation of why Radi_Tumor and Radi_Edge perform well in predicting BMV-H.

As for textural features, first-order statistics features represented one-fifth of the top 5 important radiomic features of Radi_Tumor, and high-order textural features comprised

two-fifth and four-fifth of the top 5 important radiomic features of Radi_Tumor and Radi_Edge, respectively. These important textural features, especially 4 leading radiomic features derived from the tumor edge region, correlated with PE. In the past, glioblastoma PE was regarded as a poor prognosis factor,³³ due to its association with cancer cell infiltration as a result of the destruction of the blood–brain barrier.³⁴ Similar prognostic observations for PE were also found in BMs treated with SRS.^{35,36} Tini et al. demonstrated that PE greater than 10 mm correlates with higher out-of-field recurrence ($P = .029$), and Nardone et al. found that PE extent is associated with poor OS (hazard ratio [HR]: 1.044, $P = .009$). In our study, PE also correlated with BMV-H risk (HR: 2.401, $P = .011$). The association between important textural features and PE may also explain how the model performs predictions. Additionally, Radi_Tumor wavelet-LHH_glcml_dln was associated with EGFR mutation status in the lung cancer subgroup. In previous studies, intracranial response and survival after BMs radiotherapy were significantly favorable in EGFR-mutant cohorts.^{37,38} Radiomics features have also been found to be associated with EGFR mutation status based on CT images of lung cancer³⁹ or MRI images of brain tumors.⁴⁰ Although, in our study, EGFR mutation status was not associated with BMV-H, the connection between radiographic phenotypes and molecular genotypes is worth further investigation.

Our study is not without limitations, however. First, it is a single-institution study. One benefit is that all patients underwent a similar protocol for SRS and the same MRI scan sequence. Nevertheless, validation of our results by other medical centers will face the reproducibility challenge due to variations in treatment modality and imaging construction. We suggest combining image preprocessing and analysis workflow with the standardized radiomic features provided by the Imaging Biomarker Standardization Initiative for future prospective multicenter validation.⁴¹ In addition, all lesion segmentations in our study were retrieved from GTV contoured by expert radiation oncologists or a neurosurgeon and were independently reevaluated manually for inter-reader variability. Even when using similar delineation principles from single training systems for tumor segmentations, the manual tumor boundary delineation for radiomics analysis in other validation datasets may still challenge reproducibility. Deep neural network-based auto- or semi-auto segmentation for BMs will be required in future multi-institutional evaluation. Furthermore, the few cases other than lung-origin BMs, and the lack of patients with 5 or more BMs, limit generalization of the CR model to other populations with BMs, which emphasizes the importance for external validation with a larger study.

Conclusions

We systematically investigated the BMV of patients receiving SRS by analyzing clinical data and MRI-derived radiomic features. The integrated CR model of clinical predictors and R-scores performed well to predict BMV-H based on pre-SRS information. CR-predicted BMV risk was significantly associated with survival after initial SRS and first DBF and may help

select optimal radiation treatments for BMs patients. The use of an explainable algorithm for the ML-based model allows investigating the prognostic relevance of features. Further research is warranted on the crosstalk between genotypes, radiographic phenotypes, and other clinical predictors.

Supplementary Data

Supplementary data are available at *Neuro-Oncology Advances* online.

Keywords

brain metastases velocity | distant brain failure | machine learning | neuro-oncology | radiomics

Funding

This study was supported by the grants of Ministry of Science and Technology (107-2634-F-002-016), Ministry of Science and Technology (108-2634-F-002-013), Ministry of Science and Technology (109-2634-F-002-028), and Ministry of Science and Technology (109-2314-B-002-105-MY2). The funding source had no role in study design, data collection, analysis or interpretation, report writing, or the decision to submit this paper for publication.

Conflict of interest statement. All authors declare no conflict of interest.

Authorship statement. Experimental design: C-Y.H., F.X., K-L.L., T-L.C., and W.W. Implementation: C-Y.H., Y-C.L., T-L.C., and W.W. Analysis: Y-C.L., T-L.C., and W.W. Interpretation: C-Y.H., F.X., K-L.L., T-L.C., and W.W. Writing and revision of the manuscript: All authors.

References

1. Scoccianti S, Ricardi U. Treatment of brain metastases: review of phase III randomized controlled trials. *Radiother Oncol.* 2012;102(2):168–179.
2. Johnson AG, Ruiz J, Hughes R, et al. Impact of systemic targeted agents on the clinical outcomes of patients with brain metastases. *Oncotarget.* 2015;6(22):18945–18955.
3. Brown PD, Jaeckle K, Ballman KV, et al. Effect of radiosurgery alone vs radiosurgery with whole brain radiation therapy on cognitive function in patients with 1 to 3 brain metastases: a randomized clinical trial. *JAMA.* 2016;316(4):401–409.

4. Aoyama H, Shirato H, Tago M, et al. Stereotactic radiosurgery plus whole-brain radiation therapy vs stereotactic radiosurgery alone for treatment of brain metastases: a randomized controlled trial. *JAMA*. 2006;295(21):2483–2491.
5. Chang EL, Wefel JS, Hess KR, et al. Neurocognition in patients with brain metastases treated with radiosurgery or radiosurgery plus whole-brain irradiation: a randomised controlled trial. *Lancet Oncol*. 2009;10(11):1037–1044.
6. Greene-Schloesser D, Robbins ME, Peiffer AM, Shaw EG, Wheeler KT, Chan MD. Radiation-induced brain injury: a review. *Front Oncol*. 2012;2:73.
7. Lester SC, Taksler GB, Kuremsky JG, et al. Clinical and economic outcomes of patients with brain metastases based on symptoms: an argument for routine brain screening of those treated with upfront radiosurgery. *Cancer*. 2014;120(3):433–441.
8. Aoyama H, Tago M, Shirato H; Japanese Radiation Oncology Study Group 99-1 (JROSG 99-1) Investigators. Stereotactic radiosurgery with or without whole-brain radiotherapy for brain metastases: secondary analysis of the JROSG 99-1 randomized clinical trial. *JAMA Oncol*. 2015;1(4):457–464.
9. Farris M, McTyre ER, Cramer CK, et al. Brain metastasis velocity: a novel prognostic metric predictive of overall survival and freedom from whole-brain radiation therapy after distant brain failure following upfront radiosurgery alone. *Int J Radiat Oncol Biol Phys*. 2017;98(1):131–141.
10. McTyre ER, Soike MH, Farris M, et al. Multi-institutional validation of brain metastasis velocity, a recently defined predictor of outcomes following stereotactic radiosurgery. *Radiother Oncol*. 2020;142:168–174.
11. Yamamoto M, Aiyama H, Koiso T, et al. Validity of a recently proposed prognostic grading index, brain metastasis velocity, for patients with brain metastasis undergoing multiple radiosurgical procedures. *Int J Radiat Oncol Biol Phys*. 2019;103(3):631–637.
12. Rodrigues G, Warner A, Zindler J, Slotman B, Lagerwaard F. A clinical nomogram and recursive partitioning analysis to determine the risk of regional failure after radiosurgery alone for brain metastases. *Radiother Oncol*. 2014;111(1):52–58.
13. Ayala-Peacock DN, Peiffer AM, Lucas JT, et al. A nomogram for predicting distant brain failure in patients treated with gamma knife stereotactic radiosurgery without whole brain radiotherapy. *Neuro Oncol*. 2014;16(9):1283–1288.
14. Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. 2016;34(18):2157–2164.
15. Liu Z, Zhang XY, Shi YJ, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res*. 2017;23(23):7253–7262.
16. Kickingereder P, Götz M, Muschelli J, et al. Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. *Clin Cancer Res*. 2016;22(23):5765–5771.
17. Peng L, Parekh V, Huang P, et al. Distinguishing true progression from radionecrosis after stereotactic radiation therapy for brain metastases with machine learning and radiomics. *Int J Radiat Oncol Biol Phys*. 2018;102(4):1236–1243.
18. Zhang Z, Yang J, Ho A, et al. A predictive model for distinguishing radiation necrosis from tumour progression after gamma knife radiosurgery based on radiomic features from MR images. *Eur Radiol*. 2018;28(6):2255–2263.
19. Bhatia A, Birger M, Veeraraghavan H, et al. MRI radiomic features are associated with survival in melanoma brain metastases treated with immune checkpoint inhibitors. *Neuro Oncol*. 2019;21(12):1578–1586.
20. Karami E, Soliman H, Ruschin M, et al. Quantitative MRI biomarkers of stereotactic radiotherapy outcome in brain metastasis. *Sci Rep*. 2019;9(1):19830.
21. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310–1320.
22. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104–e107.
23. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. Paper presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; June 26–July 1; 2016; San Francisco, CA:785–794.
24. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Paper presented at: Advances in Neural Information Processing Systems; December 4–9; 2017; Long Beach, CA:4765–4774.
25. Ribeiro MT, Singh S, Guestrin C. Why should i trust you? Explaining the predictions of any classifier. Paper presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; June 26–July 1; 2016; San Francisco, CA:1135–1144.
26. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56–67.
27. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565–574.
28. Zhong J, Sun Y, Peng W, Xie M, Yang J, Tang X. XGBFEMF: an XGBoost-based framework for essential protein prediction. *IEEE Trans Nanobiotechnology*. 2018;17(3):243–250.
29. Li Y, Kang K, Krahn JM, et al. A comprehensive genomic pan-cancer classification using The Cancer Genome Atlas gene expression data. *BMC Genomics*. 2017;18(1):508.
30. Sawrie SM, Guthrie BL, Spencer SA, et al. Predictors of distant brain recurrence for patients with newly diagnosed brain metastases treated with stereotactic radiosurgery alone. *Int J Radiat Oncol Biol Phys*. 2008;70(1):181–186.
31. McTyre E, Ayala-Peacock D, Contessa J, et al. Multi-institutional competing risks analysis of distant brain failure and salvage patterns after upfront radiosurgery without whole brain radiotherapy for brain metastasis. *Ann Oncol*. 2018;29(2):497–503.
32. Mouraviev A, Detsky J, Sahgal A, et al. Use of radiomics for the prediction of local control of brain metastases after stereotactic radiosurgery. *Neuro Oncol*. 2020;22(6):797–805.
33. Wu CX, Lin GS, Lin ZX, Zhang JD, Liu SY, Zhou CF. Peritumoral edema shown by MRI predicts poor clinical outcome in glioblastoma. *World J Surg Oncol*. 2015;13(1):97.
34. Jansen EP, Dewit LG, van Herk M, Bartelink H. Target volumes in radiotherapy for high-grade malignant glioma of the brain. *Radiother Oncol*. 2000;56(2):151–156.
35. Nardone V, Nanni S, Pastina P, et al. Role of perilesional edema and tumor volume in the prognosis of non-small cell lung cancer (NSCLC) undergoing radiosurgery (SRS) for brain metastases. *Strahlenther Onkol*. 2019;195(8):734–744.
36. Tini P, Nardone V, Pastina P, et al. Perilesional edema in brain metastasis from non-small cell lung cancer (NSCLC) as predictor of response to radiosurgery (SRS). *Neurol Sci*. 2017;38(6):975–982.
37. Magnuson WJ, Lester-Coll NH, Wu AJ, et al. Management of brain metastases in tyrosine kinase inhibitor-naïve epidermal growth factor receptor-mutant non-small-cell lung cancer: a retrospective multi-institutional analysis. *J Clin Oncol*. 2017;35(10):1070–1077.
38. Wang TJ, Saad S, Qureshi YH, et al. Does lung cancer mutation status and targeted therapy predict for outcomes and local control in the setting of brain metastases treated with radiation? *Neuro Oncol*. 2015;17(7):1022–1028.

39. Liu Y, Kim J, Balagurunathan Y, et al. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clin Lung Cancer*. 2016;17(5):441–448.e446.
40. Han Y, Xie Z, Zang Y, et al. Non-invasive genotype prediction of chromosome 1p/19q co-deletion by development and validation of an MRI-based radiomics signature in lower-grade gliomas. *J Neurooncol*. 2018;140(2):297–306.
41. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328–338.