




# The radiomic-clinical model using the SHAP method for assessing the treatment response of whole-brain radiotherapy: a multicentric study

Yixin Wang<sup>1,2,3</sup> · Jinwei Lang<sup>1,2</sup> · Joey Zhaoyu Zuo<sup>1,2</sup> · Yaqin Dong<sup>4</sup> · Zongtao Hu<sup>1,3</sup> · Xiuli Xu<sup>3</sup> · Yongkang Zhang<sup>3</sup> · Qinjie Wang<sup>1,2</sup> · Lizhuang Yang<sup>1,2,3</sup> · Stephen T. C. Wong<sup>5</sup> · Hongzhi Wang<sup>1,2,3</sup> · Hai Li<sup>1,2,3</sup> 

Received: 4 April 2022 / Revised: 11 May 2022 / Accepted: 16 May 2022  
© The Author(s), under exclusive licence to European Society of Radiology 2022

## Abstract

**Objective** To develop and validate a pretreatment magnetic resonance imaging (MRI)-based radiomic-clinical model to assess the treatment response of whole-brain radiotherapy (WBRT) by using SHapley Additive exPlanations (SHAP), which is derived from game theory, and can explain the output of different machine learning models.

**Methods** We retrospectively enrolled 228 patients with brain metastases from two medical centers (184 in the training cohort and 44 in the validation cohort). Treatment responses of patients were categorized as a non-responding group vs. a responding group according to the Response Assessment in Neuro-Oncology Brain Metastases (RANO-BM) criteria. For each tumor, 960 features were extracted from the MRI sequence. The least absolute shrinkage and selection operator (LASSO) was used for feature selection. A support vector machine (SVM) model incorporating clinical factors and radiomic features was used to construct the radiomic-clinical model. SHAP method explained the SVM model by prioritizing the importance of features, in terms of assessment contribution.

**Results** Three radiomic features and three clinical factors were identified to build the model. Radiomic-clinical model yielded AUCs of 0.928 (95%CI 0.901–0.949) and 0.851 (95%CI 0.816–0.886) for assessing the treatment response in the training cohort and validation cohort, respectively. SHAP summary plot illustrated the feature's value affected the feature's impact attributed to model, and SHAP force plot showed the integration of features' impact attributed to individual response.

**Conclusion** The radiomic-clinical model with the SHAP method can be useful for assessing the treatment response of WBRT and may assist clinicians in directing personalized WBRT strategies in an understandable manner.

## Key Points

- Radiomic-clinical model can be useful for assessing the treatment response of WBRT.
- SHAP could explain and visualize radiomic-clinical machine learning model in a clinician-friendly way.

**Keywords** Magnetic resonance imaging · Neoplasm metastasis · Machine learning · Radiotherapy · Game theory

## Abbreviations

3D 3 dimensional

ADC Apparent diffusion coefficient

AUC

Area under the curve

CET1-w

Contrast-enhanced T1-weighted

CI

Confidence interval

✉ Hongzhi Wang  
wanghz@hfcas.ac.cn

✉ Hai Li  
hli@cmpt.ac.cn

<sup>1</sup> Anhui Province Key Laboratory of Medical Physics and Technology, Institute of Health and Medical Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, People's Republic of China

<sup>2</sup> University of Science and Technology of China, Hefei 230026, People's Republic of China

<sup>3</sup> Department of Oncology, Hefei Cancer Hospital, Chinese Academy of Sciences, Hefei 230031, People's Republic of China

<sup>4</sup> Department of Radiation Oncology, The First Affiliated Hospital of Anhui Medical University, Hefei 230022, People's Republic of China

<sup>5</sup> Department of Systems Medicine and Bioengineering, Houston Methodist Cancer Center, Weill Cornell Medical College, Houston, TX 77030, USA

CR	Complete response
DCE MRI	Contrast-enhanced magnetic resonance imaging
DS-GPAs	Disease-specific Graded Prognostic Assessments
ECM	Extracranial metastases
ER	Estrogen receptor
glcm	Gray-level co-occurrence matrix
glszm	Gray-level size zone matrix
HER2	Human epidermal growth factor receptor 2
I	Idn
KPS	Karnofsky Performance Score
LASSO	Least absolute shrinkage and selection operator
LGLZE	Low Gray Level Zone Emphasis
M	Median
MRI	Magnetic resonance imaging
NSCLC	Non-small cell lung cancer
ORR	Objective response rate
OS	Overall survival
PACS	Picture archiving and communication system
PD	Progressive disease
PR	Partial response
PR	Progesterone receptor
ROC	Receiver-operating characteristic
SD	Stable disease
SHAP	SHapley Additive exPlanations
SRS	Stereotactic radiosurgery
SVM	Support vector machine
WBRT	Whole brain radiotherapy

## Introduction

Brain metastases are a serious manifestation of cancer and are associated with poor clinical outcomes [1]. Despite recent advances in radiation oncology, neurosurgery, and systemic therapies, identifying an effective treatment for an individual patient with brain metastases remains a challenge [2–4].

Whole-brain radiotherapy (WBRT) irradiates the whole brain and targets patients with brain metastases who are not suitable for surgery or radiosurgery [1]. WBRT has the advantage of reducing the rate of recurrence and the emergence of new brain lesion [5]. In addition, according to clinical trials, patients with brain metastases who have good treatment responses, such as partial response (PR) or complete response (CR), usually have a higher overall survival (OS) [6, 7]. However, the objective response rate (ORR) of WBRT is not high, and nearly half of the patients cannot benefit from WBRT [8]. Meanwhile, the potential benefits of WBRT must be balanced against radiation-related side effects [9]. Therefore, precisely assessing the treatment response before WBRT is very important.

At present, clinical factors, such as primary tumor type and disease-specific Graded Prognostic Assessments (DS-GPAs)

which is an objective prognostic scoring system for patients with brain metastases, are commonly used to assist the radiation oncologists in identifying proper WBRT strategies, whereas they are seldom applied to assess the WBRT treatment response directly [10–14]. With the development of advanced medical image techniques, diffusion MRI outlined viable tumor volume with linear mixed effect model; apparent diffusion coefficient (ADC) value with binomial regression; and water exchange rate constant metrics with Pearson correlation analysis have been proposed as alternatives to assess the treatment response of brain metastases radiotherapy. Although these models fulfill the demand of assessing the treatment response, they ignore the comprehensive information that can be obtained from MRI. Furthermore, these studies with insufficient validation which cannot satisfy clinical requirements. [15–17].

Radiomics could provide a better understanding of cancer pathophysiology for clinical decision-making by using a large number of quantitative image features to characterize the tumor heterogeneity [18]. Several studies have reported the successful use of radiomics to assess the treatment response of Gamma Knife radiosurgery or stereotactic radiosurgery (SRS) therapy in patients with brain metastases [19–21]. However, the efficacy of radiomic approach on the discrimination of WBRT treatment response remains unknown. More importantly, the effectiveness of radiomics approach is influenced by many factors, such as imaging hardware configuration, image acquisition, image postprocessing, and software implementation; radiomics would encounter the problem of generalizability and would not be able to translate radiomic model into clinical practice [18, 22]. Therefore, a good model must be fully validated, especially external validation, in order to be applied in the clinic.

In fact, many radiomic models have been developed and achieved good results in limited dataset or applications [23, 24]. However, radiation oncologists are hesitant to adopt such models due to the unclear internal mechanisms or unexplained features [25]. In other words, interpretability largely hinders the general application of the radiomic model. The SHapley Additive exPlanations (SHAP) method may help resolve this problem. SHAP is a unified framework of six different methods used for interpreting predictions, it defines a class of additive feature importance measures and theoretical results [26]. SHAP has several advantages, including illustrating the importance of the features and their impact on the overall prediction model and understanding the importance of individual features to the model output [27, 28]. The combination of SHAP and radiomics may illustrate models in an explicable way [29–31].

Thus, we constructed radiomic-clinical machine learning models to assess the treatment response of WBRT and verified the generalizability of models by external validation. Meanwhile, we combined models with the SHAP method to

explain and visualize our models. It is hypothesized that radiomic-clinical models with the SHAP method could be useful and understandable in assessing the treatment response of WBRT.

## Materials and methods

### Study population

This retrospective clinical study was conducted in Hefei Cancer Hospital, Chinese Academy of Sciences (cohort 1), and the First Affiliated Hospital of Anhui Medical University (cohort 2). The study was conducted in accordance with the Declaration of Helsinki, and ethics review approval was obtained by the participating centers. A total of 228 patients with a histopathologic diagnosis of non-small cell lung cancer (NSCLC) or breast cancer from the two centers were included. The inclusion criteria and exclusion criteria are shown in Supplementary Information 1. We used cohort 1 as the training cohort to train models for individual assessment. Then we used cohort 2 as the validation cohort to validate models.

From the electronic medical record, demographic and clinical information were retrieved, including the following: gender, age, primary tumor type, radiation dose, biological equivalent dose, chemotherapy use, DS-GPAs scores, and follow-up days.

### WBRT

WBRT was delivered with volumetric modulated arc therapy, and WBRT used 6 MV X-ray generated by medical accelerators (Elekta Axesse (4D) in the training cohort, Varian 23EX in the validation cohort). WBRT modalities were classified as follows: 40 Gy/20f, 37.5 Gy/15f, and 30 Gy/10f. Tumors received a boost dose ranging from 10 to 20 Gy through simultaneous or sequential radiotherapy to the BM lesions.

### Assessment of WBRT treatment response

Each patient in the present study had a follow-up MRI for response assessment, typically between 6 and 8 weeks after the completion of radiation therapy. The Response Assessment in Neuro-Oncology Brain Metastases (RANO-BM) criteria were used to determine an outcome of CR (lesions that were not present on follow-up MRI), PR (> 30% reduction in the sum longest diameter of brain lesions), SD (neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD), and PD (> 20% increase in the sum longest diameter of brain lesions) [32].

ORR was widely demonstrated for assessing the treatment response, the proportion of patients with PR and CR

in all patients was defined as ORR [33]. Therefore, we included patients with PR/CR as the responding group and patients with PD/SD as the non-responding group. The detailed process of patient enrollment and grouping is shown in Fig. 1.

### Clinical factors

Considering that these clinical factors were relatively important in clinical practice, primary tumor type, sex, and DS-GPAs scores were used for clinical model construction regardless of their statistical significance in the training cohort and validation cohort. The DS-GPAs scores (Table 1) were calculated for all patients based on primary tumor type [11, 34]. At the same time, univariate analysis was used to assess the association between other clinical factors and WBRT treatment response in the training cohort and validation cohort, respectively. Significant clinical factors with  $p < 0.05$  in the training cohort were identified as potential clinical risk factors and were included for clinical model building.

### MRI data acquisition, image postprocessing, tumor segmentation, and feature extraction

All images were retrieved from a picture archiving and communication system (PACS). All patients from cohort 1 and cohort 2 underwent an MRI protocol, which included a contrast-enhanced T1-weighted (CET1-w) 3-dimensional (3D) sequence.

The detailed processes of MRI data acquisition, image postprocessing, tumor segmentation, and feature extraction were described in Supplementary Information 2.

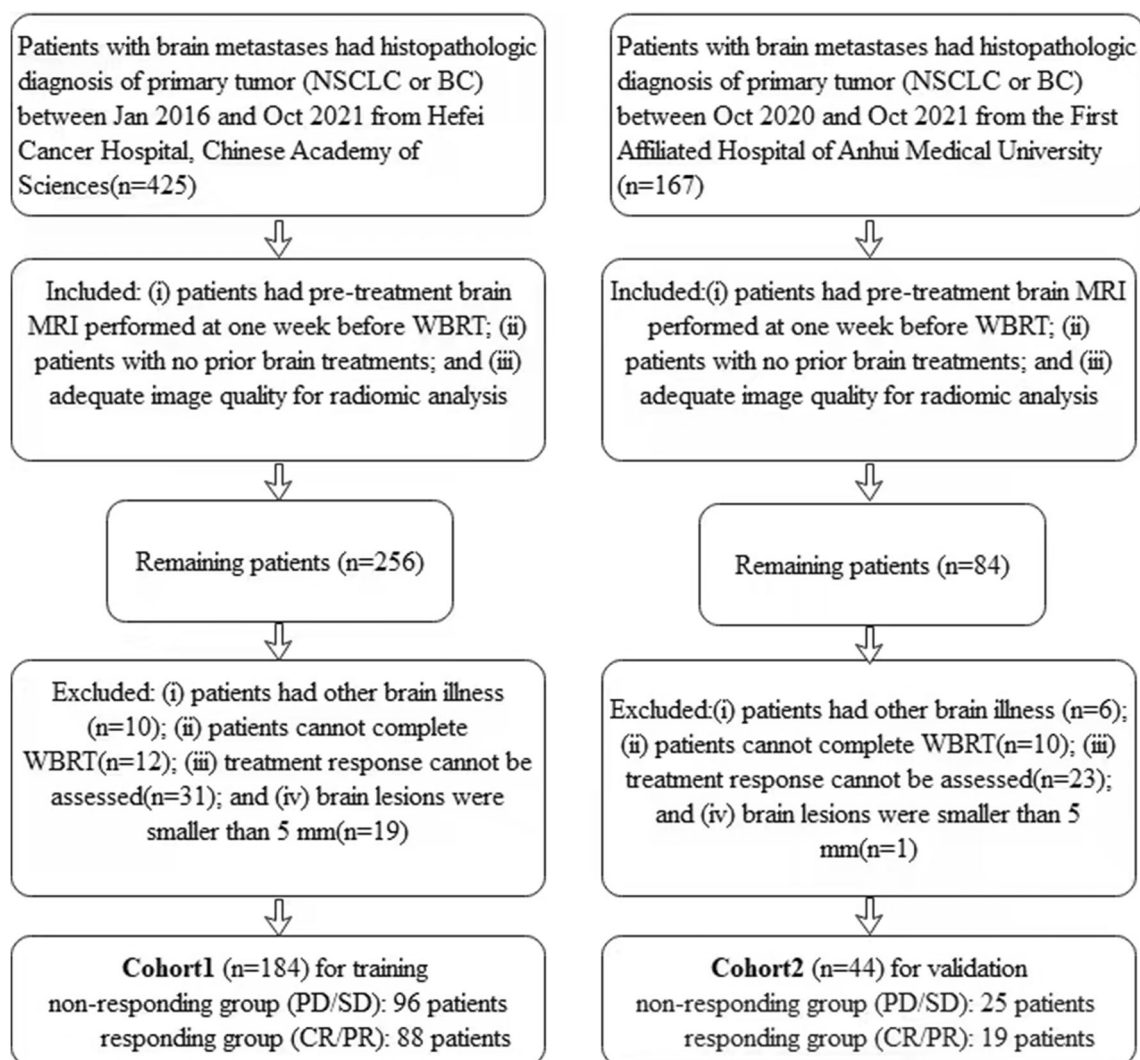
### Quantitative radiomic analysis

For patients with multiple metastases, an average radiomic feature value of respective metastases was assigned to assess WBRT treatment response at the level of the individual patient.

Feature selection was performed with the training cohort. Four steps were used for feature selection to reduce any bias of results and potential overfitting. The detailed process of quantitative radiomic analysis could be seen in Supplementary Information 3.

### Model building

We constructed a clinical model and radiomic-based models (radiomic model and radiomic-clinical model). A widely used machine learning classifier of support vector machine (SVM) was used to classify patients into the PD/SD or CR/PR groups [35]. SVM with two kernels (linear and radial basis function) models were conducted to optimize the



**Fig. 1** Flow chart of patient enrollment and grouping

hyper-parameters in the training cohort by using a grid search method and 5-fold cross-validation. Then model with the

highest area under curve (AUC) value was chosen for computing the performance of the validation cohort.

**Table 1** Definition of diagnosis-specific graded prognostic assessments indexes for patients with different primary tumor

Primary tumor	Prognostic factors	GPA scoring criteria				
		0	0.5	1	1.5	2
NSCLC	Age	> 60	50–60	< 50		
	KPS	< 70	70–80	90–100		
	EMC	Present		Absent		
	No. of BMs	> 3	2–3	1		
Breast cancer	Age	> 60	< 60			
	KPS	< 50	60	70–80	90–100	
	Subtype	Basal		LumA	Her2	LumB

Subtype definitions: *Basal*, triple negative; *LumA*, ER/PR positive, HER2 negative; *LumB*, triple positive; *HER2*, ER/PR negative, HER2 positive; *BM*, brain metastasis; *ECM*, extracranial metastases; *ER*, estrogen receptor; *HER2*, human epidermal growth factor receptor 2; *KPS*, Karnofsky Performance Score; *NSCLC*, non-small cell lung carcinoma; *PR*, progesterone receptor



Receiver-operating characteristic (ROC) curve, specificity, sensitivity, precision, and AUC with a 95% confidence interval (95%CI) were calculated to assess the performance of SVM models [36]. The performance of different ROC curves in the training cohort and the validation cohort was then compared by the DeLong test. The calibration curves were used to assess the calibration of the SVM models in the training and validation cohort [37].

## Model explanation and visualization

SHAP originated from a coalitional game theory that had been introduced to eliminate the black-box effect of machine-learning models, thereby enabling the identification and prioritization of features that determine compound classification [26–28]. Shapely (SHAP) value was the average expected marginal contribution of one player on the model after considering all possible combinations had been considered, and it was applied primarily to circumstances in which the contributions of each player were unequal, but players worked together to obtain the payoff [26, 28].

We used a SHAP kernel explainer to evaluate feature attributions, assuming that each feature of the patient was a player in a game where the prediction was the payout. The contribution of each player indicated the feature's importance to the prediction. The workflow is shown in Fig. 2.

## Statistical analysis

All statistical analyses were performed using Python version 3.5 (<https://www.python.org>). Continuous variables were expressed as mean (SD) and compared using an unpaired, two-tailed *t*-test (continuous variables with a normal distribution) or Mann–Whitney test (continuous variables without a

normal distribution). Categorical variables were compared using the  $\chi^2$  test or Fisher exact test. In all test analyses,  $p < 0.05$  was considered statistically significant. The packages used in the study are listed in Supplementary Information 4.

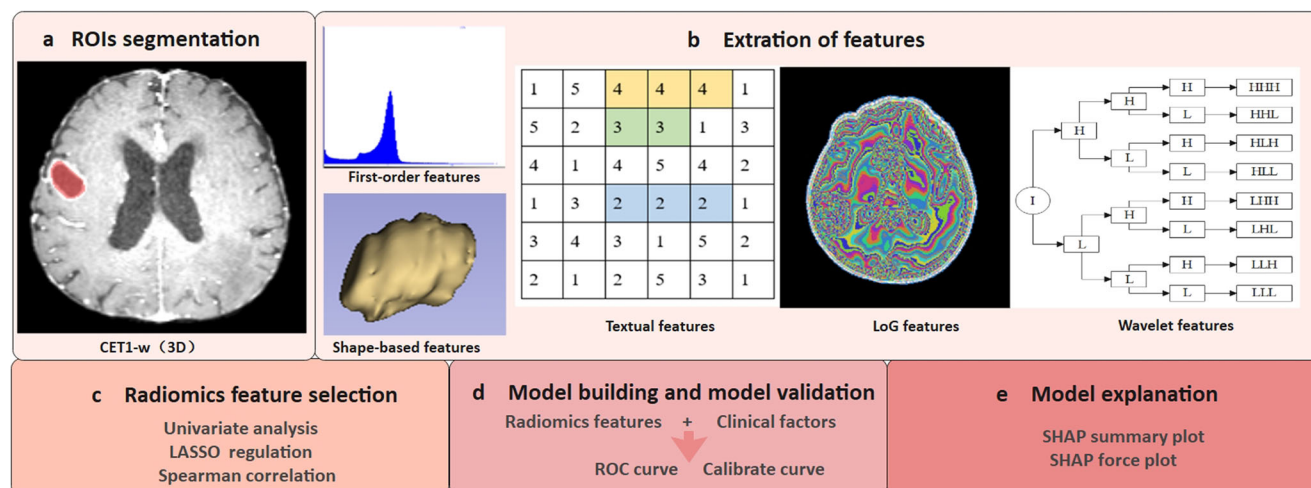
## Results

### Patient characteristics

A total of 184 eligible patients with 512 brain metastases were enrolled from cohort 1. Among these, 96 patients with brain metastases were PD/SD (non-responding group), and 88 patients were CR/PR (responding group). A total of 44 eligible patients with 81 brain metastases were enrolled in cohort 2. Among these, 25 patients with brain metastases were PD/SD (non-responding group); and 19 patients were CR/PR (responding group). No significant differences were found in clinical characteristics between cohort 1 and cohort 2. Detailed clinical characteristics from two cohorts are shown in Supplementary Table 4.

### Performance of clinical model to assess WBRT treatment response

Univariate analysis (Table 2) showed that no clinical factors were significantly related to the treatment response. Thus, primary tumor type, sex, and DS-GPAs scores were used for clinical model construction. The optimal SVM model used a linear kernel. Performances of the clinical model in the training cohort and validation cohort are shown in Table 3. The ROC curves with the clinical model in the training and validation cohorts are shown in Supplementary Fig. 1.



**Fig. 2** Workflow of constructing and validating radiomic-clinical model. **a** Segmentation of ROI. **b** Radiomic features. **c** Univariate analysis, LASSO, and Spearman correlation were used for feature selection. **d**

ROC curve and calibration curve were used to assess and validate the performance of different models. **e** SHAP summary plot and SHAP force plot were used to explain the contribution of features

**Table 2** Clinical characteristics comparison between non-responding and responding groups of training cohort and validation cohort

Characteristic	Training cohort			Validation cohort		
	Responding group (N = 88)	Non-responding group (N = 96)	p value	Responding group (N = 19)	Non-responding group (N = 25)	p value
Age, years	59.943 ± 10.612	59.521 ± 11.664	0.058	56.210 ± 8.97	53.84 ± 11.93	0.577
Sex (%)			0.314			0.339
Male	48 (54.54)	58 (60.42)		7 (36.84)	14 (56.00)	
Female	40 (45.46)	38 (39.58)		12 (63.16)	11 (44.00)	
Primary tumor type (%)			0.161			0.971
Non-small cell lung cancer	66 (62.26)	81 (80.19)		15 (60.00)	21 (80.77)	
Breast cancer	22 (20.75)	15 (14.85)		4 (16.00)	4 (15.38)	
Dose of BMs, Gy	38.60 ± 5.34	40.78 ± 5.93	0.472	39.11 ± 6.04	41.00 ± 7.22	0.334
BED of BMs, Gy	46.48 ± 6.04	48.77 ± 5.93	0.958	47.90 ± 5.78	49.67 ± 7.41	0.147
Chemotherapy use (%)			0.477			0.856
Yes	55 (62.5)	54 (56.25)		9 (52.63)	10 (40.00)	
No	33 (37.5)	42 (43.75)		10 (47.37)	15 (60.00)	
DS-GPAs scores	1.80 ± 0.85	1.86 ± 0.83	0.804	1.73 ± 0.63	1.54 ± 0.75	0.646
Image follow-up, days	47.09 ± 4.20	47.27 ± 4.42	0.474	49.10 ± 3.99	49.00 ± 4.35	0.851

*BED*, biological equivalent dose; *BM*s, brain metastases; *DS-GPAs*, diagnosis-specific graded prognostic assessments

The clinical model with three factors yielded AUCs of 0.650 (95%CI 0.579–0.721) in the training cohort. In the validation cohort, the clinical model yielded an unsatisfying AUC of 0.598 (95%CI 0.589–0.607).

### Radiomic feature selection

For each tumor, a total of 960 features were extracted from the CET1-w (3D) MR sequence. After univariate analysis, 548 features were selected. Considering the fact that a minimum of 10 patients per individual radiomic feature was investigated, as well as the sample size in this study, the coefficients of some radiomic features were subsequently compressed to zero by adding penalty terms [38]. Therefore, after LASSO regression, three features (Table 4) were selected. LASSO with 10-fold cross-validation and the LASSO coefficient are shown in Supplementary Fig. 2. Then, we calculated the Spearman correlation coefficients (Supplementary Fig. 3) between each pair of features, and no pair of features were excluded.

### Performance of radiomic-based models to assess WBRT treatment response

The optimal SVM models used a linear kernel. The performances of radiomic-based models in the training cohort and the validation cohort are shown in Table 5. The ROC curves with radiomic-based models in the training and the validation cohorts are shown in Supplementary Fig. 4 and Supplementary Fig. 5.

Radiomic-based models yielded satisfying integrated AUCs of 0.928, which were better than those of the clinical model in the training cohort. In the validation cohort, radiomic model and radiomic-clinical model also yielded good AUCs of 0.837 (95%CI 0.827–0.847) and 0.851 (95%CI 0.816–0.886), which were better than those of the clinical model. The ROC curves of SVM models in the training cohort and validation cohort are shown in Fig. 3. The *p* values of Delong test were both smaller than 0.01 when comparing the radiomic-based models with the clinical model.

Calibration plots graphically showed good agreements of radiomic-based models between the actual rate and predicted

**Table 3** Performances of clinical model to assess WBRT treatment response in the training cohort and validation cohort

Clinical model	AUC (95%CI)	SENS	SPEC	Precision	Accuracy
Training cohort	0.650 (0.579–0.721)	0.655	0.407	0.543	0.536
Validation cohort	0.598 (0.589–0.607)	0.840	0.263	0.600	0.591

*AUC*, area under curve; *SENS*, sensitivity; *SPEC*, specificity; *CI*, confidence interval

**Table 4** Radiomic features after feature selection

MR sequence	Feature name	Abbreviation
CET1-w (3D)	original_firstorder_Median	CET1-w(3D)_firstorderM
	log-sigma-1mm-3D_glcml_Idn	CET1-w(3D)_sigma1_glcml
	log-sigma-1mm-3D_glszm_LowGrayLevelZoneEmphasis	CET1-w(3D)_sigma1_glszmLGLZE

CET1-w, contrast-enhanced T1-weighted; 3D, 3 dimensional; *glcml*, gray-level co-occurrence matrix; *glszm*, gray-level size zone matrix

probability both in the training cohort (Supplementary Fig. 6) and validation cohort (Supplementary Fig. 7).

### Explanation and visualization of radiomic-clinical model

SHAP provided a quantitative explanation for the SVM method. SHAP summary plots offered a visually concise figure by representing the range and distribution of importance that feature had on the model's output and by relating the feature's value to the feature's impact. Features were sorted by their global importance first. Each dot representing the SHAP value of each feature from a patient was plotted horizontally and was stacked vertically to show the density of the same SHAP value. Then, each dot was colored by the value of the feature, from low (blue) to high (red). In Fig. 4, we observed that CET1-w (3D)\_firstorderM was the most important feature for discriminating PD/SD classification. The density of CET1-w (3D)\_firstorderM plot showed that the SHAP values of this feature were different in the cohort, and the coloring showed that the model's output increased with decreasing value of this feature.

The force plot (Fig. 5) could interpret the assessment of a single patient. It visualized the SHAP value of a feature as a force to increase or decrease the assessment, and every prediction began with the base value (0.5566), which was the average SHAP value of all predictions. The arrow's length accounted for how much (in percentage) a particular feature contributed to the SHAP value. The arrow's color represented whether the contributions were positive (red) or negative (blue).

As shown in Fig. 5A, the SHAP value of this patient was 0.80, which was larger than the base value (0.5566), thereby indicating that we could assess this patient in the non-

responding classification. Among these features, positive (red) CET1-w (3D)\_firstorderM arrow with a value of −1.222 had a great contribution to the assessment of the non-responding group. As shown in Fig. 5B, for another patient, the SHAP value was 0.12 which was smaller than the base value (0.5566). Thus, we could assess that this patient belonged to the responding group. CET1-w(3D)\_firstorderM arrow with a value of 0.09947 made a negative (blue) effort to assess non-responding classification.

### Discussion

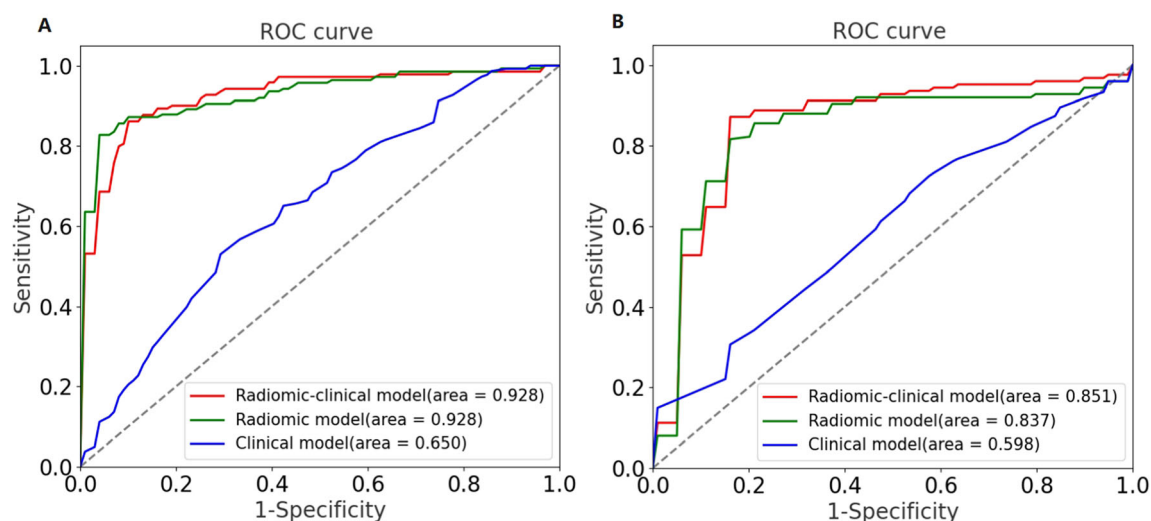
We demonstrated that radiomic-based SVM models had a good ability to assess the treatment response of WBRT, and the external validation also showed good generalizability.

We found AUCs of radiomic-based models were better than the clinical model in assessing the treatment response. DS-GPA scores were an important biomarker among the clinical factors. Patients with high DS-GPA scores were classified as the responding group. However, the clinical model with DS-GPAs scores performed poorly, and a little improvement was observed when clinical factors were added to the radiomic model. DS-GPAs are used to classify patients with brain metastases into different prognostic groups [11]. Patients with good performance status and limited metastatic lesions are usually classified by DS-GPAs as a good prognostic, which is recommended to receive SRS. However, in a previous SRS study, no significant difference was found in the overall survival of patients with different numbers of metastatic lesions, indicating that the clinical model with DS-GPAs was unreliable when used to assess the treatment response of radiotherapy [39]. Despite clinical model having a limited ability to assess the treatment response of WBRT,

**Table 5** Performances of radiomic-based models to assess WBRT treatment response in the training cohort and validation cohort

Radiomic-based models	Cohort	AUC (95%CI)	SENS	SPEC	Precision	Accuracy
Radiomic model	Training cohort	0.928 (0.909–0.947)	0.897	0.889	0.897	0.893
	Validation cohort	0.837 (0.827–0.847)	0.800	0.842	0.870	0.818
Radiomic-clinical model	Training cohort	0.928 (0.901–0.949)	0.931	0.889	0.900	0.911
	Validation cohort	0.851 (0.816–0.886)	0.880	0.842	0.880	0.864

AUC, area under curve; SENS, sensitivity; SPEC, specificity; CI, confidence interval

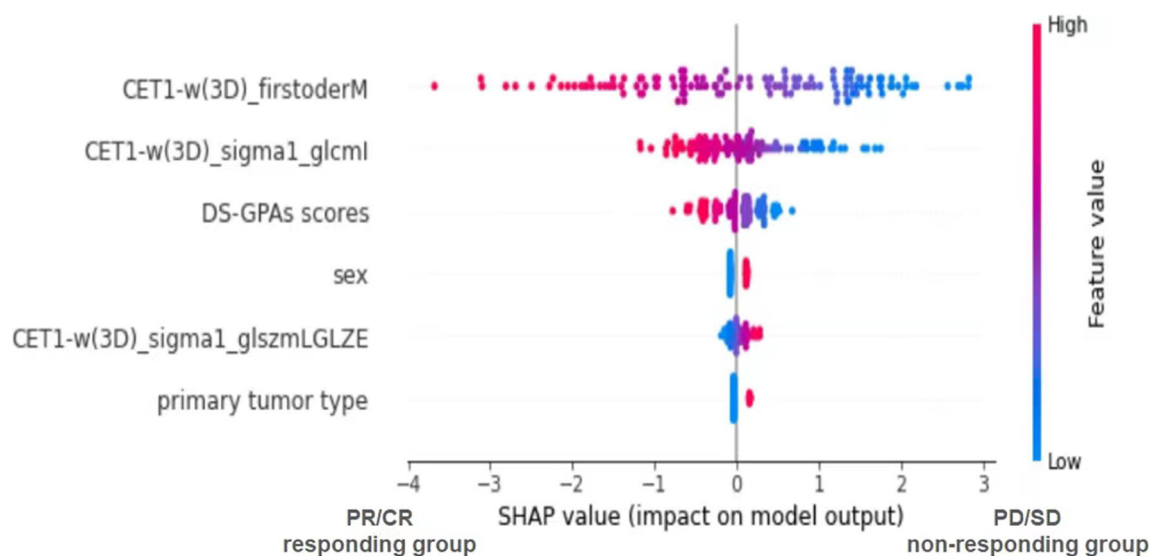


**Fig. 3** Comparison of receiver operating characteristic curves for assessing the treatment response of WBRT in the training cohort (**A**) and validation cohort (**B**)

radiomic-based models showed a good performance, which may direct radiation oncologists to personalize the treatment of patients with brain metastases

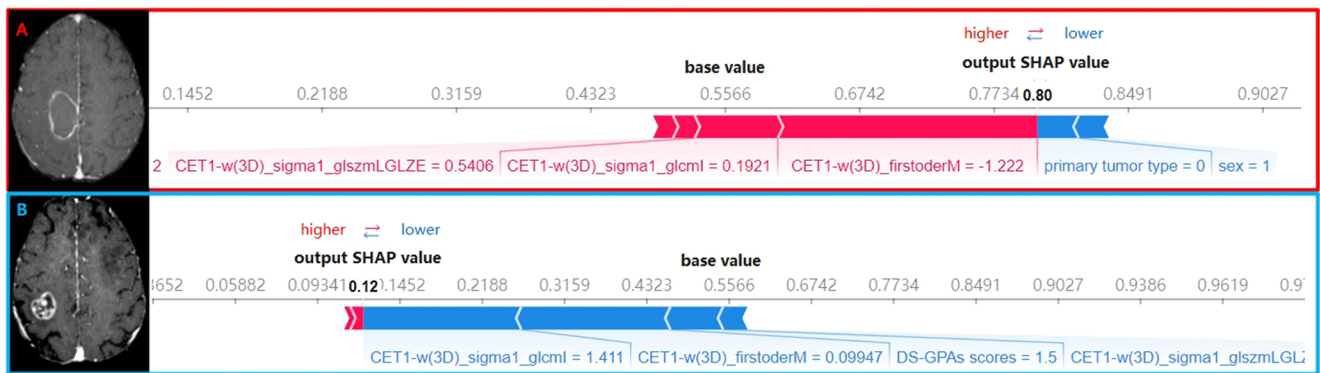
For the development and validation of radiomic-based models, we selected three radiomic features. Feature extraction inherently extracted more than a million voxels per MR imaging sequence into a handful of radiomic features, and strong radiomic features were stable in different scanners and captured distinctive tumor information [40]. According to the result of external validation, radiomic-based models showed good performance and generalizability. The underlying reason for this finding may be the fact that radiomic-based models had strong features that helped capture distinct patterns correlated with WBRT treatment response.

Strong features were obtained from CET1-w (3D) MR sequence. Different MR imaging sequences exploit different biophysical properties of brain tumors. CET1-w (3D) MR sequence could reflect enhancing regions (that show variations in blood flow) and necrosis due to gadolinium leakage from the intravascular space into the tumor [41]. CET1-w (3D) MR sequence also provides detailed anatomical information with high spatial resolution [42, 43]. Therefore, CET1-w (3D) MR sequence shows good sensitivity when used for the detection of brain metastases and is useful in clinical practice. Given the advantages of CET1-w (3D) MR sequence, we only used this MR sequence for the easy generalization of our radiomic-based models. However, the validation cohort was scanned with a different MRI protocol in the presented study,



**Fig. 4** SHAP summary plots of radiomic-clinical model. The plot illustrated the feature relevance and combined feature attributions to the model's predictive performance. M: median; LGLZE: Low Gray Level Zone Emphasis; I: Idn; DS-GPAs: disease-specific Graded Prognostic Assessments





**Fig. 5** SHAP force plots explained how the radiomic-clinical model discriminates the treatment response of two patients. The treatment response of patient A was PD/SD (non-responding group), and the treatment response of patient B was CR/PR (responding group). For instance, low feature value of CET1-w(3D)\_firstoderM contributed to the increase in

the assessment probability of the non-responding group. Patient B had a CET1-w(3D)\_firstoderM value of 0.09947, while a lower CET1-w(3D)\_firstoderM value of patient A (-1.222) contributed to assessing the non-responding group. Primary tumor type: 0 = non-small cell lung carcinoma; 1 = breast cancer. Sex: 0 = male; 1 = female

and different MRI protocols may be a risk of bias and would affect the generalizability of radiomic model; for example, the intensity of images was not quantitative. Thereby image postprocessing including resampling and intensity normalization was performed to improve the generalizability of radiomics.

Consistent with a previous study [21], our radiomic-based models built with SVM showed good performance. Although SVM-based models are widely used and powerful [44, 45], they cannot be applied to clinical practice unless the model can be explained. SHAP could provide an explanation and visualization of SVM models in a clinician-friendly way through the SHAP summary plot and SHAP force plot. From a global perspective, the SHAP summary plot could provide a visually concise figure by representing the range and distribution of importance that features impact on the model's output (the more important the feature, the wider dots' range) and by relating the feature's value to the feature's impact, from low (blue) to high (red), thereby providing a new alternative feature importance plots of the whole cohort. By contrast, the ordinary feature importance bar only offered relative importance. In this study, clinicians could clearly see how the CET1-w (3D)\_firstoderM or other features' value impacted the assessment through dots' range and color. At the same time, clinicians can simultaneously identify the strongest feature of the model. Pathophysiologically, SHAP offered us the strongest nonsemantic feature CET1-w(3D)\_firstoderM [46]. "CET1-w (3D)\_firstoderM" described the median percentile of gray values within the volume [47]. In this study, patients in the non-responding group had low "CET1-w(3D)\_firstoderM" than patients in the responding group; thus, images of non-responding tumors were rather dark. This result may indicate the lack of a gadolinium-based contrast agent due to poor vascular supply. With regard to the mechanism of radiation resistance, tumors with poor vascular support, which would result in hypoxic

conditions, usually showed resistant outcome [48]. The findings of the present study confirmed a previous study that heterogeneity and high gray levels might be associated with intertwined processes, such as chaotic vascularization [49]. To our knowledge, this study was the first to assess the WBRT treatment response by integrating radiomics with the SHAP method.

After clinicians understanding how features impacted the SVM model and the potential pathophysiological mechanism, they would like to use this model to assess individual outcomes. From the perspective of an individual patient, a local explanation of a single patient's assessment could be implemented by using a SHAP force plot. Compared with the nomogram method, which made clinicians calculate some features' value to generate total points [50], the SHAP force plot was time-saving and easier to use. Clinicians could compare the output SHAP value of a single patient with the base value directly. In the presented model, if the output SHAP value was bigger than base value, clinicians could classify this patient into the non-responding group. At the same time, clinicians could realize how features impacted an individual patient's assessment through seeing arrow's color (red color describes a force to increase the assessment for the non-responding group) and arrow's length which described how much a particular feature contributed to the assessment. As shown in Fig. 5A (red), CET1-w (3D)\_firstoderM and (red) CET1-w (3D)\_sigma1\_glcml had a positive contribution to the assessment of the non-responding group, and the arrow's length of CET1-w (3D)\_firstoderM was longer than CET1-w (3D)\_sigma1\_glcml, indicating that CET1-w (3D)\_firstoderM contributed more force to the assessment than CET1-w (3D)\_sigma1\_glcml.

The present study had some limitations. First, considering the generalizability and practicability of the model, we did not use multiparameter-MRI. However, other MR sequences may contain other biophysical properties of brain tumors that could

help assess the treatment response of WBRT. Therefore, multiparameter-MRI-based radiomics would be explored next. Second, although the situation that two features that were strongly correlated but represented different biophysiological characteristics did not happen in the present study, in order to better understand how these features contributed to the model, if such a situation occurs, we should include both features in the further analysis. Third, SHAP offered the strongest radiomic feature, but the pathological mechanism could not be associated with radiomic features directly. Local microscopic pathological image features and macroscopic histopathologic marker expression are encouraged in further analyses. Lastly, we used consensus ROIs for feature extraction to guarantee the accuracy of tumor segmentation and reduce the heterogeneity among radiation oncologists, which would result in the loss of partial tumor area information [51]. An automatic or semi-automatic tumor segmentation approach will be developed next.

## Conclusions

The proposed radiomic-clinical machine learning model that contained CET1-w (3D) MR sequence features can be useful for assessing the treatment response of WBRT. The SHAP method may assist clinicians in directing personalizing WBRT strategies in an understandable way.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00330-022-08887-0>.

**Acknowledgements** We thank the Hefei Cancer Hospital, the Chinese Academy of Science, and the First Affiliated Hospital of Anhui Medical University for their help and support in this project.

**Funding** This work was supported by the Key R&D Program of Anhui Province (201904a07020104), the Natural Science Fund of Anhui Province (2008085MC69), Collaborative Innovation Program of Hefei Science Center (2020HSC-CIP001, 2021HSC-CIP013), the General scientific research project of Anhui Provincial Health Commission (AHWJ2021b150), the Natural Science Fund of Hefei City (2021033), CAS Anhui Province Key Laboratory of Medical Physics and Technology (LMPT201904), Director's Fund of Hefei Cancer Hospital of CAS (YZJJ2019C14, YZJJ2019A04).

## Declarations

**Guarantor** The scientific guarantor of this publication is Hai Li.

**Conflict of interest** The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

**Statistics and biometry** No complex statistical methods were necessary for this paper.

**Informed consent** Written informed consent was waived by the Institutional Review Board.

**Ethical approval** Institutional Review Board approval was obtained.

## Methodology

- retrospective
- diagnostic or prognostic study
- multicenter study

## References

1. Suh JH, Kotecha R, Chao ST, Ahluwalia MS, Sahgal A, Chang EL (2020) Current approaches to the management of brain metastases. *Nat Rev Clin Oncol* 17:279–299
2. Greenspoon JN, Ellis PM, Pond G, Caetano S, Broomfield J, Swaminath A (2017) Comparative survival in patients with brain metastases from non-small-cell lung cancer treated before and after implementation of radiosurgery. *Curr Oncol* 24:e146–e151
3. Paek SH, Audu PB, Sperling MR, Cho J, Andrews DW (2005) Reevaluation of surgery for the treatment of brain metastases: review of 208 patients with single or multiple brain metastases treated at one institution with modern neurosurgical techniques. *Neurosurgery* 56:1021–1034
4. Goldberg SB, Gettinger SN, Mahajan A et al (2016) Pembrolizumab for patients with melanoma or non-small-cell lung cancer and untreated brain metastases: early analysis of a non-randomised, open-label, phase 2 trial. *Lancet Oncol* 17:976–983
5. Brown PD, Ballman KV, Cerhan JH et al (2017) Postoperative stereotactic radiosurgery compared with whole brain radiotherapy for resected metastatic brain disease (NCCTG N107C/CEC-3): a multicentre, randomised, controlled, phase 3 trial. *Lancet Oncol* 18:1049–1060
6. Brown PD, Jaeckle K, Ballman KV et al (2016) 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* 316:401–409
7. Li J, Bentzen SM, Renschler M, Mehta MP (2007) Regression after whole-brain radiation therapy for brain metastases correlates with survival and improved neurocognitive function. *J Clin Oncol* 25:1260–1266
8. Tian J, Luo Y, Xiang J, Tang J (2017) Combined treatment for non-small cell lung cancer and breast cancer patients with brain metastases with whole brain radiotherapy and temozolomide: a systematic review and meta-analysis. *J Neurooncol* 135:217–227
9. Andrews DW, Scott CB, Sperduto PW et al (2004) Whole brain radiation therapy with or without stereotactic radiosurgery boost for patients with one to three brain metastases: phase III results of the RTOG 9508 randomised trial. *Lancet* 363:1665–1672
10. Sperduto PW, Berkey B, Gaspar LE, Mehta M, Curran W (2008) A new prognostic index and comparison to three other indices for patients with brain metastases: an analysis of 1,960 patients in the RTOG database. *Int J Radiat Oncol Biol Phys* 70:510–514
11. Sperduto PW, Chao ST, Sneed PK et al (2010) Diagnosis-specific prognostic factors, indexes, and treatment outcomes for patients with newly diagnosed brain metastases: a multi-institutional analysis of 4,259 patients. *Int J Radiat Oncol Biol Phys* 77:655–661
12. Tjong MC, Mak DY, Shahi J, Li GJ, Chen H, Louie AV (2020) Current management and progress in radiotherapy for small cell lung cancer. *Front Oncol* 10:1146
13. Khalifa J, Amini A, Popat S, Gaspar LE, Faivre-Finn C, International Association for the Study of Lung Cancer Advanced Radiation Technology Committee (2016) Brain metastases from NSCLC:

- radiation therapy in the era of targeted therapies. *J Thorac Oncol* 11: 1627–1643
14. Sperduto PW, Kased N, Roberge D et al (2012) Summary report on the graded prognostic assessment: an accurate and facile diagnosis-specific tool to estimate survival for patients with brain metastases. *J Clin Oncol* 30:419–425
  15. Mahmood F, Hjorth Johannesen H, Geertsens P, Hansen RH (2020) Diffusion MRI outlined viable tumour volume beats GTV in intra-treatment stratification of outcome. *Radiother Oncol* 144:121–126
  16. Mahmood F, Johannesen HH, Geertsens P, Hansen RH (2017) Repeated diffusion MRI reveals earliest time point for stratification of radiotherapy response in brain metastases. *Phys Med Biol* 62: 2990–3002
  17. Mehrabian H, Desmond KL, Chavez S et al (2017) Water exchange rate constant as a biomarker of treatment efficacy in patients with brain metastases undergoing stereotactic radiosurgery. *Int J Radiat Oncol Biol Phys* 98:47–55
  18. Aerts HJ, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 5:4006
  19. Karami E, Ruschin M, Soliman H, Sahgal A, Stanisiz GJ, Sadeghi-Naini A (2019) An MR radiomics framework for predicting the outcome of stereotactic radiation therapy in brain metastasis. *Annu Int Conf IEEE Eng Med Biol Soc* 2019:1022–1025
  20. Huang CY, Lee CC, Yang HC et al (2020) Radiomics as prognostic factor in brain metastases treated with Gamma Knife radiosurgery. *J Neurooncol* 146:439–449
  21. Karami E, Soliman H, Ruschin M et al (2019) Quantitative MRI biomarkers of stereotactic radiotherapy outcome in brain metastasis. *Sci Rep* 9:19830
  22. Traverso A, Wee L, Dekker A, Gillies R (2018) Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys* 102:1143–1158
  23. Choi YS, Ahn SS, Chang JH et al (2020) Machine learning and radiomic phenotyping of lower grade gliomas: improving survival prediction. *Eur Radiol* 30:3834–3842
  24. Elshafeey N, Kotrotsou A, Hassan A et al (2019) Multicenter study demonstrates radiomic features derived from magnetic resonance perfusion images identify pseudoprogression in glioblastoma. *Nat Commun* 10:3170
  25. Xu Y, Liu X, Cao X et al (2021) Artificial intelligence: a powerful paradigm for scientific research. *Innovation (Camb)* 2:100179
  26. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 30 (Nips 2017) 30. 10.48550/arXiv.1705.07874
  27. Rodríguez-Pérez R, Bajorath J (2020) Interpretation of compound activity predictions from complex machine learning models using local approximations and Shapley values. *J Med Chem* 63:8761–8777
  28. Shapley LS (1988) A value for n-person games. In: Roth AE (ed) *The Shapley value: essays in honor of Lloyd S Shapley*. Cambridge University Press, Cambridge, pp 31–40
  29. Giraud P, Giraud P, Nicolas E et al (2020) Interpretable Machine Learning Model for Locoregional Relapse Prediction in Oropharyngeal Cancers. *Cancers (Basel)* 13
  30. Ma L, Xiao Z, Li K, Li S, Li J, Yi X (2020) Game theoretic interpretability for learning based preoperative gliomas grading. *Future Gener Comp Sys* 112:1–10
  31. Li R, Shinde A, Liu A et al (2020) Machine learning-based interpretation and visualization of nonlinear interactions in prostate cancer survival. *JCO Clin Cancer Inform* 4:637–646
  32. Lin NU, Lee EQ, Aoyama H et al (2015) Response assessment criteria for brain metastases: proposal from the RANO group. *Lancet Oncol* 16:e270–e278
  33. Bradbury P, Seymour L (2009) Tumor shrinkage and objective response rates: gold standard for oncology efficacy screening trials, or an outdated end point? *Cancer J* 15:354–360
  34. Sperduto PW, Kased N, Roberge D et al (2012) Effect of tumor subtype on survival and the graded prognostic assessment for patients with breast cancer and brain metastases. *Int J Radiat Oncol Biol Phys* 82:2111–2117
  35. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. *IEEE Expert* 13:18–28
  36. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36
  37. Alba AC, Agoritsas T, Walsh M et al (2017) Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 318:1377–1384
  38. Riley RD, Snell KI, Ensor J et al (2019) Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 38:1276–1296
  39. Hughes RT, Masters AH, McTyre ER et al (2019) Initial SRS for patients with 5 to 15 brain metastases: results of a multi-institutional experience. *Int J Radiat Oncol Biol Phys* 104:1091–1098
  40. Zhou M, Scott J, Chaudhury B et al (2018) Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches. *AJNR Am J Neuroradiol* 39:208–216
  41. Drevelegas A (2002) *Imaging Modalities in Brain Tumors*. In: Drevelegas A (ed) *Imaging of brain tumors with histological correlations*. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 11–25
  42. de Lange EE, Mugler JP 3rd, Bertolina JA, Gay SB, Janus CL, Brookeman JR (1991) Magnetization prepared rapid gradient-echo (MP-RAGE) MR imaging of the liver: comparison with spin-echo imaging. *Magn Reson Imaging* 9:469–476
  43. Park HJ, Lee SY, Rho MH, Kwon HJ, Kim MS, Chung EC (2015) The usefulness of the three-dimensional enhanced T1 high-resolution isotropic volume excitation MR in the evaluation of shoulder pathology: comparison with two-dimensional enhanced T1 fat saturation MR. *Br J Radiol* 88:20140830
  44. Chen BT, Jin T, Ye N et al (2020) Radiomic prediction of mutation status based on MR imaging of lung cancer brain metastases. *Magn Reson Imaging* 69:49–56
  45. Choi W, Oh JH, Riyahi S et al (2018) Radiomics analysis of pulmonary nodules in low-dose CT for early detection of lung cancer. *Med Phys* 45:1537–1549
  46. Tomaszewski MR, Gillies RJ (2021) The biological meaning of radiomic features. *Radiology* 298:505–516
  47. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14:749–762
  48. Horsman MR, Overgaard J (2016) The impact of hypoxia and its modification of the outcome of radiotherapy. *J Radiat Res* 57(Suppl 1):i90–i98
  49. Sun R, Limkin EJ, Vakalopoulou M et al (2018) A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* 19:1180–1191
  50. Zhang J, Jin J, Ai Y et al (2020) Computer tomography radiomics-based nomogram in the survival prediction for brain metastases from non-small cell lung cancer underwent whole brain radiotherapy. *Front Oncol* 10:610691
  51. Kniep HC, Madesta F, Schneider T et al (2019) Radiomics of Brain MRI: Utility in Prediction of Metastatic Tumor Type. *Radiology* 290:479–487