Radiation Oncology

RESEARCH

Open Access

Integrating peritumor and tumor CT radiomics features in predicting local control after SBRT in patients with pulmonary oligometastases

Yao Lu[1,2†], Yiqi Wang[1†], Yuxi Ding[3], Danni Chen[1,4], Wenguang He[5], Weixiang Zhong[6], Jing Yang[1,4], Senxiang Yan[1,4*], Ge Ren[2,7*] and Feng Zhao[1,4*]

Abstract

Purpose Local control prediction for patients with pulmonary oligometastases underwent stereotactic body radiotherapy (SBRT) is crucial for optimizing therapeutic strategies. This study aims to develop and validate a predictive radiomics model integrating both tumor-intrinsic and peritumoral features along with clinical factors to enhance local control prediction using a multi-center dataset.

Materials and methods We analyzed 223 tumors from 146 patients, which was divided into a training set ($n=165$) and an external validation set ($n=58$). Radiomic features from the gross tumor volume (GTV) and peritumoral regions (pGTV) representing the tumor microenvironment (TME) in CT images were extracted and combined with clinical factors to build a clinical outcome prediction model. Tumor response was classified into Favorable Response Group (FRG) and Unfavorable Response Group (URG) according to the 3-month and 1-year follow-up. Models were built using a Multilayer Perceptron (MLP) approach with SHAP analysis.

Results Model-G (with GTV features) and Model-P (with pGTV features) achieved a validation area under curve (AUC) of 0.806 and 0.708, respectively. Meanwhile, Model-GP (with GTV and pGTV features) demonstrated an improved performance with a validation AUC of 0.851, reflecting the added value of peritumoral features. The Model-GPC, which incorporated GTV, pGTV, and clinical features, achieved a best validation AUC of 0.902, demonstrating the model's ability to robustly integrate clinical and radiomic data for accurate local control prediction.

Conclusion The Model-GPC, integrating clinical and radiomic features, accurately predicts post-SBRT local control in pulmonary oligometastases. Incorporating peritumoral features and SHAP analysis enhances prediction accuracy, offering insights to optimize SBRT strategies.

†Yao Lu, and Yiqi Wang contributed equally to this study.

*Correspondence:
Senxiang Yan
yansenxiang@zju.edu.cn
Ge Ren
gary-ge.ren@polyu.edu.hk
Feng Zhao
zju_zhaofeng@zju.edu.cn

Full list of author information is available at the end of the article

Keywords Pulmonary oligometastases, Stereotactic body radiotherapy (SBRT), Radiomics, Peritumoral features, Local control prediction

Background

Pulmonary metastases are a significant clinical challenge as the lungs are one of the most common sites for distant metastasis in various cancers [1–3]. The occurrence of pulmonary metastases often signals disease progression, requiring aggressive and multidisciplinary treatment approaches [4, 5]. Despite advancements in systemic therapies, such as chemotherapy and targeted therapies, local control of pulmonary metastatic lesions remains a critical component of comprehensive cancer management, particularly in patients with limited metastatic burden commonly referred to as oligometastases [6, 7]. The concept of oligometastases, introduced by Hellman and Weichselbaum in 1995, suggests that patients with a small number of metastatic lesions might still benefit from aggressive local therapies, such as surgical resection or radiotherapy, with the potential for long-term survival or even cure [8]. stereotactic body radiotherapy (SBRT) has emerged as an effective, minimally invasive treatment modality for lung oligometastases, particularly for patients who are unsuitable for surgery due to comorbid conditions or other risk factors [9, 10]. The precise and high-dose radiation delivery of SBRT increases tumor control with minimal damage to surrounding healthy lung tissue [11, 12]. Studies have demonstrated that SBRT achieved local control rates comparable to surgical resection in patients with pulmonary oligometastases [13, 14].

However, despite its effectiveness, predicting which patients will benefit most from SBRT remains a major clinical challenge. Current methods for assessing treatment response rely primarily on post-treatment imaging and follow-up, providing limited guidance in the pre-treatment phase. This inability to predict individual patient local control accurately hinders clinicians' ability to tailor SBRT plans, potentially leading to suboptimal therapeutic decisions and outcomes.

Radiomics, a field that leverages advanced computational techniques to extract and analyze quantitative imaging features, offers a novel approach to address this issue [15]. By transforming medical imaging data into high-dimensional features, radiomics can provide detailed information on tumor characteristics, such as shape, texture, and intensity, that are often invisible to the human eye [15, 16]. This detailed imaging data, when combined with machine learning algorithms, holds the potential to develop predictive models that can forecast treatment response, thereby personalizing and optimizing therapeutic strategies for patients with pulmonary metastases [17]. While radiomics has demonstrated potential across several types of cancer, including breast, lung, and pancreatic cancers [18–20], its application in predicting SBRT local control for pulmonary metastases remains underexplored. Most of the existing studies in this area are limited by small sample sizes and single-center designs, which restrict the generalizability of their findings. For example, the largest study to date on radiomics in pulmonary oligometastases involved only 69 patients and 85 metastatic lesions [21]. Such small-scale studies, while valuable, are insufficient for developing robust predictive models that can be reliably applied across diverse patient populations. Additionally, these studies have largely focused on tumor-intrinsic features without considering the tumor's interaction with tumor microenvironment (TME). This narrow focus on the tumor itself may overlook critical information from the TME, which is known to play a significant role in influencing treatment response, particularly in metastatic diseases like pulmonary oligometastases [22–24]. The TME is a dynamic and complex area surrounding the tumor, consisting of stromal cells, immune cells, vasculature, and extracellular matrix. It plays a critical role in regulating tumor progression and treatment response through interactions between tumor cells and their microenvironment. Emerging evidence suggests that radiomic features derived from the peritumoral region may provide complementary information beyond tumor-intrinsic characteristics, potentially capturing the influence of the TME on therapeutic local control [22–24].

In this study, we aim to develop a more accurate and robust predictive model for post-SBRT local control in pulmonary metastases by integrating radiomics, clinical data, and TME features. We will validate this model using a multi-center dataset, which allows us to combine diverse and representative data to enhance the generalizability of our findings. By incorporating radiomic features from both the gross tumor volume (GTV) and the peritumoral region, which reflects the TME, we expect to capture critical tumor-environment interactions that influence treatment response. This approach, which integrates both tumor-intrinsic and peritumoral features with clinical data, will offer a more personalized and accurate prediction of post-SBRT local control for patients with pulmonary metastases.

Materials and methods

Patient cohort

This retrospective study analyzed patients who received pulmonary oligometastases SBRT at three radiation therapy centers (Qingchun, Yuhang, and Zhijiang) affiliated with the First Affiliated Hospital of Zhejiang

University School of Medicine, between 2018 and 2023. Ethical approval was obtained from our institutional review boards (No. 2024 – 0513). The patient selection flow chart is shown in Fig. 1. The inclusion criteria consisted of patients with pulmonary oligometastases, where all the number of pulmonary metastatic lesions did not exceed five. Additionally, systemic disease was required to be in a stable condition. Exclusion criteria included patients with a follow-up period of less than 12 months, those with pleural effusion or severe pneumonia, or those lacking high quality imaging. A total of 223 tumors from 146 patients were included in the study. 165 tumors from the Qingchun center were assigned as the training set, while 58 tumors from centers two (Yuhang) and three (Zhijiang) were the external validation set. All patients received 40–60 Gy of radiation, delivered in no more than 8 fractions. At our institutions, breath-hold (BH) is the standard protocol during both CT simulation and treatment delivery for thoracic radiotherapy. 4D-CT is selectively used based on clinical needs. For this study, only patients who underwent BH-based CT simulation were included to ensure consistency and minimize respiratory motion artifacts.

Clinical characteristics and response evaluation

The clinical characteristics of the included cases involved various data such as age, gender, radiation dose, fractionation, pathological classification, and tumor location/position.

In this study, tumor response was classified according to the Response Evaluation Criteria in Solid Tumors (RECIST 1.1) guidelines into complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD) [25]. The specific definitions were as follows: PR: A reduction of at least 30% in the longest diameter of the treated lesion compared with the baseline. CR: The complete disappearance of treated lesions with no abnormal residue on imaging. SD: The magnitude of change in the lesion was not sufficient to meet the criteria for PR, while did not meet the criteria for PD, typically the change of less than 30% in either direction. PD: An increase in the longest diameter of the treated lesion by at least 30% or the appearance of new lesions within the treated area. Tumor responses were evaluated using imaging at both 3 months and 1 year post-SBRT. For this study, patients were classified into two response groups: the Favorable Response Group (FRG) and the Unfavorable Response Group (URG). Patients who achieved CR or PR at both 3 months and 1 year were classified into the FRG, reflecting sustained positive local control. Conversely, patients who showed SD or PD at either 3 months or 1 year were classified into the URG, indicating less favorable local control. This classification method allows for a more comprehensive assessment of treatment effectiveness, considering both short-term and long-term local control.

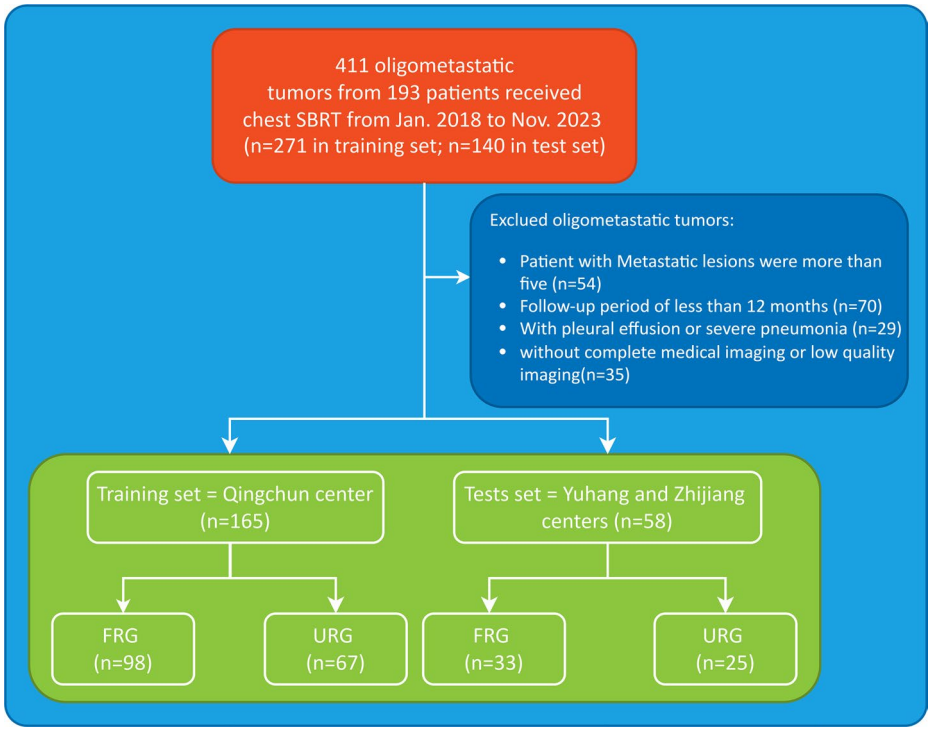

Fig. 1 Flow diagram of the study cohort. FRG, favorable response group; URG unfavorable response group

Tumor segmentation and morphological operations

Tumor regions of interest (ROIs), GTV, were manually delineated for each patient on the planning CT scans using Eclipse (Varian Medical Systems, USA) or Monaco (Elekta, Sweden) software by two experienced radiation oncologists (D.N. C. and F. Z.).

To better characterize the TME, morphological operations were applied to segment the peritumoral tissue. First, the GTV was manually outlined (Fig. 2A). Then, the outer boundary of the GTV was contracted inward by 2 mm to create a smaller region, which we defined as the core GTV (cGTV) (Fig. 2B). This core GTV represents the inner part of the tumor, closely reflecting its direct characteristics. Next, the outer boundary of the GTV was expanded outward by 2 mm to define the extended GTV (eGTV) (Fig. 2C). The region between the contracted cGTV and the expanded eGTV was defined as the peripheric GTV (pGTV) (Fig. 2D). This area is particularly important as it reflects the tumor microenvironment and may provide crucial insights into treatment response. This morphological process enabled the extraction of radiomics features from both the GTV and the pGTV, helping to capture the heterogeneity of both the tumor core and its surrounding microenvironment, which are believed to influence local control. This 2 mm contraction and 2 mm extension approach was informed by histopathological observations of TME depths from data collected at our hospital, which revealed a significant extension of the TME, with depths typically less than 4000 μm (see Table A.1 in the supplementary material). Representative pathological images from choriocarcinoma, breast cancer, colorectal cancer, and hepatocellular carcinoma, obtained from our hospital, were used to validate this approach and are shown in Fig. 3. Additionally, our methodology is supported by several studies that have examined both TME contraction and expansion, with most research focusing on a typical extension of 3–5 mm adjacent to the tumor. However, we have adopted a 2 mm contraction and 2 mm extension approach, which we believe more accurately reflects the dynamic nature of the TME, as it may appear as part of the tumor itself in imaging [26–28].

Radiomics feature extraction

Radiomics features were extracted separately from the GTV and pGTV using the PyRadiomics package (Version 3.0.1), covering shape, size, first-order statistics, texture, and filter-derived features. CT signal intensity was normalized, and images were discretized using a fixed bin width of 25. Voxel sizes were resampled to $1 \times 1 \times 1$ mm^3. A total of 1,319 features were extracted from each CT image (Table B.1). The slice thickness of the CT images was 1.25 mm.

Feature selection and model construction

The primary objective of the modeling process was to perform binary classification of tumors into FRG versus URG. Multilayer Perceptron (MLP) was employed to facilitated the creation of four distinct predictive models (Model-G, Model-P, Model-GP, Model-GPC). Each MLP model consisted of two hidden layers, respectively, using "ReLU" as the activation function. Dropout layers were incorporated after each hidden layer to mitigate overfitting. The models were trained using the Adam optimizer with a batch size of 32. To address class imbalance, a weighted binary cross-entropy loss (BCEWithLogitsLoss) was applied, with the positive class weight adjusted according to the inverse class frequency. To enhance model robustness and ensure stable performance, we applied stratified bootstrap resampling during the model training phase. Specifically, the training dataset underwent 1000 bootstrap iterations with replacement, maintaining the original class proportions in each resampling iteration (stratified bootstrap). This process aimed to prevent potential biases arising from random data splits and further validated the reliability of the training process.

Since CT scanner models have been demonstrated to be a significant factor affecting radiomics model reproducibility, statistical tests (Spearman correlation, point-biserial correlation, and chi-square test) were conducted to confirm the independence of scanner type from predictive labels. Additionally, the feature selection process incorporated multiple stages explicitly designed to minimize scanner-induced variability, beginning with a feature-level stability analysis aimed at reducing scanner effects on radiomic features. All radiomic features were first standardized using z-score normalization. Scanner-sensitive features were identified based on a combination of three criteria: [1] Spearman correlation coefficient with scanner type > 0.3 or < -0.3 [2], Mann-Whitney U test p-value < 0.05, and [3] false discovery rate (FDR)-adjusted q-value < 0.05. Features meeting at least one of these three stringent criteria were excluded from further analysis. Kernel density estimation (KDE) was performed to evaluate variance consistency across different scanners.

As the next stage, a two-step feature selection process was deployed separately on GTV's and pGTV's radiomics features datasets (Fig. A.1). Non-significant features were excluded through the Mann-Whitney U test ($p < 0.05$). Features exhibiting low correlation with the target label were removed by using the Spearman correlation coefficient (threshold of 0.8). The threshold of 0.8 was chosen based on prior radiomics studies identifying it as a widely accepted cut-off for strong correlation [29, 30]. Compared to 0.9, it more strictly eliminates highly correlated features, thereby reducing redundancy and multicollinearity more effectively while still retaining

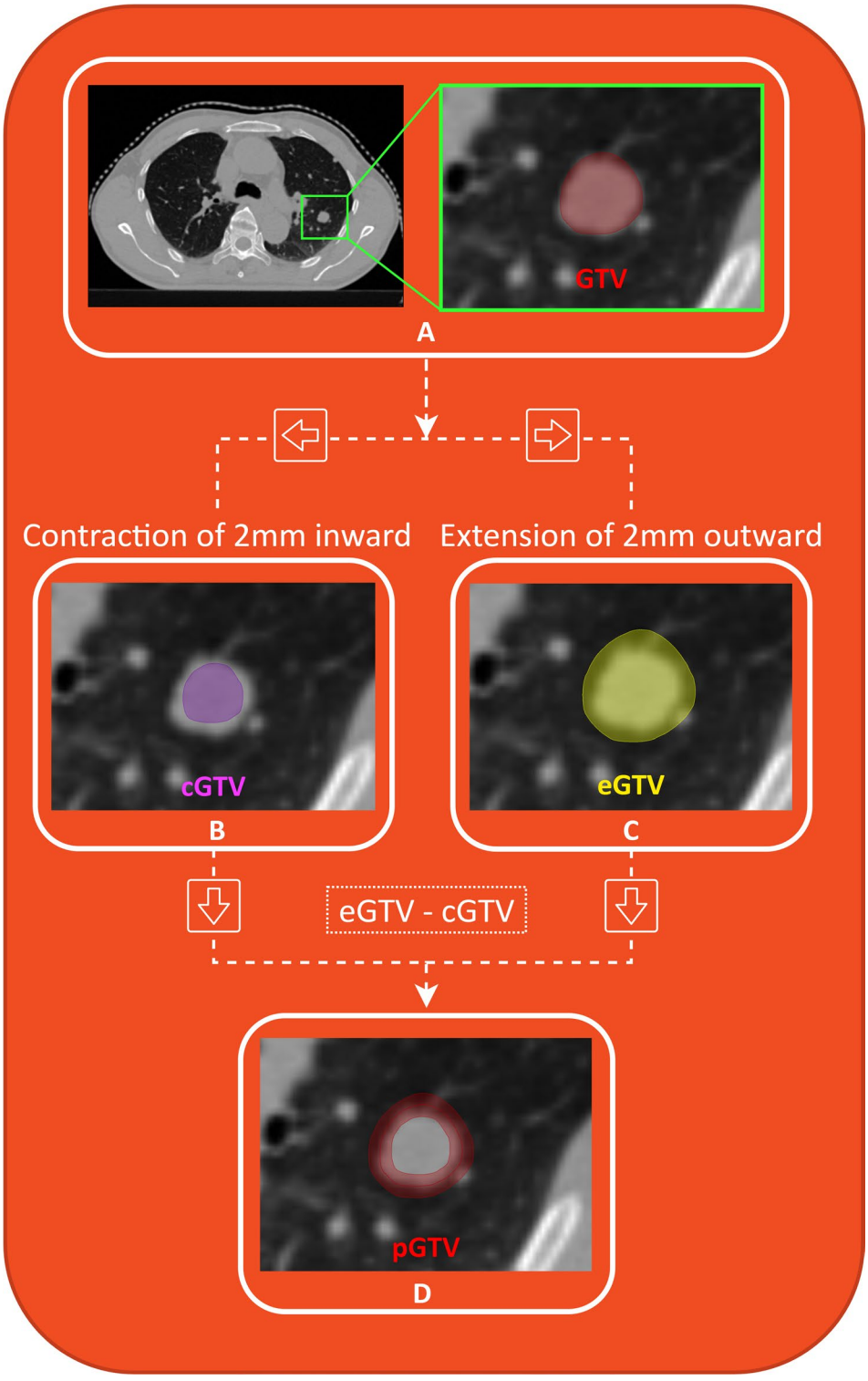

Fig. 2 Tumor segmentation and morphological operations. (**A**) Gross tumor volume (GTV); (**B**) Contraction of 2 mm inward from the outer boundary of the GTV to create the core GTV (cGTV); (**C**) Extension of 2 mm outward from the outer boundary of the GTV to create the extended GTV (eGTV); (**D**) Peripheric GTV (pGTV), representing the region between the eGTV and the cGTV

Table 1 Summary of model performance metrics

Model	Model-G		Model-*P*		Model-GP		Model-GPC	
Dataset	Training	Validation	Training	Validation	Training	Validation	Training	Validation
AUC	0.887	0.806	0.874	0.708	0.892	0.851	0.945	0.902
(95%CI)	(0.844–0.927)	(0.689–0.902)	(0.830–0.915)	(0.589–0.817)	(0.842–0.938)	(0.754–0.936)	(0.913–0.970)	(0.836–0.959)
Accuracy	0.829	0.7959	0.818	0.694	0.879	0.843	0.898	0.858
(95%CI)	(0.776–0.879)	(0.707–0.879)	(0.758–0.867)	(0.5690–0.810)	(0.842–0.921)	(0.776–0.914)	(0.861–0.933)	(0.793–0.931)
Precision	0.829	0.782	0.856	0.697	0.849	0.807	0.872	0.836
(95%CI)	(0.737–0.910)	(0.681–0.903)	(0.713–0.929)	(0.569–0.846)	(0.803–0.895)	(0.727–0.892)	(0.825–0.922)	(0.750–0.938)
Specificity	0.712	0.652	0.781	0.481	0.746	0.619	0.790	0.747
(95%CI)	(0.493–0.881)	(0.400–0.880)	(0.418–0.910)	(0.0174–0.800)	(0.657–0.836)	(0.520–0.840)	(0.702–0.881)	(0.590–0.920)
Sensitivity	0.909	0.905	0.843	0.882	0.970	0.957	0.972	0.942
(95%CI)	(0.827–0.990)	(0.788–1.00)	(0.776–0.969)	(0.727–1.00)	(0.939–0.9898)	(0.879–1.00)	(0.929–1.00)	(0.849–1.00)
F1 Score	0.867	0.839	0.849	0.778	0.905	0.877	0.919	0.886

essential variability for model generalizability. Model-G and Model-P were individually used for analyzing the selected radiomics features from different spatial regions (GTV and pGTV). To prioritize those features contributing most to the model predictions, SHapley Additive exPlanations (SHAP) analysis was applied as the second part of the feature selection process, guiding the fusion of both models (Model-G and Model-P) as the combined Model-GP. The top three features with the highest average absolute SHAP values for each model were selected, resulting in a total of six features being chosen as the final feature set for training Model-GP. The comprehensive Model-GPC includes the combination of features from Model-GP with clinical factors selected using Mann-Whitney U and chi-square tests.

Model evaluation and statistical analyses

Model performance was comprehensively evaluated using multiple metrics, including the area under the receiver operating characteristic curve (ROC-AUC) with 95% confidence intervals (CIs), accuracy, F1 score, specificity, sensitivity (recall), and precision. In all classification-related evaluations, the FRG was defined as the positive class, and the URG as the negative class. These metrics assessed the models' classification capability from various aspects, covering overall accuracy and predictive precision for positive (FRG) and negative (URG) classes.

To validate the generalizability and robustness of model performance, A stratified bootstrap procedure with 1000 resamples validation procedure was implemented. For each bootstrap iteration, the evaluation metrics mentioned above were calculated. The mean values and corresponding 95% confidence intervals of these metrics were reported, ensuring the observed high performance was not a consequence of random favorable data splits but indicative of stable predictive capability. The DeLong test was used to assess the differences in ROC-AUCs of different models. To contextualize the predictive performance of our proposed models, we additionally collected performance metrics reported in representative published radiomics-based outcome prediction studies. These external benchmark metrics were summarized to facilitate a comparative interpretation of our model's results.

Additionally, SHAP analysis was incorporated to gain deeper insights into the predictive mechanisms of the models. The contribution was quantified of each feature to the model's predictions, thereby enhancing the interpretability of the models.

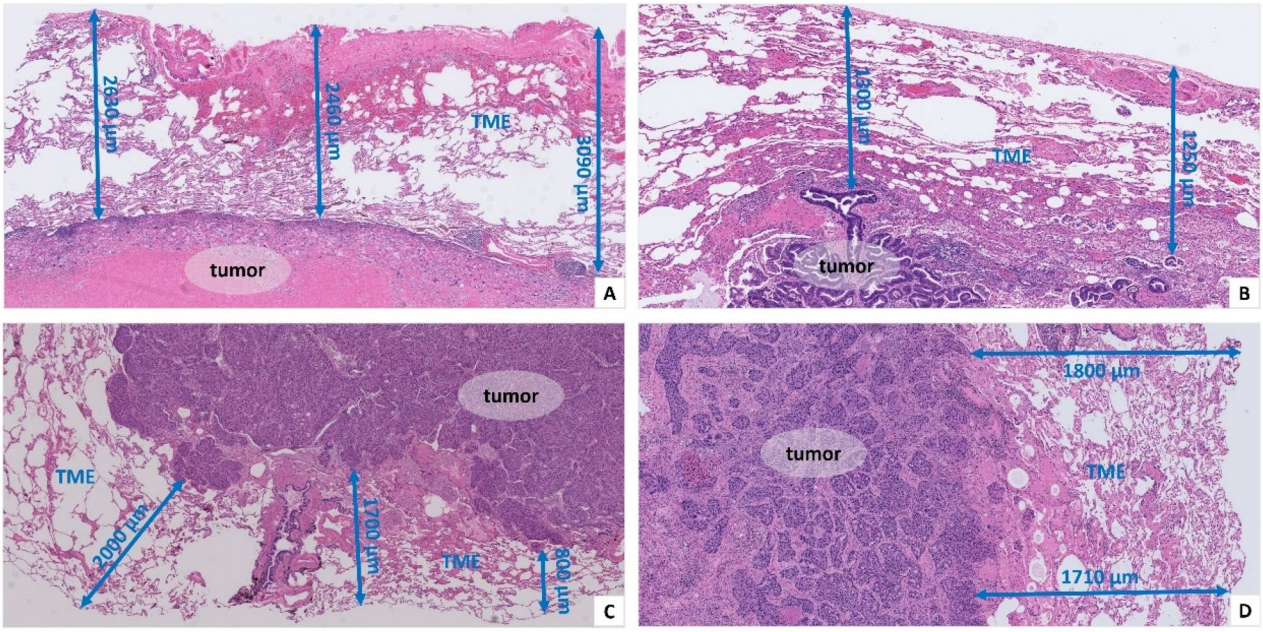

Fig. 3 Representative pathological images of lung metastases from choriocarcinoma (**A**), colorectal cancer (**B**), hepatocellular carcinoma (HCC) (**C**), and breast cancer (**D**). Each image highlights the tumor (denoted by the black "tumor") and the surrounding tumor microenvironment (TME) (denoted by the blue "TME"), with the depth of the TME region measured and indicated in micrometers (μm) (blue line). These measurements reveal a significant extension of the TME surrounding each tumor, with depths ranging from 800 μm to 3090 μm

Results

Patient characteristics

A total of 146 patients (95 males and 51 females) were recruited between 2018 and 2023, with a median age of 61 years (range, 24 to 81 years). Since some patients received radiotherapy for more than one tumor, the total number of analyzed tumors was 223 ($n = 165$ in training set; $n = 58$ in external validation set). In the following follow-up, 92 tumors achieved CR and 39 as PR; these 131 metastases were considered as the FRG. 83 tumors were categorized as SD, and 9 were identified as PD; these 92 metastases were considered as the URG. The most frequently used dose-fractionation regimens in the SBRT cohort were 56 Gy in seven fractions (29.0% of patients, $n = 64$; $BED_{10} = 100.8$), 50 Gy in five fractions (28.5% of patients, $n = 63$; $BED_{10} = 100.0$), 60 Gy in ten fractions (10.4% of patients, $n = 23$; $BED_{10} = 96$), 50 Gy in four fractions (5.43% of patients, $n = 12$; $BED_{10} = 112.5$), and 56 Gy in eight fractions (3.17% of patients, $n = 7$; $BED_{10} = 95.2$). The clinical characteristics of the participants are summarized in Table C.1.

Feature selection and model construction

Radiomic features were selected from both the GTV and the pGTV to build predictive models for post-SBRT local control. After performing the feature selection process, 23 features and 22 features were identified and used to construct Model-G and Model-P, respectively. The KDE plots (Supplementary Fig.B.1) illustrated highly similar variance distributions across scanners, indicating negligible inter-scanner variability. Reproducibility assessments demonstrated no significant association between scanner type and predictive labels (Spearman $\rho = 0.037$, $p = 0.578$; point-biserial $r = 0.037$, $p = 0.578$; chi-square test, $p = 0.678$), confirming scanner independence from classification outcomes.

Figure 4A-D presents the SHAP analysis results of the features following the completion of model construction. The bar plot illustrates the global importance of features, ranked by the absolute SHAP values, while the summary plot displays the distribution of feature SHAP values across all samples. Each point in the summary plot represents a sample, with red and blue indicating the magnitude of the feature values. Finally, we individually selected the top three features contributing most to Model-G and Model-P to form the feature set for Model-GP. For Model-GPC, the feature set was expanded by adding the only significant clinical characteristic, curative dose, to the Model-GP final feature set.

In the SHAP analysis, Model-G shows that the texture features of the tumor region (GTV) are the primary drivers of model predictions (Fig. 4A). In Model-P, the texture features of peritumoral region (pGTV) reflecting the heterogeneity of the TME areas begin to make important contributions to the model's performance (Fig. 4B). By integrating features from Model-G and Model-P, Model-GP further demonstrates how the tumor and peritumoral features synergistically affect the combined model

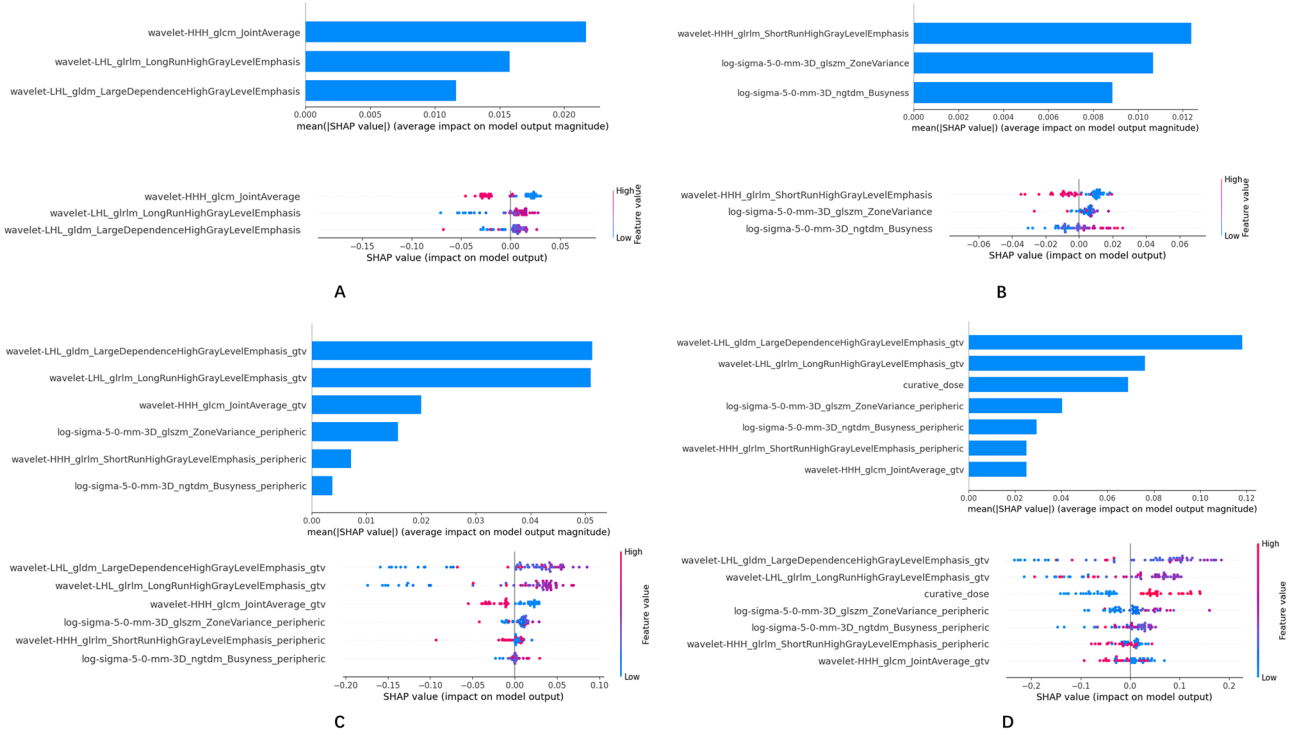

Fig. 4 SHAP value bar plots and summary plots for models. The figure shows the global feature importance ranked by absolute SHAP values, as well as the distribution of SHAP values across all samples: A: Model-G (GTV), **B**: Model-P (pGTV), C: Model-GP (GTV and pGTV), and D: Model-GPC (GTV, pGTV, and clinical parameters)

(Fig. 4C). In the final hybrid Model-GPC (Fig. 4D), the addition of clinical parameters demonstrates the synergistic effect of dose and imaging features. Although wavelet-LHL_gldm_LargeDependenceHighGrayLevelEmphasis_gtv (SHAP value = 0.1181) and wavelet-LHL_glrlm_LongRunHighGrayLevelEmphasis_gtv (SHAP value = 0.0762) remain the most important features, the newly added curative_dose feature becomes another significant variable, while simultaneously altering the synergistic effect among the other features.

Performance of predictive models

The predictive performance of models was assessed using various metrics, including AUC, Accuracy, Precision, Specificity, Sensitivity, and F1 Score on both the training and validation datasets (Table 1). For Model-G, the AUC on the training set and external validation set were 0.887 (95% CI, 0.844 to 0.927) and 0.806 (95% CI, 0.689 to 0.902), respectively (Table 1). The assessing metric also suggests that Model-G is more willing to predict the samples to the FRG. Meanwhile, Model-P presents AUCs of 0.874 (95% CI, 0.830 to 0.915) and 0.708 (95% CI, 0.589 to 0.817) for the training set and external validation set, respectively (Table 1). Model-GP outperforms both Model-G and Model-P ($p < 0.05$), with AUCs of 0.892 (95% CI, 0.842 to 0.938) and 0.851 (95% CI, 0.754 to 0.936) on the training and external validation sets, respectively, demonstrating stronger generalization ability (Fig. 5; Table 1). Results indicate that Model-GP has a very strong ability to detect positive class samples (FRG), with almost no missed detections (sensitivity close to 0.96), but its specificity still lags behind its sensitivity.

Among all the models, Model-GPC significantly outperforms other models (all $p < 0.05$), with AUCs of 0.945 (95% CI, 0.913 to 0.970) and 0.902 (95% CI, 0.836 to 0.959) on the training and external validation sets, respectively (Fig. 5 and Table D.1). In the external validation set, Model-GPC achieved a specificity of 0.747, an accuracy of 0.858, and an F1 score of 0.886. These results suggest that Model-GPC offers a strong balance between sensitivity and specificity, making it a promising tool for predicting post-SBRT local control in pulmonary metastases. To further contextualize these results, we compared our model's performance metrics with representative published radiomics-based predictive models. As summarized in Table 2, our model achieved higher AUC, sensitivity, and F1 score than previous models.

Discussion

To achieve precise radiotherapy for lung cancer, clinicians have continuously pursued improved local control. However, as SBRT has been increasingly used in patients with oligometastatic disease, the statistically observed pCR rate is only 60%, far below previous expectations

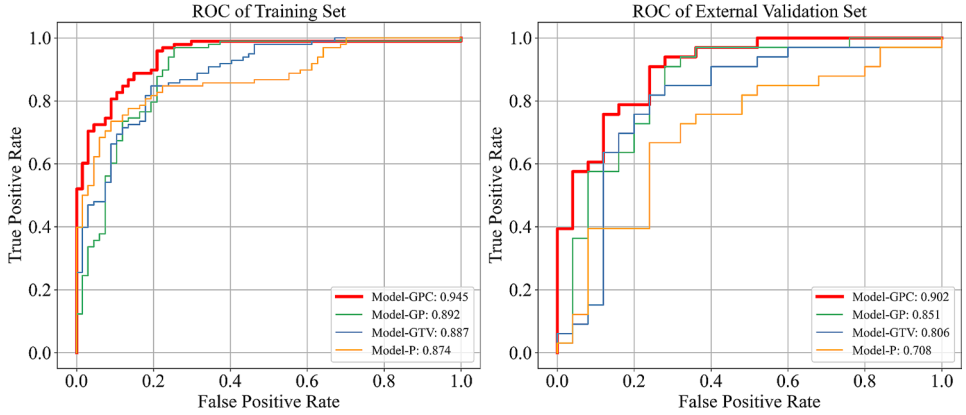

Fig. 5 ROC curve comparison of models on training and external validation sets

Table 2 Summary of published radiomics-based predictive models

	Algorithm	Method	AUC (95% CI)	Accuracy (95%CI)	Specificity (95%CI)	Sensitivity (95%CI)	F1 score
This Study	Multilayer Perceptron	CT Radiomics (GTV + pGTV)	0.902 (0.836–0.959)	0.858 (0.793–0.931)	0.747 (0.590–0.920)	0.942 (0.849–1.00)	0.886
Cheung, B. M. F., et al. (2021).	Support Vector Machine	CT Radiomics (GTV)	0.857 (0.789–0.915)	N/A	N/A	N/A	N/A
Avanzo, M., et al. (2021).	Convolutional Neural Network	Radiomics (CT + BED)	0.843 (0.785–0.902)	0.819 (0.763–0.863)	0.605 (0.455–0.727)	0.900 (0.845–0.948)	N/A
Cilla, S., et al. (2023).	Classification and Regression Tree	CT Radiomics (GTV)	0.753 (0.675–0.836)	0.792 (0.756–0.827)	N/A	0.754 (0.692–0.815)	0.694
Cilla, S., et al. (2023).	Logistic Regression	CT Radiomics (GTV)	0.707 (0.633–0.817)	0.644 (0.610–0.679)	N/A	0.635 (0.606–0.665)	0.638

[31]. In Hamamoto's study, the local control rate was poorer for metastatic and more advanced tumors [32]. Therefore, predicting the efficacy of SBRT for lung cancer has become crucial, positively impacting medical decision-making and patient management [33]. In an era where clinical features and radiomics are extensively used to predict treatment outcomes, the performance of predictive models has become increasingly crucial [34]. However, relying solely on clinical features and traditional radiomics may not fully capture the impact of tumor heterogeneity on treatment response. Consequently, incorporating additional effective information beyond the core tumor region has become a focal point of research efforts.

The aim of this study was to develop a robust radiomics model for predicting post-SBRT local control in pulmonary metastases, integrating both tumor intrinsic and peritumoral radiomic features along with clinical data. Therefore, in addition to addressing feature-level variability associated with different scanners, we further assessed the consistency of radiomic feature variance across scanner types. The overall variance distributions under different scanners were then compared using KDE. The resulting KDE curves showed substantial overlap, indicating no major variance shift attributable to scanner differences. These findings support the stability of the selected features and suggest that inter-scanner variability at the variance level had minimal influence on model performance. To further verify that scanner type did not confound the classification outcomes, we examined the association between scanner types and the ground-truth labels using multiple statistical tests. The Spearman correlation coefficient between scanner type and label was low ($\rho = 0.037$), and the corresponding p-value was not significant ($p = 0.578$), indicating no meaningful monotonic relationship. Similar results were observed using the point-biserial correlation ($r = 0.037$, $p = 0.578$). In addition, the chi-square test based on the contingency table of labels and scanner types also yielded a non-significant result ($p = 0.678$). These findings suggest that the label distribution was not biased by scanner type, further supporting the reliability and generalizability of our model across different imaging sources.

In this study, we evaluated the performance of four different MLP models in predicting radiotherapy local control, with a particular focus on the novel radiomics model that incorporates intratumoral and peritumoral radiomic features as well as clinical factors. The results demonstrate that incorporating features from the peritumoral region significantly improves the model's predictive performance, highlighting the pivotal role of the TME in local control prediction. The TME is a complex

and dynamic region of stromal cells surrounding cancer cells. The interactions between various cells within the TME and with cancer cells play a critical role in tumorigenesis and progression [24, 35]. These interactions can regulate cancer cell invasiveness, promote tumor growth, and drive metastasis [24]. The peripheric GTV radiomic features, observed from the surrounding tumor region, also highlight the morphological, textural, and metabolic information of this area. Our results suggest that these quantified radiomic features are possibly related to the TME and contribute positively to the predictive performance of the models.

The multilayer perceptron developed in this study achieved a higher AUC, accuracy, sensitivity, specificity, and F1 score among the listed approaches. Reported AUC values for prior models, including support vector machines, convolutional neural networks, and logistic regression, ranged from 0.707 to 0.857 [36–38]. While differences in study design, feature selection strategies, and patient populations should be considered, the higher performance metrics observed in our study suggest a potential advantage of the proposed model in radiomics-based outcome prediction.

While complex algorithmic tools, such as support vector machines, random forest, and neural networks, have been widely used to predict local control and assess patients with promising performance [37–40], the combination of numerous high-dimensional features has exacerbated the model's complexity, resulting in a 'black-box' nature [41]. This 'black-box' nature makes it difficult to intuitively understand the contributions of different features and their combinations to the final prediction [41]. To mitigate this 'black-box' issue, an increasing number of radiomics studies have employed the SHAP method to provide interpretability to models, enabling the observation of the specific impact of features on model outputs [42–44].

In this study, SHAP analysis revealed the varying impacts of multiple features on model predictions. These features not only reflect the texture characteristics of the tumor itself and peripheral regions but also highlight the significant role of treatment dose in lesion response. According to the SHAP analysis results, we found that in the hybrid models, GTV-related features remain the most significant contributors to the model's predictions.

GTV-related features reflect the tissue density and texture characteristics of the tumor core. For example, the feature "wavelet-LHL_gldm_LargeDependenceHighGrayLevelEmphasis_gtv" represents the dependence and uniformity of high gray levels, with higher values potentially indicating a higher or more compact cellular density in the tumor core [45, 46]. In pulmonary oligometastatic lesions, such features often suggest a higher degree of malignancy or greater radiosensitivity [47]. On the other hand, "wavelet-HHH_glcm_JointAverage_gtv" further describes the texture consistency of the tumor core region through the joint average gray level, with higher values potentially corresponding to denser tumor tissue [48–50]. Our SHAP results indicate that higher "JointAverage" values were potentially associated with greater challenges in achieving effective local control.

Secondly, the texture features of the tumor peripheral region provide potential information for understanding the tumor microenvironment. "Wavelet-HHH_glrlm_ShortRunHighGrayLevelEmphasis_peripheric" provides the information about continuous pixels on the direction of the same grayscale run the spatial distribution [51]. This index can reflect the signal intensity and texture changes of the tumor margin. "log-sigma-5-0-mm-3D_glszm_ZoneVariance_peripheric" was used to quantify the variation of the consistency area in the periphery of the tumor [51]. Meanwhile, "log-sigma-5-0-mm-3D_ngtdm_Busyness_peripheric" reflects the rapid gray-level changes between the central pixel or voxel and its neighbors [51]. All these features reflect the texture heterogeneity or invasiveness of the peritumoral region and complementary to each other. Research indicates that heterogeneity of tumor boundary could be related to the biological behavior, infiltration degree, or immune infiltration status of metastatic tumors [52, 53]. We hold the opinion that GTV-related features, covering the imaging information of the entire lesion, provide an average characterization of the tumor. This "averaging" effect may obscure the specificality of texture or gray-level properties to the peritumoral region, thereby weakening the distinctive heterogeneity of the peritumoral region. Therefore, by incorporating pGTV-related features into the model, Model-GP and Model-GPC enable to retain the overall information of the tumor while conducting separate analysis of the peritumoral region, thereby enhancing the classification ability of the Model.

It's worth noting that the contribution of curative dose to the model highlights its critical role in achieving effective local control after SBRT. Patients receiving $BED_{10} \geq 100$ Gy typically have a higher likelihood of achieving the treatment goals. In contrast, treatment with $BED_{10} < 100$ Gy, influenced by other clinical factors (such as lesion complexity and patient's physical tolerance), may limit the effectiveness of the treatment. This result suggests that, during treatment planning, more attention should be given to the rational planning of the dose. In particular, when the patient's condition allows, prioritizing the use of curative doses should be considered to maximize the treatment success rate and improve patient prognosis.

Despite the promising results, several limitations remain. Although our study involves a larger cohort compared to previous studies, the sample size is still

moderate, which may limit the generalizability of the model to more diverse patient populations. Additionally, our study primarily relied on radiomic features derived from CT images, which may not fully capture all tumor characteristics. Future studies could incorporate other imaging modalities, such as MRI or PET, to provide a more comprehensive analysis of the tumor and its microenvironment. While the current methods proved effective in mitigating scanner-related biases, the integration of deep learning-based domain adaptation techniques, such as CycleGAN, holds promise for achieving even greater robustness in multi-institutional applications. In addition, the lack of test–retest imaging is acknowledged as a limitation and will be addressed in future work through repeated scanning or additional harmonization strategies. Finally, while SHAP analysis provides valuable insights into the interpretability of our model, the "black-box" nature of the MLP still poses challenges. Future research could focus on developing more transparent models or incorporating additional interpretability techniques to improve clinical applicability and decision-making support of predictive models.

Conclusions

In conclusion, our study demonstrates the value of integrating radiomics, clinical data, and tumor microenvironment features to predict post-SBRT local control in pulmonary metastases. The inclusion of the peritumoral region significantly improved model performance, offering a more personalized approach to treatment planning. As SBRT continues to gain popularity for treating pulmonary metastases, predictive models like Model-GPC could play a critical role in improving patient selection and treatment optimization.

Abbreviations

AUC	Area under curve
ROC-AUC	Area under the receiver operating characteristic curve
CIs	Confidence intervals
CR	Complete response
cGTV	Core gross tumor volume
eGTV	Extended gross tumor volume
FDR	False discovery rate
FRG	Favorable Response Group
GTV	Gross tumor volume
KDE	Kernel density estimation
MLP	Multilayer Perceptron
PD	Progressive disease
PR	Partial response
pGTV	Peritumoral regions
RECIST	Response Evaluation Criteria in Solid Tumors
ROIs	Tumor regions of interest
SBRT	Stereotactic body radiotherapy
SD	Stable disease
SHAP	SHapley Additive exPlanations
TME	Tumor microenvironment
URG	Unfavorable Response Group

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13014-025-02712-w.

Supplementary Material 1

Author contributions

Y.L. and Y.W. contributed equally to this work. Y.L. and Y.W. conducted the study design, data collection, analysis, and interpretation, and wrote the main manuscript text. Y.D. and D.C. were responsible for image data acquisition and preprocessing. W.H. performed image analysis and selection. J.Y. contributed to data organization, feature extraction, and methodological design. W.Z. collected, analyzed, and measured pathological images. S.Y., G.R., and F.Z. supervised the study and provided critical revisions. All authors reviewed and approved the final manuscript.

Funding

This study has been supported by the National Natural Science Foundation of China (82171890 & 81701683).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval

Institutional Review Board approval was obtained. The study was approved by the Clinical Research Ethics Committee of the First Affiliated Hospital, Zhejiang University School of Medicine (No. 2024 – 0513).

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Radiation Oncology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310003, P.R. China
[2]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong KongKowloon, China
[3]School of Basic Medical Sciences and Forensic Medicine, Hangzhou Medical College, Hangzhou, Zhejiang 310053, P.R. China
[4]Cancer Center, Zhejiang University, Hangzhou, Zhejiang 310014, P.R. China
[5]Department of Radiology, The First Affiliated Hospital, Zhejiang University School of Medicine, No. 79 Qingchun Road, Hangzhou, Zhejiang 310003, P.R. China
[6]Department of Pathology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310003, P.R. China
[7]Department of Clinical Oncology, The University of Hong Kong-Shenzhen Hospital, Shenzhen, P.R. China

Received: 13 May 2025 / Accepted: 11 August 2025
Published online: 14 August 2025

References

1. Mangiameli G, Cioffi U, Alloisio M, Testori A. Lung metastases: current surgical indications and new perspectives. Front Surg. 2022;9:884915.
2. Pastorino U. Lung metastasectomy: why, when, how. Crit Rev Oncol Hematol. 1997;26(3):137–45.
3. Xie S, Wu Z, Qi Y, Wu B, Zhu X. The metastasizing mechanisms of lung cancer: recent advances and therapeutic challenges. Biomed Pharmacother. 2021;138:111450.

4. Comacchio GM, Melan L, Zambello G, Mammana M, Rea F. Lung metastases from hepatocellular carcinoma: multidisciplinary approach—narrative review. AME Surg J. 2021;2.
5. David EA, Clark JM, Cooke DT, Melnikow J, Kelly K, Canter RJ. The role of thoracic surgery in the therapeutic management of metastatic Non-Small cell lung cancer. J Thorac Oncol. 2017;12(11):1636–45.
6. Chan M, Palma D, Barry A, Hope A, Moore R, O'Neil M, et al. Practical considerations for the implementation of a stereotactic body radiation therapy program for Oligo-Metastases. Adv Radiat Oncol. 2021;6(1):100499.
7. Zhu Z, Ni J, Cai X, Su S, Zhuang H, Yang Z, et al. International consensus on radiotherapy in metastatic non-small cell lung cancer. Transl Lung Cancer Res. 2022;11(9):1763–95.
8. Hellman S, Weichselbaum RR, Oligometastases. J Clin Oncol. 1995;13(1):8–10.
9. Begum H, Swaminath A, Lee Y, Fahim C, Bramson J, Naqvi A, et al. The histologic effects of neoadjuvant stereotactic body radiation therapy (SBRT) followed by pulmonary metastasectomy-rationale and protocol design for the post SBRT pulmonary metastasectomy (PSPM) trial. Transl Cancer Res. 2022;11(4):918–27.
10. Pasalic D, Lu Y, Betancourt-Cuellar SL, Taku N, Mesko SM, Bagley AF, et al. Stereotactic ablative radiation therapy for pulmonary metastases: improving overall survival and identifying subgroups at high risk of local failure. Radiother Oncol. 2020;145:178–85.
11. Tsang MW. Stereotactic body radiotherapy: current strategies and future development. J Thorac Dis. 2016;8(Suppl 6):S517–27.
12. Buchberger DS, Videtic GMM. Stereotactic body radiotherapy for the management of Early-Stage Non-Small-Cell lung cancer: A clinical overview. JCO Oncol Pract. 2023;19(5):239–49.
13. Chang JY, Verma V, Feng L, Roth JA. SABR for operable stage I non-small-cell lung cancer: comparison to surgery – Authors' reply. The Lancet Oncology. 2021;22(12):e537-e8.
14. Kong FS, Moiseenko V, Zhao J, Milano MT, Li L, Rimner A, et al. Organs at risk considerations for thoracic stereotactic body radiation therapy: what is safe for lung parenchyma?? Int J Radiat Oncol Biol Phys. 2021;110(1):172–87.
15. Kocak B, Durmaz ES, Ates E, Kilickesmez O. Radiomics with artificial intelligence: a practical guide for beginners. Diagn Interv Radiol. 2019;25(6):485–95.
16. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563–77.
17. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48(4):441–6.
18. Liang W, Yang P, Huang R, Xu L, Wang J, Liu W, et al. A combined nomogram model to preoperatively predict histologic grade in pancreatic neuroendocrine tumors. Clin Cancer Res. 2019;25(2):584–94.
19. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. Clin Cancer Res. 2016;22(21):5256–64.
20. Li H, Zhu Y, Burnside ES, Drukker K, Hoadley KA, Fan C, et al. MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of mammaprint, oncotype DX, and PAM50 gene assays. Radiology. 2016;281(2):382–91.
21. Cheung BMF, Lau KS, Lee VHF, Leung TW, Kong FS, Luk MY, et al. Computed tomography-based radiomic model predicts radiological response following stereotactic body radiation therapy in early-stage non-small-cell lung cancer and pulmonary oligo-metastases. Radiat Oncol J. 2021;39(4):254–64.
22. Khorrami M, Prasanna P, Gupta A, Patil P, Velu PD, Thawani R, et al. Changes in CT radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in Non–Small cell lung cancer. Cancer Immunol Res. 2020;8(1):108–19.
23. Vaidya P, Bera K, Patil PD, Gupta A, Jain P, Alilou M, et al. Novel, non-invasive imaging approach to identify patients with advanced non-small cell lung cancer at risk of hyperprogressive disease with immune checkpoint Blockade. J Immunother Cancer. 2020;8(2):e001343.
24. Arneth B. Tumor Microenvironment Med. 2020;56(1):15.
25. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45(2):228–47.
26. Tian C, Hu Y, Li S, Zhang X, Wei Q, Li K et al. Peri- and intra-nodular radiomic features based on 18F-FDG PET/CT to distinguish lung adenocarcinomas from pulmonary granulomas. Front Med. 2024;11.
27. Beig N, Khorrami M, Alilou M, Prasanna P, Braman N, Orooji M, et al. Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas. Radiology. 2019;290(3):783–92.
28. Lin P, Xie W, Li Y, Zhang C, Wu H, Wan H et al. Intratumoral and peritumoral radiomics of MRIs predicts pathologic complete response to neoadjuvant chemoimmunotherapy in patients with head and neck squamous cell carcinoma. J Immunother Cancer. 2024;12(11).
29. Chen Z, Xu L, Zhang C, Huang C, Wang M, Feng Z, et al. CT radiomics model for discriminating the risk stratification of gastrointestinal stromal tumors: a multi-class classification and multi-center study. Front Oncol. 2021;11:654114.
30. Zhang X, Song X, Li G, Duan L, Wang G, Dai G, et al. Machine learning radiomics model for external and internal respiratory motion correlation prediction in lung tumor. Technol Cancer Res Treat. 2022;21:15330338221143224.
31. Palma DA, Nguyen TK, Louie AV, Malthaner R, Fortin D, Rodrigues GB, et al. Measuring the integration of stereotactic ablative radiotherapy plus surgery for Early-Stage Non-Small cell lung cancer: A phase 2 clinical trial. JAMA Oncol. 2019;5(5):681–8.
32. Hamamoto Y, Kataoka M, Yamashita M, Nogami N, Sugawara Y, Kozuki T, et al. Factors affecting the local control of stereotactic body radiotherapy for lung tumors including primary lung cancer and metastatic lung tumors. Jpn J Radiol. 2012;30(5):430–4.
33. Bade BC, Blasberg JD, Mase VJ Jr., Kumbasar U, Li AX, Park HS, et al. A guide for managing patients with stage I NSCLC: deciding between lobectomy, segmentectomy, wedge, SBRT and ablation-part 3: systematic review of evidence regarding surgery in compromised patients or specific tumors. J Thorac Dis. 2022;14(6):2387–411.
34. Cheung BMF. Radiomics in stereotactic body radiotherapy for non-small cell lung cancer: a systematic review and radiomic quality score study. Radiat Oncol J. 2024;42(1):4–16.
35. Yoon NK, Maresh EL, Shen D, Elshimali Y, Apple S, Horvath S, et al. Higher levels of GATA3 predict better survival in women with breast cancer. Hum Pathol. 2010;41(12):1794–801.
36. Avanzo M, Gagliardi V, Stancanello J, Blanck O, Pirrone G, El Naqa I, et al. Combining computed tomography and biologically effective dose in radiomics and deep learning improves prediction of tumor response to robotic lung stereotactic body radiation therapy. Med Phys. 2021;48(10):6257–69.
37. Cheung BMF, Lau KS, Lee VHF, Leung TW, Kong F-MS, Luk MY, et al. Computed tomography-based radiomic model predicts radiological response following stereotactic body radiation therapy in early-stage non-small-cell lung cancer and pulmonary oligo-metastases. Radiat Oncol J. 2021;39(4):254–64.
38. Cilla S, Pistilli D, Romano C, Macchia G, Pierro A, Arcelli A, et al. CT-based radiomics prediction of complete response after stereotactic body radiation therapy for patients with lung metastases. Strahlenther Onkol. 2023;199(7):676–85.
39. Fodor A, Mori M, Tummineri R, Broggi S, Deantoni CL, Mangili P, et al. CT radiomic predictors of local relapse after SBRT for lung oligometastases from colorectal cancer: a single Institute pilot study. Strahlenther Onkol. 2023;199(5):477–84.
40. Cheung BMF, Lau JKS, Luk MY, Yuen KK. 361P radiomic model predicting radiological response after thoracic stereotactic body radiotherapy regardless of tumor histology and staging. Ann Oncol. 2020;31:S1380.
41. Luo Y, Tseng HH, Cui S, Wei L, Ten Haken RK, El Naqa I. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. BJR Open. 2019;1(1):20190021.
42. Raptis S, Tsougos I, Theodorou K, Ilioudis C, Edts. harmonizing radiomics and interpretable AI: precision and transparency in oncological prognostication. 2024 IEEE international symposium on biomedical imaging (ISBI); 2024 27–30 May 2024.
43. Pang XL, Chen XB, Feng LL, Hong MP, Xie P, Wei KK et al. Interpretable artificial intelligence to predict treatment response in rectal cancer. Int J Radiat Oncol Biol Phys. 2024;120(2 Suppl):e663.
44. Raptis S, Ilioudis C, Theodorou K. From pixels to prognosis: unveiling radiomics models with SHAP and LIME for enhanced interpretability. Biomed Phys Eng Express. 2024;10(3).
45. Ganeshan B, Abaleke S, Young RC, Chatwin CR, Miles KA. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. Cancer Imaging. 2010;10(1):137–43.
46. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to Decode the radiographic phenotype. Cancer Res. 2017;77(21):e104–7.

47. Schiavo F, Toma-Dasu I, Kjellsson Lindblom E. The impact of heterogeneous cell density in hypoxic tumors treated with radiotherapy. Adv Exp Med Biol. 2023;1438:121–6.
48. Long QY, Wang FY, Hu Y, Gao B, Zhang C, Ban BH, et al. Development of the interpretable typing prediction model for osteosarcoma and chondrosarcoma based on machine learning and radiomics: a multicenter retrospective study. Front Med (Lausanne). 2024;11:1497309.
49. Yu H, Buch K, Li B, O'Brien M, Soto J, Jara H, et al. Utility of texture analysis for quantifying hepatic fibrosis on proton density MRI. J Magn Reson Imaging. 2015;42(5):1259–65.
50. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for High-Throughput image-based phenotyping. Radiology. 2020;295(2):328–38.
51. Mayerhoefer ME, Materka A, Langs G, Haggstrom I, Szczypinski P, Gibbs P, et al. Introduction to radiomics. J Nucl Med. 2020;61(4):488–95.
52. Diaz-Cano SJ. Tumor heterogeneity: mechanisms and bases for a reliable application of molecular marker design. Int J Mol Sci. 2012;13(2):1951–2011.
53. Knoche SM, Larson AC, Sliker BH, Poelaert BJ, Solheim JC. The role of tumor heterogeneity in immune-tumor interactions. Cancer Metastasis Rev. 2021;40(2):377–89.

Publisher's note