



CT-based radiomics prediction of complete response after stereotactic body radiation therapy for patients with lung metastases

Savino Cilla¹ · Domenico Pistilli¹ · Carmela Romano¹ · Gabriella Macchia² · Antonio Pierro³ · Alessandra Arcelli⁴ · Milly Buwenge⁴ · Alessio Giuseppe Morganti^{4,5} · Francesco Deodato^{2,6}

Received: 6 February 2023 / Accepted: 11 April 2023 / Published online: 31 May 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany 2023

Abstract

Purpose Stereotactic body radiotherapy (SBRT) is a key treatment modality for lung cancer patients. This study aims to develop a machine learning-based prediction model of complete response for lung oligometastatic cancer patients undergoing SBRT.

Materials and methods CT images of 80 pulmonary oligometastases from 56 patients treated with SBRT were analyzed. The gross tumor volumes (GTV) were contoured on CT images. Patients that achieved complete response (CR) at 4 months were defined as responders. For each GTV, 107 radiomic features were extracted using the Pyradiomics software. The concordance correlation coefficients (CCC) between the region of interest (ROI)-based radiomics features obtained by the two segmentations were calculated. Pairwise feature interdependencies were evaluated using the Spearman rank correlation coefficient. The association of clinical variables and radiomics features with CR was evaluated with univariate logistic regression. Two supervised machine learning models, the logistic regression (LR) and the classification and regression tree analysis (CART), were trained to predict CR. The models were cross-validated using a five-fold cross-validation. The performance of models was assessed by receiver operating characteristic curve (ROC) and class-specific accuracy, precision, recall, and F1-measure evaluation metrics.

Results Complete response was associated with four radiomics features, namely the surface to volume ratio (SVR; $p=0.003$), the skewness (Skew; $p=0.027$), the correlation (Corr; $p=0.024$), and the grey normalized level uniformity (GNLU; $p=0.015$). No significant relationship between clinical parameters and CR was found. In the validation set, the developed LR and CART machine learning models had an accuracy, precision, and recall of 0.644 and 0.750, 0.644 and 0.651, and 0.635 and 0.754, respectively. The area under the curve for CR prediction was 0.707 and 0.753 for the LR and CART models, respectively.

Conclusion This analysis demonstrates that radiomics features obtained from pretreatment CT could predict complete response of lung oligometastases following SBRT.

Keywords Radiomics · Machine learning · Lung cancer · SBRT · Treatment response

Two authors share the seniorship: Alessio Giuseppe Morganti and Francesco Deodato.

✉ Dr. Savino Cilla
savinocilla@gmail.com, savino.cilla@gemellimolise.it

¹ Gemelli Molise Hospital, Medical Physics Unit, Largo Gemelli 1, 86100 Campobasso, Italy

² Radiation Oncology Unit, Gemelli Molise Hospital, Campobasso, Italy

³ Radiology Unit, Gemelli Molise Hospital, Campobasso, Italy

⁴ Radiation Oncology Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy

⁵ Department of Experimental, Diagnostic, and Specialty Medicine—DIMES, Alma Mater Studiorum, Università di Bologna, Diagnostic, Italy

⁶ Istituto di Radiologia, Università Cattolica del Sacro Cuore, Roma, Italy

Introduction

Advancements in imaging and radiation planning and delivery techniques have allowed widespread implementation of stereotactic body radiation therapy (SBRT) as a gold standard of treatment for incurable early-stage non-small cell lung cancer (NSCLC) and lung metastases [1, 2]. In the latter case, use of SBRT allows for high ablative doses with the potential for longer local control and overall survival. In particular, a few investigations have reported local control of isolated or few lung metastases to be 70–100% at 1 year [3–5].

However, it has recently been reported that the pathological complete response (CR) rate following thoracic SBRT was lower than previously predicted, not only in early-stage non-small cell lung cancer [6] but also in the lung oligometastatic scenario [7–9].

Therefore, there is an urgent need for a patient stratification approach that should be able to identify the patients who are at a higher risk of incomplete/partial response (or at a higher risk of recurrence) after SBRT for lung metastases. Such an approach could help to design treatment plans tailored to the specific disease profile of the patient, thus improving the therapeutic window by allowing, prior to treatment, the choice of dose intensification or additional systemic therapy.

There is currently still a paucity of accurate predictive models for clinical outcomes of lung SBRT [10–12]. In the scenario of pulmonary metastasis, Tanadini-Lang et al. [10] have developed a nomogram that predicts 2-year overall survival after SBRT. Karnofsky performance index, type of primary tumor, control of the primary tumor, maximum diameter, and number of metastases were found to be significant prognostic factors. The calculated concordance index for the nomogram was 0.73. In the case of primary lung tumors, Baker et al. [11] developed and validated a nomogram to predict overall survival following SBRT for early-stage lung cancer. The resulting nomogram was based on six prognostic factors: age, sex, Karnofsky performance status, operability, Charlson comorbidity index, and tumor diameter, providing a concordance index of 0.64. Lastly, Ye et al. [12] created a pretreatment prognostic nomogram for patients with stage I NSCLC receiving SBRT. The authors showed that 2-year locoregional control and progression-free survival were significantly related to tumor size and the maximum standardized uptake value. The concordance indexes for the nomograms were greater than 0.8 and 0.7 in the model and validation groups, respectively.

In recent years, radiomics, a novel approach able to extract a large number of quantitative features from medical images, has shown great potential for assessing the tumor phenotype [13]. These mathematical features can be used as imaging biomarkers for diagnosis, staging, prognosis,

and prediction of tumor response. Although radiomics for pulmonary metastasis is still mostly exploratory [14], a few studies have been successful in obtaining radiomics signatures for prediction of nodal relapse and recurrence rate [15–18] or overall survival [19–22]. However, the clinical impact of the various developed algorithms is still lower than expected, mainly due to the poor reproducibility of the results and the lack of external validation on data from other institutions [23]. Studies focused on the efficacy of SBRT treatments in terms of treatment response after SBRT are much rarer. Cheung et al. [24] identified two radiomics predictors of treatment response (the skewness and the root mean squared) in patients affected by pulmonary oligometastases treated with SBRT and developed a machine learning model for prediction purposes. Their support vector machine model supplied an accuracy of 74.8%.

Because of the large worldwide implementation of SBRT for lung oligometastases, new powerful predictive models based on radiomics analysis would be welcome to differentiate between tumors responsive and non-responsive to treatment.

The aim of the current study was to perform an exploratory CT-based radiomics analysis of lung metastases able to identify the radiomics features that can predict radiological and functional response after SBRT. Thereafter, two machine learning models were generated to predict the treatment response after SBRT based on selected radiomics and clinical features.

Materials and methods

Patients

From January 2010 to December 2020, 56 patients with 80 lung oligometastases were treated with SBRT in our institution. Patients were enrolled within a previous prospective phase I–II study in primary or oligometastatic cancer patients (DESTROY-1) [25], approved by the Catholic University Institutional Review Board (Destroy-1: P#594/CE). All patients signed written informed consent.

Patients with performance status ECOG >3 or medical problems unrelated to malignancy which would limit full compliance with the study and those with previous SBRT were excluded.

CT image acquisition, segmentation, and planning

Computed tomography (CT) simulation was performed with a 128-slice scanner (Brilliance 128, Philips Healthcare, Best, the Netherlands). CT images were acquired with 2-mm thick slices. All patients were simulated in supine position. All lesions were manually segmented by a senior

radiologist (A.P.) and a senior radiation oncologist (F.D.) on the treatment planning system.

The clinical target volume (CTV) was defined as the gross tumor volume (GTV) identified on the planning CT and/or CT-PET and/or MRI. The internal target volume (ITV) was defined on the basis of respiratory excursions analysis (free breathing or abdominal compression or deep-inspiration breath-hold). Based on the quantification of patients' setup reproducibility, the setup margin was set at 3 mm, as previously described [25]. Image-guided radiotherapy was used for patient setup at each fraction.

All patients were treated with a prescription dose of 40–50 Gy in five consecutive fractions, equivalent to a BED₁₀ ranging from 80 to 100 Gy. SBRT plans were generated using the fixed-field intensity-modulated (IMRT) or the volumetric modulated arc therapy (VMAT) technique with 6-MV X-ray energy beams. A uniform method for selection of the prescription isodose surface (IDS) was adopted. For each plan, the IDS was selected as the greatest IDS fulfilling the two following criteria: 95% of the PTV volume reached 100% of the prescription dose and 99% of the PTV reached a minimum of 90% of the prescription dose. The maximum dose within the GTV should not exceed 140% of the prescribed dose. Plans were calculated using the collapsed cone convolution algorithm and a dose grid size of 1.5 mm.

Treatment response evaluation

The primary clinical outcome of this study was the radiological treatment response, which was evaluated by morphological contrast-enhanced CT 4 months after SBRT. This timing was considered suitable for both early and late responder malignancies, according to the RECIST (Response Evaluation Criteria in Solid Tumors) system v1.0 [26]. Complete response (CR) was defined as the disappearance of the lesions at CT scan; a reduction greater than 30% was considered partial response (PR); any growing lesion not clearly ascribable to fibrosis was reported as progression of disease (PD).

Radiomics analysis

The CT DICOM images containing the segmentation of the GTV were used to extract all the features for radiomics analysis. Image processing was performed using the software package Pyradiomics 2.0.1 [27]. This software allows the extraction of standardized features as defined by the IBSI (Image Biomarker Standardization Initiative) [28]. Preprocessing included resampling of the CT images to isotropic voxels of 1 mm³ using a B-spline interpolation function. A total of 107 features were automatically extracted for each lesion, including 19 first-order statis-

tics features (intensity histogram, IH), 26 shape-based histogram features, and texture features (gray-level co-occurrence matrix, GLCM, 24 features; gray-level run-length matrix, GLRLM, 16 features; gray-level size-zone matrix, GLSZM, 16 features; neighboring gray-tone difference matrix, NGTDM, 5 features; and gray-level dependence matrix, GLDM, 14 features).

Features selection, modeling, and statistical analysis

The following three steps were performed to avoid collinearity, reduce dimensionality, and identify only the necessary radiomics features. First, we calculated the concordance correlation coefficients (CCC) between the ROI-based radiomics features obtained by the two segmentations for 30 randomly chosen lesions performed by a radiologist and a radiation oncologist. The stable features were defined as the features having a CCC > 0.90 and were included in the subsequent analysis.

Secondly, we assessed the pairwise correlations between the radiomics features using the Spearman rank coefficient (ρ). All features having $|\rho| \geq 0.8$ were considered as redundant and were eliminated from subsequent analysis, maintaining only the features with the strongest association with the clinical outcome in terms of univariate analysis.

Finally, the relative importance of the remaining features was determined by a stepwise backward elimination approach. In this process, a feature is considered for subtraction from the set of explanatory variables based on the Akaike information criterion (AIC).

The Mann–Whitney U test was used to assess the association between the treatment response and the selected radiomics features. Statistical significance was set for p -values of 0.05 or lower.

The final significant features were finally used to build two models for the classification of responsive or non-responsive lesions, including logistic regression (LR) and classification and regression tree analysis (CART).

LR is a classical machine learning algorithm that is usually used for binary classification tasks. This model provides the class membership probability for one of the two categories in the dataset (0 or 1) using a logistic equation. Fast training is a well-known advantage of LR, and input variables can be either discrete or continuous.

The classification and regression tree (CART) is a decision tree-based method widely used in datamining. It is able to reveal data relationships and automatically searches for patterns to uncover hidden structure even in highly complex data. The typical representation of the CART model is a binary tree in which each root node represents a single input feature and a split point on that feature. The leaf nodes of the tree contain an output variable that is used to

make a prediction. The best splits were identified by the Gini impurity (GI) index.

The Gini index varies between values 0 and 1, where 0 expresses the purity of classification (i.e., all the elements belong to a specified class) and 1 indicates the random distribution of elements across various classes. The value of 0.5 of the Gini index shows an equal distribution of elements over some classes.

The models were cross-validated using a five-fold cross-validation. This is a resampling procedure primarily used to estimate the skill of a machine learning model on unseen data. This approach involves randomly dividing the set of observations into k groups (five folds in our case) of approximately equal size. The first fold is treated as a validation set, and the method is fit on the remaining k–1 folds. The process was repeated ten times with the aim of reducing the variance of the cross-validation results and the chance of getting too optimistic results from only one run.

The performance of the models was assessed using receiver operating characteristic curves (ROCs), together with accuracy, precision, recall, and F1-measure evaluation metrics.

Statistical analysis including machine learning training and testing was performed using the XLSTAT statistical packages (Addinsoft, New York, NY, USA).

Results

Patients

Fifty-six patients (38 males, 18 females) were recruited from 2010 to 2020. Median age was 72 years (range 38–90 years). The total number of analyzed lung metastases was 80, of which 36 achieved complete response (group CR). Another 31 metastases achieved partial response, 10 were considered as stable disease, and 3 were considered as progressive disease; these 44 metastases were considered non-complete responders (group NCR). Most metastases were from pulmonary (44 metastases) and colorectal (20 metastases) tumors. In terms of histology, the large majority were adenocarcinoma (35 lesions), followed by squamous cell carcinomas (7 lesions). Table 1 reports the patients’ characteristics.

Table 1 Patient characteristics

| | |
|---|--|
| No. of patients | 56 |
| Median age, years (range) | 72.0 (37.5–90.0) |
| Gender, M/F | 38/18 |
| Smoker, Y/N | 23/33 |
| No. of lesions | 80 |
| Location, central/peripheral | 17/63 |
| Number of lesions, patients | 1: 46 patients 2: 12 patients 3: 3 patients 4: 1 patient |
| Number of lesions, primary tumor | 44: lung 11: rectum 9: colon 8: gynecological 2: breast 2: skin 4: other |
| Median GTV, cc (range) | 5.3 (0.2–61.9) |
| Median PTV, cc (range) | 27.5 (4.9–215.7) |
| Median prescribed dose, Gy (range) | 50 (30–50) |
| Median prescribed isodose, percentage (range) | 81 (73–89) |
| Median BED, Gy (range) | 100 (48–100) |

GTV Gross Tumor Volume, PTV Planning Target Volume, BED Biological Equivalent Dose

Radiomics features and selection

A total of 107 radiomics features were extracted from each lesion.

The overall mean CCC was 0.97 ± 0.14 . The mean CCC for shape features between the two clinicians was 0.98 ± 0.08 , indicating favorable interobserver reproducibility for the segmentation of the lung metastases. The mean CCCs for the intensity and texture features were ≥ 0.89 in all the 30 analyzed metastases.

The pairwise Spearman correlation analysis between features revealed a high level of interdependence in the overall dataset, with over 85% correlating with at least one other feature at $|r| > 0.8$. As a result of these two steps, we obtained a primary feature set of 15 from 107.

Table 2 Diagnostic accuracy statistics for the four features highly correlated with complete response

| | Area under the curve (AUC) [CI95%] | Sensitivity | Specificity | Positive predictive value (PPV) | Negative predictive value (NPV) | Accuracy |
|------|------------------------------------|-------------|-------------|---------------------------------|---------------------------------|----------|
| SVR | 0.692 [0.578–0.806] | 0.639 | 0.591 | 0.650 | 0.750 | 0.700 |
| Skew | 0.645 [0.525–0.764] | 0.676 | 0.684 | 0.561 | 0.667 | 0.613 |
| Corr | 0.648 [0.527–0.769] | 0.750 | 0.545 | 0.574 | 0.727 | 0.638 |
| GLNU | 0.659 [0.543–0.775] | 0.667 | 0.614 | 0.585 | 0.692 | 0.638 |

SVR Surface Volume Ratio, Skew Skewness, Corr correlation and GLNU Grey normalized level uniformity

Fig. 1 Box-and-whisker plots for the four radiomics features associated with the response outcome at univariate analysis: **a** surface volume ratio, **b** skewness, **c** correlation, **d** grey level nonuniformity. *CR* complete response and *noCR* non-complete response

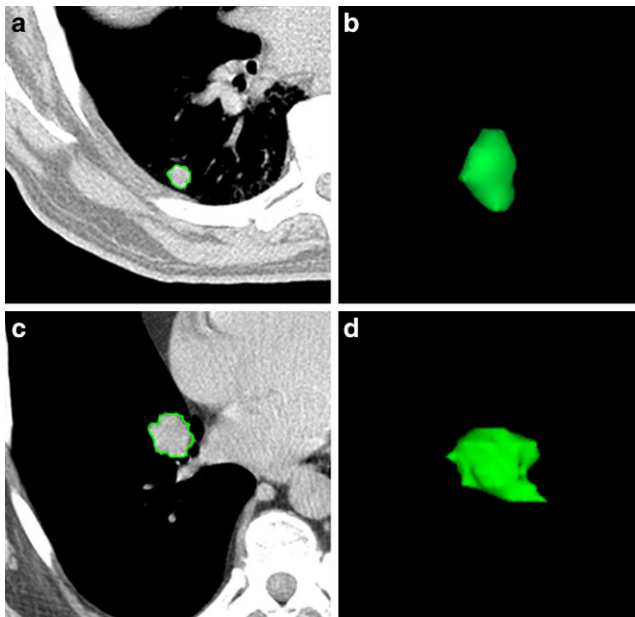
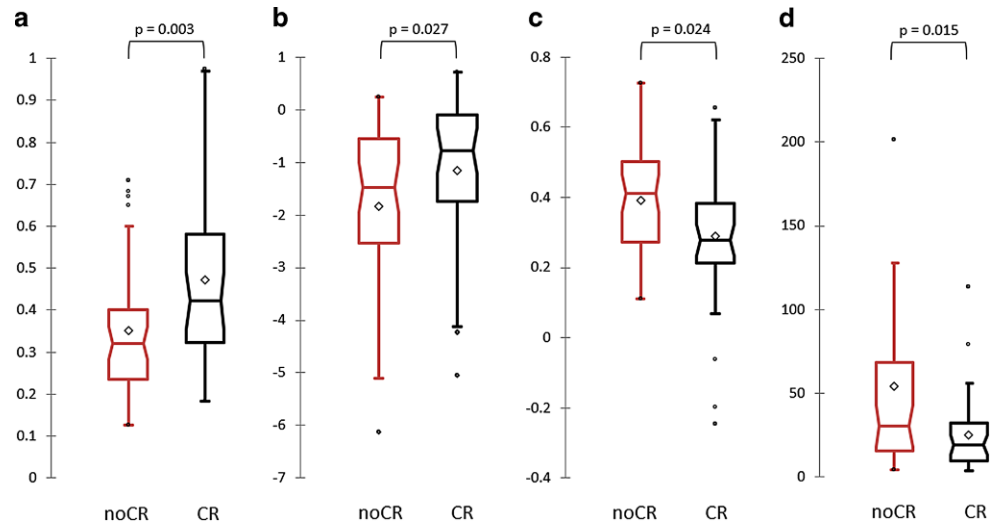


Fig. 2 Examples of two lesions on axial CT and 3D plots and their respective radiomic feature values: **a, b** lung lesion in a 78-year-old male patient who evolved with complete response, surface volume ratio: 0.687, skewness: 0.108, correlation: -0.246 , grey level nonuniformity: 3.316; **c, d** lung lesion in a 75-year-old male patient who evolved with stable disease, surface volume ratio: 0.343, skewness: -1.911 , correlation: 0.220, grey level nonuniformity: 353.81

Following this, based on the stepwise backward elimination approach, four final features, namely the surface to volume ratio (SVR), the skewness (Skew), the correlation (Corr), and the grey normalized level uniformity (GNLU), were selected for constructing the predictive models. The univariate analysis (Mann–Whitney U test) for the association of these four radiomics features with the two lesion groups reported statistical significance with $p < 0.05$. The detailed diagnostic accuracy statistics for these four features are reported in Table 2 in terms of area under the curve

(AUC), sensitivity, specificity, positive and negative predictive value, and accuracy. The relative box-and-whisker plots are shown in Fig. 1 according to the response class.

Two lesions reporting different responses are shown in Fig. 2.

In particular, other than obvious macroscopic differences in terms of dimension and shape, the visual assessment of lesion texture does not allow any kind of reliable evaluation. The smaller lesion in Fig. 2a was associated with a complete response while the larger lesion in Fig. 2b was associated with stable disease after SBRT.

Of note, no clinical variable was found to be associated with complete response. The only non-radiomics variable associated with the two response groups was the volume of the lesion, with a smaller volume significantly associated with a complete response. However, as expected, this variable was found to be collinear with the SVR feature and was eliminated from subsequent analysis.

Table 3 Performance of the logistic regression (LR) and classification and regression tree (CART) classifiers for the training and internal validation datasets

| | LR | CART |
|-----------------------|---------------------|---------------------|
| <i>Training set</i> | | |
| Accuracy | 0.691 (0.667–0.715) | 0.792 (0.756–0.827) |
| Precision | 0.640 (0.615–0.665) | 0.784 (0.698–0.871) |
| Recall | 0.680 (0.658–0.702) | 0.803 (0.727–0.879) |
| F1 score | 0.659 (0.639–0.680) | 0.785 (0.761–0.809) |
| <i>Validation set</i> | | |
| Accuracy | 0.644 (0.610–0.679) | 0.750 (0.717–0.783) |
| Precision | 0.644 (0.600–0.687) | 0.651 (0.597–0.705) |
| Recall | 0.635 (0.606–0.665) | 0.754 (0.692–0.815) |
| F1 score | 0.638 (0.610–0.665) | 0.694 (0.668–0.719) |

Fig. 3 Receiver operating characteristic curve for the logistic regression (LR; **a**) and classification and regression tree (CART; **b**) models

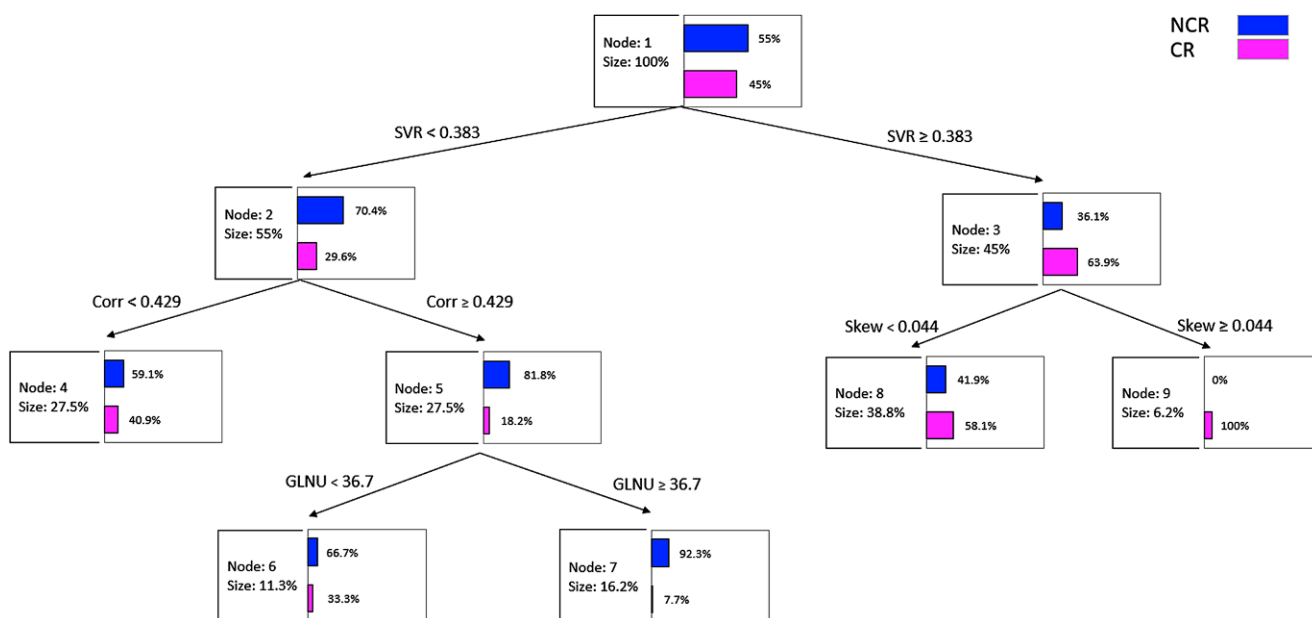
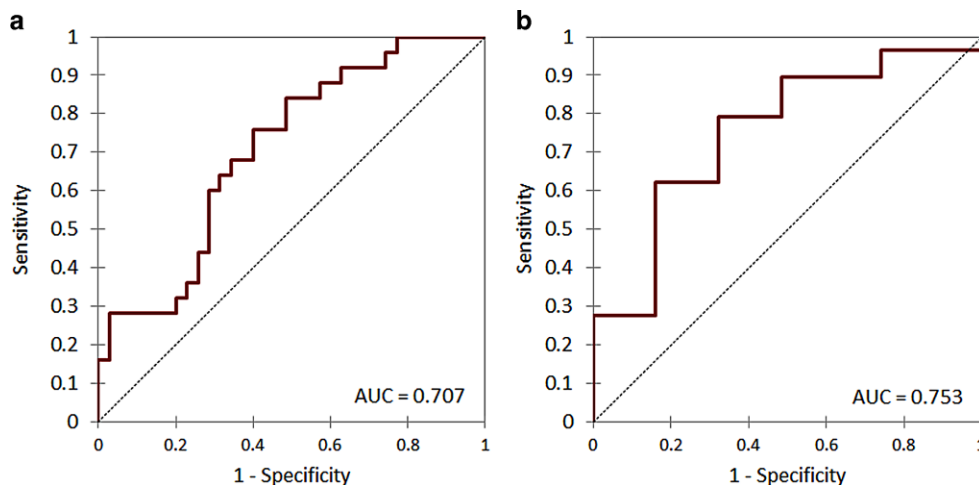


Fig. 4 Classification and regression tree analysis (CART) for the most significant radiomics features

Machine learning models

The classification performance for the LR and CART models are reported in Table 3 for the training and internal validation datasets. The CART classifier reported the best performance, with accuracy, precision, recall, and F1 score equal to 79.2%, 78.4%, 80.3%, and 78.5%, respectively. The receiver operating characteristic (ROC) curves of the two models for the cross-validated results are shown in Fig. 3.

The areas under the curves (AUC) were 0.707 (95% CI: 0.633–0.817) and 0.753 (95% CI: 0.675–0.836) for the LR and CART models, respectively.

The CART classification tree for the most informative variables is displayed in Fig. 4.

This figure well demonstrates the interpretability of the CART model, i.e., its ability to help human experts to understand the cause of the decisions adopted by the artificial intelligence algorithm.

Discussion

Lung cancer is a very heterogeneous disease, at both cellular and histological levels [29]. Substantial variation can be observed not just between individuals but also across lesions in the same patient and between various parts of the same lesion. Consequently, study of the clinical characteristics alone might may fail to capture the heterogeneity of the lung lesions. Nowadays, it has been proven that tumors with high heterogeneity are to be considered more aggressive [30] and that tumors with irregular spiculated edges

are typical of more infiltrating tumors [31]. Overall, these findings imply that lesions with denser internal areas and less heterogeneity may respond more favorably.

Moreover, differentiating primary lung cancer from lung metastasis is of the utmost clinical importance because of the significant differences in prognostic and therapeutic implications [32].

From this point of view, radiomics has major potential to provide reliable predictions in cancer outcomes thanks to its ability to describe histology and genetic footprint and to capture intratumoral tumor heterogeneity [33]. However, radiomics is a relatively recent field of research and its potential clinical use for lung metastases is still a niche subfield, as demonstrated by the scarcity of related studies.

The main objective of the present study was to explore the feasibility of radiomics for predicting a complete response of lung metastases in patients undergoing SBRT. Aiming to improve the predictive model, we also attempted to integrate clinical features such as age, smoking status, tumor histology, and stage, but no clinical variable was found to be correlated with the lesion response. While none of the clinical features were prognostic of complete response, we found four radiomics features (surface-to-volume ratio, skewness, correlation, and grey level nonuniformity) to be significantly correlated with complete response, highlighting the potential of radiomics analysis for outcome prediction after SBRT of lung metastases. The surface-to-volume ratio was the shape feature with higher correlation with complete response. In particular, we found that a lesion showing higher values of SVR suggests a higher probability of complete response after SBRT. This was not an unexpected result because it has been well reported that patients with large-volume and less-spherical tumors have a worse prognosis in terms of overall survival and locoregional control [34, 35].

Interestingly, also skewness and correlation were identified in previous radiomic studies as a major radiomics signature for lung tumors. To our knowledge, only one study performed by Cheung et al. [24], evaluated the role of radiomics for prediction purposes following SBRT in pulmonary oligometastases. The authors reported that skewness and root mean square were predictors of radiological treatment response with an accuracy of 74.8%. On the other hand, several investigations focused on the application of radiomics to primary lung tumors, to identify predictive radiomics features. Bousabarah et al. [22] reported on the application of CT radiomics analysis in a cohort of 110 inoperable early-stage NSCLC patients treated with SBRT aiming to predict local control, overall, and disease-free survival. In particular, the authors reported that lung tumors with high skewness (i.e., lesions with lower Hounsfield values dominant) have a lower risk for local failure and increased survival probability, and lung tumors with a high correla-

tion (a measure of tissue homogeneity) were more likely to cause local failure. These results overlap perfectly with our findings. Coroller et al. [36] identified skewness as one of the features significantly associated with pathological gross residual disease following chemoradiation in early-stage lung cancer. Similarly, Chong et al. [37] demonstrated that skewness was predictive of pathological non-responder patients following chemoradiation in lung adenocarcinoma. Lastly, Caruso et al. [38] tested the ability of CT texture analysis on chest CT to identify malignant lung nodules. The authors reported that in addition to kurtosis, the most significant difference in the comparison between texture parameters of malignant and benign nodules was the skewness, with values statistically higher in malignant lesions.

As there is growing evidence from different radiomics studies that the skewness value may be a signature for poor prognosis, some researchers are actually exploring the biological basis underlying radiomic features that may have a specific radiological appearance [39–41]. In particular, skewness has been found to be associated with *KRAS* mutations in NSCLC [39] that characterize lung cancers with poor prognosis and treatment resistance [40]. A current hypothesis is that the poorer prognosis observed in “skewed” tumors may be at least partially driven by *KRAS* mutation, for example by the induction of a cancer stem cell-like phenotype leading to radioresistance [41].

Using these four top-ranked features, we trained and evaluated two machine learning models using LR and CART analysis methods. The best diagnostic performance was attained by the CART classifier, reporting an accuracy, precision, and recall of 0.792, 0.784, and 0.803 in the training set and 0.750, 0.651, and 0.754 in the evaluation set, respectively. The obtained CART model has the notable advantage of being easily interpretable, unlike other methods of machine learning or deep learning. With four radiomics covariates, we modeled the CART using nine nodes and three depth levels in order to produce a straightforward and compact decision tree. The CART analysis chose the lesions with an SVR greater than 0.383 to be associated with complete response. The skewness and correlation subsequently played a significant role in increasing the classification rate among the lesions, with high skewness and low correlation values favoring complete response. Because of this high classification accuracy, we think this CART model could represent a valuable tool for radio-oncologists when prescribing effective treatment options in an SBRT strategy for lung lesions.

Despite encouraging results, a few limitations of this study should be highlighted and discussed. Firstly, this is a retrospective study and the size of the patient cohort (80 lesions) is not overly large. This is a typical drawback of many quantitative radiomics investigations. For this reason, we choose a simple binary classification for the response

outcome (complete response vs. all other responses) to allow a reasonable sample size for each group. In the future, our aim is to gather more lesions in order to better stratify the response beyond the proposed binary classification (i.e., complete response, partial response, stable disease, and progressive disease). Secondly, our research was limited to a single institution, meaning that image acquisition, target definition, and the evaluation of clinical outcomes may differ between different institutions. With respect to target definition, accurate lesion segmentation is essential to ensure the reliability of the extracted features. For this reason, in order to quantify the reproducibility of the features, we calculated the CCC for the radiomic features extracted from the two sets of segmentations obtained by two independent physicians. The very high CCC values (mean 0.972) are a signature of reliable target definition. Also, the intra- and interobserver uncertainties in the assessment of RECIST response may have led to misclassification of the overall response [42]. However, it has been well reported that the reproducibility of response is much higher for lung lesions with respect to other sites [43]. Lastly, the limited cohort of lesions did not allow evaluation of the performance of the models in an independent dataset of patients. To partially overcome this limitation, we performed a reliable internal validation with a five-fold cross-validation process. This is a well-known resampling technique used to evaluate machine learning models based on a limited data sample, i.e., to estimate how the model is expected to perform in predicting data not used during the training of the model. On the other hand, it must be also underlined that cross-validation cannot strictly evaluate the selection bias in retrospective studies, thus potentially translating into an overestimation of prediction performance. Nonetheless, a recent study focused on the development of radiomic models for classification of malignant lung tumors reported that internally validated models with AUC > 0.89 had AUC of 0.82 in external validation [44]. In conclusion, even if an external validation of our results with an independent dataset is needed to confirm the robustness and generalizability of our model, our study presents a reliable proof-of-concept that the radiomics analysis may improve the accuracy necessary for prediction of SBRT treatment response.

Funding The authors received no specific financial support for the research, authorship, and/or publication of this article.

Author Contribution Concept and design: SC. Conduct: SC, DP, FD, CR, GM. Supervision: AGM. Data acquisition: DP, CR, AP, AA, MB. Statistical analysis: SC. Critical review: SC, FD, GM, AGM. Manuscript drafting, editing: all authors. Revision and final approval: all authors.

Declarations

Conflict of interest S. Cilla, D. Pistilli, C. Romano, G. Macchia, A. Pierro, A. Arcelli, M. Buwenge, A.G. Morganti, and F. Deodato declare that they have no competing interests.

Ethical standards All procedures performed in studies involving human participants or on human tissue were in accordance with the ethical standards of the institutional and/or national research committee and with the 1975 Helsinki declaration and its later amendments or comparable ethical standards. The study received approval at the Gemelli Molise Hospital Institutional Review Board. Informed consent: not applicable; patients signed informed consent to the treatment procedure, but the study was retrospective.

References

1. Timmerman R, Paulus R, Galvin J et al (2010) Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA* 303(11):1070–1076. <https://doi.org/10.1001/jama.2010.261>
2. Virbel G, Le Fèvre C, Noël G, Antoni D (2021) Stereotactic body radiotherapy for patients with lung oligometastatic disease: a five-year systematic review. *Cancers* 13(14):3623. <https://doi.org/10.3390/cancers13143623>
3. Ricardi U, Filippi AR, Guarneri A, Ragona R, Mantovani C, Giglioli F et al (2012) Stereotactic body radiation therapy for lung metastases. *Lung Cancer* 75:77–81. <https://doi.org/10.1016/j.lungcan.2011.04.021>
4. Okunieff P, Petersen AL, Philip A, Milano MT, Katz AW, Boros L et al (2006) Stereotactic Body Radiation Therapy (SBRT) for lung metastases. *Acta Oncol* 45:808–817. <https://doi.org/10.1080/02841860600908954>
5. Nuytens JJ, van der Voort van Zyp NCMG, Verhoef C, Maat A, van Klaveren RJ, van der Holt B et al (2015) Stereotactic body radiation therapy for oligometastases to the lung: a phase 2 study. *Int J Radiat Oncol Biol Phys* 91:337–343. <https://doi.org/10.1016/j.ijrobp.2014.10.021>
6. Palma DA, Nguyen TK, Louie AV et al (2019) Measuring the integration of stereotactic ablative radiotherapy plus surgery for early-stage non-small cell lung cancer: a phase 2 clinical trial. *JAMA Oncol* 5:681–688. <https://doi.org/10.1001/jamaoncol.2018.6993>
7. Frakulli R, Salvi F, Balestrini D et al (2015) Stereotactic radiotherapy in the treatment of lung metastases from bone and soft-tissue sarcomas. *Anticancer Res* 35(10):5581–5586
8. Lin Q, Zhou N, Zhu X et al (2022) Outcomes of SBRT for lung oligo-recurrence of non-small cell lung cancer: a retrospective analysis. *J Radiat Res* 63(2):272–280. <https://doi.org/10.1093/jrr/rrab118>
9. Salama JK, Hasselle MD, Chmura SJ et al (2012) Stereotactic body radiotherapy for multisite extracranial oligometastases: final report of a dose escalation trial in patients with 1 to 5 sites of metastatic disease. *Cancer* 118(11):2962–2970. <https://doi.org/10.1002/ncr.26611>
10. Tanadini-Lang S, Rieber J, Filippi AR et al (2017) Nomogram based overall survival prediction in stereotactic body radiotherapy for oligo-metastatic lung disease. *Radiother Oncol* 123:182–188. <https://doi.org/10.1016/j.radonc.2017.01.003>
11. Baker S, Bakunina K, Duijm M et al (2020) Development and external validation of a nomogram to predict overall survival following stereotactic body radiotherapy for early-stage lung cancer. *Radiat Oncol* 15(1):89. <https://doi.org/10.1186/s13014-020-01537-z>
12. Ye L, Shi S, Zeng Z et al (2018) Nomograms for predicting disease progression in patients of Stage I non-small cell lung cancer.

- cer treated with stereotactic body radiotherapy. *Jpn J Clin Oncol* 48(2):160–166. <https://doi.org/10.1093/jco/hyx179>
13. Lambin P, Rios-Velazquez E, Leijenaar R et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48:441–446. <https://doi.org/10.1016/j.ejca.2011.11.036>
 14. Gabelloni M, Faggioni L, Fusco R et al (2023) Radiomics in lung metastases: a systematic review. *J Pers* 13:225. <https://doi.org/10.3390/jpm13020225>
 15. Franceschini D, Cozzi L, De Rose F et al (2020) A radiomic approach to predicting nodal relapse and disease-specific survival in patients treated with stereotactic body radiation therapy for early-stage non-small cell lung cancer. *Strahlenther Onkol* 196(10):922–931. <https://doi.org/10.1007/s00066-019-01542-6>
 16. Lafata KJ, Hong JC, Geng R et al (2019) Association of pretreatment radiomic features with lung cancer recurrence following stereotactic body radiation therapy. *Phys Med Biol* 64(2):25007. <https://doi.org/10.1088/1361-6560/aaf5a5>
 17. Li Q, Kim J, Balagurunathan Y et al (2017) CT imaging features associated with recurrence in non-small cell lung cancer patients after stereotactic body radiotherapy. *Radiat Oncol* 12(1):158. <https://doi.org/10.1186/s13014-017-0892-y>
 18. Oikonomou A, Khalvati F, Tyrrell PN et al (2018) Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy. *Sci Rep* 8(1):4003. <https://doi.org/10.1038/s41598-018-22357-y>
 19. Huynh E, Coroller TP, Narayan V et al (2016) CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. *Radiother Oncol* 120(2):258–266. <https://doi.org/10.1016/j.radonc.2016.05.024>
 20. Li H, Galperin-Aizenberg M, Pryma D et al (2018) Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy. *Radiother Oncol* 129(2):218–226. <https://doi.org/10.1016/j.radonc.2018.06.025>
 21. Li Q, Kim J, Balagurunathan Y et al (2017) Imaging features from pretreatment CT scans are associated with clinical outcomes in non-small cell lung cancer patients treated with stereotactic body radiotherapy. *Med Phys* 44(8):4341–4349. <https://doi.org/10.1002/mp.12309>
 22. Bousabarah K, Temming S, Hoevels M et al (2019) Radiomic analysis of planning computed tomograms for predicting radiation-induced lung injury and outcome in lung cancer patients treated with robotic stereotactic body radiation therapy. *Strahlenther Onkol* 195:830–842. <https://doi.org/10.1007/s00066-019-01452-7>
 23. van Timmeren JE, Carvalho S, Leijenaar RTH et al (2019) Challenges and caveats of a multi-center retrospective radiomics study: an example of early treatment response assessment for NSCLC patients using FDG-PET/CT radiomics. *PLoS ONE* 14(6):e217536. <https://doi.org/10.1371/journal.pone.0217536>
 24. Cheung BMF, Lau KS, Lee VHF et al (2021) 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* 39(4):254–264. <https://doi.org/10.3857/roj.2021.00311>
 25. Deodato F, Macchia G, Cilla S et al (2019) Dose escalation in extracranial stereotactic ablative radiotherapy (DESTROY-1): A multiarm Phase I trial. *Br J Radiol* 91:20180422. <https://doi.org/10.1259/bjr.20180422>
 26. Therasse P, Arbuck SG, Eisenhauer EA et al (2000) New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 92(3):205–216. <https://doi.org/10.1093/jnci/92.3.205>
 27. Van Griethuysen JJ, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. *Can Res* 77(21):e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
 28. Zwanenburg A, Vallieres M, Abdalah MA et al (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 295(2):328–338. <https://doi.org/10.1148/radiol.2020191145>
 29. de Sousa VML, Carvalho L (2018) Heterogeneity in lung cancer. *Pathobiology* 85(1–2):96–107. <https://doi.org/10.1159/000487440>
 30. Ferreira Junior JR, Koenigkam-Santos M, de Vita Graves C et al (2021) Quantifying intratumor heterogeneity of lung neoplasms with radiomics. *Clin Imaging* 74:27–30. <https://doi.org/10.1016/j.clinimag.2020.12.017>
 31. Dou TH, Coroller TP, van Griethuysen JJM et al (2018) Peritumoral radiomics features predict distant metastasis in locally advanced NSCLC. *PLoS ONE* 13(11):e206108. <https://doi.org/10.1371/journal.pone.0206108>
 32. Voulaz E, Novellis P, Rossetti F et al (2020) Distinguishing multiple lung primaries from intra-pulmonary metastases and treatment implications. *Expert Rev Anticancer Ther* 20:985–995. <https://doi.org/10.1080/14737140.2020.1823223>
 33. Avanzo M, Stancanello J, Pirrone G et al (2020) Radiomics and deep learning in lung cancer. *Strahlenther Onkol* 196:879–887. <https://doi.org/10.1007/s00066-020-01625-9>
 34. Davey A, van Herk M, Faivre-Finn C et al (2019) Is tumour sphericity an important prognostic factor in patients with lung cancer? *Radiother Oncol* 143:73–80. <https://doi.org/10.1016/j.radonc.2019.08.003>
 35. Salguero FJ, Belderbos JSA, Rossi MMG et al (2013) Microscopic disease extensions as a risk factor for loco-regional recurrence of NSCLC after SBRT. *Radiother Oncol* 109:26–31. <https://doi.org/10.1016/j.radonc.2013.08.028>
 36. Coroller TP, Agrawal V, Narayan V et al (2016) Radiomic phenotype features predict pathological response in non-small cell lung cancer. *Radiother Oncol* 119:480–486. <https://doi.org/10.1016/j.radonc.2016.04.004>
 37. Chong Y, Kim JH, Lee HY et al (2014) Quantitative CT variables enabling response prediction in neoadjuvant therapy with EGFR-TKIs: are they different from those in neoadjuvant concurrent chemoradiotherapy? *Plos One* 9:e88598. <https://doi.org/10.1371/journal.pone.0088598>
 38. Caruso D, Zerunian M, Daffina J et al (2021) Radiomics and functional imaging in lung cancer: the importance of radiological heterogeneity beyond FDG PET/CT and lung biopsy. *Eur J Radiol* 142:109874. <https://doi.org/10.1016/j.ejrad.2021.109874>
 39. Weiss GJ, Ganeshan B, Miles KA et al (2014) Noninvasive image texture analysis differentiates Kras mutation from pan-wildtype NSCLC and is prognostic. *Plos One* 9:e100244. <https://doi.org/10.1371/journal.pone.0100244>
 40. Guan JL, Zhong WZ, An SJ et al (2013) KRAS mutation in patients with lung cancer: a predictor for poor prognosis but not for EGFR-TKIs or chemotherapy. *Ann Surg Oncol* 20:1381–1388. <https://doi.org/10.1245/s10434-012-2754-z>
 41. Wang M, Han J, Marcar L et al (2017) Radiation resistance in KRAS mutated lung cancer is enabled by stem-like properties mediated by an osteopontin-EGFR pathway. *Cancer Res* 77:2018–2028. <https://doi.org/10.1158/0008-5472.CAN-16-0808>
 42. Muenzel D, Engels HP, Bruegel M et al (2012) Intra- and inter-observer variability in measurement of target lesions: implication on response evaluation according to RECIST 1.1. *Radiol Oncol* 46(1):8–18. <https://doi.org/10.2478/v10019-012-0009-z>
 43. McErlan A, Panicek DM, Zabor EC et al (2013) Intra- and inter-observer variability in CT measurements in oncology. *Radiology* 269(2):451–459. <https://doi.org/10.1148/radiology.13122665>

44. Garau N, Paganelli C, Summers P et al (2020) External validation of radiomics-based predictive models in low-dose CT screening for early lung cancer diagnosis. *Med Phys* 47(9):4125–4136. <https://doi.org/10.1002/mp.14308>

author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the