ORIGINAL ARTICLE

# Imaging-Based Biomarkers Predict Programmed Death-Ligand 1 and Survival Outcomes in Advanced NSCLC Treated With Nivolumab and Pembrolizumab: A Multi-Institutional Study

Sevinj Yolchuyeva, PhD,[a] Elena Giacomazzi, MSc,[a] Marion Tonneau, PhD,[b,c] Fabien Lamaze, PhD,[d] Michele Orain, MT,[d] François Coulombe, MD,[d] Julie Malo, CCRP,[b] Wiam Belkaid, PhD,[b] Bertrand Routy, MD, PhD,[b,e] Philippe Joubert, MD, PhD,[d,f] Venkata S. K. Manem, MSc, PhD[a,d,g,*]

[a]*Department of Mathematics and Computer Science, Université du Québec à Trois Rivières, Quebec, Canada*
[b]*Centre de Recherche du Centre Hospitalier Universitaire de Montréal, Quebec, Canada*
[c]*Université de Médecine de Lille, Lille, France*
[d]*Quebec Heart & Lung Institute Research Center, Quebec, Canada*
[e]*Centre Hospitalier Universitaire de Montréal, Hemato-Oncology Service, Quebec, Canada*
[f]*Department of Molecular Biology, Medical Biochemistry and Pathology, Laval University, Quebec, Canada*
[g]*Centre de Recherche du CHU de Québec - Université Laval, Quebec, Canada*

Received 31 May 2023; revised 18 October 2023; accepted 8 November 2023
Available online - 18 November 2023

## ABSTRACT

**Background:** Although the immune checkpoint inhibitors, nivolumab and pembrolizumab, were found to be promising in patients with advanced NSCLC, some of them either do not respond or have recurrence after an initial response. It is still unclear who will benefit from these therapies, and, hence, there is an unmet clinical need to build robust biomarkers.

**Methods:** Patients with advanced NSCLC (N = 323) who were treated with pembrolizumab or nivolumab were retrospectively identified from two institutions. Radiomics features extracted from baseline pretreatment computed tomography scans along with the clinical variables were used to build the predictive models for overall survival (OS), progression-free survival (PFS), and programmed death-ligand 1 (PD-L1). To develop the imaging and integrative clinical-imaging predictive models, we used the XGBoost learning algorithm with ReliefF feature selection method and validated them in an independent cohort. The concordance index for OS, PFS, and area under the curve for PD-L1 was used to evaluate model performance.

**Results:** We developed radiomics and the ensemble radiomics-clinical predictive models for OS, PFS, and PD-L1 expression. The concordance indices of the radiomics model were 0.60 and 0.61 for predicting OS and PFS and area under the curve was 0.61 for predicting PD-L1 in the validation cohort, respectively. The combined radiomics-clinical model resulted in higher performance with 0.65, 0.63, and 0.68 to predict OS, PFS, and PD-L1 in the validation cohort, respectively.

**Conclusions:** We found that pretreatment computed tomography imaging along with clinical data can aid as predictive biomarkers for PD-L1 and survival end points. These imaging-driven approaches may prove useful to expand the therapeutic options for nonresponders and improve the selection of patients who would benefit from immune checkpoint inhibitors.

***Corresponding author.**

***Disclosure: The authors declare no conflict of interest.***

**Address for correspondence: Venkata S. K. Manem, MSc, PhD, Centre de Recherche du CHU de Québec - Université Laval, 9 Rue McMahon, Quebec, QC G1R 3S3, Canada. E-mail: venkata.manem@crchudequebec.ulaval.ca**

**Cite this article as: Yolchuyeva S, Giacomazzi E, Tonneau M, et al. Imaging-based biomarkers predict programmed death-ligand 1 and survival outcomes in advanced NSCLC treated with nivolumab and pembrolizumab: a multi-institutional study. *JTO Clin Res Rep*. 2023;4:100602.**



**ISSN: 2666-3643**

**https://doi.org/10.1016/j.jtocrr.2023.100602**

*Keywords:* Radiomics; Machine learning; Immunotherapy; Biomarker discovery

## Introduction

Immune checkpoint inhibitors (ICIs) have modified the therapeutic landscape to treat several aggressive cancer types, including NSCLC. Compared with conventional cytotoxic therapies, ICIs targeting programmed cell death protein-1 (PD-1) significantly extend the overall survival (OS) in patients with a broad variety of malignancies.[1–3] ICIs have now become the standard of care for patients presented with advanced NSCLC. In addition, the recent Food and Drug Administration approvals of nivolumab and pembrolizumab represent the beginning of a significant paradigm shift for treating patients with NSCLC.[4,5] Although there have been several clinical trials and studies that revealed the survival benefits of ICIs than conventional therapies,[2–7] a subset of patients do not respond to these therapies or have a relapse of cancer after an initial response. So far, it is unclear as to who will benefit from these therapies or what are the mechanisms driving the treatment failure. Importantly, the primary resistance to ICIs remains unpredictable and exceeds 60%, whereas secondary resistance rates approach 100%, highlighting the need to develop novel predictive biomarkers.

Programmed death-ligand 1 (PD-L1) is the only approved test to guide a patient's treatment with ICIs, and currently it affects the management of cancer. Although clinical trials have found that PD-L1 expression above 50% is associated with increased likelihood of response to checkpoint inhibitors, it neither guarantees durable response in patients with a high PD-L1 tumor expression nor eliminates the possibility of response in tumors with low PD-L1 expression. There have been efforts from the scientific community to develop novel predictive biomarkers to identify and select patients with advanced NSCLC who will receive clinical benefit from ICIs. Building a robust and reproducible data-driven biomarker can potentially spare these patients from a prolonged administration of these expensive compounds and from unnecessary toxicities.

In recent years, computational imaging approaches, such as radiomics, have achieved success in quantifying the characteristics of tumors using medical imaging data. Radiomics is a high-throughput computational tool that aims to provide a deeper understanding into the subvisual characteristics of a tumor.[8,9] Radiomics-based techniques have a number of benefits, including the fact that they are noninvasive and the features can be derived from routine medical images.[10,11] These make them ideal for translating to clinical practice as the method does not involve invasive procedures to predict or monitor a patient's response to various therapeutic interventions. Studies in the literature have found that radiomics can be used to build diagnostic or predictive biomarkers in a variety of malignancies, including advanced solid tumors, melanoma, and other types of NSCLC.[12–14] There have been several research efforts in the literature that focused on building imaging-based biomarkers leveraging medical imaging data.

In a study by Ligero et al.,[15] the authors developed a radiomics signature from baseline (pretreatment) computed tomography (CT) scans to predict response to anti–PD-1 and PD-L1 in patients with advanced solid tumors. Zhao et al.[16] developed a radiomics and a combined radiomics-clinical model to predict PD-L1 in patients with NSCLC using positron emission tomography imaging. In another retrospective study done by Liu et al.,[17] the authors developed a radiomics model to predict the clinical outcomes of patients with advanced NSCLC treated with nivolumab. Three machine learning classifiers were trained and tested on a relatively small data set with only 46 cases. Although the leave-one-out cross-validation was applied for building the classifiers to avoid biases, the robustness and effectiveness of their model still need to be evaluated on larger data sets. To predict OS and progression-free survival (PFS) in patients with NSCLC treated with first-line pembrolizumab, Zerunian et al.[18] developed a CT-derived radiomics signature. Nevertheless, their model was validated on a sample size of 21 patients, and a study with larger cohorts was needed to confirm the results in a clinical setting. In the study of Franzese et al.,[19] radiomics and clinical features were investigated for their predictive value relative to locoregional failure, PFS, OS, and the built multivariate models. They used five radiomic features along with two clinical features significantly predictive for both PFS and OS. The limitation of their study is that they built the models using a small cohort of patients from a single institution and did not separate their data into training and testing subgroups, thus affecting both the robustness and reproducibility of the results.

Studies in literature have revealed the utility of radiomics as a promising noninvasive biomarker strategy for patients with advanced NSCLC amenable to ICIs, although most of them have been based on a single-center retrospective study, or the sample size was small to validate the predictive models. Moreover, none of these studies have attempted to build survival and

PD-L1–predictive models integrating imaging and clinical features from a multicenter perspective to predict PFS, OS, and PD-L1 in patients treated with ICIs nivolumab and pembrolizumab. On the basis of this premise, in this study, we aimed to develop radiomics and ensemble radiomics-clinical models for predicting OS, PFS, and PD-L1. We used two independent cohorts from different institutions to develop and validate the radiomics biomarkers, which will make them more generalizable and clinically relevant.

## Materials and Methods

### *Description of Cohorts*

This retrospective study included 323 patients presented with advanced NSCLC and treated with nivolumab or pembrolizumab between 2015 and 2021. The cohorts were obtained from two thoracic oncology reference centers, the Institut Universitaire de Cardiologie et de Pneumologie de Québec (Quebec Heart and Lung Institute, IUCPQ) and the Centre Hospitalier Universitaire de Montréal (CHUM). The samples used in this study came from the IUCPQ's Quebec Respiratory Health Network Tissue Bank (https://rsr-qc.ca/biobanque/). A research consent form was obtained from each patient, and the study was approved by the institutional review board at the two academic institutions where patient data were acquired (MP-10-2020-3397/CÉR CHUM: 19.390). All patients with advanced NSCLC treated with ICIs and who had a pre-ICI CT scan were eligible for retrospective review. Response Evaluation Criteria in Solid Tumors version 1.1 was used to assess tumor response. The progression was defined by the qualitative evaluation of the radiologist. All patients were followed until death or until the date of censoring at the last time the subject was known to be alive, that is, in January 2022. To assess disease progression after treatment initiation, at least one breath-hold chest pre-ICI CT scan (2 mo before treatment) and one post-treatment CT scan (every 8 wk for nivolumab and every 9 wk for pembrolizumab) were available. The two centers' data set included 323 patients in total, with 174 patients from the discovery cohort (CHUM) and 149 patients from the in-house validation cohort (IUCPQ) who were administered with pembrolizumab or nivolumab.

### *PD-L1, PFS, and OS Assessment*

On the surface of activated T-cells, PD-1 is expressed and down-regulates the T-cell activity by binding to its ligands, PD-L1 and PD-L2.[20,21] Immunohistochemistry labeling (Dako Autostainer) was used to assess the PD-L1–positive tumor cells (or tumor proportion score [TPS]) with the 22C3 clone as part of standard patient management after lung cancer diagnosis (pharmDx kit). To calculate PD-L1 expression, the TPS value was used, expressed as a percentage of tumor cells with positive membranous staining ranging from 0% to 100%. According to the clinical cutoff for PD-L1 TPS, each tumor was reclassified as 1%, 1% to 49.9%, or more than 50%. In this study, we combined the categories 1% and 1% to 49% into a singular class. With this, we ended up with two groups, namely, PD-L1 less than 50% and PD-L1 more than or equal to 50%, and considered it as a binary classification task.

The period of time between the start of treatment and the point at which the disease progresses is known as PFS. It is determined by the number of days between the start of treatment and the date of illness progression, death from any cause, or the final follow-up (censored). Imaging techniques such as CT, magnetic resonance imaging, or positron emission tomography/CT are frequently used to confirm the progression of the disease. OS is the period of time from the date of diagnosis to the date of death from any cause or to the date of censoring at the last time the subject was known to be alive. We considered the survival end points, OS and PFS, as a regression-based task.

### *CT Scan Annotation*

A radiation oncologist or radiologist manually annotated each primary lesion on deidentified CT scans obtained 2 months before immunotherapy administration. To extract the radiomics features, the region of interest (ROI) was determined as follows: (1) After CT scan alignment, mathematically-based denoising followed by the chest segmentation was used for the connected regions; (2) the contour for the lung was roughly segmented and after skin boundary detection, the pulmonary parenchyma was refined; and finally, (3) the lung nodule ROI was identified, which was independent of the size, position, and spreading near or through the pleura because of the relative symmetry of the lung.

### *Extraction of Radiomics Features and Preprocessing*

PyRadiomics (version 3.0.1),[22] an open-source Python tool for extracting custom radiomics characteristics from CT scan data region of interests, was used. These are interpretable engineered features of the tumor that comply with the image biomarker standardization initiative. All scans' slice thicknesses were interpolated to voxel sizes of $1 \times 1 \times 1$ mm$^3$, and features were computed in three dimensional. The radiomics characteristics were extracted and categorized into four primary categories in a traditional pipeline using

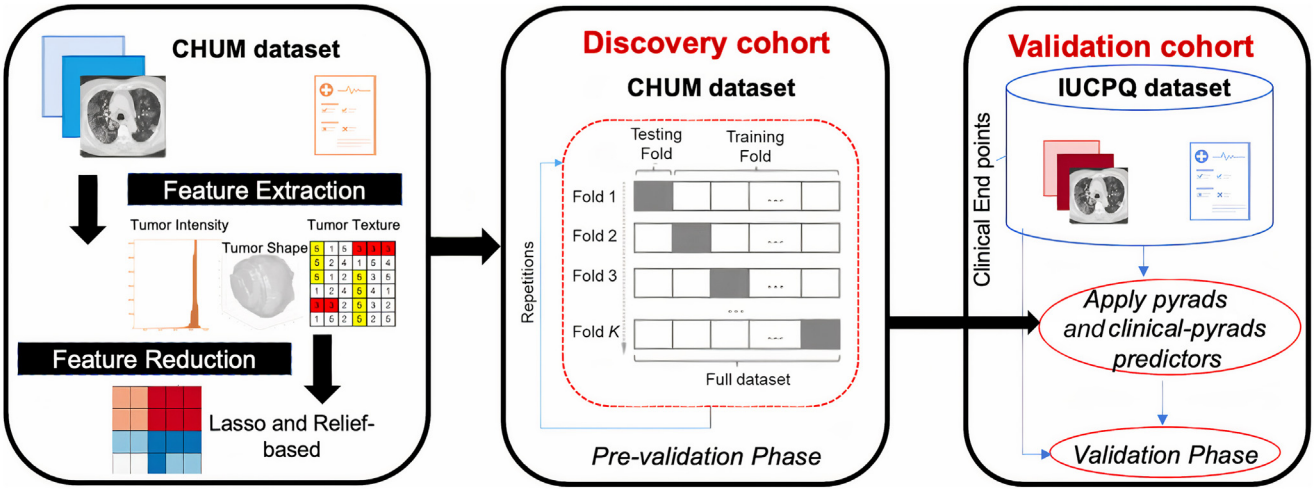


**Figure 1.** Analysis pipeline of the study. CHUM, Centre Hospitalier Universitaire de Montréal; IUCPQ, Institut Universitaire de Cardiologie et de Pneumologie de Québec (Quebec Heart and Lung Institute).

PyRadiomics, including tumor intensity-based, shape-based, texture-based, and wavelet-based features. Intensity-based features are first-order statistical features that quantify tumor intensity characteristics on the basis of a histogram of all voxel intensity values. Shape-based characteristics, such as sphericity or compactness, describe the shape of the tumor. Texture-based characteristics describe the texture changes within the tumor volume. This was accomplished by the use of a gray-level co-occurrence matrix, a run-length gray-level matrix, or a gray-level size-zone matrix to cluster voxels with similar appearance. After applying a Laplace of Gaussian transformation to the image, wavelet-based features were derived on the intensity and textural data.

### *Model Development*

Several preprocessing operations were carried out before the feature selection, and machine learning procedures were implemented. Standard-Scaler from Sklearn was used to normalize the entire data set.[23] After fitting the scalar to the data, the values are changed so that the mean is 0 and the SD is 1. To construct imaging-based prediction models for all tasks, we used the ReliefF-based feature selection method, which is frequently used to identify the most relevant features.[24,25] This method is based on the original Relief-based algorithm.[24] The main idea behind ReliefF is to find the most relevant features by estimating their quality using the feature-weighting scheme.[26] The ReliefF has a few advantages over other feature selection methods. First, it is simple and efficient, and it can handle large data sets with a large number of features. Second, it is sensitive to both linear and nonlinear interactions between features and can identify feature dependencies that might not be captured by other methods. Finally, it is robust to noise and missing values, as it only considers the differences between instances and their neighbors rather than the actual values of the features.

To build the predictive model, we used the well-known XGBoost machine learning algorithm. XGBoost was first introduced by Chen and Guestrin,[27] which stands for "eXtreme Gradient Boosting." It is an ensemble approach that is based on the gradient boosting algorithm.[28,29] The gradient boosting algorithm was improved to make the predictive model work better and to cut down on the computational time by making it run in parallel. Extreme is used to describe how much better it works than the original gradient boosting algorithm. By putting together several weak models, such as decision trees, ensemble methods like gradient boosting make stronger models. In XGBoost, errors in existing models are fixed by adding new models. This makes the overall performance better. XGBoost is easy to use, has low computational costs, and is known to be good at making predictions.[30–32]

### *Analysis Framework*

Figure 1 illustrates the analysis pipeline for developing the radiomics model and the integrated radiomics-clinical model using the feature selection method and machine learning approach. The CHUM data set was used as the discovery cohort, whereas the IUCPQ data set was used for validation. On the discovery data set, we first performed the prevalidation phase, which consisted of 10 iterations/repetition of fivefold cross-validation for the predictive model presented previously. The performance of the model was evaluated using the concordance index (C-index) and the area under the receiver operating characteristic curve (AUC). The C-index is

defined as the probability that two variables will rank a random pair of samples in the same order. A random predictor would yield a C-index of 0.5, whereas a perfect predictor yields a C-index of 1. In simple words, it is a generalization of the AUC.

### Radiomics Model Development

In general, we performed the following steps to develop the predictive models leveraging imaging and clinical data:

- The empty rows were removed from CHUM and IUCPQ cohort.
- Data standardization: All variables were required to have a mean of zero and a SD of one for both cohorts.
- We used Lasso regression for the OS and PFS tasks and Lasso Logistic regression with penalty "l1" for the PD-L1 challenge, as our first-step feature selection approach. Using the primary feature selection Relief and the XGBoost machine learning approach independently, Lasso and Lasso Logistic regression are applied to the entire training cohort. Nevertheless, we used grid search with fivefold cross-validation to determine the best alpha for Lasso regression and C for Logistic regression. After training Lasso on the entire CHUM cohort, we eliminated features with zero coefficients.
- We incrementally chose features for the ReliefF feature selection method and XGBoost machine learning method from the generated features (including those maintained after the LASSO). This was carried out in the discovery data set's prevalidation phase, which consisted of fivefold cross-validation with 10-times repetitions. For each fold, the ReliefF feature selection approach was used at this stage.
- The highest C-index for OS, PFS, and AUC for PD-L1 was used to determine the best number of features using Relief-based method.
- GridSearchCV from SciKit-Learn was used to conduct a cross-validation for XGBoost's hyperparameter tuning. Scikit-learn version 1.0.2, Python version 3.9.13, XGBoost version 1.6.2, and Skrebate version 0.62 were used to implement the ReliefF feature selection and the XGBoost regressor and classifier, respectively.
- We used the final model with the best features and best hyperparameters to test the model's performance on the validation data set.

### Radiomics-Clinical Model

After the selection of the best radiomics features with the highest C-index for OS, PFS, and AUC for PD-L1, we integrated the statistically relevant clinical features. We combined Eastern Cooperative Oncology Group (ECOG) and sex for model building of OS/PFS and first-line treatment for the PD-L1 task. The model's performance was evaluated on the discovery data set using fivefold cross-validation with 10-times repeats. In other words, the radiomics-clinical model was built using both radiomic and significant clinical variables to predict continuous values of OS, PFS, and binary values for PD-L1 expression.

### Survival Analysis

The prognostic value of radiomics-only signatures was assessed using a log-rank test for Kaplan-Meier survival curves, as implemented in the survcomp R package. Median value was used to split the data into two groups, whereas OS and relapse-free survival were used as the end points.

## Results

### Patient Characteristics

Table 1 displays the clinical characteristics of the two cohorts used in the current study. Continuous data were presented as mean ± SD, whereas categorical data were presented as counts and percentages. For OS and PFS, the CHUM discovery cohort consists of 174 patients with a mean age of 66.2 years (9.3). Of the patients in the smoking category, 63% were former smokers, whereas 30% are current smokers. The IUCPQ validation cohort consists of 149 patients with a mean age of 67.7 years (7.3). Of these patients, 70.5% were former smokers, whereas 24.2% are current smokers. For PD-L1, the CHUM discovery cohort consists of 146 patients with a mean age of 66.6 years (9.4). Of the patients in the smoking category, 64.4% were former smokers, whereas 30.2% are current smokers. The IUCPQ validation cohort consists of 121 patients with a mean age of 67.7 years (7.0). Of these patients, 69.4% were former smokers, whereas 24.8% are current smokers. The number of nonsmoker patients for PD-L1, PFS, and OS was less than 10% in both cohorts.

We retrospectively selected data from patients who had been treated with at least one cycle of nivolumab or pembrolizumab at the CHUM and IUCPQ centers. The number of patients was 90 and 84 who were treated with nivolumab or pembrolizumab, respectively, on the CHUM cohort, and 82 and 67 on the IUCPQ cohort for the OS and PFS task. For PD-L1, 83 patients were treated with nivolumab and 63 patients were treated with pembrolizumab in the CHUM cohort; 81 patients were treated with nivolumab and 40 patients were treated with pembrolizumab in the IUCPQ cohort. The range of OS values was 0.3 to 67.3 months in the CHUM data set and 0.4 to 59.4 months in the IUCPQ data set. In addition, the range of PFS values was 0.13 to 62.1 months in the

**Table 1.** Clinical Characteristics of Discovery and Validation Cohorts

| Clinical Features | OS and PFS | | PD-L1 | |
|---|---|---|---|---|
| Characteristics | CHUM (Discovery) | IUCPQ (Validation) | CHUM (Discovery) | IUCPQ (Validation) |
| No. of samples | 174 | 149 | 146 | 121 |
| Age (mean) | 66.2 ± 9.3 | 67.7 ± 7.3 | 66.6 ± 9.4 | 67.7 ± 7.0 |
| Sex, n (%) | | | | |
| Female | 87 (50.0) | 73 (49) | 74 (50.6) | 63 (52.0) |
| Male | 87 (50.0) | 76 (51) | 72 (49.4) | 58 (48.0) |
| Smoking status, n (%) | | | | |
| Former | 110 (63.0) | 105 (70.5) | 94 (64.4) | 84 (69.4) |
| Current | 52 (30.0) | 36 (24.2) | 44 (30.1) | 30 (24.8) |
| Never | 12 (7.0) | 8 (5.4) | 8 (5.5) | 7 (5.8) |
| ECOG status, n (%) | | | | |
| 0 | 49 (28.2) | 44 (29.5) | 45 (48.6) | 37 (30.5) |
| 1 | 89 (51.1) | 94 (63.0) | 71 (30.8) | 75 (62.0) |
| 2 | 32 (18.4) | 8 (5.4) | 26 (17.8) | 6 (5.0) |
| 3 | 4 (2.3) | 3 (2.1) | 4 (2.7) | 3 (2.5) |
| First-line treatment, n (%) | | | | |
| No | 120 (69.0) | 94 (63.1) | 92 (63.0) | 67 (55.3) |
| Yes alone | 54 (31.0) | 48 (32.2) | 54 (37.0) | 48 (39.7) |
| Yes combined | - | 7 (4.7) | - | 6 (5.0) |
| Subtype group, n (%) | | | | |
| Adenocarcinoma | 140 (80) | 112 (75) | 118 (81) | 91 (75.2) |
| Squamous | 26 (15) | 19 (12.8) | 20 (13.8) | 14 (11.6) |
| Other | 8 (5) | 18 (12.2) | 8 (5.2) | 16 (13.2) |
| IO type | | | | |
| Nivolumab | 90 (52) | 82 (55) | 83 (57) | 81 (66) |
| Pembrolizumab | 84 (48) | 67 (45) | 63 (43) | 40 (34) |
| OS (mean) | 18.4 mo | 21.1 mo | - | - |
| PFS (mean) | 13.4 mo | 13.1 mo | - | - |
| PD-L1 expression, n (%) | | | | |
| <50% | | | 57 (39) | 40 (33) |
| >50% | - | - | 89 (61) | 81 (67) |

CHUM, Centre Hospitalier Universitaire de Montréal; ECOG, Eastern Cooperative Oncology Group; IO, immunotherapy; IUCPQ, Institut Universitaire de Cardiologie et de Pneumologie de Québec; OS, overall survival; PD-L1, programmed death-ligand 1; PFS, progression-free survival.

CHUM data set and 0.13 to 57.7 months in the IUCPQ data set.

### *Univariate Analysis of Clinical Variables*

The clinical variables used in this study were the following: age, sex, ECOG status, smoking status, subtype group, and first-line treatment of patients. To evaluate the association between OS and PFS and PD-L1 and clinical factors, we performed univariate analyses using Pearson correlation for continuous values independently. The *p* values resulting from the univariate analysis are presented in Table 2. From our analysis, we found sex and ECOG status to be clinically significant variables associated with the OS and PFS as continuous values. First-line treatment is a clinically significant variable associated with PD-L1 as a continuous value.

### *Radiomics Features*

Using the PyRadiomics pipeline, a total of 851 radiomics features were extracted from the segmented tumor areas of the pretreatment CT scans of two distinct cohorts to develop predictive noninvasive radiomics biomarkers. The CHUM cohort was used for feature selection and model training, whereas the IUCPQ cohort was used to evaluate the performance of the model for the end points OS, PFS, and PD-L1.

**Table 2.** Univariate Analysis Between Clinical Characteristics and OS, PFS, and PD-L1

| | OS | PFS | PD-L1 |
|---|---|---|---|
| Characteristics | *p* Value | *p* Value | *p* Value |
| Age | 0.89 | 0.93 | 0.89 |
| Sex | 0.03 | 0.02 | 0.34 |
| Smoking status | 0.44 | 0.43 | 0.25 |
| ECOG status | 1e-6 | 0.001 | 0.008 |
| First-line treatment | 0.29 | 0.27 | 0.000004 |
| Subtype group | 0.21 | 0.42 | 0.11 |

ECOG, Eastern Cooperative Oncology Group; OS, overall survival; PD-L1, programmed death-ligand 1; PFS, progression-free survival.

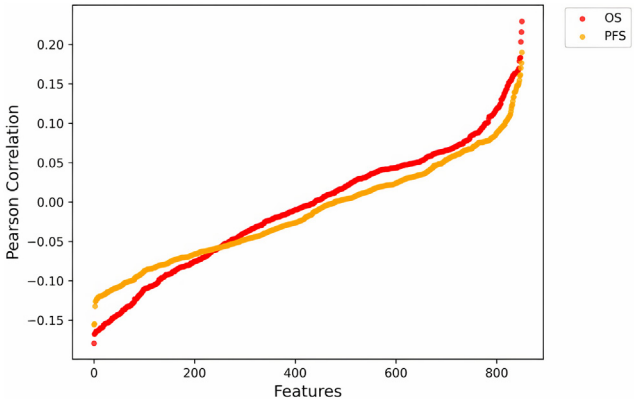


**Figure 2.** Pearson correlation of radiomics features with OS and PFS in the discovery cohort. OS, overall survival; PFS, progression-free survival.

### *Association of Radiomics Features With Survival End Points*

We used the Pearson correlation to compute the association between radiomics features and the clinical end points OS and PFS in the discovery cohort (Fig. 2). The highest correlation values were 0.22 and 0.18 and the lowest correlation values were −0.18 and −0.16 for OS and PFS, respectively. Overall, we found that 412 and 372 features were associated in the positive direction with the end points, whereas 439 and 479 features were associated negatively with OS and PFS, respectively. We found that for both OS and PFS, the top 10 positively correlated features were from wavelet-based group and the top 10 negatively correlated features were from the wavelet-based group (nine of them) and one from the texture-based group.

### *Model Evaluation*

In this section, we present the results of the three objectives presented in this study. After applying the LASSO feature reduction method, features with zero coefficients were removed and we ended up with 21 features for OS, 22 features for PFS, and 35 features for the PD-L1 task.

The performance of the radiomics and radiomics-clinical models for OS is illustrated in Figure 3. Figure 3*A* illustrates the C-index values obtained from the discovery cohort, and Figure 3*B* illustrates the C-index values obtained during the validation phase. To build the radiomics-clinical integrative model, we used only the significant features (from Table 2), that is, sex and ECOG status. The final set of features to predict the OS was all from the wavelet-based group. The radiomics model resulted in approximately 0.599 in the discovery set and approximately 0.554 in the validation cohort. With the ensemble radiomics-clinical model, the C-index was enhanced from 0.599 to 0.651 in the discovery cohort and from 0.554 to 0.591 in the validation cohort. Overall, we found that the radiomics-clinical model resulted in a better model performance than radiomics model alone.

Figure 4 presents the performance of the radiomics and the ensemble radiomics-clinical models to predict the PFS. The panels (Fig. 4*A*) and (Fig. 4*B*) present the C-index in the discovery and validation cohorts, respectively. For developing the radiomics-clinical predictive model, we utilized the significantly associated features with the PFS (from Table 2), i.e., sex and ECOG status.

Similar to the OS, the C-index was slightly improved with the integrated radiomics-clinical model compared with the baseline radiomics model. The model accuracy improved from 0.608 to 0.63 in the discovery cohort and from 0.563 to 0.595 in the validation cohort. The final set of features to predict the PFS was a combination from the wavelet-, texture-, and intensity-based groups.

The performance of the radiomics and the integrated radiomics-clinical model for predicting the PD-L1 expression is presented in Figure 5. To develop the radiomics-clinical predictive model, we used only those features that were significantly associated with the PD-L1 expression (from Table 2), that is, first-line treatment and ECOG status. The AUC was significantly improved with the ensemble radiomics-clinical model in comparison with the radiomics model alone—increased from 0.62 to 0.78 for the discovery cohort (Fig. 5*A*) and from 0.61 to 0.696 (Fig. 5*B*) for the validation cohort. Overall, our results indicate that the radiomics-clinical model has a better predictive ability compared with radiomics model alone and captures a unique fingerprint by leveraging routine CT scans and clinical data. Importantly, this will help the clinicians to expand the treatment horizons toward greater precision to treat patients with NSCLC with ICI therapies.

### *Prognostication of Radiomics Signatures*

Furthermore, the prognostic value of imaging-derived signatures was assessed using a log-rank test using Kaplan-Meier survival curves (Fig. 6). The left and right panels present the results for OS (Fig. 6*A*) and PFS (Fig. 6*B*), respectively. The group with lower than median (OS or PFS) value has poor outcome, whereas the group with higher than median (OS or PFS) value has good outcome. We obtained significant values for both the survival end points with 0.00022 and 0.013, respectively. Through these findings, we believe that the developed imaging-based models have clinical impact and can stratify patients into low- and high-survival groups.

## Discussion

Despite the fact that ICIs have changed the treatment landscape of many tumor types, response rates remain

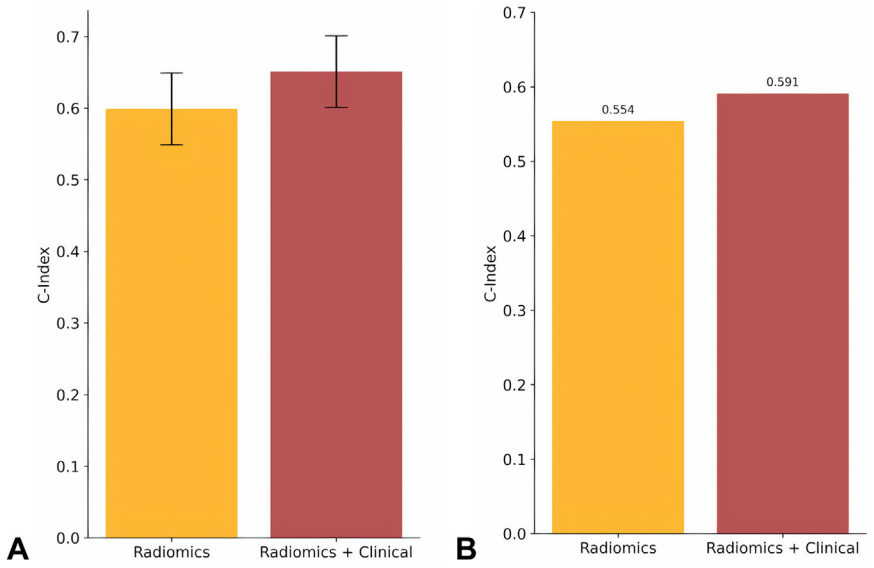


**Figure 3.** Performance of radiomics and the ensemble radiomics-clinical models to predict the OS. (*A*) C-index in the discovery cohort and (*B*) C-index in the validation cohort. C-index, concordance index; OS, overall survival.

suboptimal, and better patient stratification is required. The activation of the immune system can lead to durable responses and better response in some but not in all patients with advanced NSCLC. In addition, the recent Food and Drug Administration approvals of nivolumab and pembrolizumab present a paradigm shift in the treatment fabric for patients with advanced NSCLC.[4,5] In this regard, there is a pressing clinical need to identify patients who are most likely to respond to ICIs and have better survival outcomes. Furthermore, predicting those unresponsive patients to ICIs early may allow the physicians to choose other effective therapeutic interventions and improve their survival, while reducing the costs and toxicities associated with these compounds. Although there have been efforts to build data-driven biomarkers for ICIs, there is a dire need to build robust predictive models for clinical translation. The recent advent of radiomics through quantitative image analysis has been gaining interest in oncology as a novel strategy for predicting treatment response.[33–36] Nevertheless, the development of imaging-based signatures that can be robust and generalizable across academic centers has been a bottleneck to adopt radiomics in clinical practice.

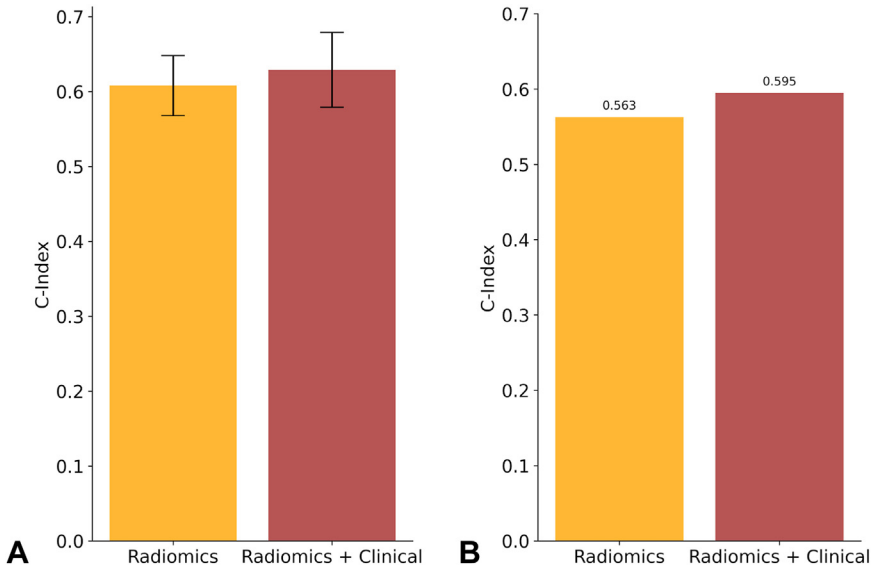


**Figure 4.** Performance of radiomics and the ensemble radiomics-clinical models to predict the PFS. (*A*) C-index in the discovery cohort and (*B*) C-index in the validation cohort. C-index, concordance index; PFS, progression-free survival.

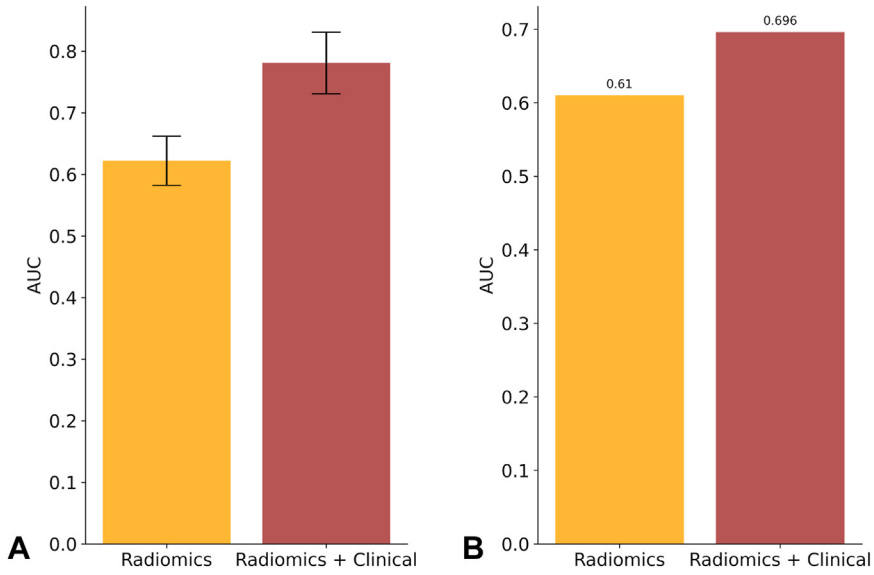


**Figure 5.** Performance of radiomics and the ensemble radiomics-clinical models to predict PD-L1. (*A*) AUC in the discovery cohort and (*B*) AUC in the validation cohort. AUC, area under the curve; PD-L1, programmed death-ligand 1.

Most of the studies in literature have revealed the utility of radiomics as a biomarker strategy for patients with NSCLC amenable to ICIs. Despite these promising retrospective studies, the adoption of imaging-based models in clinical workflows is limited. In a study led by Zerunian et al.,[18] the authors built the predictive models for PFS and OS using 21 patients from a single institution who were treated with pembrolizumab only. Despite obtaining an AUC of close to 0.72, major limitations of this study include the sample size and the potential biases in using only texture features. Considering the interpretation of the texture features, this reflects the tumor aggressiveness rather than biomarkers of response to ICIs, thus, limiting the model's applicability. In another retrospective study, Liu et al.[17] developed radiomics model to predict survival outcomes in a cohort of 46 patients with NSCLC treated with nivolumab alone. The models were trained on a small data set with leave-out-one–fold cross-validation, suggesting a need to test for the robustness on diverse datasets. Finally, Trebeschi et al.[12] leveraged radiomics features to assess ICI response in patients with NSCLC and melanoma. Although their model was able to predict progressive disease in NSCLC with a good accuracy, it achieved poor performance on both pulmonary and hepatic melanoma lesions (AUC of 0.55). To assess the performance of radiomics models, they used 133 and 70 patients in the discovery and test sets from a single institution, lacking sufficient validation and generalizability. To overcome the above-mentioned limitations, in this study, we considered diverse data sets from two academic institutions to train and test the predictive models. To the

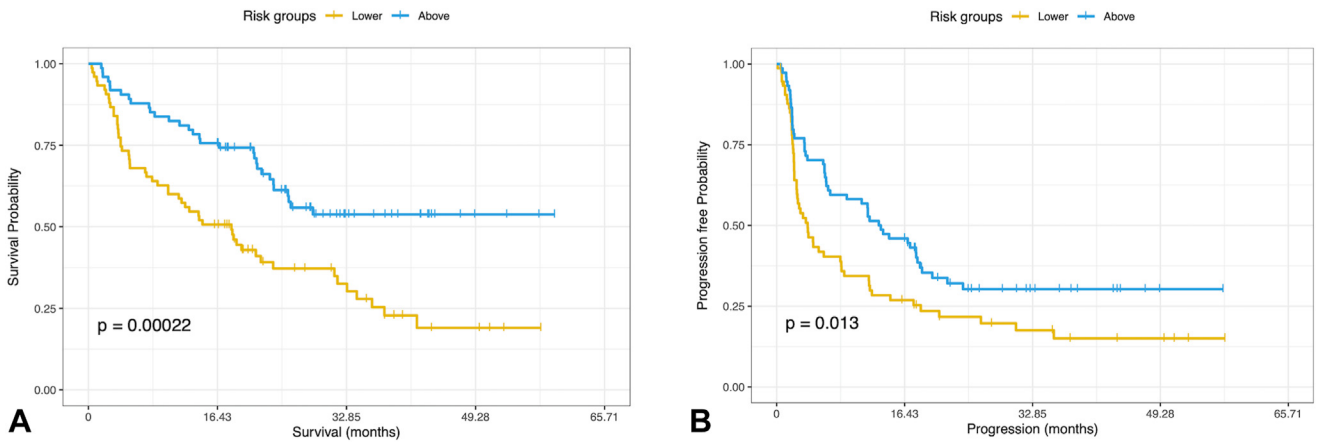


**Figure 6.** Survival analyses of clinical end points. (*A*) OS and (*B*) PFS. OS, overall survival; PFS, progression-free survival.

best of our knowledge, no study has been published on radiomics biomarkers predicting three clinical end points in patients with NSCLC treated with nivolumab and pembrolizumab using a large cohort. The significant number of patients treated with these two compounds and trained/evaluated in two institutions (n = 323) distinguishes our study from prior studies.

On the basis of this premise, we leveraged pretreatment CT imaging data along with the clinical variables to develop predictive biomarkers for three clinical end points—OS, PFS, and PD-L1 expression. To build the predictive models, we used a total of 323 patients across two institutions. The PyRadiomics platform was used to extract 851 radiomics features from the pre-ICI CT scan data, which were then used to develop the predictive models. To decrease the dimensionality of the features, we first used the LASSO feature selection strategy followed by the ReliefF feature selection method. The XGBoost machine learning method was used to build the models for all the clinical end points—OS, PFS, and PD-L1 expression. For building the ensemble radiomics-clinical model, we used only the clinically significant features. Clinical factors, ECOG, and sex were used for model building of OS/PFS; ECOG status and first-line treatment were used for the PD-L1 task. For OS, we found the C-index for the baseline radiomics model to be approximately 0.599 in the discovery set and 0.554 in the validation cohort. With the integrated radiomics-clinical model, the performance of the model improved from 0.599 to 0.651 on the discovery cohort and from 0.554 to 0.591 on the validation cohort. To predict PFS, the C-index of the radiomics model was found to be 0.608 and 0.563 in the discovery and validation cohorts, respectively. The C-index of the ensemble radiomics-clinical model significantly improved from 0.608 to 0.63 in the discovery data set and from 0.563 to 0.595 in the validation data set. For predicting PD-L1 expression, the AUC of the ensemble radiomics-clinical model significantly improved from 0.62 to 0.78 in the discovery data set and from 0.61 to 0.696 in the validation data set. We hypothesize that the model performance can be further improved by harmonizing image acquisition parameters across institutions.

Nevertheless, our study has its limitations. First, all samples were limited to primary tumors only and did not consider any patients with multiple lesions. Second, our models were validated on the retrospective cohort that may have caused potential biases in the cohort design. Last, few studies evaluated the potential impact of gray-level discretization and the need for harmonization of various imaging protocols arising from the two centers on the predictive model development. We hope to address these limitations in our future studies. In conclusion, our findings suggest that CT-based radiomics signatures from pretreatment ICI scans of patients with advanced NSCLC can potentially act as predictive biomarkers for PD-L1 expression and survival outcomes (for both OS and PFS). Through the developed biomarkers, clinicians can assess the treatment outcome for a given patient, accordingly prescribe alternative therapeutic interventions, and closely monitor those patients with a higher risk of cancer recurrence. Albeit, additional prospective validation of these noninvasive predictive biomarkers is required to precisely define their clinical translatability. Motivated by the results and the wide availability of routine CT scans for patients on cancer immunotherapy, we believe that this study is a step toward building robust and reproducible imaging-based models that could potentially be implemented in the clinical workflow.

## CRediT Authorship Contribution Statement

**Sevinj Yolchuyeva:** Formal analysis; Software; Methodology; Roles/Writing—original draft; Writing—review and editing.

**Venkata S.K. Manem:** Conceptualization; Supervision; Roles/Writing—original draft; Writing—review and editing.

**Marion Tonneau:** Data curation; Writing—review and editing.

**Fabien Lamaze:** Data curation; Writing—review and editing.

**Michele Orain:** Data curation; Writing—review and editing.

**François Coulombe:** Data curation; Writing—review and editing.

**Julie Malo:** Data curation; Writing—review and editing.

**Wiam Belkaid:** Data curation; Writing—review and editing.

**Bertrand Routy:** Data curation; Writing—review and editing.

**Philippe Joubert:** Data curation; Writing—review and editing.

**Elena Giacomazzi:** Methodology; Roles/Writing—original draft; Writing—review and editing.

## Data Availability

Data presented in this study are not publicly available at this time but may be obtained from the corresponding author, Venkata Manem, on reasonable request.

## Acknowledgments

Manem holds a salary support award from the Fonds de Recherche du Québec—Santé (FRQS: Quebec Foundation for Health Research) and was supported by the FRQS

Foundation grant from the Quebec Heart & Lung Institute Research Center. Joubert is supported by a salary support award from the Fond de Recherche du Québec—Santé and is a member of the Quebec Respiratory Health Research Network. Routy is supported by a salary support award from the Fond de Recherche du Québec—Santé. The authors thank the Imagia-Canexia-Health company for providing the annotated images, which was done through their Evidens platform. We thank the Quebec Heart & Lung Institute Research Center biobank for providing the specimens.

## References

1. Krawczyk PA, Zhou Q, Dziadziuszko R, Leighl N. *Issues and challenges in NSCLC immunotherapy*. Lausanne, Switzerland: Frontiers Media SA; 2021.
2. Cui P, Li R, Huang Z, et al. Comparative effectiveness of pembrolizumab vs. nivolumab in patients with recurrent or advanced NSCLC. *Sci Rep*. 2020;10:13160.
3. Chang KC, Shao SC, Chen HY, Chan YY, Fang YF. Comparative effectiveness and safety of standard-dose and low-dose pembrolizumab in patients with non-small-cell lung cancer: a multi-institutional cohort study in Taiwan. *Cancers*. 2022;14:1157.
4. Grigg C, Rizvi NA, Rizvi NA. PD-L1 biomarker testing for non-small cell lung cancer: truth or fiction? *J Immunother Cancer*. 2016;4:48.
5. Otano I, Ucero AC, Zugazagoitia J, Paz-Ares L. At the crossroads of immunotherapy for oncogene-addicted subsets of NSCLC. *Nat Rev Clin Oncol*. 2023;20:143–159.
6. Torasawa M, Yoshida T, Yagishita S, et al. Nivolumab versus pembrolizumab in previously-treated advanced non-small cell lung cancer patients: a propensity-matched real-world analysis. *Lung Cancer*. 2022;167:49–57.
7. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet*. 2016;387:1540–1550.
8. Colen RR, Rolfo C, Ak M, et al. Radiomics analysis for predicting pembrolizumab response in patients with advanced rare cancers. *J Immunother Cancer*. 2021;9:e001752.
9. Sun R, Limkin EJ, Vakalopoulou M, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol*. 2018;19:1180–1191.
10. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. *Radiol Med*. 2021;126:1296–1311.
11. Lévi-Strauss T, Tortorici B, Lopez O, et al. Radiomics, a promising new discipline: example of hepatocellular carcinoma. *Diagnostics (Basel)*. 2023;13:1303.
12. Trebeschi S, Drago SG, Birkbak NJ, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol*. 2019;30:998–1004.
13. Brunese L, Mercaldo F, Reginelli A, Santone A. Formal methods for prostate cancer Gleason score and treatment prediction using radiomic biomarkers. *Magn Reson Imaging*. 2020;66:165–175.
14. Whybra P. Standardisation and optimisation of radiomic techniques for the identification of robust imaging biomarkers in oncology. Cardiff University. 2021. https://orca.cardiff.ac.uk/id/eprint/144441/2/Thesis_pw%20%281%29.pdf. Accessed November 29, 2023.
15. Ligero M, Garcia-Ruiz A, Viaplana C, et al. A CT-based radiomics signature is associated with response to immune checkpoint inhibitors in advanced solid tumors. *Radiology*. 2021;299:109–119.
16. Zhao X, Zhao Y, Zhang J, Zhang Z, Liu L, Zhao X. Predicting PD-L1 expression status in patients with non-small cell lung cancer using [$^{18}$F]FDG PET/CT radiomics. *EJNMMI Res*. 2023;13:4.
17. Liu C, Gong J, Yu H, Liu Q, Wang S, Wang J. A CT-based radiomics approach to predict nivolumab response in advanced non-small-cell lung cancer. *Front Oncol*. 2021;11:544339.
18. Zerunian M, Caruso D, Zucchelli A, et al. CT based radiomic approach on first line pembrolizumab in lung cancer. *Sci Rep*. 2021;11:6633.
19. Franzese C, Lillo S, Cozzi L, et al. Predictive value of clinical and radiomic features for radiation therapy response in patients with lymph node-positive head and neck cancer. *Head Neck*. 2023;45:1184–1193.
20. Han Y, Liu D, Li L. PD-1/PD-L1 pathway: current researches in cancer. *Am J Cancer Res*. 2020;10:727–742.
21. Dang TO, Ogunniyi A, Barbee MS, Drilon A. Pembrolizumab for the treatment of PD-L1 positive advanced or metastatic non-small cell lung cancer. *Expert Rev Anticancer Ther*. 2016;16:13–20.
22. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77:e104–e107.
23. Garreta R, Moncecchi G. *Learning Scikit-Learn: Machine Learning in Python*. Birmingham, United Kingdom: Packt Publishing Ltd; 2013.
24. Kononenko I. Estimating attributes: analysis and extensions of RELIEF. In: Bergadano F, Raedt L, eds. *Machine Learning: ECML-94*. Berlin, Germany: Springer; 1994:171–182.
25. Urbanowicz RJ, Olson RS, Schmitt P, Meeker M, Moore JH. Benchmarking relief-based feature selection methods for bioinformatics data mining. *J Biomed Inform*. 2018;85:168–188.
26. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief-based feature selection: introduction and review. *J Biomed Inform*. 2018;85:189–203.
27. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD'16. Association for Computing Machinery; 2016:785–794.
28. González-Recio O, Jiménez-Montero JA, Alenda R. The gradient boosting algorithm and random boosting for genome-assisted evaluation in large data sets. *J Dairy Sci*. 2013;96:614–624.

29. Mayr A, Binder H, Gefeller O, Schmid M. The evolution of boosting algorithms. *Methods Inf Med*. 2014;53:419–427.
30. Mao B, Zhang L, Ning P, et al. Preoperative prediction for pathological grade of hepatocellular carcinoma via machine learning-based radiomics. *Eur Radiol*. 2020;30:6924–6932.
31. Li J, Shi Z, Liu F, et al. XGBoost classifier based on computed tomography radiomics for prediction of tumor-infiltrating CD8+ T-cells in patients with pancreatic ductal adenocarcinoma. *Front Oncol*. 2021;11:671333.
32. Wang X, You X, Zhang L, et al. A radiomics model combined with XGBoost may improve the accuracy of distinguishing between mediastinal cysts and tumors: a multicenter validation analysis. *Ann Transl Med*. 2021;9:1737.
33. Wen Q, Yang Z, Dai H, Feng A, Li Q. Radiomics study for predicting the expression of PD-L1 and tumor mutation burden in non-small cell lung cancer based on CT images and clinicopathological features. *Front Oncol*. 2021;11:620246.
34. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
35. Yolchuyeva S, Giacomazzi E, Tonneau M, et al. A radiomics-clinical model predicts overall survival of non-small cell lung cancer patients treated with immunotherapy: a multicenter study. *Cancers*. 2023;15:3829.
36. Yolchuyeva S, Giacomazzi E, Tonneau M, et al. Radiomics approaches to predict PD-L1 and PFS in advanced non-small cell lung patients treated with immunotherapy: a multi-institutional study. *Sci Rep*. 2023;13:11065.