European Journal of Cancer 183 (2023) 142–151

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Computed tomography-based radiomics for the differential diagnosis of pneumonitis in stage IV non-small cell lung cancer patients treated with immune checkpoint inhibitors

Fariba Tohidinezhad [a], Dennis Bontempi [a,b,c], Zhen Zhang [a], Anne-Marie Dingemans [d], Joachim Aerts [e], Gerben Bootsma [f], Johan Vansteenkiste [g], Sayed Hashemi [h], Egbert Smit [i], Hester Gietema [c], Hugo JWL. Aerts [b,c,j], Andre Dekker [a], Lizza E.L. Hendriks [d], Alberto Traverso [a], Dirk De Ruysscher [a,*]

[a] *Department of Radiation Oncology (Maastro Clinic), School for Oncology and Reproduction (GROW), Maastricht University Medical Center, Maastricht, the Netherlands*
[b] *Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, Boston, MA, USA*
[c] *Department of Radiology and Nuclear Medicine, Maastricht University Medical Center, Maastricht, the Netherlands*
[d] *Department of Pulmonary Diseases, School for Oncology and Reproduction (GROW), Maastricht University Medical Center, Maastricht, the Netherlands*
[e] *Department of Pulmonary Medicine, School of Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands*
[f] *Department of Pulmonary Diseases, Zuyderland Hospital, Heerlen, the Netherlands*
[g] *Department of Respiratory Oncology, University Hospital KU Leuven, Leuven, Belgium*
[h] *Department of Pulmonary Medicine, Amsterdam UMC, VU University Medical Center, Amsterdam, the Netherlands*
[i] *Netherlands Cancer Institute, Amsterdam, the Netherlands*
[j] *Departments of Radiation Oncology and Radiology, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA*

Received 15 December 2022; received in revised form 29 January 2023; accepted 29 January 2023
Available online 9 February 2023

KEYWORDS
Immunotherapy;
Pneumonitis;
Image analysis;

Abstract ***Introduction:*** Immunotherapy-induced pneumonitis (IIP) is a serious side-effect which requires accurate diagnosis and management with high-dose corticosteroids. The differential diagnosis between IIP and other types of pneumonitis (OTP) remains challenging due to similar radiological patterns. This study was aimed to develop a prediction model to

* *Corresponding author*: Department of Radiation Oncology (Maastro Clinic), School for Oncology and Reproduction (GROW), Maastricht University Medical Center, Dr Tanslaan 12, 6229 ET, Maastricht, the Netherlands.
E-mail address: dirk.deruysscher@maastro.nl (D. De Ruysscher).

https://doi.org/10.1016/j.ejca.2023.01.027

Radiomics;
Prediction model

differentiate IIP from OTP in patients with stage IV non-small cell lung cancer (NSCLC) who developed pneumonitis during immunotherapy.

Methods: Consecutive patients with metastatic NSCLC treated with immunotherapy in six centres in the Netherlands and Belgium from 2017 to 2020 were reviewed and cause-specific pneumonitis events were identified. Seven regions of interest (segmented lungs and spheroidal/cubical regions surrounding the inflammation) were examined to extract the most predictive radiomic features from the chest computed tomography images obtained at pneumonitis manifestation. Models were internally tested regarding discrimination, calibration and decisional benefit. To evaluate the clinical application of the models, predicted labels were compared with the separate clinical and radiological judgements.

Results: A total of 556 patients were reviewed; 31 patients (5.6%) developed IIP and 41 patients developed OTP (7.4%). The line of immunotherapy was the only predictive factor in the clinical model (2nd versus 1st odds ratio = 0.08, 95% confidence interval:0.01–0.77). The best radiomic model was achieved using a 75-mm spheroidal region of interest which showed an optimism-corrected area under the receiver operating characteristic curve of 0.83 (95% confidence interval:0.77–0.95) with negative and positive predictive values of 80% and 79%, respectively. Good calibration and net benefits were achieved for the radiomic model across the entire range of probabilities. A correct diagnosis was provided by the radiomic model in 10 out of 12 cases with non-conclusive radiological judgements.

Conclusion: Radiomic biomarkers applied to computed tomography imaging may support clinicians making the differential diagnosis of pneumonitis in patients with NSCLC receiving immunotherapy, especially when the radiologic assessment is non-conclusive.

1. Introduction

Immunotherapy has changed the treatment paradigm for patients with non-small cell lung cancer (NSCLC) by evoking the immune system's response against tumours [1]. Immune checkpoint inhibitors (ICIs) are the most used class of immunotherapy agents, which provoke an immune reaction against cancer by blockade of inhibitory receptors, including programmed death-1 (PD-1), its ligand (PD-[L]1), or cytotoxic T-lymphocyte-associated protein-4 [2]. Currently, there are five PD-[L]1 blocking agents (nivolumab, pembrolizumab, durvalumab, atezolizumab and cemiplimab) and one cytotoxic T-lymphocyte-associated protein-4 antibody (ipilimumab), which have been approved to be used as monotherapy or combination therapy with or without chemotherapy as the first-, second-, or consolidation-line of treatment in patients with unresectable NSCLC [3]. However, an artificially reinvigorated immune system may result in autoimmune reactions that can potentially harm healthy tissues, resulting in poor prognosis and decreased quality of life in patients treated with ICIs.

Immunotherapy-induced pneumonitis (IIP) is a potentially fatal side-effect with an incidence rate of 4–10% in patients with NSCLC treated with ICIs [4]. Many cases of IIP are reversible, but an early diagnosis is crucial for immediate administration of corticosteroids and immunotherapy cessation. However, patients who receive ICIs can also develop other types of pneumonitis (OTP), such as bacterial, viral and fungal infections, sarcoid-like pulmonary reactions and radiation-induced pneumonitis (if they were previously treated with radiotherapy [RT]). These patients will require different treatment approaches, usually without any indication for immunotherapy discontinuation or high-dose corticosteroids administration with its associated side-effects. Therefore, overlapping clinical manifestations of IIP and OTP are a diagnostic challenge.

Making a differential diagnosis based on the computed tomography (CT) images is error-prone due to similar radiological patterns [5]. The most commonly reported radiographic findings for IIP are ground glass opacities, consolidations, bronchiectasis, inter-lobular septal thickening and pleural effusions, which can also be seen in patients with OTP. As of today, according to the latest published guidelines, there is no 'gold standard' for the differential diagnosis of pneumonitis in patients receiving ICIs [6,7]. The analysis of the cellular composition and culture of bronchoalveolar lavage fluid is only advised to assess the possibility of infection. However, its large-scale use is hampered because it is burdensome for patients, not always possible due to respiratory insufficiency of the patient, expensive, time-consuming and not well standardised. Therefore, an objective, non-invasive and scalable approach for discriminating the patients with IIP would have a direct and significant impact on clinical practice.

Radiomics is a field of translational research aiming to extract minable quantitative measures from medical images using mathematical algorithms. Typical radiomic features include shape, intensity properties and texture quantifiers (i.e. spatial arrangement between neighbouring pixels or voxels) [8]. Radiomics can be coupled with machine learning because of its capability to handle massive datasets to establish a diagnosis, response assessment and prognostication. Together, they will be more likely to achieve the primary goal of providing practical, individually tailored information via point-of-care decision support systems. However, less focus has been devoted to using radiomics for predicting treatment-related side-effects [9]. Published papers on the risk factors of IIP are limited to case reports and narrative reviews [4,10].

In this study, we developed a prediction model for estimating the differential diagnosis of pneumonitis (IIP versus OTP) and tested it in patients with stage IV NSCLC treated with anti-PD-[L]1 agents in a multicentre clinical cohort. To optimise this goal, we: (1) used seven Regions of Interest (ROIs) in the lungs that were either automatically segmented using deep learning or guided by the physician's indication to identify the most predictive radiomic features; (2) developed clinical, radiomic and combined models using presumptive clinical parameters and the most prognostic radiomic features; (3) performed internal validation with bootstrapping to assess the models' discrimination power, calibration and clinical decisional benefit; (4) compared the predicted labels with separate judgements made by the respiratory oncologists and radiologists.

2. Materials and methods

2.1. Study design and setting

A prospectively collected cohort (NCT03305380) included all the patients with stage IV NSCLC who developed pneumonitis (IIP or OTP) while receiving anti-PD-[L]1 antibodies as the first- or second-line treatment from six centres in the Netherlands and Belgium (Maastricht University Medical Centre, Netherlands Cancer Institute, University Hospital Leuven, Amsterdam University Medical Centres, Erasmus Medical Centre Rotterdam and Zuyderland Hospital Sittard) between November 2017 and October 2020. No explicit exclusion criteria were applied for the initial patient selection. Ethical approval for the study was obtained from the medical ethical review board of Zuyderland Hospital (17-N-87). The requirement for informed consent was waived due to the observational nature of the study and anonymised data collection.

Fig. 1 shows an overview of the study workflow from data collection to deployment of the prediction model. Below is a detailed description of each step.

2.2. Radiomic features

Converted minable format of the CT images (.nrrd) were resampled to the same number of slices and isotropic voxel size, as recommended by the Image Biomarker Standardisation Initiative. The following ROIs were used to extract the radiomic features: (A) the segmented lungs using a previously published deep learning-based automated method (https://github.com/JoHof/lungmask) and revised by a medical physicist, (B) isotropic cubes of N mms, with N = {50, 100, 150}mm and (C) isotropic spheroids of N mms, with N = {25, 50, 75}mm. The cubical and spheroidal ROIs were centred around the seed point of the inflammation, which was determined by a pulmonologist (LH) with more than 10 years of experience, as the suspected region of the highest inflammation. For the marginal inflammations, the cubes/spheroids were cropped to be in the lung masks. For each ROI, a total of 18 first order features were extracted to represent the distribution of pixel intensities. Following texture features were also calculated: n = 22 Gray Level Co-occurrence Matrix (GLCM), n = 16 Gray Level Run Length Matrix (GLRLM), n = 16 Gray Level Size Zone Matrix (GLSZM) and n = 14 Gray Level Dependence Matrix (GLDM) features. In addition, eight wavelet-decomposition images were generated: HHH, HHL, HLH, LHH, LLH, LHL, HLL, LLL, where 'H' means 'high-pass filter' and 'L' means 'low-pass filter'. Z-normalisation was performed on all radiomic features to reduce the variability of data. Radiomic features were extracted using the open-source software PyRadiomics (v.3.0.1).

2.3. Clinical parameters

The following potential clinical predictors associated with immunotherapy exposure and IIP were collected: demography (age, gender and body mass index), clinical history (chronic obstructive pulmonary disease, diabetes mellitus, cardiovascular diseases, smoking status and performance status), tumour pathology (squamous or non-squamous), treatment history (RT and chemotherapy) and immunotherapy regimen (type of anti-PD-[L]1 medication, line of treatment, dose and schedule).

2.4. Outcome assessment

CT scans were routinely obtained at baseline, 6 weeks, 12 weeks and afterwards every 12 weeks to measure the treatment response. In case of suspicious pneumonitis (IIP/OTP), additional CT scans were taken. If a pneumonitis was seen on CT, further blood tests, physical examinations, sputum cultures or bronchoalveolar lavage were also performed according to the physician's discretion. The initial judgement of the respiratory oncologist (IIP or OTP based on first presentation) and

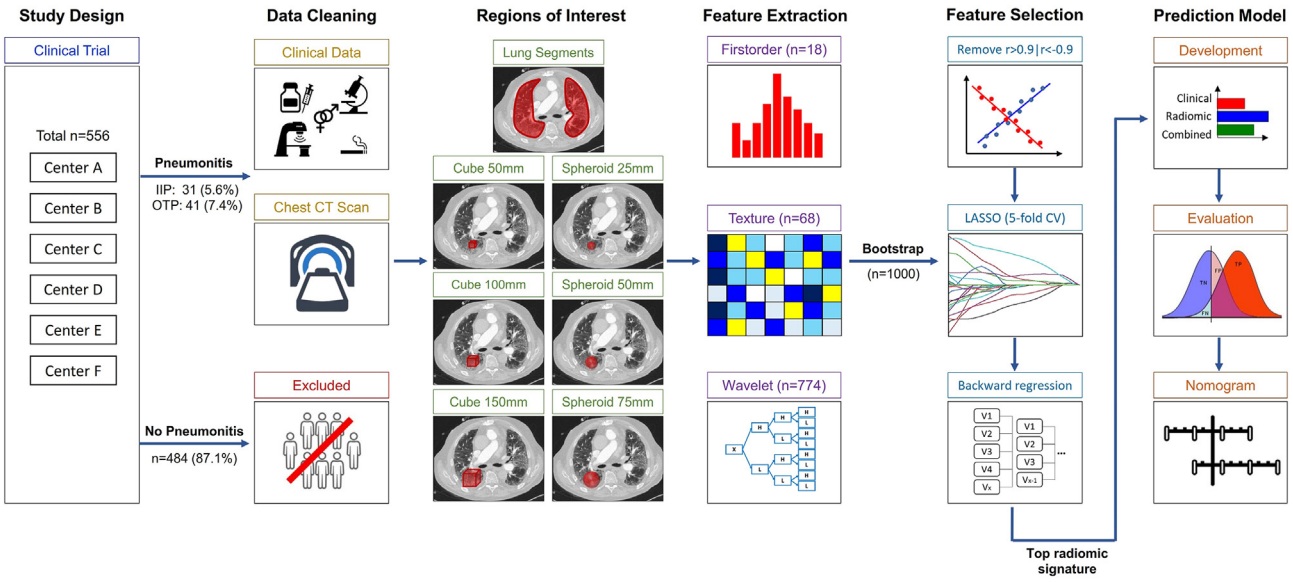

Fig. 1. Overview of the study workflow to develop and evaluate prediction models for immunotherapy-induced pneumonitis versus other types of pneumonitis (IIP versus OTP) in patients with stage IV non-small cell lung cancer treated with immune checkpoint inhibitors. IIP, immunotherapy-induced pneumonitis; OTP, other types of pneumonitis.

the assessment of the radiologist based on the CT images were also recorded to be compared with the prediction models. The final diagnosis ('gold standard') was determined by the treating respiratory oncologist based on the course of disease and outcome of the administered treatments.

2.5. Model development

The radiomic feature selection process was adapted from previous studies published by our research group [11,12]. In summary, the following four steps were followed to select the most predictive radiomic features. First, 1000 bootstrap samples with replacement were drawn from the original cohort. In each bootstrap sample, the pairwise mean absolute correlations were calculated to minimise the number of highly correlated features ($r > 0.9$ or $r < -0.9$) in an unsupervised manner. Second, the least absolute shrinkage and selection operator (LASSO) embedded with logistic regression using 5-fold cross-validation was applied on the 1000 samples to sort the features according to how frequently they were retained by the LASSO. Next, the top six features were arbitrarily selected with respect to the decrease in frequency of the selected features. It should be noted that among the same features with different wavelet decompositions, the one with the highest frequency was selected. Moreover, the selected top radiomic features were compared for different number of bootstrap samples. The six selected features were used to perform stepwise backward logistic regression on the same 1000 bootstrap samples. Finally, the top most frequent signature (more than one radiomic feature) was arbitrarily selected to build the final model. The original cohort was used to fit the coefficients of the final model.

The clinical model was trained using the regular stepwise backward logistic regression on the 15 candidate predictors. Significant predictors in the clinical and radiomic models with the highest effect estimates based on odds ratio (OR) were selected to build the combined model. To avoid overfitting, maximum number of three predictors were considered to be included in a single model (event per variable>10) [13].

2.6. Model evaluation

Discrimination power of the prediction models was estimated using the area under the receiver operating characteristic curve (AUC). The sensitivity, specificity, positive predictive value, negative predictive value and accuracy were also calculated based on the threshold determined by the Youden index method. Calibration, which gauges the agreement between the predicted probabilities and actual outcomes, was evaluated using the graphical assessment of the calibration power (scatter plot where $x = y$ line denotes perfect calibration) [14]. Internal validation with 1000 bootstrap samples was performed to estimate the statistical optimism of the AUCs and calibration slopes using the method recommended in the Transport Reporting of a Multivariable Prediction Model for Individual Prognosis of Diagnosis guidelines [15,16]. The estimated optimisms were subtracted from the original AUCs and calibrations to obtain the optimism-corrected performance measures.

The decision curve analysis was used to visualise the decisional benefit of models considering 'Treat All' and 'Treat None' as the benchmarking strategies. This method calculates the net benefit as a function of relative harms related to the false predictions across a range of threshold probabilities. The model with the highest net benefit at a particular threshold is optimal, regardless of the size of the difference [17].

To enhance the use of the prediction model in future diagnostic research, a nomogram was developed to predict the probability of having IIP based on significant clinical and radiomic features. All analyses were performed in R v.4.2.2 (R Foundation for Statistical Computing).

3. Results

3.1. Clinical characteristics

Seventy-two out of 556 patients, with a mean age of 62.4 ± 9.9 years (range: 25–81) and 60% (n = 43) male gender developed pneumonitis during immunotherapy treatment. A total of 31 (5.6%) patients with IIP and 41 patients with OTP (7.4%) were identified. The major pathological types in this group were adenocarcinoma (n = 48, 67%) and squamous cell carcinoma (n = 20, 28%). The most frequent comorbidities were cardiovascular diseases (n = 23, 32%), chronic obstructive pulmonary disease (n = 18, 25%) and diabetes mellitus (n = 7, 10%). More than half of the patients were ex-smokers (n = 52, 72%) and most patients (n = 64, 89%) had a good performance status of 0 or 1 at the time of immunotherapy initiation. Nivolumab (n = 59, 82%) and pembrolizumab (n = 10, 14%) were the most frequently used anti-PD-[L]1 agents. Patients received the immunotherapy regimen on a 2-week (n = 60, 83%) or a 3-week (n = 12, 17%) basis. Summary statistics per diagnosis (IIP and OTP) are shown in Table 1.

Table 1
Descriptive statistics of the demographic and treatment variables.

Characteristic	Total (n = 72)[a]	IIP (n = 31)[a]	OTP (n = 41)[a]	P-value[b]
Age				0.092
≤60	24 (33%)	7 (23%)	17 (41%)	
>60	48 (67%)	24 (77%)	24 (59%)	
Male gender	43 (60%)	18 (58%)	25 (61%)	0.8
WHO performance status				0.6
0	28 (39%)	11 (35%)	17 (41%)	
1	36 (50%)	15 (48%)	21 (51%)	
2	8 (11%)	5 (16%)	3 (7.3%)	
Body mass index (kg/m^2)				0.2
<25	45 (62%)	17 (55%)	28 (68%)	
≥25	27 (38%)	14 (45%)	13 (32%)	
Tumour pathology				0.7
Squamous	20 (28%)	8 (26%)	12 (29%)	
Non-squamous	52 (72%)	23 (74%)	29 (71%)	
Smoking				>0.9
Current	12 (17%)	5 (16%)	7 (17%)	
Former	52 (72%)	23 (74%)	29 (71%)	
Never	8 (11%)	3 (9.7%)	5 (12%)	
COPD	18 (25%)	9 (29%)	9 (22%)	0.5
Diabetes mellitus	7 (9.7%)	2 (6.5%)	5 (12%)	0.7
Cardiovascular diseases	23 (32%)	10 (32%)	13 (32%)	>0.9
Radiotherapy				0.11
None	43 (60%)	20 (65%)	23 (56%)	
Low-dose (<60 Gy)	9 (12%)	6 (19%)	3 (7.3%)	
High-dose (≥60 Gy)	20 (28%)	5 (16%)	15 (37%)	
Chemotherapy	22 (31%)	7 (23%)	15 (37%)	0.2
Line of immunotherapy				0.038
First-line	7 (9.7%)	6 (19%)	1 (2.4%)	
Second-line	65 (90%)	25 (81%)	40 (98%)	
Anti-PD-[L]1 medication				0.2
Nivolumab	59 (82%)	23 (74%)	36 (88%)	
Pembrolizumab	10 (14%)	7 (23%)	3 (7.3%)	
Other	3 (4.2%)	1 (3.2%)	2 (4.9%)	
Anti-PD-[L]1 dose (mg)				0.7
<225	35 (49%)	16 (52%)	19 (46%)	
≥225	37 (51%)	15 (48%)	22 (54%)	
Anti-PD-[L]1 schedule				0.2
Once every 2 weeks	60 (83%)	24 (77%)	36 (88%)	
Once every 3 weeks	12 (17%)	7 (23%)	5 (12%)	

[a] n (%), Median (IQR).
[b] Pearson's Chi-squared test, Fisher's exact test or Wilcoxon rank sum test. Abbreviations: COPD, chronic obstructive pulmonary disease; IIP, immunotherapy-induced pneumonitis; OTP, other types of pneumonitis; PD-[L], programmed death (ligand); WHO, world health organization.

3.2. Prediction models

Among the candidate clinical predictors, age (>60 versus ≤ 60 OR = 3.02, 95% confidence interval (CI): 0.96 to 9.5, P = 0.059) and line of immunotherapy (2nd versus 1st OR = 0.08, 95% CI: 0.01 to 0.77, P = 0.028) were found to be the most predictive variables. Using different cubical/spheroidal ROIs as well as the segmented lungs, the best radiomic signature was produced using the 75-mm spheroidal ROI (Supplementary Material S1). Comparing the seven ROIs, the radiomic signature based on the 75-mm spheroidal ROI showed the best calibration which was conformed to the ideal line across the entire range of probabilities. Moreover, despite the 75-mm spheroidal ROI, the radiomic signatures based on the segmented lungs, cubical or small spheroidal ROIs (25-mm and 50-mm) showed multiple lower net benefits than their associated combined models.

Using the 75-mm spheroidal ROI, the frequency of the top features selected by the LASSO as well as the minimum, maximum and mean ORs on 1000 bootstrap samples are shown in Supplementary Material S2-Table S1. As shown in Supplementary Material S2-Figure S2,

Table 2
Clinical, radiomic and combined models for predicting immunotherapy-induced pneumonitis (versus other types of pneumonitis) in patients with stage IV non-small cell lung cancer treated with immune checkpoint inhibitors.

	OR	95% Confidence interval	p-value
Clinical model			
(Intercept)	3.55	0.39 to 32.06	0.258
Age (>60 versus ≤ 60)	3.02	0.96 to 9.50	0.059
Line of immunotherapy (2nd versus 1st)	0.08	0.01 to 0.77	0.028
Radiomic model			
(Intercept)	0.57	0.30 to 1.10	0.096
HLH first order mean	0.46	0.22 to 0.95	0.035
log.sigma.1.0.mm.3D GLSZM zone entropy	5.66	2.10 to 15.22	0.001
HHH GLDM small dependence low grey level emphasis	3.82	1.47 to 9.91	0.006
Combined model			
(Intercept)	1.45	0.14 to 14.52	0.752
Line of immunotherapy (2nd vs 1st)	0.40	0.04 to 4.14	0.442
log.sigma.1.0.mm.3D GLSZM zone entropy	4.47	1.77 to 11.30	0.002
HHH GLDM small dependence low grey level emphasis	2.92	1.23 to 6.96	0.016

Abbreviations GLDM, grey level dependence matrix; GLSZM, grey level size zone matrix; HHH, high-pass filter on three axes; OR, odds ratio.

the top six selected radiomic features did not change for the bootstrap samples of more than or equal to 500, indicating that b = 1000 is a sufficient number of bootstrap samples for feature selection. The final radiomic signature was built using the following features: HLH first order mean (OR = 0.46, 95% CI: 0.22 to 0.95, P = 0.035), log.sigma.1.0.mm.3D glszm zone entropy (OR = 5.66, 95% CI: 2.10 to 15.22, P = 0.001) and HHH gldm small dependence low grey level emphasis (OR = 3.82, 95% CI: 1.47 to 9.91, P = 0.006). Details of the selected radiomic features with CT scans representing the lowest and highest values are shown in Supplementary Material S2-Table S3.

The line of immunotherapy from the clinical model and the two radiomic features with the highest OR in the radiomic model were used to build the combined model: line of immunotherapy (2nd versus 1st OR = 0.40, 95% CI: 0.04 to 4.14, P = 0.442), log.sigma.1.0.mm.3D glszm zone entropy (OR = 4.47, 95% CI: 1.77 to 11.30, P = 0.002) and HHH gldm small dependence low grey level emphasis (OR = 2.92, 95% CI: 1.23 to 6.96, P = 0.016). The details of the clinical, radiomic and combined models are shown in Table 2.

3.3. Model evaluation

As shown in Fig. 2A, the radiomic model with an optimism-corrected AUC of 0.83 (95% CI: 0.77–0.95) outperformed the clinical model (AUC = 0.66, 95% CI: 0.56–0.78). Pairwise comparison of AUCs using bootstrap samples showed a significant difference between the clinical and radiomic models (0.66 versus 0.83, 95% CI: 0.04 to 0.31, P = 0.003). The negative predictive value and positive predictive value of the radiomic model was 80% and 79%, respectively. The calibration power of the three models is shown in Fig. 2B. While the clinical model did not cover the low-risk and high-risk patients, the radiomic model showed good calibration for the entire range of predicted probabilities. In addition, the decision curve analysis (Fig. 2C) revealed the higher net benefit of the radiomic model across the entire range of threshold probabilities in comparison with the clinical and combined models.

3.4. Comparison of the predicted labels with clinical and radiological judgements

While the radiologists judged 35 (49%) and 25 (35%) cases as IIP and OTP, they found 12 (17%) patients as non-conclusive cases, where the definitive diagnosis could not be made based on the CT findings. The respiratory oncologists also considered 34 (47%), 21 (29%) and 17 (24%) cases as IIP, OTP and non-conclusive, respectively. The radiomic model provided correct diagnoses in 10 (83%) patients with non-conclusive radiological judgements. Moreover, the radiomic model correctly classified 14 (82%) patients whom were initially considered as non-conclusive cases by the respiratory oncologists (Fig. 3).

The nomogram of the combined model, presenting the probability estimation for a sample patient is shown in Fig. 4. To enhance the reporting quality of the paper, all data have been reported in line with the Transport Reporting of a Multivariable Prediction Model for Individual Prognosis of Diagnosis statement (Supplementary Material S2-Table S4) [18].

4. Discussion

4.1. Interpretation of the findings

To our knowledge, this study represents the first patient-specific risk prediction algorithm to estimate the probability for IIP diagnosis versus OTP in patients with stage IV NSCLC who have developed pneumonitis while receiving anti-PD-[L]1 agents.

The clinical model revealed a significant predictive value of the line of immunotherapy indicating that the patients who received immunotherapy as the first-line treatment were significantly at higher risk to develop an IIP. Previous studies have confirmed that patients receiving ICI as the first-line treatment are at higher risk of developing immune-related adverse events [19,20]. However, no explicit evidence was found to assess the predictive value of line of immunotherapy for IIP

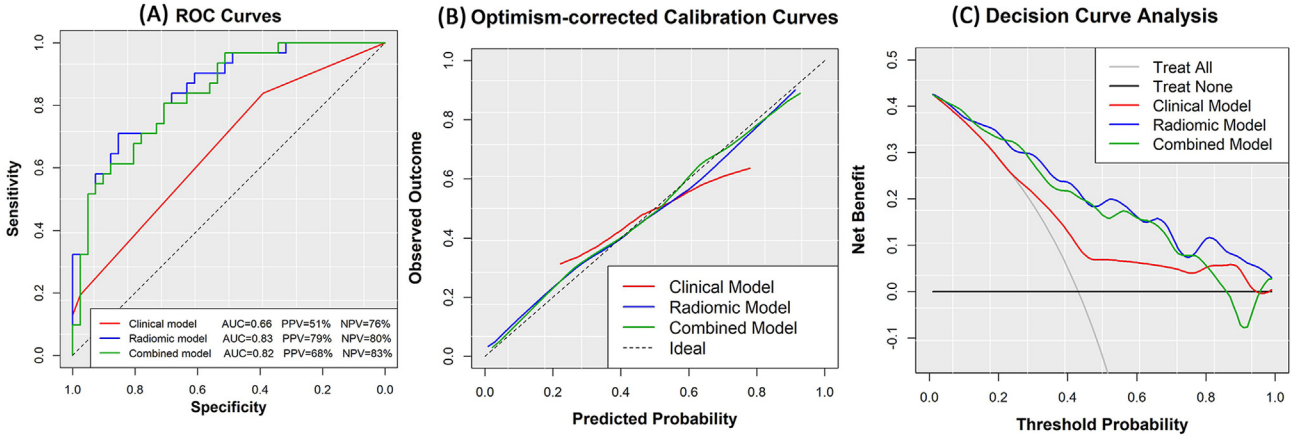

Fig. 2. The receiver operating characteristic (ROC) curves, calibration plots, decision curve analysis of the clinical, radiomic and combined models for predicting immunotherapy-induced pneumonitis (versus other types of pneumonitis) in patients with stage IV non-small cell lung cancer treated with immune checkpoint inhibitors.

against OTP. Moreover, contradictory findings have been reported regarding the predictive value of the age on immune-related pneumonitis [21–24], which may explain its marginal significance value in the current study.

Seven experiments on different ROIs were performed to test whether the ROI definition approach will affect the predictive value of the radiomic features [25]. Although the discrimination and calibration results were comparable for different ROIs, we found that the entire field of lungs may result in negative net benefits for high-risk patients in clinical decision making. In other words, inclusion of non-inflammatory regions can mislead the diagnosis by radiomic features. Among different ROIs, the 75-mm geometric spheroid resulted in the best radiomic signature in terms of calibration and decisional benefit, which implies its favourable predictive value at the patient level. Moreover, since manual ROI delineation is cumbersome and usually prone to inter-reader disagreement, a semi-automated approach (taking the manual-based centre of the inflammatory region and automatically extending it to a specific-sized region) is an optimal solution for practical use.

The texture features that were found to be associated with higher risk of IIP included heterogeneity and irregularity (higher zone entropy and higher small dependence low grey level emphasis), which is concordant with the association between heterogeneous textures and IIP in previous case reports [26]. The predictive value of the radiomic feature on the local intensity variation (first order mean) is in line with the previously published studies showing the predictive value of density features for the discrimination of parenchymal lesions of coronavirus lung infection and radiation pneumonitis [27–29]. Interestingly, we found that GLSZM zone entropy was the most predictive feature in six experimented ROIs, which implies the robustness of its predictive value.

Despite several papers that have found the high advantage of clinical features in the combined models in different clinical settings [30–32], in this study, we found that radiomic features took over the main predictive role, diminishing the small association provided by the only significant clinical factor. This finding may suggest using the qualitative radiologic features (e.g. number of involved lobes, focality of radiological changes, ground-glass opacity, consolidation, etc.) as a complementary set of candidate predictors to improve the clinical judgement.

4.2. Comparison to similar studies

According to a recent review [33], this is the first study evaluating the impact of radiomics for diagnosing IIP on prospectively collected patients in a multicentre clinical series. At the time of writing, two retrospective studies used radiomics to discern the aetiology of pneumonitis.

Chen *et al.* performed a study to test the utility of radiomics in identifying the aetiology of pneumonitis (n = 23 after ICI, n = 29 after RT and n = 30 after ICI + RT). They extracted 93 radiomic features (first order, GLCM, GLDM, GLRLM, GLSZM and NGTDM) from the bilateral whole lungs on diagnostic CTs both before ICI initiation and at the time of pneumonitis diagnosis. Random forest was used to build the prediction model on features selected by the LASSO. An AUC of 0.76 was achieved after testing the model on 30% of the dataset [34].

Like the above study, Cheng *et al.* developed a CT-based radiomic approach to identify the pneumonitis after different treatments (n = 28 after ICI, n = 31 after RT and n = 14 after ICI + RT). However, they used an inflammatory lesion as the ROI, which was manually annotated by experienced radiation oncologists. The intensity, GLCM and Bag of Word radiomic features

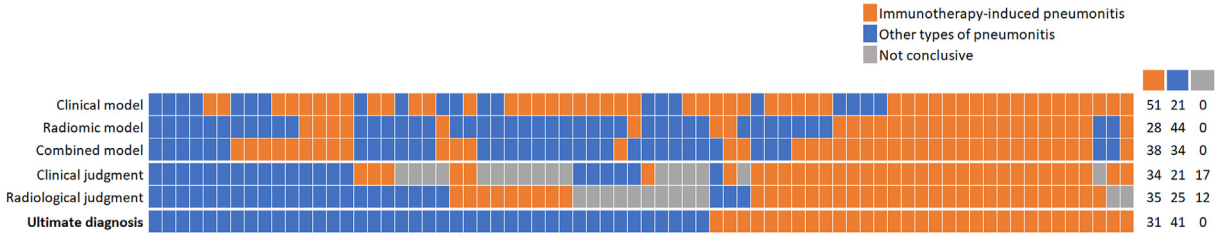

Fig. 3. Comparison of the predictions provided by the clinical, radiomic and combined models with the judgements made by the respiratory oncologists and radiologists. Ultimate diagnoses were determined based on the course of the disease.

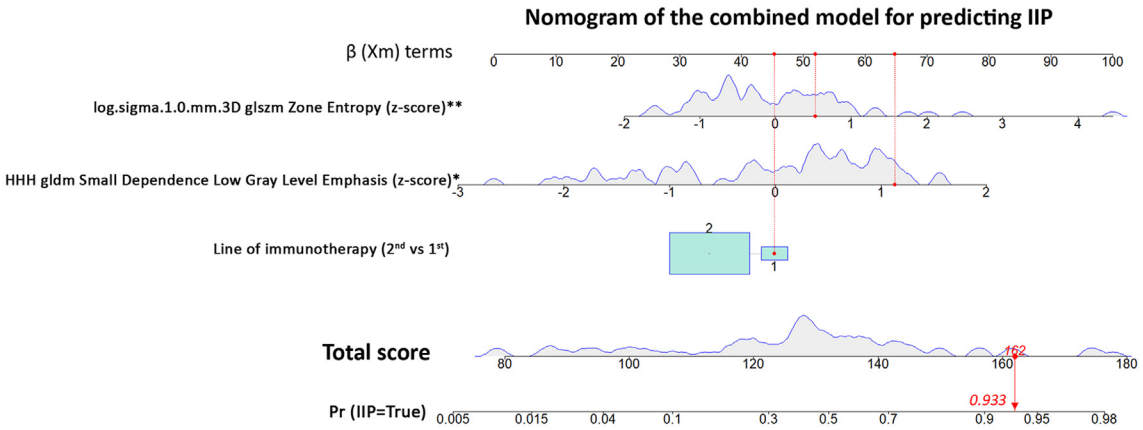

Fig. 4. The radiomic-clinical nomogram incorporated three factors of line of immunotherapy, log.sigma.1.0.mm.3D GLSZM zone entropy and HHH GLDM small dependence low grey level emphasis. GLDM, grey level dependence matrix; GLSZM, grey level size zone matrix.

were fed into logistic regression, random forest and support vector machine as the machine learning algorithms. The best AUC of 0.9 was achieved by applying logistic regression on Bag of Word features using 10-fold cross-validation approach [35].

We, therefore, believe that our study has added value compared with the two previously published ones: first, inclusion of pneumonitis in the ICI-only and RT-only treated patients does not solve this diagnostic challenge [36]. Second, given the event per candidate predictor ratio of less than 10, the previously developed prediction models are at high-risk of being overfitted on the development sample. Third, performing split or cross validation on the data from a single institution usually produces overoptimistic AUC, which might be difficult to achieve on datasets from different centres. Fourth, AUC is essential but not enough to show the predictive value of a prediction model. Previous studies have proved that while AUC hardly changes, calibration drifts easily across different samples [37]. Thus, it is necessary to evaluate the models at patient-level. Fifth, fully manual ROI delineation will impede the reusability of prediction models in clinical practice. Sixth, the predictive value of clinical variables in this study can help future investigators to carefully choose the set of candidate predictors.

4.3. Limitations

The following limitations should be noted when interpreting the results from this report. First, the lack of gold-standard confirmation for IIP and OTP might have hampered the diagnostic accuracy. However, the ultimate diagnosis was made based on the course of disease by the respiratory oncologists. Second, the potential predictive value of other biomarkers (e.g. lab tests) was not captured. However, as also stated in a recent European Society for Medical Oncology Clinical Practice Guideline on immunotherapy toxicity, there are no gold-standard biomarkers [6,7]. Third, due to few number of events per centre, external validation was not possible. However, bootstrapping-adjusted performance measures were used to correct the optimistic results of internal validation. Fourth, the small cohort size remains as an important caveat of this study. However, the sample size is on par with previous papers in this subject [34,35] and still seems to be reasonable given the low incidence rate of IIP and OTP in patients with stage IV NSCLC.

4.4. Implications for practice and research

The major practical contribution of this study is that it presents an empirical semi-automated data-driven

prediction model, which has provided acceptable patient-level predictions for identifying the IIP events. Highlighting the results found by comparing the prediction models with the clinical and radiological judgements, we believe that our model can provide useful predictions when the radiological and clinical judgements are non-conclusive. The current research is timely, given the increasing use of ICIs among patients with (lung) cancer. Moreover, due to the multicentre design of the trial, the results are likely to be pertinent to many similar institutions. However, further prospective clinical trials are needed to investigate the efficacy of the proposed prediction model on discriminative diagnosis of the pneumonitis.

5. Conclusions

The findings of our study suggest that the radiomic signature provided more predictive value than the clinical variables on the differential diagnosis of pneumonitis in patients with NSCLC receiving immunotherapy. Implementation of the proposed model could complement clinicians' intuition and improve the targeting of treatment interventions in an individualised patient-specific fashion.

Funding source

The study was funded by an unrestricted investigator-initiated study grant from Bristol-Myers Squibb (2016-NPS-0226).

Credit statement

Fariba Tohidinezhad: Conceptualisation, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization. **Dennis Bontempi:** Conceptualisation, Methodology, Investigation, Writing - Review & Editing. **Zhen Zhang:** Conceptualisation, Methodology, Writing - Review & Editing. **Anne-Marie Dingemans:** Conceptualisation, Resources, Writing - Review & Editing. **Joachim Aerts:** Conceptualisation, Resources, Writing - Review & Editing. **Gerben Bootsma:** Conceptualisation, Resources, Writing - Review & Editing. **Johan Vansteenkiste:** Conceptualisation, Resources, Writing - Review & Editing. **Sayed Hashemi:** Conceptualisation, Resources, Writing - Review & Editing. **Egbert Smit:** Conceptualisation, Resources, Writing - Review & Editing. **Hester Gietema:** Conceptualisation, Validation, Writing - Review & Editing. **Hugo JWL Aerts:** Conceptualisation, Methodology, Writing - Review & Editing. **Andre Dekker:** Conceptualisation, Methodology, Writing - Review & Editing. **Lizza E.L. Hendriks:** Conceptualisation, Methodology, Investigation, Resources, Writing - Review & Editing, Supervision. **Alberto Traverso:** Conceptualisation, Methodology, Investigation, Writing - Review & Editing, Supervision. **Dirk De Ruysscher:** Conceptualisation, Methodology, Investigation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to express our gratitude to data managers from all the centres for their help in collecting data and answering our queries.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2023.01.027.

References

[1] Forde PM, Kelly RJ, Brahmer JR. New strategies in lung cancer: translating immunotherapy into clinical practice. Clin Cancer Res 2014;20:1067–73. https://doi.org/10.1158/1078-0432.CCR-13-0731.

[2] D'Incecco A, Andreozzi M, Ludovini V, Rossi E, Capodanno A, Landi L, et al. PD-1 and PD-L1 expression in molecularly selected non-small-cell lung cancer patients. Br J Cancer 2015;112:95–102. https://doi.org/10.1038/bjc.2014.555.

[3] Lim SM, Hong MH, Kim HR. Immunotherapy for non-small cell lung cancer: current landscape and future perspectives. Immune Netw 2020;20. https://doi.org/10.4110/in.2020.20.e10.

[4] Zhong L, Altan M, Shannon VR, Sheshadri A. Immune-related adverse events: pneumonitis. Immunotherapy 2020;1244:255–69. https://doi.org/10.1007/978-3-030-41008-7_13.

[5] Rashdan S, Minna JD, Gerber DE. Diagnosis and management of pulmonary toxicity associated with cancer immunotherapy. Lancet Respir Med 2018;6:472–8. https://doi.org/10.1016/S2213-2600(18)30172-3.

[6] Haanen J, Obeid M, Spain L, Carbonnel F, Wang Y, Robert C, et al. Management of toxicities from immunotherapy: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol 2022;33:1217–38. https://doi.org/10.1016/j.annonc.2022.10.001.

[7] Haanen JBaG, Carbonnel F, Robert C, Kerr KM, Peters S, Larkin J, et al. Management of toxicities from immunotherapy: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2017;28:iv119–42. https://doi.org/10.1093/annonc/mdx225.

[8] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278:563–77. https://doi.org/10.1148/radiol.2015151169.

[9] Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. Radiomics and artificial intelligence for biomarker and prediction model development in Oncology. Comput Struct Biotechnol J 2019;17:995–1008. https://doi.org/10.1016/j.csbj.2019.07.001.

[10] Sears CR, Peikert T, Possick JD, Naidoo J, Nishino M, Patel SP, et al. Knowledge gaps and research priorities in immune checkpoint inhibitor-related pneumonitis. An official American thoracic society research statement. Am J Respir Crit Care Med 2019;200:e31–43. https://doi.org/10.1164/rccm.201906-1202ST.

[11] Zhang Z, Wang Z, Yan M, Yu J, Dekker A, Zhao L, et al. Radiomics and dosiomics signature from whole lung predicts radiation pneumonitis: a model development study with prospective external validation and decision-curve analysis. Int J Radiat Oncol Biol Phys 2022. S0360-3016(22)03189-3 https://doi.org/10.1016/j.ijrobp.2022.08.047.

[12] Compter I, Verduin M, Shi Z, Woodruff HC, Smeenk RJ, Rozema T, et al. Deciphering the glioblastoma phenotype by computed tomography radiomics. Radiother Oncol 2021;160:132–9. https://doi.org/10.1016/j.radonc.2021.05.002.

[13] Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 2007;165:710–8. https://doi.org/10.1093/aje/kwk052.

[14] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128–38. https://doi.org/10.1097/EDE.0b013e3181c30fb2.

[15] Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016;69:245–7. https://doi.org/10.1016/j.jclinepi.2015.04.005.

[16] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–73. https://doi.org/10.7326/M14-0698.

[17] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74. https://doi.org/10.1177/0272989X06295361.

[18] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55–63. https://doi.org/10.7326/M14-0697.

[19] Zhao L, Li Y, Jiang N, Song X, Xu J, Zhu X, et al. Association of blood biochemical indexes and antibiotic exposure with severe immune-related adverse events in patients with advanced cancers receiving PD-1 inhibitors. J Immunother 2022;45:210–6. https://doi.org/10.1097/CJI.0000000000000415.

[20] Chen X, Nie J, Dai L, Hu W, Zhang J, Han J, et al. Immune-related adverse events and their association with the effectiveness of PD-1/PD-L1 inhibitors in non-small cell lung cancer: a real-world study from China. Front Oncol 2021;11:607531. https://doi.org/10.3389/fonc.2021.607531.

[21] Yamaguchi T, Shimizu J, Hasegawa T, Horio Y, Inaba Y, Yatabe Y, et al. Pre-existing pulmonary fibrosis is a risk factor for anti-PD-1-related pneumonitis in patients with non-small cell lung cancer: a retrospective analysis. Lung Cancer 2018;125:212–7. https://doi.org/10.1016/j.lungcan.2018.10.001.

[22] Suresh K, Voong KR, Shankar B, Forde PM, Ettinger DS, Marrone KA, et al. Pneumonitis in non-small cell lung cancer patients receiving immune checkpoint immunotherapy: incidence and risk factors. J Thorac Oncol 2018;13:1930–9. https://doi.org/10.1016/j.jtho.2018.08.2035.

[23] Cho JY, Kim J, Lee JS, Kim YJ, Kim SH, Lee YJ, et al. Characteristics, incidence, and risk factors of immune checkpoint inhibitor-related pneumonitis in patients with non-small cell lung cancer. Lung Cancer 2018;125:150–6. https://doi.org/10.1016/j.lungcan.2018.09.015.

[24] Kenmotsu H, Sakai F, Kato T, Kusumoto M, Baba T, Kuwano K, et al. Nivolumab-induced interstitial lung disease (ILD) in Japanese patients with non-small cell lung cancer: a study on risk factors using interim results of post-marketing all-case surveillance. J Clin Orthod 2017;35: 9078–9078. https://doi.org/10.1200/JCO.2017.35.15_suppl.9078.

[25] Jensen LJ, Kim D, Elgeti T, Steffen IG, Hamm B, Nagel SN. Stability of radiomic features across different region of interest sizes-A CT and MR phantom study. Tomography 2021;7:238–52. https://doi.org/10.3390/tomography7020022.

[26] Mojsak D, Dębczyński M, Kuklińska B, Moniuszko-Malinowska A, Mróz RM. The many faces of immune checkpoint inhibitor-associated pneumonitis: 4 case reports. Am J Case Rep 2022;23:e936420–1. https://doi.org/10.12659/AJCR.936420. e936420-7.

[27] Gülbay M, Özbay BO, Mendi BAR, Baştuğ A, Bodur H. A CT radiomics analysis of COVID-19-related ground-glass opacities and consolidation: is it valuable in a differential diagnosis with other atypical pneumonias? PLoS One 2021;16:e0246582. https://doi.org/10.1371/journal.pone.0246582.

[28] Anthony GJ, Cunliffe A, Castillo R, Pham N, Guerrero T, Armato SG, et al. Incorporation of pre-therapy 18F-FDG uptake data with CT texture features into a radiomics model for radiation pneumonitis diagnosis. Med Phys 2017;44:3686–94. https://doi.org/10.1002/mp.12282.

[29] Kawahara D, Imano N, Nishioka R, Ogawa K, Kimura T, Nakashima T, et al. Prediction of radiation pneumonitis after definitive radiotherapy for locally advanced non-small cell lung cancer using multi-region radiomics analysis. Sci Rep 2021;11:16232. https://doi.org/10.1038/s41598-021-95643-x.

[30] Luo L-M, Huang B-T, Chen C-Z, Wang Y, Su C-H, Peng G-B, et al. A combined model to improve the prediction of local control for lung cancer patients undergoing stereotactic body radiotherapy based on radiomic signature plus clinical and dosimetric parameters. Front Oncol 2021;11:819047. https://doi.org/10.3389/fonc.2021.819047.

[31] Chee CG, Yoon MA, Kim KW, Ko Y, Ham SJ, Cho YC, et al. Combined radiomics-clinical model to predict malignancy of vertebral compression fractures on CT. Eur Radiol 2021;31:6825–34. https://doi.org/10.1007/s00330-021-07832-x.

[32] Chang C, Zhou S, Yu H, Zhao W, Ge Y, Duan S, et al. A clinically practical radiomics-clinical combined model based on PET/CT data and nomogram predicts EGFR mutation in lung adenocarcinoma. Eur Radiol 2021;31:6259–68. https://doi.org/10.1007/s00330-020-07676-x.

[33] Castello A, Castellani M, Florimonte L, Urso L, Mansi L, Lopci E. The role of radiomics in the era of immune checkpoint inhibitors: a new protagonist in the jungle of response criteria. J Clin Med 2022;11:1740. https://doi.org/10.3390/jcm11061740.

[34] Chen X, Sheikh K, Nakajima E, Lin CT, Lee J, Hu C, et al. Radiation versus immune checkpoint inhibitor associated pneumonitis: distinct radiologic morphologies. Oncol 2021;26:e1822–32. https://doi.org/10.1002/onco.13900.

[35] Cheng J, Pan Y, Huang W, Huang K, Cui Y, Hong W, et al. Differentiation between immune checkpoint inhibitor-related and radiation pneumonitis in lung cancer by CT radiomics and machine learning. Med Phys 2022;49:1547–58. https://doi.org/10.1002/mp.15451.

[36] Wang H, Guo X, Zhou J, Li Y, Duan L, Si X, et al. Clinical diagnosis and treatment of immune checkpoint inhibitor-associated pneumonitis. Thorac Cancer 2020;11:191–7. https://doi.org/10.1111/1759-7714.13240.

[37] Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. J Am Med Inf Assoc 2017;24:1052–61. https://doi.org/10.1093/jamia/ocx030.