Clinical Radiology 78 (2023) e377–e385

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

# Deep learning for predicting the risk of immune checkpoint inhibitor-related pneumonitis in lung cancer

M. Cheng [a,1], R. Lin [b,1], N. Bai [b,1], Y. Zhang [a], H. Wang [a], M. Guo [a], X. Duan [a], J. Zheng [c,***,2], Z. Qiu [b,**,2], Y. Zhao [a,*,2]

[a] *Department of Internal Medical Oncology, Harbin Medical University Cancer Hospital, Harbin Medical University, Harbin, Heilongjiang Province, China*

[b] *College of Information and Computer Engineering, Northeast Forestry University, Harbin, Heilongjiang Province, China*

[c] *Department of Radiology, Harbin Medical University Cancer Hospital, Harbin Medical University, Harbin, Heilongjiang Province, China*

ARTICLE INFORMATION

*Article history:*
Received 11 July 2022
Received in revised form
14 November 2022
Accepted 20 December 2022

AIM: To develop and validate a nomogram model that combines computed tomography (CT)-based radiological factors extracted from deep-learning and clinical factors for the early predictions of immune checkpoint inhibitor-related pneumonitis (ICI–P).

MATERIALS AND METHODS: Forty ICI–P patients and 101 patients without ICI–P were divided randomly into the training ($n$=113) and test ($n$=28) sets. The convolution neural network (CNN) algorithm was used to extract the CT-based radiological features of predictable ICI–P and calculated the CT score of each patient. A nomogram model to predict the risk of ICI–P was developed by logistic regression.

RESULTS: CT score was calculated from five radiological features extracted by the residual neural network-50-V2 with feature pyramid networks. Four predictors of ICI–P in the nomogram model included a clinical feature (pre-existing lung diseases), two serum markers (absolute lymphocyte count and lactate dehydrogenase), and a CT score. The area under curve of the nomogram model in the training (0.910 versus 0.871 versus 0.778) and test (0.900 versus 0.856 versus 0.869) sets was better than the radiological and clinical models. The nomogram model showed good consistency and better clinical practicability.

CONCLUSION: The nomogram model that combined CT-based radiological factors and clinical factors can be used as a new non-invasive tool for the early prediction of ICI–P in lung cancer patients after immunotherapy with low cost and low manual input.



* Guarantor and correspondent: Y. Zhao, Department of Internal Medical Oncology, Harbin Medical University Cancer Hospital, Harbin Medical University, 150 Ha Ping Road, Harbin, Heilongjiang Province, 150081, China. Tel.: +86 139 0481 1741.

** Guarantor and correspondent: Z. Qiu, College of Information and Computer Engineering, Northeast Forestry University, 26 He Xing Road, Harbin, Heilongjiang Province, 150006, China. Tel: +86 139 0360 1207.

*** Guarantor and correspondent: J. Zheng, Department of Radiology, Harbin Medical University Cancer Hospital, Harbin Medical University, 150 Ha Ping Road, Harbin, Heilongjiang Province, 150081, China. Tel: +86 136 7463 2164.

*E-mail addresses:* zhengjian_12@163.com (J. Zheng), qiuzw@nefu.edu.cn (Z. Qiu), zhaoyanbin1978@sina.com (Y. Zhao).

[1] Co-first authors.

[2] These authors contributed equally to this work.

https://doi.org/10.1016/j.crad.2022.12.013
0009-9260/

## Introduction

Lung cancer has become the leading cause of cancer-related deaths worldwide.[1] Immune checkpoint inhibitors (ICIs) have substantially improved the survival of cancer patients. ICIs are approved as first-line standard treatment in advanced non-small cell lung cancer (NSCLC) without driver mutations and extensive small cell lung cancer.[2] This has contributed to a rise in immune-related adverse events (irAEs), including ICI-related pneumonitis (ICI–P).[3] ICI–P is probably the most severe and life-threatening of all reported irAEs. The incidence of ICI–P in advanced NSCLC was 4.17% in a meta-analysis, which was much less than 12.4% in clinical practice.[4,5] The mortality of ICI–P in NSCLC could reach 28.1% in the clinical practice.[6] ICI–P is further complicated by radiotherapy and systemic diseases that may damage the lung.[7] There is currently no effective tool for the early prediction of ICI–P in ICI-treated patients.

The interdisciplinary research in the medical imaging technology of cancer and computer-aided diagnosis has become one of the hot topics with the arrival of precision medicine and the big data era.[8] As researchers began using computer algorithms to process medical images, deep-learning algorithms made a significant breakthrough in image classification, with the advantage of automatically extracting image features, computational power, extensive datasets, and diagnostic accuracy that surpassed even physicians in certain diseases.[9,10] A series of studies have shown the models that combined radiological features and deep learning applied to predict lung cancer, including the survival, diagnosis, treatment response, etc.[10–12]

The convolution neural network (CNN) is a subset of deep learning, which can effectively perform data dimension reduction for processing and retain the features of images more completely compared with an ordinary neural network.[13] The residual neural network (ResNet) CNN proposed by He *et al.* won the image field of the ILSVRC & COCO 2015 competitions and reduced the image classification recognition error to 3.57%, which was almost equivalent to the accuracy of image recognition by the human eye.[14] After that, three ResNet networks commonly used in computer vision have been proposed including ResNet-50, ResNet-101, and ResNet-50-V2. Moreover, feature pyramid networks (FPN) on the basis of ResNet-50-V2 can be added to address the problem of multi-scale feature extraction.[15] A nomogram can be used as a new non-invasive tool for the prediction of lung cancer with low cost and low manual input. Recent studies have explored the nomogram model to predict lung cancer, such as prognosis and therapeutic response, pulmonary nodules, operative mortality, epidermal growth factor receptor mutation status, etc.[16–20] This study aimed to develop and validate a nomogram model that combined computed tomography (CT)-based radiological features and CNN for the early prediction of ICI–P in lung cancer patients after immunotherapy.

## Materials and methods

### *Patients*

This retrospective study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the ethics committee of the hospital and informed consent was waived. The electronic medical records and CT data of cancer patients from January 2016 to December 2020 were reviewed and 1,157 ICI-treated lung cancer patients were found. The last follow-up time was 5 December 2021. Inclusion criteria included the following: (1) histopathologically proven lung cancer, (2) received at least one kind of ICI, and (3) complete clinical and imaging data were archived before and after immunotherapy. Exclusion criteria included the following: (1) incomplete clinical and imaging data or lost to follow-up; (2) poor image quality; (3) infectious diseases, such as bacteria, viruses, tuberculosis, fungi, and *Pneumocystis carinii*; (4) suffering from autoimmune diseases; (5) previous or coexisting other tumours. Finally, 141 patients were screened for further analysis according to the criteria. All patients were divided randomly into the training ($n$=113) and test ($n$=28) sets in a ratio of 4:1.

### *Clinical features of ICI–P*

The baseline clinical features and serum markers of all patients were collected. The baseline was defined as the recent period before the first immunotherapy. Pre-existing lung diseases included chronic obstructive pulmonary disease (COPD), pulmonary fibrosis, and other interstitial lung diseases (ILDs). The baseline clinical variables include age, sex, smoking, body mass index, Eastern Cooperative Oncology Group performance status (ECOG PS), pre-existing lung diseases, previous radiotherapy, tumour histology, tumour node metastasis stage according to the 8th edition of the American Joint Committee on Cancer, ICI type. The baseline serum markers included absolute lymphocyte count (ALC), neutrophil/lymphocyte ratio (NLR), platelet/lymphocyte ratio (PLR), lymphocyte/monocyte ratio, albumin, and lactate dehydrogenase (LDH).

ICI–P was defined as emerging new pulmonary shadows after immunotherapy excluding other pulmonary infections or tumour progression, etc. ICI–P was determined by the treating physician and the radiologists at the time of diagnosis based on clinical and radiological findings to exclude other possible diagnoses. Differential diagnoses included pulmonary infections (bacteria, viruses, fungi, mycoplasma/chlamydia, etc.), tumour progression and pseudoprogression, ILDs, radiation-induced pulmonary injury, pulmonary embolism, cardiogenic pulmonary oedema, etc.

### *Deep-learning dataset preparation and preprocessing*

CT images were collected before and after immunotherapy of all patients. Digital Imaging and Communications

in Medicine (DICOM) images were derived from the Picture Archiving and Communication Systems and imported into Materialise Mimics Medical software (version 20.0, https://www.materialise.com/). Materialize Mimics Medical software was used to read and give the true label of each CT image. Label 1 or 0 was assigned to each CT image according to the presence or absence of ICI–P lesions before training, respectively. All CT images were converted from DICOM format to Portable Network Graphic format. They were then divided into the ICI–P-positive and ICI–P-negative groups according to their labels. All CT images were qualitative assessments by two radiologists and a clinician, none of whom were blinded to the patients' final diagnosis.

The ICI–P-positive and ICI–P-negative groups included 2,373 and 2,688 CT sections, respectively. The accuracy of 4,049 CT images of 80% of patients in the training set was validated. The 1,012 CT images of 20% patients in the test set were independently tested for the final performance of the model. Data augmentation was designed to handle imbalanced datasets, augment a dataset, avoid over-fit, and improve the robustness of deep learning models in the training set by the means of flip, rotation, scale, crop, translation, etc.[13,21]

### CNN

Experiments were conducted on a server running Linux system (Ubuntu 20.04 LTS, 64-bit), using a Intel Xeon E5-2690 v3 CPU at 2.60 GHz × 48, 128 GB RAM, and the NVIDIA Corporation TU102 (GeForce RTX 2080 Ti Rev.A). The programming language Python (version 3.6.0, https://www.python.org/) and the deep-learning framework were used to develop the ICI–P recognition model.

In this study, the ResNet-50-V2 CNN was improved by introducing FPN to develop the ICI–P recognition model. The number of sections was normalised using $\mu = \frac{512}{n}$ (where μ refers to the normalised value, n represents the number of images on each CT examination) before training. ResNet-50, ResNet-101, and ResNet-50-V2 with FPN models were used for pre-experiments.[14,15] Accuracy was used as the metric to primarily compare the classification ability of the three models with different layers and versions after the pre-experiments. Ten-fold cross-validation during the training process was undertaken to continuously adjust the hyper-parameters of the model and to monitor whether over-fit occurred until the best prediction accuracy model was selected. The equations used to calculate accuracy are shown below.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

where true positive (TP) is the number of ICI–P-positive images that are correctly classified. True negative (TN) is the number of ICI–P-negative images that are correctly classified. False positive (FP) is the number of ICI–P-negative images that are misclassified as ICI-P-positive. False negative (FN) is the number of ICI–P-positive images that are misclassified as ICI-P-negative.

Finally, the predictive performance of the most accurate model was measured by drawing the receiver operating characteristic (ROC) curve and calculating the area under the ROC curve (AUC) on the test set. The patients' CT score was calculated automatically using this model. The calculation formula was as follows:

$$\text{CT score} = \mu^* \sum_{i=1}^{i=n} (i^* p(i)) \quad (2)$$

where μ refers to the normalised value, n represents the number of images on each CT scan, i refers to the predictive value for each section of each patient through the ICI–P recognition model, i value is 0 (ICI-P-negative) or 1 (ICI-P-positive), and p(i) is the predicted probability corresponding to i, $0 = p(i) \leq 1$. The equation limitation is that μ or n is not 0.

### Nomogram model construction, comparison, and evaluation

The nomogram model was developed to realise the visualisation of a logistic regression model. The variables and the weight coefficients of each variable in this model were obtained by logistic regression and the CNN algorithm. The Hosmer–Leme show test was used to evaluate the goodness-of-fit of the nomogram model in the training and test sets. The discrimination of the nomogram, radiological, and clinical models in the training and test sets were compared by the AUC. The calibration curve was drawn by the bootstrap self-sampling method to evaluate the consistency of the nomogram model in the training and test sets. Decision curve analysis was used to compare the clinical practicability of the nomogram, radiological, and clinical models in the training and test sets.

### Statistical analysis

The Mann–Whitney *U*-test and Chi-square ($\chi^2$) or Fisher's exact test were used to analyse the differences between continuous variables and categorical variables in all patients, respectively. The ROC curve of serum markers was drawn to transform it into categorical variables. The Mann–Whitney *U*-test was used to evaluate CT score differences between ICI–P-positive and ICI–P-negative patients in the training and test sets. Logistic regression was used to determine the potential risk factors of ICI–P. Data were analysed using IBM SPSS (version 25) and the rms package in R software (version 3.4.2). The confidence intervals (CIs) of the odds ratios (ORs) were 95%. A two-sided *p*-value $<0.05$ was considered significant.

## Results

### Clinical characteristics of patients

Forty ICI–P patients and 101 patients without ICI–P comprised the final cohort. The training set included 32 ICI–P-positive and 81 IC-P-negative patients, the test set included eight ICI–P-positive and 20 IC-P-negative patients.

**Table 1**
Baseline characteristics in lung cancer patients after immunotherapy.

| Variables | Training set ($n$=113) | Test set ($n$=28) | $p$-Value |
|---|---|---|---|
| Age (years) | | | 0.630 |
| ≥65 | 35 (31) | 10 (35.7) | |
| <65 | 78 (69) | 18 (64.3) | |
| Sex | | | 0.232 |
| Male | 78 (69) | 16 (57.1) | |
| Female | 35 (31) | 12 (42.9) | |
| Smoking | | | 0.834 |
| Current/former | 59 (52.2) | 14 (50) | |
| Never | 54 (47.8) | 14 (50) | |
| BMI (kg/m$^2$) | | | 0.372 |
| Underweight (≤18.5) | 8 (7.1) | 4 (14.3) | |
| Normal weight (18.5–25) | 64 (56.6) | 17 (60.7) | |
| Overweight (25–30) | 38 (33.6) | 6 (21.4) | |
| Obese (≥30) | 3 (2.7) | 1 (3.6) | |
| ECOG PS | | | 0.135 |
| ≥2 | 8 (7.1) | 5 (17.9) | |
| <2 | 105 (92.9) | 23 (82.1) | |
| Pre-existing lung diseases | | | 0.725 |
| Yes | 32 (28.3) | 7 (25) | |
| No | 81 (71.7) | 21 (75) | |
| Previous radiotherapy | | | 0.550 |
| Yes | 39 (34.5) | 8 (28.6) | |
| No | 74 (65.5) | 20 (71.4) | |
| Histology | | | 0.787 |
| Adenocarcinoma | 58 (51.3) | 17 (60.7) | |
| Squamous cell carcinoma | 43 (38.1) | 9 (32.1) | |
| Small cell carcinoma | 11 (9.7) | 2 (7.1) | |
| Adenosquamous carcinoma | 1 (0.9) | 0 (0) | |
| TNM stage | | | 0.688 |
| III–IV | 105 (92.9) | 27 (96.4) | |
| I–II | 8 (7.1) | 1 (3.6) | |
| ICI type | | | 0.571 |
| PD-1 | 100 (88.5) | 27 (96.4) | |
| PD-L1 | 12 (10.6) | 1 (3.6) | |
| PD-1 + CTLA-4 | 1 (0.9) | 0 (0) | |
| ALC ( × 10$^9$/l) | 1.67 (1.2–2.43) | 1.91 (1.45–2.52) | 0.320 |
| NLR | 2.93 (1.81–4.04) | 2.43 (1.81–3.05) | 0.271 |
| PLR | 155.14 (107.17–221.7) | 126.24 (91.65–204.52) | 0.380 |
| LMR | 3.02 (2.21–4.55) | 3.76 (2.54–4.37) | 0.468 |
| Albumin (g/l) | 40.7 (39–43.1) | 43.5 (37.50–44.5) | 0.282 |
| LDH (U/l) | 191 (161–232) | 180 (164–213) | 0.533 |
| CT score | 161.38 (115.93–196.7) | 166.38 (147.24–202.73) | 0.199 |

The data are shown as the number (percentage) or median (interquartile range).
BMI, body mass index; ECOG PS, Eastern Cooperative Oncology Group performance status; TNM, tumour node metastasis; ICI, immune checkpoint inhibitor; PD-1, programmed cell death protein-1; PD-L1, programmed death ligand-1; CTLA-4, cytotoxic T-lymphocyte-associated protein 4; ALC, absolute lymphocyte count; NLR, neutrophil/lymphocyte ratio; PLR, platelet/lymphocyte ratio; LMR, lymphocyte/monocyte ratio; LDH, lactate dehydrogenase; CT, computed tomography.

Table 1 shows that there were no significant differences in all baseline characteristics between the training ($n$=113) and test ($n$=28) sets ($p$>0.05).

### *Deep-learning experiment*

The workflow of ICI–P recognition model development, training, and validation is shown in Fig 1. First, the CT images were labelled and then divided into the ICI–P-positive and ICI–P-negative groups. Data augmentation of each CT section in the training set through eight methods, respectively, including horizontal flip, vertical flip, horizontal and vertical flip, horizontal translation, vertical translation, reduce brightness, increase brightness, and shrink 0.2 × for training (Fig 1a). Then, the FPN was added to ResNet-50-V2 to ensure that the network extracted multi-scale features with limited time and arithmetic cost. ResNet-50-V2 with FPN networks extracted five multi-scale features: Feature 3, Feature 4, Feature 5, Feature 6, and Feature 7. Five features were classified and predicted by the network and each predictive value consisted of two neurons. Finally, 10 neurons were connected by Concatenate and then input to the Softmax classifier to obtain the binary classification results of ICI–P (Fig 1b). After 100 epochs of training for the ResNet-50, ResNet-101, and ResNet-50-V2 with FPN models, respectively, the ResNet-50-V2 with FPN model showed the highest accuracy (97.22%). This model had a favourable classification performance (AUC 0.880; Fig 1c). CT scores of patients who were ICI–P positive and ICI–P negative were significantly different in the training ($p$<0.001) and test ($p$=0.001) sets.

### *Risk factors associated with ICI–P*

The risk factors of ICI–P in the training set ($n$=113) were identified by logistic regression (Table 2). The univariate and multivariate analysis showed that pre-existing lung diseases (OR: 13.932, 95% CI: 2.823–68.751; $p$=0.001), low levels of ALC (OR: 0.139, 95% CI: 0.020–0.945; $p$=0.044), high levels of LDH (OR: 6.277, 95% CI: 1.009–38.425; $p$=0.049), and high levels of CT score (OR: 1.019, 95% CI: 1.007–1.031; $p$=0.002) were independent predictors for the occurrence of ICI–P.

### *Nomogram model construction and validation*

The nomogram model included the above four independent predictors of ICI–P that were identified by logistic regression in the training set ($n$=113; Fig 2). The calculating formula was as follows:

$$\text{Nomoscore} = \text{Pre-existing lung diseases} \times 1.987 - \text{ALC} \times 0.888 + \text{LDH} \times 0.005 + \text{CT score} \times 0.019 - 4.605.$$

This nomogram is similar to a slide rule. The point of each variable is determined by drawing a line to the point axis. The risk axis corresponding to the total points of all variables was the predicted probability for the occurrence of ICI–P.

There were no significant differences in the Hosmer–Lemeshow test on the training ($p$=0.072) and test ($p$=0.332) sets. In the training set, the AUC of the nomogram, radiological, and clinical models were 0.910 (95% CI:

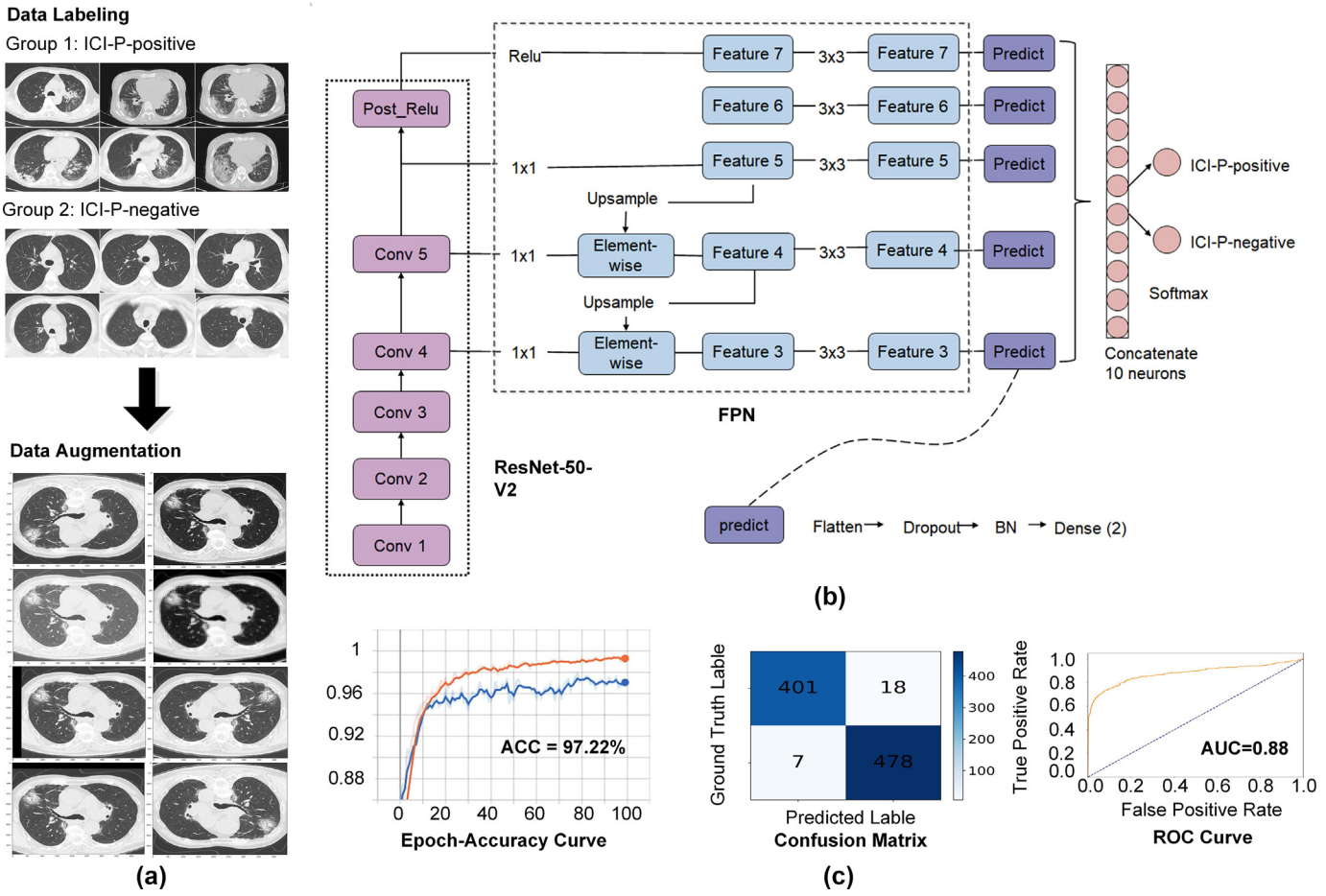


**Figure 1** Workflow of immune checkpoint inhibitor-related pneumonitis recognition model development, training, and validation. (a) CT image labelling, grouping, data augmentation. (b) ResNet-50-V2 with FPN model. (c) The accuracy, confusion matrix, ROC curve of the ResNet-50-V2 with FPN model. CT, computed tomography; ResNet, residual neural network; FPN, feature pyramid networks; ACC, accuracy; ROC, receiver operating characteristic; AUC, area under curve.

0.837–0.983), 0.871 (95% CI: 0.786–0.956), and 0.778 (95% CI: 0.698–0.858; Fig 3a). In the test set, the AUC of the nomogram, radiological, and clinical models were 0.900 (95% CI: 0.715–1.000), 0.856 (95% CI: 0.634–1.000), and 0.869 (95% CI: 0.724–1.000; Fig 3b). In the test set, the accuracy (0.928 versus 0.893 versus 0.821), sensitivity (0.750 versus 0.625 versus 0.500), and specificity (1.000 versus 1.000 versus 0.950) of the nomogram, radiological, and clinical models were analysed. The mean absolute error of the nomogram model on the calibration curve in the training (Fig 4a) and test (Fig 4b) sets were 0.025 and 0.058, separately. The decision curve demonstrated that the nomogram model provided a better net benefit than either the radiological or clinical models alone in the training (Fig 5a) and test (Fig 5b) sets.

## Discussion

The present study analysed the baseline characteristics of ICI–P. Pre-existing lung diseases, low levels of ALC, and high levels of LDH were independent risk factors of ICI–P. Previous studies have shown that age >65 years, lung cancer, PS ≥2, previous COPD or ILD, previous radical radiotherapy, ICI monotherapy, programmed cell death protein-1 with cytotoxic T-lymphocyte-associated protein 4 inhibitors were risk factors for ICI–P.[2,5,22–27] These studies including the present study may help to identify patients with known ICI–P risk factors.

Next, the CNN algorithm was used to dig deeply into the CT-based radiological features to develop a nomogram model for predicting ICI–P in lung cancer. Recent studies have combined radiomics and machine-learning classifiers to predict lung cancer. They used hand-crafted feature-extraction techniques and image segmentation, such as the least absolute shrinkage and selection operator algorithm and regions of interest sketch.[28,29] Common machine-learning classifiers in lung cancer included random forests, naive Bayes, decision trees, support vector machines, k-nearest neighbours, logistic regression, etc.[30–36] Only two recent studies Mallio *et al.* and Cheng *et al.* have used deep learning and machine learning in ICI–P to distinguish ICI–P from radiation pneumonitis.[36,37] To the authors' knowledge, just Colen *et al.* reported the proof of concept for radiomics to predict ICI–P in 2018. Their study involved only two ICI–P patients and did not incorporate clinical factors.[38]

**Table 2**
Univariate and multivariate logistic regression analysis for the risk factors of ICI–P patients in the training set.

| Variables | Univariate analysis | | Multivariate analysis | |
|---|---|---|---|---|
| | OR (95% CI) | *p*-Value | OR (95% CI) | *p*-Value |
| Age (years), (≥65 versus <65) | 0.828 (0.336–2.038) | 0.681 | | |
| Gender (male versus female) | 1.208 (0.491–2.974) | 0.681 | | |
| Smoking (current/former versus never) | 0.884 (0.39–2.004) | 0.767 | | |
| BMI ($kg/m^2$) | | | | |
| Underweight (≤18.5) | 0.341 (0.04–2.889) | 0.324 | | |
| Normal weight (18.5–25) | 1.687 (0.721–3.947) | 0.228 | | |
| Overweight or obese (≥25) | 0.733 (0.307–1.753) | 0.485 | | |
| ECOG PS (≥2 versus <2) | 4.815 (1.078–21.509) | 0.040[a] | 0.668 (0.043–10.271) | 0.772 |
| Pre-existing lung diseases | 15.62 (5.755–42.396) | 0.001[a] | 13.932 (2.823–68.751) | 0.001[a] |
| Previous radiotherapy | 1.2 (0.512–2.813) | 0.675 | | |
| Histology (adenocarcinoma versus others) | 1.106 (0.487–2.51) | 0.810 | | |
| TNM stage (III–IV versus I–II) | 0.364 (0.085–1.553) | 0.172 | | |
| ICI type (PD-1 versus others) | 1.208 (0.305–4.783) | 0.787 | | |
| ALC ( $\times 10^9$/l), (≥1.07 versus <1.07) | 0.209 (0.075–0.587) | 0.003[a] | 0.139 (0.020–0.945) | 0.044[a] |
| NLR (≥2.56 versus <2.56) | 3.395 (1.365–8.446) | 0.009[a] | 1.063 (0.175–6.457) | 0.947 |
| PLR (≥157.71 versus <157.71) | 4.582 (1.873–11.209) | 0.001[a] | 1.828 (0.312–10.701) | 0.503 |
| LMR (≥2.54 versus <2.54) | 0.608 (0.262–1.408) | 0.245 | | |
| Albumin (g/l), (≥37.70 versus <37.70) | 0.471 (0.170–1.310) | 0.149 | | |
| LDH (U/l), (≥274 versus <274) | 3.636 (1.312–10.077) | 0.013[a] | 6.277 (1.009–38.425) | 0.049[a] |
| CT score | 1.027 (1.016–1.038) | <0.001[a] | 1.019 (1.007–1.031) | 0.002[a] |

ICI–P, immune checkpoint inhibitor-related pneumonitis; BMI, body mass index; ECOG PS, Eastern Cooperative Oncology Group performance status; TNM, tumour node metastasis; ICI, immune checkpoint inhibitor; PD-1, programmed cell death protein-1; ALC, absolute lymphocyte count; NLR, neutrophil/lymphocyte ratio; PLR, platelet/lymphocyte ratio; LMR, lymphocyte/monocyte ratio; LDH, lactate dehydrogenase; CT, computed tomography; OR, odds ratio; CI, confidence interval.
[a] $p<0.05$.

They applied manually maximum relevance minimum redundancy feature selection and image segmentation to predict ICI–P.

Since the AlexNet model first applied deep learning to image classification in 2012, CNNs have become a common method to solve computer vision tasks.[39] Many network problems were fixed when updating to the latest ResNet-V2 network through the AlexNet, VGGNet, and ResNet networks.[14,39–41] FPN was introduced into the ResNet-50-V2 CNN network, which fused high-level features with bottom-level information to achieve multi-scale feature extraction and improved the prediction accuracy without reducing the prediction speed.[15] This research was not only without manual image segmentation but also could automatically extract multi-scale features, which saved the time and costs of manual image segmentation, avoided the bias of manual image segmentation and feature selection, and realised image-level ICI–P recognition with more concise and faster steps.

In the present study, the ResNet-50-V2 with FPN model showed higher accuracy (97.22%) than the ResNet-50 and ResNet-101 models. This model also displayed favourable

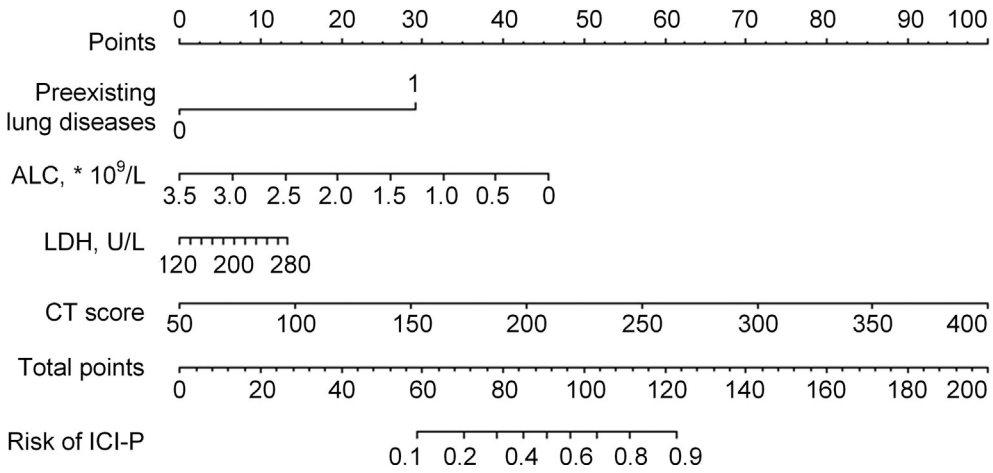


**Figure 2** Nomogram for the prediction of ICI–P in the training set. ALC, absolute lymphocyte count; LDH, lactate dehydrogenase; CT, computed tomography.

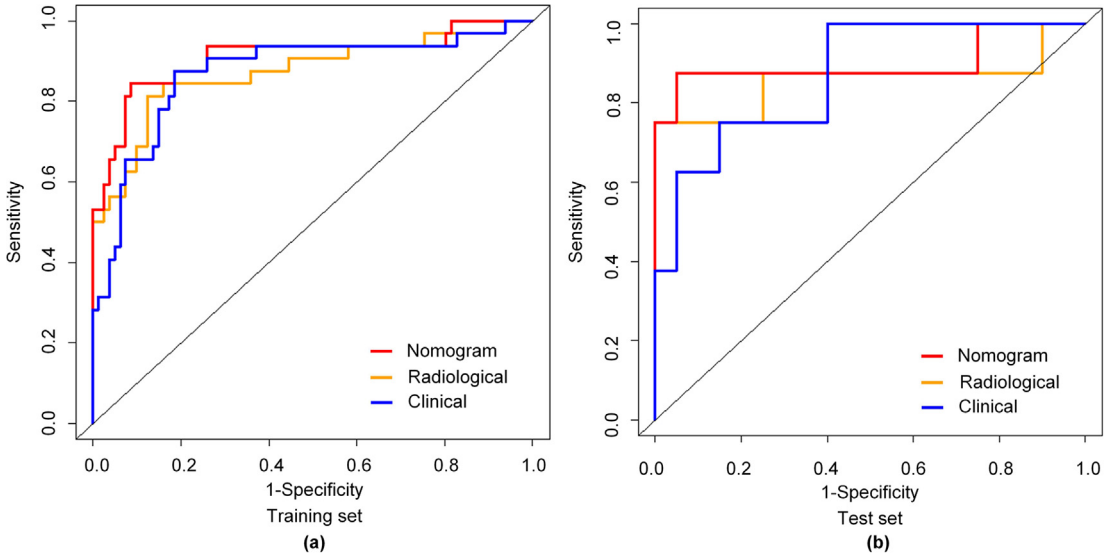


**Figure 3** ROC curves of the nomogram, radiological, and clinical models. (a) The training set. (b) The test set.

discrimination (AUC 0.880). The CT score, which was automatically calculated from this model, was used to quantify the five features extracted by the networks. The CT score showed that the ICI–P recognition model could distinguish radiological features very well ($p<0.05$). Compared with previous studies that required manual feature selection and complex formulas to calculate the radiomic score, the present study used a straightforward algorithm to automatically derive CT scores.[16,28,42]

Using logistic regression, the present study showed that pre-existing lung diseases, ALC, LDH, and CT score are predictors of ICI–P, and a nomogram model was constructed. The reported nomogram models for predicting lung cancer had AUCs of 0.660–0.890 and simply combined two types of variables, such as clinical variables and radiological variables/serum variables.[16–20,43] The present model could reach AUCs of 0.910 and 0.900 in the training and test sets, respectively, exceeding 0.900 for the first time. The nomogram model had better accuracy (0.928), higher sensitivity (0.750) and specificity (1.000) in the test set compared to the radiological and clinical models alone. The nomogram model creatively not only used deep learning to extract radiological features but also a combination of three types of variables: clinical variables, serum variables, and radiological variables. Better predictive performance (AUC 0.910) was achieved with more comprehensive biomarkers. The nomogram model also showed better consistency and clinical practicability on the calibration curve and decision curves.

The present study had some limitations. First, the sample size of the study was limited. Second, this was a single-centre retrospective study. In the future, it should be

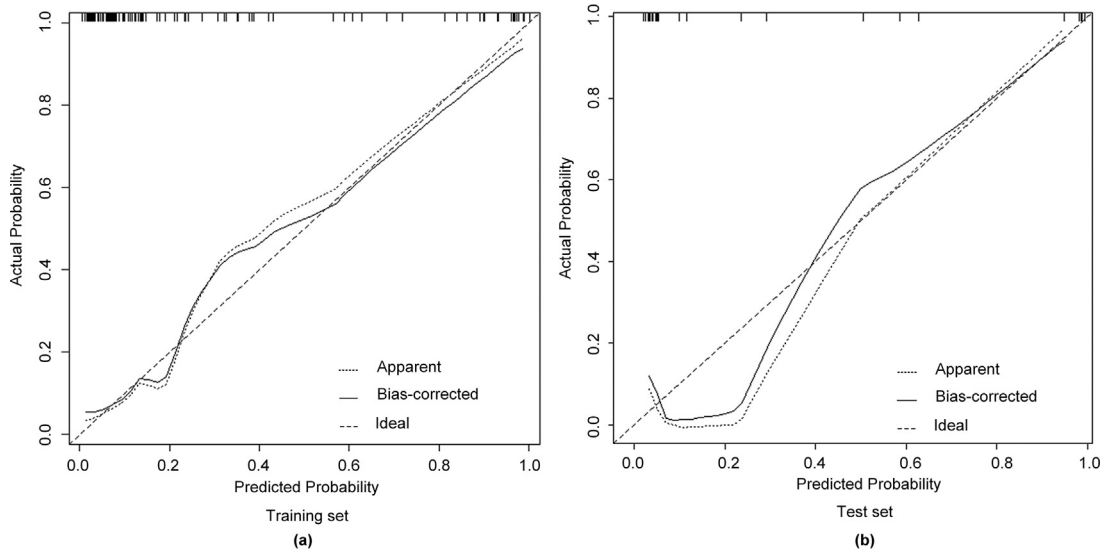


**Figure 4** Calibration curves of the nomogram model. (a) The training set. (b) The test set.

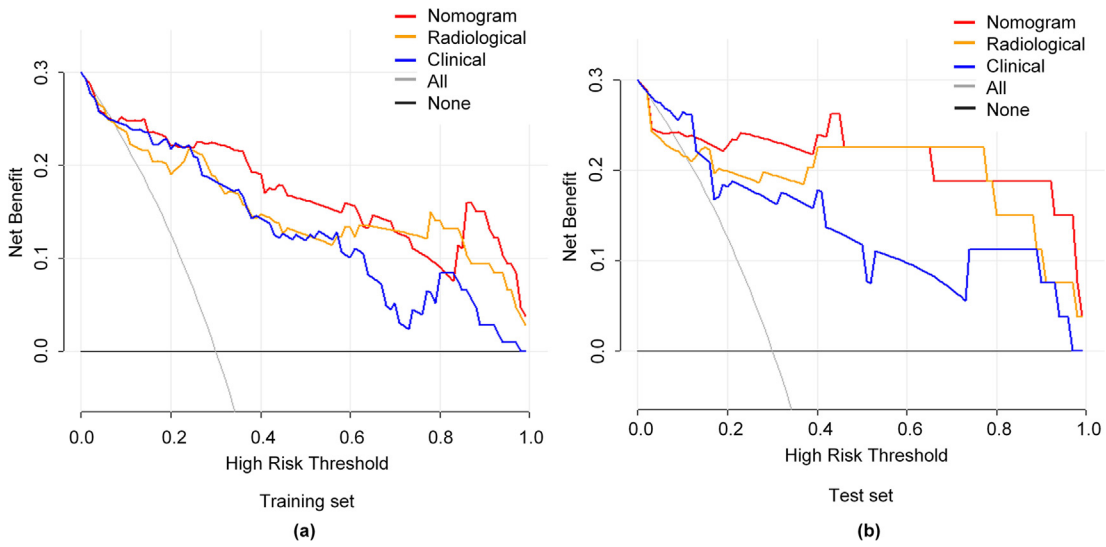


**Figure 5** Decision curve of the nomogram, radiological, and clinical models. (a) The training set. (b) The test set.

extended to multicentre prospective studies to validate the performance of the nomogram model independently. Furthermore, more time-series CT images of patients are required. The combination of deep learning and dynamic changes in the radiological features of patients are required to develop and validate models that can predict the outcome and prognosis of ICI–P. Deep learning can be used to identify human histological and genetic features that can predict the occurrence of ICI–P, which may help elucidate the underlying mechanisms of ICI–P.

In conclusion, the present study used deep learning to predict ICI–P and visualised the ICI–P recognition model using the nomogram for the first time. ResNet-50-V2 with FPN networks were used to develop and validate a nomogram model that combined CT-based radiological factors and clinical factors. The nomogram model had higher discrimination (AUC 0.910), better practicability, and good consistency on both the training and test sets compared to the radiological and clinical models alone. The present study suggests that CT-based radiological factors and clinical factors can be predictors of ICI–P. The nomogram can be used as a new non-invasive tool for the early prediction of ICI–P in lung cancer patients after immunotherapy with low cost and low manual input.

## Declaration of competing interest

The authors declare no conflict of interest.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant no. 81673024) and the Fund of Harbin Medical University Cancer Hospital (grant no. CIP-01). The authors gratefully acknowledge Jun Zhu and Zhao Jin for data collection, and Yi Zhao for reviewing the chest CT images of patients.

## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin* 2020;**70**:7–30. https://doi.org/10.3322/caac.21590.
2. Atchley WT, Alvarez C, Saxena-Beem S, *et al.* Immune Checkpoint inhibitor-related pneumonitis in lung cancer: real-world incidence, risk factors, and management practices across six health care centers in North Carolina. *Chest* 2021;**160**:731–42. https://doi.org/10.1016/j.chest.2021.02.032.
3. Sears CR, Peikert T, Possick JD, *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* 2020;**200**:e31–43. https://doi.org/10.1164/rccm.201906-1202ST.
4. Lin GF, Xu Y, Lin H, *et al.* The association between the incidence risk of pneumonitis and PD-1/PD-L1 inhibitors in advanced NSCLC: a meta-analysis of randomised controlled trials. *Int Immunopharmacol* 2021;**99**:108011, https://doi.org/10.1016/j.intimp.2021.108011.
5. Fujimoto D, Miura S, Yoshimura K, *et al.* Pembrolizumab plus chemotherapy-induced pneumonitis in chemo-naïve patients with non-squamous non-small cell lung cancer: a multicentre, retrospective cohort study. *Eur J Cancer* 2021;**150**:63–72. https://doi.org/10.1016/j.ejca.2021.03.016.
6. Huang A, Xu Y, Zang X, *et al.* Radiographic features and prognosis of early- and late-onset non-small cell lung cancer immune checkpoint inhibitor-related pneumonitis. *BMC Cancer* 2021;**21**:634. https://doi.org/10.1186/s12885-021-08353-y.
7. Naidoo J, Nishino M, Patel SP, *et al.* Immune-related pneumonitis after chemoradiotherapy and subsequent immune checkpoint blockade in unresectable stage III non-small-cell lung cancer. *Clin Lung Cancer* 2020;**21**:e435–44. https://doi.org/10.1016/j.cllc.2020.02.025.
8. Yan Y, Yao XJ, Wang SH, *et al.* A survey of computer-aided tumour diagnosis based on convolutional neural network. *Biology (Basel)* 2021;**10**:1084. https://doi.org/10.3390/biology10111084.
9. Albaradei S, Thafar M, Alsaedi A, *et al.* Machine learning and deep learning methods that use omics data for metastasis prediction. *Comput Struct Biotechnol J* 2021;**19**:5008–18. https://doi.org/10.1016/j.csbj.2021.09.001.
10. Chen CH, Lee YW, Huang YS, *et al.* Computer-aided diagnosis of endobronchial ultrasound images using convolutional neural network. *Comput Methods Programs Biomed* 2019;**177**:175–82. https://doi.org/10.1016/j.cmpb.2019.05.020.
11. Paul R, Hawkins SH, Balagurunathan Y, *et al.* Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. *Tomography* 2016;**2**:388–95. https://doi.org/10.18383/j.tom.2016.00211.

12. Xu Y, Hosny A, Zeleznik R, *et al*. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res* 2019;**25**:3266–75. https://doi.org/10.1158/1078-0432.CCR-18-2495.
13. Yamashita R, Nishio M, Do RKG, *et al*. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018;**9**:611–29. https://doi.org/10.1007/s13244-018-0639-9.
14. He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition. In: *IEEE Conference on computer Vision and pattern recognition (CVPR) 2016*; 2016. p. 770–8. https://doi.org/10.1109/CVPR.2016.90.
15. Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. In: *IEEE Conference on computer Vision and pattern recognition (CVPR) 2017*; 2017. p. 936–44. https://doi.org/10.1109/CVPR.2017.106.
16. Tu W, Sun G, Fan L, *et al*. Radiomics signature: a potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology. *Lung Cancer* 2019;**132**:28–35. https://doi.org/10.1016/j.lungcan.2019.03.025.
17. Wang R, Dai W, Gong J, *et al*. Development of a novel combined nomogram model integrating deep learning-pathomics, radiomics and immunoscore to predict postoperative outcome of colorectal cancer lung metastasis patients. *J Hematol Oncol* 2022;**15**:11. https://doi.org/10.1186/s13045-022-01225-3.
18. Wang S, Yang L, Ci B, *et al*. Development and validation of a nomogram prognostic model for SCLC patients. *J Thorac Oncol* 2018;**13**:1338–48. https://doi.org/10.1016/j.jtho.2018.05.037.
19. Lv X, Wu Z, Cao J, *et al*. A nomogram for predicting the risk of lymph node metastasis in T1-2 non-small-cell lung cancer based on PET/CT and clinical characteristics. *Transl Lung Cancer Res* 2021;**10**:430–8. https://doi.org/10.21037/tlcr-20-1026.
20. Yu X, Gao S, Xue Q, *et al*. Development of a nomogram for predicting the operative mortality of patients who underwent pneumonectomy for lung cancer: a population-based analysis. *Transl Lung Cancer Res* 2021;**10**:381–91. https://doi.org/10.21037/tlcr-20-561.
21. Pérez E, Reyes O, Ventura S. Convolutional neural networks for the automatic diagnosis of melanoma: an extensive experimental study. *Med Image Anal* 2020;**67**:101858, https://doi.org/10.1016/j.media.2020.101858.
22. Li M, Spakowicz D, Zhao S, *et al*. Brief report: inhaled corticosteroid use and the risk of checkpoint inhibitor pneumonitis in patients with advanced cancer. *Cancer Immunol Immunother* 2020;**69**:2403–8. https://doi.org/10.1007/s00262-020-02674-w.
23. Tone M, Izumo T, Awano N, *et al*. High mortality and poor treatment efficacy of immune checkpoint inhibitors in patients with severe grade checkpoint inhibitor pneumonitis in non-small cell lung cancer. *Thorac Cancer* 2019;**10**:2006–12. https://doi.org/10.1111/1759-7714.13187.
24. Jung J, Kim HY, Kim DG, *et al*. Sequential treatment with an immune checkpoint inhibitor followed by a small-molecule targeted agent increases drug-induced pneumonitis. *Cancer Res Treat* 2020;**53**:77–86. https://doi.org/10.4143/crt.2020.543.
25. Fujiwara Y, Horita N, Namkoong H, *et al*. The effect of adding immune checkpoint inhibitors on the risk of pneumonitis for solid tumours: a meta-analysis of phase III randomised controlled trials. *Eur J Cancer* 2021;**150**:168–78. https://doi.org/10.1016/j.ejca.2021.03.012.
26. Su Q, Zhu EC, Wu JB, *et al*. Risk of pneumonitis and pneumonia associated with immune checkpoint inhibitors for solid tumours: a systematic review and meta-analysis. *Front Immunol* 2019;**10**:108. https://doi.org/10.3389/fimmu.2019.00108.
27. Lin X, Deng H, Yang Y, *et al*. Peripheral blood biomarkers for early diagnosis, severity, and prognosis of checkpoint inhibitor-related pneumonitis in patients with lung cancer. *Front Oncol* 2021;**11**:698832, https://doi.org/10.3389/fonc.2021.698832.
28. Hou J, Li H, Zeng B, *et al*. MRI-based radiomics nomogram for predicting temporal lobe injury after radiotherapy in nasopharyngeal carcinoma. *Eur Radiol* 2021;**32**:1106–14. https://doi.org/10.1007/s00330-021-08254-5.
29. Cai J, Zheng J, Shen J, *et al*. A radiomics model for predicting the response to bevacizumab in brain necrosis after radiotherapy. *Clin Cancer Res* 2020;**26**:5438–47. https://doi.org/10.1158/1078-0432.CCR-20-1264.
30. Hawkins S, Wang H, Liu Y, *et al*. Predicting malignant nodules from screening CT scans. *J Thorac Oncol* 2016;**11**:2120–8. https://doi.org/10.1016/j.jtho.2016.07.002.
31. Wu W, Parmar C, Grossmann P, *et al*. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front Oncol* 2016;**6**:71. https://doi.org/10.3389/fonc.2016.00071.
32. Zhang Y, Oikonomou A, Wong A, *et al*. Radiomics-based prognosis analysis for non-small cell lung cancer. *Sci Rep* 2017;**7**:46349, https://doi.org/10.1038/srep46349.
33. Liu Y, Kim J, Balagurunathan Y, *et al*. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clin Lung Cancer* 2016;**17**:441–448.e6. https://doi.org/10.1016/j.cllc.2016.02.001.
34. Liang B, Yan H, Tian Y, *et al*. Dosiomics: extracting 3D spatial features from dose distribution to predict incidence of radiation pneumonitis. *Front Oncol* 2019;**9**:269. https://doi.org/10.3389/fonc.2019.00269.
35. Qiu Q, Xing L, Wang Y, *et al*. Development and validation of a radiomics nomogram using computed tomography for differentiating immune checkpoint inhibitor-related pneumonitis from radiation pneumonitis for patients with non-small cell lung cancer. *Front Immunol* 2022;**13**:870842, https://doi.org/10.3389/fimmu.2022.870842.
36. Cheng J, Pan Y, Huang 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.
37. Mallio CA, Napolitano A, Castiello G, *et al*. Deep learning algorithm trained with COVID-19 pneumonia also identifies immune checkpoint inhibitor therapy-related pneumonitis. *Cancers (Basel)* 2021;**13**:652. https://doi.org/10.3390/cancers13040652.
38. Colen RR, Fujii T, Bilen MA, *et al*. Radiomics to predict immunotherapy-induced pneumonitis: proof of concept. *Invest New Drugs* 2018;**364**:601–7. https://doi.org/10.1007/s10637-017-0524-2.
39. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2012:84–90. https://doi.org/10.1145/3065386.
40. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556. doi: 10.48550/arXiv.1409.1556.
41. He K, Zhang X, Ren S, *et al*. Identity mappings in deep residual networks. In: *Computer vision – ECCV 2016. Lecture notes in computer science*, vol. 9908. Cham: Springer; 2016. p. 630–45. https://doi.org/10.1007/978-3-319-46493-0_38.
42. Weng Q, Zhou L, Wang H, *et al*. A radiomics model for determining the invasiveness of solitary pulmonary nodules that manifest as part-solid nodules. *Clin Radiol* 2019;**74**:933–43. https://doi.org/10.1016/j.crad.2019.07.026.
43. Liang M, Chen M, Singh S, *et al*. Prognostic nomogram for overall survival in small cell lung cancer patients treated with chemotherapy: a SEER-based retrospective cohort study. *Adv Ther* 2021;**39**:346–59. https://doi.org/10.1007/s12325-021-01974-6.