

## RESEARCH ARTICLE

# Differentiation between immune checkpoint inhibitor-related and radiation pneumonitis in lung cancer by CT radiomics and machine learning

Jun Cheng<sup>1,2,3</sup> | Yi Pan<sup>4</sup> | Wei Huang<sup>4</sup> | Kun Huang<sup>5,6</sup> | Yanhai Cui<sup>7</sup> |  
Wenhui Hong<sup>1</sup> | Lingling Wang<sup>1</sup> | Dong Ni<sup>1,2,3</sup> | Peixin Tan<sup>4</sup>

<sup>1</sup> National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China

<sup>2</sup> Medical Ultrasound Image Computing (MUSIC) Laboratory, Shenzhen University, Shenzhen, China

<sup>3</sup> Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen, China

<sup>4</sup> Department of Radiation Oncology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China

<sup>5</sup> Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, Indiana, USA

<sup>6</sup> Regenstrief Institute, Indianapolis, Indiana, USA

<sup>7</sup> Department of Radiology, Guangdong Provincial People's Hospital/Guangdong Academy of Medical Sciences, Guangzhou, China

## Correspondence

Dong Ni, School of Biomedical Engineering, 1066 Xueyuan Avenue, Shenzhen 518055, China.

Email: [nidong@szu.edu.cn](mailto:nidong@szu.edu.cn)

Peixin Tan, Department of Radiation Oncology, 106 Zhongshan 2nd Road, Yuexiu District, Guangzhou 510080, China.

Email: [tpxsaxin@163.com](mailto:tpxsaxin@163.com)

# Jun Cheng and Yi Pan are co-first authors.

## Funding information

National Natural Science Foundation of China, Grant/Award Number: 61901275; Guangzhou Science and Technology Plan Foundation, Grant/Award Number: 2021-02-01-04-1002-0017; Shenzhen University Startup Fund, Grant/Award Number: 2019131; National Key R&D Program of China, Grant/Award Number: 2019YFC0118300; Shenzhen Peacock Plan, Grant/Award Numbers: KQTD2016053112051497, KQJSCX20180328095606003; Medical Scientific Research Foundation of Guangdong Province, China, Grant/Award Numbers: B2018031, B2020024; National Natural Science Foundation of Guangdong Provincial People's Hospital, Grant/Award Number: 8210032051

## Abstract

**Purpose:** Consolidation immunotherapy after completion of chemoradiotherapy has become the standard of care for unresectable locally advanced non-small cell lung cancer and can induce potentially severe and life-threatening adverse events, including both immune checkpoint inhibitor-related pneumonitis (CIP) and radiation pneumonitis (RP), which are very challenging for radiologists to diagnose. Differentiating between CIP and RP has significant implications for clinical management such as the treatments for pneumonitis and the decision to continue or restart immunotherapy. The purpose of this study is to differentiate between CIP and RP by a CT radiomics approach.

**Methods:** We retrospectively collected the CT images and clinical information of patients with pneumonitis who received immune checkpoint inhibitor (ICI) only ( $n = 28$ ), radiotherapy (RT) only ( $n = 31$ ), and ICI+RT ( $n = 14$ ). Three kinds of radiomic features (intensity histogram, gray-level co-occurrence matrix [GLCM] based, and bag-of-words [BoW] features) were extracted from CT images, which characterize tissue texture at different scales. Classification models, including logistic regression, random forest, and linear SVM, were first developed and tested in patients who received ICI or RT only with 10-fold cross-validation and further tested in patients who received ICI+RT using clinicians' diagnosis as a reference.

**Results:** Using 10-fold cross-validation, the classification models built on the intensity histogram features, GLCM-based features, and BoW features achieved an area under curve (AUC) of 0.765, 0.848, and 0.937, respectively. The best model was then applied to the patients receiving combination treatment, achieving an AUC of 0.896.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine

**Conclusions:** This study demonstrates the promising potential of radiomic analysis of CT images for differentiating between CIP and RP in lung cancer, which could be a useful tool to attribute the cause of pneumonitis in patients who receive both ICI and RT.

**KEYWORDS**

CT radiomics, immune checkpoint inhibitor-related pneumonitis, lung cancer, machine learning, radiation pneumonitis

## 1 | INTRODUCTION

Consolidation immunotherapy of immune checkpoint inhibitor (ICI) durvalumab following concurrent chemoradiotherapy (CCRT) is the current standard of care for patients with stage III unresectable non-small cell lung cancer (NSCLC) as the phase 3 PACIFIC trial has shown that administering ICI after CCRT significantly improved progression-free survival and overall survival compared with placebo.<sup>1,2</sup> A number of clinical trials are currently evaluating the role of concurrent/induction ICI with chemoradiotherapy in locally advanced NSCLC.<sup>3,4</sup> While ICI in conjunction with radiotherapy (RT) has shown promising prospects, treatment-related pneumonitis including radiation pneumonitis (RP) and checkpoint inhibitor-related pneumonitis (CIP), one of the most frequent and clinically challenging adverse events in the combination setting, should raise concerns.

CIP is a rare but fatal side effect with incidence ranging from 1% to 6% in any grade and <1% to ~3% in grade 3 or higher, as reported in clinical trials for advanced NSCLC patients treated with PD-1/PD-L1 inhibitors.<sup>5–8</sup> In addition to ICI, radiation also leads to lung damage and induces pneumonitis. The incidence of RP is 14% to 49% in grade 2 or higher<sup>9,10</sup> and 4% to 9% in grade 3 or higher<sup>11</sup> in NSCLC patients after radical RT within 6 months. Recent studies reported that the incidence of CIP may be higher because of the potential synergy with RT and lung injury caused by RT.<sup>12,13</sup> A second analysis of the KEYNOTE-001 study demonstrated that pneumonitis of any grade was 63% in patients who have received prior RT versus 40% in those who did not ( $p = 0.052$ ).<sup>12</sup> In the PACIFIC study, the incidence of pneumonitis of any grade was higher with consolidation ICI (33.9% vs. 24.8%).<sup>1</sup> A recent multicenter retrospective study reported a much higher incidence (81.8%) of any grade pneumonitis in a real-world cohort of patients treated with durvalumab after CCRT.<sup>14</sup>

It is very difficult and challenging for clinicians to differentiate between CIP and RP as the clinical and radiologic features of CIP are very similar to those of RP, with nonproductive cough, unresolved dyspnea, and nonspecific interstitial pneumonia in the periphery or anywhere of the lungs.<sup>15,16</sup> Differentiating between CIP and RP can have significant implications for clinical management such as the treatments

for pneumonitis and the decision to continue or restart immunotherapy.<sup>17,18</sup>

Although a few studies discussed the typical radiologic appearance of CIP and RP,<sup>8,19</sup> these radiologic findings are only suggestive because pneumonitis has a wide spectrum of radiologic appearance. For example, RP is usually, but not always, limited to the radiation field of the lung. Figure 1 shows some typical CT images of CIP and RP and two images of RP resembling CIP. In lung cancer, CT is routinely used for clinical management, including diagnosis, radiation treatment planning, and surveillance of treatment response. CT-based radiomics approaches have been successfully applied to various tasks such as differentiation between benign and malignant lesions<sup>20,21</sup>; prediction of prognosis,<sup>22,23</sup> treatment response,<sup>24–26</sup> and distance metastasis<sup>27,28</sup>; and associations between genotype and imaging phenotype.<sup>29–31</sup> There are very few studies focusing on the differentiation between CIP and RP using radiomic features, as ICI therapy has been used in lung cancer for only a few years and the incidence of CIP is relatively low.

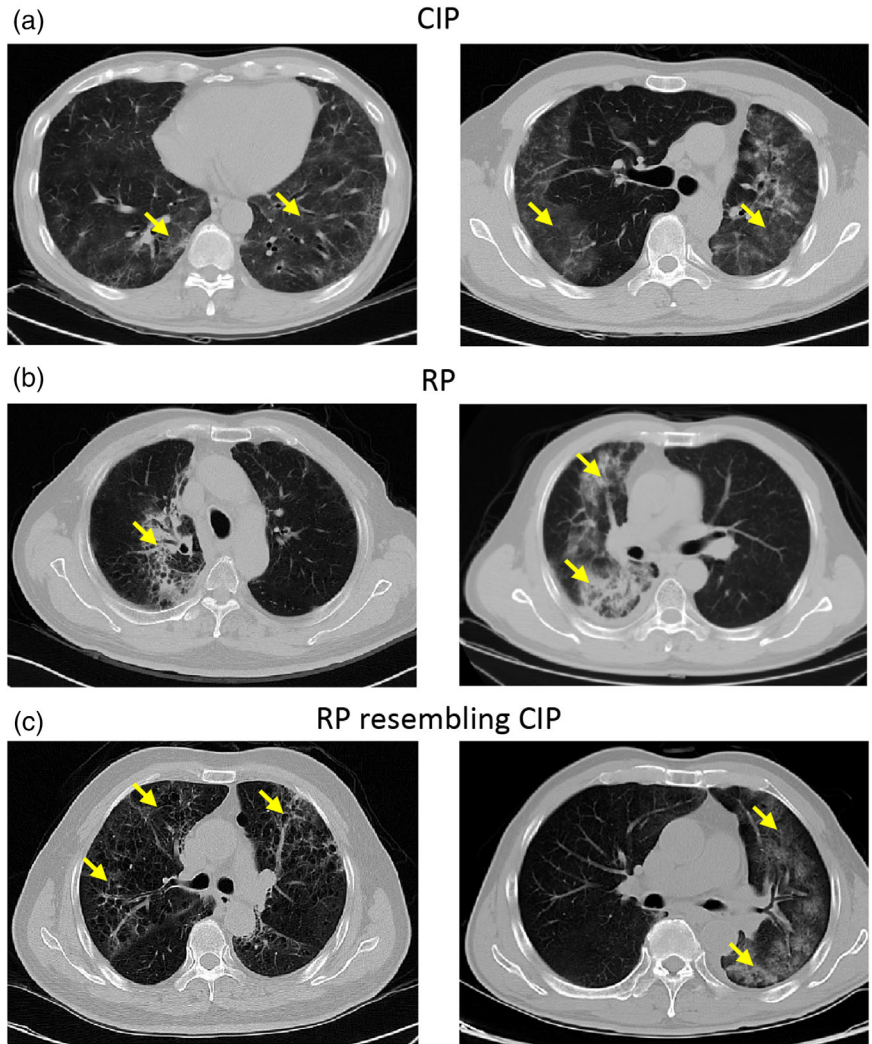
In this study, we present a CT radiomics approach to differentiate between CIP and RP in lung cancer patients. We collected three cohorts of patients with pneumonitis who received ICI only, RT only, and ICI+RT, respectively. Three different kinds of radiomic features were extracted from CT images. The utility of these radiomic features for classifying CIP and RP was first evaluated using the ICI and RT cohorts and further validated using the ICI+RT cohort.

## 2 | MATERIALS AND METHODS

### 2.1 | Patients and CT image acquisition

This retrospective study was approved by the Ethics Committee of Guangdong Provincial People's Hospital. We collected three datasets (ICI, RT, and ICI+RT datasets) which contained the CT images and clinical information of patients who developed pneumonitis. The ICI dataset consisted of 28 lung cancer patients who developed CIP after ICI therapy. Patients were excluded from the analysis if they received thoracic RT before the occurrence of CIP. The RT dataset consisted of 31 patients randomly selected from locally advanced

**FIGURE 1** Examples of CT images in patients with pneumonitis who received ICI only or RT only. (a) CT images in two patients with CIP demonstrated ground-glass and reticular opacities involving both lungs with a diffuse distribution, representing a cryptogenic organizing pneumonia pattern. Left patient also presented a nonspecific interstitial pneumonia pattern. (b) CT images in two patients with RP demonstrated reticular opacities, consolidations, and bronchiectasis. The inflammatory lesions were within radiation field and had clear boundaries. (c) CT images in two patients with RP demonstrated radiologic features resembling those of CIP. CIP, immune checkpoint inhibitor-related pneumonitis; ICI, immune checkpoint inhibitor; RP, radiation pneumonitis; RT, radiotherapy

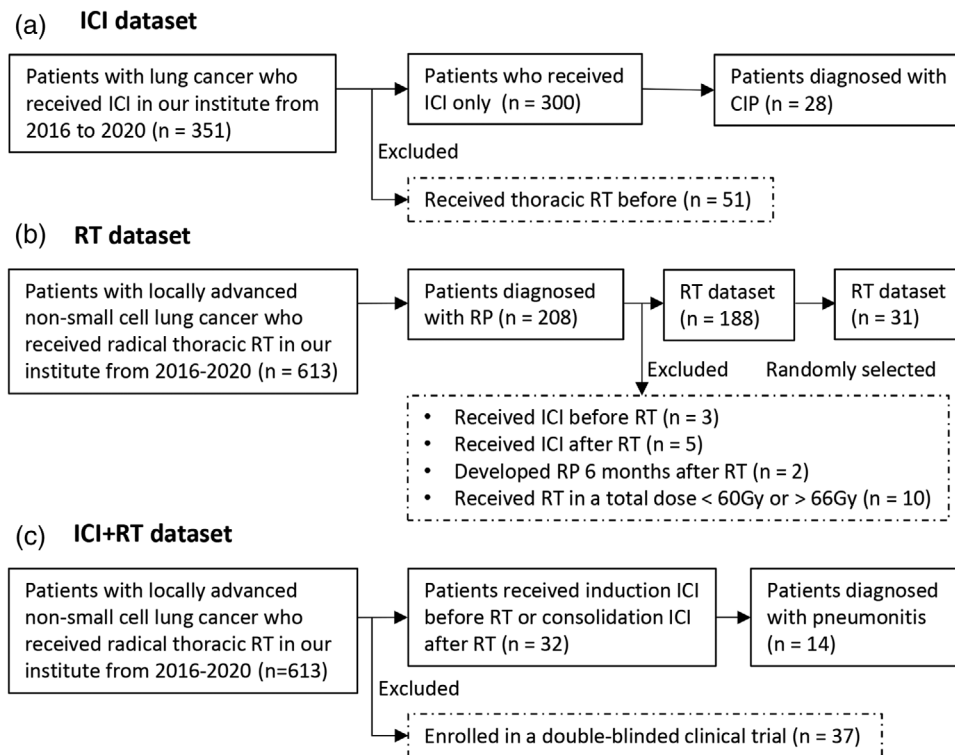


NSCLC patients who were treated with radical thoracic RT in a total dose of 60 to 66 Gy. These patients developed RP within 6 months after RT. Patients were excluded if they received ICI therapy before or after RT. The ICI+RT dataset consisted of 14 patients who developed treatment-related pneumonitis after induction ICI therapy followed by thoracic RT or consolidation ICI therapy following thoracic RT. Note that we excluded patients with clear alternative etiologies, such as proven active pulmonary infection, tuberculosis, pulmonary embolism, or tumor progression. A flowchart for preparing the patient cohorts is shown in Figure 2, and a summary of patient characteristics is provided in Table 1.

We defined CIP by (1) a treatment history of ICI therapy; (2) symptoms of nonproductive cough, unresolving dyspnea, fever, and chest pain; and (3) varied radiographic findings in a chest CT imaging, such as cryptogenic organizing pneumonia, with ground-glass or consolidative opacities in peripheral or peribronchial distribution, or nonspecific interstitial pneumonia, with ground-glass opacities and reticular opacities primar-

**TABLE 1** Clinical characteristics of the patients in the immune checkpoint inhibitor (ICI), radiation therapy (RT), and ICI+RT datasets

Characteristic	ICI	RT	ICI+RT
Patient No.	28	31	14
Sex			
Female	0	5	2
Male	28	26	12
Age (year)			
Median	62	62	62
Range	39–75	44–70	41–78
Smoking (pack-year)			
Median	40	40	26
Range	0–120	0–150	0–60
Pneumonitis grade			
Grade 1	4	19	5
Grade 2	12	3	8
Grade 3	11	9	1
Grade 4	1	0	0



**FIGURE 2** Flowchart for preparing the ICI dataset (a), RT dataset (b), and ICI+RT dataset (c). ICI, immune checkpoint inhibitor; RT, radiotherapy

ily in the peripheral and lower lungs, or pneumonitis presenting as acute interstitial pneumonia and acute respiratory distress syndrome.<sup>18,32</sup> We defined RP by (1) a treatment history of RT; (2) symptoms of shortness of breath, low-grade fever, and nonproductive cough; and (3) radiographic findings in a chest CT imaging with patchy consolidation roughly within the area of the high-dose radiation field and does not conform to normal lobar anatomy.<sup>33</sup> The grade of CIP and RP was scored by treating physicians according to the Common Terminology Criteria for Adverse Events v5.0.

The CT examinations were performed using CT scanners from different manufacturers, including Siemens (Somatom Definition Flash; Erlangen, Germany), General Electric (Lightspeed VCT 99; Waukesha, WI, USA), and Philips (iCT 256 and Ingenuity; Cleveland, Ohio, USA). Thoracic CT scans containing the entire lung were analyzed utilizing a multi-slice helical technique at 120 kVp, mean exposure of 158 mA, mean pixel spacing of 0.78 mm, and slice thickness of 5 mm.

## 2.2 | Analysis workflow

The analysis workflow of our study is shown in Figure 3, which consists of three steps. In the first step, we collected CT images and manually segmented regions of interests (ROIs), that is, inflammatory lesions. Next, three kinds of radiomic features that characterize lung tissue texture at different scales were extracted from the ROIs.

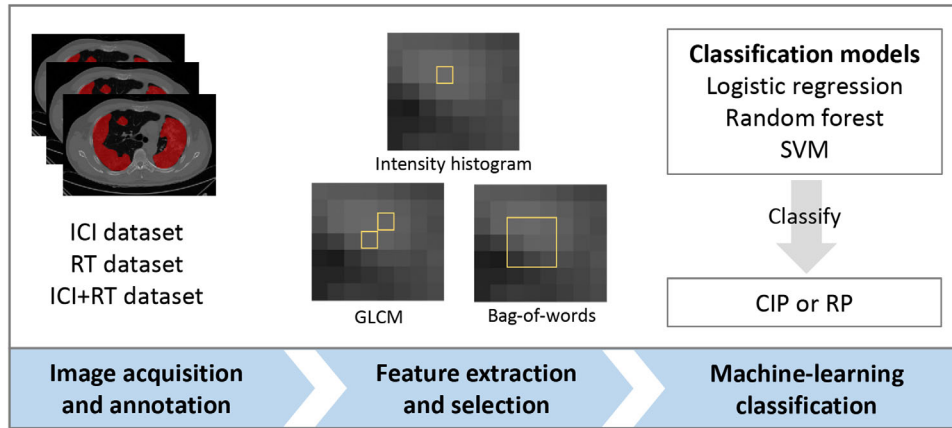
At last, we built classification models on the ICI and RT datasets. The models were first validated on the ICI and RT datasets with 10-fold cross-validation and were then tested on the ICI+RT dataset.

## 2.3 | CT image feature extraction

We extracted radiomic features from the ROIs (inflammatory lesions), which were annotated by an experienced radiation oncologist (PT) and further reviewed by a senior radiation oncologist (YP). Specifically, three feature extraction methods were employed to quantify the texture of ROIs: intensity histogram features, gray-level co-occurrence matrix (GLCM)-based features,<sup>34</sup> and bag-of-words (BoW) features.<sup>35</sup> The three kinds of features describe tissue texture at increasing scales. Intensity histogram is based on individual pixels, GLCM is based on the co-occurrence of two pixels, and BoW is based on small patches (e.g.,  $5 \times 5$  image patch) (see the illustration in Figure 3). Essentially, all three kinds of features are based on the counts of different-scale patterns, so we can simply calculate these features slice-by-slice and aggregate them across the whole CT volume.

### 2.3.1 | Intensity histogram features

To extract intensity histogram features, we first partitioned the pixel values into a specific number of equally



**FIGURE 3** Workflow scheme. Three kinds of radiomic features (intensity histogram, GLCM-based features, and bag-of-words features) were extracted from the CT images of patients who received ICI only, RT only, and ICI+RT. After feature selection, classification models were built on the selected features to classify patients into CIP or RP. CIP, checkpoint inhibitor-related pneumonitis; GLCM, gray-level co-occurrence matrix; ICI, immune checkpoint inhibitor; RP, radiation pneumonitis; RT, radiotherapy

spaced bins (i.e., pixel values were quantized to a specific number of gray levels) and then calculated the bin counts using the pixels within the ROI. The bin counts were L1-normalized (i.e., the sum is equal to 1) to remove the effect of ROIs having different sizes. The normalized bin counts were used as the final intensity histogram features.

### 2.3.2 | GLCM-based features

GLCM is commonly used to characterize the texture in images. A GLCM is a 2D histogram of co-occurring intensities (gray levels) at a given offset. There are two parameters involved in the construction of GLCM. One is the number of gray levels, and the other is the offset between the pixel of interest and its neighbor. For a given number of gray levels and the distance between two pixels, four GLCMs in four directions (0, 45, 90, and 135 degrees) were constructed. Based on each GLCM, four second-order statistical features (contrast, correlation, energy, and homogeneity) were calculated,<sup>34</sup> resulting in 16 texture features per image.

### 2.3.3 | BoW features

The BoW model is a feature representation method originally used in natural language processing and information retrieval. As its name implies, this model can represent a text or document by converting it into a bag of words, which is the occurrence counts of the most frequently used words. The BoW model has also been used in computer vision. In computer vision, the BoW model, sometimes called the bag-of-visual-words model, represents an image as a vector of occurrence counts of a vocabulary of local image features.

The vocabulary of local image features, equivalent to frequently used words in document classification, is usually generated by clustering local image features. The BoW representations can be obtained into three steps: extraction of local image features, construction of the visual vocabulary, and representation of images as the occurrence counts of visual words. In our work, we first used raw image patches as the local features. 2D image patches were densely sampled from the ROI and vectorized. Next, to create the visual vocabulary, we performed the k-means algorithm on the extracted local features. The words in the visual vocabulary were then defined as the learned cluster centers. Finally, for each patient, all its local features were assigned to one of the visual words via vector quantization based on Euclidean distance. The BoW feature representation of a patient is the L1-normalized counts of words.

## 2.4 | Machine-learning methods for classification

We first trained and tested different classifiers, including logistic regression, random forest, and linear SVM, on the ICI and RT datasets with 10-fold cross-validation. In each of the 10 rounds, we first performed feature selection and then trained the classification model based on the selected features using the training set. The learned classification model was then applied to the held-out test set to make predictions. After 10 rounds were completed, each sample was predicted with a label and a probability. We then applied the model trained on the ICI and RT datasets to the patients in the ICI+RT dataset.

For feature selection, we performed a two-sided Mann–Whitney U-test on each feature and selected

those with a  $p$ -value less than 0.05. We adopted the R package glmnet for logistic regression. Note that in all experiments, feature selection and model training were performed only using the training set, with the test set untouched.

## 2.5 | Evaluation metrics

The receiver operating characteristic (ROC) analysis was performed. The area under the ROC curve (AUC) and its 95% confidence interval were calculated using the R package pROC. We computed the Youden's index (defined as sensitivity + specificity - 1) for each of the points on the ROC curve and used the maximum value of this index as a criterion for selecting the optimal cut-off point. Then the accuracy, sensitivity, and specificity at the optimal cut-off point were reported. The accuracy is the proportion of samples being correctly classified. In our classification model, we considered RP as the positive class. Therefore, the sensitivity measures the proportion of RP cases that are correctly classified, and the specificity is the proportion of CIP cases that are correctly classified.

## 3 | RESULTS

### 3.1 | Experimental settings

Intensity histogram, GLCM based, and BoW features were extracted with different parameter settings. For the intensity histogram features, we tested different values for the gray level from 20, 40, 60, 80, and 100. For the GLCM-based features, we tested different values for the gray level from {20, 40, 60, 80, 100} and distance from {1, 2, 3, 4}. For the BoW features, we tested different values for the patch size from {3, 5, 7, 9, 11} and vocabulary size from {16, 32, 64, 128, 256}. For each kind of feature, we reported the results of the highest AUC. In addition, we also investigated whether including more boundary area would affect classification performance. To this end, we dilated the annotated ROI mask using a disk-shaped structure element with a radius of 5 and 10 pixels, respectively.

### 3.2 | Classification performance on the ICI and RT datasets

Based on each of the three kinds of radiomic features, classification models (logistic regression, random forest, and linear SVM) were trained and evaluated with 10-fold cross-validation on the ICI and RT datasets. Table 2 shows the classification performance of different features and ROI sizes. Using the originally annotated ROI, GLCM-based features outperformed intensity histogram

**TABLE 2** Classification performance of different kinds of features on the immune checkpoint inhibitor and radiation therapy datasets

Index	Intensity histogram			GLCM			BoW		
	LR	RF	SVM	LR	RF	SVM	LR	RF	SVM
$R = 0$ , Acc	0.644	0.729	0.661	0.797	0.746	0.797	0.746	0.797	0.797
$R = 0$ , Sen	0.645	0.774	0.839	0.968	0.710	0.871	0.548	0.871	0.742
$R = 0$ , Spe	0.643	0.679	0.464	0.607	0.786	0.714	0.964	0.714	0.857
$R = 0$ , AUC	0.608	0.677	0.634	0.848	0.758	0.817	0.815	0.813	0.834
$R = 5$ , Acc	0.644	0.780	0.593	0.797	0.729	0.814	0.915	0.780	0.881
$R = 5$ , Sen	0.968	0.839	0.968	0.806	0.581	0.871	0.903	0.710	0.839
$R = 5$ , Spe	0.286	0.714	0.179	0.786	0.893	0.750	0.929	0.857	0.929
$R = 5$ , AUC	0.560	0.763	0.517	0.821	0.764	0.829	0.937	0.865	0.926
$R = 10$ , Acc	0.610	0.780	0.661	0.797	0.763	0.831	0.814	0.814	0.831
$R = 10$ , Sen	0.871	0.774	0.839	0.774	0.710	0.903	0.742	0.710	0.871
$R = 10$ , Spe	0.321	0.786	0.464	0.821	0.821	0.75	0.893	0.929	0.786
$R = 10$ , AUC	0.505	0.765	0.563	0.789	0.784	0.825	0.894	0.866	0.884

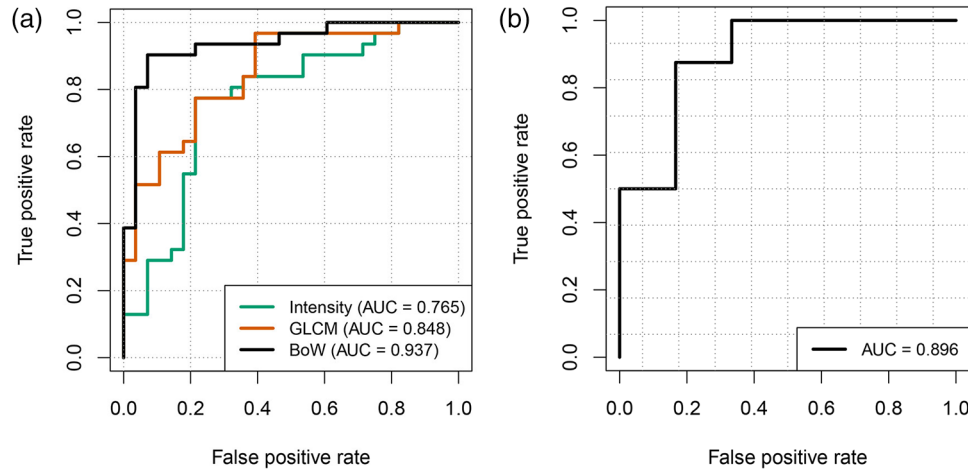
*Note:* We enlarged the ROI mask by image dilation using a disk-shaped structure element with a radius ( $R$ ) of 0, 5, and 10 pixels.  $R = 0$  means no image dilation was performed. We tested different classifiers including logistic regression (LR), random forest (RF), and linear SVM, and reported different metrics including accuracy (Acc), sensitivity (Sen), specificity (Spe), and area under ROC curve (AUC).

Abbreviations: ROC, receiver operating characteristic; ROI, region of interest.

and BoW features (AUC: 0.848 vs. 0.677 and 0.834). As the size of ROI increased, the performance of BoW features generally improved and then declined. The best performance (AUC = 0.937) was achieved when we used BoW features, logistic regression, and ROI dilation with five pixels.

Figure 4a shows the ROC curves which correspond to the best performance achieved by the three kinds of features. Using the cut-off of the classifier's output that maximized the Youden's index (sensitivity + specificity - 1), the corresponding accuracy, sensitivity, and specificity of the classifier built on BoW features were 0.915, 0.903, and 0.929, respectively. This means that 2 out of 28 patients with CIP (negative class) were misclassified into RP (positive class) while 3 out of 31 patients with RP were misclassified into CIP.

We further investigated the impact of the parameters of feature extraction methods on classification performance. The parameters of the three feature extraction methods are described in the previous section. Tables S1–S3 show the impact of different parameters for intensity histogram, GLCM based, and BoW features, respectively, when logistic regression was used. The intensity histogram features achieved the highest AUC when the number of gray levels was set to 60. The GLCM-based features achieved the best performance when the number of gray levels and the distance between two pixels were set to 60 and 3. The BoW features achieved the best performance when the patch size and vocabulary size were set to 9 and 128.



**FIGURE 4** Performance of differentiating CIP and RP (a) ROC curves for the models with the best performance on the ICI and RT datasets, using intensity histogram features, GLCM based features, and BoW features, respectively. The 95% confidence intervals are 0.638–0.892 for intensity, 0.750–0.946 for GLCM, and 0.873–1 for BoW. (b) ROC curve for classifying CIP and RP in the ICI+RT dataset. The model with the highest AUC in (a) was used. The 95% confidence interval for AUC is 0.714–1. AUC, area under curve; BoW, bag-of-words; CIP, checkpoint inhibitor-related pneumonitis; GLCM, gray-level co-occurrence matrix; ICI, immune checkpoint inhibitor; ROC, receiver operating characteristic; RP, radiation pneumonitis; RT, radiotherapy

A general observation from those tables was that performance got better when relatively larger values of the parameters were used and that, however, the performance began to decline if the parameters were too large.

### 3.3 | Assessing importance of features

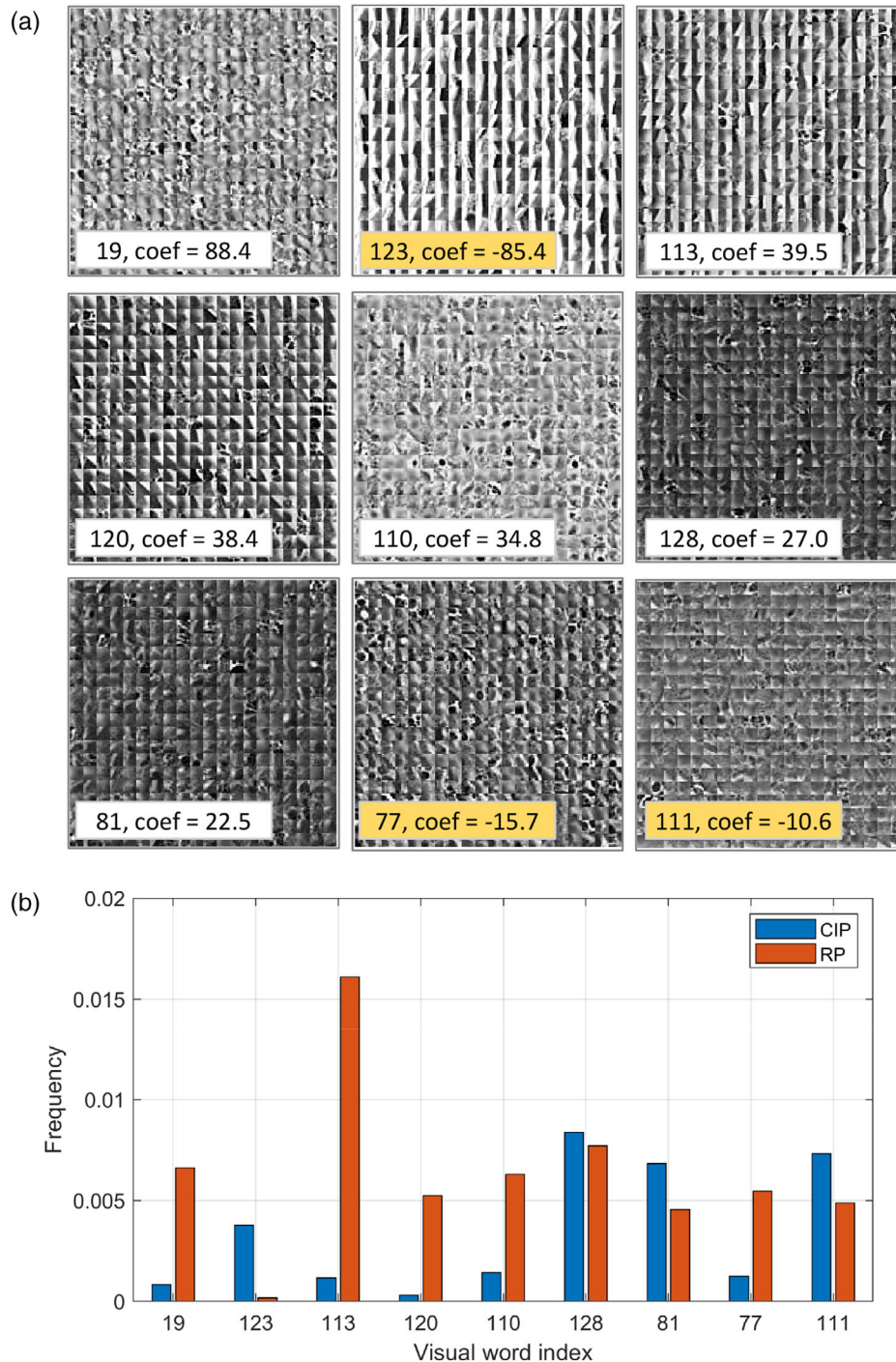
The best performance was achieved using the BoW features, logistic regression, and ROI dilation with five pixels. The patch size and vocabulary size for the BoW features were set to 9 and 128. This means the BoW features are 128-dimensional (see Section 2.3 for the details of this method). To identify the features that robustly and significantly contributed to the model, we recorded the selected features and their coefficients in each round of the 10-fold cross-validation and computed their counts of selection and mean coefficients. We selected the top nine features with the largest mean coefficients (regardless of the sign) from the features that were selected at least eight times.

The visualization of the image patches ( $9 \times 9$  pixels) that belong to each of the nine visual words is shown in Figure 5a. As we can see, the 400 ( $20 \times 20$ ) image patches in each of the nine panels present a very similar pattern. Along with each panel, the index and mean coefficient of each feature are also provided. A positive coefficient means that the corresponding image patch pattern is more likely to appear in the RP class (RP was regarded as the positive class when training classifiers), whereas a negative coefficient means that the corre-

sponding image patch pattern tends to appear more frequently in the CIP class. Figure 5b shows the average occurrence frequency of the nine visual words of three CIP patients and three RP patients that were most confidently predicted by our model. We can see clearly that for the top three most significant visual words, the 19th and 113th visual words have a much higher frequency in RP than in CIP, while the 123rd visual word is the opposite.

### 3.4 | Evaluation in patients receiving both ICI and RT

To further validate our method, we applied the classification model with the highest AUC on the ICI and RT datasets to the patients in the ICI+RT dataset, in which patients received both treatments. Performance on the ICI+RT dataset was evaluated using clinicians' diagnosis as a reference, and the cause of pneumonitis was diagnosed on the basis of radiologic features, clinical symptoms, and onset time of pneumonitis. Three radiation oncologists participated in the diagnosis independently, and the final class label of each patient was determined by majority voting. Table 3 provides a summary of each oncologist's diagnosis, final voting result, and our model's prediction. The three oncologists made the exact same diagnosis for 8 out of 14 patients (Fleiss's kappa = 0.417). Our model generalized well on the ICI+RT dataset, achieving an accuracy of 0.857 and AUC of 0.896. The corresponding ROC curve is provided in Figure 4b.



**FIGURE 5** Visualization of the nine visual words that most robustly and significantly contributed to our classification model. (a) Example image patches showing different patterns for each visual word. The index and mean coefficient of each visual word are shown in the white box. (b) Bar graph of the average occurrence frequency of the nine visual words of three CIP patients and three RP patients that were most confidently predicted by our model. The visual words with positive coefficients are more likely to appear in the positive class (i.e., RP) such as visual words 19, 113, and 120. CIP, checkpoint inhibitor-related pneumonitis; RP, radiation pneumonitis

## 4 | DISCUSSION

In the setting of concomitant ICIs with RT, the distinction between CIP and RP is crucial to subsequent treatment decisions because it is much safer for a patient to restart

ICI therapy after experiencing RP. Prior work has documented the radiologic patterns and clinical symptoms of CIP<sup>36–38</sup> and RP,<sup>39,40</sup> but the characteristics of these two kinds of pneumonitis can mimic each other. Currently, distinguishing CIP from RP poses a great challenge



**TABLE 3** Summary of radiation oncologists' diagnosis, majority voting result, and model's prediction for each patient in the ICI+RT dataset

Patient index	Oncologist 1	Oncologist 2	Oncologist 3	Majority voting	Model-predicted probability of being RP
1	CIP	CIP	RP	CIP	0.255
3	CIP	CIP	CIP	CIP	0.954
7	CIP	CIP	CIP	CIP	0.393
9	RP	CIP	CIP	CIP	0.761
13	RP	CIP	CIP	CIP	0.504
14	CIP	CIP	CIP	CIP	0.144
2	CIP	RP	RP	RP	0.892
4	RP	RP	RP	RP	0.972
5	RP	RP	RP	RP	1.000
6	RP	RP	CIP	RP	0.708
8	RP	RP	RP	RP	0.995
10	RP	RP	RP	RP	0.789
11	CIP	RP	RP	RP	0.913
12	RP	RP	RP	RP	1.000

Abbreviations: CIP, checkpoint inhibitor-related pneumonitis; ICI, immune checkpoint inhibitor; RP, radiation pneumonitis; RT, radiotherapy.

for clinicians. In this study, we investigated whether CT radiomic features can help differentiate between these two kinds of pneumonitis. We developed a workflow for the generation of a rich set of quantitative features to characterize the texture of inflammatory lesions in CT images. Based on these features, we trained a classification model using the ICI and RT datasets and applied this model to the patients in the ICI+RT dataset. Ten-fold cross validation on the training set and evaluation on the independent test set demonstrated the efficacy of our method with AUCs of 0.937 and 0.896, respectively.

Three kinds of features were tested: intensity histogram features, GLCM-based features, and BoW features. We found that the BoW features yielded the best cross-validation performance with an AUC of 0.937, followed by the GLCM-based features (AUC = 0.848) and the intensity histogram features (AUC = 0.765). The distinction of performance is expected as the three kinds of features characterize the texture of image content at different scales. Intensity histogram features are based on individual pixels and completely ignore the information of surrounding pixels, thereby leading to the worst results. GLCM-based features describe the pairwise relationships between two pixels and thus provide better results. BoW features deal with a group of adjacent pixels. Therefore, the BoW features are more informative and discriminative.

To further show the effectiveness of the radiomic features, we compared the diagnostic performance of our model and a radiation oncologist on the ICI and RT datasets. The ICI and RT datasets were used because there is no ambiguity of the cause of pneumonitis since

the patients received either ICI or RT. For a fair comparison, the oncologist made a diagnosis only based on the CT images without referring to other clinical information, which is the same as our method. The classification by the oncologist achieved an AUC of 0.777, which is inferior to our BoW feature-based model (AUC = 0.937). These results provide compelling evidence that our radiomics approach can discover quantitative and discriminative features to effectively distinguish CIP from RP, which are difficult for humans to notice.

Attributing the cause of the pneumonitis in patients receiving both ICI and RT can be a very difficult task, which can be seen from the diagnoses by the three radiation oncologists. As shown in Table 3, the oncologists made the same diagnosis for only 8 out of 14 patients (57.14%, Fleiss's kappa = 0.417). The patients with consensus among oncologists can be easily diagnosed by some clear evidence. For example, clear evidence suggesting RP includes that pneumonitis is only seen in the high-dose area and that the onset time of pneumonitis is close to the completion of RT and is far from the administration of ICI, and vice versa for the evidence suggesting CIP. However, not all patients exhibit clear evidence; findings of RP and CIP are varied, overlapped, and sometimes non-specific. This means that clinician's diagnosis for some patients in the ICI+RT dataset may not be accurate. For this reason, we train our classification model using the ICI and RT datasets, in which each patient has a definite diagnosis and solicit the diagnosis from multiple clinicians to reduce potential diagnostic bias.

To the best of our knowledge, there are very few studies focusing on this topic. We only found a rele-

vant abstract published in 2020 that used CT radiomics and machine learning for distinguishing between CIP and RP. Chen et al.<sup>41</sup> used a general package (PyRadiomics) to extract radiomic features, trained a random forest classifier in patients who received ICI ( $n = 23$ ) and RT ( $n = 29$ ) only, and tested the classifier in patients who received ICI+RT ( $n = 30$ ). The random classifier achieved an AUC of 0.79 on the training set and an AUC of 0.84 on the test set. Our method achieved better performance with AUCs of 0.937 and 0.896 on our training and test sets, respectively, which can be attributed to the more powerful radiomic features used in our method.

A limitation of the present study is that although our method was rigorously validated using 10-fold cross-validation on the training set and further tested using an independent dataset, this study was conducted using data from a single institution. Future work will focus on collecting more in-house samples and samples from different institutions as an external validation set. A prospective study is being designed to rechallenge ICI in the patients who are classified as RP cases by our model.

## 5 | CONCLUSIONS

In summary, the wide spectrum of radiologic manifestations of CIP and RP poses great diagnostic and management challenges in clinical practice. Our results demonstrated that using CT radiomics and machine learning can successfully distinguish CIP from RP with a high accuracy (AUC of 0.896 on an independent test set). This indicates that our method has the potential to be a useful tool for identifying the RP patients from the patients with pneumonitis who receive both ICI therapy and RT, which has significant implications in improving patient management.

## ACKNOWLEDGMENTS

The authors thank Haiyan Tu, Biao Huang, and Jine Zhang from Guangdong Provincial People's Hospital, China, for retrieving and reviewing the chest CT scans of patients. This study was supported by National Natural Science Foundation of China (61901275), Guangzhou Science and Technology Plan Foundation (2021-02-01-04-1002-0017), Shenzhen University Startup Fund (2019131), National Key R&D Program of China (2019YFC0118300), Shenzhen Peacock Plan (KQTD2016053112051497 and KQJSCX20180328095606003), Medical Scientific Research Foundation of Guangdong Province, China (B2018031 and B2020024), and Supporting start-up funds of National Natural Science Foundation of Guangdong Provincial People's Hospital (8210032051).

The funding sources have no involvement in the study.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## REFERENCES

1. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer. *N Engl J Med*. 2017;377(20):1919-1929. <https://doi.org/10.1056/NEJMoa1709937>.
2. Antonia SJ, Villegas A, Daniel D, et al. Overall survival with durvalumab after chemoradiotherapy in stage III NSCLC. *N Engl J Med*. 2018;379(24):2342-2350. <https://doi.org/10.1056/NEJMoa1809697>.
3. Lin SH, Lin Y, Yao L, et al. Phase II trial of concurrent atezolizumab with chemoradiation for unresectable NSCLC. *J Thorac Oncol*. 2020;15(2):248-257. <https://doi.org/10.1016/j.jtho.2019.10.024>.
4. Peters S, Felip E, Dafni U, et al. Safety evaluation of nivolumab added concurrently to radiotherapy in a standard first line chemoradiotherapy regimen in stage III non-small cell lung cancer—The ETOP NICOLAS trial. *Lung Cancer*. 2019;133:83-87. <https://doi.org/10.1016/j.lungcan.2019.05.001>.
5. Nishino M, Giobbie-Hurder A, Hatabu H, Ramaiya NH, Hodi FS. Incidence of programmed cell death 1 inhibitor-related pneumonitis in patients with advanced cancer: a systematic review and meta-analysis. *JAMA Oncol*. 2016;2(12):1607-1616. <https://doi.org/10.1001/jamaoncol.2016.2453>.
6. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 2016;375:1823-1833. <https://doi.org/10.1056/NEJMoa1606774>.
7. Gandhi L, Rodriguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med*. 2018;378:2078-2092. <https://doi.org/10.1056/NEJMoa1801005>.
8. 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(5):E435-E444. <https://doi.org/10.1016/j.clcc.2020.02.025>.
9. Curran WJ, Paulus R, Langer CJ, et al. Sequential vs concurrent chemoradiation for stage III non-small cell lung cancer: randomized phase III trial RTOG 9410. *J Natl Cancer Inst*. 2011;103(19):1452-1460. <https://doi.org/10.1093/jnci/djr325>.
10. Bradley JD, Paulus R, Komaki R, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. *Lancet Oncol*. 2015;16(2):187-199. [https://doi.org/10.1016/S1470-2045\(14\)71207-0](https://doi.org/10.1016/S1470-2045(14)71207-0).
11. Wang S, Liao Z, Wei X, et al. Association between systemic chemotherapy before chemoradiation and increased risk of treatment-related pneumonitis in esophageal cancer patients treated with definitive chemoradiotherapy. *J Thorac*

- Oncol.* 2008;3(3):277-282. <https://doi.org/10.1097/JTO.0b013e3181653ca6>.
12. Shaverdian N, Lisberg AE, Bornazyan K, et al. Previous radiotherapy and the clinical activity and toxicity of pembrolizumab in the treatment of non-small-cell lung cancer: a secondary analysis of the KEYNOTE-001 phase 1 trial. *Lancet Oncol.* 2017;18(7):895-903. [https://doi.org/10.1016/S1470-2045\(17\)30380-7](https://doi.org/10.1016/S1470-2045(17)30380-7).
  13. Voong KR, Hazell SZ, Fu W, et al. Relationship between prior radiotherapy and checkpoint-inhibitor pneumonitis in patients with advanced non-small-cell lung cancer. *Clin Lung Cancer.* 2019;20(4):e470-e479. <https://doi.org/10.1016/j.clcc.2019.02.018>.
  14. Saito G, Oya Y, Taniguchi Y, et al. Real-world survey of pneumonitis/radiation pneumonitis among patients with locally advanced non-small cell lung cancer treated with chemoradiotherapy after durvalumab approval: a multicenter retrospective cohort study (HOPE-005/CRIMSON). *J Clin Oncol.* 2020;38(15 suppl):9039-9039. [https://doi.org/10.1200/jco.2020.38.15\\_suppl.9039](https://doi.org/10.1200/jco.2020.38.15_suppl.9039).
  15. Naidoo J, Wang X, Woo KM, et al. Pneumonitis in patients treated with anti-programmed death-1/programmed death ligand 1 therapy. *J Clin Oncol.* 2017;35(7):709-717. <https://doi.org/10.1200/JCO.2016.68.2005>.
  16. Tirumani SH, Ramaiya NH, Keraliya A, et al. Radiographic profiling of immune-related adverse events in advanced melanoma patients treated with ipilimumab. *Cancer Immunol Res.* 2015;3(10):1185-1192. <https://doi.org/10.1158/2326-6066.CIR-15-0102>.
  17. Bradley J, Movsas B. Radiation pneumonitis and esophagitis in thoracic irradiation. *Cancer Treat Res.* 2006;128:43-64. [https://doi.org/10.1007/0-387-25354-8\\_4](https://doi.org/10.1007/0-387-25354-8_4).
  18. Brahmer JR, Lacchetti C, Schneider BJ, et al. Management of immune-related adverse events in patients treated with immune checkpoint inhibitor therapy: American Society of Clinical Oncology clinical practice guideline. *J Clin Oncol.* 2018;36(7):1714-1768. <https://doi.org/10.1200/JCO.2017.77.6385>.
  19. Schoenfeld JD, Nishino M, Severgnini M, Manos M, Mak RH, Hodi FS. Pneumonitis resulting from radiation and immune checkpoint blockade illustrates characteristic clinical, radiologic and circulating biomarker features. *J Immunother Cancer.* 2019;7(1):112. <https://doi.org/10.1186/s40425-019-0583-3>.
  20. Uthoff J, Nagpal P, Sanchez R, Gross TJ, Lee C, Sieren JC. Differentiation of non-small cell lung cancer and histoplasmosis pulmonary nodules: insights from radiomics model performance compared with clinician observers. *Transl Lung Cancer Res.* 2019;8(6):979-988. <https://doi.org/10.21037/tlcr.2019.12.19>.
  21. Beig N, Khorrami M, Aliou M, et al. Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas. *Radiology.* 2019;290(3):783-792. <https://doi.org/10.1148/radiol.2018180910>.
  22. Xu X, Zhang HL, Liu QP, et al. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. *J Hepatol.* 2019;70(6):1133-1144. <https://doi.org/10.1016/j.jhep.2019.02.023>.
  23. Jiang Y, Jin C, Yu H, et al. Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer. *Ann Surg.* 2020;274(6):e1153-e1161. <https://doi.org/10.1097/sla.0000000000003778>.
  24. Khorrami M, Jain P, Bera K, et al. Predicting pathologic response to neoadjuvant chemoradiation in resectable stage III non-small cell lung cancer patients using computed tomography radiomic features. *Lung Cancer.* 2019;135:1-9. <https://doi.org/10.1016/j.lungcan.2019.06.020>.
  25. Bera K, Velcheti V, Madabhushi A. Novel quantitative imaging for predicting response to therapy: techniques and clinical applications. *Am Soc Clin Oncol Educ B.* 2018;38(38):1008-1018. [https://doi.org/10.1200/edbk\\_199747](https://doi.org/10.1200/edbk_199747).
  26. Xu Y, Hosny A, Zeleznik R, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res.* 2019;25(11):3266-3275. <https://doi.org/10.1158/1078-0432.CCR-18-2495>.
  27. Chen A, Lu L, Pu X, et al. CT-based radiomics model for predicting brain metastasis in category T1 lung adenocarcinoma. *Am J Roentgenol.* 2019;213(1):134-139. <https://doi.org/10.2214/AJR.18.20591>.
  28. Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol.* 2015;114(3):345-350. <https://doi.org/10.1016/j.radonc.2015.02.015>.
  29. 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(9):1180-1191. [https://doi.org/10.1016/S1470-2045\(18\)30413-3](https://doi.org/10.1016/S1470-2045(18)30413-3).
  30. Rios Velazquez E, Parmar C, Liu Y, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res.* 2017;77(14):3922-3930. <https://doi.org/10.1158/0008-5472.CAN-17-0122>.
  31. Rizzo S, Petrella F, Buscarino V, et al. CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer. *Eur Radiol.* 2016;26(1):32-42. <https://doi.org/10.1007/s00330-015-3814-0>.
  32. Chuzi S, Tavora F, Cruz M, et al. Clinical features, diagnostic challenges, and management strategies in checkpoint inhibitor-related pneumonitis. *Cancer Manag Res.* 2017;9:207-213. <https://doi.org/10.2147/CMAR.S136818>.
  33. Bledsoe TJ, Nath SK, Decker RH. Radiation pneumonitis. *Clin Chest Med.* 2017;38(2):201-208. <https://doi.org/10.1016/j.ccm.2016.12.004>.
  34. Torheim T, Malinen E, Kvaal K, et al. Classification of dynamic contrast enhanced MR images of cervical cancers using texture analysis and support vector machines. *IEEE Trans Med Imaging.* 2014;33(8):1648-1656. <https://doi.org/10.1109/TMI.2014.2321024>.
  35. Cheng J, Huang W, Cao S, et al. Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PLoS ONE.* 2015;10(10):e0140381. <https://doi.org/10.1371/journal.pone.0140381>.
  36. Nishino M, Ramaiya NH, Awad MM, et al. PD-1 inhibitor-related pneumonitis in advanced cancer patients: radiographic patterns and clinical course. *Clin Cancer Res.* 2016;22(24):6051-6060. <https://doi.org/10.1158/1078-0432.CCR-16-1320>.
  37. Cho JY, Kim J, Lee JS, et al. Characteristics, incidence, and risk factors of immune checkpoint inhibitor-related pneumonitis in patients with non-small cell lung cancer. *Lung Cancer.* 2018;125:150-156. <https://doi.org/10.1016/j.lungcan.2018.09.015>.
  38. Cadranet J, Canellas A, Matton L, et al. Pulmonary complications of immune checkpoint inhibitors in patients with non-small cell lung cancer. *Eur Respir Rev.* 2019;28(153):190058. <https://doi.org/10.1183/16000617.0058-2019>.
  39. Bernchou U, Christiansen RL, Asmussen JT, Schytte T, Hansen O, Brink C. Extent and computed tomography appearance of early radiation induced lung injury for non-small cell lung cancer. *Radiother Oncol.* 2017;123(1):93-98. <https://doi.org/10.1016/j.radonc.2017.02.001>.
  40. Thomas R, Chen YH, Hatabu H, Mak RH, Nishino M. Radiographic patterns of symptomatic radiation pneumonitis in lung cancer patients: imaging predictors for clinical severity and outcome. *Lung Cancer.* 2020;145:132-139. <https://doi.org/10.1016/j.lungcan.2020.03.023>.
  41. Chen X, Sheikh K, Lin CT, et al. CT radiomics and machine learning for distinguishing radiotherapy vs. immune checkpoint

inhibitor-induced pneumonitis in non-small cell lung cancer patients. *Int J Radiat Oncol.* 2020;108(3):S163. <https://doi.org/10.1016/j.ijrobp.2020.07.929>.

### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** 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;49:1547–1558. <https://doi.org/10.1002/mp.15451>