



## OPEN ACCESS

## EDITED BY

Feng Gao,  
The Sixth Affiliated Hospital of Sun  
Yat-sen University, China

## REVIEWED BY

Shumao Pang,  
Guangzhou Medical University, China  
Yao Wu,  
Children's National Hospital,  
United States

## \*CORRESPONDENCE

Shasha Du,  
duss0202@163.com  
Jun Cheng,  
chengjun583@qq.com

<sup>†</sup>These authors have contributed equally  
to this work

## SPECIALTY SECTION

This article was submitted to  
Computational Physiology and  
Medicine,  
a section of the journal  
Frontiers in Physiology

RECEIVED 25 June 2022

ACCEPTED 05 July 2022

PUBLISHED 25 July 2022

## CITATION

Tan P, Huang W, Wang L, Deng G,  
Yuan Y, Qiu S, Ni D, Du S and Cheng J  
(2022), Deep learning predicts immune  
checkpoint inhibitor-related  
pneumonitis from pretreatment  
computed tomography images.  
*Front. Physiol.* 13:978222.  
doi: 10.3389/fphys.2022.978222

## COPYRIGHT

© 2022 Tan, Huang, Wang, Deng, Yuan,  
Qiu, Ni, Du and Cheng. This is an open-  
access article distributed under the  
terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Deep learning predicts immune checkpoint inhibitor-related pneumonitis from pretreatment computed tomography images

Peixin Tan<sup>1†</sup>, Wei Huang<sup>1†</sup>, Lingling Wang<sup>2,3,4†</sup>, Guanhua Deng<sup>5</sup>,  
Ye Yuan<sup>2,3,4</sup>, Shili Qiu<sup>2,3,4</sup>, Dong Ni<sup>2,3,4</sup>, Shasha Du<sup>1\*</sup> and  
Jun Cheng<sup>2,3,4\*</sup>

<sup>1</sup>Department of Radiation Oncology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, <sup>2</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>3</sup>Medical Ultrasound Image Computing (MUSIC) Laboratory, Shenzhen University, Shenzhen, China, <sup>4</sup>Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen, China, <sup>5</sup>Department of Oncology, Guangdong Sanjiu Brain Hospital, Guangzhou, China

Immune checkpoint inhibitors (ICIs) have revolutionized the treatment of lung cancer, including both non-small cell lung cancer and small cell lung cancer. Despite the promising results of immunotherapies, ICI-related pneumonitis (ICIP) is a potentially fatal adverse event. Therefore, early detection of patients at risk for developing ICIP before the initiation of immunotherapy is critical for alleviating future complications with early interventions and improving treatment outcomes. In this study, we present the first reported work that explores the potential of deep learning to predict patients who are at risk for developing ICIP. To this end, we collected the pretreatment baseline CT images and clinical information of 24 patients who developed ICIP after immunotherapy and 24 control patients who did not. A multimodal deep learning model was constructed based on 3D CT images and clinical data. To enhance performance, we employed two-stage transfer learning by pre-training the model sequentially on a large natural image dataset and a large CT image dataset, as well as transfer learning. Extensive experiments were conducted to verify the effectiveness of the key components used in our method. Using five-fold cross-validation, our method accurately distinguished ICIP patients from non-ICIP patients, with area under the receiver operating characteristic curve of 0.918 and accuracy of 0.920. This study demonstrates the promising potential of deep learning to identify patients at risk for developing ICIP. The proposed deep learning model enables efficient risk stratification, close monitoring, and prompt management of ICIP, ultimately leading to better treatment outcomes.

## KEYWORDS

immune checkpoint inhibitor-related pneumonitis, deep learning, transfer learning, contrastive learning, CT images, lung cancer

## 1 Introduction

Since the first immune checkpoint inhibitor (ICI) ipilimumab was approved by the Food and Drug Administration for treating melanoma in 2011, ICIs have become standard treatments for many cancers such as lung cancer, renal cell carcinoma, Hodgkin lymphoma, and hepatocellular carcinoma (Akinleye and Rasool, 2019; Xin Yu et al., 2019; Robert, 2020; Vaddepally et al., 2020). Although ICIs produce remarkable immune response by immune upregulation and demonstrate improved cancer-related outcomes, they induce a unique spectrum of toxicities, called immune-related adverse events (irAEs) (Martins et al., 2019; Ramos-Casals et al., 2020). These irAEs can occur in multiple organ systems, where uncontrolled immune response is generated against healthy tissue. Due to different organ systems affected, there are various types of irAEs, including dermatitis, encephalitis, uveitis, hepatitis, and pneumonitis. Among these, immune checkpoint inhibitor-related pneumonitis (ICIP) is one of the most concerned adverse events because it is potentially life-threatening (Naidoo et al., 2017).

Lung cancer is the second most common cancer and the leading cause of cancer death worldwide. ICIs have shown significant clinical benefit in the treatment of advanced non-small cell lung cancer (NSCLC) (Reck et al., 2016). The incidence of ICIP in NSCLC is 4.1% as reported in a prospective study, while some real-world studies outside of clinical trials report a much higher incidence, ranging from 7% to 19% (Sears et al., 2019; Atchley et al., 2021). The time to onset of ICIP can vary from 9 days to 24.3 months after the initiation of immunotherapy (Nishino et al., 2015; Naidoo et al., 2017), with a median time of 52.5 days (Atchley et al., 2021). Most patients with ICIP have high severity that requires hospitalization, and about 27% of them die during the treatment for ICIP (Atchley et al., 2021). Unfortunately, the pathogenesis of ICIP has not been clearly elucidated. Possible risk factors include prior thoracic radiotherapy, pulmonary comorbidities, smoking status, and PD-1 inhibitors (Howell et al., 2015; Delaunay et al., 2017; Khunger et al., 2017; Pillai et al., 2018; Winer et al., 2018). However, it is challenging to accurately predict ICIP based on these clinical risk factors. In order to improve lung cancer treatment and outcomes, there is an urgent need for early prediction of ICIP, which enables risk stratification before starting immunotherapy and allows a close monitoring of high-risk patients during treatment.

Radiomics is a rapidly evolving research area in personalized precision medicine that aims to extract informative radiomic features from medical images and relate these features to clinical and biological endpoints. Computed tomography (CT) is routinely used for diagnosing lung cancer and assessing treatment response. CT-based quantitative radiomics approaches have been successfully applied to various tasks, such as lesion classification (Naidoo et al., 2017; Gitto et al.,

2021; Xu et al., 2021), prediction of prognosis and treatment response (van Timmeren et al., 2017; Yang et al., 2019; Chetan and Gleeson, 2021), and genotype-phenotype associations (Rios Velazquez et al., 2017; Thawani et al., 2018; Zanfardino et al., 2019; Wu et al., 2021). There are very few studies focusing on the prediction of ICIP using radiomics. To the best of our knowledge, we only found two closely related ones. The first study reported a 100% accuracy of classification based on baseline chest CT images, but only two ICIP patients were enrolled (Colen et al., 2018). Mu et al. (2020) performed a radiomics analysis of PET/CT images to predict severe immune-related adverse events and achieved an area under the receiver operating characteristic curve (AUC) of 0.88 in a prospective validation cohort (Mu et al., 2020).

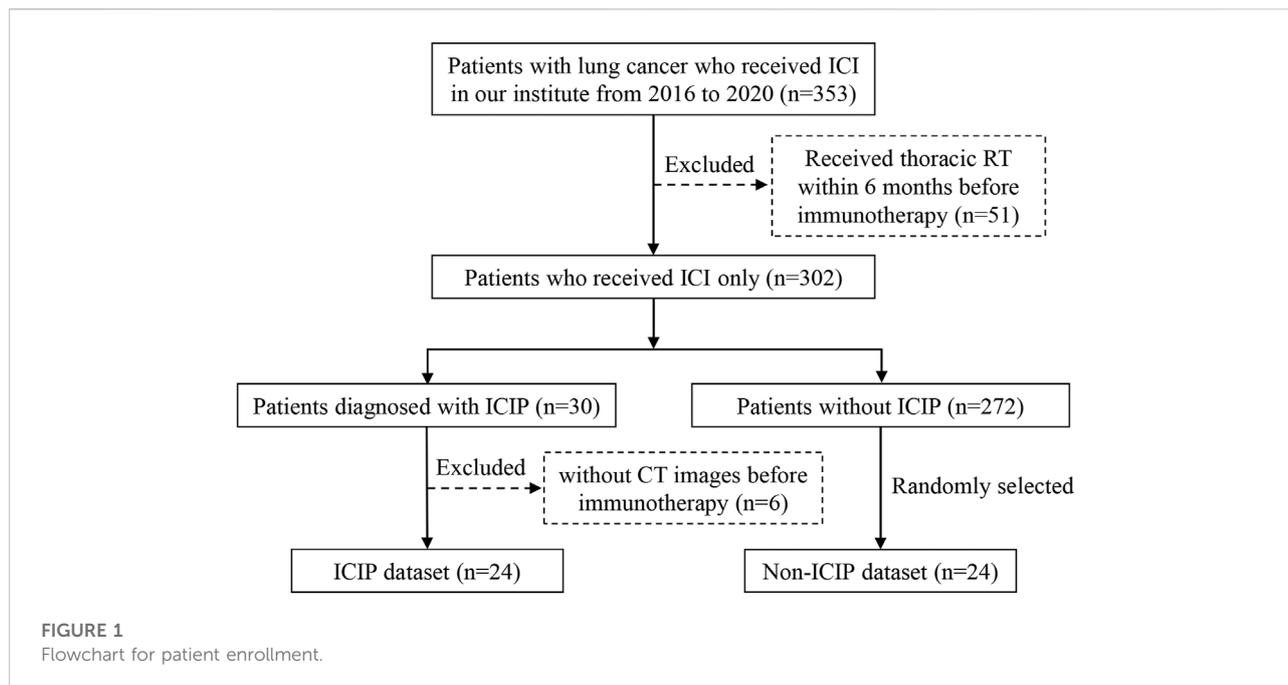
Predicting ICIP by conventional radiomics methods has two limitations regarding to the two steps in radiomics analysis pipeline. Radiomics requires first the segmentation of region of interest (ROI) and then the extraction of a fixed set of features from ROI. The first limitation is that it is unclear what region in pretreatment CT images should be used as ROI due to the lack of guidance for regional predilection of pneumonitis. The second limitation is that the predefined set of features may not be optimal for the final prediction task. Recent studies have demonstrated the excellent performance of deep learning models in computer-aided diagnosis (Yu et al., 2020; Zheng et al., 2020; Cheng et al., 2022b; Qian et al., 2022). Compared with hand-crafted features, deep learning models can directly learn discriminative features from images without prior segmentation of ROI and thus may provide a better prediction of ICIP.

In this study, we aim to develop a deep learning model based on clinical data and pretreatment chest CT images to predict the risk of ICIP in lung cancer patients. To this end, we collect a relatively large dataset consisting of ICIP and non-ICIP patients and propose a deep learning model in which multimodal data, two-stage transfer learning, and contrastive learning are used. Extensive experiments are conducted to assess the performance of different settings. The results demonstrate that the use of the aforementioned three strategies is effective and achieves state-of-the-art performance with an AUC of 0.918.

## 2 Materials and methods

### 2.1 Patients and data collection

This study was approved by the Ethics Committee of Guangdong Provincial People's Hospital, and the requirement for informed consent was waived. Figure 1 shows the detailed inclusion and exclusion criteria for preparing the patient cohort. A total of 353 lung cancer patients were treated with ICIs between January 2016 and December 2020 at our institute. We excluded 51 patients who received thoracic radiotherapy because radiotherapy can induce radiation pneumonitis which is



difficult to be distinguished from ICIP. Among the remaining patients, 30 of them developed ICIP, resulting in an incidence of 8.50% which is comparable to the data reported in previous studies (Sears et al., 2019; Atchley et al., 2021). We used the same criteria in a previous study to define ICIP (Cheng et al., 2022a). After excluding six patients who did not have CT scans before the start of immunotherapy, we finally got 24 patients for the ICIP dataset. To match the sample size of the ICIP dataset, we randomly chose 24 patients who did not develop ICIP to construct the control dataset, i.e., the non-ICIP dataset.

The collected chest CT images are within 6 months before the start of immunotherapy, which were produced by two different scanners, Philips iCT 256 and Philips ingenuity CT. Thoracic CT scans containing the whole lung were analyzed using a multi-slice helical technique at 120 kVp, mean exposure of 205 mAs, axial resolution of 5 mm, and mean in-plane resolution of 0.8174 mm.

## 2.2 Development of the deep learning model

### 2.2.1 Data preprocessing

For 3D CT scans, cropping, padding, and resizing techniques were used to convert the CT volume into a  $192 \times 192 \times 224$  matrix as the network input. We used a suitable window width for lung tissue from -500 to 1500 Hounsfield units (Zhang et al., 2020) to linearly rescale the pixel value to (0, 1) by the min-max method. Common data augmentation techniques including random flipping, noise, and affine

transformation were used. For clinical information, as shown in Table 1, categorical variables were converted to distinct numbers so as to be input to models.

### 2.2.2 Network architecture

The overall network architecture is shown in Figure 2. To predict ICIP, we first built an image network and a clinical network based on pretreatment CT images and clinical data, respectively. Due to the relatively small size of our dataset, a lightweight network, 3D ResNet18, was chosen as the backbone of the image network. Then, a multimodal fusion network was constructed by combining the nine clinical features (Table 1) and the image features learned from the image network. The clinical features and image features were fused by direct concatenation. Cross-entropy loss was used to supervise the ICIP prediction task. To enhance the prediction performance, two-stage transfer learning and contrastive learning strategies were used, which are introduced in the following sections.

### 2.2.3 Transfer learning

A two-stage transfer learning strategy inspired by (Altaf et al., 2021) was used to train our image network. We first downloaded the pre-trained model which was built using two massive natural image datasets (Kay et al., 2017; Monfort et al., 2020). The pre-trained weights may not be appropriate for our ICIP prediction task due to distributional shift between natural images and medical images. Therefore, in the second stage we fine-tuned the network using a large CT image dataset associated with

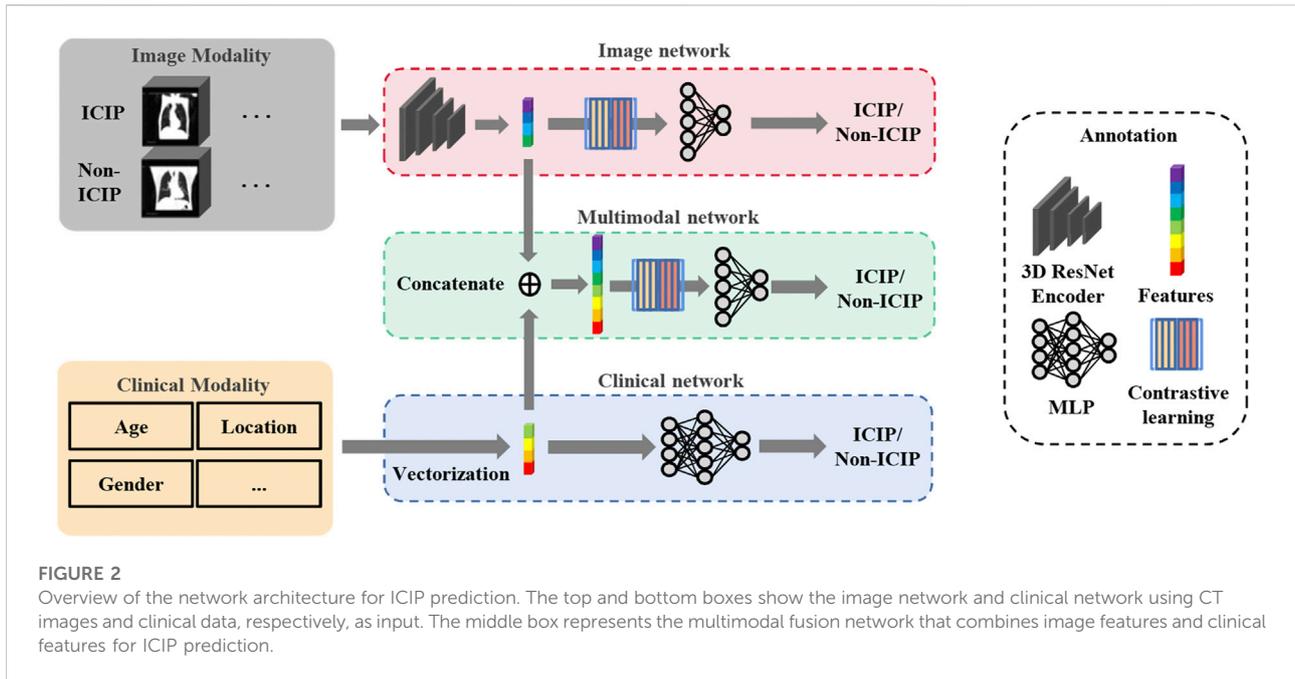
TABLE 1 Patient characteristics. *p* values less than 0.05 are highlighted with an asterisk.

Characteristic	ICIP dataset ( <i>n</i> = 24)	Non-ICIP dataset ( <i>n</i> = 24)	<i>p</i> value
<b>Sex</b>			0.022*
Female	0 (0.0%)	6 (25.0%)	
Male	24 (100.0%)	18 (75.0%)	
<b>Age</b>			0.261
Median	60	59	
Range	38–75	37–77	
<b>Lesion location</b>			0.9999
Upper left	9 (37.5%)	10 (41.7%)	
Upper right	9 (37.5%)	7 (29.2%)	
Lower left	3 (12.5%)	2 (8.3%)	
Lower right	3 (12.5%)	3 (12.5%)	
Mediastinal	0 (0.0%)	1 (4.2%)	
Middle right	0 (0.0%)	1 (4.2%)	
<b>Histologic type</b>			0.337
Adenocarcinoma	16 (66.7%)	16 (66.7%)	
Squamous cell carcinoma	7 (29.2%)	4 (16.7%)	
Adenosquamous carcinoma	0 (0.0%)	1 (4.2%)	
Small cell endocrine carcinoma	1 (4.2%)	0 (0.0%)	
Large cell endocrine carcinoma	0 (0.0%)	1 (4.2%)	
Lymphoepithelioma-like carcinoma	0 (0.0%)	2 (8.3%)	
<b>T stage</b>			0.496
T0	1 (4.2%)	1 (4.2%)	
T1	5 (20.8%)	2 (8.3%)	
T2	6 (25.0%)	11 (45.8%)	
T3	4 (16.7%)	2 (8.3%)	
T4	8 (33.3%)	8 (33.3%)	
<b>N stage</b>			0.233
N0	1 (4.2%)	2 (8.3%)	
N1	0 (0.0%)	0 (0.0%)	
N2	13 (54.2%)	7 (29.2%)	
N3	10 (41.7%)	15 (62.5%)	
<b>M stage</b>			0.461
M0	6 (25.0%)	3 (12.5%)	
M1	18 (75.0%)	21 (87.5%)	
<b>Surgery before immunotherapy</b>			0.666
Yes	4 (16.7%)	2 (8.3%)	
No	20 (83.3%)	22 (91.7%)	
<b>Radiotherapy before immunotherapy</b>			0.023*
Yes	8 (33.3%)	1 (4.2%)	
No	16 (66.7%)	23 (95.8%)	

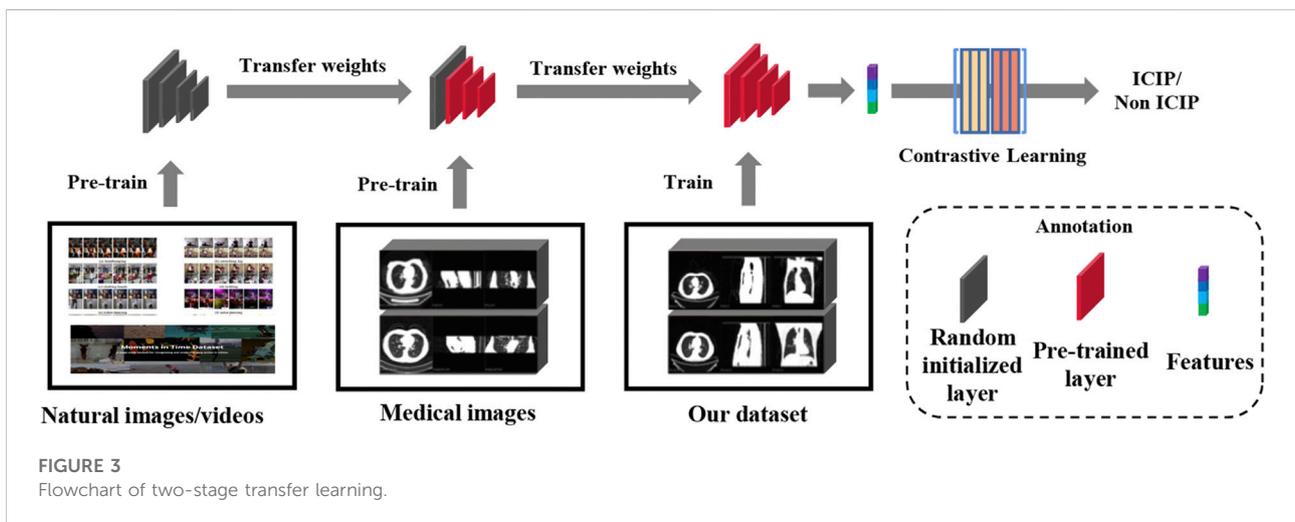
pneumonia (CC-CCII dataset) (Zhang et al., 2020). After the transfer of knowledge from a related task, the domain gap between the source and target tasks was significantly reduced. The two-stage transfer learning flowchart is shown in Figure 3.

## 2.2.4 Contrastive learning

Besides transfer learning, contrastive learning was also adopted to further boost the performance of the image network. The key idea of contrastive learning is to learn an embedding space in which positive sample pairs stay close to each



**FIGURE 2**  
Overview of the network architecture for ICIP prediction. The top and bottom boxes show the image network and clinical network using CT images and clinical data, respectively, as input. The middle box represents the multimodal fusion network that combines image features and clinical features for ICIP prediction.



**FIGURE 3**  
Flowchart of two-stage transfer learning.

other while negative ones are far apart. In essence, contrastive learning allows the model to learn high-level features about the data. The contrastive learning can be broken into three basic steps: sample pair construction, encoding, and loss minimization of representations. In our study, positive sample pairs were samples from the same class while negative sample pairs were samples from different classes. We then used 3D ResNet18 to encode the images as vector representations (Figure 2). Lastly, we maximized the similarity of the two vector representations of the positive sample pair and minimized that of the negative sample pair by minimizing a contrastive loss function. We took the

cosine as the similarity metric. The contrastive loss function is defined by the following equations:

$$sim(x_1, x_2) = \frac{x_1 \cdot x_2}{\|x_1\| \cdot \|x_2\|}, \tag{1}$$

$$loss_{CL} = \sum_x sim(x, x_-) - \sum_x sim(x, x_+). \tag{2}$$

Here,  $sim$  represents the cosine similarity metric ( $x, x_-$ ) and ( $x, x_+$ ) respectively denote the negative sample pair and positive sample pair. The overall loss function is defined by

$$loss_{all} = loss_{CE} + \gamma loss_{CL}, \tag{3}$$

where  $loss_{CE}$  is the typical cross entropy loss, and  $\gamma$  is a hyper-parameter to control the weight of contrastive learning loss.

## 2.3 Implementation details

The proposed method was implemented using PyTorch on a workstation equipped with four NVIDIA RTX A6000 GPUs (48 GB memory each). In all comparative studies, we employed ResNet18 (He et al., 2016) as the backbone. Adam optimizer with the learning rate of  $1e-4$ ,  $1e-4$ , and  $5e-5$  was employed to train the baseline model (without transfer learning), one-stage transfer learning model, and two-stage transfer learning model, respectively. The batch size and  $\gamma$  were set to 6 and 0.02, respectively, which were selected by grid search. Specifically, a finite number of values were tried, and the one with the best performance was selected. In all experiments, five-fold cross-validation was used, and the average performance was reported.

## 2.4 Statistical analysis

To compare the data distribution between the ICIP and non-ICIP datasets, Fisher's exact test was used for categorical variables, and Mann-Whitney U test was used for continuous variables. Two-tailed tests are used to determine significance at the 5% level. All statistical analyses were conducted using Statistical Product and Service Solutions (IBM SPSS, version 20.0).

To evaluate the classification performance, several typical metrics were used, including accuracy, sensitivity, specificity, precision, and F1-score. We considered ICIP as the positive class and non-ICIP as the negative class, so true positive (TP), false positive (FP), true negative (TN), and false negative (FN) can be accordingly defined. After getting the numbers of TP, FP, TN, and FN, the abovementioned performance metrics can be calculated using Eqs. 4–8. Since we used five-fold cross-validation, the average of these metrics were reported. We also used the area under the receiver operating characteristic (ROC) curve to evaluate model performance. Since every patient was tested for and only for once in five-fold cross-validation, we gathered the results across all the five folds, then plotted ROC curves, and calculated AUCs.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$F1 - score = \frac{2Precision * Specificity}{Precision + Specificity} \quad (8)$$

## 3 Results

### 3.1 Patient characteristics

Among the 48 patients, there were 42 men and 6 women with an overall mean age of 58.00 years  $\pm$  9.75 (standard deviation). We collected nine clinical characteristics for the 48 patients. Table 1 shows these characteristics separately for the ICIP and non-ICIP datasets. Among the nine characteristics, there were significant differences between the two datasets for sex (Fishers' exact test  $p$  value = 0.022) and radiotherapy before immunotherapy (Fishers' exact test  $p$  value = 0.023), whereas no significant differences were observed for the remaining characteristics.

### 3.2 Performance of deep learning model to predict Immune checkpoint inhibitors-related pneumonitis

To explore and validate the effectiveness of the key components used in our method, we conducted extensive experiments. The comparison of quantitative performance is presented in Table 2. The details of different methods are provided as follows:

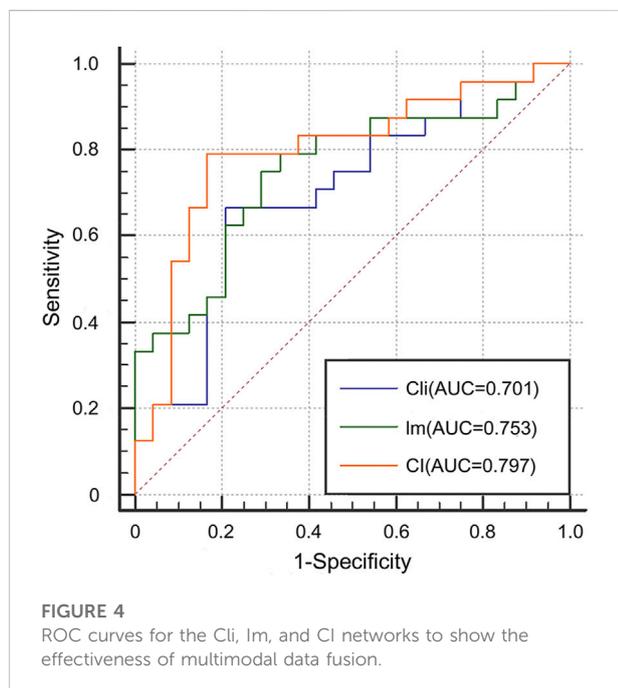
- Cli denotes the clinical network
- Im denotes the image network without using transfer learning and contrastive learning
- CI denotes the multimodal network built on both clinical data and CT images
- Im-1T denotes the image network with one-stage transfer learning
- Im-2T denotes the image network with two-stage transfer learning
- Im-2T-C denotes the image network with two-stage transfer learning and contrastive learning
- CI-2T denotes the multimodal network with two-stage transfer learning
- CI-2T-C denotes the multimodal network with two-stage transfer learning and contrastive learning

#### 3.2.1 Effectiveness of multimodal data fusion

We first evaluated the effectiveness of combining images and clinical data to predict ICIP. To this end, we compared the classification performance of the multimodal network with that of the image network and the clinical network (CI vs. Cli and Im, Table 2). As shown in Table 2, the image network achieved an

TABLE 2 Quantitative analysis of key components in our method. The best results are highlighted in bold.

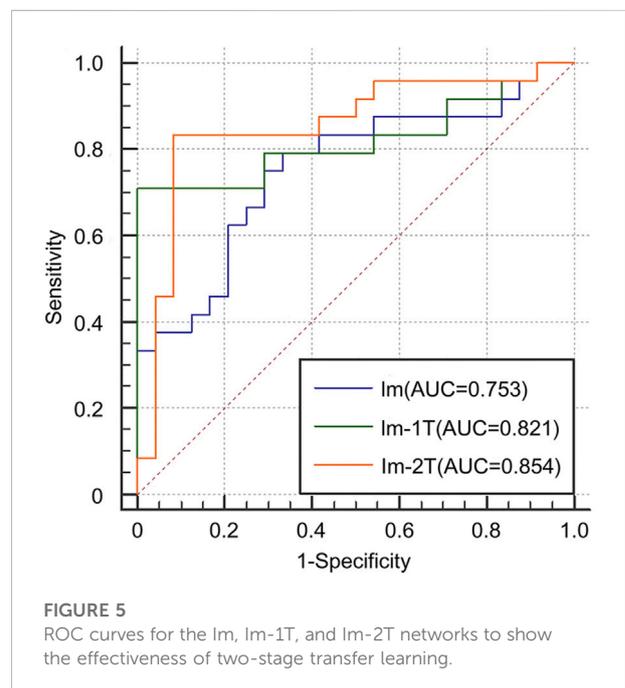
Method	AUC	Accuracy	Sensitivity	Specificity	Precision	F1-score
Cli	0.701	0.730 ± 0.045	0.660 ± 0.134	0.800 ± 0.200	0.817 ± 0.171	0.723 ± 0.041
Im	0.753	0.725 ± 0.075	0.740 ± 0.195	0.710 ± 0.175	0.728 ± 0.079	0.717 ± 0.081
CI	0.797	0.815 ± 0.078	0.790 ± 0.143	0.840 ± 0.089	0.837 ± 0.096	0.814 ± 0.079
Im-1T	0.821	0.830 ± 0.120	0.700 ± 0.200	0.960 ± 0.089	0.837 ± 0.096	0.824 ± 0.125
Im-2T	0.854	0.855 ± 0.087	0.800 ± 0.200	0.910 ± 0.125	0.920 ± 0.110	0.851 ± 0.090
Im-2T-C	0.901	0.920 ± 0.084	0.960 ± 0.089	0.880 ± 0.179	0.910 ± 0.131	0.918 ± 0.087
CI-2T	0.865	0.880 ± 0.130	0.920 ± 0.110	0.840 ± 0.167	0.860 ± 0.142	0.879 ± 0.131
CI-2T-C	0.918	0.920 ± 0.084	0.920 ± 0.110	0.920 ± 0.179	0.943 ± 0.128	0.918 ± 0.087



AUC of 0.753 (Im, Table 2), which was superior to the clinical network that yielded an AUC of 0.701 (Cli, Table 2). By utilizing both images and clinical data, the classification performance was significantly improved up to 0.797 (CI, Table 2). Other metrics in Table 2 also indicates that multimodal data fusion is beneficial to ICIP prediction. ROC curves are shown in Figure 4.

### 3.2.2 Effectiveness of two-stage transfer learning

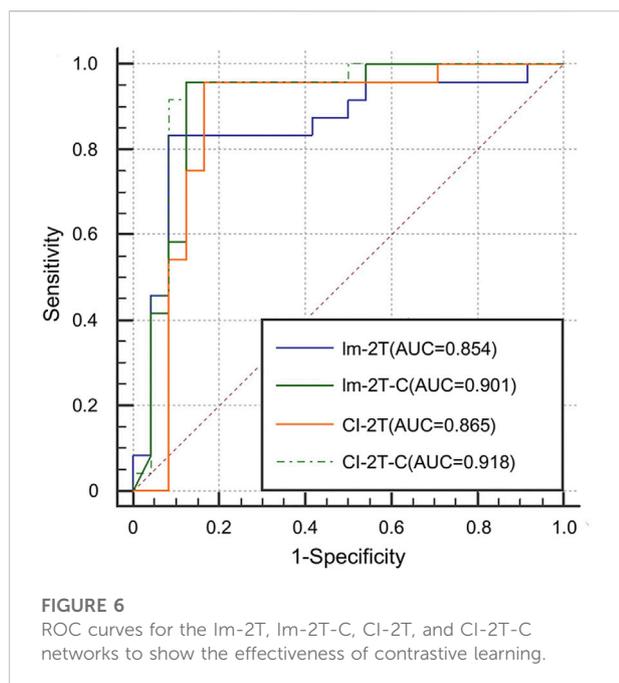
We next evaluated the effectiveness of the proposed two-stage transfer learning strategy. To this end, we first used the image network without transfer learning as the baseline and then gradually incorporated one-stage and two-stage transfer learning. Compared with the baseline model trained from scratch, one-stage transfer learning brought large performance



gain from 0.753 to 0.821 in term of AUC (Im vs. Im-1T, Table 2). Moreover, the use of the two-stage transfer learning further lifted the prediction performance. The AUC, accuracy, sensitivity, and specificity were 0.854, 0.855, 0.800, and 0.910, respectively (Im-2T, Table 2). The increasingly better performance from Im, Im-1T to Im-2T suggests that using a pre-trained network and fine-tuning on a large related dataset are essential to obtain good performance. ROC curves are shown in Figure 5.

### 3.2.3 Effectiveness of contrastive learning

Finally, we evaluated the effectiveness of the contrastive learning strategy in the image network and the multimodal network. The image network with two-stage transfer learning but without contrastive learning achieved AUC of 0.854, accuracy of 0.855, sensitivity of 0.800, and specificity of 0.910 (Im-2T, Table 2). Adding contrastive learning gave a boost in



performance. The resulting AUC, accuracy, sensitivity, and specificity were 0.901, 0.920, 0.960, and 0.880, respectively (Im-2T-C, [Table 2](#)). Similarly, performance gain was also observed when incorporating contrastive learning into the multimodal network. The multimodal network with two-stage transfer learning yielded AUC of 0.865, accuracy of 0.880, sensitivity of 0.920, and specificity of 0.840 (CI-2T, [Table 2](#)). The use of contrastive learning increased the performance by a large margin. The resulting AUC, accuracy, sensitivity, and specificity were 0.918, 0.920, 0.920, and 0.920, respectively (CI-2T-C, [Table 2](#)). ROC curves are shown in [Figure 6](#).

### 3.3 Visualization

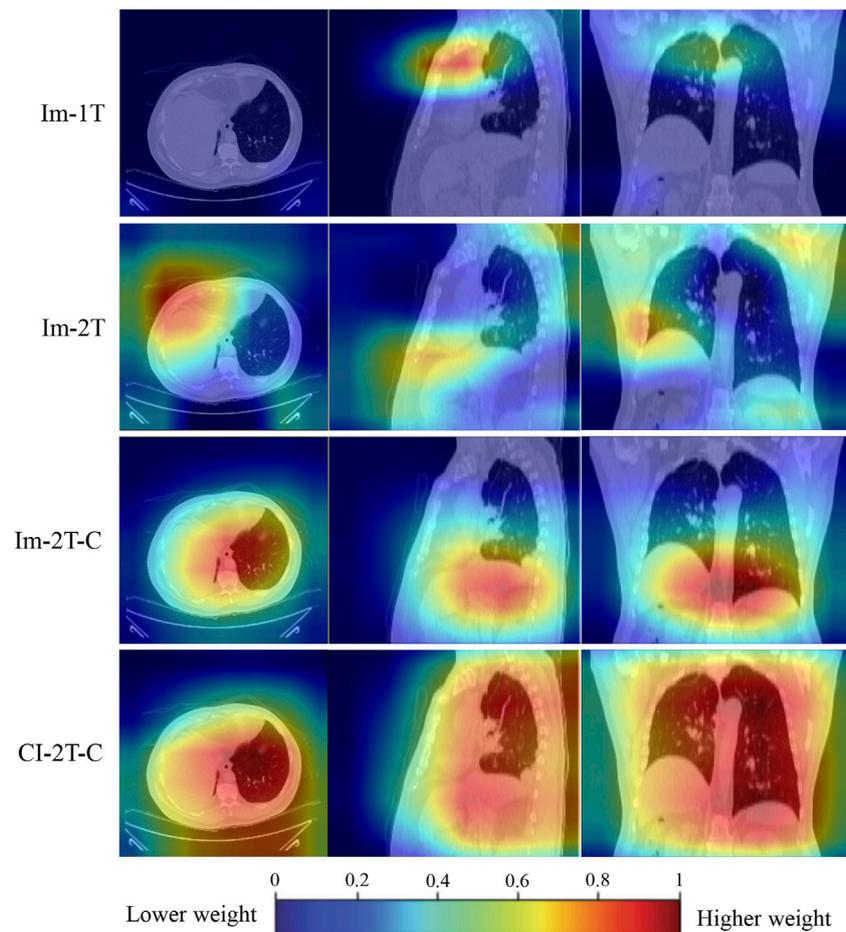
To gain an understanding of why the performance was improved when the key components such as two-stage transfer learning were introduced to the model, the class activation maps were generated by the gradient of the deep learning to highlight the important regions within the input image ([Gotkowski et al., 2021](#)). As shown in [Figure 7](#), the primary attention of the one-stage transfer learning model (Im-1T, first row) was not focused on the lung area. However, the use of two-stage transfer learning and especially contrastive learning brought more attention to the lung area (Im-2T and Im-2T-C, second and third rows). Interestingly, fusing clinical and image features made the network concentrate on the whole lung (CI-2T-C, fourth row), indicating that the whole lung is crucial and informative for ICIP prediction. This makes sense as ICIP can occur anywhere in lung.

## 4 Discussion

As ICI treatment is becoming more frequently used in lung cancer patients, an increasing number of irAEs (i.e., ICIP) are being reported. ICIP is potentially fatal. Thus, early prediction of ICIP is crucial for improving treatment outcomes. However, based on clinical factors or pretreatment CT images, it is very challenging for doctors to predict whether ICIP will occur prior to immunotherapy. Therefore, there is a critical need for an accurate and automated approach to assist doctors in identifying patients at risk for ICIP before immunotherapy, which allows personalized treatment options and reduces the number of deaths due to severe ICIP. In this study, we developed the first deep learning model for predicting ICIP using clinical information and pretreatment baseline chest CT images. In addition to the use of multimodal data, we also introduced two-stage transfer learning and contrastive learning in our model development. We evaluated our method using five-fold cross-validation on 24 ICIP patients and 24 non-ICIP patients. The results demonstrated that the deep learning model accurately differentiated between ICIP and non-ICIP patients, with an AUC of 0.918.

Few prior studies have demonstrated the utility of radiomics to predict irAEs. [Colen et al. \(2018\)](#) presented the first reported work exploring the potential of CT-based radiomics to predict patients at risk for developing ICIP and reported an AUC of 1 ([Colen et al., 2018](#)). Although the performance was extremely high, this study only included 2 ICIP cases and suffered from severe class imbalance problem. By contrast, our study used a balanced dataset consisting of 24 ICIP cases and 24 non-ICIP cases. [Mu et al.](#) proposed a PET/CT based radiomics approach to predict severe irAEs in patients with NSCLC ([Mu et al., 2020](#)). A total of 30 cases with severe irAEs and 164 control cases were curated in the patient cohorts. The radiomics approach yielded an AUC of 0.88 in the prospective validation cohort. However, this work is based on PET/CT which is not widely available in hospitals and thus has limited utility. In contrast to the traditional radiomics methods that extracted a fixed set of image features, our study proposed a deep learning model that can directly learn discriminative features from CT images and demonstrated a better performance with an AUC of 0.918.

The superiority of our method can be attributed to the use of multimodal data fusion, two-stage transfer learning, and contrastive learning in our deep learning model. The effectiveness of these key components was validated by extensive ablation studies. The multimodal data fusion model outperformed the models built on either clinical data or CT images by a large margin. This suggests that the two kinds of data harbor complementary information. Thus, the ICIP prediction task can greatly benefit from this fusion approach. Training deep learning models requires a large dataset. However, in medical applications, oftentimes, only a small dataset is available due to low incidence of disease or expensive cost of data collection. Our



**FIGURE 7**

Activation maps for a non-ICIP sample to show the most “important” regions that different methods consider. The red color represents a higher weight (i.e., more attention is paid to this region).

results confirm that transfer learning is helpful in this case. Simply using a pre-trained model learned on a large unrelated dataset (one-stage transfer learning) or subsequently retraining the model on a large related dataset (two-stage transfer learning) can improve the performance of ICIP prediction markedly. Moreover, contrastive learning can further enhance the feature representation ability by contrasting similar (positive) and dissimilar (negative) samples.

This study has several limitations. First, although our method was rigorously validated by five-fold cross-validation, the data used in this study was collected from a single institution, future efforts will concentrate on validating the findings in a larger multi-institutional cohort. Second, to maintain a balance of sample size between classes, we randomly selected a portion of patients without ICIP to match the sample size of the ICIP dataset. There might be an issue with this strategy as it does not reflect a real-world class distribution. Third, due to the retrospective nature of this study, it may be prone to biases from missing data and reliance on available medical

documentation for review. Prospective studies are needed in the future.

In conclusion, patients who will develop ICIP have subtle changes at their pretreatment baseline CT scans that could not be identified by the naked eye but could be detected by quantitative analysis. Our study presents the first deep learning model based on clinical data and CT images to predict patients at risk for developing ICIP. This model can accurately predict ICIP patients with an AUC of 0.918, which enables efficient risk stratification, close monitoring, and prompt management of ICIP. This will potentially improve ICI treatment outcomes in patients with lung cancer.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

## Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Guangdong Provincial People's Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## Author contributions

PT and JC contributed to conception and design of the study. PT, LW, WH, and GD collected the data. LW, YY, and SQ analyzed and interpreted the data. LW and PT wrote the first draft of the manuscript. DN, SD, and JC supervised the project and revised the manuscript. All authors contributed to the manuscript and approved the submitted version.

## Funding

This study was supported by National Natural Science Foundation of China (61901275, 62171290, 81972970, and 82172671), Guangzhou Science and Technology Plan Foundation

## References

- Akinleye, A., and Rasool, Z. (2019). Immune checkpoint inhibitors of PD-L1 as cancer therapeutics. *J. Hematol. Oncol.* 12, 92. doi:10.1186/s13045-019-0779-5
- Altaf, F., Islam, S. M. S., Janjua, N. K., and Akhtar, N. (2021). "Boosting deep transfer learning for COVID-19 classification," in Proceedings - International Conference on Image Processing, ICIP, Anchorage, AK, USA, September 19–22, 2021. (IEEE Computer Society), 210–214. doi:10.1109/ICIP42928.2021.9506646
- Atchley, W. T., Alvarez, C., Saxena-Beem, S., Schwartz, T. A., Ishizawa, R. C., Patel, K. P., et al. (2021). 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* 160, 731–742. doi:10.1016/j.chest.2021.02.032
- Cheng, J., Pan, Y., Huang, W., Huang, K., Cui, Y., Hong, W., et al. (2022a). Differentiation between immune checkpoint inhibitor-related and radiation pneumonitis in lung cancer by CT radiomics and machine learning. *Med. Phys.* 49, 1547–1558. doi:10.1002/mp.15451
- Cheng, J., Wang, H., Li, R., Li, X., Zhou, X., Yang, X., et al. (2022b). A two-stage multiresolution neural network for automatic diagnosis of hepatic echinococcosis from ultrasound images: a multicenter study. *Med. Phys.* 49, 3199–3212. doi:10.1002/mp.15548
- Chetan, M. R., and Gleeson, F. V. (2021). Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. *Eur. Radiol.* 31, 1049–1058. doi:10.1007/s00330-020-07141-9
- Colen, R. R., Fujii, T., Bilan, M. A., Kotrotsou, A., Abrol, S., Hess, K. R., et al. (2018). Radiomics to predict immunotherapy-induced pneumonitis: proof of concept. *Invest. New Drugs* 36, 601–607. doi:10.1007/s10637-017-0524-2
- Delaunay, M., Cadranet, J., Lusque, A., Meyer, N., Gounaut, V., Moro-Sibilot, D., et al. (2017). Immune-checkpoint inhibitors associated with interstitial lung disease in cancer patients. *Eur. Respir. J.* 50, 1700050. doi:10.1183/13993003.00050-2017
- Gitto, S., Cuocolo, R., Annovazzi, A., Anelli, V., Acquasanta, M., Cincotta, A., et al. (2021). CT radiomics-based machine learning classification of atypical cartilaginous tumours and appendicular chondrosarcomas. *EBioMedicine* 68, 103407. doi:10.1016/j.ebiom.2021.103407
- Gotkowski, K., Gonzalez, C., Bucher, A., and Mukhopadhyay, A. (2021). "M3d-CAM: a PyTorch library to generate 3D attention maps for medical deep learning,"

(2021-02-01-04-1002-0017), Shenzhen University Startup Fund (2019131), Shenzhen-Hong Kong Joint Research Program (SGDX20201103095613036), Natural Science Foundation of Guangdong Province of China (2020A1515010186), GDPH Supporting Fund (KY0120220119), and GDPH Supporting Fund for NSFC Program (8190120260). The funding sources have no involvement in the study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

in *Informatik aktuell* (Regensburg: Springer Science and Business Media Deutschland GmbH), 217–222. doi:10.1007/978-3-658-33198-6\_52

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 770–778. doi:10.1109/CVPR.2016.90

Howell, M., Lee, R., Bowyer, S., Fusi, A., and Lorigan, P. (2015). Optimal management of immune-related toxicities associated with checkpoint inhibitors in lung cancer. *Lung Cancer* 88, 117–123. doi:10.1016/j.lungcan.2015.02.007

Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., et al. (2017). The kinetics human action video dataset. *arXiv Comput. Vis. Pattern Recognit.*

Khunger, M., Rakshit, S., Pasupuleti, V., Hernandez, A. V., Mazzone, P., Stevenson, J., et al. (2017). Incidence of pneumonitis with use of programmed death 1 and programmed death-ligand 1 inhibitors in non-small cell lung cancer: a systematic review and meta-analysis of trials. *Chest* 152, 271–281. doi:10.1016/j.chest.2017.04.177

Martins, F., Sofiya, L., Sykiotis, G. P., Lamine, F., Maillard, M., Fraga, M., et al. (2019). Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. *Nat. Rev. Clin. Oncol.* 16, 563–580. doi:10.1038/s41571-019-0218-0

Monfort, M., Andonian, A., Zhou, B., Ramakrishnan, K., Bargal, S. A., Yan, T., et al. (2020). Moments in time dataset: one million videos for event understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 502–508. doi:10.1109/TPAMI.2019.2901464

Mu, W., Tunali, I., Qi, J., Schabath, M. B., and Gillies, R. J. (2020). Radiomics of 18 F fluorodeoxyglucose PET/CT images predicts severe immune-related adverse events in patients with NSCLC. *Radiol. Artif. Intell.* 2, e190063. doi:10.1148/ryai.2019190063

Naidoo, J., Wang, X., Woo, K. M., Iyriboz, T., Halpenny, D., Cunningham, J., et al. (2017). Pneumonitis in patients treated with anti-programmed death-1/programmed death ligand 1 therapy. *J. Clin. Oncol.* 35, 709–717. doi:10.1200/JCO.2016.68.2005

Nishino, M., Sholl, L. M., Hatabu, H., Ramaiya, N. H., and Hodi, F. S. (2015). Anti-PD-1-Related pneumonitis during cancer immunotherapy. *N. Engl. J. Med.* 373, 288–290. doi:10.1056/nejmc1505197

- Pillai, R. N., Behera, M., Owonikoko, T. K., Kamphorst, A. O., Pakkala, S., Belani, C. P., et al. (2018). Comparison of the toxicity profile of PD-1 versus PD-L1 inhibitors in non-small cell lung cancer: a systematic analysis of the literature. *Cancer* 124, 271–277. doi:10.1002/cncr.31043
- Qian, J., Li, R., Yang, X., Huang, Y., Luo, M., Lin, Z., et al. (2022). Hasa: hybrid architecture search with aggregation strategy for echinococcosis classification and ovary segmentation in ultrasound images. *Expert Syst. Appl.* 202, 117242. doi:10.1016/j.eswa.2022.117242
- Ramos-Casals, M., Brahmer, J. R., Callahan, M. K., Flores-Chávez, A., Keegan, N., Khamashta, M. A., et al. (2020). Immune-related adverse events of checkpoint inhibitors. *Nat. Rev. Dis. Prim.* 6, 38. doi:10.1038/s41572-020-0160-6
- Reck, M., Rodríguez-Abreu, D., Robinson, A. G., Hui, R., Csósz, T., Fülöp, A., et al. (2016). Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N. Engl. J. Med.* 375, 1823–1833. doi:10.1056/nejmoa1606774
- Rios Velazquez, E., Parmar, C., Liu, Y., Coroller, T. P., Cruz, G., Stringfield, O., et al. (2017). Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res.* 77, 3922–3930. doi:10.1158/0008-5472.CAN-17-0122
- Robert, C. (2020). A decade of immune-checkpoint inhibitors in cancer therapy. *Nat. Commun.* 11, 3801. doi:10.1038/s41467-020-17670-y
- Sears, C. R., Rivera, M. P., Camus, P., Gaga, M., Garon, E. B., Gould, M. K., et al. (2019). Knowledge gaps and research priorities in immune checkpoint inhibitor-related pneumonitis: an official American thoracic society research statement. *Am. J. Respir. Crit. Care Med.* 200, E31–E43. doi:10.1164/rccm.201906-1202ST
- Thawani, R., McLane, M., Beig, N., Ghose, S., Prasanna, P., Velcheti, V., et al. (2018). Radiomics and radiogenomics in lung cancer: a review for the clinician. *Lung Cancer* 115, 34–41. doi:10.1016/j.lungcan.2017.10.015
- Vaddepally, R. K., Kharel, P., Pandey, R., Garje, R., and Chandra, A. B. (2020). Review of indications of FDA-approved immune checkpoint inhibitors per NCCN guidelines with the level of evidence. *Cancers* 12, 738. doi:10.3390/cancers12030738
- van Timmeren, J. E., Leijenaar, R. T. H., van Elmpt, W., Reymen, B., Oberije, C., Monshouwer, R., et al. (2017). Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. *Radiother. Oncol.* 123, 363–369. doi:10.1016/j.radonc.2017.04.016
- Winer, A., Nicholas Bodor, J., and Borghaei, H. (2018). Identifying and managing the adverse effects of immune checkpoint blockade. *J. Thorac. Dis.* 10, S480–S489. doi:10.21037/jtd.2018.01.111
- Wu, H., Wu, C., Zheng, H., Wang, L., Guan, W., Duan, S., et al. (2021). Radiogenomics of neuroblastoma in pediatric patients: CT-based radiomics signature in predicting MYCN amplification. *Eur. Radiol.* 31, 3080–3089. doi:10.1007/s00330-020-07246-1
- Xin Yu, J., Hubbard-Lucey, V. M., and Tang, J. (2019). Immuno-oncology drug development goes global. *Nat. Rev. Drug Discov.* 18, 899–900. doi:10.1038/d41573-019-00167-9
- Xu, Q. Q., Shan, W. L., Zhu, Y., Huang, C. C., Bao, S. Y., Guo, L. L., et al. (2021). Prediction efficacy of feature classification of solitary pulmonary nodules based on CT radiomics. *Eur. J. Radiol.* 139, 109667. doi:10.1016/j.ejrad.2021.109667
- Yang, L., Yang, J., Zhou, X., Huang, L., Zhao, W., Wang, T., et al. (2019). Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. *Eur. Radiol.* 29, 2196–2206. doi:10.1007/s00330-018-5770-y
- Yu, J., Deng, Y., Liu, T., Zhou, J., Jia, X., Xiao, T., et al. (2020). Lymph node metastasis prediction of papillary thyroid carcinoma based on transfer learning radiomics. *Nat. Commun.* 11, 4807. doi:10.1038/s41467-020-18497-3
- Zanfardino, M., Franzese, M., Pane, K., Cavaliere, C., Monti, S., Esposito, G., et al. (2019). Bringing radiomics into a multi-omics framework for a comprehensive genotype-phenotype characterization of oncological diseases. *J. Transl. Med.* 17, 337. doi:10.1186/s12967-019-2073-2
- Zhang, K., Liu, X., Shen, J., Li, Z., Sang, Y., Wu, X., et al. (2020). Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* 181, 1360. doi:10.1016/j.cell.2020.08.029
- Zheng, X., Yao, Z., Huang, Y., Yu, Y., Wang, Y. Y., Liu, Y., et al. (2020). Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat. Commun.* 11, 1236. doi:10.1038/s41467-020-15027-z