Radiotherapy and Oncology 195 (2024) 110266

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

# Validated machine learning tools to distinguish immune checkpoint inhibitor, radiotherapy, COVID-19 and other infective pneumonitis

Sumeet Hindocha [a,b,*], Benjamin Hunter [a], Kristofer Linton-Reid [b], Thomas George Charlton [c], Mitchell Chen [d], Andrew Logan [d], Merina Ahmed [e], Imogen Locke [e], Bhupinder Sharma [f], Simon Doran [g], Matthew Orton [h], Catey Bunce [g], Danielle Power [i], Shahreen Ahmad [c], Karen Chan [c], Peng Ng [c], Richard Toshner [j], Binnaz Yasar [k], John Conibear [k], Ravindhi Murphy [l], Tom Newsom-Davis [l], Patrick Goodley [m,n], Matthew Evison [m], Nadia Yousaf [o], George Bitar [f], Fiona McDonald [o], Matthew Blackledge [p], Eric Aboagye [b], Richard Lee [a]

[a] Early Diagnosis and Detection Centre, The Royal Marsden NHS Foundation Trust, Fulham Road, London SW36JJ, UK
[b] Cancer Imaging Centre, Department of Surgery & Cancer, Imperial College London, Du Cane Road, London W12 0NN, UK
[c] Guy's Cancer Centre, Guy's and St Thomas' NHS Foundation Trust, Great Maze Pond, London, SE19RT, UK
[d] Department of Surgery and Cancer, Imperial College London, Du Cane Road, London W12 0NN, UK
[e] Lung Unit, The Royal Marsden NHS Foundation Trust, Downs Road, Sutton SM25PT, UK
[f] Department of Radiology, The Royal Marsden NHS Foundation Trust, Fulham Road, London SW36JJ, UK
[g] Institute of Cancer Research NIHR Biomedical Research Centre, London, UK
[h] Artificial Intelligence Imaging Hub, Royal Marsden NHS Foundation Trust, Downs Road, Sutton SM25PT, UK
[i] Department of Clinical Oncology, Imperial College Healthcare NHS Trust, Fulham Palace Road, London W6 8RF, UK
[j] Interstitial lung disease unit, St Bartholomews' Hospital, Barts Health NHS Trust, West Smithfield, London EC1A 7BE, UK
[k] Department of Clinical Oncology, St Batholomew's Hospital, Barts Health NHS Trust, West Smithfield, London, EC1A 7BE, UK
[l] Chelsea and Westminster Hospital, Chelsea and Westminster NHS Foundation Trust, 369 Fulham Road, London SW10 9NH, UK
[m] Lung Cancer & Thoracic Surgery Directorate, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Greater Manchester, UK
[n] Division of Immunology, Immunity to Infection & Respiratory Medicine, University of Manchester, Manchester, UK
[o] Lung Unit, The Royal Marsden NHS Foundation Trust, Fulham Road, London SW36JJ, UK
[p] Radiotherapy and Imaging, Institute of Cancer Research, 123 Old Brompton Road, London SW7 3RP, UK

## ABSTRACT

*Background:* Pneumonitis is a well-described, potentially disabling, or fatal adverse effect associated with both immune checkpoint inhibitors (ICI) and thoracic radiotherapy. Accurate differentiation between checkpoint inhibitor pneumonitis (CIP) radiation pneumonitis (RP), and infective pneumonitis (IP) is crucial for swift, appropriate, and tailored management to achieve optimal patient outcomes. However, correct diagnosis is often challenging, owing to overlapping clinical presentations and radiological patterns.

*Methods:* In this multi-centre study of 455 patients, we used machine learning with radiomic features extracted from chest CT imaging to develop and validate five models to distinguish CIP and RP from COVID-19, non-COVID-19 infective pneumonitis, and each other. Model performance was compared to that of two radiologists.

*Results:* Models to distinguish RP from COVID-19, CIP from COVID-19 and CIP from non-COVID-19 IP out-performed radiologists (test set AUCs of 0.92 vs 0.8 and 0.8; 0.68 vs 0.43 and 0.4; 0.71 vs 0.55 and 0.63 respectively). Models to distinguish RP from non-COVID-19 IP and CIP from RP were not superior to radiologists but demonstrated modest performance, with test set AUCs of 0.81 and 0.8 respectively. The CIP vs RP model performed less well on patients with prior exposure to both ICI and radiotherapy (AUC 0.54), though the radiologists also had difficulty distinguishing this test cohort (AUC values 0.6 and 0.6).

*Conclusion:* Our results demonstrate the potential utility of such tools as a second or concurrent reader to support oncologists, radiologists, and chest physicians in cases of diagnostic uncertainty. Further research is required for patients with exposure to both ICI and thoracic radiotherapy.

* Corresponding author.
*E-mail address:* sumeet.hindocha@rmh.nhs.uk (S. Hindocha).

https://doi.org/10.1016/j.radonc.2024.110266
Received 8 September 2023; Received in revised form 27 March 2024; Accepted 31 March 2024
Available online 4 April 2024

## Introduction

Pneumonitis is a well-described, potentially disabling or fatal adverse effect associated with both checkpoint inhibitor immunotherapy and thoracic radiotherapy [1]. Immune checkpoint inhibitors (ICI), which stimulate an immune response against tumour cells by blocking inhibitory receptors such as Programmed Death-1 (PD-1), its ligand (PD-L1), or Cytotoxic T-lymphocytic-associated protein-4 (CTLA-4), have revolutionised cancer treatment [2–4]. However, artificial augmentation of the immune system can result in several side effects including pneumonitis, which has a reported average incidence of 3–6 %, but up to 10 % when an anti PD-1/PD-L1 antibody is combined with anti CTLA-4 antibody, and as high as 19 % in some reported studies [1,5–9].

Radiotherapy is integral to the management of many cancers. Approximately 40 % of curative treatments involve radiotherapy, either alone or in combination with surgery or systemic therapy, and it is also highly effective for palliation in recurrent or advanced disease. Furthermore, a growing body of evidence supports the use of high-dose radiotherapy for oligometastatic disease [10]. Thoracic radiotherapy can result in radiation-induced lung disease encompassing radiation pneumonitis (RP) and lung fibrosis, influenced by the total lung radiation dose. Incidence of symptomatic RP is between 5 and 58 % and risk is increased with concurrent chemotherapy or after administration of systemic anticancer therapy in a previously irradiated lung [1,8,9,11]. Subsequent analysis of KEYNOTE-001 demonstrated a 23 % higher all-grade pneumonitis in patients treated with Pembrolizumab (a PD-1 inhibitor) who had received prior radiotherapy compared to those who had not ($p = 0.052$) [12]. Incidence of all-grade pneumonitis was higher in patients treated with adjuvant durvalumab (a PD-L1 inhibitor) after concurrent chemoradiotherapy compared to those receiving placebo (33.9 % vs 24.8 %) [13].

Significant morbidity and mortality can result from both checkpoint inhibitor pneumonitis (CIP) and RP. Both conditions require prompt management, usually with steroids and may necessitate hospital admission for oxygen therapy, and even intubation and ventilation. Pneumonitis may return on weaning steroids or with reintroduction of radiotherapy or ICI [1,11,14–16].

Severe cases require interruption or cessation of ICI or radiation therapy, which may compromise efficacy, resulting in loss of tumour control and worse outcomes. Misdiagnosis of RP or CIP can prohibit subsequent re-challenge of therapy, which may also significantly alter outcomes. Delayed diagnosis and appropriate management can lead to reduced functional status and quality of life. Early and accurate diagnosis is therefore crucial. However, this is often complicated by a wide window of onset, varied clinical presentation and a broad spectrum of radiologic patterns, which overlap with infective pneumonitis (pneumonia). This challenge has been made more apparent since COVID-19. With RP, imaging features may be limited to the radiation field and not conform to lobar anatomy, however this is not always the case, possibly due to a progressive immunological response [8,9,17,18]. [8,111,19].

Bronchoscopy with bronchoalveolar lavage or biopsy may be recommended to rule out alternative causes such as infection or tumour progression, however widespread uptake is limited in practice due to additional risk, patient preference, resource constraints and delay to diagnosis. Even when performed, cytology and tissue samples may be inadequate for diagnosis [16,17,20]. For these reasons, in practice, patients are often treated empirically with antibiotics to cover for infective pneumonitis (IP) [1,16]. However, treatment without diagnostic confirmation is not optimal and could promote antimicrobial resistance or disruption of protective microbiota [21]. A more accurate approach to non-invasively and rapidly distinguish CIP, RP and IP is therefore of significant clinical value. This would facilitate prompt, tailored management and, may allow earlier reintroduction of ICI or radiation therapy where an alternative diagnosis is confirmed.

Radiomics, the automated extraction and machine learning analysis of high-dimensional quantitative features from medical images, may present a solution. These features can be correlated with clinical outcomes to develop decision support tools that have demonstrated prognostic ability in numerous studies [22–31]. Employing machine learning techniques on radiomic features may identify key patterns predictive of CIP, RP, and IP.

In this study, we developed and validated five machine learning tools, comparing handcrafted radiomic features, to distinguish:

1. CIP vs non-COVID-19 infective pneumonitis
2. CIP vs COVID-19
3. RP vs non-COVID-19 infective pneumonitis
4. RP vs COVID-19
5. CIP vs RP

## Methods

This work was reviewed by the Royal Marsden Hospital Committee for Clinical Research and approved by the UK Health Research Authority (reference number: 283611), ClinicalTrials.gov identifier: NCT04721444. Some data for this study were obtained from the National COVID-19 Chest Imaging Database (NCCID) [32] (Research Ethics Committee reference number: 20/LO/0688). NCCID is a centralized database containing medical images of hospital patients from over 25 centres across the UK. Data used in the study were deidentified and therefore patient consent was not required, as per Health Research Authority and Research Ethics Council approvals.

Clinical data and accompanying CT images demonstrating pneumonitis were collected from six UK NHS Trusts (BLINDED) and the NCCID. Comprehensive inclusion and exclusion criteria were implemented to curate five distinct cohorts of patients with either:

- A confirmed diagnosis of pneumonitis attributed to ICI alone (CIP cohort): 103 patients
- A confirmed diagnosis of pneumonitis attributed to radiotherapy alone (RP cohort): 111 patients
- A confirmed diagnosis of non-COVID-19 infective pneumonia alone (IP cohort): 106 patients
- A confirmed diagnosis of COVID-19 alone (COVID-19 cohort): 120 patients
- Pneumonitis in the context of prior ICI and radiotherapy exposure (mixed ICI + RT test cohort): 15 patients

For example, for the CIP cohort, patients must have received an ICI in the three months prior to data collection and presented with new radiological lung changes confirmed on the CT, of a severity and distribution consistent with CIP, that are not incompatible with lower respiratory tract infection. Patients must not have had radiotherapy involving the thorax in the 12 months prior to presentation, and where there was documented clinical concern for infection, they must have undergone investigations to rule this out. Similarly, for the RP cohort, patients must have completed a course of radiotherapy involving the thorax (e.g. lung, breast, oesophageal radiotherapy) in the 12 months prior to presentation, and have not received an ICI. They must have presented with new radiological lung changes on CT, of a severity and distribution consistent with RP or early fibrosis (not established fibrosis). These changes should be of severity and distribution that are not incompatible with lower respiratory tract infection. For the infective pneumonitis cohort, patients presented prior to 2020 and thus did not have COVID-19. Ground truth was determined by a multidisciplinary team including clinical oncologists, medical oncologists, respiratory physicians, and radiologists. Full details of the inclusion and exclusion criteria for each cohort are listed in the Supplementary Material.

Example images demonstrating the difficulty of disease differentiation, for each cohort are presented in Fig. 1. In A1, A2, B1 and B2, there

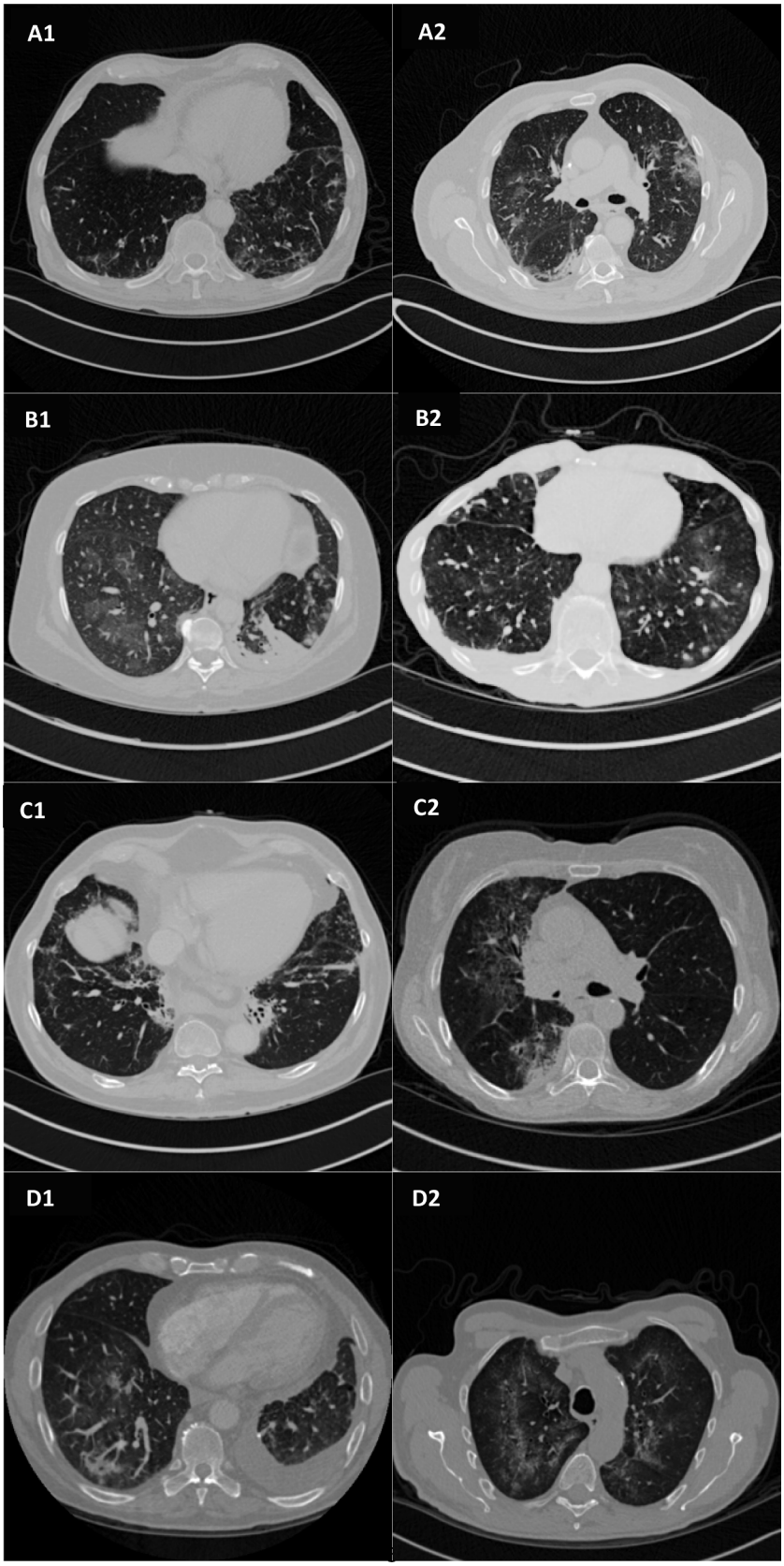


**Fig. 1.** Two example axial CT slices from each cohort, demonstrating the difficulty of disease differentiation based on imaging alone, with non-specific radiological features and pattern present in all cases. A1 & A2) COVID-19, B1 & B2) CIP, C1 & C2) RP, D1 & D2) non-COVID-19 IP.

are bilateral multi-focal ground glass opacifications. There are additionally unilateral consolidations +/- atelectases in the dependent part of the lungs in A2 and B1, and unilateral small pleural effusions in B1 and B2. C1 demonstrates paramediastinal consolidations. In C2, there is unilateral ground glass and posterior consolidation in the right lung. In these cases, these features conform to the irradiated field but as described earlier, this might not always be the case, particularly in cases treated with stereotactic ablative radiotherapy [33]. In D1 and D2, there are bilateral ground glass opacifications. In D1, there are additionally multifocal consolidations visible in the right lower lobe and a unilateral pleural effusion on the left.

Following pseudonymisation, CT images were resampled to 1x1x1mm using trilinear interpolation, a conservative approach recommended by the Image Biomarker Standardisation Initiative (IBSI) [34], and then converted from DICOM to NIfTI format using MRIcroGL (version 1.2). Images were reviewed with CT reports to identify regions of pneumonitis. These were then contoured over three consecutive slices by a clinical oncologist (SH) to form regions of interest (ROIs) using ITK-Snap (version 3.8.0) [35], for feature extraction. Where pneumonitis was diffuse, the whole lobe was contoured.

Several software programs are available for radiomic feature extraction, either open-source, commercial or those developed in-house [36]. In this study we used Pyradiomics 3.0.1, an open-source, IBSI-compliant python package for the extraction of radiomics data, without wavelet filtering [34,37]. A total of 107 handcrafted features were extracted from each ROI using 25 Hounsfield unit bins using Pyradiomics. All statistical analysis was undertaken in R Studio (version 1.4.1106).

As ROI masks did not pertain to a specific shape (e.g. that of a specific tumour or nodule), size and shape features were removed, leaving 93 Pyradiomic features for each patient. Features were standardised by centering on the mean and dividing by the standard deviation (Z-score normalisation) [38]. Twenty patients from each of the CIP, RP, IP, and COVID-19 cohorts, were allocated to hold-out test sets at random, with the remaining number of patients in each cohort assigned to training sets. For each of the models, respective training cohorts and test cohorts were combined. For example, for the CIP vs COVID-19 prediction model, the CIP and COVID-19 training sets were combined and the CIP and COVID-19 test sets were combined. Feature reduction steps were employed to reduce data dimensionality. Highly correlated features were removed, and univariable logistic regression with False Discovery Rate correction was used to select highly significant features for model development. Parameters for these steps are listed in Supplementary Table 1.

Finally, LASSO regression with 10-fold cross-validation was used to select the lambda value of the minimum cross-validated error (lambda.min) before model fitting. The weighted sum of features with non-zero regression coefficients gave the radiomic predictive vector (RPV) for pneumonitis type. Features contributing to the RPV for each model along with their weights are listed in Supplementary Table 2. The optimal Receiver-Operator Characteristic (ROC) curve cut-off to maximise the training set accuracy was used to predict pneumonitis type in the training and test sets. For the CIP vs RP model, this was further validated on a second test comprising the 15 patients presenting with pneumonitis in the context of both prior ICI and radiotherapy treatment (mixed test cohort). The Area Under the Curve (AUC) with confidence intervals was calculated for each model. Classification metrics including Balanced Accuracy, F1 score, sensitivity, specificity, positive and negative predictive value were also recorded and are detailed in Supplementary Table S3.

Test sets for each model were independently reviewed by two board-certified clinical radiologists (MC, AL) with 7 and 4 years of experience respectively, who were asked to assign a binary label to each scan in a particular test set. For example, for the CIP vs COVID-19 test set, the option was either CIP or COVID-19. The radiologists were blinded to clinical data and the number of a particular class in the test set but were able to review the full CT scans in multi-planar reformat. AUC values and classification metrics were calculated for each radiologist's performance (classification metrics are detailed in Supplementary Table S4).

**Table 1**
Training and test set sizes for each of the models.

| Model | Training Set Size | Test Set Size |
|---|---|---|
| CIP vs non-COVID-19 IP | 169 (83 CIP, 86 non-COVID-19 IP) | 40 (20 each) |
| CIP vs COVID-19 | 183(83 CIP, 100 COVID-19) | 40 (20 each) |
| RP vs non-COVID-19 IP | 177(91 RP, 86 non-COVID-19 IP) | 40 (20 each) |
| RP vs COVID-19 | 191(91 RP, 100 COVID-19) | 40 (20 each) |
| CIP vs RP | 174(83 CIP, 91 RP) | 40 (20 each) |

**Table 2**
Training and test set AUC values and Confidence Intervals for each of the models together with test set AUC values for the radiologists (R1 = radiologist 1, R2 = radiologist 2).

| Model | Set | AUC | CI | R1 AUC | R2 AUC |
|---|---|---|---|---|---|
| CIP vs non-COVID-19 IP | Train | 0.79 | 0.72—0.87 | | |
| | Test | 0.71 | 0.53—0.86 | 0.55 | 0.63 |
| CIP vs COVID-19 | Train | 0.70 | 0.62—0.76 | | |
| | Test | 0.68 | 0.49—0.84 | 0.43 | 0.4 |
| RP vs non-COVID-19 IP | Train | 0.85 | 0.79—0.91 | | |
| | Test | 0.81 | 0.64—0.94 | 0.93 | 0.88 |
| RP vs COVID-19 | Train | 0.85 | 0.79—0.90 | | |
| | Test | 0.92 | 0.81—0.99 | 0.80 | 0.80 |
| CIP vs RP | Train | 0.77 | 0.70—0.84 | | |
| | Test | 0.80 | 0.64—0.92 | 0.90 | 0.85 |
| | Mixed ICI + RT Test | 0.54 | 0.22—0.86 | 0.60 | 0.60 |

## Results

A total of 455 patients were included in the study, across the five cohorts (Fig. 2). Numbers of patients included in each model training and test set are described in Table 1.

AUC values with 95 % confidence intervals for each model and test set AUC values for each radiologist are listed in Table 2. The models to distinguish RP from COVID-19, CIP from COVID-19 and CIP from non-COVID-19 IP out-performed the radiologists (test set AUCs of 0.92 vs 0.8 and 0.8; 0.68 vs 0.43 and 0.4; 0.71 vs 0.55 and 0.63 respectively). The models to distinguish RP vs non-COVID-19 IP and CIP vs RP demonstrated modest performance, with test set AUCs of 0.81 and 0.8 respectively, though these were not superior to the radiologists. the CIP vs RP model and both radiologists did not agree with the MDT ground truth labels for the mixed ICI + RT test cohort, though the radiologists performed slightly better with AUC values of 0.6 and 0.6 vs the model's 0.54. ROC curves for each of the models are shown in Fig. 3.

For the CIP vs RP model, four texture features were selected: original_glrlm_GrayLevelNonUniformity, original_glrlm_GrayLevelVariance, original_glszm_SmallAreaEmphasis and original_ngtdm_Coarseness. A Gray Level Run Length Matrix (GLRLM) quantifies gray level runs - the number of consecutive pixels with the same gray level value. Gray Level Non-uniformity measures variability of gray level intensity values in the image, with lower values indicating more homogeneity in intensity, and Variance measures the variance in gray level intensity for the runs. A Gray Level Size Zone (GLSZM) measures gray level zones in an image. Zones are the number of connected voxels that share the same gray level intensity. Small Area Emphasis measures the distribution of small size zones, with larger values indicating more smaller zones and more fine textures. A Neighbouring Gray Tone Difference Matrix (NGTDM) quantifies the difference between a gray value and the average gray value of its neighbours within a particular distance. Coarseness measures the average difference between the centre voxel and its neighbourhood, indicating spatial rate of change. Higher values indicate

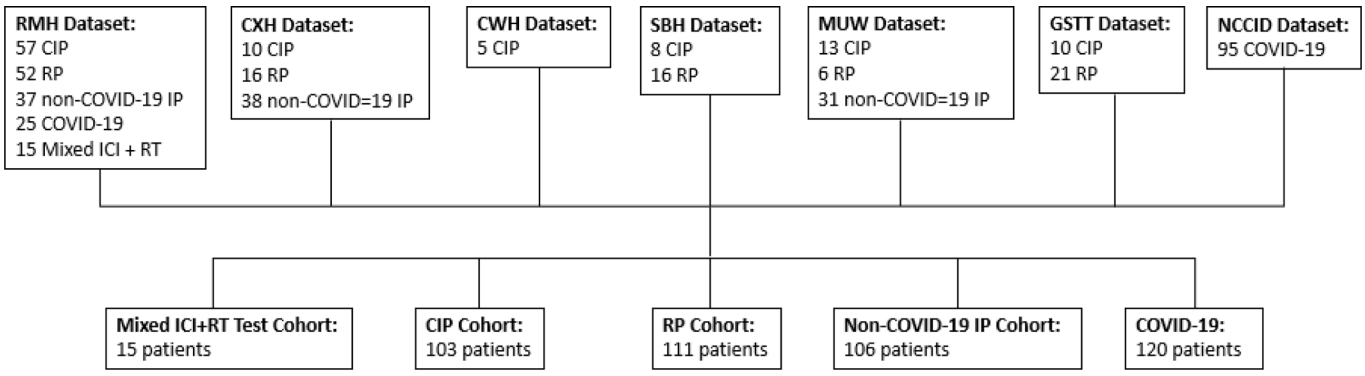


**Fig. 2.** Numbers of patients from each of the six NHS Trusts and the NCCID allocated to each cohort. RMH = Royal Marsden Hospital NHS Foundation Trust, CXH = Imperial College Healthcare NHS Trust - Charing Cross Hospital, CWH = Chelsea & Westminster Hospital NHS Foundation Trust, SBH = Barts Health NHS Trust, MUW = Manchester University Hospital NHS Foundation Trust - Wythenshawe Hospital, GSTT = Guy's & St Thomas' NHS Foundation Trust.

lower spatial change rates and a locally more uniform texture. Features for each of the models, with their respective weights, are listed in Supplementary Table 2.

## Discussion

The ability to accurately distinguish causes of pneumonitis is crucial to instigate appropriate, tailored management and avoid unnecessary intervention or interruption of ICI or radiotherapy, thus increasing likelihood of optimal outcomes. However, accurate diagnosis is often difficult due to overlapping and varied clinical presentation and radiologic patterns, as well as a wide window of onset. Imaging-based artificial intelligence tools that can non-invasively detect subtle differences between different pneumonitis aetiologies to support multi-disciplinary clinical decision-making are therefore of great value. In this multicentre study of 455 patients, we have developed and tested radiomic models to distinguish CIP and RP from COVID-19, non-COVID-19 IP, and from each other. These models were then benchmarked against two board-certified clinical radiologists.

Since the emergence of COVID-19, numerous studies have described radiomic and deep-learning models that differentiate COVID-19 from other IP [39–46], however very few have focused on distinguishing CIP from RP, and we were not able to find any that focused specifically on distinguishing either CIP or RP from COVID-19 or non-COVID-19 IP, as we have done.

Tohidinezhad et al. developed a model to distinguish CIP from other types of pneumonitis, including both infection and radiotherapy, achieving an AUC of 0.83. However, the cohort was limited to 72 patients receiving anti-PD-1/PD-L1 antibodies for Stage IV non-small cell lung cancer (NSCLC) and was not externally tested [20]. Chen et al. compared RP vs CIP in patients with NSCLC with a training set of 55 patients (29 RP and 23 CIP). Their model tested on the CIP vs RP alone training set reached an AUC 0.76 (CI 0.63–0.90), vs our training and test set values of 0.77 and 0.80 [47]. They also tested their model on 30 patients that had been treated with both radiotherapy and ICI therapy. In a differing approach to our methodology, and as no gold-standard diagnostic criteria exists to definitively distinguish pneumonitis cause in this cohort, they compared model results to a multi-disciplinary consensus attribution, however a protocol or assessment scale for reaching consensus was not described. Cheng et al. developed a "Bag of Words" model, also in patients with NSCLC, with a training set of 59 (28 CIP and 31 RP) achieving an AUC of 0.937 with 10-fold cross validation. The model was validated on a limited cohort of 14 patients treated with both radiotherapy and ICI therapy achieving an AUC of 0.896[9]. Qiu et al. developed a nomogram to distinguish CIP and RP in patients with NSCLC using 11 radiomic and two imaging features – bilateral lung changes and presence of sharp borders, achieving a validation set AUC of 0.947[48].

Our models were developed using radiomic features extracted from three CT slices whilst the clinical radiologists were able to review the entire CT scan. Despite this, the models to distinguish CIP from infection and RP vs COVID-19 outperformed the clinical radiologists and the RP vs non-COVID-19 IP and CIP vs RP models demonstrated modest performance (AUC values 0.81 and 0.8 respectively). It is possible that model performance could be improved with the ability to leverage the feature space of the whole CT, and/or through the use of deep learning.

Our CIP vs RP model failed to differentiate on the test set treated with both ICI and radiotherapy however the radiologists also had difficulty correctly differentiating this test set. This may be due to the difficulty in establishing ground truth labels for this cohort as well as the existence of crossed immune signalling pathways in both radiation and ICI induced lung injury, reflecting shared pathology and imaging features [49]. Additionally, there is a possibility of pneumonitis in this cohort being due to ICI induced radiation recall pneumonitis which is suggested to have distinct radiological patterns and immunological mechanisms to that of direct radiation pneumonitis [50,51]. Development of and scoring against a robust multidisciplinary consensus of ground truth labels would be worth exploring.

Chen et al. found that qualitative imaging features such as number of involved lobes, distribution of pneumonitis change and presence of a sharp border improved model performance[47]. Similarly, integration of clinical data with radiomic features improved performance of classification models in our previous work[24], however clinical and qualitative imaging features were not available for inclusion.

Whilst our study has larger training and test sets compared to those described above, and is not limited to only patients with NSCLC, our sample sizes may be considered relatively small in the context of machine learning and may be responsible for wider confidence intervals for the test set AUCs. This is a consequence of our robust and comprehensive inclusion and exclusion criteria, designed to ensure cohorts were not cross contaminated – i.e. a patient in the CIP cohort did not have concurrent infection. However, future work would benefit from larger training and test sets. In particular, the number of patients with pneumonitis in the context of prior exposure to ICI and radiotherapy is limited in our study. Future research should focus on this cohort as increasing numbers of patients are likely to receive both thoracic radiotherapy and ICI, for example with consolidation Durvalumab following concurrent chemoradiotherapy for Stage III NSCLC. Additional limitations of this work include the retrospective nature of the study, as well as the dichotomous nature of the models; a multi-class model would be preferable for maximal clinical utility.

Despite this, to our knowledge, this is the first study to present models for both CIP and RP vs COVID-19 and other infective pneumonitis. Our models are developed using CT imaging from a CIP cohort of patients treated with both anti PD-1/PD-L1 and CTLA-4 antibodies and our CIP and RP cohorts consist of patients with several tumour types

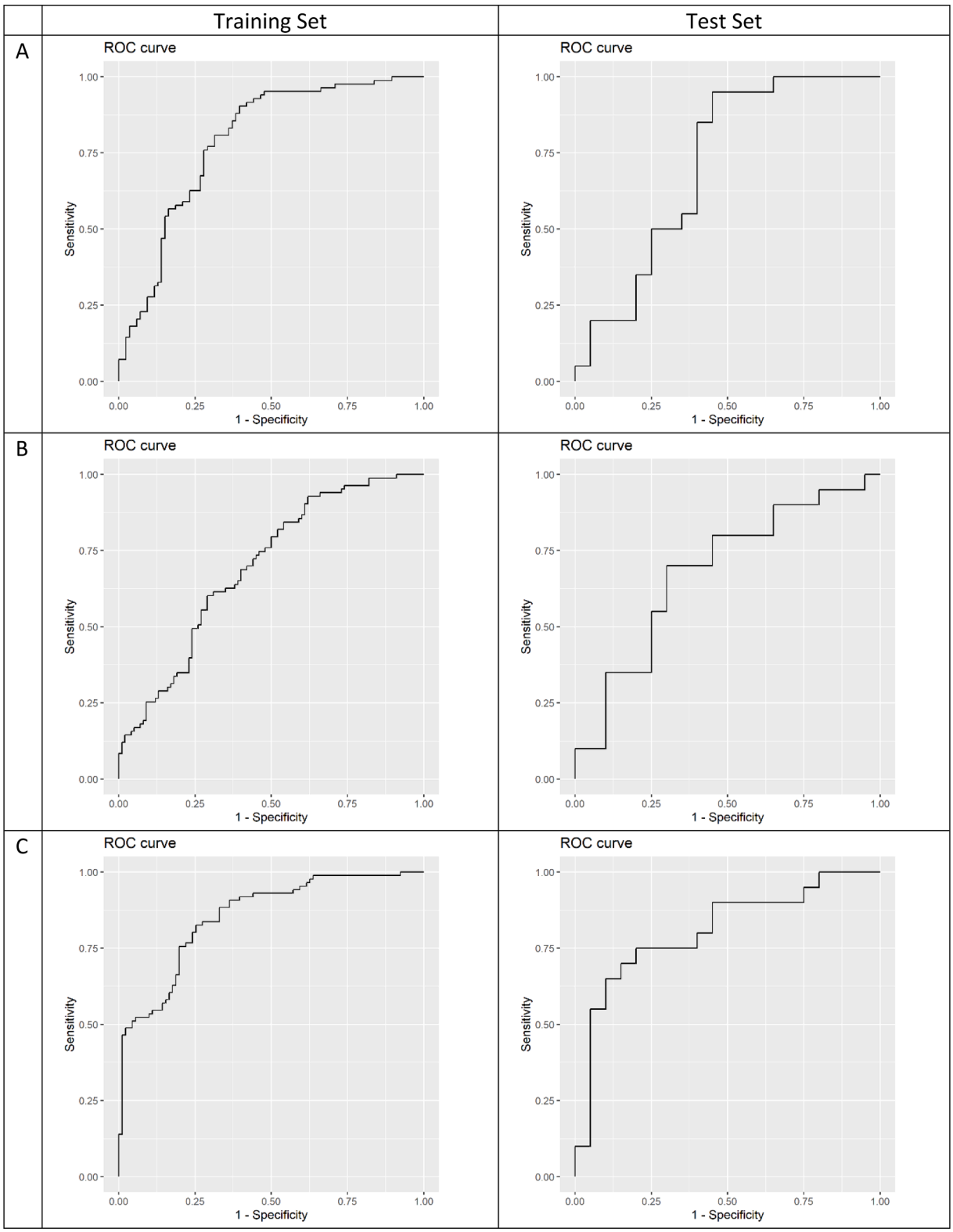


**Fig. 3.** ROC curves demonstrating the trade-off between sensitivity (true positive rate) and 1-Specificity (false positive rate) at varying classification thresholds, for each model. Left panel: training sets; right panel: test sets. A) CIP vs non-COVID-19 IP model B) CIP vs COVID-19 model. C) RP vs non-COVID-19 IP model. D) RP vs COVID-19 model. E) CIP vs RP model.

beyond NSCLC, thus extending their clinical application. We have also employed larger, multi-centre cohorts for model development, compared to previous single-centre studies focusing on CIP vs RP discussed above. Furthermore, our models are externally tested and benchmarked against clinical radiologists, with encouraging AUC values that reflect potential ability to act as a second reader, supporting oncologists, radiologists, and chest physicians in determining pneumonitis aetiology in cases with diagnostic uncertainty. Future work could

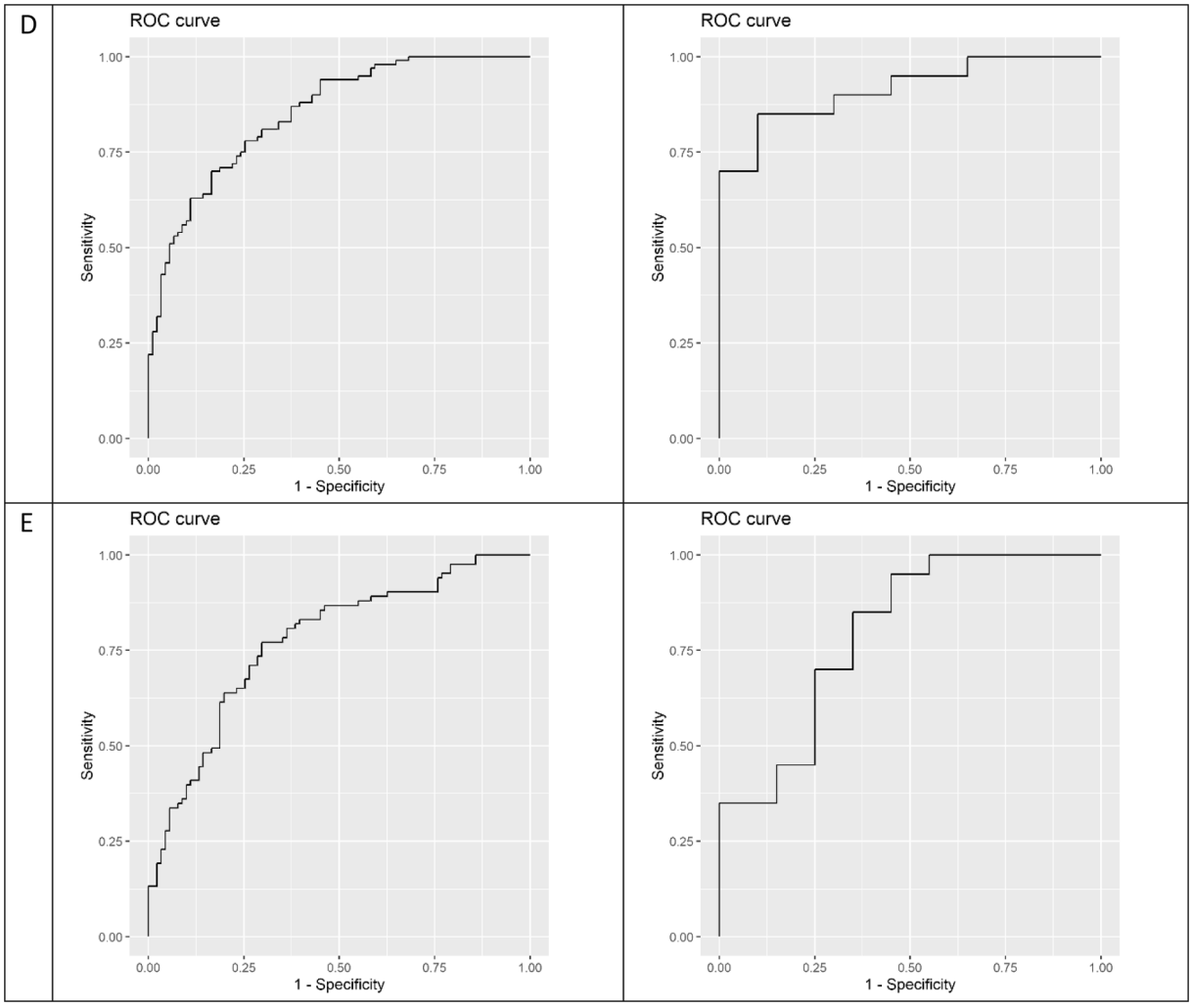


**Fig. 3.** (*continued*).

explore deep learning as an alternative to handcrafted radiomics, multi-class model architectures, and incorporation of other biomarkers to aid model performance.

## CRediT authorship contribution statement

**Sumeet Hindocha:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Benjamin Hunter:** Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Kristofer Linton-Reid:** Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Thomas George Charlton:** Data curation. **Mitchell Chen:** Writing – review & editing, Formal analysis. **Andrew Logan:** Writing – review & editing, Formal analysis. **Merina Ahmed:** Writing – review & editing, Conceptualization. **Imogen Locke:** Writing – review & editing. **Bhupinder Sharma:** Writing – review & editing, Conceptualization. **Simon Doran:** Writing – review & editing. **Matthew Orton:** Writing – review & editing, Data curation. **Catey Bunce:** Writing – review & editing. **Danielle Power:** Writing – review & editing, Data curation. **Shahreen Ahmad:** Writing – review & editing, Data curation. **Karen Chan:** Writing – review & editing, Data curation. **Peng Ng:** Writing – review & editing, Data curation. **Richard Toshner:** Writing – review & editing, Data curation. **Binnaz Yasar:** Writing – review & editing, Data curation. **John Conibear:** Writing – review & editing, Data curation. **Ravindhi Murphy:** Writing – review & editing, Data curation. **Tom Newsom-Davis:** Writing – review & editing, Data curation. **Patrick Goodley:** Writing – review & editing, Data curation. **Matthew Evison:** Writing – review & editing, Data curation. **Nadia Yousaf:** Writing – review & editing, Conceptualization. **George Bitar:** Writing – review & editing, Formal analysis. **Fiona McDonald:** Writing – review & editing. **Matthew Blackledge:** Writing – review & editing, Data curation. **Eric Aboagye:** Supervision, Methodology, Conceptualization. **Richard Lee:** Supervision, Project administration, Methodology, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [RL is funded by the Royal Marsden and ICR NIHR BRC, Royal Marsden Cancer charity and SBRI (collaborating with QURE.AI). RL's institution receives compensation for time spent in a secondment role for the NHS England in lung cancer screening and National Institute of Health and Care Research. He has received research funding from CRUK, Innovate UK (co funded by GE Healthcare, Roche and Optellum), SBRI, RM Partners Cancer Alliance and NIHR (coapplicant with Optellum). He has received honoraria from CRUK and undertakes personal private practice. Dr Merina Ahmed declares research funding from BMS and MSD, honoraria from Astra Zeneca and is on the advisory board for BMS and the data monitoring committee for Astra Zeneca. Dr Fiona McDonald is a speaker for and declares advisory board fees from Astra Zeneca. Matthew Blackledge receives funding from Cancer Research UK, paid to

his institution. Simon J Doran receives a salary from Cancer Research UK as the manager of the National Cancer Imaging Translational Accelerator (NCITA) Repository Unit. The Repository Unit provides the repository resource used to store the data for this project, together with the curation expertise and charges the PI of the project a fee for its use. Mitchell Chen is funded by the NIHR (Grant No. CL-2021–21-005) and Academy of Medical Sciences (Grant No. SGL026\1024). Sumeet Hindocha is funded by the UKRI CDT in AI for Healthcare https://ai4health.io (Grant No. P/S023283/1), by Imperial College London and by the Royal Marsden & Institute of Cancer Research NIHR Biomedical Research Centre. Benjamin Hunter is funded by Cancer Research UK (grant reference C309/A31316) and Royal Marsden Partners].

## Acknowledgements

This study represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, London, and by the Royal Marsden Cancer Charity. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Sumeet Hindocha is funded by the UKRI CDT in AI for Healthcare http://ai4health.io (Grant No. P/S023283/1), by Imperial College London and by the Royal Marsden & Institute of Cancer Research NIHR Biomedical Research Centre.

Benjamin Hunter is funded by Cancer Research UK (grant reference C309/A31316) and Royal Marsden Partners.

Mitchell Chen is funded by the NIHR (Grant No. CL-2021-21-005) and Academy of Medical Sciences (Grant No. SGL026\1024).

RL is funded by the Royal Marsden and ICR NIHR BRC, Royal Marsden Cancer charity and SBRI (collaborating with QURE.AI). RL's institution receives compensation for time spent in a secondment role for the NHS England in lung cancer screening and National Institute of Health and Care Research. He has received research funding from CRUK, Innovate UK (cofunded by GE Healthcare, Roche and Optellum), SBRI, RM Partners Cancer Alliance and NIHR (coapplicant with Optellum). He has received honoraria from CRUK and undertakes personal private practice.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2024.110266.

## References

[1] Hindocha S, et al. Immune checkpoint inhibitor and radiotherapy-related pneumonitis: an informatics approach to determine real-world incidence, severity. Management, and Resource Implications Front Med (Lausanne) 2021;8:2039.
[2] Bersanelli M, Buti S. From targeting the tumor to targeting the immune system: transversal challenges in oncology with the inhibition of the PD-1/PD-L1 axis. World Journal of Clinical Oncology 2017;8:37–53.
[3] Kruger S, et al. Advances in cancer immunotherapy 2019 - latest trends. J Exp Clin Cancer Res 2019;38:268.
[4] Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. Nat Rev Immunol 2020 20:11 2020,;20:651–68.
[5] Delaunay M, et al. Management of pulmonary toxicity associated with immune checkpoint inhibitors. Eur Respir Rev 2019;28.
[6] Suresh K, et al. Pneumonitis in non-small cell lung cancer patients receiving immune checkpoint immunotherapy: incidence and risk factors. J Thorac Oncol 2018;13:1930–9.
[7] Porcu M, et al. Immunotherapy associated pulmonary toxicity: biology behind clinical and radiological features. Cancers 2019;11.
[8] Arroyo-Hernández M, et al. Radiation-induced lung injury: current evidence. BMC Pulm Med 2021;21:1–12.
[9] Cheng J, 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.
[10] Beaton L, Bandula S, Gaze MN, Sharma RA. How rapid advances in imaging are defining the future of precision radiation oncology. Br J Cancer 2019;120:779.
[11] Jain V, Berman AT. Radiation pneumonitis: old problem, new tricks. Cancers 2018;10.
[12] Shaverdian N, 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;18:895.
[13] Antonia SJ, et al. Durvalumab after chemoradiotherapy in stage III non–small-cell lung cancer. N Engl J Med 2017;377:1919–29.
[14] Pozzessere C, et al. Relationship between pneumonitis induced by immune checkpoint inhibitors and the underlying parenchymal status: a retrospective study. ERJ Open Res 2020;6:00165–2019.
[15] Schoenfeld JD, et al. Pneumonitis resulting from radiation and immune checkpoint blockade illustrates characteristic clinical, radiologic and circulating biomarker features. J Immunother Cancer 2019;7:112.
[16] Haanen J, et al. Management of toxicities from immunotherapy: ESMO clinical Practice guideline for diagnosis, treatment and follow-up ☆. Ann Oncol 2022;33:1217–38.
[17] Bledsoe TJ, Nath SK, Decker RH. Radiation pneumonitis. Clin Chest Med 2017;38:201–8.
[18] Huang J, Chen X, Xia B, Ma S. Advances in CT features and radiomics of checkpoint inhibitor-related pneumonitis: a short review. Front Immunol 2023;14:164.
[19] Naidoo J, et al. Pneumonitis in patients treated with anti-programmed death-1/programmed death ligand 1 therapy. J Clin Oncol 2017;35:709–17.
[20] Tohidinezhad F, et al. Computed tomography-based radiomics for the differential diagnosis of pneumonitis in stage IV non-small cell lung cancer patients treated with immune checkpoint inhibitors. Eur J Cancer 2023;183:142–51.
[21] Christaki, E., Marcou, M. & Tofarides, A. Antimicrobial Resistance in Bacteria: Mechanisms, Evolution, and Persistence. *Journal of Molecular Evolution 2019 88:1* **88**, 26–40 (2019).
[22] Colen RR, et al. Radiomics to predict immunotherapy-induced pneumonitis: proof of concept. Invest New Drugs 2018;36:601–7.
[23] Hunter B, Hindocha S, Lee RW. The role of artificial intelligence in Early cancer diagnosis. Cancers (Basel) 2022;14.
[24] Hindocha, S. *et al.* Gross tumour volume radiomics for prognostication of recurrence & death following radical radiotherapy for NSCLC. *npj Precision Oncology 2022 6:1* **6**, 1–11 (2022).
[25] Hosny A, Aerts HJ, Mak RH. Handcrafted versus deep learning radiomics for prediction of cancer therapy response. The Lancet Digital Health 2019;1:e106–7.
[26] Aerts HJWL, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5.
[27] Coroller TP, et al. Radiomic-based pathological response prediction from Primary tumors and lymph nodes in NSCLC. J Thorac Oncol 2017;12:467–76.
[28] Lambin P, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–62.
[29] Gardin I, et al. Radiomics: principles and radiotherapy applications. Crit Rev Oncol Hematol 2019;138:44–50.
[30] Liu Z, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. Theranostics 2019;9:1303–22.
[31] Vial A, et al. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. Transl Cancer Res 2018;7:803–16.
[32] NHSX. National COVID-19 Chest Imaging Database (NCCID). https://www.nhsx.nhs.uk/covid-19-response/data-and-covid-19/national-covid-19-chest-imaging-database-nccid/ (2021).
[33] Aoki, T. *et al.* Evaluation of Lung Injury after Three-dimensional Conformal Stereotactic Radiation Therapy for Solitary Lung Tumors: CT Appearance1. *10.1148/radiol.2301021226* **230**, 101–108 (2004).
[34] Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. Radiology 2016;295:328–38.
[35] ITK-SNAP Home. http://www.itksnap.org/pmwiki/pmwiki.php.
[36] Fornacon-Wood I, et al. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. Eur Radiol 2020;30:6241–50.
[37] Van Griethuysen JJM, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104–7.
[38] Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. Eur J Nucl Med Mol Imaging 2019;46:2638–55.
[39] Wang H, et al. Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures. Eur J Nucl Med Mol Imaging 2021;48:1478–86.
[40] Jia LL, et al. Artificial intelligence model on chest imaging to diagnose COVID-19 and other pneumonias: a systematic review and meta-analysis. Eur J Radiol Open 2022;9.
[41] Alsharif W, Qurashi A. Effectiveness of COVID-19 diagnosis and management tools: a review. Radiography (Lond) 2021;27:682–7.
[42] Bouchareb Y, et al. Artificial intelligence-driven assessment of radiological images for COVID-19. Comput Biol Med 2021;136:104665.
[43] Piccialli, F., di Cola, V. S., Giampaolo, F. & Cuomo, S. The Role of Artificial Intelligence in Fighting the COVID-19 Pandemic. *Information Systems Frontiers 2021 23:6* **23**, 1467–1497 (2021).
[44] Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Singh V. Application of deep learning for fast detection of COVID-19 in X-rays using nCOVnet. Chaos Solitons Fractals 2020;138:109944.
[45] Sedik A, et al. Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections. Viruses 2020;12.
[46] Amyar A, Modzelewski R, Li H, Ruan S. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation. Comput Biol Med 2020;126.
[47] Chen X, et al. Radiation versus immune checkpoint inhibitor associated pneumonitis: distinct radiologic morphologies. Oncologist 2021;26:e1822.

[48] Qiu Q, Xing L, Wang Y, Feng A, Wen Q. 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.
[49] Zhang Z, et al. Crossed pathways for radiation-induced and immunotherapy-related lung injury. Front Immunol 2021;12.
[50] Teng F, Li M, Yu J. Radiation recall pneumonitis induced by PD-1/PD-L1 blockades: mechanisms and therapeutic implications. BMC Med 2020;18.
[51] McGovern K, Ghaly M, Esposito M, Barnaby K, Seetharamu N. Radiation recall pneumonitis in the setting of immunotherapy and radiation: a focused review. Future Sci OA 2019;5.