



An Artificial Intelligence Platform for the Radiologic Diagnosis of Pulmonary Sarcoidosis: An Initial Pilot Study of Chest Computed Tomography Analysis to Distinguish Pulmonary Sarcoidosis from a Negative Lung Cancer Screening Scan

Marc A. Judson¹ · Jianwei Qiu² · Camille L. Dumas³ · Jun Yang⁴ · Brion Sarachan² · Jhimli Mitra²

Received: 14 September 2023 / Accepted: 31 October 2023 / Published online: 14 November 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Purpose To determine the reliability of an artificial intelligence, deep learning (AI/DL)-based method of chest computer tomography (CT) scan analysis to distinguish pulmonary sarcoidosis from negative lung cancer screening chest CT scans (*Lung Imaging Reporting and Data System* score 1, Lung-RADS score 1).

Methods Chest CT scans of pulmonary sarcoidosis were evaluated by a clinician experienced with sarcoidosis and a chest radiologist for clinical and radiologic evidence of sarcoidosis and exclusion of alternative or concomitant pulmonary diseases. The AI/DL based method used an ensemble network architecture combining Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The method was applied to 126 pulmonary sarcoidosis and 96 Lung-RADS score 1 CT scans. The analytic approach of training and validation of the AI/DL method used a *fivefold* cross-validation technique, where 4/5th of the available data set was used to train a diagnostic model and tested on the remaining 1/5th of the data set, and repeated 4 more times with non-overlapping validation/test data. The probability values were used to generate Receiver Operating Characteristic (ROC) curves to assess the model's discriminatory power.

Results The sensitivity, specificity, positive and negative predictive value of the AI/DL method for the 5 folds of the training/validation sets and the entire set of CT scans were all over 94% to distinguish pulmonary sarcoidosis from LUNG-RADS score 1 chest CT scans. The area under the curve for the corresponding ROC curves were all over 97%.

Conclusion This AI/DL model shows promise to distinguish sarcoidosis from alternative pulmonary conditions using minimal radiologic data.

Keywords Sarcoidosis · Interstitial lung disease · Chest CT scan · Artificial intelligence · Machine learning

Introduction

The lung is the most common organ involved with sarcoidosis with a frequency of 90 percent in most series [1, 2]. In addition, pulmonary sarcoidosis is problematic to diagnose,

with an average delay of 3 months between symptom onset and diagnosis and a delay of more than one year in 20 percent of cases [3]. The delay in the diagnosis of pulmonary sarcoidosis has been attributed to the lack of specificity of its presenting symptoms that are often found with several common alternative pulmonary diseases [3, 4]. Significant delays in the diagnosis of pulmonary sarcoidosis may lead to significant disease-related morbidity as well as inappropriate treatment of alternative conditions.

There is currently no gold-standard diagnostic test for sarcoidosis. The diagnosis of sarcoidosis requires a compatible clinical presentation, histologic evidence of non-caseating granulomatous inflammation, and exclusion of other disorders capable of producing similar histology and clinical features [5]. However, the diagnosis of sarcoidosis

✉ Marc A. Judson
judsonm@amc.edu

¹ Division of Pulmonary and Critical Care Medicine, Department of Medicine, Albany Medical College, MC-91; 16 New Scotland, Albany, NY 12208, USA

² GE Healthcare, Troy, NY, USA

³ Cardiovascular Division, Department of Radiology, Albany Medical College, Albany, NY, USA

⁴ Albany Medical College, Albany, NY, USA

is never completely secure, because the diagnostic criteria of “a compatible clinical presentation” and “exclusion of other disorders capable of producing similar histology and clinical features” are subjective clinical decisions that cannot be rigorously defined and are dependent on the subjective opinions of the medical provider [6, 7].

Although it was previously thought that a tissue biopsy was the gold-standard diagnostic test for diffuse lung disease, certain radiologic features have reached the level of diagnostic specificity for certain pulmonary disorders. Idiopathic pulmonary fibrosis (IPF) is the prototypical diffuse lung disease where the diagnosis can be established on the basis of the clinical presentation and lung imaging findings [8]. Several chest computed tomography (CT) features of pulmonary sarcoidosis are thought to be highly specific for the disease [9], although their diagnostic power has not been formally tested.

There is increasing evidence that artificial intelligence (AI) has the potential to provide clinical radiologic assessment at an expert level [10]. In the past several years, AI and machine learning tools have been extensively constructed to reliably diagnose lung diseases [11–15], including interstitial lung diseases [13]. The establishment of objective radiologic criteria for the diagnosis of pulmonary sarcoidosis has the potential to accelerate the diagnostic process as well as avoid invasive biopsy procedures. Herein we present an AI/Deep Learning (DL)-based method designed to diagnose pulmonary sarcoidosis based on chest CT imaging features. We present data from a pilot study using this platform to distinguish CT scans from pulmonary sarcoidosis patients from those of smokers who had negative lung cancer screening chest exams.

Methods

This research was approved by the Albany Medical College Institutional Review Board (study number 6039). The purpose of this study was to determine the sensitivity and specificity of a machine learning AI platform to identify chest CT scan images of pulmonary sarcoidosis patients versus chest CT scan images obtained from patients who underwent lung cancer screening where the scan was interpreted as showing no evidence of lung malignancy (defined below in the data section). This research involved first identifying chest CT scans for analysis and then subjecting them to the AI/DL method.

Data

The chest CT scans for this study were identified and screened consecutively as follows. The chest CT scans of pulmonary sarcoidosis patients ($n = 126$) were obtained

either from an institution-approved clinical database or through the radiology records at Albany Medical Center. An author with extensive experience in pulmonary sarcoidosis (MAJ) carefully reviewed the clinical records of these patients and confirmed their diagnosis using established international criteria [5]. Subsequently, a board-certified radiologist (CD) with expertise in chest CT scan interpretation reviewed the chest CT scans of these patients to confirm that their chest CT scan findings were consistent with pulmonary sarcoidosis. In cases where sarcoidosis patients had multiple chest CT scans, the first scan showing significant disease was selected in order to minimize the possibility of developing a second pulmonary condition. For all sarcoidosis patients whose CT scans were selected for analysis, the clinician excluded them if they had clinical evidence of a concomitant additional lung disease. Similarly, the chest CT radiologist excluded chest CT scans from these sarcoidosis patients with radiologic evidence of a concomitant additional pulmonary disease or where the chest CT findings were inconsistent with sarcoidosis. No CT scan of a pulmonary sarcoidosis patient was excluded because of a specific radiographic form of the disease (e.g., fibrocystic disease, micro-nodularity), because we desired our model to learn to distinguish the chest imaging findings of all pulmonary sarcoidosis cases from those of other pulmonary disorders. The CT scans of the controls ($N = 96$) were obtained from patients cared for at Albany Medical Center who had undergone lung cancer screening. The criteria for patients to undergo chest CT scan screening for lung cancer changed in 2021 [16]. Therefore, these patients ranged from 50 to 80 years old, had at least a 20 pack-year history of cigarette smoking, and were either currently smoking or had quit smoking within 15 years [16, 17]. The CT chest exams from these patients were low-dose lung cancer screening studies which received a Lung-RADS score (Lung Imaging Reporting and Data System score) of “1” or “negative for lung cancer” [18]. These CT scans were either normal exams without evidence of an acute or chronic pulmonary disease or revealed minimal findings of chronic smoking-related changes, such as mild emphysema.

The AI Method to Diagnose Pulmonary Sarcoidosis

We have developed an AI/Deep Learning (DL)-based method, which is an ensemble network architecture combining Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), to classify pulmonary sarcoidosis vs. Lung-RADS score 1 from 3D-chest CT volume. CNNs have an inherent capability to learn discriminative features within convolutional blocks for classification tasks from image patches. However, with more recent advancements in DL, ViTs have become popular in building robust classification models—sometimes outperforming CNNs [19].

Unlike CNNs that capture only local information of the image within the receptive field of the convolutional filters, ViTs capture global dependencies for contextual understanding within an image for a classification task. However, one limitation of ViTs is that they require a large amount of data to train the model [20, 21]. The combination of CNNs and ViTs in an image recognition method dramatically reduces the number of test images required for learning [20, 22, 23]. Furthermore, CNNs are superior to ViTs in capturing local contextual information which was another motivation for combining these two techniques. In addition, our method only requires knowledge of the diagnosis and image data. No further human interaction is required, such as identifying regions of interest.

The overall analytic approach of this study consisted of training and validation of the AI/DL classification method using a K -fold cross-validation technique [24], where $K = 5$, i.e., 4/5th of the available dataset was used to train a diagnostic model and tested on the remaining 1/5th of the dataset and repeated four more times with non-overlapping validation/test data. For each validation fold, similar numbers of sarcoidosis and healthy lungs were maintained. Table 1 shows the data that were used in each fold of the five-fold cross-validation. The AI/DL framework was developed using Python 3.7.16, PyTorch 1.8.1, and using a NVIDIA V-100

Table 1 Training and validation data partitions, across each cross-validation fold for different lung conditions

Cross-Validation Fold #	Training data		Validation data	
	Sarcoidosis	Lung-RADS score of 1	Sarcoidosis	Lung-RADS score of 1
1	100	77	26	19
2	103	74	23	22
3	96	82	30	14
4	103	75	23	21
5	102	76	24	20

Total number of subjects: sarcoidosis—126; Healthy controls—96
LUNG-RADS Lung imaging reporting and data system

Table 2 Sensitivity, specificity, positive predictive value, and negative predictive value as evaluation of the AI model performance across 5 folds of cross-validation

Cross-Validation Fold #	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC
1	0.98	1.00	0.95	0.96	1.00	1.00
2	0.93	0.91	0.95	0.95	0.91	0.98
3	0.95	1.00	0.86	0.94	1.00	0.99
4	0.93	0.87	1.00	1.00	0.88	0.97
5	0.93	0.92	0.95	0.96	0.90	0.98
Mean ± Std. Dev	0.95 ± 0.02	0.94 ± 0.05	0.94 ± 0.05	0.96 ± 0.02	0.94 ± 0.05	0.98 ± 0.01

Graphics Processing Units (GPU), enabled with CUDA 10.1 and CUDNN 8.0.5. The five-fold cross-validation was performed using scikit-learn 1.0.2.

Statistical Analysis

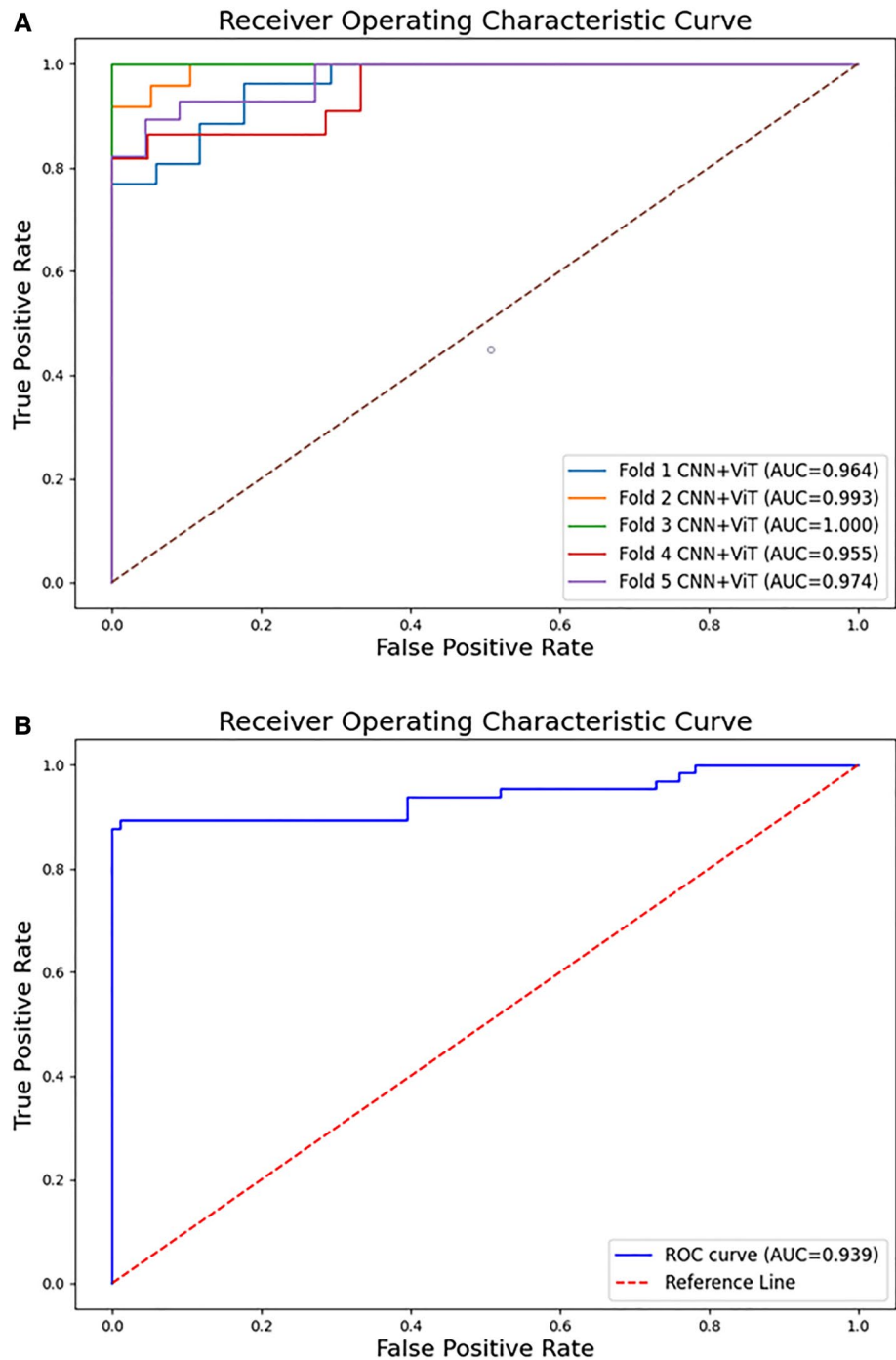
The AI/DL method takes each 3D CT scan as input. It then extracts, manipulates, and reduces these inputs to a set of features in the CNN + ViT framework with a mathematical sigmoid function that provides probabilistic values belonging to the class label of sarcoidosis. A probability value greater than 0.5 was assigned a label of sarcoidosis and probability value of ≤ 0.5 was assigned to the label of Lung-RADS score 1. Therefore, a binary classification decision was made for each input test CT scan. The probability values were also used to generate the Receiver Operating Characteristic (ROC) curves that measure discrimination power of the predictive classification model. The area under the curve (AUC) from the Receiver Operating Characteristic (ROC) curve was computed examining the proportion of true positives versus the proportion of false positives at different probability cutoffs. The performance metrics of the AI/DL method for the diagnosis of sarcoidosis vs Lung-RADS score 1 subjects were computed for each fold using the following equations: Sensitivity = $TP/(TP + FN)$; Specificity = $TN/(TN + FP)$; PPV = $TP/(TP + FP)$; NPV = $TN/(TN + FN)$; and Accuracy = $(TP + TN)/(TP + FP + FN + TN)$.

Results

The validation results for the five-fold cross-validation are presented in Table 2. High values of performance metrics for the diagnosis of sarcoidosis were achieved in all folds of cross-validation. The overall sensitivity, specificity, positive predictive value, negative predictive value, and accuracy for the model to distinguish sarcoidosis from Lung-RADS score 1 were at least 94 percent.

Figure 1A and B shows the ROCs for each fold of validations and the overall validation set, respectively. High AUCs for each of the 5 validation folds and the overall

Fig. 1 **A** Receiver operating characteristic (ROC) curves for all cross-validation folds for classifying pulmonary sarcoidosis from **Lung** Imaging Reporting and Data System (Lung-RADS) score = 1. **B**: Combined receiver operating characteristic (ROC) curve across all validation folds for classifying pulmonary sarcoidosis from **Lung** Imaging Reporting and Data System (Lung-RADS) score = 1



validation set were all at least 97%. We also constructed training/validation loss and accuracy curves for each of the 5 folds (Fig. 1) that demonstrated well-converged training-validation loss curves in each of the folds. This suggests that our model was optimally fitted in each fold. While the loss was tracked for training convergence, the model (epoch) with the best validation accuracy was chosen for prediction on the test data for the fold.

Discussion

We found that our artificial intelligence/deep learning method of chest CT scan analysis accurately distinguished CT scans of pulmonary sarcoidosis patients from those with a Lung-RADS score of 1 on a lung cancer screening examination. The sensitivity, specificity, positive, and

negative predictive value, and accuracy of this method were all over 94%. The area under the ROC curve of over 97% suggests that our method can reliably distinguish these two conditions radiologically.

Artificial intelligence and machine learning have tremendous potential to be useful for chest medical image analysis and interpretation [25, 26]. These techniques have already been shown to be as or more accurate than radiologists in the detection of lung nodules [27], tuberculosis [28], and pneumonia [29]. We suspect that artificial intelligence/machine learning chest imaging platforms will be a particularly valuable assessment and diagnostic tool for interstitial lung diseases because multidisciplinary conferences attended by clinicians, pathologists, and radiologists are now considered the standard of care in the management of these diseases [30]. In particular, these platforms should be very useful for pulmonary sarcoidosis, where the diagnosis is commonly delayed [3] and based on subjective criteria [5]. AI methods could serve as an excellent screening tool prior to a final read by a radiologist.

The AI/DL method that we used is particularly useful in the radiographic diagnosis of ILD for several reasons. First, unlike ViT methods which require a large quantity of data, the combination of CNNs and ViTs vastly decreases the data required for training [20–23]. This is important in the case of ILDs as many of them are relatively rare diseases and a large number of ILD radiographic images are not available for machine learning. Second, this method does not require segmentation of lung parenchymal regions of interest as a preprocessing step. This allows the method to be developed without human direction. This may allow for novel approaches in radiographic diagnosis that may equal or surpass the current clinical approach. We believe that the use of ViTs is a critical component of our method, because many ILDs are distinguished on the basis of the specific location of the radiographic abnormalities relative to anatomic structures within the thorax. ViTs explicitly capture relative positional information along with image features for classification tasks. Finally, many ILDs such as sarcoidosis have no known cause and no standardized diagnostic test and therefore, the diagnosis of these ILDs is based on clinical judgment. It is conceivable that an analytic diagnostic approach to the radiographic features of these ILDs may surpass clinical judgment and lead to the establishment of a diagnostic standard based in chest imaging findings.

Our analysis has some limitations. First, our sarcoidosis and lung cancer screening patients were all from one medical center. Furthermore, this pilot study included a relatively small population with only 126 pulmonary sarcoidosis patients. It is possible that these patients were not representative of a universal sample of individuals with these conditions. Second, it is possible that these patients were misdiagnosed or had additional or alternative pulmonary

diagnoses. However, we believe this was not common, as we rigorously searched for these conditions. Third, in this pilot analysis, we only distinguished pulmonary sarcoidosis from lung cancer screening patients with Lung-RADS score 1 chest CT scans. It is possible that other lung diseases might mimic the radiologic features of sarcoidosis more closely and it may be more problematic to distinguish pulmonary sarcoidosis from such diseases. These limitations suggest that future studies should analyze the diagnostic power of our AI/DL platform in a multicenter trial with multiple non-sarcoidosis diseases as alternative conditions. Fourth, although there was no referral or selection bias in the study, demographic, race, ethnicity, gender, CT machine vendor, and CT reconstruction biases remain. AI models are typically unaware of the biases and can lead to faulty predictions unless the data selection for training the AI represents all the variations possible. All these above factors are human biases that are probably introduced in the AI model, which is then training with unrepresentative data can lead to faulty predictions. One method to mitigate this bias is by engaging a human in the loop i.e., a radiologist to review the model's predictions confirming sarcoidosis. Finally, although we did not observe over-fitting in the K-fold cross-validation method, the lack of validation on multicenter data is a potential limitation of this study.

In summary, we have demonstrated that our AI/DL method can reliably distinguish CT scans of pulmonary sarcoidosis patients from those with a Lung-RADS scores of 1 on a lung cancer screening examination. Our method is capable of being applied to any specific lung disease. We plan to test our method to distinguish pulmonary sarcoidosis from a variety of other pulmonary diseases.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00408-023-00655-1>.

Author Contributions MAJ contributed to study design, data collection, and writing and editing of the manuscript. JQ contributed to study design, data collection and statistical analysis, and writing and editing the manuscript. CLD contributed to study design, data collection, and writing and editing of the manuscript. JY contributed to data collection and writing and editing of the manuscript. BS contributed to study design and writing and editing of the manuscript. JM contributed to study design, data collection, and writing and editing of the manuscript.

Funding There was no funding for this research.

Declarations

Competing Interests MAJ has received grants for his institution from Mallinckrodt, aTyr Pharmaceuticals, and Foundation for Sarcoidosis Research. No other author has any competing interests. JQ is an employee of General Electric HealthCare. BS is an employee of General Electric HealthCare JM is an employee of General Electric HealthCare. JY: No competing interests.

Ethical Approval This study was approved by the Albany Medical Center Institutional Review Board (study number 6039).

References

- Baughman RP, Teirstein AS, Judson MA et al (2001) Clinical characteristics of patients in a case control study of sarcoidosis. *Am J Respir Crit Care Med* 164(10 Pt 1):1885–1889
- Judson MA, Boan AD, Lackland DT (2012) The clinical course of sarcoidosis: presentation, diagnosis, and treatment in a large white and black cohort in the United States. *Sarcoidosis Vasc Diffuse Lung Dis* 29(2):119–127
- Judson MA, Thompson BW, Rabin DL et al (2003) The diagnostic pathway to sarcoidosis. *Chest* 123(2):406–412
- Judson MA (2023) The management of sarcoidosis in the 2020s by the primary care physician. *Am J Med* 136(6):534–544
- Crouser ED, Maier LA, Wilson KC et al (2020) Diagnosis and detection of sarcoidosis. An Official American Thoracic Society Clinical Practice Guideline. *Am J Respir Crit Care Med* 201(8):e26–e51
- Putman M, Patel JJ, Dua A (2022) There is no diagnosis of exclusion in rheumatology. *Rheumatology (Oxford)* 62(1):1–2
- Judson MA (2018) The diagnosis of sarcoidosis: attempting to apply rigor to arbitrary and circular reasoning. *Chest* 154(5):1006–1007
- Raghu G, Remy-Jardin M, Myers JL et al (2018) Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice Guideline. *Am J Respir Crit Care Med* 198(5):e44–e68
- Tana C, Donatiello I, Coppola MG et al (2020) CT findings in pulmonary and abdominal sarcoidosis. Implications for diagnosis and Classification. *J Clin Med* 9(9):3028
- Langlotz CP (2019) Will artificial intelligence replace radiologists? *Radiol Artif Intell* 1(3):e190058
- Frix AN, Cousin F, Refaee T et al (2021) Radiomics in lung diseases imaging: state-of-the-art for clinicians. *J Pers Med* 11(7):602
- De Giacomi F, Raghunath S, Karwoski R, Bartholmai BJ, Moua T (2018) Short-term automated quantification of radiologic changes in the characterization of idiopathic pulmonary fibrosis versus nonspecific interstitial pneumonia and prediction of long-term survival. *J Thorac Imaging* 33(2):124–131
- Furukawa T, Oyama S, Yokota H et al (2022) A comprehensible machine learning tool to differentially diagnose idiopathic pulmonary fibrosis from other chronic interstitial lung diseases. *Respirology* 27(9):739–746
- Baghdadi N, Maklad AS, Malki A, Deif MA (2022) Reliable sarcoidosis detection using chest X-rays with efficient nets and stain-normalization techniques. *Sensors* 22(10):3846
- Thattaamuriyil Padmakumari L, Guido G, Caruso D et al (2022) The role of chest CT radiomics in diagnosis of lung cancer or tuberculosis: a pilot study. *Diagnostics* 12(3):739
- Ritzwoller DP, Meza R, Carroll NM et al (2021) Evaluation of population-level changes associated with the 2021 US preventive services task force lung cancer screening recommendations in community-based health care systems. *JAMA Netw Open* 4(10):e2128176
- Moyer VA (2014) Screening for lung cancer: US preventive services task Force recommendation statement. *Ann Intern Med* 160(5):330–338
- Pinsky PF, Gierada DS, Black W et al (2015) Performance of lung-RADS in the national lung screening trial: a retrospective assessment. *Ann Intern Med* 162(7):485–491
- Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR* 2020; 2020.
- Wu Y, Qi S, Sun Y, Xia S, Yao Y, Qian W (2021) A vision transformer for emphysema classification using CT images. *Phys Med Biol* 66:245016
- Wu H, Xiao B, Codella N, et al. Cvt: Introducing convolutions to vision transformers. Paper presented at: *IEEE/CVF International Conference on Computer Vision* 2021.
- D'Ascoli S, Touvron H, Leavitt ML, Morcos A, S., Biroli G, Sagun L. ConViT: Improving vision transformers with soft convolutional inductive biases. Paper presented at: *International Conference on Machine Learning* 2021.
- Maurício J, Domingues I, Bernardino J (2023) Comparing vision transformers and convolutional neural networks for image classification: a literature review. *Appl Sci* 13:5521
- Hastie T, Tibshirani R, Friedman J (2009) Model assessment and selection. In: Hastie T, Tibshirani R, Friedman J (eds) *The elements of statistical learning: data mining, inference, and prediction*, 2nd edn. Springer-Verlag, Berlin
- Liu F, Tang J, Ma J et al (2021) The application of artificial intelligence to chest medical image analysis. *Intelligent Medicine* 1:104–117
- Chassagnon G, Vakalopoulou M, Paragios N, Revel MP (2020) Artificial intelligence applications for thoracic imaging. *Eur J Radiol* 123:108774
- Nam JG, Park S, Hwang EJ et al (2019) Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology* 290(1):218–228
- Hwang EJ, Park S, Jin KN et al (2019) Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw Open* 2(3):e191095
- Rajpurkar P, Irvin J, Ball RL et al (2018) Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* 15(11):e1002686
- Lee CT (2022) Multidisciplinary meetings in interstitial lung disease: polishing the gold standard. *Ann Am Thorac Soc* 19(1):7–9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.