



Towards the adoption of quantitative computed tomography in the management of interstitial lung disease

Simon L.F. Walsh¹, Jan De Backer², Helmut Prosch³, Georg Langs^{3,4}, Lucio Calandriello⁵, Vincent Cottin⁶, Kevin K. Brown⁷, Yoshikazu Inoue⁸, Vasilios Tzilas⁹ and Elizabeth Estes¹⁰, on behalf of the Open Source Imaging Consortium (OSIC)

¹National Heart and Lung Institute, Imperial College, London, UK. ²FLUIDDA, Antwerp, Belgium. ³Medical University of Vienna, Vienna, Austria. ⁴contextflow GmbH, Vienna, Austria. ⁵Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy. ⁶National Reference Center for Rare Pulmonary Diseases, Louis Pradel Hospital, Hospices Civils de Lyon, Claude Bernard University Lyon 1, UMR 754, Lyon, France. ⁷Department of Medicine, National Jewish Health, Denver, CO, USA. ⁸Clinical Research Center, National Hospital Organization Kinki-Chuo Chest Medical Center, Sakai City, Japan. ⁹5th Respiratory Department, Chest Diseases Hospital Sotiria, Athens, Greece. ¹⁰Open Source Imaging Consortium (OSIC), Saugatuck, MI, USA.

Corresponding author: Simon L.F. Walsh (s.walsh@imperial.ac.uk)



Shareable abstract (@ERSpublications)

Quantitative CT software has huge potential to develop imaging biomarkers to improve the management of fibrosing ILDs, but optimising their adoption will require technical and societal challenges to be addressed. <https://bit.ly/481ZoQA>

Cite this article as: Walsh SLF, De Backer J, Prosch H, *et al.* Towards the adoption of quantitative computed tomography in the management of interstitial lung disease. *Eur Respir Rev* 2024; 33: 230055 [DOI: 10.1183/16000617.0055-2023].

Copyright ©The authors 2024

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

Received: 23 March 2023
Accepted: 31 Jan 2024

Abstract

The shortcomings of qualitative visual assessment have led to the development of computer-based tools to characterise and quantify disease on high-resolution computed tomography (HRCT) in patients with interstitial lung diseases (ILDs). Quantitative CT (QCT) software enables quantification of patterns on HRCT with results that are objective, reproducible, sensitive to change and predictive of disease progression. Applications developed to provide a diagnosis or pattern classification are mainly based on artificial intelligence. Deep learning, which identifies patterns in high-dimensional data and maps them to segmentations or outcomes, can be used to identify the imaging patterns that most accurately predict disease progression. Optimisation of QCT software will require the implementation of protocol standards to generate data of sufficient quality for use in computerised applications and the identification of diagnostic, imaging and physiological features that are robustly associated with mortality for use as anchors in the development of algorithms. Consortia such as the Open Source Imaging Consortium have a key role to play in the collation of imaging and clinical data that can be used to identify digital imaging biomarkers that inform diagnosis, prognosis and response to therapy.

Knowledge past and present

Interstitial lung diseases (ILDs) comprise a vast group of pulmonary disorders. All patients with idiopathic pulmonary fibrosis (IPF) and some patients with other fibrosing ILDs (fILDs) develop progressive pulmonary fibrosis (PPF), characterised by worsening of lung function, symptoms and quality of life, and high mortality [1]. fILDs have a variable clinical course [2–5]. The introduction of effective therapies [6–9] has increased the need for biomarkers that predict outcomes in patients with fILDs, but likely added to the complexity of predicting outcomes. Forced vital capacity (FVC), the pulmonary function biomarker routinely used in clinical trials, has limitations, including a low sensitivity to small changes and the need for serial measurement to determine longitudinal changes that inform prognosis [10, 11]. In addition, since antifibrotic therapies slow FVC decline, the incremental benefit of new drugs will require increasingly sensitive biomarkers to capture treatment effects.

High-resolution computed tomography (HRCT) plays a key role in the diagnosis and management of fILDs [1]. Classification of the pattern evident on HRCT, in particular the identification of the usual



interstitial pneumonia (UIP) pattern, and assessment of disease extent, allow the identification of patients at higher risk of progression regardless of the underlying disease [12–18]. However, qualitative visual assessment of HRCT is limited by high inter-observer variability and shows low sensitivity in the assessment of changes over time [19–22]. Honeycombing, a feature of UIP on HRCT and a predictor of mortality, is often misclassified, particularly when there is coexistent emphysema [19, 23]. Composite scoring systems that integrate HRCT findings with demographic information and pulmonary function tests (PFTs) can improve mortality predictions [24–26] but have not been widely implemented.

The shortcomings of visual assessment have motivated the development of computer-based tools to analyse and quantify disease on HRCT. Software for quantification of disease extent is usually referred to as quantitative CT (QCT). These software packages have moved from histogram- or threshold-based analysis of lung density to software based on feature engineering and machine learning (ML) [27]. Basic patterns quantified by this software include reticulation, ground-glass attenuation and honeycombing, while some tools assess specific parameters; for example, subtypes of low-attenuation areas are quantified by CALIPER (Computer-Aided Lung Informatics for Pathology Evaluation and Ratings) [28]. Unlike visual scoring, these applications provide objective and reproducible results and are sensitive to longitudinal changes in disease extent [26, 29, 30]. Computerised applications can also detect changes in lung volume [31, 32]. QCT parameters are predictive of measures of ILD progression, including FVC decline and mortality [18, 31, 33–38]. Importantly, QCT tools are able to identify prognostic biomarkers that are not detectable to human eyes, such as the vessel-related structure measured by CALIPER, which is a strong predictor of mortality in patients with IPF [34].

Functional respiratory imaging (FRI) enables the quantification of functional parameters in addition to attenuation-based analysis [39]. By applying a computational flow simulation, airways resistance can be calculated. FRI performs volumetric structural analysis of airways and vascular structures (figure 1). The airway volume calculated by FRI provides a way to quantify traction bronchiectasis and has been shown to be predictive of functional decline in patients with IPF, including those with preserved FVC [40].

Applications developed to provide a diagnosis or pattern classification of HRCT images are mainly based on artificial intelligence (AI) [41]. Initially these tools were mainly used for recognising specific features, such as honeycombing or reticulation, with high accuracy on small patches of images. In the last few years, software able to identify patterns on whole CT images has become available [42]. Software based on AI that provides a radiological classification of HRCT based on guideline criteria for a UIP pattern has demonstrated a performance equivalent to that of expert thoracic radiologists [43].

Most of the QCT parameters described here have been validated against PFTs and patient outcomes [30, 34, 36, 44–47]. The use of derivation–validation cohort study designs has helped to prevent spurious conclusions being drawn from models overfitted on a single population [34, 43]. The sensitivity of these tools for identifying patients at high risk of progressive disease means that they could facilitate enrichment of clinical trial populations, reducing the sample size and follow-up time required to detect treatment effects with confidence and so reducing costs [34].

Unmet needs: imaging frontiers in fibrotic lung disease

The advent of antifibrotic therapies has created an urgent need for biomarkers that reliably predict disease behaviour and response to treatment. Disease monitoring, characterisation of early disease and prediction of progressive fibrosis represent three clinical problems for which QCT may provide solutions.

Disease monitoring

QCT consistently outperforms human-based CT evaluation. However, many traditional QCT tools rely on “feature engineering” with the computer quantifying pre-specified CT patterns after training by expert radiologists, a process that is time-consuming, requires high-level domain expertise and incorporates the lack of reliability of visual analysis. Further it requires that the image features best reflecting disease extent or progression be known *a priori* and does not allow for the discovery of novel imaging biomarkers that may be clinically significant but are undetectable visually. These difficulties are overcome if image features are learned automatically using a general-purpose learning procedure. Deep learning is a form of ML that identifies patterns in high-dimensional data, such as the voxel data in medical images, and maps those patterns to pre-specified end-points such as diagnosis or disease progression [43, 48–50]. Data-driven textural analysis combines an unsupervised approach based on cluster analysis of voxel data and a convolutional neural network to detect and quantify fibrosis on CT (figure 2) [30, 51, 52].

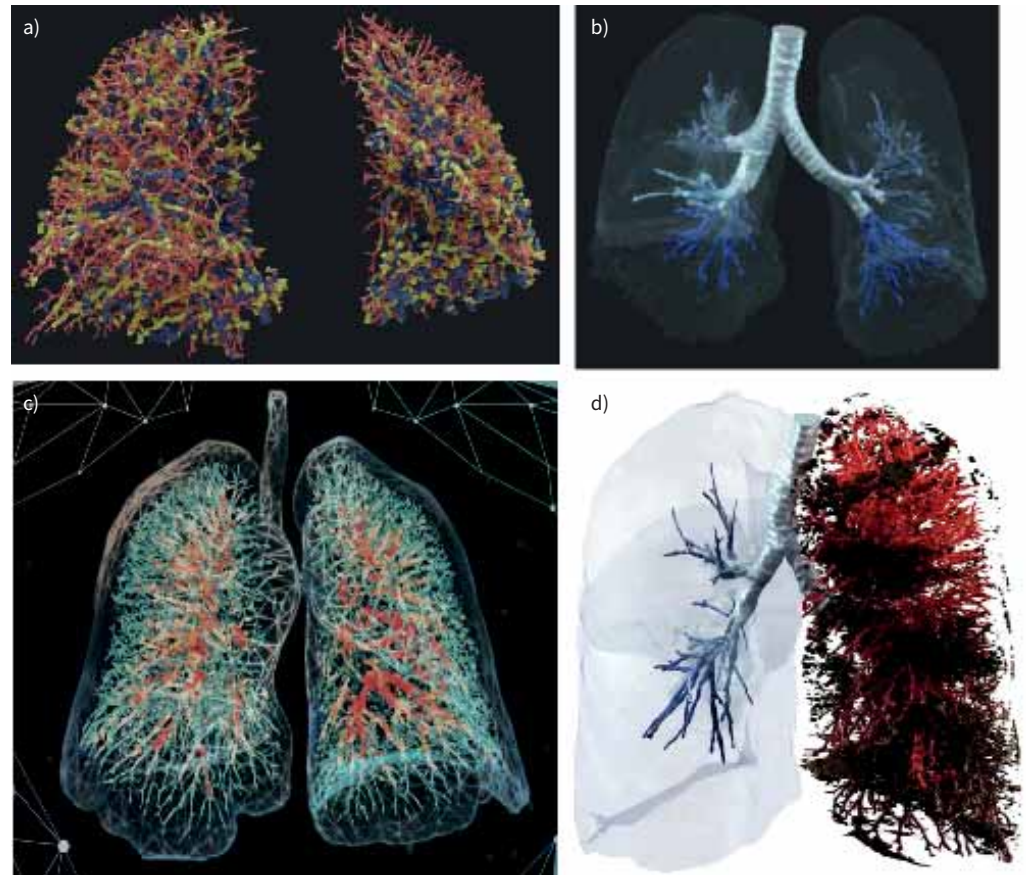


FIGURE 1 Images obtained using functional respiratory imaging, which combines high-resolution computed tomography and computational fluid dynamics technology. **a)** Blood vessel volumes in a patient with idiopathic pulmonary fibrosis (IPF). **b)** Airways in a patient with IPF. **c)** Blood vessel volumes in a patient with coronavirus disease 2019. **d)** Airways, blood vessel volumes and emphysema in a patient with COPD. With permission from Jan De Backer (FLUIDDA).

Predicting progressive pulmonary fibrosis

PPF or the “progressive fibrotic phenotype” are umbrella terms that describe patients with an ILD other than IPF that has progressed despite management based on “standard” therapy for that ILD (e.g. immunosuppressive therapy for many ILDs). Criteria for the identification of PPF are generally based on decline in lung function, worsening radiological fibrosis and deterioration in symptoms, although slightly different criteria have been used in different studies [9, 53–60]. In May 2022, a group of international associations published criteria for the identification of PPF in patients with ILDs other than IPF, acknowledging that these require further validation [1]. PPF is associated with poor outcomes, including high mortality [55, 56–61] and a drawback of the current criteria is that patients must show significant progression, which is likely irreversible, before PPF is identified and treatment can be started. Thus, the lack of reliable predictive data at baseline represents an unmet need for the effective management for patients with ILDs. Predicting PPF using CT can be viewed as a pattern-recognition problem amenable to computer-based analysis, *i.e.* by anchoring computer model training to future disease behaviour. This might permit the identification of CT biomarkers or phenotypes that stratify patients into groups based on outcome.

Characterisation of interstitial lung abnormalities (ILAs)

ILAs present a challenging imaging problem. Longitudinal cohort studies have demonstrated shared clinical and genetic associations between ILAs that are incidentally detected on CT and IPF [62–64]. However, the prevalence of ILAs greatly exceeds that of IPF, meaning that only a minority of individuals with ILA develop IPF. Currently, it is not possible to predict which ILAs will progress. As with the diagnosis of ILDs, classification of ILAs is based on visually defined CT patterns, not disease behaviour.

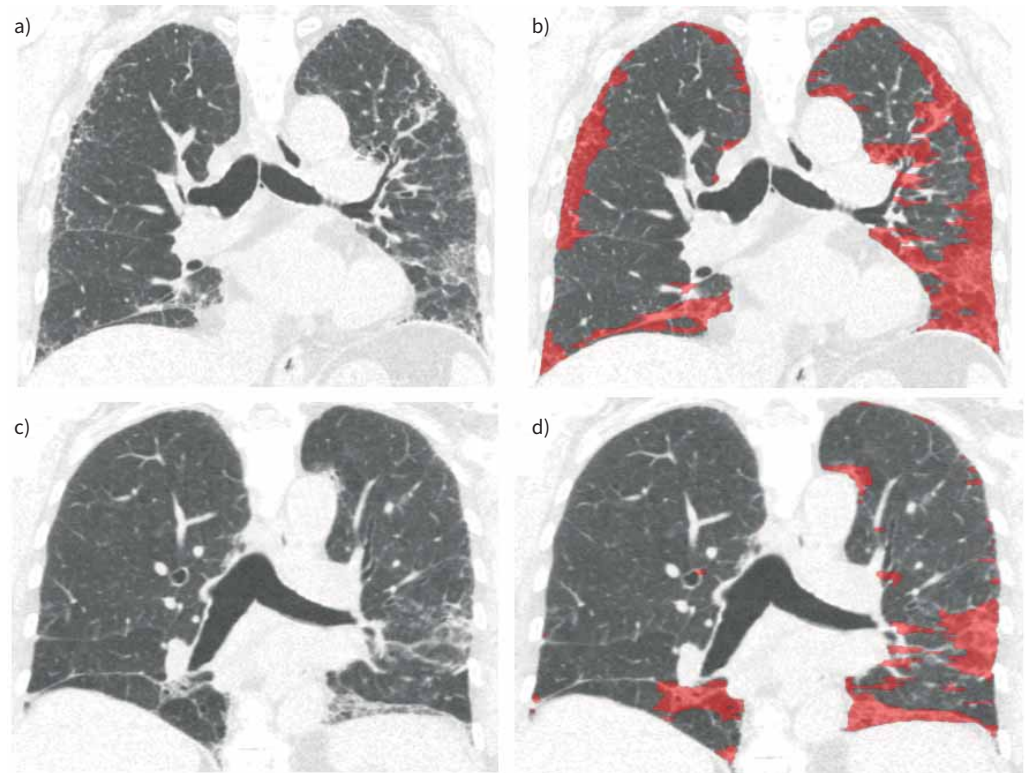


FIGURE 2 Coronal reconstructions with and without data-driven textural analysis (DTA) overlay of **a)** a patient with idiopathic pulmonary fibrosis (DTA score=36.0) and **b)** a patient with nonspecific interstitial pneumonia (DTA score=14.4). With permission from Stephen Humphries (UCLA, Los Angeles, CA, USA).

However, unlike in established fILD, where there are agreed definitions for CT patterns, the current definition of an ILA captures essentially any CT pattern that exceeds 5% in any lung zone [65]. This makes it challenging to develop prognostic computer models based on ILA morphology. Novel ML-based image analysis approaches may provide a solution by anchoring algorithmic training to the behaviour of ILAs rather than to their morphology, avoiding the need to make *a priori* assumptions about the prognostic value of individual ILA patterns.

Obstacles and potential solutions

Quantity and quality of data

A significant challenge to capitalising on novel approaches to image analysis in fILDs is that legacy imaging repositories are often of insufficient quality or quantity to drive algorithm development [66]. This is particularly true in ILAs, where cohort enrichment strategies are likely to be needed. As visual assessment is regarded as relatively insensitive to subtle changes on CT, high-quality longitudinal CT datasets in fILDs are scarce. The insufficient quantity of data means that datasets need to be combined, but even this may not solve the problem of inadequate diversity (and so generalisability) of the data set. Solving these problems will require the development and implementation of imaging protocol standards that will provide data of sufficient quality for use in computerised applications and the use of consortia such as the Open Source Imaging Consortium (OSIC) to collate data from multiple institutions in different countries in a centralised way.

Consensus on anchors

A further obstacle to the development of QCT tools is the lack of agreement on the reference standards that should be used for training the computer. This lack of adequate anchors applies to diagnosis, imaging features and physiological changes. While multidisciplinary team (MDT) discussions are regarded as the gold standard for diagnosis of ILDs and have been shown to improve diagnostic confidence for IPF [67, 68], there is no consensus on the composition or procedure for conducting an MDT discussion [69]. In addition, MDT discussions have not been shown to be associated with improved outcomes. Imaging features used as anchors

are subject to issues of high inter-observer variability, poor reproducibility and low sensitivity to change [20, 22]. Physiological anchors such as FVC and diffusing capacity are subject to variability and missing data and there is no consensus on what constitutes “progression” or “treatment response”. The solution to these problems will be the collection of more data on hard end-points, as well as on diagnostic, imaging and physiological features robustly associated with mortality, to use as anchors in the development of algorithms.

Governance issues

The data governance considerations that apply to the development of QCT applications for fILDs are by no means unique to this field. To improve patient care across diseases, there is a need to make healthcare systems “data ready”, *i.e.* structured in a way that enables the data needed to develop algorithms to be easily pulled from the system. However, this must be done in a way that ensures the security and confidentiality of the data and respects patients’ privacy. The potential for bias in the data collected (leading to inadequate diversity and so to a lack of generalisability of the algorithms developed) also needs to be considered. The solution to these challenges lies in the centralised collection of data with the best governance standards applied, clarity on data ownership and communication of the value of “big data” for research that will ultimately benefit patients.

Open-source data in a closed ecosystem

Initiatives that aim to effect change may be hindered by issues that are not only financial, legal or technological, but also arise from ingrained beliefs, emotions and behaviours. Challenges to the creation of an open-source repository of HRCT scans from patients with fILDs included a historically closed ecosystem, regulatory unknowns and the fear that accompanies data sharing. In creating the OSIC Cloud Data Repository (<https://repository.osicild.org/repository>), the necessity for patient anonymisation/pseudonymisation was complicated by the need to link patients’ clinical data to their imaging data. OSIC’s solution was a three-pronged workflow that 1) created an offline workflow, including a 38-page electronic case report form, to anonymise data prior to uploading, 2) provided a secure environment in which the data could be accessed for quality control purposes by contributors *via* an encryption key prior to allowing wider access, and 3) allowed OSIC to review data prior to being uploaded into the database.

A core principle of OSIC is to combine varied perspectives. Collaboration among experts in pulmonology, thoracic radiology and ML/AI from multiple institutions is the key to addressing multifarious challenges. OSIC is driven by several cross-functional working groups designed to address the biggest issues. These include data privacy and architecture, regulatory framework, publication, research and ethics, biomarker development, and a patient working group. All work is conducted under the premise that OSIC exists to make “radical progress” on behalf of patients, by finding digital imaging biomarkers that inform diagnosis, prognosis and response to therapy.

The OSIC operating model was built with incentives aimed at upending the closed ecosystem that discouraged sharing data. Small start-ups are offered the same access to data as large pharmaceutical companies. Thus, a small institution can offer 500 anonymised scans to the database in exchange for access to 15 000 scans for their research project. Algorithms produced by OSIC will be made available to all. The status of the OSIC Data Repository is summarised in figure 3 and updated regularly at: www.osicild.org/dr-about.html

Perception that QCT has no impact on management decisions

One of the reasons quoted by radiologists and pulmonologists to explain the limited adoption of QCT is that they are unclear how the results, *i.e.* the numbers, would impact clinical decisions. This reflects the limitations of the available data. Knowing the percentage of the lung affected by honeycombing or airway volumes in millilitres is unlikely to change physicians’ behaviour unless there is evidence that acting on these numbers improves clinical outcomes. However, this does not mean that the numbers have no value. Conventional visual assessments of CT scans by radiologists are generally characterised by qualitative expressions such as “slight dilation of the airways” or “severe reticulation in the lower lobes”. These qualitative descriptions often differ substantially between radiologists [70] and it is hard to say if features are more or less “severe” in a follow-up scan. Thus, even in the absence of prognostic algorithms, including numbers in the descriptions of CT scans can facilitate follow-up.

Clinical outcomes might be improved if physicians had tools indicating that clinical worsening, in terms of symptoms or functional decline, is imminent. In such a case, the patient could be monitored more intensely or treatment could be adjusted. There is an increasing volume of data to suggest that this may be possible. In a study evaluating a texture-based QCT tool in patients with IPF, the extent of reticular opacity was an independent predictor of decline in FVC [33]. A study using CALIPER showed that an extent of ILD

Overview of OSIC Cloud Data Repository, November 2023

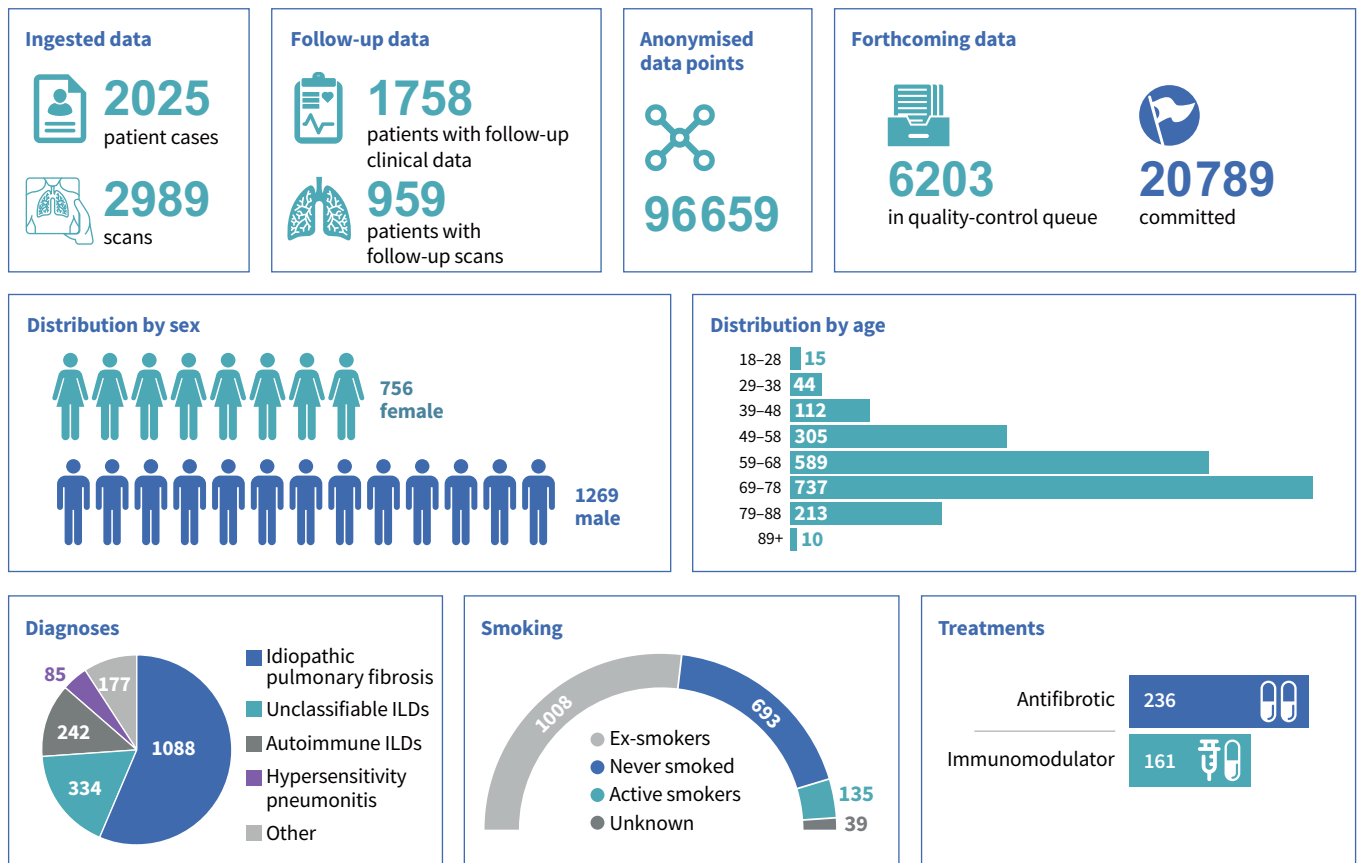


FIGURE 3 Status of the Open Source Imaging Consortium (OSIC) Cloud Data Repository. ILD: interstitial lung disease.

>20% at baseline was predictive of prognosis in patients with IPF [45]. Among patients with unclassifiable ILD, change in the extent of fibrosis measured using CALIPER was a stronger predictor of outcome than any other baseline or longitudinal variable [18]. Change in the severity of traction bronchiectasis measured using CALIPER predicted mortality in patients with IPF independent of FVC decline [71].

A more ambitious role of QCT would be to identify the optimal treatment pathway. To achieve this, CT-based end-points need to be included in treatment trials to allow identification of potential responder phenotypes. The United States Food and Drug Administration (FDA) has been supportive of this initiative [72] and several drugs have been assessed using a form of QCT. For example, a phase IIIb trial of nintedanib in patients with IPF showed a smaller increase in quantitative lung fibrosis score over 6 months in patients who received nintedanib than placebo [73]. A phase II trial in patients with IPF showed a smaller increase in quantitative lung fibrosis score over 24 weeks in patients who received pamrevlumab than placebo [74]. A phase II trial of an autotaxin inhibitor in patients with IPF showed stabilisation in specific airway volume in the treatment group compared to deterioration in the placebo group [75]. Scleroderma Lung Study I showed that in patients with ILD associated with systemic sclerosis, treatment with cyclophosphamide was associated with a reduction in quantitative ILD score at 1 year compared with placebo [76].

Perception that QCT/AI-based tools are a “black box”

The implementation of AI in radiology has been slowed by doubts regarding the quality and safety of the applications and a lack of trust in the results they provide [77, 78]. These are in large part a consequence of the limited “explainability” of the predictive models used by many ML algorithms, which are fundamentally perceived as a “black box” [79]. In addition, there are concerns that the algorithms may be biased due to confounders in the training data (“Clever Hans solutions”) [80] or the diversity of healthcare across geographies [81]. These concerns need to be addressed to increase trust in this new technology

[82, 83]. There are several approaches to help understand the reason for individual prediction results or the types of relationship in the training data captured by models [84]. In models deriving classification or prediction from imaging data, backpropagation approaches such as GradCAM [85] or guided backpropagation [86] mark the regions responsible for a particular prediction. Methods that enable the calculation of uncertainty estimates exist within the ML framework [87–89]. Data visualisation techniques and even semantics can also help enhance the interpretability of the results of ML models [79]. Also important is evaluation of the comparative effectiveness of ML against other methodologies.

Over time, the role of ML in automating the assessment of imaging data will be complemented by its role in improving understanding of the biology of disease and treatment response. Prediction models that integrate patterns in imaging data and other variables may change our capability to steer individual treatment. The integration of ML at earlier stages of drug development and in the study of biological mechanisms may have a massive impact. Here, ML can contribute techniques for the generation of hypotheses based on large realistic patient populations and their paths through disease and treatment. ML will also play a role in clinical trials, both for linking multivariate patient profiles to the prediction of response and to exploit these relationships to inform further studies and *in vitro* experiments. Here, interpretability will be a key requirement to translate representations in models to tools for understanding mechanisms.

Regulatory hurdles

For a long time, it was unclear under which regulation CT post-processing software should be categorised, but it is now clear that this type of software is subject to the medical device regulation. In practice, this implies that companies need to obtain FDA 510k or *de novo* and ISO13485/CE clearance for US and EU market authorisation, respectively. Other large markets such as Japan and Australia require their own clearances, but once FDA and EU clearance have been obtained, the other approval processes are fairly straightforward. Regulatory frameworks are highly geared towards physical devices. Although there are efforts to update these frameworks, especially focusing on AI, it remains a challenge for companies to reconcile regulatory requirements with efficient operations.

The FDA 510k clearance pathway requires a company to identify a so-called “predicate device”, *i.e.* an already cleared device similar to the new device. If the company can demonstrate equivalence, additional clinical evidence is not required. At present, most QCT packages that are cleared limit their context of use to providing “raw data”, without providing advice or claims about diagnosis, prognosis or treatment selection. Radiologists and/or clinicians are then able to use the data for clinical decision-making based on their expertise. However, it is not always clear to radiologists and clinicians what the quantifications mean clinically. This problem could be solved in part by education on the clinical utility of QCT parameters. If the clinical interpretation of visual reads is known, understanding the utility of QCT parameters, such as the quantification of lung nodules and textures like honeycombing and ground glass opacity, should be straightforward. However, demonstrating the utility of QCT parameters in improving clinical outcomes will require further prospective clinical trials, which are expensive. In addition, to generate a return on investment, the improvement in clinical outcomes would need to be translated into adequate reimbursement by governments and insurers for the QCT algorithms. While this process has been well established for therapeutics, it is still in its infancy for software-based diagnostics and prognostics. An additional challenge for the EU market is the Medical Device Regulation [90], which remains challenging even for experienced regulatory professionals. While it will take a few years to establish best practice guidelines, this provides an opportunity to align regulation with developments such as AI.

Conclusions

Identifying biomarkers that allow reliable prediction of disease course or response to therapy remains an unmet need in the management of patients with fILDs. HRCT is a compelling target for biomarker research. Driven by the shortcomings associated with visual assessment, over the past decade, there has been a surge of interest in computer-based methods for quantification of fibrotic lung disease on HRCT. Research has accelerated with the arrival of AI technology, which has the potential to identify clinically important imaging biomarkers without human supervision, including those that are visually inaccessible. Computer-based image analysis might be applied to disease monitoring, the prediction of disease behaviour and the characterisation of ILAs. However, to optimise the implementation of this technology, several technical and societal challenges need to be addressed. Large and diverse imaging repositories must be established to drive research into ML. Such repositories require bespoke clinical governance frameworks to manage data sharing and privacy concerns and quantify potential biases. Well-defined reference standards for diagnosis and disease progression, against which computer algorithms can be trained, must be agreed. Intuitive methods for enhancing the interpretability of algorithms are needed to alleviate the perception that AI-based tools are

“black boxes”. Finally, “light-touch” regulatory pathways, which ensure proper validation of algorithms while facilitating their implementation into clinical care, are needed.

Points for clinical practice

- QCT provides a more reliable assessment of patterns and extent of disease on HRCT than visual assessment.
- QCT parameters predict disease progression and mortality in patients with fILDs.
- Application of QCT software requires that HRCT be performed using standard protocols.
- The provision of HRCT scans and clinical data collected in clinical practice to central repositories will enable the development of algorithms that inform diagnosis and prognosis.

Questions for future research

- Which QCT parameters most reliably predict the progression of fILDs?
- Can QCT be used to predict response to treatment?
- How do features identified on QCT relate to the underlying pathophysiology of fILDs?
- Which QCT parameters should be used as end-points in clinical trials?
- Should QCT be used in the monitoring of patients with fILDs?
- How can ML and AI algorithms be made “understandable” to radiologists and clinicians?
- How should new technologies be regulated within clinical governance frameworks?

Provenance: Submitted article, peer reviewed.

Acknowledgements: Writing assistance for this article was provided by Wendy Morris of Fleishman-Hillard, London, UK, supported by OSIC.

Conflicts of interest: All the authors are members of OSIC. In addition, S.L.F. Walsh reports relationships with Boehringer Ingelheim (BI), Bracco, FLUIDDA, Galapagos, OncoArendi, Roche and Sanofi-Genzyme. J. De Backer owns shares in FLUIDDA. H. Prosch reports grants from BI; payment for presentations from AstraZeneca, Bristol Myers Squibb (BMS), BI, Janssen, Merck Sharp & Dohme (MSD), Novartis, Roche/InterMune, Sanofi, Siemens and Takeda; support for travel from BI; and has served on an Advisory Board for BMS, BI, MSD, Roche/InterMune and Sanofi. G. Lungs reports payment for presentations from Novartis; research support from Novartis and NVIDIA; and is co-founder, shareholder and Chief Scientist of contextflow. L. Calandriello has served on an Advisory Board for and received payment for presentations from BI. V. Cottin reports grants from BI; consulting fees from BI, FibroGen, Galapagos, Galecto, PureTech, RedX, Roche and Shionogi; payment for presentations and support for attending meetings from BI and Roche; and has served on a Data Safety Monitoring Board or Advisory Board for Celgene, BMS, Galapagos and Roche/Promedior. K.K. Brown reports grants from NHLBI; consultancy fees, speaker fees, support for travel and/or has served as an advisor or on the data monitoring committee for AbbVie, Biogen, Blade Therapeutics, BI, BMS, CSL Behring, DevPro Biopharma, Dispersol, Eleven P15, Galapagos, Galecto, Huitai Biomedicine, Humanetics, Pliant, Redx Pharma, Sanofi, Third Pole and Translate Bio; and he holds a leadership role with the Fleischner Society. Y. Inoue reports grants from the Japanese Ministry of Health, Labour, and Welfare and the Japan Agency for Medical Research and Development; payment for presentations from BI, Kyorin, Shionogi, GlaxoSmithKline, and ThermoFisher; and has served as a consultant or steering committee member for BI, Galapagos, Roche, Taiho, CSL Behring, Vicore Pharma and Savara. V. Tzilas reports no disclosures. E. Estes reports no disclosures.

Support statement: The work of G. Lungs and H. Prosch was partially funded by the Austrian Science Fund (FWF ONSET P 35189-B).

References

- 1 Raghu G, Remy-Jardin M, Richeldi L, *et al.* Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2022; 205: e18–e47.
- 2 Ley B, Bradford WZ, Vittinghoff E, *et al.* Predictors of mortality poorly predict common measures of disease progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016; 194: 711–718.
- 3 Adegunsoye A, Oldham JM, Chung JH, *et al.* Phenotypic clusters predict outcomes in a longitudinal interstitial lung disease cohort. *Chest* 2018; 153: 349–360.

- 4 Hyldgaard C, Ellingsen T, Hilberg O, *et al.* Rheumatoid arthritis-associated interstitial lung disease: clinical characteristics and predictors of mortality. *Respiration* 2019; 98: 455–460.
- 5 Hoffmann-Vold AM, Allanore Y, Alves M, *et al.* Progressive interstitial lung disease in patients with systemic sclerosis-associated interstitial lung disease in the EUSTAR database. *Ann Rheum Dis* 2021; 80: 219–227.
- 6 Richeldi L, du Bois RM, Raghu G, *et al.* Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2071–2082.
- 7 King TE Jr, Bradford WZ, Castro-Bernardini S, *et al.* A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2083–2092.
- 8 Distler O, Highland KB, Gahlemann M, *et al.* Nintedanib for systemic sclerosis-associated interstitial lung disease. *N Engl J Med* 2019; 380: 2518–2528.
- 9 Flaherty KR, Wells AU, Cottin V, *et al.* Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med* 2019; 381: 1718–1727.
- 10 Paterniti MO, Bi Y, Rekić D, *et al.* Acute exacerbation and decline in forced vital capacity are associated with increased mortality in idiopathic pulmonary fibrosis. *Ann Am Thorac Soc* 2017; 14: 1395–1402.
- 11 Brown KK, Inoue Y, Flaherty KR, *et al.* Predictors of mortality in subjects with progressive fibrosing interstitial lung diseases. *Respirology* 2022; 27: 294–300.
- 12 Flaherty KR, Thwaite EL, Kazerooni EA, *et al.* Radiological versus histological diagnosis in UIP and NSIP: survival implications. *Thorax* 2003; 58: 143–148.
- 13 Lynch DA, Godwin JD, Safrin S, *et al.* High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. *Am J Respir Crit Care Med* 2005; 172: 488–493.
- 14 Edey AJ, Devaraj AA, Barker RP, *et al.* Fibrotic idiopathic interstitial pneumonias: HRCT findings that predict mortality. *Eur Radiol* 2011; 21: 1586–1593.
- 15 Walsh SL, Sverzellati N, Devaraj A, *et al.* Chronic hypersensitivity pneumonitis: high resolution computed tomography patterns and pulmonary function indices as prognostic determinants. *Eur Radiol* 2012; 22: 1672–1679.
- 16 Walsh SL, Sverzellati N, Devaraj A, *et al.* Connective tissue disease related fibrotic lung disease: high resolution computed tomographic and pulmonary function indices as prognostic determinants. *Thorax* 2014; 69: 216–222.
- 17 Adegunsoye A, Oldham JM, Bellam SK, *et al.* Computed tomography honeycombing identifies a progressive fibrotic phenotype with increased mortality across diverse interstitial lung diseases. *Ann Am Thorac Soc* 2019; 16: 580–588.
- 18 Jacob J, Bartholmai BJ, Rajagopalan S, *et al.* Unclassifiable-interstitial lung disease: outcome prediction using CT and functional indices. *Respir Med* 2017; 130: 43–51.
- 19 Watadani T, Sakai F, Johkoh T, *et al.* Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology* 2013; 266: 936–944.
- 20 Walsh SL, Calandriello L, Sverzellati N, *et al.* Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax* 2016; 71: 45–51.
- 21 Taha N, D'Amato D, Hosein K, *et al.* Longitudinal functional changes with clinically significant radiographic progression in idiopathic pulmonary fibrosis: are we following the right parameters? *Respir Res* 2020; 21: 119.
- 22 Widell J, Lidén M. Interobserver variability in high-resolution CT of the lungs. *Eur J Radiol Open* 2020; 7: 100228.
- 23 Akira M, Inoue Y, Kitaichi M, *et al.* Usual interstitial pneumonia and nonspecific interstitial pneumonia with and without concurrent emphysema: thin-section CT findings. *Radiology* 2009; 251: 271–279.
- 24 Ley B, Elicker BM, Hartman TE, *et al.* Idiopathic pulmonary fibrosis: CT and risk of death. *Radiology* 2014; 273: 570–579.
- 25 Chahal A, Sharif R, Watts J, *et al.* Predicting outcome in idiopathic pulmonary fibrosis: addition of fibrotic score at thin-section CT of the chest to gender, age, and physiology score improves the prediction model. *Radiol Cardiothorac Imaging* 2019; 1: e180029.
- 26 Sverzellati N, Silva M, Seletti V, *et al.* Stratification of long-term outcome in stable idiopathic pulmonary fibrosis by combining longitudinal computed tomography and forced vital capacity. *Eur Radiol* 2020; 30: 2669–2679.
- 27 Wu X, Kim GH, Salisbury ML, *et al.* Computed tomographic biomarkers in idiopathic pulmonary fibrosis. The future of quantitative analysis. *Am J Respir Crit Care Med* 2019; 199: 12–21.
- 28 Jacob J, Bartholmai BJ, Rajagopalan S, *et al.* Automated quantitative computed tomography versus visual computed tomography scoring in idiopathic pulmonary fibrosis: validation against pulmonary function. *J Thorac Imaging* 2016; 31: 304–311.
- 29 Kim HJ, Brown MS, Chong D, *et al.* Comparison of the quantitative CT imaging biomarkers of idiopathic pulmonary fibrosis at baseline and early change with an interval of 7 months. *Acad Radiol* 2015; 22: 70–80.
- 30 Humphries SM, Yagihashi K, Huckleberry J, *et al.* Idiopathic pulmonary fibrosis: data-driven textural analysis of extent of fibrosis at baseline and 15-month follow-up. *Radiology* 2017; 285: 270–278.
- 31 Si-Mohamed SA, Nasser M, Colevray M, *et al.* Automatic quantitative computed tomography measurement of longitudinal lung volume loss in interstitial lung diseases. *Eur Radiol* 2022; 32: 4292–4303.

- 32 Strambu IR, Seemayer CA, Fagard LMA, et al. GLPG1205 for idiopathic pulmonary fibrosis: a phase 2 randomised placebo-controlled trial. *Eur Respir J* 2023; 61: 2201794.
- 33 Park HJ, Lee SM, Song JW, et al. Texture-based automated quantitative assessment of regional patterns on initial CT in patients with idiopathic pulmonary fibrosis: relationship to decline in forced vital capacity. *Am J Roentgenol* 2016; 207: 976–983.
- 34 Jacob J, Bartholmai BJ, Rajagopalan S, et al. Predicting outcomes in idiopathic pulmonary fibrosis using automated computed tomographic analysis. *Am J Respir Crit Care Med* 2018; 198: 767–776.
- 35 Jacob J, Bartholmai BJ, Rajagopalan S, et al. Serial automated quantitative CT analysis in idiopathic pulmonary fibrosis: functional correlations and comparison with changes in visual CT scores. *Eur Radiol* 2018; 28: 1318–1327.
- 36 Humphries SM, Mackintosh JA, Jo HE, et al. Quantitative computed tomography predicts outcomes in idiopathic pulmonary fibrosis. *Respirology* 2022; 27: 1045–1053.
- 37 Pan J, Hofmanninger J, Nenning KH, et al. Unsupervised machine learning identifies predictive progression markers of IPF. *Eur Radiol* 2023; 33: 925–935.
- 38 Yeo J, Yoon SH, Kim JY, et al. Quantitative interstitial lung disease scores in idiopathic inflammatory myopathies: longitudinal changes and clinical implications. *Rheumatology* 2023; 62: 3690–3699.
- 39 Hajian B, De Backer J, Vos W, et al. Functional respiratory imaging (FRI) for optimizing therapy development and patient care. *Expert Rev Respir Med* 2016; 10: 193–206.
- 40 Clukers J, Lanclus M, Mignot B, et al. Quantitative CT analysis using functional imaging is superior in describing disease progression in idiopathic pulmonary fibrosis compared to forced vital capacity. *Respir Res* 2018; 19: 213.
- 41 Anthimopoulos M, Christodoulidis S, Ebner L, et al. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging* 2016; 35: 1207–1216.
- 42 Gao M, Bagci U, Lu L, et al. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Comput Methods Biomech Biomed Eng Imaging Vis* 2018; 6: 1–6.
- 43 Walsh SLF, Calandriello L, Silva M, et al. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018; 6: 837–845.
- 44 Salisbury ML, Lynch DA, van Beek EJ, et al. Idiopathic pulmonary fibrosis: the association between the adaptive multiple features method and fibrosis outcomes. *Am J Respir Crit Care Med* 2017; 195: 921–929.
- 45 Romei C, Tavanti LM, Taliani A, et al. Automated computed tomography analysis in the assessment of idiopathic pulmonary fibrosis severity and progression. *Eur J Radiol* 2020; 124: 108852.
- 46 Oh JH, Kim GHJ, Cross G, et al. Automated quantification system predicts survival in rheumatoid arthritis-associated interstitial lung disease. *Rheumatology* 2022; 61: 4702–4710.
- 47 Volkmann ER, Tashkin DP, Roth MD, et al. Early radiographic progression of scleroderma: lung disease predicts long-term mortality. *Chest* 2022; 161: 1310–1319.
- 48 Walsh SLF, Humphries SM, Wells AU, et al. Imaging research in fibrotic lung disease: applying deep learning to unsolved problems. *Lancet Respir Med* 2020; 8: 1144–1153.
- 49 Shaish H, Ahmed FS, Lederer D, et al. Deep learning of computed tomography virtual wedge resection for prediction of histologic usual interstitial pneumonitis. *Ann Am Thorac Soc* 2021; 18: 51–59.
- 50 Walsh SL, Mackintosh JA, Calandriello L, et al. Deep learning-based outcome prediction in progressive fibrotic lung disease using high-resolution computed tomography. *Am J Respir Crit Care Med* 2022; 206: 883–889.
- 51 Humphries SM, Swigris JJ, Brown KK, et al. Quantitative high-resolution computed tomography fibrosis score: performance characteristics in idiopathic pulmonary fibrosis. *Eur Respir J* 2018; 52: 1801384.
- 52 Oh AS, Lynch DA, Swigris JJ, et al. Deep learning-based fibrosis extent on CT predicts outcome of fibrosing interstitial lung disease independent of visually assessed CT pattern. *Ann Am Thorac Soc* 2023; 21: 218–227.
- 53 Maher TM, Corte TJ, Fischer A, et al. Pirfenidone in patients with unclassifiable progressive fibrosing interstitial lung disease: a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet Respir Med* 2020; 8: 147–157.
- 54 Behr J, Prasse A, Kreuter M, et al. Pirfenidone in patients with progressive fibrotic interstitial lung diseases other than idiopathic pulmonary fibrosis (RELIEF): a double-blind, randomised, placebo-controlled, phase 2b trial. *Lancet Respir Med* 2021; 9: 476–486.
- 55 Gagliardi M, Berg DV, Heylen CE, et al. Real-life prevalence of progressive fibrosing interstitial lung diseases. *Sci Rep* 2021; 11: 23988.
- 56 Nasser M, Larrieu S, Si-Mohamed S, et al. Progressive fibrosing interstitial lung disease: a clinical cohort (the PROGRESS study). *Eur Respir J* 2021; 57: 2002718.
- 57 Simpson T, Barratt SL, Beirne P, et al. The burden of progressive fibrotic interstitial lung disease across the UK. *Eur Respir J* 2021; 58: 2100221.
- 58 Khor YH, Farooqi M, Hambly N, et al. Patient characteristics and survival for progressive pulmonary fibrosis using different definitions. *Am J Respir Crit Care Med* 2023; 207: 102–105.
- 59 Pugashetti JV, Adegunsoye A, Wu Z, et al. Validation of proposed criteria for progressive pulmonary fibrosis. *Am J Respir Crit Care Med* 2023; 207: 69–76.

- 60 Takei R, Brown KK, Yamano Y, *et al.* Prevalence and prognosis of chronic fibrosing interstitial lung diseases with a progressive phenotype. *Respirology* 2022; 27: 333–340.
- 61 Brown KK, Martinez FJ, Walsh SLF, *et al.* The natural history of progressive fibrosing interstitial lung diseases. *Eur Respir J* 2020; 55: 2000085.
- 62 Hunninghake GM, Hatabu H, Okajima Y, *et al.* MUC5B promoter polymorphism and interstitial lung abnormalities. *N Engl J Med* 2013; 368: 2192–2200.
- 63 Jin GY, Lynch D, Chawla A, *et al.* Interstitial lung abnormalities in a CT lung cancer screening population: prevalence and progression rate. *Radiology* 2013; 268: 563–571.
- 64 Araki T, Putman RK, Hatabu H, *et al.* Development and progression of interstitial lung abnormalities in the Framingham Heart Study. *Am J Respir Crit Care Med* 2016; 194: 1514–1522.
- 65 Hatabu H, Hunninghake GM, Richeldi L, *et al.* Interstitial lung abnormalities detected incidentally on CT: a position paper from the Fleischner Society. *Lancet Respir Med* 2020; 8: 726–737.
- 66 Willemink MJ, Koszek WA, Hardell C, *et al.* Preparing medical imaging data for machine learning. *Radiology* 2020; 295: 4–15.
- 67 Walsh SLF, Wells AU, Desai SR, *et al.* Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case-cohort study. *Lancet Respir Med* 2016; 4: 557–565.
- 68 Walsh SLF, Maher TM, Kolb M, *et al.* Diagnostic accuracy of a clinical diagnosis of idiopathic pulmonary fibrosis: an international case-cohort study. *Eur Respir J* 2017; 50: 1700936.
- 69 Richeldi L, Launders N, Martinez F, *et al.* The characterisation of interstitial lung disease multidisciplinary team meetings: a global study. *ERJ Open Res* 2019; 5: 00209-2018.
- 70 Garrana SH, Som A, Ndakwah GS, *et al.* Comparison of chest CT findings of COVID-19, influenza, and organizing pneumonia: a multireader study. *AJR Am J Roentgenol* 2021; 217: 1093–1102.
- 71 Jacob J, Aksman L, Mogulkoc N, *et al.* Serial CT analysis in idiopathic pulmonary fibrosis: comparison of visual features that determine patient outcome. *Thorax* 2020; 75: 648–654.
- 72 Food and Drug Administration. Qualification Process for Drug Development Tools Guidance for Industry and FDA Staff. Date last accessed: 26 August 2022. Date last updated: November 2020. www.fda.gov/media/133511/download
- 73 Lancaster L, Goldin J, Trampisch M, *et al.* Effects of nintedanib on quantitative lung fibrosis score in idiopathic pulmonary fibrosis. *Open Respir Med J* 2020; 14: 22–31.
- 74 Richeldi L, Fernández Pérez ER, Costabel U, *et al.* Pamrevlumab, an anti-connective tissue growth factor therapy, for idiopathic pulmonary fibrosis (PRAISE): a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet Respir Med* 2020; 8: 25–33.
- 75 Maher TM, van der Aar EM, van de Steen O, *et al.* Safety, tolerability, pharmacokinetics, and pharmacodynamics of GLPG1690, a novel autotaxin inhibitor, to treat idiopathic pulmonary fibrosis (FLORA): a phase 2a randomised placebo-controlled trial. *Lancet Respir Med* 2018; 6: 627–635.
- 76 Kim HJ, Tashkin DP, Gjertson DW, *et al.* Transitions to different patterns of interstitial lung disease in scleroderma with and without treatment. *Ann Rheum Dis* 2016; 75: 1367–1371.
- 77 Strohm L, Hehakaya C, Ranschaert ER, *et al.* Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. *Eur Radiol* 2020; 30: 5525–5532.
- 78 Huisman M, Ranschaert E, Parker W, *et al.* An international survey on AI in radiology in 1041 radiologists and radiology residents part 2: expectations, hurdles to implementation, and education. *Eur Radiol* 2021; 31: 8797–8806.
- 79 Reyes M, Meier R, Pereira S, *et al.* On the interpretability of artificial intelligence in radiology: challenges and opportunities. *Radiol Artif Intell* 2020; 2: e190043.
- 80 Lapuschkin S, Wäldchen S, Binder A, *et al.* Unmasking Clever Hans predictors and assessing what machines really learn. *Nat Commun* 2019; 10: 1096.
- 81 McCradden MD, Joshi S, Anderson JA, *et al.* Patient safety and quality improvement: ethical principles for a regulatory approach to bias in healthcare machine learning. *J Am Med Inform Assoc* 2020; 27: 2024–2027.
- 82 Vellido A. Societal issues concerning the application of artificial intelligence in medicine. *Kidney Dis* 2019; 5: 11–17.
- 83 Cuttillo CM, Sharma KR, Foschini L, *et al.* Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digit Med* 2020; 3: 47.
- 84 Holzinger A, Langs G, Denk H, *et al.* Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* 2019; 9: e1312.
- 85 Selvaraju RR, Cogswell M, Das A, *et al.* Grad-CAM: visual explanations from deep networks *via* gradient-based localization. *Int J Comput Vis* 2019; 128: 336–359.
- 86 Oramas J, Wang K, Tuytelaars T. Visual explanation by interpretation: improving visual feedback capabilities of deep neural networks. *arXiv* 2017; preprint [<https://doi.org/10.48550/arXiv.1712.06302>].
- 87 Pawlowski N, Brock A, Lee MCH, *et al.* Implicit weight uncertainty in neural networks. *arXiv* 2017; preprint [<https://doi.org/10.48550/arXiv.1711.01297>].

- 88 Carter B, Bileschi M, Smith J, *et al.* Critiquing protein family classification models using sufficient input subsets. *J Comput Biol* 2020; 27: 1219–1231.
- 89 Gustafsson FK, Danelljan M, Schön TB. Evaluating scalable Bayesian deep learning methods for robust computer vision. *arXiv* 2019; preprint [<https://doi.org/10.48550/arXiv.1906.01620>].
- 90 European Parliament, Council of the European Union. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. Date last updated: 20 March 2023. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=uriserv:OJ.L_.2017.117.01.0001.01.ENG