Imaging research in fibrotic lung disease; applying deep learning to unsolved problems

Simon L F Walsh, Stephen M Humphries, Athol U Wells, Kevin K Brown**

Over the past decade, there has been a groundswell of research interest in computer-based methods for objectively quantifying fibrotic lung disease on high resolution CT of the chest. In the past 5 years, the arrival of deep learning-based image analysis has created exciting new opportunities for enhancing the understanding of, and the ability to interpret, fibrotic lung disease on CT. Specific unsolved problems for which computer-based imaging analysis might provide solutions include the development of reliable methods for assisting with diagnosis, detecting early disease, and predicting disease behaviour using baseline imaging data. However, to harness this technology, technical and societal challenges must be overcome. Large CT datasets will be needed to power the training of deep learning algorithms. Open science research and collaboration between academia and industry must be encouraged. Prospective clinical utility studies will be needed to test computer algorithm performance in real-world clinical settings and demonstrate patient benefit over current best practice. Finally, ethical standards, which ensure patient confidentiality and mitigate against biases in training datasets, that can be encoded in machine-learning systems will be needed as well as bespoke data governance and accountability frameworks to encourage buy-in from health-care professionals, patients, and the public.

Lancet Respir Med 2020

Published Online
February 25, 2020
https://doi.org/10.1016/S2213-2600(20)30003-5

*These authors contributed equally

National Heart and Lung Institute, Imperial College, London, UK (S L F Walsh MD); **Quantitative Imaging Laboratory, Department of Radiology** (S M Humphries PhD) **and Division of Pulmonary and Critical Care Medicine** (Prof K K Brown MD), **National Jewish Health, Denver, CO, USA; and Interstitial Lung Disease Unit, Royal Brompton Hospital, London, UK** (Prof A U Wells MD)

Correspondence to:
Dr Simon L F Walsh MD, National Heart and Lung Institute, Imperial College, London SW3 6LY, UK
s.walsh@imperial.co.uk

Introduction

The advent of antifibrotic therapy in idiopathic pulmonary fibrosis (IPF) and its extension to other progressive fibrotic lung diseases in ongoing trials has created an urgent need for biomarkers that reliably predict disease behaviour and treatment response.[1–4] In the past decade, there has been a surge of interest in computer-based methods for objectively quantifying fibrotic lung disease on high resolution CT of the chest. This interest has been amplified by problems with reproducibility, high interobserver variability, and the relative insensitivity of visual CT assessment.[5–8] However, traditional image processing also relies heavily on human input to train the computer, which is time consuming, requires high-level domain expertise, and might incorporate some of the unreliability associated with vision-based image analysis.

Medical imaging research has recently entered an accelerated phase with the arrival of artificial intelligence (AI) technology and, in particular, deep learning.[9–12] The advantage of deep learning over traditional image analysis lies in its potential to identify clinically important imaging variables without human supervision, including, crucially, visually undetectable CT variables. However, to harness this technology, technical and societal challenges must be overcome. In this Personal View, we introduce deep learning and an overview of how it might be applied to important unsolved imaging problems in fibrotic lung disease.

Deep learning

For more than 20 years, computer-based quantification of disease on CT, known as quantitative CT, has been used in a variety of fibrotic lung diseases, and in the past few years has consistently outperformed human-based CT evaluation.[13–17] However, despite the apparent benefits of quantitative CT, most traditional quantitative CT tools rely on some form of feature engineering with the computer quantifying prespecified CT patterns only after training by expert radiologists.[14,15,18] This method has three critical limitations. First, the reliability of the training process might be reduced by the subjective nature of CT interpretation.[6,8] Second, feature engineering requires that the image features best reflecting disease extent, disease progression, or response to therapy are known a priori. Feature engineering does not include patterns that might be

Key messages

- In fibrotic lung disease, reliable biomarkers that allow detection of early disease, predict disease progression, and facilitate monitoring of therapeutic response are urgently needed.
- Advances in deep learning technology have created new opportunities in medical imaging analysis.
- A key advantage of deep learning is that it can bypass the need for supervised training and learn the most important features directly from the imaging data. In principle, this approach can be used to develop novel imaging biomarkers as wells as generate accurate imaging analysis tools.
- Specific imaging challenges in fibrotic lung disease research to which deep learning might be applied include improved imaging diagnostics, classification of limited interstitial lung abnormalities based on disease behaviour, and prediction of the progressive fibrotic phenotype using baseline imaging data.
- However, the scarcity of large diverse imaging biorepositories to power deep learning research in fibrotic lung disease represents a major obstacle.
- International collaborative efforts between academia, industry, and philanthropy to address this challenge are underway.

clinically significant, but undetectable by the human eye. Lastly, software training requires manual labelling of images which is time consuming and requires high-level domain expertise. These difficulties are overcome if important image features can be learned automatically using a general-purpose learning procedure; this is the principal advantage of deep learning.[19]

Deep learning is a form of machine learning that efficiently identifies patterns in high dimensional data, such as the voxel data in medical images, and maps these structures to endpoints such as diagnosis and future disease progression. Deep learning is not a new concept; its statistical and mathematical roots were planted in the 1940s and 1950s with the development of computational algorithms known as artificial neural networks. Rapid advances in specialised computer processors known as graphics processing units have created convolution neural networks developed for optimal image classification, and are now the focus of intense research. In respiratory medicine, high accuracy convolution neural networks have been successfully applied to lung cancer detection, predicting mortality in patients with chronic obstructive pulmonary disease and classifying fibrotic lung disease on CT scans.[9,20–22]

The process of training a deep learning algorithm can be illustrated with the example of wanting to predict the presence of histological usual interstitial pneumonia (UIP) on the basis of CT scan appearances. First, a large training dataset of fibrotic lung disease CT scans is needed, with each CT scan paired with a known histopathological label according to the presence or absence of histological UIP.[23] Training the algorithm is an iterative process. For each scan, the CT pattern predicted by the current state of the algorithm is compared with the preassigned histological category for that CT scan. In discordant cases, the algorithm computes its error on that scan and modifies its internal parameters to minimise future discordance through a process called backpropagation. This training procedure is repeated for all images in the training dataset many times over, each reanalysis resulting in incremental improvements in algorithmic predictions with each iteration (figure 1). With a long enough training period on a large, diverse dataset, the algorithm can accurately predict the presence of histological UIP on the basis of CT patterns in an independent, previously unseen, validation patient cohort.

During training, the most discriminative CT features of each diagnostic category are automatically amplified and learned, while irrelevant variations between scans such as imaging technique, are suppressed. The result is an algorithm that is simultaneously sensitive to subtle CT features essential for accurate classification while remaining insensitive to unimportant variations. This behaviour is autonomous; in our example, although the algorithm is not explicitly programmed to recognise basal honeycombing as being highly predictive of histological UIP, during training it is very likely that this pattern will be identified through algorithm learning. Crucially, the algorithm might also learn to identify discriminatory CT features which predict the presence of histological UIP not visible to the radiologist. The advantage of this approach is that there are no a priori assumptions on the selection of CT features that accurately classify the disease, allowing for the possibility that new, important CT features might be disclosed. Attention or saliency mapping can identify regions in an image that contribute to an algorithms output from an individual scan or image (figure 2a–c).[24] Dimensionality reduction algorithms in machine learning can also be used to visualise the way the algorithm classifies disease. This approach, for example, has been used to develop novel deep learning-based phenotypes of skin lesions, which was published in *Nature*.[12] Therefore, as well as enhancing image classification, deep learning might also generate new knowledge about the images it learns to classify.

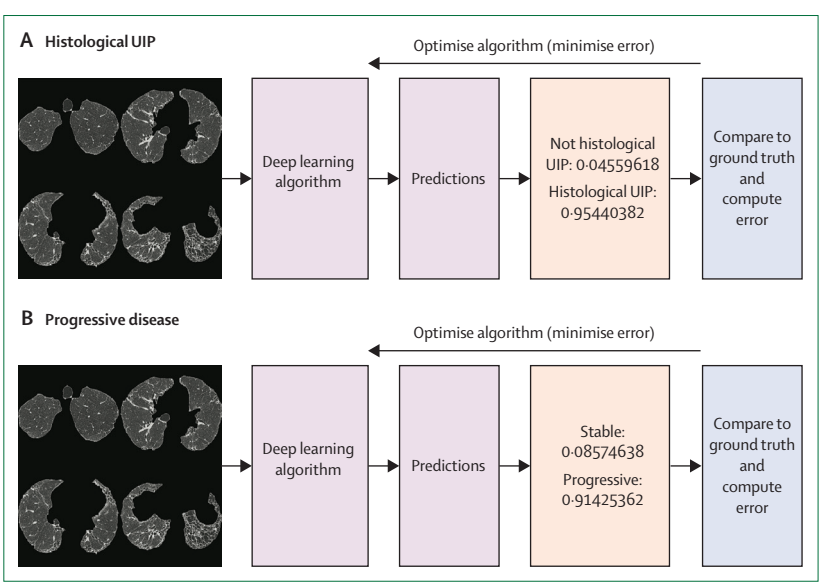

Figure 1: **Simplified illustration of the deep learning process for categorising a CT image**
(A) Categorisation of a CT image based on current American Thoracic Society, European Respiratory Society, Japanese Respiratory Society, and Latin American Thoracic Association guidelines for a UIP pattern. The algorithm input in this example is a montage consisting of four CT images. The algorithm outputs probabilities across all categories, summing to 1. The result is compared with the ground truth label for the image (histological UIP) and an algorithm called backpropagation adjusts the internal parameters, called weights, of the deep learning algorithm, to minimise the error on the input image. The process is repeated for all training samples until the algorithm prediction accuracy cannot be improved further. The trained deep learning algorithm is then tested on a separate dataset of images that are not part of the training dataset. (B) The same process could be applied to CT images using outcome as the training label. UIP=usual interstitial pneumonia.

Applications of deep learning in fibrotic lung disease

Diagnostic support

In the diagnosis of fibrotic interstitial lung disease (ILD), an accurate radiological diagnosis remains a core goal. This goal is particularly true in IPF, where the need for surgical lung biopsy is predicated on accurate CT

 Published online February 25, 2020 https://doi.org/10.1016/S2213-2600(20)30003-5

interpretation in individual patients. In selected patients, a UIP or probable UIP pattern on CT might obviate the need for surgical lung biopsy. Therefore, CT can have considerable influence on subsequent management decisions and if misinterpreted, might lead to unnecessary surgical lung biopsy. Clinical trials also rely heavily on consistent and reproducible CT interpretation for patient enrolment. In the INPULSIS and ASCEND studies, IPF diagnosis was established on the basis of a central review of CT scans and the INBUILD study incorporated cohort enrichment with cases of UIP-like disease based on CT appearances[3].

However, formulating a radiological diagnosis often requires fine judgements on the presence or absence of specific features and their distribution into CT patterns, a process highly susceptible to interobserver variability, even in the hands of expert radiologists. The result is that the use of diagnostic CT criteria, however well defined, can be severely attenuated by inconsistent interpretation. This issue was highlighted by a study reporting only moderate agreement between more than one hundred radiologists applying guideline criteria for a UIP pattern.[8] Paradoxically, greater disagreement was observed between thoracic radiologists with more than 20 years' experience (k=0·39, SD 0·15) compared with those with less than 10 years' experience (k=0·50, 0·17) These findings reflect the challenging nature of CT interpretation but show that, perhaps unsurprisingly, perceptual differences exist between individuals, which are independent of training or experience and therefore not easily overcome. This inconsistency creates opportunities for automated, imaging-based decision support. In a study a deep learning algorithm loosely based on a neural network architecture developed by Google, achieved expert thoracic radiologist-level performance assigning a radiological diagnosis on the basis of the joint IPF guideline criteria for a UIP pattern.[9] In principle, technology such as this could be deployed cheaply where access to imaging expertise is scarce as well as providing consistent patient stratification for clinical trials, thereby reducing screen failures and cost.

Early detection of clinically significant fibrotic lung disease

Interstitial lung abnormalities (ILAs) present a challenging clinical problem (figure 3). Emerging data from longitudinal lung cancer and cardiovascular cohort studies have consistently disclosed some shared clinical and genetic associations between incidentally detected ILAs on CT and IPF.[25,26] ILAs are associated with aging,[27–29] are more common in smokers[25,30,31] and patients older than 50 years expressing MUC5B promoter polymorphism positivity,[29] and ILA progression is associated with greater serial pulmonary function decline when compared to stable ILAs.[32] However, ILA prevalence exceeds that of IPF by almost two orders of magnitude meaning that only a small proportion of ILA are likely to be progressive. Reliably predicting which ILAs will progress is not currently possible although evidence has shown, perhaps unsurprisingly, that in research subjects, those with ILA classified as representing definite fibrosis, UIP or probable UIP are associated with increased mortality.[25]

As with diagnosis in established fibrotic lung disease, current ILA classification is based on visually-defined morphology, not disease behaviour. However, unlike in established fibrotic lung disease, where CT pattern definitions are agreed upon, the current ILA definition is an umbrella term that captures any CT pattern which exceeds 5% in any lung zone (figure 3).[25] If progressive ILA subgroups are to be identified, this definition will undoubtedly require refinement which might not be straightforward for several reasons. First, ILA categories that are too rigorously defined might lead to prognostic imprecision, given our current incomplete understanding

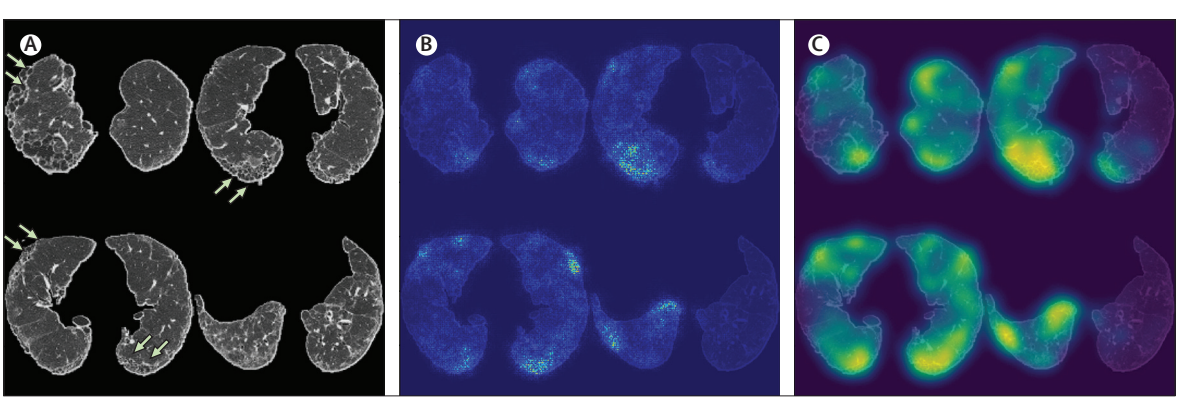

Figure 2: **Saliency mapping**
(A) Four slice axial montage generated from a patient with usual interstitial pneumonia, depicting honeycombing (green arrows) in the apical segment of the right lower lobe (upper right image slice), apical segment of the left lower lobe (lower left image slice), and in the right upper lobe (upper and lower left images.
(B) Saliency map generated by a deep learning algorithm, Systematic Objective Fibrotic Imaging analysis Algorithm (known as SOFIA),[9] highlighting pixels from figure 2a leading to a diagnosis of usual interstitial pneumonia (UIP; outputted probabilities: UIP=0·989, probable UIP=0·003, indeterminate=0·002, and alternative diagnosis= 0·006). The map shows that regions of honeycombing in the lower lobes (depicted as hotspots) contributed most to the algorithm's diagnosis.
(C) Saliency map of figure 2b following the application of a Gaussian smoothing filter to reduce image noise.

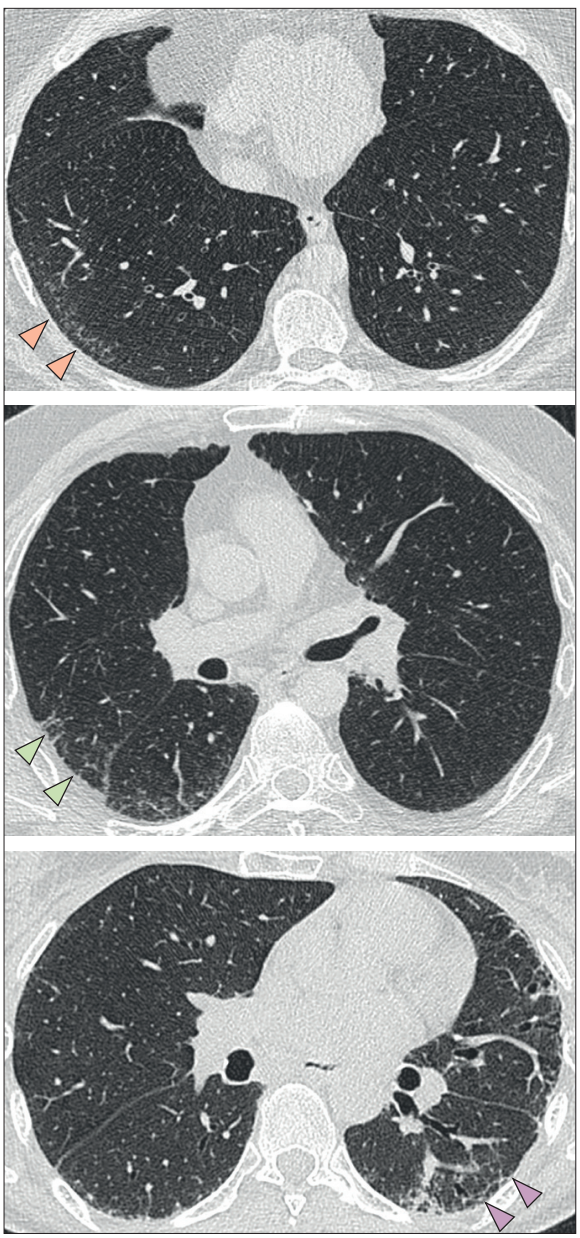

***Figure 3*: CT images showing different subpleural interstitial lung abnormalities**

Interstitial lung abnormalities (ILAs) encompasses a range of CT abnormalities and is defined as non-dependent changes affecting more than 5% of any lung zone, including reticular or ground-glass abnormalities, diffuse centrilobular nodularity, nonemphysematous cysts, honeycombing, or traction bronchiectasis.[25] Determining which of these abnormalities will develop into a progressive fibrotic lung disease remains a challenge to managing patients with incidentally detected ILAs. Ground glass opacity (red arrows), reticular opacity (green arrows), and reticular opacity with traction bronchiectasis (pink arrows).

of ILAs.[33] ILA distribution in three dimensions might have major prognostic significance but can be particularly challenging to classify visually when disease is limited. Similarly, the subjective nature of pattern recognition on CT might be amplified when the abnormalities are sparse or poorly defined. Perhaps most importantly, even if these difficulties can be overcome, ILA categorisation based on vision-based CT patterns are unlikely to precisely map to disease behaviour.

A solution to this problem might be found in deep learning-based analysis; algorithmic training could be anchored to ILA behaviour with no a priori assumptions as to the importance of individual ILA patterns. However, to achieve this, the difficulty of powering deep learning-based analyses of ILA will need to be overcome. This difficulty does not exist in established fibrotic lung disease; in one recent IPF cohort, the prevalence of progressors, defined as a 10% forced vital capacity (FVC) decline within 1 year, was 50%.[34] However, the prevalence of progressive ILAs is far lower, recently reported as 7·3% of patients in the AGES-Reykjavik study,[35] meaning that cohort enrichment, possibly using a serum biomarker, will be needed for effective algorithm training. Data from the Framingham Heart Study cohort, reporting associations between elevated serum galectin-3 concentrations, the likelihood of ILAs, and restrictive pulmonary physiology provide a conceptual basis for how this enrichment might be achieved.[36]

Predicting progressive fibrotic lung disease

The so-called progressive fibrotic phenotype is a collective term for patients with inexorably progressive disease despite conventional therapy for their specific ILD, irrespective of their clinical or histospecific diagnosis.[37–41] The progressive fibrotic phenotype was the focus of an expert group perspective as well as several published positive clinical trials of antifibrotic therapy in patients with non-IPF fibrotic lung disease, including INBUILD.[3,42–44] A challenge to translating the progressive fibrotic phenotype concept to individual patients is that there is no reliable means for predicting progression of fibrotic disease using baseline data. The early identification of the progressive fibrotic phenotype would allow clinicians to initiate therapies to slow or prevent progression at the earliest opportunity, without the need to delay intervention until progression has been clinically observed. Equally importantly, treatment could be withheld with confidence in inherently stable disease. The lack of reliable discriminatory data at baseline is possibly the most urgent unmet challenge for effective management for patients with progressive fibrotic lung disease.[34,45,46]

A major obstacle to addressing this challenge is that the existing data to identify the likely histospecific diagnosis is not sufficiently adaptable to the accurate prediction of future disease behaviour. Broad separations can be made between IPF and non-IPF disorders, and between UIP-like and non-UIP-like disease, but these cohort distinctions are imprecise. Patients with IPF have a variable clinical course with some progressing slowly while others experience rapid decline.[47,48] Some non-IPF patients with UIP patterns on CT have a good long-term outcome; in a study of rheumatoid arthritis related

fibrotic lung disease, almost 30% of patients with radiological UIP had a favourable prognosis.[49] In non-IPF patients without UIP-like appearances on CT, there are no data available that reliably predict outcome (figures 4a–b). Several studies have reported that more extensive fibrosis on CT at baseline is associated with increased risk of progression, but this might simply reflect the fact that disease has repeatedly progressed in the past and is, therefore, more likely to progress in future.[5,50,51] The identification of the progressive fibrotic phenotype is especially problematic when baseline disease extent is less severe.

From an imaging perspective, this challenge can be viewed as a pattern recognition problem amenable to deep learning-based analysis because it allows the deep learning training process to be anchored to an objective outcome. In our previous example, each CT scan in the training data was labelled subjectively as being associated with histological UIP, probable UIP, indeterminate findings, or features indicative of an alternative diagnosis. Alternatively, CT scans for training could be examined against future disease behaviour in individual patients, such as FVC decline at a predefined time point (figure 1). This approach has two distinct advantages. First, it bypasses subjective visual CT assessment entirely. Second, because there is no need to make a priori assumptions on imaging patterns likely to predict progressive disease, this approach allows for the possibility that CT patterns or features, currently unknown or imperceptible visually, could be discovered. In principle, novel CT phenotypes which allow stratification of patients with the same histospecific diagnosis, into different outcome-based groups could be developed.

The use of a deep learning-approach to CT classification at the voxel level in the quantification of specific patterns such as honeycombing or ground-glass opacities is possible—ie, deep learning-based quantitative CT. An advance in the field, termed Data-driven Textural Analysis, combines an unsupervised approach based on cluster analysis of voxel data and a convolution neural network to detect and quantify fibrosis on CT (figure 5). This feature learning method bypasses the need for feature engineering and might provide new opportunities for monitoring disease progression and therapeutic response using deep learning-based quantitative CT tools, trained to quantify changes in specific CT patterns over time.[52,53] A challenge to training and testing algorithms for disease monitoring is that legacy imaging repositories often do not contain longitudinal imaging datasets. Historically, visual assessment has been reported as relatively insensitive to subtle changes on CT and this has exacerbated the problem of insufficient imaging data for training; routine follow up imaging has not been recommended.[54]

It should be understood that the accuracy of optimal deep-learning algorithms might be increased by integration with non-CT variables. Combining variables from different domains to create multidimensional

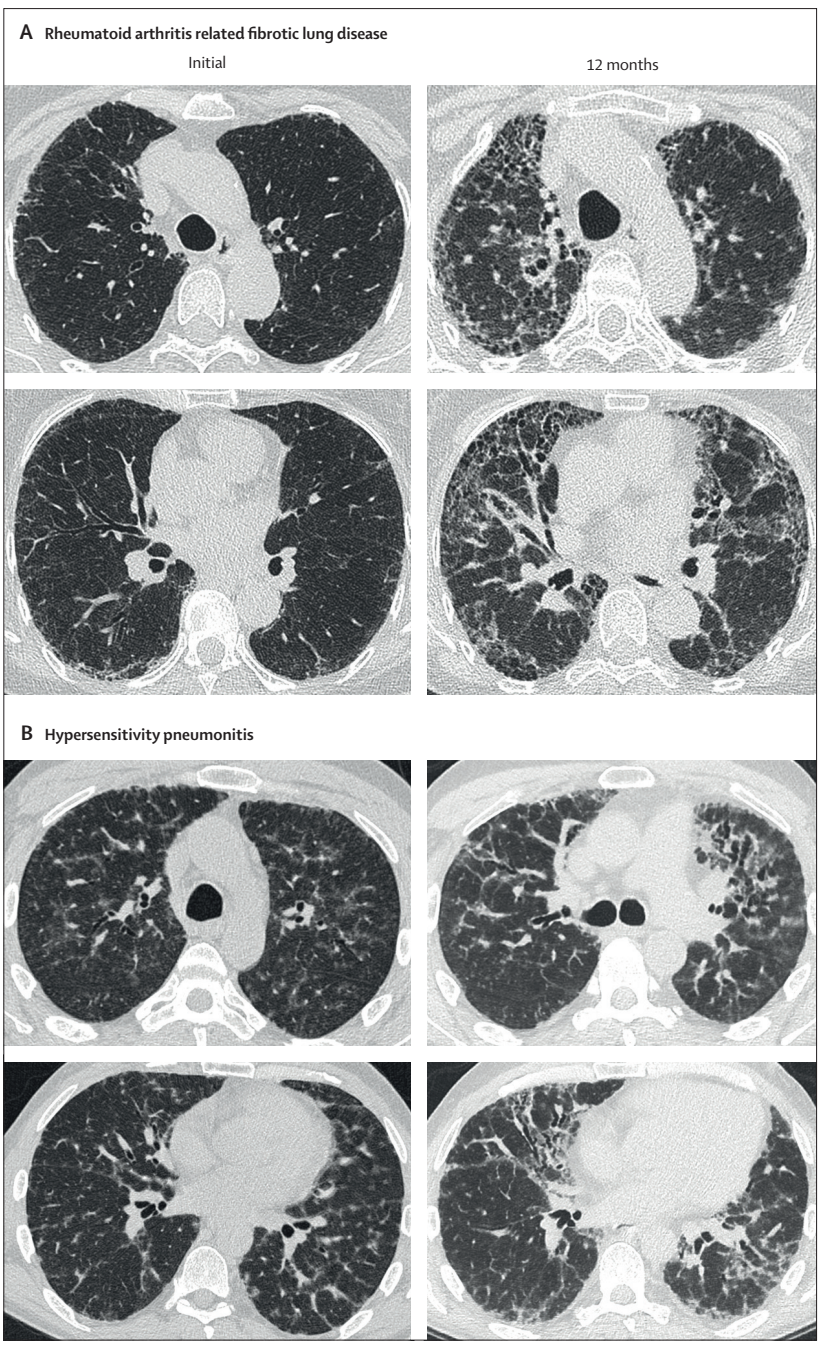

Figure 4: **Non-idiopathic pulmonary fibrosis progressive fibrotic lung disease**
(A) Axial CT images taken of a man aged 64 years man with rheumatoid arthritis related fibrotic lung disease. There is subpleural reticulation with areas of limited traction bronchiectasis but no convincing evidence of honeycombing. Axial CT images taken from the same patient 12 months later showing marked progression of fibrosis with subpleural honeycombing and severe traction bronchiectasis. Appearances now meet guideline criteria for usual interstitial pneumonia (UIP). (B) Axial CT images taken from a man aged 56 years with a multidisciplinary diagnosis of hypersensitivity pneumonitis. There is peribronchovascular reticulation and ground-glass opacities with relative sparing of the subpleural lung. There are no definitive fibrotic features. Axial CT images taken from the same patient 12 months later showing rapid progression of disease with severe traction bronchiectasis in the left upper lobe. These images show the progressive fibrotic phenotype in a patient with hypersensitivity pneumonitis and non-UIP fibrosis on CT.

staging algorithms in ILD has proven more fruitful than focusing on the stand alone value of variables in isolation.[55–58] Suggesting that the final step in developing deep learning-based technology for baseline prediction of disease behaviour might be the amalgamation of deep-learning algorithmic outputs with other omic-based biomarkers to enhance prognostic evaluation.[34,46]

For more on **Tensorflow** see https://www.tensorflow.org

Facilitating deep learning research in fibrotic lung disease

For more on **Kaggle** see https://www.kaggle.com/

How this technology will affect future clinical practice and drug development in progressive fibrotic lung diseases is worth considering. For this to happen, two major obstacles must be overcome. First, collaboration between researchers in this field has historically been relatively limited. This problem is highlighted by the fact that despite almost 20 years of traditional quantitative CT research, there are no studies which compare the different quantitative CT tools that are available using common imaging datasets; competition between developers and a desire to protect intellectual property might be at the root of this shortcoming. In contrast, open-source research promotes free access to machine learning software which might then be modified and customised by anyone. This collaborative approach to development facilitates a more diverse design perspective and ultimately, more stable machine learning tools are produced. GitHub, an open-source repository which hosts the code for millions of projects including machine learning software, epitomises this open-source ethos. In addition to allowing public access to these projects, GitHub archives past versions of source code so that researchers can follow the development of a piece of software, making it an invaluable teaching tool. Google's Tensorflow, a machine learning library of software at the core of its AI research and production pipelines is one such open-source project and has been applied to several recent machine learning-based imaging studies in medicine.[9,11,12]

For more on **GitHub** see https://github.com/

In addition to data and software sharing, open-science competitions are a key component of the machine learning development process. Kaggle is an online platform for machine learning and data science contests, which promotes collaboration between industry and academia by hosting data and challenging users to solve well-defined machine learning problems. These competitions are typically diverse in scope, ranging from medical applications such as lung nodule detection on CT to passenger screening at airport security. The power of this crowd-sourcing approach stems from the fact that often a multitude of different algorithmic solutions exist to any single machine learning problem. By engaging thousands of participants all working on the same challenge and using the same datasets, the most accurate algorithmic solution is found more efficiently. Importantly, once a contest is over, participants share their methodology and software for the benefit of the machine learning community.

The second obstacle is that currently no large fibrotic lung disease specific-imaging datasets for machine learning research exist. Two medical imaging studies, for diabetic retinopathy and for the classification of skin lesions as benign or malignant, involving deep learning demonstrate the scale of the problem; in each study more than 100 000 images were available for algorithm development.[9–12] By contrast, even the largest referral centres for fibrotic lung disease do not have access to imaging datasets of this order of magnitude, making a collaborative effort to develop a centralised imaging and clinical data biorepository for these disorders crucial. An additional, important role of this repository would be to provide a single, diverse dataset for computer algorithm generalisability to be evaluated and different algorithms to be compared. This is of relevance in fibrotic lung disease, where algorithms that are deployed in clinical practice will be expected to cope with the diversity of CT imaging protocols and CT scanners that now exist.

In 2018, the Open Source Imaging Consortium (OSIC) was founded as a cooperative effort between academia, industry, and philanthropy to specifically address these challenges by generating and curating a diverse repository of CT chest images and clinical data from patients with fibrotic lung disease and facilitating cooperative research efforts between academia and industry through the development and execution of machine learning competitions focused on digital biomarker development in progressive fibrotic lung disease. Algorithms developed on this biorepository are made open source for anyone to develop and refine further. The incentive for members of

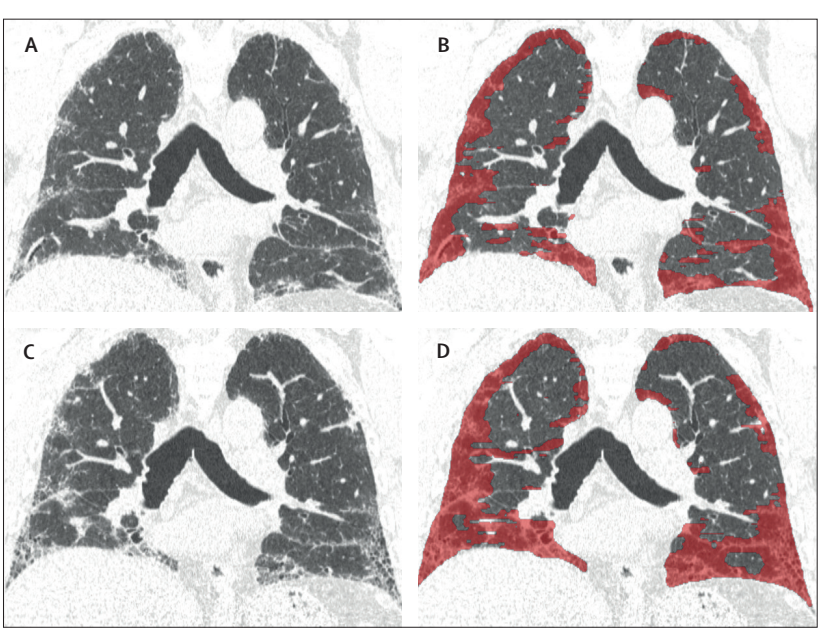

Figure 5: **Sequential quantitative CT fibrosis measurement in a patient with idiopathic pulmonary fibrosis**
(A) Baseline coronal image and (B) regions of fibrosis (red) classified by data-driven texture analysis on baseline CT. (C) Coronal image from 14 month follow up CT and (D) regions of fibrosis (red) classified by data-driven texture analysis at follow-up. The extent of fibrosis increased from 39% to 52%, and the patient deteriorated physiologically, with decrease in forced vital capacity from 66% predicted to 45%.

OSIC to share data comes from the access they are granted to much larger quantities of data for their own imaging research. Increasing available data, encouraging collaboration, and showcasing digital biomarker research challenges to new computer science participants will also help to accelerate research in this field.

Challenges to implementation

The implementation of deep learning-based technology in routine clinical practice will require substantial buy-in from health-care professionals and patients. As with any new drug or technology, this will require prospective clinical utility studies that test algorithm performance in real world clinical settings and show patient benefit over current best practice. Several unique challenges first need to be addressed. These challenges are both technical and societal.

First, the relative opacity of neural networks, upon which deep learning is based, is often viewed as a substantial weakness of this technology sometimes referred to as the black box phenomenon. The complexity that allows a neural network to identify patterns in large quantities of data can also obscure its reasoning. This relative obscurity might create barriers to implementation in medicine, particularly when algorithmic decision-making is based on features contained within the images which are invisible to human observers. The complexity of the deep learning process can also hamper algorithm development, which often relies on understanding why an algorithm misclassifies specific images. It should be noted that there are many black boxes in medicine–eg, the exact mechanism of action of some routinely used drugs. When a physician makes decisions based on intuition

Challenge	Potential solution
In established fibrotic lung disease, there is an urgent unmet need for biomarkers that can reliably predict progressive disease and response to treatment in an individual patient. Machine learning, particularly deep learning-based biomarker development, offers new opportunities to address this need.	Increased collaboration between clinical medicine and the artificial intelligence (AI) community must be encouraged and promoted. This might include the training of crossdisciplinary experts who can connect cutting edge biomedical imaging analysis techniques with important unanswered clinical questions. Large datasets are needed for successful deep learning algorithm training, which will also require collaboration between research groups. The development of multidisciplinary consortia focused on AI research in fibrotic lung disease such as The Open Source Imaging Consortium will accelerate innovation.
Interstitial lung abnormalities (ILA) represent a heterogeneous group of abnormalities on CT, which creates barriers to ILA research.	Consensus guidelines on ILA classification are needed to facilitate computer-based analysis of these lesions on CT.
Data heterogeneity remains a key issue. Medical imaging equipment, software, and procedures have been developed to enable visual interpretation of exams. Different acquisition and reconstruction protocols are selected to optimise visual assessment and, to some extent, are based on radiologist preference.	Consensus guidelines on imaging CT protocols optimised for computer-based analysis are needed including more access to source data from imaging devices.
Deep learning algorithms are complex input–output functions and operate like black boxes, meaning their complexity obscures their reasoning.	Making deep learning algorithms more interpretable is an area of ongoing research and will be essential if these technologies are to be successfully integrated into clinical practice.
For deep learning, it is currently not clear what the optimum approach to data labelling is in the setting of fibrotic lung disease.	At the current stage of technological development research might best be focused on quantifying interpretable variables such as the extent of fibrosis on a CT scan. In the future, as deep learning algorithm interpretability improves, algorithms trained on variables not derived from the images (such as physiological decline and genomic or proteomic signatures) might generate novel image-based biomarkers.
Buy in from patients and the public have been identified as key enablers of AI technology in health care.	Diagnostic and prognostic deep-learning algorithms need to be evaluated in the setting of multidisciplinary team meetings. Clinical utility studies that clearly demonstrate the benefit of incorporating algorithms' outputs into multidisciplinary decision making are needed. Robust and transparent data management will be essential to gain public and patient trust. This might involve establishing bespoke data governance frameworks. Gaps in regulatory and approval processes need to be addressed to accommodate emerging technologies. Active patient and public participation will be important. Citizens' juries, for example, can help to define the role AI will have in patient care. A strong narrative around the importance of data sharing for accelerating research needs to be developed and encouraged. Data governance frameworks must be a light touch to protect patient privacy while at the same time allowing innovation to flourish. Clarity around data ownership will be necessary, particularly when collaborating with industry.

Figure 6: **Challenges and potential solutions to deep learning imaging research in fibrotic lung disease**

backed by years of experience, this process is also to some extent, inscrutable. However, developing better methods for visualising novel imaging biomarkers generated by deep learning algorithms will be necessary to appraise their biological plausibility before they can be successfully integrated into clinical practice.

Second, a technical challenge to deep learning model training are the memory constraints associated with commercially available graphics processing units, which currently cannot accommodate training a deep neural network on a complete volumetric CT dataset. This means that down sampling (selecting a subset of the whole dataset that is more manageable in terms of computational load) of the CT data is required. In a 2018 study, attempts were made to overcome this problem by generating up to 500 unique 4-slice montages per CT scan.[9] Although this approach meant that most of the data contained within the volumetric CT was accessed by the algorithm, it also means that intricate or subtle relationships in the data between contiguous slices was almost certainly lost during the down sampling procedure. Furthermore, any process that results in sampling might introduce bias into the training data and therefore some argue that over processing of imaging pretraining is undesirable; training data should contain all of the noise encountered in routine clinical practice (such as different CT protocols and CT scanner models). In the aforementioned study, the algorithm was trained on data from two institutions using five different CT protocols.[9]

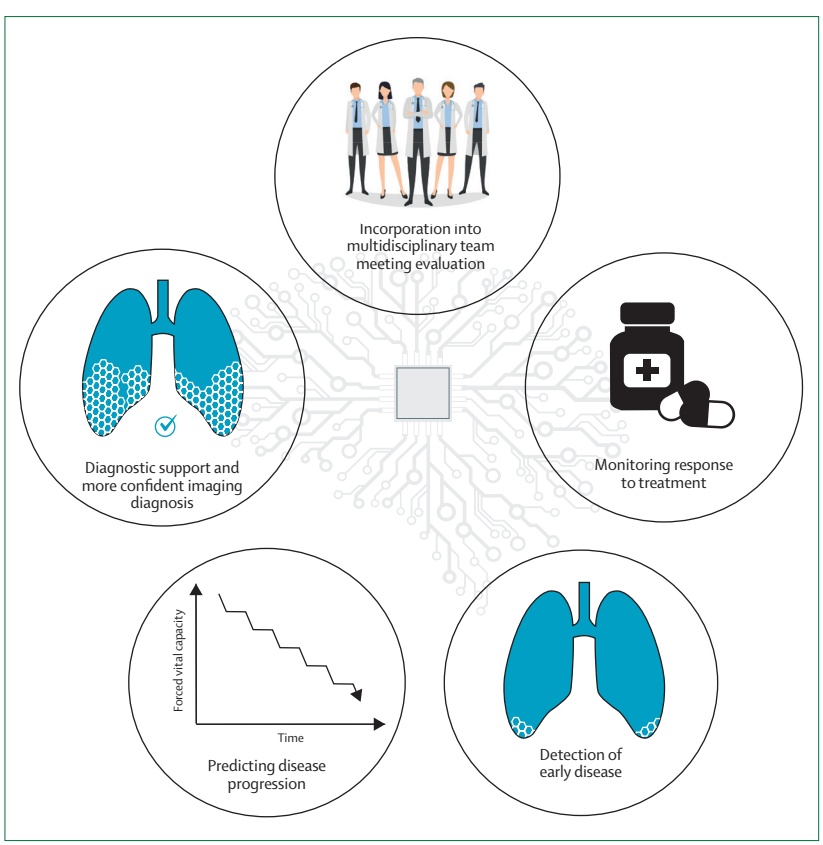

Figure 7: **Applications of deep learning to imaging research in fibrotic lung disease**

Deep learning algorithms also come with unique risks because of their potential to amplify biases in training data and their vulnerability to data manipulation. Missing or unbalanced data not only affect algorithm performance but also have the potential to reinforce a wide variety of disparities in healthcare in ways that are not immediately obvious. For example, in certain subgroups of patients, particularly in rare disorders, insufficient sample sizes might make it difficult to apply deep learning-based analysis techniques, therefore excluding these patients from improvements in clinical care that this technology might provide. Algorithm development can also be manipulated so that predictions are biased toward recommending specific actions, such as using an expensive test provided by a third-party provider.

In surveys, establishing an ethical framework to build and preserve trust and transparency has been identified as an important enabler of artificial intelligence implementation in healthcare. This will likely require bespoke data governance and accountability frameworks to address the unique challenges deep learning algorithm development presents. These frameworks will need to facilitate transparency in algorithm development and provide guidelines regarding data ownership, privacy and sharing. Balancing appropriate regulation at the same time as facilitating rapid innovation will also create challenges. The Verifiable Data Audit, a digital ledger that records all interactions with patient data including when and by whom data is accessed, is one suggested framework. Independent review panels such as the Topol Review have been commissioned by the UK government to advise on how artificial intelligence will change the roles of clinical staff in the future and what the implications of those changes will be.[59] Citizens' Juries, which promote patient and public involvement, have also been held to discuss how artificial intelligence technology should be incorporated into healthcare. Medical curricula and training will undoubtedly be effected and new subspecialisms that combine clinical medicine, bioinformatics, and computer science might develop. Finding ways to encode ethical standards in machine-learning systems will be essential to build trust (figure 6).

Future directions

The application of AI to healthcare is a rapidly evolving field requiring close collaboration between the AI community and clinical medicine. Deep learning is a powerful technology that can readily be applied to medical imaging classification; the challenge is identifying clinical problems that need solving and developing tools that are likely to have a real effect in clinical practice. At the current stage of technological development, deep learning research in fibrotic lung disease might best be focused on quantifying interpretable variables such as changes in the extent of fibrosis on a CT scan over time or in response to

Search strategy and selection criteria

References for this Personal View were identified through searches of PubMed and the computer science and engineering database, IEEE Xplore Digital Library on March 31, 2019, using the search terms, "machine learning", "deep learning", "idiopathic pulmonary fibrosis", "fibrotic lung disease", "quantitative computed tomography", "prognosis", "interstitial lung abnormalities", "progressive fibrotic phenotype", "diagnostic accuracy", "diagnostic guidelines", and "misdiagnosis" for articles published in English between Jan 1, 2011, and March 31, 2019. Following peer review, these searches were extended to November 30, 2019.

therapy. As algorithm interpretability improves, deep learning could be used to develop novel image-based biomarkers which are imperceptible visually. Computer-generated radiological phenotypes, responsive to specific therapies, might also be identified, facilitating precision medicine. Following this, diagnostic and prognostic algorithms for use in non-fibrotic ILD might be the next step. Most importantly, it must be established that incorporating these algorithms into clinical decision-making during multidisciplinary team discussions translates to improved patient outcomes (figure 7).

For this progress to occur the mismatch that exists between current technology and the availability of large fibrotic lung disease specific imaging datasets for deep learning research represents one of the biggest obstacles to future progress. This mismatch is particularly true in the case of ILA where the relatively low prevalence of progressive lesions will require massive imaging datasets to power the analysis. This challenge will only be overcome through the amalgamation of imaging datasets from different centres. The incentive to share data will come from the mutual benefit collaborators gain from access to the resulting large quantity of combined data.

Finally, replacing radiologists with this technology in the near future seems unlikely. Advances in computational image analysis in the present era of deep learning are more likely to enhance the understanding of, and the ability to interpret, CT abnormalities related to fibrotic lung disease. As Gary Kasparov suggested following his defeat to the IBM chess computer Deep Blue, the best outcomes will probably come from humans and machines working together. The clear identification of areas in which AI research in medicine should concentrate its efforts and the integration of AI algorithms with data in other domains will be vital if we are to maximise the benefits of this synergistic relationship.

Contributors
SLFW had the perspective concept and wrote the manuscript. All authors edited and approved the manuscript.

Declaration of interests
SLFW reports personal fees from Sanofi-Aventis, Roche, Boehringer Ingelheim, Galapagos, Open Source Imaging Consortium, and Bracco; and grants from Boehringer Ingelheim and National Institute for Health and Research, outside of the submitted work. SMH reports grants from National Heart, Lung, and Blood Institute, personal fees from Boehringer Ingelheim, outside of the submitted work. SMH has a patent systems and methods for automatic detection and quantification of pathology using dynamic feature classification pending to National Jewish Health. KKB reports grants from National Institutes of Health, and advisory board participation for Biogen, Blade, Boehringer Ingelheim, Galapagos, Galecto, Genoa, Lifemax, MedImmune, monARC Bionetworks, Open Source Imaging Consortium, Pliant, ProMetic, Third Pole, Theravance, Three Lakes Partners, and Veracyte, outside of the submitted work. AUW declares no competing interests.

References

1 Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014; **370:** 2071–82.
2 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:** 2083–92.
3 Flaherty KR, Brown KK, Wells AU, et al. Design of the PF-ILD trial: a double-blind, randomised, placebo-controlled phase III trial of nintedanib in patients with progressive fibrosing interstitial lung disease. *BMJ Open Respir Res* 2017; **4:** e000212.
4 Distler O, Highland KB, Gahlemann M, et al. Nintedanib for systemic sclerosis-associated interstitial lung disease. *N Engl J Med* 2019; **380:** 2518–28.
5 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:** 488–93.
6 Watadani T, Sakai F, Johkoh T, et al. Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology* 2013; **266:** 936–44.
7 Saketkoo LA, Mittoo S, Huscher D, et al. Connective tissue disease related interstitial lung diseases and idiopathic pulmonary fibrosis: provisional core sets of domains and instruments for use in clinical trials. *Thorax* 2014; **69:** 428–36.
8 Walsh SL, Calandriello L, Sverzellati N, Wells AU, Hansell DM, Consort UIPO. Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax* 2016; **71:** 45–51.
9 Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018; **6:** 837–45.
10 De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018; **24:** 1342–50.
11 Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; **316:** 2402–10.
12 Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542:** 115–18.
13 Jacob J, Bartholmai BJ, Rajagopalan S, et al. Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures. *Eur Respir J* 2017; **49:** 1601011.
14 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:** 921–29.
15 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.
16 Bak SH, Park HY, Nam JH, et al. Predicting clinical outcome with phenotypic clusters using quantitative CT fibrosis and emphysema features in patients with idiopathic pulmonary fibrosis. *PLoS One* 2019; **14:** e0215303.
17 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.

18 Jacob J, Bartholmai BJ, Rajagopalan S, et al. Mortality prediction in IPF: evaluation of automated computer tomographic analysis with conventional severity measures. *Eur Respir J* 2016.

19 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521:** 436–44.

20 González G, Ash SY, Vegas-Sánchez-Ferrero G, et al. Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med* 2018; **197:** 193–203.

21 Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; **25:** 954–61.

22 Wang S, Shi J, Ye Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J* 2019; **53:** 1800986.

23 Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of idiopathic pulmonary fibrosis. an official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2018; **198:** e44–68.

24 Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. *Proc IEEE Conf on Comput Vis and Pattern Recognit* 2016. **2016:** 2921–29.

25 Putman RK, Hatabu H, Araki T, et al. association between interstitial lung abnormalities and all-cause mortality. *JAMA* 2016; **315:** 672–81.

26 Hobbs BD, Putman RK, Araki T, et al. Overlap of genetic risk between interstitial lung abnormalities and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2019; **200:** 1402–13.

27 Washko GR, Hunninghake GM, Fernandez IE, et al. Lung volumes and emphysema in smokers with interstitial lung abnormalities. *N Engl J Med* 2011; **364:** 897–906.

28 Sverzellati N, Guerci L, Randi G, et al. Interstitial lung diseases in a lung cancer screening trial. *Eur Respir J* 2011; **38:** 392–400.

29 Hunninghake GM, Hatabu H, Okajima Y, et al. MUC5B promoter polymorphism and interstitial lung abnormalities. *N Engl J Med* 2013; **368:** 2192–200.

30 Lederer DJ, Enright PL, Kawut SM, et al. Cigarette smoking is associated with subclinical parenchymal lung disease: the Multi-Ethnic Study of Atherosclerosis (MESA)-lung study. *Am J Respir Crit Care Med* 2009; **180:** 407–14.

31 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:** 563–71.

32 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:** 1514–22.

33 Walsh SL, Hansell DM. Diffuse interstitial lung disease: overlaps and uncertainties. *Eur Radiol* 2010; **20:** 1859–67.

34 Maher TM, Oballa E, Simpson JK, et al. An epithelial biomarker signature for idiopathic pulmonary fibrosis: an analysis from the multicentre PROFILE cohort study. *Lancet Respir Med* 2017; **5:** 946–55.

35 Putman RK, Gudmundsson G, Axelsson GT, et al. Imaging patterns are associated with interstitial lung abnormality progression and mortality. *Am J Respir Crit Care Med* 2019; **200:** 175–83.

36 Ho JE, Gao W, Levy D, et al. Galectin-3 Is Associated with restrictive lung disease and interstitial lung abnormalities. *Am J Respir Crit Care Med* 2016; **194:** 77–83.

37 Jegal Y, Kim DS, Shim TS, et al. Physiology is a stronger predictor of survival than pathology in fibrotic interstitial pneumonia. *Am J Respir Crit Care Med* 2005; **171:** 639–44.

38 Latsi PI, du Bois RM, Nicholson AG, et al. Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. *Am J Respir Crit Care Med* 2003; **168:** 531–37.

39 Gimenez A, Storrer K, Kuranishi L, Soares MR, Ferreira RG, Pereira CAC. Change in FVC and survival in chronic fibrotic hypersensitivity pneumonitis. *Thorax* 2018; **73:** 391–92.

40 Solomon JJ, Chung JH, Cosgrove GP, et al. Predictors of mortality in rheumatoid arthritis-associated interstitial lung disease. *Eur Respir J* 2016; **47:** 588–96.

41 Goh NS, Hoyles RK, Denton CP, et al. Short-term pulmonary function trends are predictive of mortality in interstitial lung disease associated with systemic sclerosis. *Arthritis Rheumatol* 2017; **69:** 1670–78.

42 Wells AU, Brown KK, Flaherty KR, Kolb M, Thannickal VJ, Group IPFCW. What's in a name? That which we call IPF, by any other name would act the same. *Eur Respir J* 2018; **51:** 1800692.

43 Flaherty KR, Wells AU, Cottin V, et al. Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med* 2019; **381:** 1718–27.

44 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* 2019; published online Sept 2019. **8:** 147–57.

45 Ley B, Brown KK, Collard HR. Molecular biomarkers in idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* 2014; **307:** L681–91.

46 Maher TM, Stowasser S, Nishioka Y, et al. Biomarkers of extracellular matrix turnover in patients with idiopathic pulmonary fibrosis given nintedanib (INMARK study): a randomised, placebo-controlled study. *Lancet Respir Med* 2019; **7:** 771–79.

47 Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. *Lancet* 2017; **389:** 1941–52.

48 Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011; **183:** 431–40.

49 Jacob J, Hirani N, van Moorsel CHM, et al. Predicting outcomes in rheumatoid arthritis related interstitial lung disease. *Eur Respir J* 2019; **53:** 1800869.

50 Walsh SL, Sverzellati N, Devaraj A, Keir GJ, Wells AU, Hansell DM. Connective tissue disease related fibrotic lung disease: high resolution computed tomographic and pulmonary function indices as prognostic determinants. *Thorax* 2014; **69:** 216–22.

51 Walsh SL, Sverzellati N, Devaraj A, Wells AU, Hansell DM. Chronic hypersensitivity pneumonitis: high resolution computed tomography patterns and pulmonary function indices as prognostic determinants. *Eur Radiol* 2012; **22:** 1672–79.

52 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.

53 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:** 270–78.

54 Flaherty KR, Mumford JA, Murray S, et al. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med* 2003; **168:** 543–48.

55 Goh NS, Desai SR, Veeraraghavan S, et al. Interstitial lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med* 2008; **177:** 1248–54.

56 Walsh SL, Wells AU, Sverzellati N, et al. An integrated clinicoradiological staging system for pulmonary sarcoidosis: a case-cohort study. *Lancet Respir Med* 2014; **2:** 123–30.

57 Ley B, Elicker BM, Hartman TE, et al. Idiopathic pulmonary fibrosis: CT and risk of death. *Radiology* 2014; **273:** 570–79.

58 Ley B, Ryerson CJ, Vittinghoff E, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* 2012; **156:** 684–91.

59 NHS Health Education England. The Topol Review. Preparing the healthcare workforce to deliver the digital future. 2019. https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf (accessed Jan 15, 2020).