Disentangling Computed Tomography Pattern and Extent to Estimate Prognosis in Fibrosing Interstitial Lung Diseases

Clinicians generally have two objectives when evaluating patients with interstitial lung disease (ILD). The first is to assimilate clinical and phenotypic information to confer a singular ILD diagnosis, which broadly informs medical management and estimated prognosis. The second objective is to identify personal factors that necessitate alterations to the standard diagnosis-based management approach. Several studies have demonstrated that information gained from computed tomography (CT) can advance both ILD evaluation goals. The radiographic ILD pattern correlates with the histologic pattern and is clearly relevant to diagnostic categorization (1–3). In addition, specific CT features, such as radiographic honeycombing, can refine the individual prognosis within diagnostic subtypes (4). However, challenges with human interpretation of CT scans are well documented and relate to worldwide shortages of skilled readers, lengthy reads required to evaluate disease severity/extent, and the often limited agreement attained by expert readers. Recently, researchers have sought to overcome these limitations by devising automated systems to reduce variability and improve reproducibility. Some of these algorithms have focused on quantifying the extent of involved lung parenchyma as a surrogate for disease severity (5, 6), although others have focused on automating pattern recognition to improve ILD classification (7). Both approaches offer an opportunity to streamline ILD diagnostic evaluation while still informing downstream management.

In this issue of the *Journal*, Humphries and colleagues (pp. 1121–1131) developed and applied a deep learning algorithm using multi-instance learning (MIL) of ILD CT scans to estimate the probability of a histologic usual interstitial pneumonitis (UIP) pattern (8). Their algorithm was trained on a combination of ILD and non-ILD CTs using histopathologic UIP or confident radiographic UIP categorization as ground truth. The algorithm was tested and then validated on two independent cohorts with mixed ILD diagnoses. As a continuous marker, the MIL-UIP algorithm demonstrated superior discriminatory power for predicting histologic UIP in both validation cohorts (area under the receiver operating characteristic curve, 0.77 and 0.79) compared with visual CT assessment alone (area under the receiver operating characteristic curve, 0.65 and 0.71). Moreover, after dichotomizing the MIL-UIP score at the 0.5 threshold, the categorical score produced a higher sensitivity (0.70) than visual CT assessment (0.61) but comparable specificity (MIL-UIP 0.73 and visual CT assessment 0.76) for histologic UIP in the Chicago cohort.

The authors also assessed the prognostic ability of MIL-UIP in the mixed ILD validation cohorts. They found that the MIL-UIP algorithm was associated with differential survival even after accounting for disease extent (by data driven texture analysis) and visual assessment of the radiographic UIP pattern. However, when evaluating prognostic radiographic features in fibrosing ILD, radiographic disease extent may provide complementary information to disease patterns (e.g., UIP pattern [9]) or the presence of particular features (e.g., traction bronchiectasis or honeycombing [10]). For example, patients with extensive reticulation and overlaid ground-glass densities may have a similar prognosis to patients with a limited extent of disease if, for example, honeycombing coexists. Therefore, the authors assessed the interaction between disease extent (by data driven texture analysis) and disease pattern (by MIL-UIP probability) on mortality and rate of lung function decline. Their results suggest that higher MIL-UIP prediction scores were more prognostically useful at lesser extents of disease, whereas in cases with more extensive disease, prediction of UIP was no longer independently associated with mortality. Similar results were also found when assessing the association of MIL-UIP with survival across gender-age-physiology stages and by evaluating lung function trajectory.

Therefore, quantifying both disease extent and probability of UIP pattern may help identify patients at risk for progression despite their limited extent of disease, a conundrum all too familiar to radiologists evaluating lung cancer screening cohorts.

A curious finding in the study was that a visually assessed UIP pattern did not independently predict mortality. The authors dichotomized radiographic UIP patterns by grouping definite and probable UIP together and compared them to subjects classified as either indeterminate for UIP or an alternative diagnosis pattern. For patients with probable or definite UIP, the MIL-UIP score did not associate with mortality when adjusted for CT disease extent. As these subjects are likely to have had more extensive disease, UIP prediction may have less prognostic impact. Instead, it was in CTs visually classified as indeterminate for UIP or alternative diagnosis to UIP for which the MIL-UIP score was most useful. These findings imply a potential clinical workflow whereby CTs could be triaged using human reads of UIP pattern. Then, rather than using the radiologic UIP classification alone to indicate survival, patients in whom algorithmic UIP probability prediction would be most insightful could be prioritized.

The analyses also presuppose that a radiographic UIP pattern is associated with worse survival than non-UIP CT patterns across all ILDs. The fact that these comparisons did not significantly associate with mortality highlights a limitation of the existing UIP criteria (2), namely that a radiologic UIP pattern is by default a description of an idiopathic pulmonary fibrosis (IPF)-UIP pattern. Transposing an IPF-UIP pattern, developed to aid diagnostic classification for patients suspected of having IPF, to prognosticate non-IPF ILDs can be problematic. Non-IPF ILDs often do not display basilar or subpleural predominant traction bronchiectasis or honeycombing. Indeed,

Supported by Wellcome Trust grant 209553/Z/17/Z (J.J.), the NIHR University College London Hospitals Biomedical Research Centre (J.J.), and NHLBI grant K23HL149498 (C.A.N.).

Originally Published in Press as DOI: 10.1164/rccm.202401-0117ED on February 8, 2024

regardless of its location, honeycombing identifies patients with non-IPF ILD whose disease behaves similarly to patients with IPF (4, 11, 12). It is not difficult to picture a subject with rheumatoid arthritis–related ILD and extensive upper zone honeycombing being classified as indeterminate for UIP using IPF-UIP criteria. Yet, their survival may be similar to a patient with rheumatoid arthritis–related ILD with a classical IPF-UIP pattern. Although the application of radiologic IPF-UIP criteria to patients with non-IPF ILD may lack the same prognostic impact if used in isolation, the current study shows how IPF-UIP criteria could be usefully applied to subjects with non-IPF ILD.

The study also interrogated the deep learning algorithm using heatmaps to localize the CT region contributing most strongly to the algorithmic prediction. When these were examined visually, counterintuitively, it was the peripheral anterior lung in the middle zones that most frequently contributed to a UIP classification. Furthermore, rather than honeycomb cysts or gross traction bronchiectasis, it was subpleural irregularities and/or reticulation (often limited in extent) that most frequently associated with a UIP classification. Although these might suggest reevaluation of existing paradigms for identifying UIP visually, the results need to be interpreted with caution. Heatmaps can be unreliable and localize seemingly unrelated structures as contributing to prediction by an algorithm (13). The nonspecific nature of heatmap outputs is visible in Figure 3B (8), where the chest wall is simultaneously highlighted together with parenchymal damage as contributing toward UIP prediction.

We commend Humphries and colleagues (8) for developing an algorithm that estimates UIP probability and for disentangling the prognostic impact of both disease extent and ILD pattern. Once validated and implemented, it is easy to envision that automated CT quantification could provide valuable information for subsets of patients with fibrotic ILD who often fall between diagnostic categories yet are at risk for progressive disease. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Joseph Jacob, F.R.C.R., M.D. (Res)
Centre for Medical Image Computing
and
Department of Respiratory Medicine
University College London
London, United Kingdom

Chad A. Newton, M.D.
Division of Pulmonary and Critical Care Medicine
University of Texas Southwestern Medical Center
Dallas, Texas

ORCID IDs: 0000-0002-8054-2293 (J.J.); 0000-0001-5256-9029 (C.A.N.).

References

1. Chung JH, Chawla A, Peljto AL, Cool CD, Groshong SD, Talbert JL, *et al.* CT scan findings of probable usual interstitial pneumonitis have a high predictive value for histologic usual interstitial pneumonitis. *Chest* 2015; 147:450–459.
2. Raghu G, Remy-Jardin M, Richeldi L, Thomson CC, Inoue Y, Johkoh T, *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.
3. Raghu G, Remy-Jardin M, Ryerson CJ, Myers JL, Kreuter M, Vasakova M, *et al.* Diagnosis of hypersensitivity pneumonitis in adults: an official ATS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2020; 202:e36–e69.
4. Adegunsoye A, Oldham JM, Bellam SK, Montner S, Churpek MM, Noth I, *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.
5. Humphries SM, Yagihashi K, Huckleberry J, Rho BH, Schroeder JD, Strand M, *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.
6. Zavaletta VA, Bartholmai BJ, Robb RA. High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. *Acad Radiol* 2007;14:772–787.
7. 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: 837–845.
8. Humphries SM, Thieke D, Baraghoshi D, Strand MJ, Swigris JJ, Chae KJ, *et al.* Deep learning classification of usual interstitial pneumonia predicts outcomes. *Am J Respir Crit Care Med* 2024;209:1121–1131.
9. Walsh SLF, Mackintosh JA, Calandriello L, Silva M, Sverzellati N, Larici AR, *et al.* Deep learning-based outcome prediction in progressive fibrotic lung disease using high-resolution computed tomography. *Am J Respir Crit Care Med* 2022;206:883–891.
10. Jacob J, Bartholmai BJ, Rajagopalan S, Egashira R, Brun AL, Kokosi M, *et al.* Unclassifiable-interstitial lung disease: outcome prediction using CT and functional indices. *Respir Med* 2017;130:43–51.
11. Salisbury ML, Gu T, Murray S, Gross BH, Chughtai A, Sayyouh M, *et al.* Hypersensitivity pneumonitis: radiologic phenotypes are associated with distinct survival time and pulmonary function trajectory. *Chest* 2019;155:699–711.
12. Jacob J, Hirani N, van Moorsel CHM, Rajagopalan S, Murchison JT, van Es HW, *et al.* Predicting outcomes in rheumatoid arthritis related interstitial lung disease. *Eur Respir J* 2019;53:1800869.
13. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health* 2021;3:e745–e750.