When Deep Blue first defeated Kasparov: is a machine stronger than a radiologist at predicting prognosis in idiopathic pulmonary fibrosis?

Nicola Sverzellati¹ and Pierre-Yves Brillet²

Affiliations: ¹Radiology, Dept of Surgical Sciences, University of Parma, Parma, Italy. ²Université Paris 13, Sorbonne Paris Cité, EA2363 "Réponses cellulaires et fonctionnelles à l'hypoxie" Assistance Publique Hôpitaux de Paris (AP-HP), Dept of Radiology, Avicenne Hospital, Bobigny, France.

Correspondence: Nicola Sverzellati, Radiology, Dept of Surgical Sciences, University of Parma, V. Gramsci 14, 43126, Parma, Italy. E-mail: nicola.sverzellati@unipr.it

@ERSpublications

Quantitative CT data is prognostically more accurate than visual CT scoring in patients with IPF
<http://ow.ly/dNMB306gGoR>

Cite this article as: Sverzellati N, Brillet P-Y. When Deep Blue first defeated Kasparov: is a machine stronger than a radiologist at predicting prognosis in idiopathic pulmonary fibrosis? *Eur Respir J* 2017; 49: 1602144 [https://doi.org/10.1183/13993003.02144-2016].

Idiopathic pulmonary fibrosis (IPF) is a relentlessly progressive disease with a poor prognosis. Nevertheless, survival time varies substantially, with some patients experiencing an accelerated decline and a few living as long as 10 years after diagnosis [1]. Predicting the future for patients with IPF is challenging. Yet, this is of paramount importance given the clinical implications, such as initiation of treatment, inclusion in clinical trials, referral for lung transplantation or palliation.

Several multidimensional indexes have been developed to initially stratify the likely prognosis of patients with IPF. The composite physiologic index (CPI) is quite simple as it is made up of spirometric volumes and diffusing capacity of the lung for carbon monoxide (DLCO) [2]. CPI predicts mortality more accurately than individual functional indexes. The CPI is expressed as a continuous variable, which may represent a limitation for a comprehensive staging of IPF in routine clinical practice. On the other hand, the GAP (gender, age and physiology) index provides a simple three-category scoring system that has both research and clinical utility [3]. Intriguingly, a CT fibrosis score can replace the DLCO in the GAP model by maintaining a similar prognostic value [4]. Moreover, inclusion of serum biomarkers in risk prediction was shown to improve our ability to model important outcomes in IPF [5].

Chest CT has gained a central role in the diagnostic work-up of patients with suspected IPF since the publication of the American Thoracic Society/European Respiratory Society/Japanese Respiratory Society/Latin American Thoracic Association IPF guidelines in 2011 [6]. CT pattern is an important diagnostic entry criterion for clinical trials [7]. However, the substantial interobserver variability for both honeycombing and CT diagnostic categories (e.g. definite, possible, inconsistent), as well as the positive correlation between the likelihood of IPF and the extent of reticular opacities on CT (possibly admixed with traction bronchiectasis, even without honeycombing) in the correct clinical setting, underscore the need for accurate and objective quantification of individual IPF patterns [8, 9].

Received: Nov 02 2016 | Accepted: Nov 06 2016

Conflict of interest: Disclosures can be found alongside this article at ersjournals.com

Copyright ©ERS 2017

To date, visual scoring has been the most frequently applied method for assessing either global disease or extent of individual CT patterns. In addition, the severity of traction bronchiectasis by visual score has been shown to have a high prognostic value in patients with IPF [10]. However, interobserver variability represents the major limitation of the visual scoring.

Quantitative image analysis of interstitial lung disease is rapidly evolving, and has the potential to be useful in the assessment of disease extent in IPF in both clinical trials and clinical practice. Quantitative CT refers to the extraction and use of numerical/statistical features from CT images. By definition, quantitative CT is an objective analysis that may overcome the issue of the interobserver variability and, thus, it could potentially help to provide prognostic indexes more consistently. There are several quantitative CT systems of varying degrees of sophistication. The most widely available quantitative CT system relies on global histogram density measurements of CT images. Such measurements can also be obtained from software run on standard personal computers. In a large study cohort, the visual extent of lung fibrosis (*e.g.* reticular opacities and honeycombing) and kurtosis (*e.g.* weight of tails in the histogram) were the only independent predictors of mortality [11]. More sophisticated textural analyses have also been used to quantify the extent of individual patterns [12–14]. Parenchymal classification is applied to voxel volume unit (*e.g.* discrete volume that allows detailed characterisation of local parenchymal features) using texture analysis, computer vision-based image understanding of volumetric histogram signature mapping features, and 3D-morphology. Textural analysis is based on regions of interest selected by trained observers in the lung, according to a set of specific patterns (normal, reticular, honeycombing, *etc.*). The histogram or textural features of each volumetric region of interest are extracted, and a machine-learning algorithm is used to develop a predictive model for specific patterns [15]. This kind of software is still not commercially available on CT vendors' diagnostic workstations and such software requires high-resolution images, preferably reconstructed with parameters that reduce image noise. Such multidimensional analysis demands considerable computational power that usually requires a dedicated workstation outside of the clinical radiology workflow.

In this issue of the *European Respiratory Journal*, JACOB *et al.* [16] used CALIPER software (Computer-Aided Lung Informatics for Pathology Evaluation and Ratings), a CT image analysis tool based on the above described volumetric textural analysis, to quantify the extent of global disease, individual patterns (*e.g.* ground-glass opacification, reticular pattern and honeycombing), emphysema, and pulmonary vessels volume in a cohort of 283 consecutive patients with an IPF multidisciplinary diagnosis. The prognostic value of the CALIPER features was compared with standard visual scoring of CT features (also including traction bronchiectasis) and the CPI. The significant predictive value of the visual scoring for the fibrosing lung disease pattern (*e.g.* honeycombing, reticulation and traction bronchiectasis) on the univariate analysis is in keeping with prior studies and indeed increases consistency of findings from JACOB *et al.* [16]. However, importantly, visual scores were not retained in the multivariate model for prediction of mortality. Furthermore, two CALIPER features, namely honeycombing extent and pulmonary vessels volume, and the CPI were retained in the multivariate model for prediction of mortality. Thus, a key message of this study is that quantitative CT data is prognostically more accurate than visual CT scoring.

Given the well-known interobserver variability for the assessment of honeycombing, the development of an objective quantitative CT tool that can quantify honeycombing with prognostic value is of utmost importance. The novel pulmonary vessels volume feature cannot be visually scored and might represent a promising parameter in the evaluation of patients with IPF. Interestingly, it might add prognostic information to that provided by the presence of an enlarged pulmonary artery on CT [17]. However, the pulmonary vessels volume score quantifies the volumes of both intrapulmonary arteries and veins, thus biasing correlations with pulmonary artery disease in IPF. In fact, on linear regression analysis, a weak degree of colinearity was demonstrated between pulmonary vessels volume and right ventricular systolic pressure measured on echocardiography. Instead, colinearity with DLCO and total disease extent, either visually or CALIPER based, was substantially higher. JACOB *et al.* [16] make intriguing proposals about the positive pulmonary vessels volume signal in their discussion. However, the authors acknowledge that the exact pathophysiological mechanism that links interstitial lung disease extent to global vascular volume is not yet fully understood. Pulmonary vessels volume still needs clarification and further validation before it can be adopted as a reliable measure. In particular, data on its accuracy, robustness in relation to CT acquisition parameters and prognostic value according to fibrosis extent are needed.

In a sub-analysis, the CALIPER-CPI model slightly outperformed the CALIPER-GAP model (HR 2.23 (95% CI 1.85–2.69) *versus* HR 2.00 (95% CI 1.61–2.48)). The study results show that a combination of functional and quantitative CT parameters is a powerful noninvasive tool to predict outcome of patients with IPF. Furthermore, the CALIPER-CPI model and the CALIPER-only score were of similar prognostic value, suggesting that an automated scoring CT tool could be useful particularly when full functional assessment (*e.g.* DLCO) is not available. JACOB *et al.* [16] converted CALIPER-CPI and CALIPER-only scores into categorical scores, which is a step forward to increase clinical utility of the prognostic models. However, the

discriminative power of the models, in other words to what extent they may separate subjects with a poor outcome from those without, requires further clarification, ideally by validation in a large prospective cohort.

It seems that quantitative CT is at last on the brink of a true (rather than false) dawn, nevertheless there is much to be done before it becomes mainstream: the lack of comparative analysis between different quantitative CT software, sufficient validation and reproducibility of individual CT software results on multicentre cohorts, amongst other issues, need to be addressed. However, interest and support for quantitative CT is undeniably growing and is likely to be further piqued by emerging studies similar to the one by JACOB *et al.* [16]. The study by JACOB *et al.* [16] is the first study demonstrating the superiority of machine over radiologist to predict patient mortality. It deepens into new perspectives in both radiology and respiratory medicine. In a way, it looks like Deep Blue, the chess-playing computer, defeating Garry Kasparov for the first time...

References

- 1 Nathan SD, Shlobin OA, Weir N, *et al.* Long-term course and prognosis of idiopathic pulmonary fibrosis in the new millennium. *Chest* 2011; 140: 221–229.
- 2 Wells AU, Desai SR, Rubens MB, *et al.* Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *Am J Respir Crit Care Med* 2003; 167: 962–969.
- 3 Ley B, Ryerson CJ, Vittinghoff E, *et al.* A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* 2012; 156: 684–691.
- 4 Ley B, Elicker BM, Hartman TE, *et al.* Idiopathic pulmonary fibrosis: CT and risk of death. *Radiology* 2014; 273: 570–579.
- 5 Richards TJ, Kaminski N, Baribaud F, *et al.* Peripheral blood proteins predict mortality in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2012; 185: 67–76.
- 6 Raghu G, Collard HR, Egan JJ, *et al.* An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011; 183: 788–824.
- 7 Raghu G, Scholand MB, de Andrade J, *et al.* FG-3019 anti-connective tissue growth factor monoclonal antibody: results of an open-label clinical trial in idiopathic pulmonary fibrosis. *Eur Respir J* 2016; 47: 1481–1491.
- 8 Watadani T, Sakai F, Johkoh T, *et al.* Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology* 2013; 266: 936–944.
- 9 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.
- 10 Edey AJ, Devaraj AA, Barker RP, *et al.* Fibrotic idiopathic interstitial pneumonias: HRCT findings that predict mortality. *Eur Radiol* 2011; 21: 1586–1593.
- 11 Best AC, Meng J, Lynch AM, *et al.* Idiopathic pulmonary fibrosis: physiologic tests, quantitative CT indexes, and CT visual scores as predictors of mortality. *Radiology* 2008; 246: 935–940.
- 12 Uppaluri R, Hoffman EA, Sonka M, *et al.* Interstitial lung disease: a quantitative study using the adaptive multiple feature method. *Am J Respir Crit Care Med* 1999; 159: 519–525.
- 13 Bartholmai BJ, Raghunath S, Karwoski RA, *et al.* Quantitative computed tomography imaging of interstitial lung diseases. *J Thorac Imaging* 2013; 28: 298–307.
- 14 Yoon RG, Seo JB, Kim N, *et al.* Quantitative assessment of change in regional disease patterns on serial HRCT of fibrotic interstitial pneumonia with texture-based automated quantification system. *Eur Radiol* 2013; 23: 692–701.
- 15 Hansell DM, Goldin JG, King TE Jr, *et al.* CT staging and monitoring of fibrotic interstitial lung diseases in clinical practice and treatment trials: a position paper from the Fleischner Society. *Lancet Respir Med* 2015; 3: 483–496.
- 16 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.
- 17 Shin S, King CS, Puri N, *et al.* Pulmonary artery size as a predictor of outcomes in idiopathic pulmonary fibrosis. *Eur Respir J* 2016; 47: 1445–1451.