Artificial intelligence for pulmonary function test interpretation

To the Editor:

We read with interest the recent article by Topalovic *et al.* [1], in which they showed that artificial intelligence (AI) could interpret pulmonary function tests (PFTs) and reach a diagnosis with accuracy greater than individual pulmonologists, and approximately equal to that of an expert panel. Their work raises a number of important issues which should be further explored.

First, a number of caveats to their work should be acknowledged. PFTs, like any other investigation, can only be interpreted in the light of the pre-test probability of the diagnoses under consideration. The probabilities of different diagnoses suggested by the AI algorithm would have been heavily influenced by their prevalence within the training dataset. This may not have been fully representative of a population with undifferentiated symptoms encountered in clinical practice, since it was enriched with less common diagnoses such as neuromuscular disease and pulmonary vascular disease. Furthermore, the authors do not describe in detail how the 50 cases within the test dataset were chosen: were they truly random or to a certain extent hand-picked to ensure a broad mix of diagnoses? The AI algorithm would have been advantaged if the diagnostic prevalence in the test dataset broadly matched that of the training dataset.

The authors show that the AI algorithm could match an expert panel with respect to diagnostic ability, but that individual pulmonologists fell far short of this. While it may be over-optimistic to expect every individual pulmonologist to reach the level of an expert panel, the results do suggest that there is room for improvement in the training of pulmonologists in PFT interpretation. Patout *et al.* [2] recently reported that only 16% of French pulmonology trainees had attended a placement with a PFT laboratory, but that those who had attended such a placement performed significantly better on a written test of PFT knowledge. We consider that all pulmonology trainees should attend an accredited course on PFT interpretation and undertake a short attachment with a regional PFT laboratory.

Much of the focus of AI research has hitherto been on training AI algorithms to learn patterns from data in order to emulate or surpass human abilities. Often this has involved training the AI using human expert judgement as the gold standard (AI learning from humans). However, there has been relatively little attention paid to the transfer of knowledge from AI back to humans. The AI algorithm developed by Topalovic *et al.* [1] presents the opportunity to do exactly that. It would be relatively straightforward to develop an interface for educational use that would allow users to alter the various parameters included in the model and observe how the probabilities of the different diagnoses change accordingly. Moreover, it would be worthwhile exploring the multidimensional space of the model with an expert's hand, in order to discern whether the algorithm has uncovered completely novel associations between PFT patterns and disease, or simply recapitulated the multitude of patterns which are already known to experts. As AI becomes increasingly embedded into medical practice it is important that the diagnoses and treatment decisions emerging from computer algorithms are presented together with a rationale or explanation that is understandable to humans. This will allow clinicians to sense-check the computer output as well as ensuring continued two-way learning between artificial and human intelligence.

We encourage the authors to make the detailed structure and parameters of their model publicly available to allow others to investigate its properties. We consider that manuscripts reporting the results of AI research should specify their models in detail within a supplementary appendix in order to allow others to replicate their research. Open source code should also be made available within an online repository. As the field matures it is likely that standard checklists will be developed by the scientific publishing

@ERSpublications
Artificial intelligence holds promise as an aid to interpreting complex medical data, such as pulmonary function tests. Transparency and reproducibility are just as important as diagnostic accuracy for the success of this technology. http://bit.ly/2vYjZ8w

Cite this article as: Gonem S, Siddiqui S. Artificial intelligence for pulmonary function test interpretation. *Eur Respir J* 2019; 53: 1900638 [https://doi.org/10.1183/13993003.00638-2019].

community to ensure that all relevant aspects of AI research are reported, particularly in clinical and other domain-specific journals.

Sherif Gonem[1] and Salman Siddiqui[2,3]
[1]Dept of Respiratory Medicine, Nottingham University Hospitals NHS Trust, Nottingham, UK. [2]NIHR Leicester Biomedical Research Centre (Respiratory Theme), Glenfield Hospital, Leicester, UK. [3]Dept of Respiratory Sciences, University of Leicester, Leicester, UK.

Correspondence: Sherif Gonem, Dept of Respiratory Medicine, Nottingham City Hospital, Hucknall Road, Nottingham, NG5 1PB, UK. E-mail: sg330@le.ac.uk

Received: March 29 2019 | Accepted: April 11 2019

Conflict of interest: S. Gonem reports personal fees from 3M and Anaxsys, outside the submitted work. S. Siddiqui reports grants from Sir Jules Thorne Trust, NIHR-EME, Anaptys Bio, MRC/EPSRC and Napp, personal fees from AstraZeneca, GSK, Novartis, Boehringer Ingelheim, Mundipharma and Owlstone Medical, outside the submitted work.

Support statement: This work was supported by the National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR or the Department of Health. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Topalovic M, Das N, Burgel P-R, *et al.* Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur Respir J* 2019; 53: 1801660.
2 Patout M, Sesé L, Gille T, *et al.* Does training respiratory physicians in clinical respiratory physiology and interpretation of pulmonary function tests improve core knowledge? *Thorax* 2018; 73: 78–81.

From the authors:

We appreciate the attentive comments of S. Gonem and S. Siddiqui in their correspondence addressing our article [1]. They further highlight the apparent opportunities and the need to create a space for the introduction of novel technologies, such as our work in the field of pulmonary function testing (PFT).

We agree with the authors that disease prevalence within the training dataset will influence the outcome provided by the artificial intelligence (AI), which is why real-life situations should be considered when building the initial training dataset. It is common knowledge that classes (diseases) with few cases in the training data or unbalanced datasets may result in a poor approximation. Therefore, careful consideration should be taken to secure a large enough sample for each disease, as we did in our work. It is essential to add that the final model performance will also depend on the data pre-processing steps (such as data cleaning, normalisation, transformation, feature extraction and selection) that were taken and final model parameter tuning and optimisation [2]. This all goes into the final equation for which the algorithm will be able to recognise specific patterns that each disease carries in the PFT data. Since the prevalence of different diseases does indeed affect the overall accuracy of the algorithm, we attempted to study the disease-specific sensitivities and specificities, which are prevalence-independent test statistics [3]. The results are shown in figure 3 of the original manuscript, where the positive predictive value was usually highest for the low prevalent diseases.

The selection of 50 cases in the study was carried out over 5 days in a random week in which we randomly selected 10 cases with complete PFT performed on that day. As cases were selected at a university hospital and reference centre for respiratory diseases, a broad spectrum of disease classes was obtained.

@ERSpublications
The opportunities that may arise from using AI software for interpretation of PFTs and help in diagnosing lung diseases are evident. The idea of a complementary partnership between doctors and AI-supported systems sounds very promising. http://bit.ly/2JE60gb

Cite this article as: Topalovic M, Das N, Janssens W. Artificial intelligence for pulmonary function test interpretation. *Eur Respir J* 2019; 53: 1900782 [https://doi.org/10.1183/13993003.00782-2019].

S. Gonem and S. Siddiqui noted that all studies today focus on the knowledge that a computer can receive from humans, without any consideration for the opposite. We firmly believe in the benefits that this may create for our community, for the education of trainees and the quality of the PFT outcome and, most importantly, for the patients themselves. We can only speculate on where the future will take us, but the idea of a complementary partnership between humans and AI-supported systems sounds very promising. It may be even illusionary to believe that this will not be the case in many fields of medical diagnostic applications. During the development of our AI software, we anticipated the value it may bring and the need to make it available for the global community. Therefore, the AI software presented in the study is already available for everyday use at the University Hospital of Leuven *via* our interface. To date, more than 5000 complete PFTs have already been analysed by our software in daily clinical practice. These are real patients visiting lung function testing facilities. This number increases daily, and we are hoping to accelerate it over the coming months by adding new lung function testing laboratories. The initial and straightforward advantage of such software is that it can save a lot of the time spent on analysing PFTs in the laboratory. Moreover, the outcome of the software can be considered as an accurate second opinion (similar to automated ECG interpretations), thereby increasing the quality of the interpretations themselves. It is evident that with the appropriate setup the AI software can further enhance its performance with respect to accuracy, probabilities, and detection of new diseases or phenotypes.

Finally, as the authors noted, there is always some degree of scepticism of AI applications, owing to their being perceived as a closed “black box”. To increase confidence and trust in the software, our current work focuses on unveiling the complex reasoning of AI behind making each decision. This will, apart from offering probabilities for a diagnostic decision, additionally provide key factors that influenced that decision (for example, the probability for COPD might be 80%, because a patient has low forced expiratory volume in 1 s to forced vital capacity ratio, severely decreased diffusion capacity, increased airway resistance, hyperinflation and air-trapping, and the patient is an elderly heavy smoker).

Our original study is an excellent example of the potential that novel technologies such as AI have. It is important to emphasise that this potential needs to be nurtured appropriately, with careful attention to all factors: training data, machine learning methodology and the often-forgotten domain knowledge. We will continue our work in this field, with a promise to find a mechanism for the global medical community to implement AI software for interpretation of pulmonary function tests in their clinical practices.

Marko Topalovic, Nilakash Das and Wim Janssens, on behalf of the authors
Respiratory Medicine, University Hospital Leuven, Chronic Diseases, Metabolism and Ageing, KU Leuven, Leuven, Belgium.

Correspondence: Wim Janssens, Respiratory Medicine, University Hospital Leuven, Chronic Diseases, Metabolism and Ageing, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. E-mail: wim.janssens@uzleuven.be

Received: April 18 2019 | Accepted: April 19 2019

Conflict of interest: M. Topalovic is a co-founder of a spin-off company, ArtiQ. N. Das has nothing to disclose. W. Janssens is a co-founder of a spin-off company, ArtiQ.

References

1 Topalovic M, Das N, Burgel P-R, *et al.* Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur Respir J* 2019; 53: 1801660.
2 Kotsiantis S, Kanellopoulos D, Pintelas P. Data preprocessing for supervised leaning. *Int J Comput Sci* 2006; 1: 111–117.
3 Simundic AM. Measures of diagnostic accuracy: basic definitions. *EJIFCC* 2009; 19: 203–211.