



Artificial intelligence for thoracic radiology: from research tool to clinical practice

Lucio Calandriello¹ and Simon L.F. Walsh²

Affiliations: ¹Dipartimento di Diagnostica per Immagini, Radioterapia Oncologica ed Ematologia - Fondazione Policlinico Universitario A. Gemelli - IRCCS, Rome, Italy. ²National Heart and Lung Institute, Imperial College, London, UK.

Correspondence: Lucio Calandriello, Dipartimento di Diagnostica per Immagini, Radioterapia Oncologica ed Ematologia - Fondazione Policlinico Universitario A. Gemelli - IRCCS, L.go A. Gemelli 8, Rome, RM 00168, Italy. E-mail: lucio.calandriello@policlinicogemelli.it

 @ERSpublications

Deep learning software can provide decision support to radiologists. New evidence shows that these tools are almost ready for implementation in clinical practice. <https://bit.ly/3vk4h5t>

Cite this article as: Calandriello L, Walsh SLF. Artificial intelligence for thoracic radiology: from research tool to clinical practice. *Eur Respir J* 2021; 57: 2100625 [<https://doi.org/10.1183/13993003.00625-2021>].

Artificial intelligence (AI) presents an attractive opportunity for providing decision support to radiologists, who are often overburdened by the ever-increasing number of radiographs that are requested each year [1]. Interpretation errors, reporting delays and backlogs, particularly of chest radiographs (CXR), continue to be a major problem faced by busy radiology departments.

Deep learning is a branch of AI that shows particular promise, being proficient at identifying patterns in large quantities of data and mapping these patterns to simple categories, such as diagnosis, without the need for human programming [2]. This technology is particularly suited to medical imaging analysis and with the increasing workload faced by radiologists, may play an invaluable role in providing instantaneous decision support and reducing perceptual errors and reporting delays. The autonomous learning capabilities of deep learning algorithms also create an opportunity for developing novel image-based biomarkers that are not easily detected visually [3].

Several AI applications for CXR diagnosis have been developed for specific tasks, such as the detection of lung nodules [4], tuberculosis [5] or pneumothorax [6], and have demonstrated radiologist-level accuracy. However, an obstacle to implementing these tools in clinical practice is that they are designed to identify only one specific pathology and have been validated in retrospective, *in silico* settings. Also, although there has been some attempt to develop systems capable of identifying multiple different CXR abnormalities [7–9], many lack prospective validation in real-world clinical practice. Furthermore, it is unclear how easily deep learning-based applications will integrate into a radiologist's clinical workflow and what impact they will have on efficiency. These data will be critically important if deep learning technology is to be routinely adopted.

In this issue of the *European Respiratory Journal*, NAM *et al.* [10] present a deep learning algorithm trained to identify 10 abnormalities on CXR. The abnormalities selected (pneumothorax, mediastinal widening, pneumoperitoneum, nodule/mass, consolidation, pleural effusion, linear atelectasis, fibrosis, calcification and cardiomegaly) include some of the more significant findings routinely encountered in clinical practice.

Received: 1 March 2021 | Accepted: 2 March 2021

Copyright ©The authors 2021. For reproduction rights and permissions contact permissions@ersnet.org

The software was tested on an internal and external cohort of cases and showed an excellent detection accuracy, with area under the curve 0.893–1.0, depending on the patient cohort and abnormality. The algorithm also demonstrated improved sensitivity when compared to radiologists, with lower specificity. This is particularly attractive for radiologists facing increased workloads where fatigue and stress can lead to search and recognition errors during reporting [11]. By ensuring a low false-negative rate, important pathology is not missed, while false-positives can be reviewed and discarded by the radiologist.

Importantly, the authors tested its applicability in clinical practice by integrating it into the picture archiving and communication system (PACS). The algorithm boosted the detection performance of the radiologists and helped in identifying those CXRs needing more urgent attention. Using the software as a prioritisation tool allowed a significant reduction in the time-to-report of critical cases by sorting radiographs based on clinical urgency.

As with any deep learning software development, the training methodology is critical to the algorithm's performance. Although the quantity of data available for training is often a limiting factor, how accurately training data are labelled is equally important. In the study reported by NAM *et al.* [10], the software was trained with more than 146 000 radiographs, which were reviewed and labelled by board-certified radiologists. One concern with the human labelling of diagnostic images is the inevitable bias that is introduced because of the subjective nature of the evaluation; visual assessment of medical images is notoriously susceptible to interobserver disagreement [12]. This problem may be amplified when the diagnostic reference standard is not well-defined [13]. NAM *et al.* [10] attempted to mitigate this difficulty by labelling CXRs, where available, based on same-day CT findings. Only a proportion of CXRs had same-day computed tomography (CT) available, so one might expect that the algorithm performance might have been improved if more CT data had been available. This highlights a pervasive problem when developing deep learning applications for medical imaging analysis; most institutions do not have access to the requisite quantity of high-quality data for algorithm training. Several international initiatives are underway to address this challenge by creating large, diverse multi-institution repositories of imaging data to support deep learning research. Such a resource could also serve as common dataset for benchmarking algorithms as well as comparing the performance of different applications [3].

Another major obstacle to implementing this technology in clinical practice is that the complexity of deep learning algorithms limits their interpretability [14]. This issue is exacerbated when an algorithm is basing its predictions on features that human observers cannot detect. In recent years efforts have been made to improve algorithm interpretability; attribution techniques, such as saliency mapping, can help to identify which features have most influence on algorithmic decision making. NAM *et al.* [10] used probability maps in order to localise the regions representing the target abnormality on the CXR. This feature allows the radiologist to be more confident in confirming or rejecting a diagnosis made by the software.

Introducing deep learning software to the daily workflow can dramatically impact radiology reporting, just as the introduction of PACS did many years ago. However, this will take time; radiologists will need to become familiar with, and fully understand, these tools before they can be implemented in routine practice [7]. Lastly, prospective clinical utility studies are needed to test algorithm performance in real world clinical settings and demonstrate patient benefit over current best practice. The study by NAM *et al.* [10] represents a significant step in the right direction.

Conflict of Interest: L. Calandriello declares honoraria from Boehringer Ingelheim and Roche. S.L.F. Walsh declares a fellowship and honoraria from the National Institute for Health Research; consultancy fees and honoraria from Boehringer Ingelheim, Sanofi-Genzyme, Galapagos, Roche, Bracco, Fluidica, the Open Source Imaging Consortium, Oncoendri Therapeutics and Medscape; and advisory board membership for Boehringer Ingelheim, Sanofi-Genzyme, Galapagos and Roche.

References

- 1 Nakajima Y, Yamada K, Imamura K, *et al.* Radiologist supply and workload: international comparison—Working Group of Japanese College of Radiology. *Radiat Med* 2008; 26: 455–465.
- 2 Lee SM, Seo JB, Yun J, *et al.* Deep learning applications in chest radiography and computed tomography: current state of the art. *J Thorac Imaging* 2019; 34: 75–85.
- 3 Walsh SLF, Humphries SM, Wells AU, *et al.* Imaging research in fibrotic lung disease; applying deep learning to unsolved problems. *Lancet Respir Med* 2020; 8: 1144–1153.
- 4 Nam JG, Park S, Hwang EJ, *et al.* Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology* 2019; 290: 218–228.
- 5 Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017; 284: 574–582.
- 6 Park S, Lee SM, Kim N, *et al.* Application of deep learning-based computer-aided detection system: detecting pneumothorax on chest radiograph after biopsy. *Eur Radiol* 2019; 29: 5341–5348.

- 7 Hwang EJ, Park S, Jin KN, *et al.* Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw Open* 2019; 2: e191095.
- 8 Park S, Lee SM, Lee KH, *et al.* Deep learning-based detection system for multiclass lesions on chest radiographs: comparison with observer readings. *Eur Radiol* 2020; 30: 1359–1368.
- 9 Rajpurkar P, Irvin J, Ball RL, *et al.* Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* 2018; 15: e1002686.
- 10 Nam JG, Kim M, Park J, *et al.* Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs. *Eur Respir J* 2021; 57: 2003061.
- 11 Waite S, Scott J, Gale B, *et al.* Interpretive error in radiology. *AJR Am J Roentgenol* 2017; 208: 739–749.
- 12 Albaum MN, Hill LC, Murphy M, *et al.* Interobserver reliability of the chest radiograph in community-acquired pneumonia. PORT Investigators. *Chest* 1996; 110: 343–350.
- 13 Walsh SLF, Calandriello L, Silva M, *et al.* Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018; 6: 837–845.
- 14 Choo J, Liu S. Visual analytics for explainable deep learning. *IEEE Comput Graph Appl* 2018; 38: 84–92.