Editorial Commentary

Artificial intelligence, chest radiographs, and radiology trainees: a powerful combination to enhance the future of radiologists?

Carlo A. Mallio[1], Carlo C. Quattrocchi[1], Bruno Beomonte Zobel[1], Paul M. Parizel[2]

[1]Departmental Faculty of Medicine and Surgery, Università Campus Bio-Medico di Roma, Rome, Italy; [2]Department of Radiology, Royal Perth Hospital and University of Western Australia Medical School, Perth, WA, Australia

Correspondence to: Carlo A. Mallio, MD. Departmental Faculty of Medicine and Surgery, Università Campus Bio-Medico di Roma, Unit of Diagnostic Imaging and Interventional Radiology, Via Alvaro del Portillo, 21, 00128 Rome, Italy. Email: c.mallio@unicampus.it.

Comment on: Wu JT, Wong KCL, Gur Y, Ansari N, Karargyris A, Sharma A, Morris M, Saboury B, Ahmad H, Boyko O, Syed A, Jadhav A, Wang H, Pillai A, Kashyap S, Moradi M, Syeda-Mahmood T. Comparison of Chest Radiograph Interpretations by Artificial Intelligence Algorithm vs Radiology Residents. JAMA Netw Open 2020;3:e2022779.

Submitted Nov 26, 2020. Accepted for publication Dec 07, 2020.
doi: 10.21037/qims-20-1306
View this article at: http://dx.doi.org/10.21037/qims-20-1306

Work overload has become a major challenge for radiologists. The increasing demands upon radiologists' time, expertise and energy depend not only on the absolute number of imaging examinations to be performed and reported (i.e., number of patients), but also on the progressively growing complexity of imaging datasets, in terms of the number of images to be analyzed, as well as the quality of information to be processed, especially in the case of advanced imaging examinations that require post-processing and detailed interpretation (1-3).

Artificial intelligence (AI) is a breakthrough innovation involving computer-based algorithms tailored to analyze complex datasets (4,5). Moreover, AI is emerging as a potential game changer in many fields. In medical imaging for instance, AI showed promising results for lesion detection and quantification over a wide spectrum of clinical conditions, as well as speeding up workflows, improving accuracy, addressing resource scarcity, and reducing the costs of care (4-7). The most promising subset of AI is the so-called deep learning algorithm, in which the term "deep" is due to the artificial neural network architecture composed by multiple layers (8,9).

To be effective as representative learning applications, deep learning algorithms require large amounts of imaging data for training. These models, that are able to automatically learn, and then label, features on archetypal images, have been shown to robustly mirror or even outperform humans in task-specific applications in some cases? (8-10).

AI and deep learning are currently being tested for imaging processing in several anatomical regions and various clinical scenarios, including disorders of the chest (7,8). In this context, we read with great interest the recently published paper by Wu *et al.* (7), investigating the performance of AI model and human third-year radiology residents in interpreting chest radiographs. The novel deep learning AI algorithm that they tested was extensively trained with a large image database (i.e., 342,126 frontal chest radiographs), acquired at the emergency departments (ED) and urgent care settings in multiple hospitals. Antero-posterior (AP) and postero-anterior (PA) images were used to train the model, despite the comparison AI *vs.* human radiology residents was based on AP images only.

Interestingly, the major results of the study showed no significant difference between the performance of the AI algorithm and human radiology residents in terms of sensitivity (P=0.66); however, specificity [reported for AI 0.980 (95% CI, 0.980–0.981)] and positive predictive value [reported for AI 0.730 (95% CI, 0.718–0.742)] showed statistically significant greater results for the AI algorithm (both with P<0.001).

This work, based on the "humble" but impactful chest radiograph, which represents the most commonly performed imaging examination, is of seminal importance at least for three good reasons (7). Firstly, the authors demonstrated that there is great potential for radiologists to be helped in clinical routine by non-human (machine) assistants, with a diagnostic power that is at least equal to that achieved by medical residents. The use of machine assistants will be very helpful for physicians who work in smaller hospitals or situations where the number of radiology residents is not sufficient to cover all the clinical shifts. Moreover, radiologists working in academic teaching hospitals can also benefit if the AI supports and augment the resident activity. AI algorithms can serve as cognitive assistants for both radiologists and residents. Therefore, the time-consuming and cumbersome process of reading, improving and correcting preliminary interpretations and validating the report by attending radiologists can be speeded up, with additional benefits on improving workflow and reducing radiologist work overload. On the other hand, residents have a real-time report comparison, with a possible positive impact on their education. Secondly, this study constitutes a comprehensive and systematic effort to classify abnormalities that can be detected with chest radiographs (7). This is a fundamental and very "costly" step tailored to render AI algorithms useful in clinical practice. In order to do so, the authors started with a thorough best practices literature search, including Fleishner's glossary (11). Then, two expert clinical radiologists reviewed the included terminology for semantic consistency, which resulted in a lexicon of more than 11,000 unique terms, covering the space of 72 core findings on chest radiographs. Thirdly, the architecture of the deep neural network applied in the study is interesting in itself (7). The authors found that the best solution to analyze chest radiographs was to combine the advantages of pretrained features with a multiresolution image analysis through the feature pyramid network. Basically, they achieved this goal with a combination of ResNet22 (50 layers) (12) and VGGNet21 (16 layers) (13). Different solutions have been proposed to analyze chest radiographs with AI. For instance, a very promising one is CAD4TB (14), a deep learning system using image normalization and lung segmentation with U-net software, followed by patch-based analysis with convolutional neural network. This solution has been recently repurposed for detecting COVID-19 with chest radiograph and yielded a very good performance (area under the receiver operating characteristic curve =0.81) (15).

On the other hand, the paper by Wu *et al.* has also at least three limitations (7). Firstly, as fairly stated by the authors, they only focused the analysis on the frontal AP view without taking into account the lateral view. This is a remarkable limitation because the lateral view is of great importance in detecting chest abnormalities on X-ray studies. Certainly, including the lateral view in the AI algorithm would be very helpful in order to render the solution even more useful in clinical practice. Secondly, the AI model tested cases against five radiology residents from academic medical centers around the US. It is unclear whether, and if so, to what extent, the clinical data, the diagnostic question, and the patients' past medical history were taken into account. This information can change the approach of radiologists to image interpretation, making it more focused and efficient. Moreover, incorporating the clinical question and past medical history into the AI algorithm is ground for improvement of the reported findings (7). On this respect, a recent paper by Baltruschat *et al.* (16) investigated deep learning and chest radiograph classification. The authors reported a great spread in the yielded performance and concluded that the ResNet38 network with integration of non-image data for classification (i.e., age, gender and acquisition type) provided the best overall results (16). Thirdly, previous imaging studies of the patients were not taken into account. Since chest radiographs were sampled from the ED, it is reasonable to presume that several/many local patients already may have had previous imaging studies performed at the same hospital. Indeed, comparison with prior imaging studies, X-ray and/or CT of the chest, is often an essential step for radiologists that can influence detection and interpretation of the findings. Again, this is another point that might add strength to the AI model facilitating the transition from benchmark to bedside.

Finally, it is of note that the AI algorithm generally performed worse for low prevalence findings. Low prevalence of a condition has been recently pointed out as a challenge for AI models, and this can negatively affect the diagnostic performance (8). Moreover, the fine-grained findings which are difficult to detect and interpret, such as pulmonary nodules or enlarged hila, influenced the results, suggesting that there could still be the need for an expert over-reading of the images. Indeed, the radiologist plays a fundamental role in the clinical workflow, not only in terms of lesion detection, but also in terms of interpreting the imaging patterns in a clinical context, and this is a very valid approach for chest X-rays as well as for any other imaging modality.

There is no doubt that AI and deep learning will become increasingly prevalent and will help us to speed up our reporting workflow, face work overload and resources scarcity and, if driven by human intelligence, will enhance the future of radiologists. To improve this process, an increased transparency of deep learning AI algorithms, the so-called explainable AI, is highly advisable not only to obtain systems that are directly interpretable and trustworthy but also to give the end users the opportunity to improve their accuracy (17). In conclusion, as shown by Wu and coworkers (7), deep learning algorithms hold great potential to support clinical radiologists for reading chest radiographs.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Quantitative Imaging in Medicine and Surgery*. The article did not undergo external peer review.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/qims-20-1306). The authors have no conflicts of interest to declare.

References

1. Zha N, Patlas MNDR, Duszak R Jr. Radiologist burnout is not just isolated to the United States: perspectives from Canada. J Am Coll Radiol 2019;16:121-3.
2. Zeng Y, Zhu J, Wang J, Parasuraman P, Busi S, Nauli SM, Wáng YXJ, Pala R, Liu G. Functional probes for cardiovascular molecular imaging. Quant Imaging Med Surg 2018;8:838-52.
3. Mallio CA, Zobel BB, Quattrocchi CC. Evaluating rehabilitation interventions in Parkinson's disease with functional MRI: a promising neuroprotective strategy. Neural Regen Res 2015;10:702-3.
4. Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. Lancet Digit Health 2020;2:e486-8.
5. Neri E, Miele V, Coppola F, Grassi R. Use of CT and artificial intelligence in suspected or COVID-19 positive patients: statement of the Italian Society of Medical and Interventional Radiology. Radiol Med 2020;125:505-8.
6. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, Allison T, Arnaout O, Abbosh C, Dunn IF, Mak RH, Tamimi RM, Tempany CM, Swanton C, Hoffmann U, Schwartz LH, Gillies RJ, Huang RY, Aerts HJWL. Artificial intelligence in cancer imaging: Clinical challenges and applications. CA Cancer J Clin 2019;69:127-57.
7. Wu JT, Wong KCL, Gur Y, Ansari N, Karargyris A, Sharma A, Morris M, Saboury B, Ahmad H, Boyko O, Syed A, Jadhav A, Wang H, Pillai A, Kashyap S, Moradi M, Syeda-Mahmood T. Comparison of Chest Radiograph Interpretations by Artificial Intelligence Algorithm vs Radiology Residents. JAMA Netw Open 2020;3:e2022779.
8. Quattrocchi CC, Mallio CA, Presti G, Beomonte Zobel B, Cardinale J, Iozzino M, Della Sala SW. The challenge of COVID-19 low disease prevalence for artificial intelligence models: report of 1,610 patients. Quant Imaging Med Surg 2020;10:1891-3.
9. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018;18:500-10.
10. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521;436-44.
11. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NLRJ, Remy J. Fleischner Society: glossary of terms for thoracic imaging. Radiology 2008;246:697-722.
12. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition. June 27-30, 2016; Las Vegas, NV. Accessed on Nov 21st 2020. doi: 10.1109/CVPR.2016.90.
13. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. Preprint posted online April 10, 2015. Accessed on Nov 21st 2020. Available online: https://arxiv.org/abs/1409.1556
14. Murphy K, Habib SS, Zaidi SMA, Khowaja S, Khan A, Melendez J, Scholten ET, Amad F, Schalekamp S, Verhagen M, Philipsen RHHM, Meijers A, van Ginneken B. Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. Sci

Rep 2020;10:5492.
15. Murphy K, Smits H, Knoops AJG, Korst MBJM, Samson T, Scholten ET, Schalekamp S, Schaefer-Prokop CM, Philipsen RHHM, Meijers A, Melendez J, van Ginneken B, Rutten M. COVID-19 on Chest Radiographs: A Multireader Evaluation of an Artificial Intelligence System. Radiology 2020;296:E166-72.
16. Baltruschat IM, Nickisch H, Grass M, Knopp T, Saalbach A. Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification. Sci Rep 2019;9:6381.
17. Singh A, Sengupta S, Lakshminarayanan V. Explainable Deep Learning Models in Medical Image Analysis. J Imaging 2020:6:52.

Cite this article as: Mallio CA, Quattrocchi CC, Beomonte Zobel B, Parizel PM. Artificial intelligence, chest radiographs, and radiology trainees: a powerful combination to enhance the future of radiologists? Quant Imaging Med Surg 2021;11(5):2204-2207. doi: 10.21037/qims-20-1306

