Radiology

The Potential of Artificial Intelligence to Analyze Chest Radiographs for Signs of COVID-19 Pneumonia

Manuscript Type: Reviews and Commentary

Bram van Ginneken, PhD

Radboud University Medical Center

Geert Grooteplein Zuid 10, 6525 GA Nijmegen, The Netherlands

Corresponding author: B.v.G. (e-mail bram.vanginneken@radboudumc.nl)

See also the article by Wehbe et al.

It is about one year ago that a new coronavirus started to spread from Wuhan, China. The resulting pandemic is unprecedented in many ways, and one of them is the number of scientific publications it has generated. PubMed already lists over 70,000 papers on COVID-19. The first publication in *Radiology*, describing the CT appearance of COVID-19 pneumonia findings, dates from Feb 6. To date, *Radiology* has published 40 original research articles on the topic. These studies have a significant impact: the twelve most cited papers from *Radiology* from 2020 (using counts from Google Scholar) are all on COVID-19 and even the article ranked 12th has twice as many citations as the most cited *Radiology* paper from 2019. (Interestingly, the twelve most-cited 2019 *Radiology* publications are all on applications of artificial intelligence.)

From the *Radiology* COVID-19 papers published so far, 23 studies focused on CT and only six on chest radiography. This is likely related to the fact that, as noted by the Fleischner Society in their consensus statement [1] on the role of imaging in patient management during the pandemic, "chest radiography is insensitive in mild or early COVID-19". This conclusion was based on evidence from the first article on chest radiographic findings in COVID-19 patients [2]. However, many countries encourage individuals with complaints consistent with COVID-19 to quarantine at home. In such a scenario, patients presenting in a hospital may have more advanced disease, often with abnormalities visible on a chest radiograph. Because of its broad availability, low cost, and portability, chest radiography is a widely used tool to obtain an initial diagnosis while waiting for the results from molecular diagnostic tests. Radiographic imaging can also help to assess disease progression or detect the presence of other diseases.

Many healthcare providers are overburdened during this pandemic and struggle with a lack of resources for image interpretation. Artificial intelligence (AI) could provide support in the reading process.

In this issue of *Radiology*, Wehbe et al. [3] present an AI algorithm coined DeepCOVID-XR that detects COVID-19 in single frontal chest radiographs. It is not the first paper in *Radiology* attempting to do just this. In May, Murphy et al. [4] reported on a validation study of CAD4COVID-Xray, a freely available CE-marked commercial solution; I was a co-author of that study. In September, Zhang et al. [5] introduced CV19-Net. All three algorithms address the same task.

A strength of DeepCOVID-XR is that it was trained on a large multi-center data set: almost 15,000 images with over 4,000 cases that were positive for COVID-19 originating from more than 20 sites across the Northwestern Memorial Healthcare System, an organization operating in the Chicago region. Nearly all images were anterior-posterior bedside exams. The AI system was evaluated on data from a single community hospital. This was a proper external validation set: no images from that hospital had been used in training the system. A set of 300 random cases (134 positive) were presented to five radiologists to allow a direct comparison between the AI system and human experts.

A similar setup was used in the other two studies. CAD4COVID-Xray was evaluated on 454 images (223 positive) and compared with scores of six radiologists. But this system was trained with only 416 images of COVID-19 suspects of only one other hospital, although the deep learning network was pre-trained on pneumonia data from other sources. CV19-Net was trained with data from Henry Ford Health system and used in total around 5,000 images (about half positive) for training. The negative cases were from patients diagnosed with pneumonia in 2019. A drawback of this study was that the test set came from hospitals that also provided most of the training data. On 500 randomly selected test images, equally balanced between positive and negative cases, CV19-Net was compared with readings from three radiologists.

All three systems provided promising results in terms of their area under the receiver operating curve (AUC), the most commonly used metric for systems that provide a continuous output for a binary classification task – the AUC is equivalent to the chance that a random positive image receives a higher score than a random negative image in the test set. DeepCOVID-XR reached an AUC of 0.88, comparable to the consensus of the five radiologists (AUC 0.85, scores using a six-point scale). CV19-Net had an AUC of 0.94, outperforming each of the three readers who did not provide continuous scoring. CAD4COVID-Xray achieved an AUC of 0.81 and slightly outperformed the radiologists at high sensitivity cut-offs but performed slightly inferior to 4 of the 6 radiologists at their high specificity cut-off.

An interesting strategy that Wehbe et al. pursued was to design an ensemble of neural networks with diverse characteristics. They trained six different architectures that are popular today (DenseNet-121, ResNet-50, Inception, Inception-ResNet, Xception, and EfficientNet-B2) using two resolution levels (224 x 224 and 331 x 331) and two field-of-views (the entire radiograph and the image cropped around the automatically segmented lung fields). Research has shown [6] that zooming in on the lung fields may lead to slightly better performance for abnormality detection in chest radiographs. In this study, differences were minor, but the combined approach may be more robust when applied to unseen data.

An important avenue for further research mentioned by Wehbe et al. is to combine image analysis with additional input, such as demographics, vital signs, and laboratory data. A study using a simple scoring tool suggested such a multi-modal approach can increase the AUC substantially compared to imaging alone [7]. Additionally, AI analysis of chest radiographs may predict outcomes and guide patient management and interventions. A promising study on predicting intubation and mortality has just been published in *Radiology: Artificial Intelligence* [8]. Such research requires the availability of large datasets where standardized outcomes and treatment parameters are available. Collecting such data remains extremely challenging when patient management strategies are continuously adapted while we learn more about COVID-19.

To move forward and learn which approaches to automated analysis have the most potential, we should compare the performance of the various “nets” now published. Sharing the training and test data would facilitate this. But making medical data publicly available can be a complicated process, and it is something that journals like *Radiology* do not require. *Radiology* editorial guidelines do request researchers to share their code unless their study reports on commercial software, as in Murphy et al. [4]. However, these guidelines do not specify what type of code should be shared, nor are reviewers encouraged to verify that the code produces the results reported in the paper. As a result, the reusability of the shared code is often limited. Zhang et al. [5] shared code on GitHub, but their repository does not include the network’s weights and, therefore, cannot be used to process new images. Wehbe et al. have made their code available on GitHub, together with network weights and instructions on applying the networks to new data and even training the system on additional images. My group is now working on comparing the results of CAD4COVID-Xray and DeepCOVID-XR on the test set of Murphy et al. [4]. This would be the start of a more extensive external validation of several AI tools that could contribute to the global fight against COVID-19.

Bram van Ginneken is Professor of Medical Image Analysis at Radboud University Medical Center. He also works for Fraunhofer MEVIS in Bremen, Germany, and is a founder of Thirona, a company that develops software and provides services for medical image analysis. He studied Physics at Eindhoven University of Technology and at Utrecht University where he obtained his PhD in 2001 on Computer-Aided Diagnosis in Chest Radiography. He pioneered the concept of challenges in medical image analysis.

References

1. Rubin GD, Ryerson CJ, Haramati LB, et al. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. Radiology 2020;296:172-180.
2. Wong HYF, Lam HYS, Fong AHT, et al. Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. Radiology 2020;296:E72-E78.
3. Wehbe RM, Sheng J, Dutta S et al. DeepCOVID-XR: An Artificial Intelligence Algorithm to Detect COVID-19 on Chest Radiographs Trained and Tested on a Large US Clinical Dataset. Radiology. In Press.
4. Murphy K, Smits H, Knoops AJG, et al. COVID-19 on the Chest Radiograph: A Multi-Reader Evaluation of an AI System. Radiology 2020;296:E166–E172.
5. Zhang R, Xin Tie X, Qi Z, et al. Diagnosis of COVID-19 Pneumonia Using Chest Radiography: Value of Artificial Intelligence. Radiology. Published Online: Sep 24 2020.
6. Baltruschat IM, Steinmeister L, Ittrich H, et al. When does bone suppression and lung field segmentation improve chest x-ray disease classification? Proc. Int Symp Biomed Imaging 2019;1362–1366.
7. Kurstjens S, van der Horst A, Herpers R, et al. Rapid identification of SARS-CoV-2-infected patients at the emergency department using routine testing. Clin Chem Lab Med. 2020 Jun 29;58(9):1587-1593.
8. Li MD, Arun NT, Gidwani M, et al. Automated Assessment and Tracking of COVID-19 Pulmonary Disease Severity on Chest Radiographs using Convolutional Siamese Neural Networks. Radiology Artificial Intelligence. Published Online: Jul 22 2020.