

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

Bram van Ginneken, PhD

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 doctorate in 2001 on computer-aided diagnosis in chest radiography. He pioneered the concept of challenges in medical image analysis.



It was about 1 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 coronavirus disease 2019 (COVID-19). The first publication in *Radiology*, describing the CT appearance of COVID-19 pneumonia findings, dates from February 6, 2020. To date, *Radiology* has published 40 original research articles on this topic. These studies have had a substantial impact: the 12 most cited papers in *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* article from 2019. (Interestingly, the 12 most-cited 2019 *Radiology* publications are all on applications of artificial intelligence [AI].)

Of the *Radiology* COVID-19 articles published so far, 23 have focused on CT, and only six have focused on chest radiography. This is likely related to the fact that, as noted by the Fleischner Society (1) in their consensus statement on the role of imaging in patient care 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 patients with COVID-19 (2). However, many countries encourage individuals with symptoms 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 assess disease progression or help detect other diseases.

Many health care providers are overburdened during this pandemic and struggle with a lack of resources for

image interpretation. 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 on single frontal chest radiographs. It is not the first article in *Radiology* to attempt to do this. In May 2020, Murphy et al (4) reported on a validation study of CAD4COVID-x-ray, a freely available CE-marked commercial solution; I was a coauthor of that study. In September 2020, 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 multicenter data set: Almost 15 000 images with over 4000 cases that were positive for COVID-19 originating from more than 20 sites across the Northwestern Memorial Health Care System, an organization operating in the Chicago, Ill, region. Nearly all images were anteroposterior bedside radiographs. The AI system was evaluated on data from one community hospital. This was a proper external validation set: No images from that hospital had been used to train the system. A set of 300 random cases (134 positive for COVID-19) were presented to five radiologists to allow a direct comparison between the AI system and human experts.

A similar set-up was used in the other two studies. CAD4COVID-x-ray was evaluated on 454 images (223 positive for COVID-19) and compared with scores of six radiologists. However, this system was trained with only 416 images of COVID-19 suspects from 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 a total of approximately 5000 images (about half of which were positive for COVID-19) 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 yielded promising results in terms of their area under the receiver operating characteristic 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

From the Department of Radiology, Radboud University Medical Center, Geert Grooteplein Zuid 10, 6525 GA Nijmegen, the Netherlands. Received November 8, 2020; revision requested November 16; revision received November 16; accepted November 16. Address correspondence to the author (e-mail: bram.vanginneken@radboudumc.nl).

Conflicts of interest are listed at the end of this article.

See also the article by Wehbe et al in this issue.

Radiology 2021; 299:E214–E215 • <https://doi.org/10.1148/radiol.2020204238> • Content codes: **CH** **AI** • © RSNA, 2021

This copy is for personal use only. To order printed copies, contact reprints@rsna.org

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 were obtained using a six-point scale). CV19-Net had an AUC of 0.94, outperforming each of the three readers who did not perform continuous scoring. CAD4COVID-x-ray achieved an AUC of 0.81 and slightly outperformed the radiologists at high-sensitivity cutoffs but performed slightly inferior to four of the six radiologists at their high specificity cutoff.

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 (224×224 and 331×331) and two fields of view (the entire radiograph and the image cropped around the automatically segmented lung fields). Research has shown that zooming in on the lung fields may lead to slightly better performance for abnormality detection on chest radiographs (6). 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 multimodal approach can increase the AUC substantially compared with imaging alone (7). Additionally, AI analysis of chest radiographs may predict outcomes and guide patient care 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 data sets where standardized outcomes and treatment parameters are available. Collecting such data remains extremely challenging when patient care 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. However, 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 article. As a result, the reusability of the shared code is often limited. Zhang et al (5) shared their code on GitHub, but their repository does not include the network weights; therefore, it cannot be used to process new images. Wehbe et al have made their code available on GitHub, together with network weights and instructions on how to apply the networks to new data and even how to train the system on additional images. My group is now working on comparing the results of CAD4COVID-x-ray 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 artificial intelligence tools that could contribute to the global fight against coronavirus disease 2019.

Disclosures of Conflicts of Interest: B.v.G. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: institution received grants from Thirona, Delft Imaging, Siemens Healthineers, and MeVis Medical Solutions; institution receives royalties from Thirona, Delft Imaging, and MeVis Medical Solutions; owns Thirona stock. Other relationships: disclosed no relevant relationships.

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(1):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(2):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 U.S. Clinical Data Set. *Radiology* 2021;299:E167–E176.
4. Murphy K, Smits H, Knoops AJG, et al. COVID-19 on Chest Radiographs: A Multireader Evaluation of an Artificial Intelligence System. *Radiology* 2020;296(3):E166–E172.
5. Zhang R, Tie X, Qi Z, et al. Diagnosis of COVID-19 Pneumonia Using Chest Radiography: Value of Artificial Intelligence. *Radiology* 2020. 10.1148/radiol.2020202944. Published online September 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? In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, April 8–11, 2019. Piscataway, NJ: IEEE, 2019; 1362–1366.
7. Kurstjens S, van der Horst A, Hoppers R, et al. Rapid identification of SARS-CoV-2-infected patients at the emergency department using routine testing. *Clin Chem Lab Med* 2020;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. *Radiol Artif Intell* 2020;2(4):e200079 <https://doi.org/10.1148/ryai.2020200079>.