

Artificial Intelligence of COVID-19 Imaging: A Hammer in Search of a Nail

Ronald M. Summers, MD, PhD

The coronavirus disease 2019 (COVID-19) pandemic has irrevocably altered our personal and professional lives. More than 1 million people globally have died of the virus. New spikes in infections are occurring worldwide as I write this editorial. The economic impact has been devastating.

Early in the pandemic, expectations were raised that chest CT and radiography might play a crucial role in first-line diagnosis of COVID-19 (1). Over time, the reverse transcription polymerase chain reaction test became more sensitive, and clinicians' understanding of the disease and how to treat it improved. Chest CT and radiography has moved to a secondary role. As of this writing, chest CT is not recommended as a first-line test by the American College of Radiology (2).

The pandemic reached the United States in January 2020. By March 2020, manuscripts using artificial intelligence (AI) for evaluation of COVID-19 on chest radiographs and CT scans began appearing on preprint servers such as arXiv (3). Within a month, there were over a dozen such manuscripts. Simultaneously, major journals including *Radiology* began publishing AI articles on COVID-19 (4,5). Most found that these AI systems had high sensitivity for detection of lung opacities due to COVID-19. Subsequent works showed that AI could distinguish COVID-19 from other types of pneumonia (6).

The pace of publication of AI articles on COVID-19 is increasing. As of this writing, there are more than 500 such manuscripts on arXiv (Google Scholar search: COVID-19 ["CT" OR "X-ray"] ["machine learning" OR "deep learning" OR "artificial intelligence"] site:arxiv.org) and more than 200 articles on PubMed (COVID-19 ["CT" OR "X-ray"] ["machine learning" OR "deep learning" OR "artificial intelligence"]). The articles are appearing not only in radiology clinical journals but also in technical and general interest scientific journals (7,8). There is a clear appetite for such research despite its repetitiveness and unclear path to clinical utility.

Similar enthusiasm for applying AI to pneumonia occurred before COVID-19. During the H1N1 pandemic, an AI system used support vector machine classifiers and texture analysis to detect lung opacities at chest CT (9). The Radiological Society of North America (RSNA) pneumonia detection challenge in 2018 led to more than 1000 teams competing to submit the most effective AI systems for pneumonia detection on chest radiographs (10). But few of those same systems were subsequently applied to COVID-19 (as determined by the paucity of

citations to Shih et al [11] on arXiv and PubMed). It is surprising that those systems were not repurposed.

To a large extent, the large quantity and rapid publication of articles on AI for COVID-19 are emblematic of current trends in other areas of radiology AI. It is now so much easier to design and conduct a radiology AI experiment. The only prerequisite seems to be possession of a large data set. The AI software tools are available online for free. There are abundant tutorials and recipes for AI research on images in general and radiology images in particular (12).

Even the data needed for training and testing AI systems are often not limiting factors. Free online COVID-19 radiology data sets are proliferating. These online data sets include the Cancer Imaging Archive, the British National COVID-19 Chest Imaging Database, and the Valencian Region Medical ImageBank (13–16). The RSNA has developed the Medical Imaging and Data Resource Center, or MIDRC, and RSNA International COVID-19 Open Radiology Database resources and attained buy-in from radiology departments to contribute images from patients with COVID-19 (17,18). At the time of this writing, MIDRC data are not yet online.

Public data sets vary in quality and utility. For example, some of the data sets include images but no ancillary data such as annotations, demographics, laboratory results, or outcome data. Some of the annotations are basic such as COVID-19 or no COVID-19. Others are more comprehensive, including anatomic labeling or results of diagnostic antibody tests. Unfortunately, many comprehensive data sets are private and not available to the public. Some AI code and models needed for external use are posted online (19). But there is little evidence to date from investigators independent of the original research teams whether published AI systems generalize to their patients' data.

The three major classes of COVID-19 AI research include binary diagnosis (COVID-19 present or absent) (4), segmentation and quantification of the abnormal lung opacities (20), and distinguishing COVID-19 from non-COVID-19 pneumonias (6). Binary diagnosis was one of the first applications studied in depth. More limited areas of investigation include prediction of future need for oxygen therapy or intubation (21), prediction of acute respiratory distress syndrome development (22), generalizability to multinational patient populations (8), integration of imaging and clinical information (23), analysis of serial imaging (24), tailoring steroid treatment (25), and

From the Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Department of Radiology and Imaging Sciences, National Institutes of Health Clinical Center, 10 Center Dr, Bldg 10, Room 1C224D, MSC 1182, Bethesda, MD 20892-1182. Received November 6, 2020; revision requested November 10; final revision received November 10; accepted November 13. **Address correspondence** to the author (e-mail: rms@nih.gov).

Supported by the Intramural Research Program of the National Institutes of Health Clinical Center.

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

Radiology 2020; 00:1–3 • <https://doi.org/10.1148/radiol.20202042226> • Content codes:  • ©RSNA, 2020

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

mortality prediction (26). By using natural language processing of clinical reports, opacities on radiology images were included in machine learning models to predict need for intensive care unit admission (27).

A recent illustrative example of a binary diagnostic task is the study by Zhang et al (28). The authors studied chest radiographs from 2060 patients with COVID-19 pneumonia and 3148 patients with non-COVID-19 pneumonia. On the test set, their AI system had an area under the receiver operating characteristic curve (AUC) of 0.92 and sensitivities and specificities of 88% and 79%, or 78% and 89% for high sensitivity or high specificity operating thresholds. On a subset of 500 chest radiographs, their AI system achieved an AUC of 0.94 compared with an AUC of 0.85 for three experienced thoracic radiologists. This AUC is typical for these binary diagnostic tasks.

An illustrative mortality prediction study is that by Mushtaq et al (26). This was a single institution study of 697 adults with COVID-19 infection confirmed by reverse transcription polymerase chain reaction who presented to the emergency department. A commercial AI system analyzed patients' initial chest radiographs and outputted a score indicating percentage of lung involvement. The score was predictive of mortality (hazard ratio, 2.60) and critical COVID-19 (admission to the intensive care unit or deaths occurring before intensive care unit admission; hazard ratio, 3.40).

The basic technical approach of COVID-19 AI research is similar. The first step is to collect a sufficiently large data set. The goalposts determining what is sufficiently large keeps shifting to larger numbers of scans. Typical data sets range from hundreds to thousands of patients' scans. Whether larger data sets are necessary has received inadequate scrutiny. Some published studies include non-COVID scans for training because they are abundant. The data need accurate labels. For some published studies, the labels are binary—the study is positive or negative for COVID-19. For other studies, a segmentation label is manually drawn to identify the extent of the lung abnormality. Some studies use rectangular bounding boxes rather than the more labor-intensive free-form segmentations. The next step is to divide the scans into separate training, validation, and test sets. The machine learning software is taught by using the labeled training data, periodically run on the validation set for fine tuning, and then run only once on the test set. It is ideal to also have an external test data set of patients from a different demographic or institution than the one used for training. High performance on the external test data set increases confidence that the AI is generalizable to new patient populations.

There are many choices for the particular deep learning architecture. Some articles pitted different architectures against one another to determine which was most accurate. One such work showed little difference in performance by the different architectures (29). Example deep learning architectures proven successful for detecting COVID-19 include EfficientNet (6), U-Net (30), ResNet (31), and Inf-Net (32). Some studies preprocess the images to segment the lungs before analysis for pulmonary parenchymal opacities (5). U-Net is a very popular deep learning architecture for lung segmentation (33).

How does one put this deluge of articles into context? It seems unlikely that an AI system would detect many patients with COVID-19 who had a negative reverse transcription polymerase chain reaction test. Anecdotes will occur. But from a general perspective, this is unlikely to propel dissemination of the AI technology. What about distinguishing COVID-19 from other viral pneumonias? It seems unlikely that clinical decision making would depend on the recommendations of AI, given more definitive laboratory tests are available. Could AI lead to a fully automated interpretation? This has not been the focus of COVID-19 imaging AI to date. Multitask approaches that identify multiple abnormalities at chest imaging besides opacities will be needed, such as universal lesion detection (34,35). What about mortality prediction? Hazard ratios on the order of 2 to 3, as found in the article by Mushtaq et al, are generally insufficient for clinical decision making. While it is possible that prediction of an adverse outcome could lead to more aggressive treatment, it could also lead to unnecessary costs and adverse effects.

We are beginning to understand the many risk factors for severe COVID-19 infection and death. These include the presence of underlying conditions such as respiratory or cardiovascular disease, hypertension or diabetes, advanced age, and male sex (36). Thus, there are many opportunities for AI systems that assess or incorporate information about these diseases or patient demographics.

We are also beginning to understand some of the nonpulmonary manifestations of COVID-19 (37). These include hepatic and renal injury, neurologic illnesses, and a coagulopathy leading to thrombi. The thrombotic complications can occur anywhere in the body including the mesentery leading to bowel ischemia (38). These nonpulmonary findings are suitable and desirable targets for AI systems.

What are the current needs of AI systems for COVID-19 and CT and chest radiography? Public challenges or competitions pitting different AI systems against one another would enable “apples-to-apples” comparisons of performance. More observer performance experiments are necessary to determine whether AI improves clinical interpretation according to reader experience level and reading paradigm (first, concurrent, or second reader). Prospective outcome studies are necessary to determine whether the use of AI leads to changes in patient care, shortened hospitalizations, and reduced morbidity and mortality. Nonradiology clinical information will need to be routinely incorporated into AI models. Assessment of risk and progression of the chronic sequela of COVID-19 infection is necessary. A prospective randomized controlled trial would be exemplary.

It is time to move beyond studies showing that AI can detect opacities at CT or chest radiography—this is now well established. Instead, there is a great need for AI systems, based on a combination of imaging, laboratory, and clinical information, that provide actionable predictions otherwise unavailable or less accurate without AI.

Disclosures of Conflicts of Interest: R.M.S. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: receives royalties from iCAD, ScanMed, Philips, PingAn, and Translation Holdings. Other relationships: has cooperative research and development agreement with PingAn; received GPU card donations from NVIDIA.

References

- Fang Y, Zhang H, Xie J, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* 2020;296(2):E115–E117.
- ACR Recommendations for the use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection. <https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection>. Published 2020. Updated March 22, 2020. Accessed November 1, 2020.
- Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P. COVID-CT-Dataset: A CT Scan Dataset about COVID-19. *ArXiv e-prints* 2020:arXiv:2003.13865. [preprint] <https://arxiv.org/abs/2003.13865>. Posted March 30, 2020. Accessed November 1, 2020.
- Li L, Qin L, Xu Z, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* 2020;296(2):E65–E71.
- Murphy K, Smits H, Knoop AJG, et al. COVID-19 on Chest Radiographs: A Multi-reader Evaluation of an Artificial Intelligence System. *Radiology* 2020;296(3):E166–E172.
- Bai HX, Wang R, Xiong Z, et al. Artificial Intelligence Augmentation of Radiologist Performance in Distinguishing COVID-19 from Pneumonia of Other Origin at Chest CT. *Radiology* 2020;296(3):E156–E165.
- Zhou L, Li Z, Zhou J, et al. A Rapid, Accurate and Machine-Agnostic Segmentation and Quantification Method for CT-Based COVID-19 Diagnosis. *IEEE Trans Med Imaging* 2020;39(8):2638–2652.
- Harmon SA, Sanford TH, Xu S, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun* 2020;11(1):4080.
- Yao J, Dwyer A, Summers RM, Mollura DJ. Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification. *Acad Radiol* 2011;18(3):306–314.
- RSNA Pneumonia Detection Challenge (2018). <https://www.rsna.org/en/education/ai-resources-and-training/ai-image-challenge/RSNA-Pneumonia-Detection-Challenge-2018>. Published 2018. Accessed November 1, 2020.
- Shih G, Wu CC, Halabi SS, et al. Augmenting the National Institutes of Health Chest Radiograph Dataset with Expert Annotations of Possible Pneumonia. *Radiol Artif Intell* 2019;1(1):e180041.
- Erickson BJ. Magician's Corner: How to Start Learning about Deep Learning. *Radiol Artif Intell* 2019;1(4):e190072.
- Jacob J, Alexander D, Baillie JK, et al. Using imaging to combat a pandemic: rationale for developing the UK National COVID-19 Chest Imaging Database. *Eur Respir J* 2020;56(2):2001809.
- Chest Imaging with Clinical and Genomic Correlates Representing a Rural COVID-19 Positive Population (COVID-19-AR). <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70226443#70226443171ba531fc374829b21d3647e95f532c>. Published 2020. Updated September 15, 2020. Accessed November 1, 2020.
- CT Images in COVID-19. <https://wiki.cancerimagingarchive.net/display/Public/CT+Images+in+COVID-19>. Published 2020. Updated September 15, 2020. Accessed November 1, 2020.
- Vayá MdI, Saborit JM, Montell JA, et al. BIMCV COVID-19+: a large annotated dataset of RX and CT images from COVID-19 patients. *arXiv preprint arXiv:2006.01174* [preprint]. <https://arxiv.org/abs/2006.01174>. Posted June 1, 2020. Accessed November 1, 2020.
- RSNA To Collaborate on Open-Source COVID-19 Medical Image Database. <https://www.rsna.org/news/2020/july/covid-19-midrc>. Published 2020. Updated August 5, 2020. Accessed November 1, 2020.
- Langlotz CP. My Turn: Will Artificial Intelligence Play a Role in Imaging of COVID-19? <https://www.rsna.org/news/2020/September/Langlotz-My-Turn-COVID-19-And-AI>. Published 2020. Updated September 22, 2020. Accessed November 1, 2020.
- COVID-CT. <https://github.com/UCSD-AI4H/COVID-CT>. Published 2020. Updated May 11, 2020. Accessed November 1, 2020.
- Li Z, Zhong Z, Li Y, et al. From community-acquired pneumonia to COVID-19: a deep learning-based method for quantitative analysis of COVID-19 on thick-section CT scans. *Eur Radiol* 2020;30(12):6828–6837.
- Lanza E, Muglia R, Bolengo I, et al. Quantitative chest CT analysis in COVID-19 to predict the need for oxygenation support and intubation. *Eur Radiol* 2020;30(12):6770–6778.
- Wang Y, Chen Y, Wei Y, et al. Quantitative analysis of chest CT imaging findings with the risk of ARDS in COVID-19 patients: a preliminary study. *Ann Transl Med* 2020;8(9):594.
- Mei X, Lee HC, Diao KY, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat Med* 2020;26(8):1224–1228.
- Pu J, Leader JK, Bandos A, et al. Automated quantification of COVID-19 severity and progression using chest CT images. *Eur Radiol* 2020. 10.1007/s00330-020-07156-2. Published online August 13, 2020.
- Su Y, Han Y, Liu J, et al. Tailoring steroids in the treatment of COVID-19 pneumonia assisted by CT scans: three case reports. *J XRay Sci Technol* 2020;28(5):885–892.
- Mushtaq J, Pennella R, Lavallo S, et al. Initial chest radiographs and artificial intelligence (AI) predict clinical outcomes in COVID-19 patients: analysis of 697 Italian patients. *Eur Radiol* 2020. 10.1007/s00330-020-07269-8. Published online September 18, 2020.
- Hao B, Sutudian S, Wang T, et al. Early prediction of level-of-care requirements in patients with COVID-19. *eLife* 2020;9:e60519.
- 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.
- Bressem KK, Adams LC, Erxleben C, Hamm B, Niehues SM, Vahldiek JL. Comparing different deep learning architectures for classification of chest radiographs. *Sci Rep* 2020;10(1):13590.
- Wang M, Xia C, Huang L, et al. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. *Lancet Digit Health* 2020;2(10):e506–e515.
- Wang J, Bao Y, Wen Y, et al. Prior-Attention Residual Learning for More Discriminative COVID-19 Screening in CT Images. *IEEE Trans Med Imaging* 2020;39(8):2572–2583.
- Fan DP, Zhou T, Ji GP, et al. Inf-Net: Automatic COVID-19 Lung Infection Segmentation From CT Images. *IEEE Trans Med Imaging* 2020;39(8):2626–2637.
- Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Cham, Switzerland: Springer, 2015; 234–241.
- Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, July 21–26, 2017. Piscataway, NJ: IEEE, 2017; 3462–3471.
- Yan K, Wang X, Lu L, Summers RM. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J Med Imaging (Bellingham)* 2018;5(3):036501.
- Mesas AE, Cavero-Redondo I, Álvarez-Bueno C, et al. Predictors of in-hospital COVID-19 mortality: A comprehensive systematic review and meta-analysis exploring differences by age, sex and health conditions. *PLoS One* 2020;15(11):e0241742.
- Gupta A, Madhavan MV, Sehgal K, et al. Extrapulmonary manifestations of COVID-19. *Nat Med* 2020;26(7):1017–1032.
- Bhayana R, Som A, Li MD, et al. Abdominal Imaging Findings in COVID-19: Preliminary Observations. *Radiology* 2020;297(1):E207–E215.