Lessons learned in transitioning to AI in the medical imaging of COVID-19

Issam El Naqa,[a,b,*] **Hui Li**,[b,c] **Jordan Fuhrman**,[b,c] **Qiyuan Hu**,[b,c] **Naveena Gorre**,[a,b] **Weijie Chen**,[b,d] and **Maryellen L. Giger**[b,c]

[a]Moffitt Cancer Center, Department of Machine Learning, Tampa, Florida, United States

[b]The University of Chicago, Medical Imaging Data and Resource Center, Chicago, Illinois, United States

[c]The University of Chicago, Department of Radiology, Chicago, Illinois, United States

[d]US FDA, CDRH, Office of Science and Engineering Laboratories, Division of Imaging, Diagnosis, and Software Reliability, Silver Spring, Maryland, United States

Abstract. The coronavirus disease 2019 (COVID-19) pandemic has wreaked havoc across the world. It also created a need for the urgent development of efficacious predictive diagnostics, specifically, artificial intelligence (AI) methods applied to medical imaging. This has led to the convergence of experts from multiple disciplines to solve this global pandemic including clinicians, medical physicists, imaging scientists, computer scientists, and informatics experts to bring to bear the best of these fields for solving the challenges of the COVID-19 pandemic. However, such a convergence over a very brief period of time has had unintended consequences and created its own challenges. As part of Medical Imaging Data and Resource Center initiative, we discuss the lessons learned from career transitions across the three involved disciplines (radiology, medical imaging physics, and computer science) and draw recommendations based on these experiences by analyzing the challenges associated with each of the three associated transition types: (1) AI of non-imaging data to AI of medical imaging data, (2) medical imaging clinician to AI of medical imaging, and (3) AI of medical imaging to AI of COVID-19 imaging. The lessons learned from these career transitions and the diffusion of knowledge among them could be accomplished more effectively by recognizing their associated intricacies. These lessons learned in the transitioning to AI in the medical imaging of COVID-19 can inform and enhance future AI applications, making the whole of the transitions more than the sum of each discipline, for confronting an emergency like the COVID-19 pandemic or solving emerging problems in biomedicine. © *2021 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JMI.8.S1.010902]

1 Introduction

1.1 *Motivation*

The coronavirus disease 2019 (COVID-19) is an ongoing pandemic caused by severe acute respiratory syndrome coronavirus 2. It was first identified in late 2019 in Wuhan, China, and later declared as a pandemic by the World Health Organization (WHO), with its full impact yet to be understood, from preparedness and testing, to diagnosis and vaccination.[1–6] Currently, reverse transcription polymerase chain reaction (RT-PCR) is the reference standard for COVID-19 patient diagnosis. In addition to molecular testing, medical imaging modalities, including chest x-ray (CXR),[7,8] chest CT,[9–11] and chest ultrasound,[12–15] have been used in the diagnosis and management of COVID-19 patients.

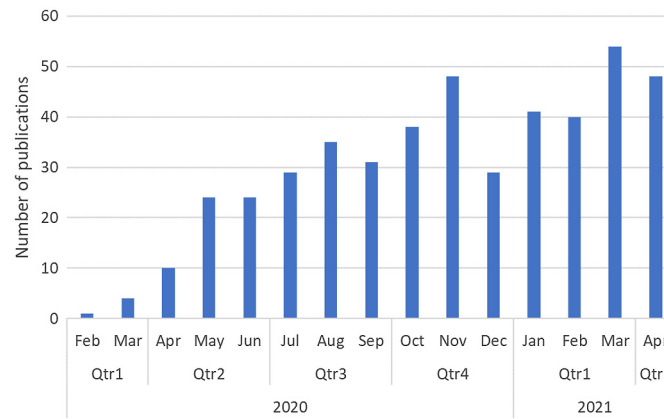*Address all correspondence to Issam El Naqa, issam.elnaqa@moffitt.org

Fig. 1 Number of publications on COVID-19 with a PubMed search of "COVID" and "imaging" and ("machine learning" or "deep learning" or "AI").

Challenges associated with this disease have also identified new opportunities for developing better diagnostic tools and therapeutics that are expected to survive this pandemic and lead to a brighter future in medical imaging analytics.[1,16] Among the many lessons learned from this pandemic is the development of efficacious predictive diagnostics using medical imaging analyses[17–24] and advanced artificial intelligence (AI) tools.[25–28] There has been a tremendous uptake in the utilization of AI for analyzing chest radiographs and computed tomography lung images for COVID-19, as presented in Fig. 1. However, many of these studies replicate similar analyses, with limitations on datasets and validation methods and, thus, unclear findings and paths to clinical utility.[29] Moreover, earlier studies have utilized pediatric datasets, whereas most of the COVID-19 patients have been adults, which created an age bias effect.[30,31]

Despite AI's known potentials in radiological sciences,[32] issues related to image quality, preprocessing, representative training/tuning/testing [The "training/validation/testing" terminology has been popularly used, especially in the computer science community. However, the terminology has been found to be confusing for many in the medical imaging and clinical communities. The "validation" refers to using a dataset (typically small) to examine the performance of a trained model under certain hyperparameters and tune the hyperparameters based on the model performance on that dataset. Therefore, we use "training/tuning/testing."] datasets, choices of improper AI algorithms and structures and interpretability of AI findings seem to have contributed to irreproducible and conflicting results in the literature.[33] Though this convergence of many forces to solve a global pandemic is admirable, to meet its intended goals, it has to be conducted not in a single step but as a transitional process requiring collective efforts among medical imaging professionals and data analytic experts to bring to bear the best of these fields in solving biomedical challenges such as the diagnose and prognosis of COVID-19 pandemic.

1.2 *Types of Transitions*

As noted earlier, there is a need to bring varying scientific and clinical expertise to develop better AI tools for COVID-19 diagnosis and prognosis, in which medical imaging will continue to play a pivotal role. Depending on the practitioner's background and expertise, there are three types of career transitions that can be extended from the COVID-19 AI era as discussed below and summarized in Fig. 2.

(1) Transition of an AI expert with no imaging background into AI of medical imaging data, which is typically the case of a computer scientist with interest in medical imaging.

(2) Transition of a medical imaging clinician (e.g., radiologist) into AI of medical imaging.

(3) Transition of combined (1) and (2) of AI of medical imaging into confronting an emerging challenge such as AI of COVID imaging.
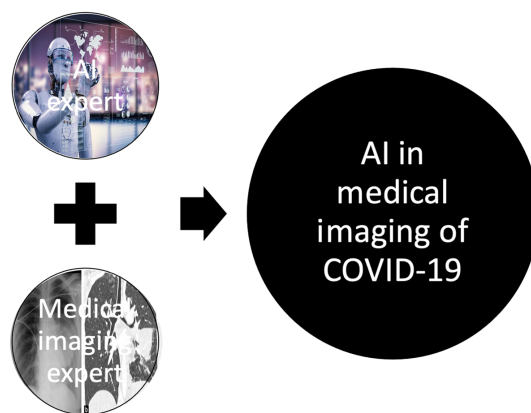
Fig. 2 Expertise transitions in AI of medical images of COVID-19.

2 Potential Use Cases and Issues

2.1 *AI of Non-Imaging Data to AI of Medical Imaging Data*

In addition to AI, knowledge of both medical presentation and imaging physics is necessary for successfully developing AI applications in medical imaging. This generally applies to computer scientists who are AI experts but with no previous background in imaging in general or medical imaging specifically. In this scenario, the AI expert may overly simplify the treatment of medical images, with their 2D and 3D pixels or components (i.e., slices), as deterministic independent data elements on which AI tools can be directly adapted, without realizing the intricates of medical imaging modalities, particularly various sources of variability that may affect the robustness/ generalizability of AI algorithms. For instance, akin to medical imaging are the distinct multi-dimensional nature of 2D (e.g., CXRs) or 3D images (e.g., CT), relative or absolute quantitative nature, the different manufacturers or models of image acquisition systems or the different acquisition parameters (e.g., mAs and kVp) for a particular system, image reconstruction methods (e.g., analytical or iterative), the need for clinical expertise in annotation/truthing (as compared with crowdsourcing in natural image annotation), and the specific clinical context that needs to be considered part of the evaluation process.[34] Among the many potential problems, we highlight the following cases due to their prevalence in the reported literature.

2.1.1 *Leakage between training and testing, in transitioning from AI to medical imaging AI*

Understanding the nature of medical images and the clinical context in which they are used is a necessary prerequisite for any successful application of an AI algorithm for diagnostic or prognostic medical purposes. How the training models should be built and how the data will be used in that process can be a main source of confusion and error and subsequent bias as well.[35] For instance, in a diagnostic learning task, a CT scan with many images could be erroneously perceived and then evaluated on a 2D slice-by-slice analysis based on an inaccurate assumption of independence between slices from the same patient while they are typically correlated. As such, slices from the same patient may be included in both training and testing sets, thereby yielding overly optimistic testing results due to this data leakage. Similarly, in the case of prognostic studies for predicting death or ventilation need, it is sometime unclear whether images from the same patient at different time points were mistakenly used for both training and testing.[36]

2.1.2 *Feature/model selection and multiple testing*

In traditional machine learning models based on feature engineering, feature selection is an essential process in the building of computational tools for the AI interpretation of medical images. This process is prone to statistical bias and potentially overly optimistic results if not

done properly; for example, if the entire dataset was first used for feature selection and then partitioned into apparently non-overlapping subsets for training and testing, the test results would be optimistically biased.[37] Proper methods should be used to avoid such pitfalls, such as the use of nested cross validation or other resampling or independent testing methods in algorithm development and evaluation.[38] Although the new generation of deep learning algorithms bypasses the feature selection process by embedding the data representation as part of the algorithm training itself [e.g., the convolutional layers in the convolutional neural networks (CNNs)], proper safeguards should still be applied when tuning a large number of hyperparameters in this case (i.e., model selection).

Another commonly encountered issue is false positive findings due to multiple testing, i.e., when multiple hypotheses are tested without appropriate statistical adjustment, a truly null hypothesis may be found significant due purely to chance. An infamous problem in this area was cited in the analysis of fMRI data from an image of a dead Atlantic salmon. Using standard acquisition, preprocessing, and analysis techniques, the authors showed that active voxel clusters could be observed in the dead salmon's brain when using uncorrected statistical thresholds for multiple comparisons, which led to such false positives, i.e., active voxels clusters in dead fish.[39] To avoid such pitfalls, appropriate statistical methods must be prespecified in a study to control the type I error rate.[40,41]

2.1.3 Erroneous labeling of the clinical task

This arises when there is no clarity in the description of the clinical problem.[29] For instance, a clinical problem might be posed as a diagnostic problem for differentiating between COVID-19 cases and control cases. In this scenario, the control cohort might be healthy individuals or might be individuals with other lung diseases such as pneumonia (in which presentation of the disease on the medical image might mimic that of COVID-19). Those two different control cohorts will yield a wide range of classification performances and, thus, confusion on the ultimate clinical use of the AI.

In another context, confusion can arise in the purpose of the AI algorithm. Is the AI algorithm for identifying which images have COVID-19 presentations (classification task) or is the AI algorithm for localizing regions of COVID-19 involvement within an image (segmentation task)? This can also manifest in the performance evaluation. These clinical tasks will require different evaluation metrics and clarification of sensitivity and/or specificity, especially when imbalances arise in the dataset.

2.1.4 Mismatches between training data and clinical task

An AI algorithm trained in a non-medical context may not transfer well to medical imaging AI, in part due to restrictive deidentification or HIPAA concerns, which may limit some of the data availability. For instance, the required data may be fragmented or distributed across multiple systems. This is in addition to issues related to the need for proper data curation processes to ensure that the necessary quality is available. Different hospitals may apply their CT protocols differently such as variations in image acquisition systems and protocols (i.e., slice thickness and reconstruction) indicating the need for harmonization. Otherwise, the process may be subject to the "garbage in garbage out" pitfall.

On the other hand, transfer learning,[42,43] with fine tuning and feature extraction, has proven to be quite useful in AI of medical imaging, especially for tasks related to image analysis such as segmentation, denoising, and classification (leading to diagnosis and other tasks). Therefore, to mitigate the problem of mismatch between the intended training process and the clinical task, there needs to be transparent communication between the AI and the medical imaging domain experts.

2.1.5 Role of continuous learning

Several dynamic algorithms such as reinforcement learning tend to continuously update their training by exploring the environment, which could work well in a safe setting such as a board

game but may have legal and ethical ramifications in the real world (e.g., self-driving accidents) and especially in the context of medicine, where life or death decisions are being made. The traditional approach to "continuous" learning is implemented through manually controlled periodic updates to "locked" algorithms.[44] Recent technology advances in continuous learning AI have the potential to provide fully automatic updates by a computer with the promise of adapting continuously to changes in local data.[45] However, how to assess such an adaptive AI system to ensure safety and effectiveness in the real world is an open question, and there are several ongoing research efforts and discussions on this topic across academia and regulatory agencies to ensure that the deployed AI algorithm is still correctly performing the originally intended task.[44–46] This process may require an AI monitoring program and periodic evaluations over the lifetime of the algorithm.

2.2 Medical Imaging Clinician to AI of Medical Imaging

The ready availability and user-friendliness of recent deep learning software codes have allowed novice to experts in AI to rapidly develop, apply, and/or evaluate AI on sets of medical imaging data. However, lack of sufficient technical expertise on medical image acquisition, subtleties in AI algorithm engineering, and statistics may result in potentials for bias, inappropriate statistical tests, and erroneous interpretations of findings.

2.2.1 Application of pretrained models

As machine learning methods are increasingly developed for medical imaging applications, there are numerous models made available by investigators either publicly or through collaborations. However, direct applications of off-the-shelf models are often unsuccessful. Needless to say, models trained for a specific task, e.g., detecting a certain disease on a certain imaging modality, are not suitable for being directly applied to a different task, because useful features for performing tasks can be vastly different. Moreover, even when the task in question remains the same, directly applying off-the-shelf models developed on one dataset often fails to generalize their performance when applied to another dataset. There are several reasons for the lack of robustness across different datasets for machine learning models in medical imaging applications, including differences in imaging acquisition protocols, scanners, patient populations, and image preprocessing methods. Features extracted by machine learning models, both human-engineered features and automatically learned features by deep learning, are often sensitive to these differences between the source domain and the target domain. Therefore, it is good practice to preprocess the images in the target domain the same way as for the development set, harmonize differences in images resulting from image acquisition and population characteristics, and fine-tune the pretrained model on a subset of the images in the target domain when appropriate.

Another significant issue is cross-institution generalizability of deep learning CNNs, i.e., a CNN trained with data from one clinical site or institution is found to have a significant drop in performance on data from another site. This has been frequently reported in the recent literature, for instances, on the classification of abnormal chest radiographs,[47] on the diagnosis of pneumonia,[48] and on the detection of COVID-19 from chest radiographs.[49] In addition to the possible reasons discussed in the previous paragraph, a particular issue found by these authors is that CNNs have readily learned features specific to sites (such as metal tokens placed on patients at a specific site) rather than the pathology information in the image. As such, proper testing procedures that mimic real-world implementation are crucial before any AI clinical deployment.[50]

2.2.2 Use of texture analysis software packages

Often off-the-shelf texture analysis software packages are used to quantify the variation perceived in a medical image, the output of which is subsequently related to disease states using AI algorithms.[51] Assessment of texture provides quantitative values of pattern intensity, spatial frequency content (i.e., coarseness), and directionality.[52] Caution is needed, however, since, as noted by Foy et al.,[51] algorithm parameters, such as those used for normalization of pixel values

prior to processing, can vary between software packages that are calculating the same mathematical feature. An example is in the calculation of texture features from gray-level co-occurrence matrices (GLCM), in which GLCM parameters may be unique to the specific software package. Entropy, for example, involves the calculation of a logarithm, which might be to the base e, the base 2, or the base 10, with each yielding different output texture values of entropy.[51] There have been attempts to standardize the calculation of radiomic features.[53] It is important to understand both the mathematical formulation of the radiomic feature as well as the software package implementation to achieve the intended results as recommended by the Image Biomarker Standardization Initiative.[53]

2.2.3 Overfitting and bias in development of AI systems

The novice developer of AI should be careful to avoid overfitting and bias in AI algorithms, often caused by limited datasets in terms of number, distribution, and collection sites of cases.[54] While even independent testing using cases from a single institution may yield promising performance levels, the AI algorithm may not be assumed to be generalizable until it is tested on an independent multi-institutional dataset with results given in terms of performance in the specific task including its variation (reproducibility and repeatability) as well as assessment of whether or not the algorithms are biased. Sharing of data in such instances may be challenging; therefore, the availability of public resources such as the Medical Imaging Data and Resource Center (MIDRC) or the utilization of emerging methods based on federated learning would be helpful in this process.[55]

2.3 AI of Medical Imaging to AI of COVID-19 Imaging

The situation of an AI/CAD researcher with medical imaging experience transitioning to COVID-19 applications is anticipated to be the most likely of the aforementioned scenarios. The researcher is generally familiar with common radiological practices, algorithms, and applications of AI to medical images but has focused on other problems such as deep learning applied to image reconstruction or interpretation for non-COVID-19 disease. However, there are aspects of COVID-19 AI research that may be unfamiliar to the researcher. We discuss three instances here but note that there are other unique aspects to COVID-19 imaging; these three were selected due to their frequent occurrence and interest.

2.3.1 Observer variability of COVID-19

Observer variability refers to the concept of different readers assessing the same input image and producing different results.[56,57] While not uncommon in medical imaging, the degree to which observer variability may impact an AI model either during training or evaluation is dependent on the imaging task.[58–62] For example, in COVID-19 detection problems, the standard ground truth identifying disease positivity is RT-PCR, which may have less variability compared with imaging.[63,64] Alternatively, medical image segmentation tasks generally utilize a ground truth mask delineated by expert radiologists. Several studies have analyzed observer variability for tasks related to COVID-19 assessment, including disease detection, diagnosis, segmentation, and prognosis, including detection of abnormalities on CXRs and segmentation of disease on CT scans, even in other diagnostic tests such as RT-PCR.[63–69] Methods that can alleviate complications related to observer variability include statistical resampling techniques, such as bootstrapping or imbalance correction such as synthetic minority over-sampling technique and using appropriate evaluation metrics such as precision–recall and receiver operating characteristic analyses.[70,71] Furthermore, implementation of clearly defined task-relevant scores such as CO-RADS can standardize evaluations and reduce the impact of such interobserver variability.[72,72]

2.3.2 Understanding COVID-19 imaging protocols

Most diseases have existed for many years and thus have undergone extensive investigation of determining appropriate imaging protocols for detection, diagnosis, and prognosis evaluations.

However, protocols for COVID-19 patient imaging have rapidly changed over the past year as researchers investigated optimal practices for a variety of situational evaluations.[73] Further, different radiological societies may have inconsistent or conflicting recommendations based on regional and national standards.[74–80] The effect of these phenomena is twofold. (1) Existing COVID-19 imaging datasets may contain images that would not be appropriate for general application due to outdated image acquisition protocols and thus should not be included in AI training without careful curation, especially if acquired externally to the region or institution in which the AI system will be utilized. (2) Task selection for an AI system should be performed based on local practice, which in turn should be based on radiological society or WHO recommendations. For example, most imaging societies do not recommend the use of CT for COVID-19 screening due to the increased patient dose, with likely no information gain compared with RT-PCR and chest radiography; thus AI algorithms trained for COVID-19 detection may prioritize detection on CXR rather than CT for widespread implementation. While these studies may still provide some value, AI algorithms should generally be developed with appropriate task selection and data curation that is relevant with current image acquisition guidelines for their intended application.

2.3.3 *Biased COVID-19 datasets and models*

One of the most common obstacles in training unbiased medical imaging AI systems is insufficient relevant patient imaging data.[81] This obstacle is exacerbated due to the brief existence of COVID-19 as the total pipeline of image acquisition, curation, and labeling for ML usage can be time-consuming. During the pandemic, several teams have attempted to utilize publicly available datasets of both COVID-19 and non-COVID-19 patient images to develop generalized models for clinical use.[35] Thus two key problems currently are posed: (1) data, particularly private data, are generally skewed such that the amount of COVID-19 positive data is notably smaller than COVID-19 negative data and (2) the use of publicly available datasets can lead to a high risk of model bias, either due to a lack of information (such as consistent confirmation of COVID-19 diagnosis through RT-PCR and/or other clinical tests) or through crossover of images between databases leading to biased evaluations.[64] Further, some public datasets include images not in DICOM format (e.g., images scrubbed from published papers). Thus it is critically important to understand the source of the imaging data and evaluate appropriately. For example, evaluation should be performed on completely independent private (when possible) and public datasets, such as the image testing data in the MIDRC.

3 AI Using Clinical, Histopathology, and Genomics with Imaging Data

In this section, we discuss the transition from AI of medical imaging to AI of COVID-19 imaging in the clinical context of pan- or multi-omics.[82] Clinicians nowadays are equipped with a wealth of information that may include patient demographics and history, routine biochemical and hematological data, genetic/genomic testing, and multi-modality imaging exams. Two tracks of research in such a clinical context hold great promise to benefit patients and our healthcare system: correlation and integration. Correlation of different types of information not only may offer better understanding of the disease characteristics but also may allow for replacement of invasive/expensive testing with non-invasive and less costly imaging exams or identify the role of imaging in different clinical tasks or stages of patient management. On the other hand, when multiple types of measurements are found to be complementary, optimal integration of them is of great interest for better patient outcomes.[83] Despite these potential benefits, lessons have been learned from the pre-COVID-19 era mostly in the area of AI for cancer diagnosis and treatment:[32] limited datasets and lack of validation data, lack of expandability, and the need for rigorous statistical analysis, to name a few.[26,84–86] Like many other diseases,[52,87–90] there are many clinical tasks related to COVID-19 that can be interrogated using information extracted from medical images. Below we discuss the potential role of AI for COVID-19 imaging in the clinical context of integrative modeling with pan- or multi-omics.

3.1 *Radiomic Analysis*

3.1.1 *COVID-19 detection/diagnosis*

Although RT-PCR is the reference standard for COVID-19 diagnosis, molecular testing may not be readily available or may lead to false negative results for patients under investigation of COVID-19, and thus under these circumstances, CXR/CT can be critically useful in the diagnosis of the disease. CXR/CT in COVID-19 diagnosis has moderate sensitivity and specificity[91] in current radiological practice possibly because of clinicians' limited experience in reading COVID-19 cases. AI/ML models can be trained using these images with RT-PCR as the reference standard. These trained models can be applied to patients to make rapid diagnosis when only imaging data are available.[92–94] AI/ML models have been shown to be able to help clinicians differentiate COVID-19 induced pneumonia from other types of pneumonia, including common and unique radiographic features differentiation.[26] Thus AI/ML models including both deep learning and human-engineered radiomics have the potential to aid clinicians with improved diagnostic performance.

3.1.2 *COVID-19 severity assessment/prognosis*

Baseline pulmonary CXR/CT imaging characteristics can be used to assess the severity/extent of disease.[95] A severity score derived from lung images can be used for triaging the patient to decide home discharge, regular hospital admission, or intensive care unit admission.[96] The imaging characteristics can be used for prognosis assessment[97] and evaluation of the extent of disease,[11] and it has potential to predict the length of hospital stay.[98]

Although COVID-19 is seen as a disease that primarily affects the lungs, it may also damage other organs, including heart[93,99] and brain,[93,99–102] and many long-term COVID-19 effects are still unknown. The COVID-19 patients have to be closely monitored even after their recovery, when imaging and AI may play a pivotal role.

3.1.3 *Response to treatment*

Baseline pulmonary CXR/CT imaging characteristics can also be used for patient management to determine the therapeutic treatment plan. The temporal changes can be used to monitor the patients' response to the treatment and to inform the decision for discharge.[97] Lung ultrasound has also been suggested as the imaging tool for COVID-19 pandemic in disease triage, diagnosis, prognosis, severity assessment, and monitoring the treatment response.[13]

3.2 *Radiogenomic Analysis*

Radiomics is high-throughput extraction of a large number of quantitative features from medical images.[103] By integrating quantitative image features with clinical, pathological, and genomic data, i.e., radiogenomics, we will be able to study the relationship or association between imaging-based phenotypes and genomic biomarkers. This would potentially allow us to discover radiomic features that can be used as diagnostic or prognostic biomarkers for monitoring the patients, assessing response to the treatment, and predicting the risk of recurrence. Studying the association between image-based phenotypes and underlying genetic mechanisms[104] may help better understand the disease mechanism and potentially precision medicine of COVID-19 or similar diseases in the future.

However, small datasets will be the challenge for radiogenomic studies since there are only limited cases with comprehensive data, including imaging, clinical, pathological, genomic, and other-omic data available for this relatively new disease. In addition, the batch effect has to be considered with diverse data with both imaging and genomic data. Data harmonization/calibration has to be performed prior to any radiogenomic analysis. Moreover, some data resources such as lung pathology images are limited to postmortem evaluations.[105] Nevertheless, radiogenomics is a valuable tool for linking imaging finding to the underlying biological mechanisms of disease.
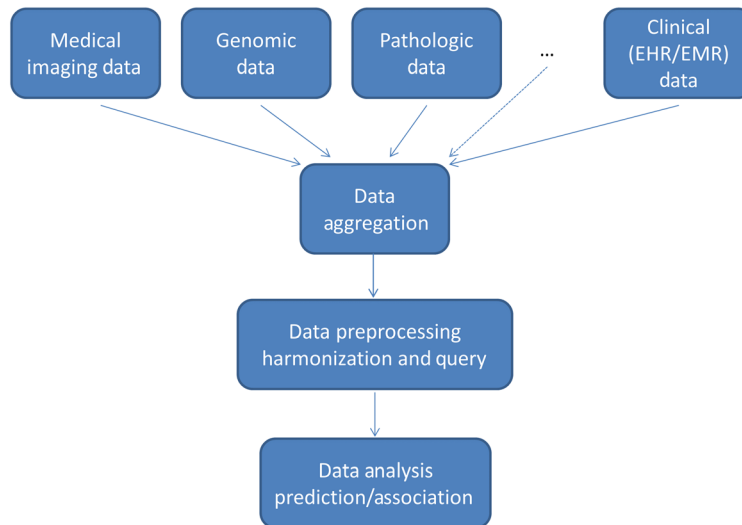
Fig. 3 Schematic diagram of multi-omics analysis.

3.3 *How to Pull Clinical Data from Different Sources*

Medical imaging data are well structured and stored with an industry standard, such as DICOM format. However, other required information, including clinical records, genomics, and pathological data, which are required to carry out scientific investigation and answer relevant clinical questions, is not as well structured as medical images. Major efforts are needed to extract relevant information from these different sources to solve clinical questions and advance the knowledge to better understand the disease. The preprocessing, such as data aggregation and data harmonization, is needed for these not-well-structured data to be suitable for use in the subsequent AI development.[106]

Medical imaging data, even if in DICOM format, may still need to be harmonized since imaging data collected from different institutions may be acquired using different acquisition protocols and using machines from different manufacturers. Pathology slides have varying staining protocols and imaging formats; however, advances in digital pathology promise to alleviate many of these issues.[107] Genomic data from different sources may need to be harmonized/corrected to minimize the batch effect. Natural language processing algorithms may be needed to efficiently process unstructured clinical records.[108] All of these preprocessing and query procedures must be done prior to performing any prediction and association analysis to advance the knowledge of disease (Fig. 3).

4 Conclusions

Though several recommendations have emerged to improve the reporting of AI for biomedical applications[109–113] and imaging specifically,[114] little have been yet developed on the career transition process across radiology, medical imaging physics, and computer science. We recognize the great potential of imaging and AI in a variety of applications related to COVID-19 (such as those discussed in Sec. 3), and three transitions were identified with the following precautions.

1. AI of non-imaging data to AI of medical imaging data

   - The imaging nature (multi-slice) may increase the risk of leakage between training and testing sets and should be cautiously addressed.
   - Though feature and model selections are similar across AI algorithms, given the large number of free and intercorrelated parameters in imaging applications, they would require additional safeguards to avoid bias.
   - Imaging tasks can share similar synonymous terminology that would require comprehensive understanding for successful implementation. This also needs to be addressed in the development and preparation of the corresponding training datasets.

- The nature of medicine is that its data continue to evolve and change over time, which imposes further needs for continuous learning protocols.

2. Medical imaging clinician to AI of medical imaging

- The application of pretrained models (transfer learning) needs to be consistent and informed by the clinical task at hand.
- Packaged AI and texture analysis, though they are handy, would require training before their proper use in clinical applications.
- AI systems are prone to overfitting and bias; therefore, dedicated training and consultation with experts is helpful to ensure generalizability and reproducibility.

3. AI of medical imaging to AI of COVID-19 imaging

- COVID-19 is a complex disease and observer variability should be considered in the application of AI algorithms.
- Understanding COVID-19 imaging protocols is needed to mitigate acquisition inconsistencies and to properly harmonize data.
- There are many different COVID-19 datasets that are available; however, little standardization had gone into them, which would require proper preprocessing for successful application to yield a reproducible AI model.

Disclosures

M. L. G. is a stockholder in R2 Technology/Hologic and QView, receives royalties from Hologic, GE Medical Systems, MEDIAN Technologies, Riverain Medical, Mitsubishi, and Toshiba, and was a cofounder in Quantitative Insights (now consultant to Qlarity Imaging). H. L. receives royalties from Hologic. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interest that would reasonably appear to be directly and significantly affected by the research activities. I. E. N. is a deputy editor for Medical Physics and reports a relationship with Scientific Advisory Endectra, LLC.

Acknowledgments

References

1. D. C. Angus, A. C. Gordon, and H. Bauchner, "Emerging lessons from COVID-19 for the US clinical research enterprise," *JAMA* **325**(12), 1159–1161 (2021).
2. C. Del Rio and P. Malani, "COVID-19 in 2021-continuing uncertainty," *JAMA* **325**(14), 1389–1390 (2021).
3. G. Kampf and M. Kulldorff, "Calling for benefit-risk evaluations of COVID-19 control measures," *Lancet* **397**(10274), 576–577 (2021).
4. B. M. Kuehn, "Most patients hospitalized with COVID-19 have lasting symptoms," *JAMA* **325**(11), 1031 (2021).
5. C. J. L. Murray and P. Piot, "The potential future of the COVID-19 pandemic: will SARS-CoV-2 become a recurrent seasonal infection?" *JAMA* **325**(13), 1249–1250 (2021).
6. G. R. Wilensky, "2020 revealed how poorly the US was prepared for COVID-19-and future pandemics," *JAMA* **325**(11), 1029–1030 (2021).

7. A. Jacobi et al., "Portable chest x-ray in coronavirus disease-19 (COVID-19): a pictorial review," *Clin. Imaging* **64**, 35–42 (2020).

8. H. Y. F. Wong et al., "Frequency and distribution of chest radiographic findings in patients positive for COVID-19," *Radiology* **296**(2), E72–E78 (2020).

9. T. C. Kwee and R. M. Kwee, "Chest CT in COVID-19: what the radiologist needs to know," *Radiographics* **40**(7), 1848–1865 (2020).

10. H. Shi et al., "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study," *Lancet Infect. Dis.* **20**(4), 425–434 (2020).

11. W. Zhao et al., "Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study," *Am. J. Roentgenol.* **214**(5), 1072–1077 (2020).

12. A. W. E. Lieveld et al., "Diagnosing COVID-19 pneumonia in a pandemic setting: lung ultrasound versus CT (LUVCT)—a multicentre, prospective, observational study," *ERJ Open Res.* **6**(4), 00539–02020 (2020).

13. G. Soldati et al., "Is there a role for lung ultrasound during the COVID-19 pandemic?" *J. Ultrasound Med.* **39**(7), 1459–1462 (2020).

14. J. C. G. de Alencar et al., "Lung ultrasound score predicts outcomes in COVID-19 patients admitted to the emergency department," *Ann. Intensive Care* **11**, 6 (2021).

15. Y. Zhang et al., "Lung ultrasound findings in patients with coronavirus disease (COVID-19)," *Am. J. Roentgenol.* **216**(1), 80–84 (2021).

16. B. S. Tan et al., "RSNA international trends: a global perspective on the COVID-19 pandemic and radiology in late 2020," *Radiology* **299**, 204267 (2020).

17. A. Jain et al., "Imaging of coronavirus disease (COVID-19), a pictorial review," *Pol. J. Radiol.* **86**, e4–e18 (2021).

18. R. M. Kwee, H. J. A. Adams, and T. C. Kwee, "Chest CT in patients with COVID-19: toward a better appreciation of study results and clinical applicability," *Radiology* **298**(2), E113–E114 (2021).

19. N. Lessmann et al., "Automated assessment of COVID-19 reporting and data system and chest CT severity scores in patients suspected of having COVID-19 using artificial intelligence," *Radiology* **298**(1), E18–E28 (2021).

20. A. Mirahmadizadeh et al., "Sensitivity and specificity of chest computed tomography scan based on RT-PCR in COVID-19 diagnosis," *Pol. J. Radiol.* **86**, e74–e77 (2021).

21. S. Schiaffino et al., "CT-derived chest muscle metrics for outcome prediction in patients with COVID-19," *Radiology* **300**, 204141 (2021).

22. M. Cellina, M. Panzeri, and G. Oliva, "Chest radiography features help to predict a favorable outcome in patients with coronavirus disease 2019," *Radiology* **297**(1), E238 (2020).

23. S. E. Miller, "Visualization of SARS-CoV-2 in the lung," *N. Engl. J. Med.* **383**(27), 2689 (2020).

24. H. N. Şendur, "Debate of chest CT and RT-PCR test for the diagnosis of COVID-19," *Radiology* **297**(3), E341–E342 (2020).

25. R. Zhang et al., "Diagnosis of coronavirus disease 2019 pneumonia by using chest radiography: value of artificial intelligence," *Radiology* **298**(2), E88–E97 (2021).

26. H. X. Bai et al., "Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT," *Radiology* **296**(3), E156–E165 (2020).

27. B. van Ginneken, "The potential of artificial intelligence to analyze chest radiographs for signs of COVID-19 pneumonia," *Radiology* **299**, 204238 (2020).

28. R. M. Wehbe 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* **299**, 203511 (2020).

29. R. M. Summers, "Artificial intelligence of COVID-19 imaging: a hammer in search of a nail," *Radiology* **298**(3), E162–E164 (2021).

30. V. Parcha et al., "A retrospective cohort study of 12,306 pediatric COVID-19 patients in the United States," *Sci. Rep.* **11**(1), 10231 (2021).

31. C. O. Serrano et al., "Pediatric chest x-ray in covid-19 infection," *Eur. J. Radiol.* **131**, 109236 (2020).

32. I. El Naqa et al., "Artificial Intelligence: reshaping the practice of radiological sciences in the 21st century," *BJR* **93**(1106), 20190855 (2020).

33. J. D. López-Cabrera et al., "Current limitations to identify COVID-19 using artificial intelligence with chest x-ray imaging," *Health Technol.* **11**(2), 411–424 (2021).

34. W. L. Bi et al., "Artificial intelligence in cancer imaging: clinical challenges and applications," *CA Cancer J. Clin.* **69**(2), 127–157 (2019).

35. M. Roberts et al., "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans," *Nat. Mach. Intell.* **3**(3), 199–217 (2021).

36. L. Wynants et al., "Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal," *BMJ* **369**, m1328 (2020).

37. R. Simon et al., "Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification," *J. Natl. Cancer Inst.* **95**(1), 14–18 (2003).

38. Y. Luo et al., "Development of a fully cross-validated Bayesian network approach for local control prediction in lung cancer," *IEEE Trans. Radiat. Plasma Med. Sci.* **3**, 232–241 (2019).

39. C. M. Bennett, G. L. Wolford, and M. B. Miller, "The principled control of false positives in neuroimaging," *Soc. Cogn. Affect Neurosci.* **4**(4), 417–422 (2009).

40. S. Y. Chen, Z. Feng, and X. Yi, "A general introduction to adjustment for multiple comparisons," *J. Thorac Dis.* **9**(6), 1725–1729 (2017).

41. A. Farcomeni, "A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion," *Stat. Methods Med. Res.* **17**(4), 347–388 (2008).

42. H. P. Chan et al., "Deep learning in medical image analysis," *Adv. Exp. Med. Biol.* **1213**, 3–21 (2020).

43. H. Ravishankar et al., "Understanding the mechanisms of deep transfer learning for medical images," in *Proc. Int. Workshop Large-Scale Annotation Biomed. Data Expert Label Syn.*, Cham (2016).

44. "Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)—discussion paper and request for feedback."

45. O. S. Pianykh et al., "Continuous learning AI in radiology: implementation principles and early applications," *Radiology* **297**(1), 6–14 (2020).

46. J. Feng, S. Emerson, and N. Simon, "Approval policies for modifications to machine learning-based software as a medical device: a study of bio-creep," *Biometrics* **77**(1), 31–44 (2021).

47. I. Pan, S. Agarwal, and D. Merck, "Generalizable inter-institutional classification of abnormal chest radiographs using efficient convolutional neural networks," *J. Digit. Imaging* **32**(5), 888–896 (2019).

48. J. R. Zech et al., "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study," *PLoS Med.* **15**(11), e1002683 (2018).

49. K. B. Ahmed et al., "Discovery of a generalization gap of convolutional neural networks on COVID-19 x-rays classification," *IEEE Access* **9**, 72970–72979 (2021).

50. I. El Naqa, "Prospective clinical deployment of machine learning in radiation oncology," *Nat. Rev. Clin. Oncol.* **18**, 605–606 (2021).

51. J. J. Foy et al., "Variation in algorithm implementation across radiomics software," *J. Med. Imaging* **5**(4), 044505 (2018).

52. M. L. Giger, N. Karssemeijer, and J. A. Schnabel, "Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer," *Annu. Rev. Biomed. Eng.* **15**, 327–357 (2013).

53. A. Zwanenburg et al., "The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology* **295**(2), 328–338 (2020).

54. C. P. Langlotz et al., "A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/the academy workshop," *Radiology* **291**(3), 781–791 (2019).

55. M. J. Sheller et al., "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Sci. Rep.* **10**(1), 12598 (2020).
56. P. Brennan and A. Silman, "Statistical methods for assessing observer variability in clinical measures," *BMJ* **304**(6840), 1491–1494 (1992).
57. Z. B. Popović and J. D. Thomas, "Assessing observer variability: a user's guide," *Cardiovasc. Diagn. Ther.* **7**(3), 317–324 (2017).
58. K. D. Hopper et al., "Analysis of interobserver and intraobserver variability in CT tumor measurements," *Am. J. Roentgenol.* **167**(4), 851–854 (1996).
59. L. Joskowicz et al., "Inter-observer variability of manual contour delineation of structures in CT," *Eur. Radiol.* **29**(3), 1391–1399 (2019).
60. S. E. Luijnenburg et al., "Intra-observer and interobserver variability of biventricular function, volumes and mass in patients with congenital heart disease measured by CMR imaging," *Int. J. Cardiovasc. Imaging* **26**(1), 57–64 (2010).
61. A. McErlean et al., "Intra- and interobserver variability in CT measurements in oncology," *Radiology* **269**(2), 451–459 (2013).
62. S. J. van Riel et al., "Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management," *Radiology* **277**(3), 863–871 (2015).
63. Y. Fang et al., "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," *Radiology* **296**(2), E115–E117 (2020).
64. H. Kim, H. Hong, and S. H. Yoon, "Diagnostic performance of CT and reverse transcriptase polymerase chain reaction for coronavirus disease 2019: a meta-analysis," *Radiology* **296**(3), E145–E155 (2020).
65. D. Byrne et al., "RSNA expert consensus statement on reporting chest CT findings related to COVID-19: interobserver agreement between chest radiologists," *Can. Assoc. Radiol. J.* **72**(1), 159–166 (2021).
66. F. Ciccarese et al., "Diagnostic accuracy of North America Expert Consensus Statement on reporting CT findings in patients suspected of having COVID-19 infection: an Italian Single-Center Experience," *Radiol. Cardiothor. Imaging* **2**(4), e200312 (2020).
67. Z. Li 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.* **30**(12), 6828–6837 (2020).
68. Y. Lichter et al., "Lung ultrasound predicts clinical course and outcomes in COVID-19 patients," *Intensive Care Med.* **46**(10), 1873–1883 (2020).
69. S. Rajaraman et al., "Analyzing inter-reader variability affecting deep ensemble learning for COVID-19 detection in chest radiographs," *PLoS One* **15**(11), e0242301 (2020).
70. N. Japkowicz and M. Shah, *Evaluating Learning Algorithms: A Classification Perspective*, Cambridge University Press, Cambridge, New York (2011).
71. N. V. Chawla et al., "SMOTE: synthetic minority over-sampling technique," *J. Artif. Intell. Res.* **16**, 321–357 (2002).
72. M. Prokop et al., "CO-RADS: a categorical CT assessment scheme for patients suspected of having COVID-19—definition and evaluation," *Radiology* **296**(2), E97–E104 (2020).
73. E. A. Akl et al., "Use of chest imaging in the diagnosis and management of COVID-19: a WHO rapid advice guide," *Radiology* **298**(2), E63–E69 (2021).
74. "The role of CT in patients suspected with COVID-19 infection," The Royal College of Radiologists.
75. "ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection."
76. L. Fan and S. Liu, "CT and COVID-19: Chinese experience and recommendations concerning detection, staging and follow-up," *Eur. Radiol.* **30**, 5214–5216 (2020).
77. A. M. Foust et al., "International expert consensus statement on chest imaging in pediatric COVID-19 patient management: imaging findings, imaging study reporting, and imaging study recommendations," *Radiol. Cardiothor. Imaging* **2**(2), e200214 (2020).
78. G. D. Rubin et al., "The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society," *Radiology* **296**(1), 172–180 (2020).

79. S. Simpson et al., "Radiological Society of North America expert consensus document on reporting chest CT findings related to COVID-19: endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA," *Radiol. Cardiothor. Imaging* **2**(2), e200152 (2020).

80. Q. Yang et al., "Imaging of coronavirus disease 2019: a Chinese expert consensus statement," *Eur. J. Radiol.* **127**, 109008 (2020).

81. L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," arXiv:171204621 (2017).

82. I. El Naqa, "Biomedical informatics and panomics for evidence-based radiation therapy," *WIREs Data Min. Knowl. Discovery* **4**(4), 327–340 (2014).

83. S. Cui, R. K. T. Haken, and I. El Naqa, "Integrating multiomics information in deep learning architectures for joint actuarial outcome prediction in non-small cell lung cancer patients after radiation therapy," *Int. J. Radiat. Oncol. Biol. Phys.* **110**(3), 893–904 (2021).

84. M. Avanzo et al., "Machine and deep learning methods for radiomics," *Med. Phys.* **47**(5), e185–e202 (2020).

85. I. El Naqa et al., "Machine learning and modeling: data, validation, communication challenges," *Med. Phys.* **45**(10), e834–e840 (2018).

86. I. El Naqa et al., "Radiation therapy outcomes models in the era of radiomics and radiogenomics: uncertainties and validation," *Int. J. Radiat. Oncol. Biol. Phys.* **102**(4), 1070–1073 (2018).

87. N. Emaminejad et al., "Fusion of quantitative image and genomic biomarkers to improve prognosis assessment of early stage lung cancer patients," *IEEE Trans. Biomed. Eng.* **63**(5), 1034–1043 (2016).

88. M. A. Mazurowski et al., "Predicting outcomes in glioblastoma patients using computerized analysis of tumor shape: preliminary data," *Proc. SPIE* **9785**, 97852T (2016).

89. Y. Peng et al., "Validation of quantitative analysis of multiparametric prostate MR images for prostate cancer detection and aggressiveness assessment: a cross-imager study," *Radiology* **271**(2), 461–471 (2014).

90. H. Yoshida et al., "Computerized detection of colonic polyps at CT colonography on the basis of volumetric features: pilot study," *Radiology* **222**(2), 327–336 (2002).

91. S. H. Yoon et al., "Chest radiographic and CT findings of the 2019 novel coronavirus disease (COVID-19): analysis of nine patients treated in Korea," *Korean J. Radiol.* **21**(4), 494 (2020).

92. X. Mei et al., "Artificial intelligence-enabled rapid diagnosis of patients with COVID-19," *Nat. Med.* **26**(8), 1224–1228 (2020).

93. Y.-Y. Zheng et al., "COVID-19 and the cardiovascular system," *Nat. Rev. Cardiol.* **17**(5), 259–260 (2020).

94. Q. Hu, K. Drukker, and M. L. Giger, "Role of standard and soft tissue chest radiography images in COVID-19 diagnosis using deep learning," *Proc. SPIE* **11597**, 1159704 (2021).

95. Z. Tang et al., "Severity assessment of COVID-19 using CT image features and laboratory indices," *Phys. Med. Biol.* **66**(3), 035015 (2021).

96. H. W. Kim et al., "The role of initial chest x-ray in triaging patients with suspected COVID-19 during the pandemic," *Emerg. Radiol.* **27**(6), 617–621 (2020).

97. J. D. Fuhrman et al., "Cascaded deep transfer learning on thoracic CT in COVID-19 patients treated with steroids," *J. Med. Imaging* **8**(S1), 014501 (2021).

98. A. Bernheim et al., "Chest CT findings in coronavirus disease-19 (COVID-19), relationship to duration of infection," *Radiology* **295**(3), 200463 (2020).

99. E. J. Topol, "COVID-19 can affect the heart," *Science* **370**(6515), 408–409 (2020).

100. A. Mahammedi et al., "Brain and lung imaging correlation in patients with COVID-19: could the severity of lung disease reflect the prevalence of acute abnormalities on neuroimaging? A global multicenter observational study," *AJNR Am. J. Neuroradiol.* **42**(6) 1008–1016 (2021).

101. L. Mao et al., "Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan, China," *JAMA Neurol.* **77**(6), 683 (2020).

102. M. Marshall, "How COVID-19 can damage the brain," *Nature* **585**(7825), 342–343 (2020).

103. R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology* **278**(2), 563–577 (2016).

104. Y. Zhu et al., "Deciphering genomic underpinnings of quantitative MRI-based radiomic phenotypes of invasive breast carcinoma," *Sci. Rep.* **5**(1), 17787 (2015).

105. F. Calabrese et al., "Pulmonary pathology and COVID-19: lessons from autopsy. The experience of European Pulmonary Pathologists," *Virchows Arch.* **477**(3), 359–372 (2020).

106. D. B. Larson et al., "Regulatory frameworks for development and evaluation of artificial intelligence-based diagnostic imaging algorithms: summary and recommendations," *J. Am. Coll. Radiol.* **18**(3), 413–424 (2021).

107. S. W. Jahn, M. Plass, and F. Moinfar, "Digital pathology: advantages, limitations and emerging perspectives," *J. Clin. Med.* **9**(11), 3697 (2020).

108. S. Locke et al., "Natural language processing in medicine: a review," *Trends Anaesthesia Critical Care* **38**, 4–9 (2021).

109. G. S. Collins and K. G. M. Moons, "Reporting of artificial intelligence prediction models," *Lancet* **393**(10181), 1577–1579 (2019).

110. B. Norgeot et al., "Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist," *Nat. Med.* **26**(9), 1320–1324 (2020).

111. S. C. Rivera et al., "Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension," *Lancet Digit. Health* **2**(10), e549–e560 (2020).

112. V. Sounderajah et al., "Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: the STARD-AI steering group," *Nat. Med.* **26**(6), 807–808 (2020).

113. I. El Naqa et al., "AI in medical physics: guidelines for publication," *Med. Phys.* **48**(9), 4711–4714 (2021).

114. J. Mongan et al., "Checklist for artificial intelligence in medical imaging (CLAIM), a guide for authors and reviewers," *Radiol. Artif. Intell.* **2**(2), e200029 (2020).

Biographies of the authors are not available.