ScienceDirect

On the Role of Artificial Intelligence in Medical Imaging of COVID-19

Jannis Born[1,2,], David Beymer[3,*], Deepta Rajan[3], Adam Coy[3,4], Vandana V. Mukherjee[3], Matteo Manica[1], Prasanth Prasanna[5,3], Deddeh Ballah[6,3], Pallav L. Shah[7], Emmanouil Karteris[8], Jan L. Robertus[9], Maria Gabrani[1], Michal Rosen-Zvi[10]*

[1]IBM Research Europe, Zurich, Switzerland
[2]Department for Biosystems Science & Engineering, ETH Zurich, Switzerland
[3]IBM Almaden Research Center, CA, USA
[4]Vision Radiology, Dallas, TX, USA
[5]Department of Radiology and Imaging Sciences, University of Utah Health Sciences Center, Salt Lake City, Utah
[6]Department of Radiology, Seton Medical Center, Daly City, CA, USA
[7]NIHR Respiratory Biomedical Research Unit, Imperial College, London, UK
[8]College of Health, Medicine and Life Sciences, Brunel University London, UK
[9]National Heart & Lung Institute, Imperial College London, UK
[10]IBM Research Haifa, Israel
* Corresponding authors: jab@zurich.ibm.com, beymer@us.ibm.com

ARTICLE INFO

Article history:
Received 00 December 00
Received in revised form 00 January 00
Accepted 00 February 00

Keywords:
COVID-19
Medical imaging
Artificial intelligence

ABSTRACT

During the COVID-19 pandemic, lung imaging takes a key role in addressing the magnified need of speed, cost, ubiquity and precision in medical care. The rise of artificial intelligence induced a quantum leap in medical imaging: AI has now proven equipollent to healthcare professionals in several diseases and the potential to save time, cost and increase coverage. But AI-accelerated medical imaging must still fully demonstrate its ability in remediating diseases such as COVID-19.

We identify key use cases of lung imaging for COVID-19, comparing CT, X-Ray and ultrasound imaging from clinical and AI perspectives. We perform a systematic, manual survey of 197 related publications that reveals a disparity in the focus of the AI and clinical communities, caused by data availability and the lack of collaboration, and in modality trends, driven by ubiquity. Last, challenges in AI-acceleration and ways to remediate them are discussed and future research goals are identified.

© 2020 xxxxxxxx. Hosting by Elsevier B.V. All rights reserved.

* *Corresponding author.* Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: author@institute.xxx

Peer review under responsibility of xxxxx.

Hosting by Elsevier

xxxx-xxxx/$ – see front matter © 2020 xxxxxxxx. Hosting by Elsevier B.V. All rights reserved.

ScienceDirect

1. Introduction

The COVID-19 pandemic has created a desperate need of ubiquitous, accurate, low-cost and fast tests, and lung imaging is a key complementary tool in the diagnosis and management of COVID-19 [1], [2]. To automate the processing of the deluge of clinical data, recent advances in artificial intelligence (AI) have been exploited by building deep learning (DL) models that match or exceed performance of healthcare professionals across a diverse set of detection and monitoring tasks [3]. The field of AI in medical imaging (MI) is flourishing in the context of COVID-19 as is evident by our study that revealed a doubling in the number of pre-print and published papers in 2020 compared to 2019 (see Fig. 1). Consequently, the hopes are high that AI can advance rapid patient stratification of suspected COVID-19 patients by automatically classifying lung screenings and thus empowering medical staff, especially in resource-limited areas without testing facilities.

In this paper, we review the current progress in the development of AI technologies for MI to assist in addressing the COVID-19 pandemic, discuss how AI meets the identified gaps and share observations regarding the speed, geographical diversity and relevancy of these developments.

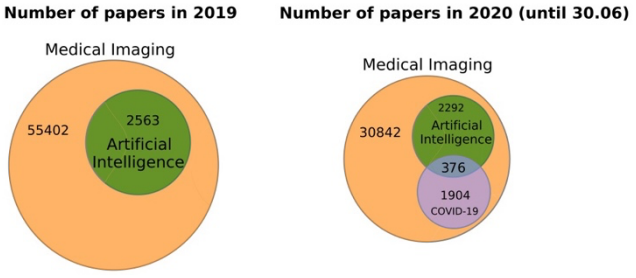

Fig. 1. AI on medical imaging has received extraordinary attention in 2020 and witnessed a multitude of interdisciplinary projects attempting to blend AI technologies with knowledge from MI in order to combat COVID-19. A keyword search on PubMed and preprint servers reveald 2563 papers on AI and MI in 2019, while the first half of 2020 has seen already 2292 such papers.

1.1. State of Artificial Intelligence in Radiology

Radiologists play a crucial role in interpreting medical images for the diagnosis and prognosis of disease. Although AI technologies have recently demonstrated performance that matches radiologists' accuracy in a number of specific tasks, it remains unclear if radiologists who adopt AI-assistance will replace those that do not. Rapid progress of AI in healthcare has been accelerated by the availability of large, diverse and qualified datasets/annotations. For example, breast cancer screening has reported tremendous success in using AI for retrospective detection due to 10 years of effort in large-scale data collection which resulted in OPTIMAM*, a database with a total cohort of >150,000 clients. Using this or other databases, several notable studies have been published comparing performances of DL-based mammography screening solutions to healthcare professionals [4]–[7]. Similarly, significant progress has been made in diagnosing lung conditions using chest X-rays (CXR) and computed tomography (CT), driven by access to open source datasets and labels. For example, DL-based approaches outperform radiologists in detecting several pulmonary conditions from CXR [8] and malignancy of lung nodules in low dose CT [9]. However, several key challenges limit the feasibility of adopting these solutions in practice, namely: (i) poor model generalization due to systemic biases, (ii) lack of model interpretability and (iii) non-scalable image annotation processes.

Related Work - Applications of AI on Imaging to Manage COVID-19
The current pandemic caused a vivid controversy about AI-based diagnosis. While "CT imaging with AI could serve as a surrogate for doctors when fast judgement is needed" [10], by May 2020 there was a lack of supporting evidence to justify such an approach [11]. This sheds a despondent light on the role of novel technologies to combat pandemics where rapid solutions against new diseases are needed. We believe that AI is expected to provide assistance in interpreting lung images in the context of COVID-19. The best systems however will have to incorporate demographics, hospitalization data, and patient history as features along with imaging [12]. Notably, all systems until May 2020 were rated at high risk of bias and overfitting due to non-representative cohort selection along with limited access to imaging data and vague reporting [12]. However, several recent studies evidence the value of AI-assisted disease management for diagnosis [13] and disease severity [14].

Motivation and Contributions

* https://medphys.royalsurrey.nhs.uk/omidb/

* *Corresponding author.* Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: author@institute.xxx
Peer review under responsibility of xxxxx.

Hosting by Elsevier
xxxx-xxxx/$ – see front matter © 2020 xxxxxxxx. Hosting by Elsevier B.V. All rights reserved.

The recent acceleration of publications intersecting AI and imaging for COVID-19, brings a need for rigorous comparative evaluation of papers to summarize and highlight trends to a broad clinical audience. Previous review papers on COVID-19 either focused on a technical assessment of AI in imaging [15], elaborated on the role of imaging [1], or appraised predictive models agnostic of imaging [12]. In contrast, this paper attempts to bridge clinical and technical perspectives by providing a comprehensive overview to guide researchers towards working on the most pressing problems in automating lung image analysis for COVID-19.

Our work uncovers insights by conducting a thorough meta-analysis of 197 COVID-19-related publications. We find that, despite strong evidence for benefits of lung imaging in the containment of COVID-19, there exists a lack of coordination among scientific efforts towards addressing valuable problems. For example, X-Ray data has received disproportionally high attention from AI researchers compared to clinicians, while exploration of AI on ultrasound is neglected. Moreover, about 80% of papers focused on designing AI models for detecting COVID-19 though clinicians have called out its restricted utility for screening. Overall, we observe a two-to-four-fold increase in the number of publications related to AI for lung imaging in Q2 of 2020; causally evoked by COVID-19.

The overall paper is organized as follows: sections 1. and 2. discuss the clinical role of lung imaging for COVID-19, an overview of lung imaging modalities and current state-of-the-art in AI. In section 3, we uncover insights and trends from the meta-analysis of 197 COVID-19-related publications. Finally, sections 4. and 5. describes key challenges for clinical adoption of AI, summarizes our analysis and presents recommendations for the future.

Throughout this paper we focus on the lung as the primary organ of SARS-CoV-2 infection but note the significance of extrapulmonary manifestations that include thrombosis, gastrointestinal symptoms, kidney injury, dermatologic and neurological complications [16].

1.2. Role of Lung Imaging in COVID-19

The two most important use cases of lung imaging for the COVID-19 pandemic are: diagnosis/detection and severity assessment. For a more detailed elaboration on the role of imaging for COVID-19 see [1].

Diagnosis/Detection

COVID-19 may present with diverse symptoms, most commonly flu-like symptoms including dry cough, rhinorrhoea, fever, fatigue and a sore throat [2]. To address the heterogeneity of the clinical presentation, a variety of tests is necessary. The most prominent test relies on the identification of viral RNA in reverse transcriptase polymerase chain reaction (RT-PCR). The viral genetic material is obtained from nasopharyngeal swabs and is considered an accurate though imperfect test [2]. Serological tests identifying anti-SARS-CoV-2 IgM and IgG antibodies are also used [17], but as antibody development can take up to 5 days in humans, they are inappropriate for detection of early infections [18]. Over the disease process, the maximal sensitivity of RT-PCR is obtained 8 days after infection with 0.8 [19]. Compared to diagnostic tests, CT imaging may be more sensitive at a single time point for the diagnosis of COVID-19 [20]. However, while chest CT may have improved sensitivity, the American College of Radiology (ACR) conservatively discourages to use imaging for diagnosis[†]. The two main diagnostic challenges are the differentiation from non-COVID-19 viral pneumonia [21] and the detection of asymptomatic patients with unaffected lungs.

Severity Assessment

Early in the pandemic chest CT was used to identify radiological manifestations [22]. Currently, monitoring disease progression is less important in patient management apart from diagnosing complications such as pulmonary embolisms and superimposed bacterial pneumonias or sequelae of intubation like pneumothoracies [23].

Imaging findings of COVID-19 patients have been shown to correlate with disease severity [24]. Retrospectively analyzed imaging findings on chest CTs that were compared with disease severity, revealed increased occurrence of consolidation, linear opacities, crazy-paving pattern and bronchial wall thickening in severe patients at a higher frequency than in non-severe COVID-19 patients. The CT findings that correlated with worse symptoms included a respiratory rate greater than 30 breaths per minute, oxygen saturation of 93% or less in a resting state and PaO2/FiO2 ratio of 300 mmHg or less, respiratory failure or a requirement for mechanical ventilation, shock or organ requiring intensive care unit monitoring and treatment [22].

2. Overview of Imaging Modalities and its AI Applications

In this section, we provide an overview of CT, X-ray, and ultrasound imaging. Per modality, we highlight recent art in AI solutions and summarize performance and evaluation schemes. Finally, we compare and contrast the role of the modalities in handling COVID-19. Details on MRI and Digital Pathology can be found in the Supplementary Material.

2.1. Computed Tomography (CT)

High resolution CT remains the gold standard imaging technique for thoracic evaluation. The main abnormalities observed in common and severe COVID-19 cases respectively were ground glass opacities (GGOs) and patchy consolidation surrounded by GGOs. CT scanning accurately assessed the severity of COVID-19 and helped monitoring disease transformation among different clinical conditions [25]. COVID-19 pneumonia manifests with chest CT imaging abnormalities, even in asymptomatic patients, with rapid evolution from focal unilateral to diffuse bilateral GGOs that progressed to or co-existed with consolidations within 1–3 weeks [26]. Crucially, patients with a positive RT-PCR test can have a normal CT. Some reports find up to 50% of false negatives when scanned two days after onset of flu-like symptoms [27].

The visual features of GGOs and consolidation lend themselves to visual analysis by DL networks, and several CT analysis systems have emerged from both academia and industry [13], [28], [29]. These networks typically ingest a raw CT scan and output a differential patient diagnosis. Performance-wise, CT networks can be impressive and report a patient level detection with AUC of 0.92 [13] or even 0.96 [29]. To improve robustness, some systems include a lung segmentation stage that focuses attention on the lung for later analysis stages [14], [28]. Some segmentation-based networks output pixelwise-labelled tissue maps of GGO or consolidation regions, which can be informative in grading disease severity or tracking progression over time. This can be achieved by segmenting anatomical landmarks with deep reinforcement learning and

[†] https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection

computing percentage of opacity and lung severity score as complementary severity measures [14].

CT datasets for training diagnosis networks are usually labelled using the RT-PCR test result for COVID-19 and consist of several hundreds to few thousand samples. Training for tissue segmentation is more labour intensive, typically employing manual region labelling by radiologists leading to sample sizes in the hundreds. COVID-19 dataset sizes are expected to grow with new repositories and consortiums such as RSNA[‡], Imaging COVID-19 AI Initiative[§], UK NHS NCCID[**].

Some studies demonstrated that radiologists' performance improves upon consultation of AI: Junior radiologists along with AI can perform as well as mid-senior radiologist [30] and radiologists' sensitivity and specificity can improve by nearly 10% through AI [31]. Similarly, in a blind review with six radiologists' decisions with AI support were superior to radiologists alone (sensitivity 88% vs. 79%, specificity 91% vs. 88%), although the AI model alone performed the best (95% and 96%) [32]. Others used a sequential process composed of an abnormal-slice-selection model and a disease-diagnosis model [13]. They achieved equal sensitivity, but lower specificity compared to a senior radiologist (279 patients). Notably, using a combination of clinicopathological features and CT scans, the model correctly classified 68% of positive patients who presented normal CT scans according to the radiologists. Using a large CT database of 3,777 patients, [30] developed a differentially diagnosing AI system comparable to experienced radiologists. Their system is globally available to combat COVID-19 and highlights the benefits of AI not only in diagnostic assistance but also in prognosing critical illness.

2.2. Chest Radiography (CXR)

Chest X-ray is an integral step in the clinical workflow and often the first-line examination when a patient presents with chest pain, shortness of breath or other cardiopulmonary symptoms. CXR has a low sensitivity and specificity for identifying COVID-19; [33] reported that 89% of 636 COVID-19 patients exhibited normal or only mildly abnormal CXR. Notably, CXR should not be routinely performed in patients presenting with mild to moderate symptoms but may be useful during diagnostic ambiguity[††]. Differing from mild cases, a bedside CXR is still considered the standard of care for both diagnostic applications in the ICU and monitoring changes in clinical status. The temporal course of findings in COVID-19, from development of consolidation on CT and opacities on X-ray is well documented [34].

A multitude of DL network-based solutions were built to distinguish between COVID-19 and other pulmonary conditions in CXR and their performances vary heavily. An early study reports 87% accuracy (sensitivity 0.85, specificity 0.92) for 3-class classification and use, like many others, heatmaps to locate affected regions [35]. Luz et al. (2020) focus on computationally light-weight classification (sensitivity 0.97) that is claimed to scale to high-resolution images and enable mobile analysis [36]. To circumvent the low availability of data, weakly-labelled data augmentation with paediatric non-COVID-19-viral pneumonia CXR samples can improve model performance [37].

Despite the relatively larger size of today's CXR datasets, severe class imbalances, simultaneous prediction of multiple diseases and uncertainty in automatic labelling [8] pose significant challenges to build AI models. Successful solutions adopt model ensemble approaches and classify pulmonary and cardiac biomarkers and conditions (e.g. pleural effusion or cardiomegaly) instead of diseases to obtain high AUC scores of 0.93 [38]. However, simpler models combining self-training, knowledge distillation and semi-supervised learning techniques can achieve comparable performances using only ~10% of labelled data [39].

2.3. Lung ultrasound (LUS)

LUS is widely considered a surrogate lung imaging modality that can reduce usage of ionizing radiation techniques at emergency rooms and resembles a sensitive bedside monitoring tool, especially for the critically-ill [40]. Over the past decades, increasing evidence for the diagnostic value of US in pulmonary conditions such as ARDS or pneumothorax has led to increased recommendation of LUS [41], [42]. Especially for resource-limited settings such as during triage or in low-income countries, LUS is superior to CXR in diagnosing pneumonia [43] or COVID-19 [44] and may resemble a viable first-line examination method [21]. SARS-CoV-2 induces patterns characteristic for viral pneumonia, most prevalently B-lines (vertical, comet-tail artifacts arising from the pleural line [45]) and pleural line abnormalities such as pleural thickening [46]. Findings suggest high diagnostic sensitivity that is comparable to CT scans and can help with diagnosis and subsequent monitoring of the infection [47]. Point-by-point correspondence between LUS and CT patterns can be made and even structured into a disease timeline [24]. This growing body of positive evidence has led many to advocate for an amplified role of LUS during the current pandemic [48], [49] as safe and portable alternative to CT.

However, US has received scant attention from the AI community given the positive clinical evidence. With the rise of DL in computer vision, learning-based approaches to US became more popular [50] and a plethora of advanced processing techniques cope with the wealth of US transducers and probes [51]. More than 50 papers were published about DL on US in 2017 [50], but sparse work has been done on LUS. B-line detection is the most commonly solved task and most works used non-learning-based approaches [52]. But [53] showed that weakly-supervised DL can detect B-lines and used neural attention maps to highlight the most significant regions of the probe that drove the prediction. This paved the road toward more interpretable models that could support clinical decision processes for a variety of pulmonary conditions with B-line involvement. Others confirmed that deep learning can quantify B-lines [54] or subpleural pulmonary lesions [55]. In the context of COVID-19, [56] performed a semantic segmentation of the lung to recognize pathologies like B-lines with high accuracy. They also predicted COVID-19 severity on a scale from 0-3 on 277 videos from 35 patients and achieved 0.7 positive predictive value and 0.6 recall. LUS thus not only has potential to enable severity assessment, but can also facilitate automatic differential diagnosis of COVID-19 from pneumonia [57]. The work continues in [58] and provides an unsupervised mechanism to extract pulmonary biomarkers which are deemed useful by physicians. The reported sensitivity of 0.98 and specificity of 0.91 for COVID-19 provides first evidence that automatic differentiation of COVID-19 from other pulmonary conditions is feasible. In triage situations LUS may thus be used to rapidly assist the patient stratification process and guide downstream molecular or radiologic

[‡] https://www.rsna.org/covid-19

[§] https://imagingcovid19ai.eu

[**] https://www.nhsx.nhs.uk/covid-19-response/data-and-information-governance/national-covid-19-chest-imaging-database-nccid/

[††] For more guidelines see www.acep.org/corona/covid-19-field-guide/

testing. As the absence of skilled personnel is the main preventer of an increased spread of US in the developing world [59], a medical decision support tool building upon existing and available technologies is of utmost importance.

2.4. Modality Comparison for Screening COVID-19

In order for imaging to be effective at scale, it must ideally be ubiquitous, accurate, low-cost and fast, and preferably provide high-quality data, depend on portable devices and be non-invasive.

Table 1 – A lung imaging modality comparison for COVID-19.

	CT	X-RAY	US
COST	1600$ [60]	370$ [61]	140$ [61]
BENEFIT	High sensitivity, high specificity when prevalence high [62]	Fast, broadly available	Mobile, radiation-free, available, cheap
DRAW-BACK	High radiation and cost, availability, patient relocation, infection risk	Low sensitivity, radiation	Operator dependent, longer than CXR, operator infection risk,
CLINICAL ROLE	Diagnosis support	First-line examination for diagnosis	Triage; first line for point-of-care monitoring

The usage of CT is restricted to modern medical facilities (~30k CT scanner are globally available [63]). The practical advantages of LUS are numerous and it is recommended during triage [49]. Its acceptance as the preferred modality in Italy [64], may promote a more prominent role in the future. For further details see Table 1.

Studies Evaluating CT, X-ray and LUS

CXR is notoriously less sensitive than CT [65]. Weinstock et al. [33] found that 90% of COVID-19 patients exhibit normal or mildly abnormal CXR. A multimodal imaging study of CT, CXR and LUS reported B-lines in LUS as most consistent pathological pattern [66]. A direct comparison of LUS and CXR revealed that LUS has a significantly higher sensitivity for COVID-19 diagnosis than CXR whereas no significant differences were found in specificity [44]. There is conflicting evidence regarding a comparison of LUS and CT and a dissent about the implications of LUS as a first-line examination method. Indeed, Yang et al. [67] found a higher sensitivity than CT in detecting all examined patterns (regional AIP, AIS, pleural effusion and consolidations). But others reported relatively low sensitivity (0.69 and 0.78) for mild and moderate cases of COVID-19 [68]. Similar accuracy of LUS and CT was reported in [47], with LUS identifying signs of COVID-19 more likely than CT (78% vs. 72%). This has caused a vivid debate on the role of LUS for the COVID-19 pandemic [48], [49], [69].

Major AI-studies that benchmark modalities by their usability for automatic segmentation, diagnosis or mere pattern detection are still underway. Several works have compared the predictive power of CT and CXR data from heterogenous data sources, but a clinical study with multi-imaging data from the same patients is needed to confirm the radiologic findings. In one work that combined COVID-19 CT and CXR data, no clear benefit was found from the multimodal integration [70].

3. Meta-Analysis of COVID-19 Literature

To discover trends from the overwhelming research activities interfacing areas of COVID-19, AI and MI, we conducted a detailed meta-analysis of publications selected through a keyword search on PubMed and preprint servers such as *arXiv*, *bioRxiv* and *medRxiv*. For an in-depth description of the publication selection procedure, see Appendix A. We quantify the publication efforts, summarize trends and highlight key findings from a thorough manual analysis of 197 papers.

3.1. The Coronavirus-induced Rise of AI in Lung Imaging

There has been a stable growth of papers on AI on both lung and breast imaging over the years 2017-2019 (see Fig. 2A). In 2020, the rise of lung-related papers has been accelerated by COVID-19 causing an additional ~200 papers (111% growth rate).

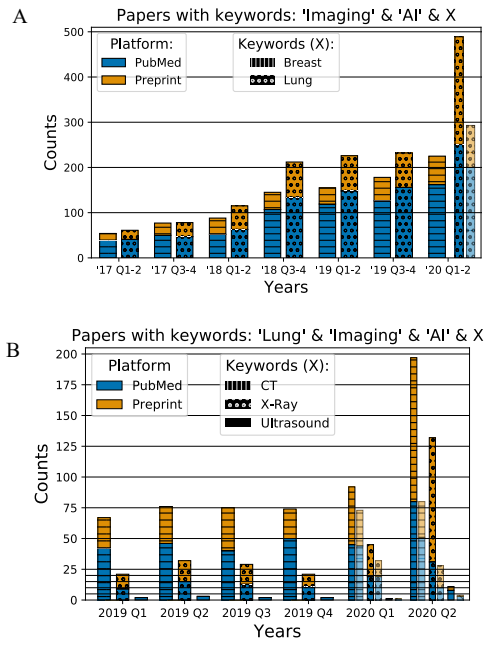

Fig. 2: Number of papers per keyword and platform. *Top:* Paper counts using AI on breast or lung imaging. *Bottom:* Paper counts comparing the usage of AI on lung imaging modalities. Shaded bars exclude COVID-19 related papers. For each keyword, multiple synonyms were used (for details see appendix).

To compare the impact of individual modalities, Figure 2B shows that 2019 witnessed a stable trend of ~100 publications per quarter on AI on lung imaging with shares of 72% for CT, 25% for CXR and 2% for LUS. Post the COVID-19 outbreak, numbers soared to 138 (340) papers in Q1 (Q2) 2020. This rise was causally evoked by COVID-19 as excluding papers mentioning COVID-19 would have resulted in a continuation of the stable trend (see shaded bars) of a hypothetical 106 (Q1) and 112 (Q2) publications. Indicating a shift in the focus of research, the first half of 2020 produced more papers for each modality than all of 2019. The simpler and more accessible modalities X-Ray and US gained shares and make up 37% and 3% of all papers in 2020 compared to 60% CT papers. For Q2, X-Ray had the highest fraction of COVID-19 papers (79%), and the growth can be primarily ascribed to preprints (370% increase) rather than peer-reviewed publications (86%).

3.2. Key Insights from the Manual Review of 197 Papers

To complement the quantitative analysis, we reviewed 197 publications for qualitative analysis (see Fig. A1 for the inclusion workflow). In sum, the modality-specific COVID-19 papers of Fig. 2B were combined with papers containing the keywords 'AI, medical imaging and COVID-19'.

Disparity between Clinical and AI Communities: Of the 1904 papers about MI and COVID-19 (see Fig 1B), 1072 are specific to modalities as shown in Fig. 3A, indicating a dominance of CT in clinical papers and a roughly even relevance of X-Ray and US imaging. By using publication counts as an indirect indicator on scientific practice, we observe a mismatch in the focus of the AI community in comparison to the clinical community as illustrated by the distribution of papers per modality in Fig. 3B.

Focus of AI for COVID-19 Misaligned with the ACR: In addition, the vast majority (77%) of papers focused on detection of COVID-19 over tasks like severity and prognosis (Fig. 4A). Note that a key concern was the unanimous use of only imaging data for training the algorithms (52% CXR, see Fig 3B). This trend is in contrast to the ACR recommendations that appraise imaging as an inconclusive test for COVID-19 detection. However, only a mere ~6% of papers exploited multimodal data towards building their AI models.
While data availability is vital to AI and explains the community's focus on X-rays, success in creating impactful and practical healthcare AI solution hinges upon close collaboration among radiologists and AI experts. Pivoting AI focus to utilize multimodal data of patient cohorts would be increasingly effective and invaluable in the containment of COVID-19.

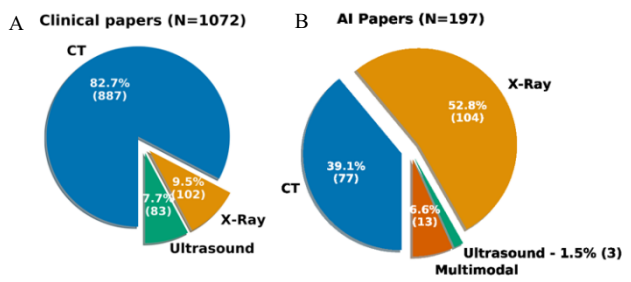

Fig. 3: Imaging modality comparison during the COVID-19 pandemic. The AI community (B) publishes disproportionally more papers on X-Ray compared to clinicians (A), whereas CT and Ultrasound are vastly underrepresented. Multimodal papers used more than one imaging modality.

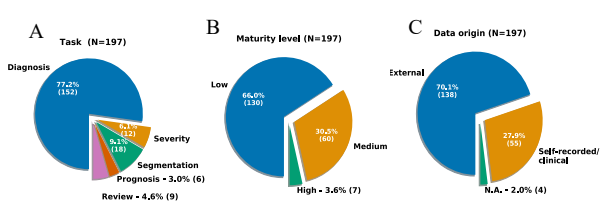

Fig. 4: Distribution of papers on AI and medical imaging during the COVID-19 pandemic. Relative proportions for task (A), quality (B) and data origin (C) are given. Papers were processed manually.

Most AI Solutions for COVID-19 have Low Maturity: To qualify the maturity of publications, we chose three parameters: (i) size of patient data used for training and evaluation, (ii) deployment and adoption of solutions in hospitals and (iii) complexity of the AI framework. We find only hily mature 7 studies, of which none utilized X-Rays. Further, papers that used CXR data were reported as lowest in average quality and highest in preprint ratio. For details on the task and maturity as distributed by modality see appendix (Fig. A2).

Overuse of Small Incomprehensive Public Datasets: We observe that only ~27% of papers used proprietary or clinical data (Fig 4C), while over 70% analyzed publicly available databases, usually comprised of no more than a few hundred images from heterogenous sources/devices without detailed patient information.

USA and China Lead in AI; China and Italy Lead in Data: The leading countries for AI expertise were China and USA with India and Canada as runners-ups. Instead, data was shared by countries hit early from the pandemic, i.e. China and, to a lesser extent, Italy, USA and Spain (see Fig 5). However, the geographical diversity of author contributions was remarkable. First-authors from 34 countries and 6 continents contributed to the research, evidencing a global collaborative spirit (Fig 5B).

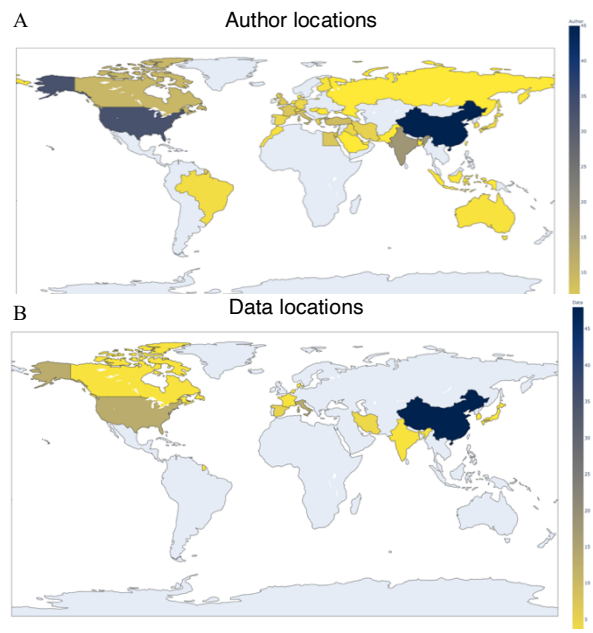

Fig. 5. Geographical depictions on the locations of main authors of papers on the AI meta-analysis (A) and the locations of the COVID-19 imaging data they used (B).

4. Outlook – Challenges in AI Acceleration

When applying AI in clinical settings, practical and socio-environmental factors like interrupted staff workflow, insufficient hardware infrastructure and trust are the main inhibitors of AI integration [71]. Here, we highlight some of the most delicate technical challenges connecting MI research and clinical deployment.

AI Robustness and Generalization
Clinical AI models are plagued by generalization issues and often fail to sustain performance in new environments, owing to inherent biases in model or data, lack of real-world representation and small size of training

datasets. Generally, model robustness can be improved by leveraging continual learning and domain adaptation techniques which attempt to learn domain invariant representations by aligning the data distribution between source and target environments [72]. A domain shift is usually caused by changes in the data distribution arising from discrepancies in population demographics, missing measurements, or label imbalance. However, studies show that for clinical data, domain shifts are often associated with task shifts due to complex interactions between predictor variables and comorbidities [73] thereby rendering sophisticated domain adaption techniques insufficient.

AI Interpretability

The combined lack of robustness and interpretability poses steep challenges for the adoption of AI models in clinical practice. The susceptibility of AI to adversarial attacks gave rise AI safety and explainability [74]. A humanly-interpretable access to the model's decision process is crucial to hone trust in AI, especially in medical applications where reasoning is inductive, sensitive decisions are made and patients expect plausible hypotheses from physicians. Most current interpretability tools focus on generating explanations which highlight patterns learned from the data but do not translate model decisions in human-understandable forms. Recently, counterfactual reasoning combined with uncertainty quantification principles have opened doors to building models integrating reliability into the optimization process [75]. This enables model introspection and facilitates human-in-the-loop analysis.

Economic factors

COVID-19 has created a sharp rise in demand of respiratory care devices with a compound annual growth rate (CAGR) of 261% between 2019 and 2020[‡‡]. The prevalence of the different imaging modalities varies across the globe for operational, financial and care-practice reasons (see Fig. A3). Key challenges are to meet the increased demands and deliver the right AI technologies into the right regions. The demand of US systems appears to increase and a switch in market lead from CT to US is predicted in North America. This development is driven by the adoption of handheld devices and home healthcare services that reduce hospitalization cost[§§]; a trend further enhanced by COVID-19.

Portable X-Ray or Ultrasound equipment is predestined for rapid on-device analysis and already used to triage at patient homes [49]. Future research will need to explore the feasibility of AI for real-time analysis. The ubiquitous presence of mobile, Bluetooth-equipped devices paves a road to absorb the absence of experienced doctors during triage. Under appropriate guidance, a wireless US coupled to a smartphone app can offer an alternative to CXR and CTs for defined tasks particularly in remote areas. But user interpretability and expertise will remain a challenge toward analytic automatization.

Data

Standardization of data across devices and streamlining of medical examination are challenging. Whenever AI integration is anticipated, execution protocols such as the BLUE protocol for LUS [45] should be followed with stringent care. While the use of US has increased exponentially in developing countries, lack of training is the highest barrier [59]. Software infrastructure to instruct physicians such as massive-open-online-courses are needed to accelerate AI-assisted diagnosis but also to collect high-quality, annotated data on large scale. Further, building AI systems that can integrate data collected across the clinical disease journey, especially multimodal imaging modalities, will be required to leverage the full power of AI and address the ACR recommendations.

5. Conclusion

To summarize, we have provided an extensive comparison of lung imaging modalities for COVID-19 with an emphasis on AI. We highlight that imaging is not recommended for COVID-19 diagnosis by the ACR due to overlapping symptoms with viral infections. CT should be reserved for cases with inconclusive findings in initial assessments via CXR and LUS and LUS can be performed as first-line examination during triage [21]. Notably, clinical diagnosis should by no means solely base on algorithmic predictions. DL models combining CT and clinical features start to match [13] or surpass [31] performance of senior radiologists in detecting COVID-19, but the road toward clinical application is long and difficult.

Our meta-analysis revealed that AI efforts predominantly focus on CXR and to a lesser extent on CT imaging. LUS instead is clinically not yet as relevant, but the economic and diagnostic impact is expected to grow. US has received little attention in AI despite its wide, global availability which is a key factor for dealing pandemics, especially in remote and undeveloped areas where supply chains are long, and repurposing of existing technologies is the sharpest weapon for prompt pandemic alleviation. Given that LUS is also more sensitive than CXR [44], comparable to CT [47], and has evident practical advantages, it is presumably the modality with the highest improvement potential for medical image analysis in the near future.

Author contributions

MRZ, DB, JLR, EK and MG conceived the presented work. MRZ conceived the meta-analysis and supervised this project, and MRZ and MG set the high-level objectives of this work. MM and JB developed the software to perform paper keyword searches. JB, DM, DR, AC, VM, EK and MG manually reviewed papers. JB analyzed the results and JB and DR created the figures. All authors contributed toward the interpretation and improvement of the analysis. DB and JB led and distributed the manuscript writing efforts. JB, DB, AC, DR, MG, EK and MRZ wrote full, individual sections of the manuscript. All authors reviewed initial versions and contributed significantly to the different sections of the manuscript and approved the submitted version.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

JB acknowledges funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 826121.

[‡‡] Analyzed with data from: https://www.marketsandmarkets.com/, *Respiratory Care Devices Market - Forecast To 2025*

[§§] https://www.grandviewresearch.com/industry-analysis/ultrasound-device-market

References

[1] D. Dong *et al.*, "The role of imaging in the detection and management of COVID-19: a review," *IEEE Rev. Biomed. Eng.*, 2020.

[2] G. Pascarella *et al.*, "COVID-19 diagnosis and management: a comprehensive review," *J. Intern. Med.*, vol. 288, no. 2, pp. 192–206, Aug. 2020.

[3] X. Liu *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *Lancet Digit. Heal.*, vol. 1, no. 6, pp. e271–e297, Oct. 2019.

[4] H. E. Kim *et al.*, "Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study," *Lancet Digit. Heal.*, vol. 2, no. 3, pp. e138–e148, Mar. 2020.

[5] A. Akselrod-Ballin *et al.*, "Predicting breast cancer by applying deep learning to linked health records and mammograms," *Radiology*, vol. 292, no. 2, pp. 331–342, Jun. 2019.

[6] S. M. McKinney *et al.*, "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, no. 7788, pp. 89–94, Jan. 2020.

[7] E. P. V. Le, Y. Wang, Y. Huang, S. Hickman, and F. J. Gilbert, "Artificial intelligence in breast imaging," *Clinical Radiology*, vol. 74, no. 5. W.B. Saunders Ltd, pp. 357–366, 01-May-2019.

[8] J. Irvin *et al.*, "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, pp. 590–597, Jul. 2019.

[9] D. Ardila *et al.*, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nat. Med.*, vol. 25, no. 6, pp. 954–961, Jun. 2019.

[10] B. McCall, "COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread," *Lancet Digit. Heal.*, vol. 2, no. 4, pp. e166--e167, 2020.

[11] A. Laghi, "Cautions about radiologic diagnosis of COVID-19 infection driven by artificial intelligence," *Lancet Digit. Heal.*, vol. 2, no. 5, p. e225, 2020.

[12] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, p. 18, Apr. 2020.

[13] X. Mei *et al.*, "Artificial intelligence--enabled rapid diagnosis of patients with COVID-19," *Nat. Med.*, pp. 1–5, 2020.

[14] S. Chaganti *et al.*, "Automated Quantification of CT Patterns Associated with COVID-19 from Chest CT," *Radiol. Artif. Intell.*, vol. 2, no. 4, p. e200048, Jul. 2020.

[15] D. Shi, Feng and Wang, Jun and Shi, Jun and Wu, Ziyan and Wang, Qian and Tang, Zhenyu and He, Kelei and Shi, Yinghuan and Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," *IEEE Rev. Biomed. Eng.*, 2020.

[16] A. Gupta *et al.*, "Extrapulmonary manifestations of COVID-19," *Nat. Med.*, vol. 26, no. 7, pp. 1017–1032, Jul. 2020.

[17] M. Lisboa Bastos *et al.*, "Diagnostic accuracy of serological tests for covid-19: Systematic review and meta-analysis," *BMJ*, vol. 370, p. 2516, Jul. 2020.

[18] Y. W. Tang, J. E. Schmitz, D. H. Persing, and C. W. Stratton, "Laboratory diagnosis of COVID-19: Current issues and challenges," *Journal of Clinical Microbiology*, vol. 58, no. 6. American Society for Microbiology, 01-Jun-2020.

[19] L. M. Kucirka, S. A. Lauer, O. Laeyendecker, D. Boon, and J. Lessler, "Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction--Based SARS-CoV-2 Tests by Time Since Exposure," *Ann. Intern. Med.*, 2020.

[20] T. Ai *et al.*, "Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," *Radiology*, p. 200642, 2020.

[21] J. B. Mendel, J. T. Lee, and D. Rosman, "Current Concepts Imaging in COVID-19 and the Challenges for Low and Middle Income Countries The Journal of Global Radiology Imaging findings," *J. Glob. Radiol.*, 2020.

[22] M. Li *et al.*, "Coronavirus Disease (COVID-19): Spectrum of CT Findings and Temporal Progression of the Disease," *Academic Radiology*, vol. 27, no. 5. Elsevier USA, pp. 603–608, 01-May-2020.

[23] World Health Organization, "Use of chest imaging in COVID-19: a rapid advice guide, 11 June 2020," 2020.

[24] M. J. Fiala, "Ultrasound in COVID-19: a timeline of ultrasound findings in relation to CT," *Clin. Radiol.*, 2020.

[25] S. M. H. Tabatabaei, H. Talari, F. Moghaddas, and H. Rajebi, "Computed Tomographic Features and Short-term Prognosis of Coronavirus Disease 2019 (COVID-19) Pneumonia: A Single-Center Study from Kashan, Iran," *Radiol. Cardiothorac. Imaging*, vol. 2, no. 2, p. e200130, Apr. 2020.

[26] H. Shi *et al.*, "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study," *Lancet Infect. Dis.*, vol. 20, no. 4, pp. 425–434, Apr. 2020.

[27] W. Yang and F. Yan, "Patients with RT-PCR-confirmed COVID-19 and normal chest CT," *Radiology*, vol. 295, no. 2, pp. E3--E3, 2020.

[28] O. Gozes *et al.*, "Rapid AI Development Cycle for the Coronavirus (covid-19) Pandemic: Initial Results for Automated Detection & Patient Monitoring Using Deep Learning CT Image A nalysis," *arXiv Prepr. arXiv2003.05037*, 2020.

[29] L. Li *et al.*, "Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct," *Radiology*, p. 200905, 2020.

[30] K. Zhang *et al.*, "Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography," *Cell*, vol. 181, no. 6, pp. 1423-1433.e11, Jun. 2020.

[31] J. Song *et al.*, "End-to-end automatic differentiation of the coronavirus disease 2019 (COVID-19) from viral pneumonia based on chest CT," *Eur. J. Nucl. Med. Mol. Imaging*, 2020.

[32] H. X. Bai *et al.*, "AI Augmentation of Radiologist Performance in Distinguishing COVID-19 from Pneumonia of Other Etiology on Chest CT," *Radiology*, p. 201491, Apr. 2020.

[33] M. B. Weinstock, A. Echenique, J. W. Russell, and others, "Chest x-ray findings in 636 ambulatory patients with COVID-19 presenting to an urgent care center: a normal chest x-ray is no guarantee.," *J Urgent Care Med*, no. 14 (7), pp. 13–18, 2020.

[34] M. Carotti *et al.*, "Chest CT features of coronavirus disease

2019 (COVID-19) pneumonia: key points for radiologists," vol. 125, pp. 636–646, 2020.

[35] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, p. 103792, Jun. 2020.

[36] E. Luz, P. L. Silva, R. Silva, and G. Moreira, "Towards an Efficient Deep Learning Model for COVID-19 Patterns Detection in X-ray Images," *arXiv Prepr. arXiv2004.05717*, 2020.

[37] S. Rajaraman and S. Antani, "Weakly Labeled Data Augmentation for Deep Learning: A Study on COVID-19 Detection in Chest X-Rays," *Diagnostics*, vol. 10, no. 6, p. 358, May 2020.

[38] H. H. Pham, T. T. Le, D. Q. Tran, D. T. Ngo, and H. Q. Nguyen, "Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels," 2020.

[39] D. Rajan, J. J. Thiagarajan, A. Karargyris, and S. Kashyap, "Self-Training with Improved Regularization for Few-Shot Chest X-Ray Classification," 2020.

[40] F. Mojoli, B. Bouhemad, S. Mongodi, and D. Lichtenstein, "Lung ultrasound for critically ill patients," *American Journal of Respiratory and Critical Care Medicine*, vol. 199, no. 6. American Thoracic Society, pp. 701–714, 15-Mar-2019.

[41] M. A. Chavez *et al.*, "Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis," *Respir. Res.*, vol. 15, no. 1, p. 50, 2014.

[42] J.-E. Bourcier, S. Braga, and D. Garnier, "Lung ultrasound will soon replace chest radiography in the diagnosis of acute community-acquired pneumonia," *Curr. Infect. Dis. Rep.*, vol. 18, no. 12, p. 43, 2016.

[43] L. E. Ellington *et al.*, "Lung ultrasound as a diagnostic tool for radiographically-confirmed pneumonia in low resource settings," *Respir. Med.*, vol. 128, pp. 57–64, 2017.

[44] E. M. S.-P. J.R. Pare, I. Camelo, K.C. Mayo, M.M. Leo, J.N. Dugas, K.P. Nelson, W.E. Baker, F. Shareef, R.N. Mitchell, M. Patrica, "Point-of-care Lung Ultrasound Is More Sensitive than Chest Radiograph for Evaluation of COVID-19," *West. J. Emerg. Med. Integr. Emerg. Care with Popul. Heal.*, 2020.

[45] D. A. Lichtenstein, "Lung ultrasound in the critically ill: the BLUE protocol," *Ann. Intensive Care*, 2015.

[46] C. Xing, Q. Li, H. Du, W. Kang, J. Lian, and L. Yuan, "Lung ultrasound findings in patients with COVID-19 pneumonia," *Crit. Care*, vol. 24, no. 1, p. 174, Apr. 2020.

[47] Y. Tung-Chen *et al.*, "Correlation between Chest Computed Tomography and Lung Ultrasonography in Patients with Coronavirus Disease 2019 (COVID-19)," *Ultrasound Med. Biol.*, vol. 0, no. 0, 2020.

[48] D. Buonsenso, D. Pata, and A. Chiaretti, "COVID-19 outbreak: less stethoscope, more ultrasound," *Lancet Respir. Med.*, 2020.

[49] M. J. Smith, S. A. Hayward, S. M. Innes, and A. Miller, "Point-of-care lung ultrasound in patients with COVID-19--a narrative review," *Anaesthesia*, 2020.

[50] S. Liu *et al.*, "Deep Learning in Medical Ultrasound Analysis: A Review," *Engineering*, vol. 5, no. 2. Elsevier Ltd, pp. 261–275, 01-Apr-2019.

[51] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proc. IEEE*, vol. 108, no. 1, pp. 11–29, 2019.

[52] R. Moshavegh, K. L. Hansen, H. Moller-Sorensen, M. B. Nielsen, and J. A. Jensen, "Automatic Detection of B-Lines in In Vivo Lung Ultrasound," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 66, no. 2, pp. 309–317, Feb. 2019.

[53] R. J. G. van Sloun and L. Demi, "Localizing B-lines in lung ultrasonography by weakly-supervised deep learning, in-vivo results," *IEEE J. Biomed. Heal. Informatics*, 2019.

[54] X. Wang, J. S. Burzynski, J. Hamilton, P. S. Rao, W. F. Weitzel, and J. L. Bull, "Quantifying lung ultrasound comets with a convolutional neural network: Initial clinical results," *Comput. Biol. Med.*, vol. 107, pp. 39–46, Apr. 2019.

[55] Y. Xu *et al.*, "Boundary Restored Network for Subpleural Pulmonary Lesion Segmentation on Ultrasound Images at Local and Global Scales," *J. Digit. Imaging*, 2020.

[56] S. Roy *et al.*, "Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound," *IEEE Trans. Med. Imaging*, 2020.

[57] J. Born *et al.*, "POCOVID-Net: Automatic Detection of COVID-19 From a New Lung Ultrasound Imaging Dataset (POCUS)," *arXiv Prepr. arXiv2004.12084*, 2020.

[58] J. Born, N. Wiedemann, G. Brändle, C. Buhre, B. Rieck, and K. Borgwardt, "Accelerating COVID-19 Detection with Explainable Ultrasound Image Analysis," in *Under review at NeurIPS 2020*, 2020.

[59] S. Shah, B. A. Bellows, A. A. Adedipe, J. E. Totten, B. H. Backlund, and D. Sajed, "Perceived barriers in the use of ultrasound in developing countries," *Crit. Ultrasound J.*, vol. 7, no. 1, pp. 1–5, 2015.

[60] W. C. Black *et al.*, "Cost-effectiveness of CT screening in the national lung screening trial," *N. Engl. J. Med.*, vol. 371, no. 19, pp. 1793–1802, Nov. 2014.

[61] B. P. Jones *et al.*, "Feasibility and safety of substituting lung ultrasonography for chest radiography when diagnosing pneumonia in children: a randomized controlled trial," *Chest*, vol. 150, no. 1, pp. 131–138, 2016.

[62] B. Xu *et al.*, "Chest CT for detecting COVID-19: a systematic review and meta-analysis of diagnostic accuracy," *Eur. Radiol.*, 2020.

[63] M. Castillo, "The industry of CT scanning." Am Soc Neuroradiology, 2012.

[64] L. Vetrugno *et al.*, "Our Italian experience using lung ultrasound for identification, grading and serial follow-up of severity of lung involvement for management of patients with COVID-19," *Echocardiography*, vol. 37, no. 4. Blackwell Publishing Inc., 2020.

[65] H. Y. F. Wong *et al.*, "Frequency and Distribution of Chest Radiographic Findings in COVID-19 Positive Patients," *Radiology*, p. 201160, Mar. 2019.

[66] P. Lomoro *et al.*, "COVID-19 pneumonia manifestations at the admission on chest ultrasound, radiographs, and CT: single-center study and comprehensive radiologic literature review,"

Eur. J. Radiol. open, p. 100231, 2020.

[67] Y. Yang, Y. Huang, F. Gao, L. Yuan, and Z. Wang, "Lung ultrasonography versus chest CT in COVID-19 pneumonia: a two-centered retrospective comparison study from China," 2020.

[68] W. Lu *et al.*, "A Clinical Study of Noninvasive Assessment of Lung Lesions in Patients with Coronavirus Disease-19 (COVID-19) by Bedside Ultrasound," *Ultraschall der Medizin*, vol. 41, no. 3, pp. 300–307, Jun. 2020.

[69] J. C. H. Cheung and K. N. Lam, "POCUS in COVID-19: pearls and pitfalls," *The Lancet Respiratory Medicine*, vol. 8, no. 5. Lancet Publishing Group, p. e34, 01-May-2020.

[70] S. El-Bana, A. Al-Kabbany, and M. Sharkas, "A Multi-Task Pipeline with Specialized Streams for Classification and Segmentation of Infection Manifestations in COVID-19 Scans," *medRxiv*, 2020.

[71] E. Beede *et al.*, "A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.

[72] R. Venkataramani, H. Ravishankar, and S. Anamandra, "Towards continuous domain adaptation for medical imaging," in *Proceedings - International Symposium on Biomedical Imaging*, 2019, vol. 2019-April, pp. 443–446.

[73] J. J. Thiagarajan, D. Rajan, and P. Sattigeri, "Understanding Behavior of Clinical Models under Domain Shifts," *arXiv Prepr. arXiv1809.07806 (2018).*, Sep. 2018.

[74] M. Reyes *et al.*, "On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities," *Radiol. Artif. Intell.*, vol. 2, no. 3, p. e190043, May 2020.

[75] J. J. Thiagarajan, B. Venkatesh, D. Rajan, and P. Sattigeri, "Improving Reliability of Clinical Models using Prediction Calibration," *MICCAI UNSURE Work.*, 2020.