



Artificial Intelligence–assisted chest X-ray assessment scheme for COVID-19

Krithika Rangarajan^{1,2} · Sumanyu Muku² · Amit Kumar Garg¹ · Pavan Gabra¹ · Sujay Halkur Shankar¹ · Neeraj Nischal¹ · Kapil Dev Soni¹ · Ashu Seith Bhalla¹ · Anant Mohan¹ · Pawan Tiwari¹ · Sushma Bhatnagar¹ · Raghav Bansal¹ · Atin Kumar¹ · Shivanand Gamanagati¹ · Richa Aggarwal¹ · Upendra Baitha¹ · Ashutosh Biswas¹ · Arvind Kumar¹ · Pankaj Jorwal¹ · Shalimar¹ · A. Shariff¹ · Naveet Wig¹ · Rajeshwari Subramaniam¹ · Anjan Trikha¹ · Rajesh Malhotra¹ · Randeep Guleria¹ · Vinay Namboodiri³ · Subhashis Banerjee² · Chetan Arora²

Received: 19 September 2020 / Revised: 27 November 2020 / Accepted: 11 December 2020

© European Society of Radiology 2021

Abstract

Objectives To study whether a trained convolutional neural network (CNN) can be of assistance to radiologists in differentiating Coronavirus disease (COVID)–positive from COVID-negative patients using chest X-ray (CXR) through an ambispective clinical study. To identify subgroups of patients where artificial intelligence (AI) can be of particular value and analyse what imaging features may have contributed to the performance of AI by means of visualisation techniques.

Methods CXR of 487 patients were classified into [4] categories—normal, classical COVID, indeterminate, and non-COVID by consensus opinion of 2 radiologists. CXR which were classified as “normal” and “indeterminate” were then subjected to analysis by AI, and final categorisation provided as guided by prediction of the network. Precision and recall of the radiologist alone and radiologist assisted by AI were calculated in comparison to reverse transcriptase-polymerase chain reaction (RT-PCR) as the gold standard. Attention maps of the CNN were analysed to understand regions in the CXR important to the AI algorithm in making a prediction.

Results The precision of radiologists improved from 65.9 to 81.9% and recall improved from 17.5 to 71.75 when assistance with AI was provided. AI showed 92% accuracy in classifying “normal” CXR into COVID or non-COVID. Analysis of attention maps revealed attention on the cardiac shadow in these “normal” radiographs.

Conclusion This study shows how deployment of an AI algorithm can complement a human expert in the determination of COVID status. Analysis of the detected features suggests possible subtle cardiac changes, laying ground for further investigative studies into possible cardiac changes.

Key Points

- Through an ambispective clinical study, we show how assistance with an AI algorithm can improve recall (sensitivity) and precision (positive predictive value) of radiologists in assessing CXR for possible COVID in comparison to RT-PCR.
- We show that AI achieves the best results in images classified as “normal” by radiologists. We conjecture that possible subtle cardiac in the CXR, imperceptible to the human eye, may have contributed to this prediction.
- The reported results may pave the way for a human computer collaboration whereby the expert with some help from the AI algorithm achieves higher accuracy in predicting COVID status on CXR than previously thought possible when considering either alone.

Keywords Artificial intelligence · Radiograph · COVID

Krithika Rangarajan and Sumanyu Muku have an equal contribution.

✉ Krithika Rangarajan
krithikarangarajan86@gmail.com;
drkrithikarangarajan.aiims@gmail.com

¹ All India Institute of Medical Sciences, New Delhi, India

² Indian Institute of Technology, Delhi, New Delhi, India

³ Indian Institute of Technology, Kanpur, Kanpur, India

Abbreviations

AI	Artificial intelligence
BSTI	British Thoracic Imaging Society
COVID	Coronavirus disease 2019 (COVID-19)
CXR	Chest X-ray
PACS	Picture Archiving and Communication System
RT-PCR	Reverse transcriptase-polymerase chain reaction

Introduction

Over 25 million people globally have tested positive for Coronavirus disease 2019 (COVID-19) as of 31 August 2020 [1]. The pandemic is still at its peak in many countries around the world, with a large number of patients infected in resource-constrained countries. The current gold standard test for severe acute respiratory syndrome coronavirus (SARS-CoV-2) is RT-PCR which has limited sensitivity and takes time to process. Rapid antigen tests suffer from low sensitivities ranging between 11.1 and 81.8% [2, 3]. With the peak of cases shifting to the developing world, availability of test-kits, protective equipment for the healthcare workers, etc. is becoming even more scarce [4].

Chest radiography (CXR) is among the most common investigations performed world over and accounts for 25% of total diagnostic imaging procedures [5]. It is a portable, inexpensive, and a safe modality, which is widely used to assess the extent of lung involvement in a wide variety of thoracic pathologies. It is widely available in hospital set-ups even in small peripheral centres and involves minimal contact with the patient. However, CXR has not been seen to be sensitive or specific for changes related to COVID-19. While many patients do not show any changes on chest radiography (to an expert eye) [6], those who do show changes are difficult to differentiate from other forms of pneumonia.

Deep learning has shown supra-human performance for image classification tasks, not only for natural images [7] but also in some medical imaging scenarios as well [8]. Typically, the strengths and weaknesses of human readers and machines have been complementary [9, 10] and collaboration between the expert and AI system tends to yield the best results [11]. We therefore explored the potential of an AI-based system as an aid to a radiologist in classifying chest radiographs, as this could provide the much-needed bedside, widely available inexpensive diagnostic tool.

With the release of several datasets of CXR from COVID-positive patients across the world, several researchers have attempted to build such a system for classification. However, only two studies had tested the algorithms on the real world, clinical data in a hospital setting [12, 13]. All others had tested on a held-out subset of curated publicly available data. There is therefore a need to validate the algorithm in a prospective hospital-based setting. In addition, none of the studies has identified subsets of patients where the use of AI can help experts or analysed the pattern of abnormalities picked up in making the prediction.

We developed and validated an algorithm that could differentiate the CXR of COVID-19-positive patients from COVID-19-negative patients. We showed how the deployment of such a tool can improve the recall (sensitivity) and precision (positive predictive value (PPV)) of radiologists in identifying COVID-positive

patients in routine clinical practice. In addition, we found that AI performed particularly well where the CXR was completely normal. Analysis of the attention maps revealed possible changes in cardiac configuration, laying ground for further studies.

Materials and methods

Data used in the study

Institutional Review Board clearance was obtained prior to the start of the study (reference number IEC-242-17.04.2020). This was an ambispective study. Recruitment of patients for training, testing, and in the ambispective clinical study is summarised in Fig. 1.

In the retrospective phase (18th April to 27th May 2020), patients who tested positive for COVID-19 by RT-PCR and underwent chest radiographs within 24 h of the test were included. Only the chest radiograph at presentation was taken into consideration. All patients underwent anteroposterior (AP) view radiographs since only portable radiography within a COVID isolation ward is performed in our institution. The COVID-negative radiographs in this phase came from consecutive (AP view) chest radiographs performed in our institution between the 1st to the 30th of January 2018. In this subset, only AP views were chosen in order to exclude bias, since all COVID-positive radiographs were AP views. Radiographs from January 2018 were chosen to negate the possibility of false-negative RT-PCR studies.

In the prospective phase, all patients presenting to the outpatient or emergency department of the All India Institute of Medical Sciences (AIIMS), New Delhi, from 1 July 2020 to 30 July 2020 with the following inclusion criteria were considered for the study: (1) Patients who have undergone throat/nasal swab for COVID-19 RT-PCR within the institution, (2) have undergone AP view radiograph within 24 h of the RT-PCR, and (3) gave informed consent for the use of the imaging data for this study.

COVID test reports (only RT-PCR) were obtained from the Electronic Medical Record (EMR) of the hospital and X-ray images were accessed through the Picture Archiving and Communication System (PACS) of the institute. As our institution also serves as a quarantine centre, some patients who were asymptomatic or mildly symptomatic also underwent CXR in order to triage patients who may need admission, as judged appropriate in specific clinical situations. Since patients presented in different phases of the illness, the timing of radiograph with respect to the onset of symptoms or date of exposure could not be standardised. A total of 177 CXR included in this study were COVID-positive on RT-PCR and 310 were COVID-negative on RT-PCR.

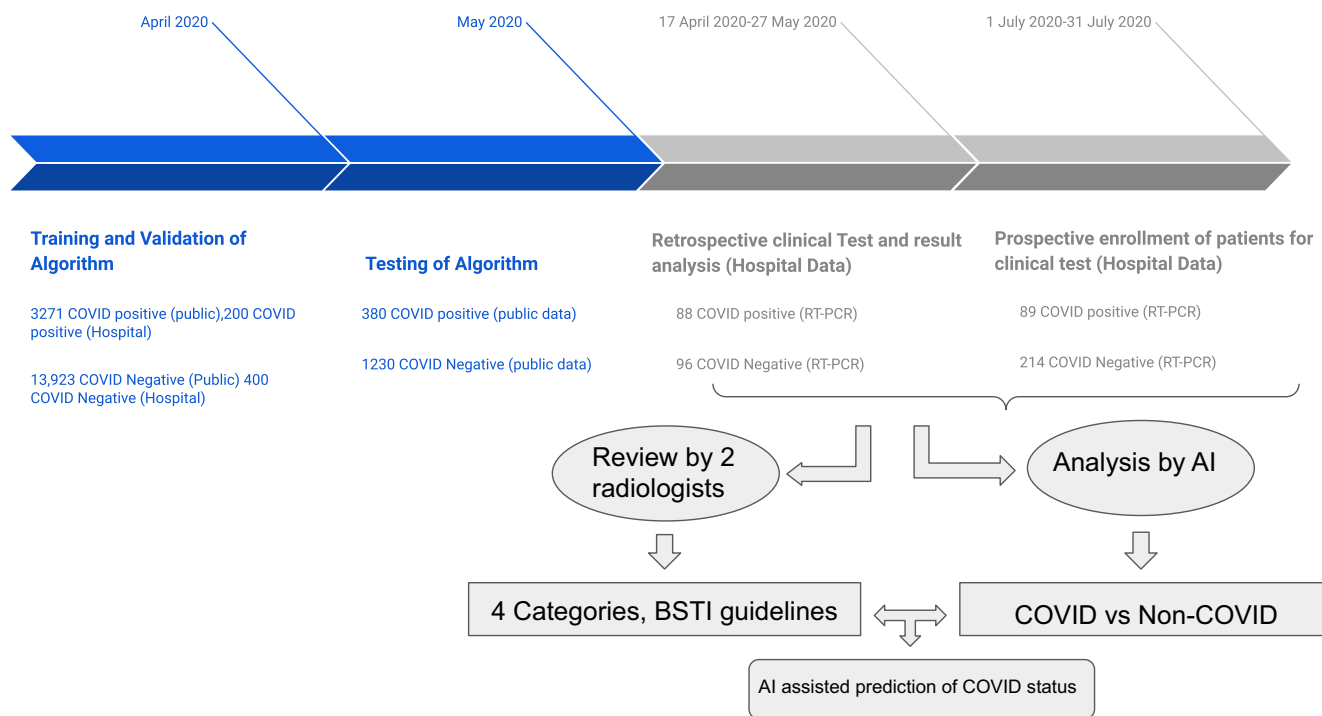


Fig. 1 Phases of the selection of patients for training and testing the CNN

AI analysis of radiographs

We designed and trained our own deep neural network, referred to as COVID-AID network for the study. The network was trained to predict the COVID status of the patient based on CXR. Details of the data used, network architecture, training strategies, and validation of network on public data are given in the supplementary appendix. The algorithm is open-sourced, and a link to the network is provided at the end of this manuscript. The performance of the model was assessed by means of an AuROC curve and precision-recall statistics. None of the images used in the clinical study was used either for training or for validation of the AI algorithm.

Radiologist analysis of radiographs

All radiographs were classified by consensus opinion of 2 radiologists (with 9 and 3 years of experience in chest imaging) as per the British Society for Thoracic Imaging (BSTI) guidelines into one of [4] categories—A: normal, B: indeterminate for COVID-19, C: classical for COVID-19, or D: non-COVID-19 [14].

AI-assisted radiologist analysis of radiograph

We assumed that when the radiologist could call the CXR as being “C: classical COVID” or “D: non-COVID” as in the BSTI guidelines, he/she was fairly certain of his/her diagnosis; therefore, we took his/her opinion as the final opinion. In cases where the radiologist classified the CXR as “A: normal”, he/

she had absolutely no guiding criteria to determine whether this was COVID/non-COVID; therefore, the opinion of AI was taken as final in this case. The strategy is also justified, because AI performed particularly well in this subset as seen in Table 3. For the subset where the human expert classified a radiograph as B: indeterminate, we used the confidence score of AI to guide whether a final categorisation may be given.

In this phase, therefore, CXR classified as C: “classical for COVID-19” and D: “non-COVID” remained in those categories, but the A: “normal” and B: “indeterminate” radiographs were reclassified after considering the predictions made by the AI algorithm and the corresponding confidence scores. In A: “normal” category, the prediction made by the AI algorithm was accepted; that is, those classified as COVID-positive by the algorithm were reclassified into the “C: classical COVID” category and those classified as COVID-negative by the AI technique were reclassified into the “D: non-COVID” category. For the cases originally in the B: indeterminate category, the prediction made by the AI algorithm was accepted wherever the confidence score of the algorithm was above 70%. We arrived at a figure of 70 % confidence based on the observation that in the publicly available data used for testing the algorithm, predictions had high accuracy where the confidence was higher than 70 percent.

Evaluation metrics

For evaluating the performance of the radiologist alone, as well as radiologist + AI, RT-PCR of the patient within 24 h

of the CXR was considered the gold standard. Only definite predictions of “C: classical COVID” or “D: non-COVID” were considered positive and negative, respectively, for the calculation of precision and recall.

Precision/positive predictive value = total number of patients predicted as C: classical COVID who were also COVID-positive as per RT-PCR/total number of patients predicted as C: classical COVID.

Recall/sensitivity = total number of patients predicted as C: classical COVID which were also COVID-positive as per RT-PCR/total number of COVID-positive patients in the study as per RT-PCR.

Specificity = total number of patients predicted as D: non-COVID which were also COVID-negative as per RT-PCR/total number of COVID-negative patients in the study as per RT-PCR.

Results

The AuROC curve of the algorithm (AI alone) on the public data and complete data collected from our hospital is given in Fig. 2 a and b respectively. The algorithm attained an AuROC of 89.3% for the COVID class in the dataset from our hospital. The overall accuracy of the algorithm (for both COVID-positive and COVID-negative) was 78% on this dataset.

The distribution of images into each of the [4] categories is given in Table 1. As seen in Table 1, prior to the use of the AI algorithm, a large number of images fell into the A: “normal” or B: “indeterminate” category, where the radiologist could not determine with any confidence whether the changes could be attributed to COVID or not. After running the proposed AI algorithm, 100% of images classified as normal by the

radiologist and 23% of images (where the AI confidence score was over 70%, as described in the section above) classified as indeterminate by the radiologist could be reclassified into a definitive category (C: COVID, or D: non-COVID). Precision and recall of the consensus opinion of the radiologist without and with the assistance of the AI tool are given in Table 2.

As seen in Table 2, the radiologists performed well in this classification if we were to include only images assigned a definitive category. However, since a large number of images were assigned either “normal” or “indeterminate category”, when the entire dataset was considered, the precision of the radiologist alone was 65.9% and recall was only 17.5%. In addition to AI, the precision and recall improved to 81.9 and 71.75% respectively.

It may be noted that the network performed particularly well on images classified as being completely normal by the radiologist. Figure 2c and Table 3 shows the ROC curve for images classified as A: normal by the radiologist.

The number of COVID-positive and COVID-negative (as per RT-PCR) patients classified in the [4] categories before and after assistance by AI is given in Table 4 and Table 5 respectively. The performance of the AI algorithm on CXR classified as A: normal is given as a confusion matrix in Table 3.

As seen in Tables 4 and 5, the specificity of the radiologist increased from merely 20 (62 correct COVID-negative predictions/ 310 COVID-negative patients in the dataset) to 68.7% (213/310).

Analysis of patterns detected by the AI algorithm: could cardiac changes hold the key?

As seen in Fig. 2c and Table 3, the algorithm attains very high accuracy in differentiating between COVID-positive and

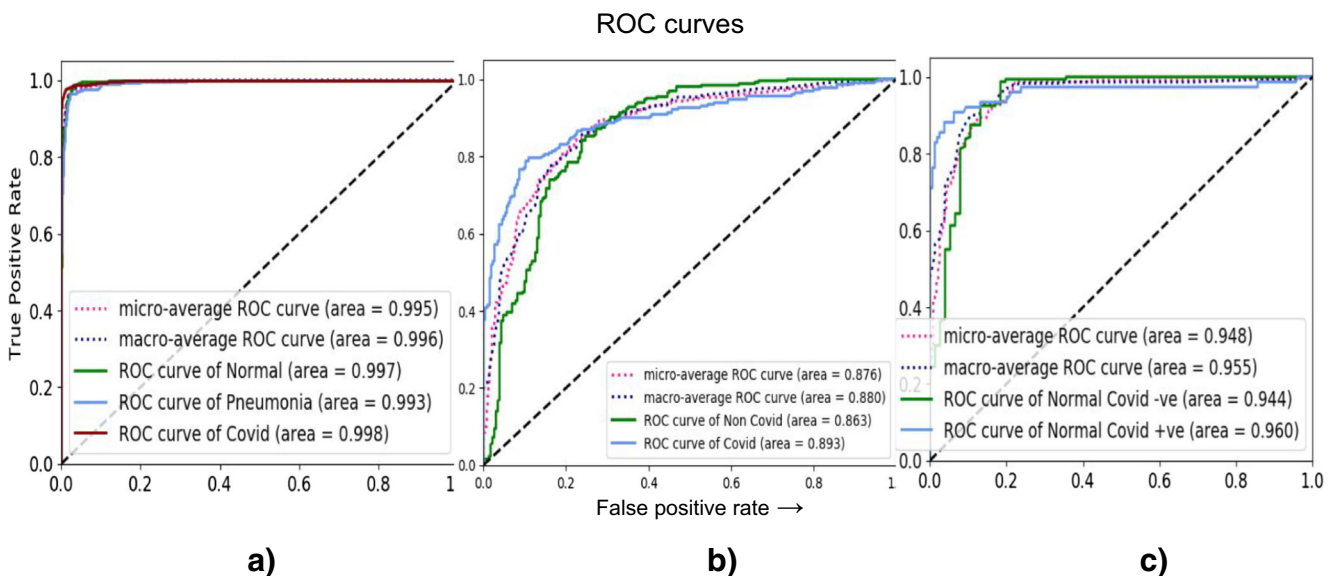


Fig. 2 Receiver operating curve plotting the performance of our model on publicly available data (a), ambispective hospital data (b), and on the subset of images (c) considered normal by the radiologist

Table 1 Distribution of CXR into 4 categories by radiologists, before and after assistance by the AI algorithm. The number of images classified in each category has been mentioned in the table

	Normal	Non-COVID	Indeterminate for COVID	Classical COVID
Radiologist alone	236	66	138	47
Radiologist+AI	0	66+160 (classified normal by radiologist)	106	47+76 (classified as normal by radiologist) 32+ (classified indeterminate by radiologist)

COVID-negative patients in the subset classified as being A: normal by the human expert. In order to understand the reason for such high accuracy, we analysed the proposed deep neural network using various visualisation techniques. The analysis highlights what regions of the radiograph the network focuses on while making a particular prediction. We used RISE visualisations and analysed local crops used for predictions. The details of the methods of visualisations used are described in detail in the supplementary [appendix](#).

These visualisations were analysed by a radiologist, along with the RT-PCR report and the confidence scores of the network. For patients who had lung changes of pneumonia, we found the network to focus on the correct regions of the radiograph, as judged by the radiologist. An example of RISE visualisations and corresponding local crop (see details of our network architecture given in supplementary [appendix](#) for description of “local crop”) used by our deep neural network is given in Fig. 3.

We were particularly interested in the visualisations for radiographs adjudged as being normal by the radiologist. We found that the AI model consistently focussed on the cardiac region in most patients and it was correctly classified as being COVID positive. The analysis of the confidence scores revealed that the network was 98–100% confident in making the prediction when it focussed on this particular region in patients with their radiographs classified as A: normal by the radiologist in our study. Figure 4 shows the visualisations described, along with the confidence scores for prediction in one such radiograph. All images were then seen again by a radiologist to look for any changes in the carinal angle, cardiothoracic ratio, and subtle signs of cardiac involvement, but no definite changes could be discerned conclusively.

In order to confirm our hypothesis of the cardiac shadow being the determining factor, we performed a few experiments. First, in the subset of images adjudged

as A: normal, we forced the network to focus on the cardiac region by taking predictions only from a central crop into consideration for the local model (see the details of the proposed network architecture in supplementary [appendix](#)). We found that 75% of the wrong predictions made by the AI model were corrected by this intervention alone. Second, we trained a second deep neural network to segment out the lungs out in a CXR in order to focus exclusively on the cardiac region (supplementary material, Fig. 5). We found that there was no drop in accuracy, though no improvement was seen in this case.

Both the above experiments supported our hypothesis that subtle changes in the cardiac configuration likely aided in a specific diagnosis of COVID positivity even in patients who did not show any lung changes on chest radiographs.

Experiments to analyse cardiac changes on other modalities

ECG-gated CT scans, echocardiography, and cardiac MRI could all have potentially yielded some clues to throw light into the possible cardiac changes. However, we do not routinely perform CT scans for all COVID-positive patients in our institution, unless their clinical condition warrants one. These CTs are not ECG gated. In our prospective dataset, only 11 COVID-positive patients underwent non-cardiac gated chest CT scans during this time. An analysis of these CT scans revealed a dilated main pulmonary artery in 8 out of 11 patients, left atrial anteroposterior diameters of over [4] cm in 3 patients, and minimal pericardial effusion in 3 patients. However, all these patients had significant pulmonary abnormalities; therefore, these CTs cannot be considered representative of the population. Due to logistic reasons, we could only perform echocardiography in

Table 2 Precision and recall of radiologist alone and radiologist plus AI

	Precision	Recall
Radiologist (here, true-positive prediction is COVID-positive patient classified as classical COVID or non-COVID are considered for metric calculator)	65.9	88.5
Radiologist (true-positive prediction is COVID-positive patient classified as classical COVID; all scans in the study categorised in any of the 4 categories were considered for metric calculation)	65.9	17.5
Radiologist +AI (criteria for calculation same as the row)	81.9	71.75

Table 3 Confusion matrix of the AI algorithm in images labelled as “normal” by the radiologist. Precision and recall of the AI algorithm are tabulated in percentage

Ground truth	Predicted	
	Normal Covid -ve	Normal Covid +ve
Normal Covid -ve	151 (0.94)	9 (0.06)
Normal Covid +ve	9 (0.12)	67 (0.88)

	Precision	Recall	F1-score	Total number
Normal CXR, COVID Negative	0.94	0.84	0.94	160
Normal CXR, COVID positive	0.88	0.88	0.88	76
Accuracy			0.92	236

a small subset of patients. Only 8 patients who had normal chest radiographs underwent echocardiography. We found

minimal pericardial effusion in 2 patients and mild left atrial enlargement in 2 patients.

Table 4 Performance of radiologist alone. Rows represent classification by the radiologist into 4 categories. 2nd and third columns show the RT-PCR results (ground truth) against which the predictions of the radiologist have been judged

Radiologist classification	RT-PCR	
	COVID-positive (RT-PCR)	COVID-negative (RT-PCR)
Classical COVID	31	16
Normal	76	160
Indeterminate	66	72
Non-COVID	4	62

Table 5 For Radiologist + AI. Rows represent classification by the radiologist + AI into 4 categories. 2nd and third columns show the RT-PCR results (ground truth) against which the predictions of the radiologist + AI has been judged

Radiologist classification	RT-PCR	
	COVID-positive	COVID-negative
Classical COVID	31+67*+29^=127	16+9*+3^=28
Normal	0	0
Indeterminate	37	69
Non-COVID	4+9*=13	62+151*=213

*Represents X-rays reclassified from normal to definitive category by addition of AI and ^ represent X-rays reclassified from indeterminate to definitive category

Discussion

In this work, we explored the role of an artificial intelligence algorithm in differentiating COVID-positive from COVID-negative patients on chest radiographs. We found that AI was particularly effective where the trained eye could not identify any abnormality with certainty. To the best of our knowledge, such efficacy in differentiating COVID-positive from COVID-negative patients in radiographs that look normal to the expert eye has not been reported previously. This algorithm has been tested in 3 phases, 2 of these being in real-world hospital settings. We also attempted to understand the “super-human” prediction and put forward a hypothesis on possible subtle cardiac involvement even in patients with mild disease (adjudged by normal radiographs) which has not been suggested previously.

In order to ensure that our AI algorithm is not “overfitting”, which is a common problem with deep neural networks, we conducted a variety of experiments. First, we ensured that data used for training was obtained from multiple sources, with radiographs of patients from all over the world. In addition, we performed extensive clinical tests, first on a held-out public dataset and second on a dataset selected retrospectively and finally implemented it prospectively. This is significant as it proves the efficacy of our model on unseen data in a real-world clinical setting. We also extensively checked for bias due to confounding factors known to influence the appearance of chest X-rays. We used only anteroposterior view

radiographs in the second and third phase of the test and checked the age and gender distributions for differences between the COVID-positive and COVID-negative subgroups.

With the release of numerous publicly available datasets, research in AI-based solutions on COVID-19 images has bloomed. However, testing these algorithms in real-world clinical scenarios is absolutely essential before any possible deployment [15]. In the real world, poor-quality radiographs cannot be excluded. Also, in the real world, other pathologies, such as cancer, effusions due to other comorbidities, and chest wall pathologies, cannot be excluded, unlike in curated datasets. In our study, we used all radiographs, without excluding any radiographs based on quality or pathology. This is very significant, as all previous studies specifically excluded radiographs believed to be of inferior quality. In real-world clinical practice, particularly when portable radiography is used, as in COVID, this may result in the exclusion of a large number of radiographs. This also explains our lower accuracy in our hospital dataset in comparison to the publicly available data, which is usually curated to include only good-quality radiographs.

AI algorithms also suffer from the problem of being “black box” algorithms; i.e. it is difficult to decipher why the system makes a particular decision. It was crucial to overcome this in our case, as the algorithm reliably picked up a finding that was not easily visible to the human eye. We have performed several experiments towards the “explainability” of these algorithms. All our experiments with RISE visualisations, attention models, feeding the proposed model only crops of the original images,



Fig. 3 Example of RISE visualisations and corresponding local crop used by our deep neural network in a patient where lung changes are seen. As seen in this image, the network correctly focuses on appropriate changes in the lung

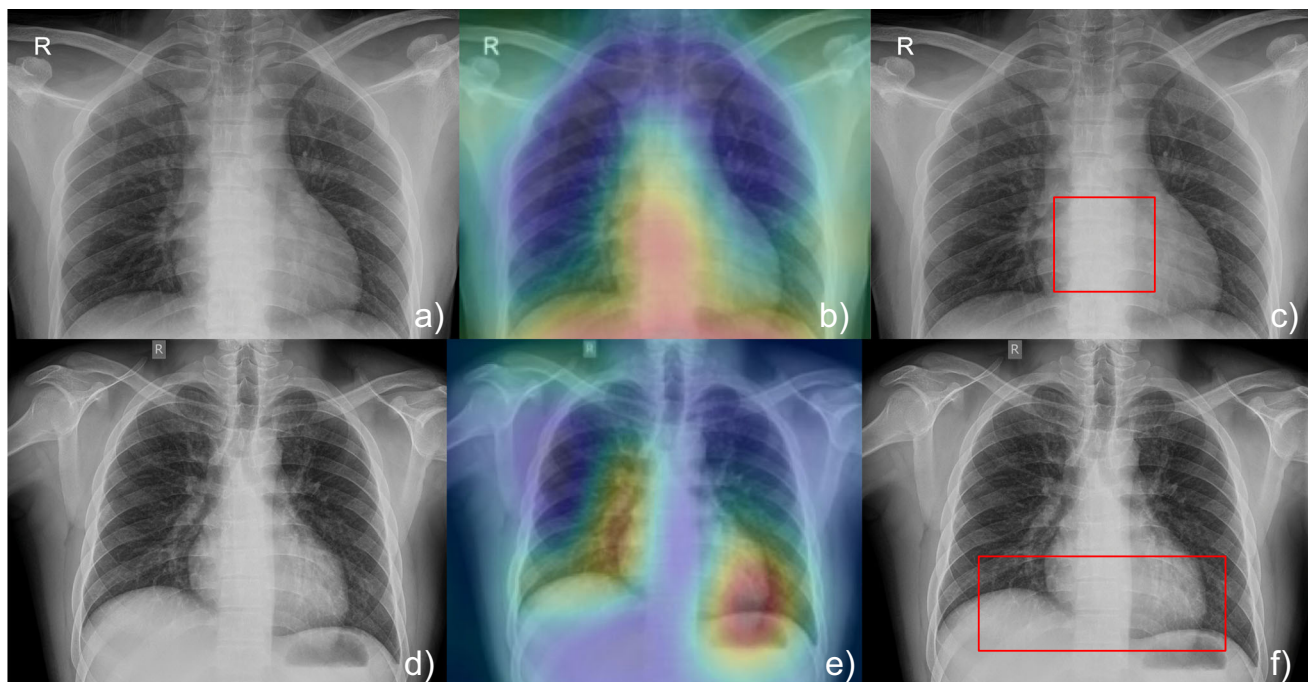


Fig. 4 Top row shows a chest X-ray (**a**) from a COVID-positive patient adjudged as being normal by a radiologist. The CNN classified this as being COVID-positive. The RISE visualisation (**b**) shows network attention in the cardiac region; the local crop (**c**) is also focussed on the region of the heart just below the carina. The bottom row shows a radiograph of a

COVID-negative patient adjudged as normal (**d**) by the radiologist and correctly classified as COVID-negative by the network. In this case, the visualisation shows attention at patchy distributed locations. Both predictions were made with 100% confidence. This pattern of visualisation was consistent in most radiographs in the test set

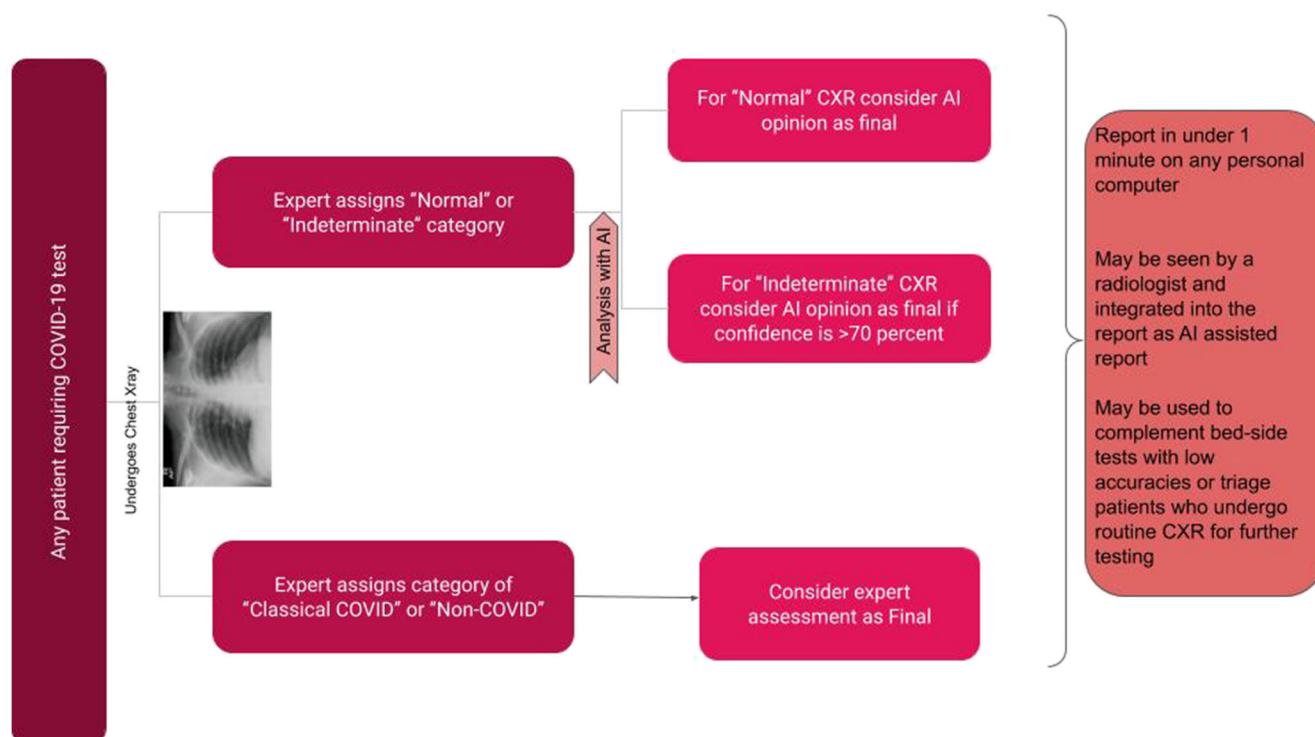


Fig. 5 Possible workflow in patients for determination of COVID-19 status on the basis of the radiograph

and segmenting out the lung tissue indicated that subtle changes in cardiac configuration may hold the key. Our findings have been supported by the work of Puntmann et al [16, 17] where the investigators found signal changes within the myocardium of over 70% of patients after recovery, even in patients who did not suffer from severe forms of the disease.

Our study has several limitations—we were unable to perform a thorough cardiac evaluation due to logistic reasons. The conclusions on cardiac involvement, therefore, only lay ground for further investigation of the cardiac manifestations, rather than being conclusions in themselves. In addition, the present study had a small sample size. Larger clinical studies would therefore be needed for the confirmation of our findings. The timing of the radiograph in relation to the onset of illness also needs to be elucidated in more detail in further clinical studies.

This paper has 2 major contributions. First, we provide a mechanism for man-machine interaction where we show that the combination can significantly improve our diagnostic ability. The ability to detect COVID in apparently normal chest radiographs can potentially transform clinical care. This tool has been tested retrospectively as well as prospectively in separate clinical cohorts of patients to exclude possible biases. The use of this tool could dramatically rationalise the number of tests that is needed to be performed per day, particularly in resource-constrained settings. We give a suggested workflow for the use of the algorithm in a clinical scenario (Fig. 5). Second, the hypothesis of possible cardiac involvement needs further investigation and potentially adds a new dimension to our understanding of the pathophysiology of the disease.

We hereby open-source our tool, hoping researchers around the world may test it in their clinical settings. The tool also introduces a possibility of human—AI collaboration in such a way that AI builds on the limitations of the human expert to provide significant benefit to the patient.

Source code and trained model are available on the link below:

https://github.com/sumanyumuku98/CovidAid_V2

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00330-020-07628-5>.

Acknowledgments We sincerely thank the research section of the All India Institute of Medical Sciences for facilitating the research process, data collection, and helping investigators from multiple disciplines interact smoothly for completion of the project. We would like to thank the research section of the Indian Institute of Technology Delhi for allocating us high priority in the high-performance computing cluster, without which the project would not have been possible. We acknowledge the contribution of our data entry operator, Hema Malhotra, who meticulously compiled the data for the project. We would like to thank all the frontline workers in the institute—the faculty, residents, staff, radiography technicians for working tirelessly through these tough times, and contributing to this project in the process.

Funding The project was funded by an intramural grant within the All India Institute of Medical Sciences, New Delhi.

Compliance with ethical standards

Guarantor Krithika Rangarajan, Assistant Professor, Radiology, AIIMS.

Conflict of interest The authors have no conflicts of interest to declare. The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Either written informed consent was obtained or verbal recorded consent was obtained from patients due to risk of infection (this was stated and approved by the ethics committee).

Ethical approval Institutional Review Board approval was obtained.

Study subjects or cohorts overlap Some parameters of the same study subjects or cohorts are currently being studied in many other projects which have been approved. However, the CXR (used in this study) of this cohort is not currently being studied in any other AI-based study.

Methodology

- ambispective study
- study of diagnostic accuracy
- performed at one institution

References

1. Coronavirus Update (Live): 23,818,500 Cases and 817,090 Deaths from COVID-19 Virus Pandemic - Worldometer. [cited 2020 Aug 25]. Available from: <https://www.worldometers.info/coronavirus/>. Accessed 25 Aug 2020
2. Scohy A, Anantharajah A, Bodéus M, Kabamba-Mukadi B, Verroken A, Rodriguez-Villalobos H (2020) Low performance of rapid antigen detection test as frontline testing for COVID-19 diagnosis. *J Clin Virol* 129:104455. <https://doi.org/10.1016/j.jcv.2020.104455>
3. Mak GC, Cheng PK, Lau SS et al (2020) Evaluation of rapid antigen test for detection of SARS-CoV-2 virus. *J Clin Virol* 104500: 129. <https://doi.org/10.1016/j.jcv.2020.104500>
4. Weinstein MC, Freedberg KA, Hyle EP, Paltiel AD (2020) Waiting for certainty on Covid-19 antibody tests — at what cost? *N Engl J Med* 383(6):e37. <https://doi.org/10.1056/NEJMp2017739>
5. United Nations (ed) (2000) Sources and effects of ionizing radiation: United Nations Scientific Committee on the effects of atomic radiation: UNSCEAR 2000 report to the general assembly, with scientific annexes. United Nations, New York 2 p
6. Wong HYF, Lam HYS, Fong AH-T et al (2020) Frequency and distribution of chest radiographic findings in patients positive for COVID-19. *Radiology*. 296(2):E72–E78. <https://doi.org/10.1148/radiol.2020201160>
7. Russakovsky O, Deng J, Su H et al (2015) ImageNet large scale visual recognition challenge. *Int J Comput Vis* 115(3):211–252

8. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 542(7639):115–118. <https://doi.org/10.1038/nature21056>
9. Ker J, Wang L, Rao J, Lim T (2018) Deep learning applications in medical image analysis. *IEEE Access* 6:9375–9389
10. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL (2018) Artificial intelligence in radiology. *Nat Rev Cancer* 18(8): 500–510. <https://doi.org/10.1038/s41568-018-0016-5>
11. Patel BN, Rosenberg L, Willcox G et al (2019) Human-machine partnership with artificial intelligence for chest radiograph diagnosis. *NPJ Digit Med* 2(1):1–10
12. Hwang EJ, Kim H, Yoon SH, Goo JM, Park CM (2020) Implementation of a deep learning-based computer-aided detection system for the interpretation of chest radiographs in patients suspected for COVID-19. *Korean J Radiol*. <https://doi.org/10.3348/kjr.2020.0536> [cited 2020 Aug 25];21
13. Murphy K, Smits H, Knoops AJG et al (2020) COVID-19 on chest radiographs: a multireader evaluation of an artificial intelligence system. *Radiology*. 296(3):E166–E172. <https://doi.org/10.1148/radiol.2020201874>
14. Johnstone A. Thoracic imaging in COVID-19 infection. :28. available from https://www.bsti.org.uk/media/resources/files/BSTI_COVID-19_Radiology_Guidance_version_2_16.03.20.pdf
15. Bachtiger P, Peters NS (2020) Walsh S machine learning for COVID-19—asking the right questions. *Lancet Digit Health* 2(8): e391–e392. [https://doi.org/10.1016/S2589-7500\(20\)30162-X](https://doi.org/10.1016/S2589-7500(20)30162-X)
16. Yancy CW, Fonarow GC (2020) Coronavirus disease 2019 (COVID-19) and the heart—is heart failure the next chapter? *JAMA Cardiol*. Available from: <https://doi.org/10.1001/jamacardio.2020.3575>
17. Puntmann VO, Carerj ML, Wieters I, et al (2020) Outcomes of cardiovascular magnetic resonance imaging in patients recently recovered from coronavirus disease 2019 (COVID-19). *JAMA Cardiol*. Available from: <https://doi.org/10.1001/jamacardio.2020.3557>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.