ORIGINAL ARTICLE

Radiographic findings in COVID-19: Comparison between AI and radiologist

Arsh Sukhija, Mangal Mahajan, Priscilla C Joshi, John Dsouza, Nagesh D N Seth, Karamchand H Patil[1]

Departments of Radiodiagnosis and Imaging and [1]Community Medicine, Bharati Vidyapeeth Deemed to be University Medical College and Hospital, Pune, Maharashtra, India

Correspondence: Dr. Mangal Mahajan, Department of Radiodiagnosis, Bharati Vidyapeeth Deemed to be University Medical College and Hospital, Pune - 411 043, Maharashtra, India. E-mail: drmangalmahajan@gmail.com

Abstract

Context: As the burden of COVID-19 enhances, the need of a fast and reliable screening method is imperative. Chest radiographs plays a pivotal role in rapidly triaging the patients. Unfortunately, in low-resource settings, there is a scarcity of trained radiologists. **Aim:** This study evaluates and compares the performance of an artificial intelligence (AI) system with a radiologist in detecting chest radiograph findings due to COVID-19. **Subjects and Methods:** The test set consisted of 457 CXR images of patients with suspected COVID-19 pneumonia over a period of three months. The radiographs were evaluated by a radiologist with experience of more than 13 years and by the AI system (NeuraCovid, a web application that pairs with the AI model COVID-NET). Performance of AI system and the radiologist were compared by calculating the sensitivity, specificity and generating a receiver operating characteristic curve. RT-PCR test results were used as the gold standard. **Results:** The radiologist obtained a sensitivity and specificity of 44.1% and 92.5%, respectively, whereas the AI had a sensitivity and specificity of 41.6% and 60%, respectively. The area under curve for correctly classifying CXR images as COVID-19 pneumonia was 0.48 for the AI system and 0.68 for the radiologist. The radiologist's prediction was found to be superior to that of the AI with a *P* VALUE of *0.005*. **Conclusion:** The specificity and sensitivity of detecting lung involvement in COVID-19, by the radiologist, was found to be superior to that by the AI system.

Key words: Artificial intelligence; chest radiographs; COVID pneumonia; rapid triaging

Introduction

The first case of the novel coronavirus 2019 (COVID-2019) was reported in Wuhan city of Hubei province of China in December 2019. Infections caused by this betacoronavirus continue to increase worldwide. The outbreak was officially recognized as a pandemic on 11th March 2020.[1-3] As of date, there are more than 30 million reported cases of COVID-19 worldwide with more than 5.2 million cases only in India.[4] As these trends continue to increase and considering that a large number of patients are asymptomatic (or have very mild symptoms), rapid triaging of the suspected patients has become imperative.[5,6]

Access this article online	
Quick Response Code:	Website: www.ijri.org
	DOI: 10.4103/ijri.IJRI_777_20

The diagnostic method used for detecting COVID-19 cases is reverse transcriptase-polymerase chain reaction (RT-PCR) testing, which detects SARS-CoV-2 RNA from specimens obtained through nasopharyngeal or oropharyngeal swabs.[7] Although RT-PCR testing is highly specific, it has its own limitations. It has a prolonged turnaround time, is expensive, laborious and complicated manual process that is in short supply. Furthermore, the sensitivity of RT-PCR testing is highly variable and depends on how and when the specimen was collected.[8,9]

Cite this article as: Sukhija A, Mahajan M, Joshi PC, Dsouza J, Seth ND, Patil KH. Radiographic findings in COVID-19: Comparison between AI and radiologist. Indian J Radiol Imaging 2021;31:S87-93.

Received: 18-Sep-2020 **Revised:** 20-Sep-2020
Accepted: 15-Oct-2020 **Published:** 23-Jan-2021

Chest radiographs (CXR) and computed tomography (CT) analyzed by radiologists are alternative screening methods that have been used for COVID-19 screening to look for findings associated with COVID viral infection.[10,11] While CT has a very high sensitivity in detecting COVID infection[12,13] there is always a risk of virus transmission during the transport of patient and also while performing the scan. It has a low specificity as compared to the RT PCR. Furthermore, lack of availability in parts of the world makes CT scan a less commonly used investigation.[14]

CXR is likely the most commonly utilized modality for diagnosing and assessing the prognosis of COVID-19 patients.[11,12] CXR enables rapid triaging of the suspected COVID-19 patients as it is a fast and relatively inexpensive imaging modality. CXR systems are readily available and accessible in almost all clinical settings as they are considered as standard equipment in most of the healthcare systems. Lastly, the existence of portable CXR systems allows imaging to be performed within an isolation ward therefore significantly reducing the risk of transmission during transport as opposed to imaging with fixed systems such as CT.[11]

On CXR, COVID-19 manifests with a wide spectrum of patterns.[15,16] These can be identified as patchy or diffuse haziness, ground glass opacities, reticulo–nodular opacities and consolidation. The disease may show basal or peripheral predominance, is usually bilateral or can be widespread. Grading the severity based on percentage of the lung involvement is also important for management of the patients.[17,18] The findings on CXR often overlap with the non-COVID pneumonias (e.g., bacterial, fungal or other viral).[15,16] Therefore, there is a need to correctly identify and distinguish the two on CXR by an experienced radiologist since the findings can be sometimes subtle.

Unfortunately, there is a severe shortage of experienced radiologists in both developed and underdeveloped regions to allow for precise interpretation of such images. Therefore, computer-aided diagnostic systems that can aid radiologists to rapidly and accurately interpret radiography images and help detect lung involvement in COVID-19 play an important role.

An AI tool may aid the radiologist or may help in screening the radiographs in the absence of a radiological expertise. Recent studies have shown that few AI systems are capable of detecting COVID-19 changes on chest radiographs with performance comparable to that of the radiologist.[19-21] In this study we evaluated the performance of an artificial intelligence (AI) system for the detection of COVID-19 lung involvement based on the CXR and compared the same with the findings by an experienced radiologist.

NeuraCovid is an artificial intelligence software developed by Neura Health Inc. (founded in June 2019). This AI model was built using 16,756 chest radiograph images from over 13645 patients. This software uses an authentication using Google account login and by using AP or PA Chest radiography input, it predicts and classifies the outcomes based as normal, pneumonia or COVID-19. It also provides HeatMap of the given radiograph to give insight into how AI made its decision. It runs in secure HIPAA compliant Google Cloud and also saves the history of all uploads, inferences and payments.

Subjects and Methods

A total of 457 CXR images were used in the test set. Both anteroposterior and posteroanterior projections of the chest were obtained with mobile units (Shimadzu make MUX-10 unit and Philips Basic HF 4003 model). Informed written consent was taken. The data set included the patients referred to the Department of Radiodiagnosis and Imaging, B.V.D.U.M.C and Hospital, Pune, Maharashtra, India for suspected COVID pneumonia. The CXR images of these patients were collected and stored following the local guidelines.

All the 457 CXR images were evaluated by the experienced radiologist. Without being aware of the COVID status of the patients the radiologist was able to assign these images into three categories: normal or no findings, findings consistent with non-COVID pneumonia and findings consistent with COVID pneumonia. The images were subsequently evaluated by the AI system which predicted the outcome based on the highest likelihood percentage values in the same categories. There was no input required on the COVID status of the patient before generating the inferences. After generating the inference there was an option to add the COVID status of the patient for the purpose of training the software. The data was tabulated in an Excel spreadsheet for each case by a third person. Neither the radiologist nor the AI had access to each other's findings beforehand. The results were evaluated and compared with various statistical tests.

Statistical analysis

All statistical analysis was done using SPSS software with version 25.0.

Receiver operating characteristic (ROC) curve analysis was used to test the performance of the AI system and radiologist. Area under the ROC curve (AUC) was evaluated. Sensitivity and specificity were calculated. The expert radiologist's findings and the AI prediction were used as the screening test as opposed to the RT-PCR results, which were used as reference gold-standard.

The McNemar test was used to obtain a statistically significant difference in the performance of the radiologist and AI system in detecting lung involvement in COVID-19 pneumonia.

Throughout the results, 5% level of significance was used and all results were shown with 95% of confidence. *P* value <0.05 was considered as significant.

Kappa statistics was used to find out the agreement between two variables.

The artificial intelligence system

NeuraCovid is a web application used to detect COVID-19 from CXR images developed by Neura Health Inc. (founded in June 2019). The application pairs with the AI model COVID-Net, built and released by University of Waterloo, Canada. NeuraCovid was released under AGPL license to make the COVID-Net AI Model more easily accessible and feasible for public use. This application is to be used for investigational purpose only.

COVID-Net is a deep learning neural network designed for the detection of COVID-19 cases from CXR images that is open source and available to the general public. COVID-Net is one of the first open source network designs for COVID-19 detection from CXR images at the time of initial release. COVIDx is an open access dataset that was generated using 13,975 CXR images across 13,870 patient cases. COVIDx is the largest open access benchmark dataset available in terms of the number of publicly available COVID-19 positive cases.

The COVID-Net AI model team has also computed the test accuracy, as well as sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for each infection type, on the COVIDx dataset. It was observed that COVID-Net has a good sensitivity for COVID-19 cases (91% sensitivity). Secondly, COVID-Net achieved high PPV for COVID-19 cases (98.9% PPV), which indicates very few false positive rates.[19] This high PPV is important given that too many false positives would increase the burden for the healthcare system due to the need for additional PCR testing and additional care. Therefore, based on these results, it can be seen that COVID-Net performs well as a whole in

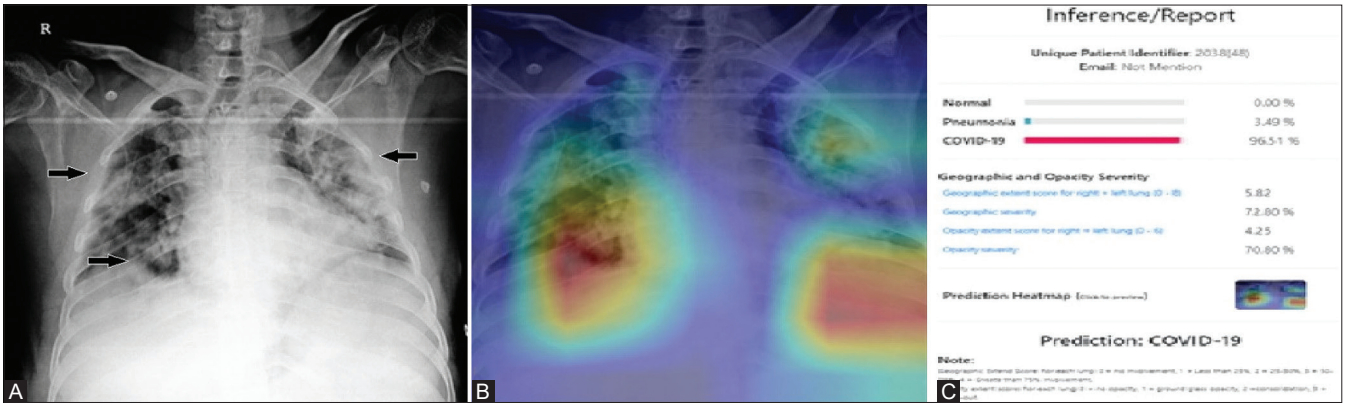

Figure 1 (A-C): This figure demonstrates the inference generated by the AI system in a 48-year-old RT-PCR positive male patient. (A) Plain PA chest radiograph of the patient shows multiple areas of inhomogeneous opacities predominantly in the peripheral zones (arrows) and was classified as COVID pneumonia by the radiologist. (B) Heat Map image of the corresponding radiograph showing areas of involvement. (C) Inference report generated by the AI system predicting it as COVID pneumonia with a 96% likelihood ratio. Note the geographic and opacity severity scores generated by the system

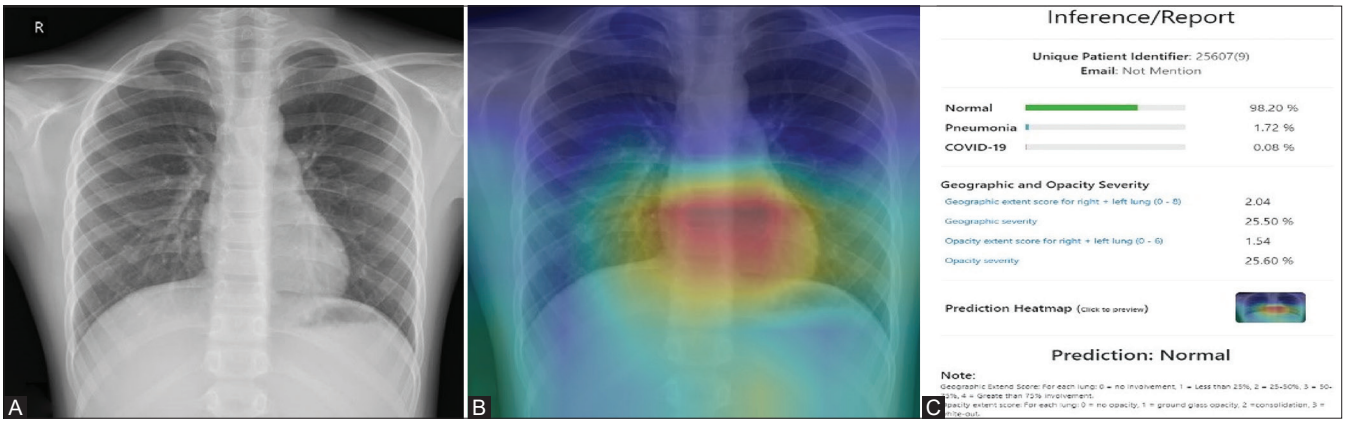

Figure 2 (A-C): (A) PA chest radiograph of a 9-year-old RT-PCR negative male patient shows no significant lung abnormality. (B) The Heat Map image of the corresponding radiograph shows no infected areas in both the lung fields. (C) The inference generated by the AI system predicts the radiograph as normal with a 98% likelihood percentage

detecting COVID-19 cases from CXR images. The accuracy for the large COVID- Net Model was assessed to be 93.3%.[19]

The NeuraCovid AI model runs in secure HIPAA compliant Google Cloud. The input image along with the demographic details of the patient like age and gender are uploaded to the cloud platform where the machine learning model runs the inference and saves the results in a database. The results are then provided to the user on the user interface. History of uploaded images and Inferences on those images are also available for user to refer for later time.

The NeuraCovid software uses an authentication using Google account login. It uses AP or PA Chest radiograph images, age and gender inputs, to predict and classify the outcomes as normal (no infection), non-COVID pneumonia (e.g., viral, bacterial, fungal, etc.) and COVID-19 viral pneumonia with likelihood percentage values for each of these classes. The reason for choosing these three possible predictions is that NeuraCovid assisted the clinicians to decide which patients should be prioritized for RT-PCR testing for COVID-19 case confirmation. This feature also helps the physicians to decide the treatment since COVID-19 and non-COVID19 pneumonias have different treatment plans.

The software also provides a Heat Map of the given radiograph to give insight into how AI arrived at a conclusion. Prediction Heat Map image shows the infected portion of lungs for COVID-19 or pneumonia. Additionally, NeuraCovid provides a footnote on the inference that gives a detailed description about geographic and opacity severity scoring system [Figure 1].

Results

The test set included 457 CXR images, of which 295 (64.5%) were proven COVID-19 RT-PCR positives and 162 were proven negative. The mean age of the patients was 41.5 (+/- 15.4) years (55.1% male). The radiologist correctly labelled 131 images as COVID positive and 150 images as

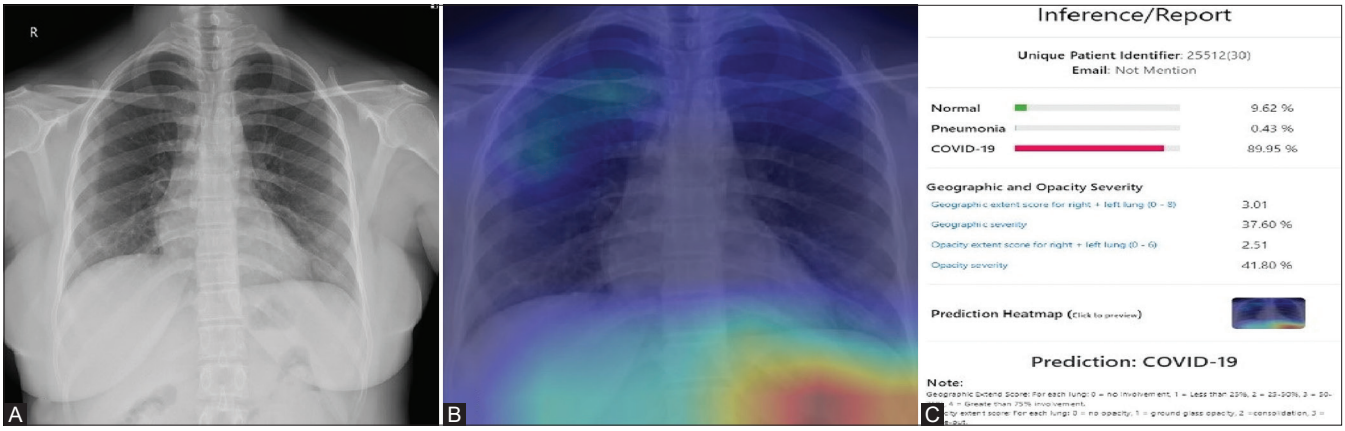

Figure 3 (A-C): (A) Shows a PA chest radiograph of a 30-year-old RT-PCR negative female patient with no obvious lung opacities (labelled as negative by the radiologist). (B) The misinterpreted Heat Map image of the corresponding radiograph shows infected areas which is seen the right upper and mid zone and outside the lung fields. (C) The inference generated by the AI system predicts the normal radiograph as COVID with a likelihood of 89% (false positive)

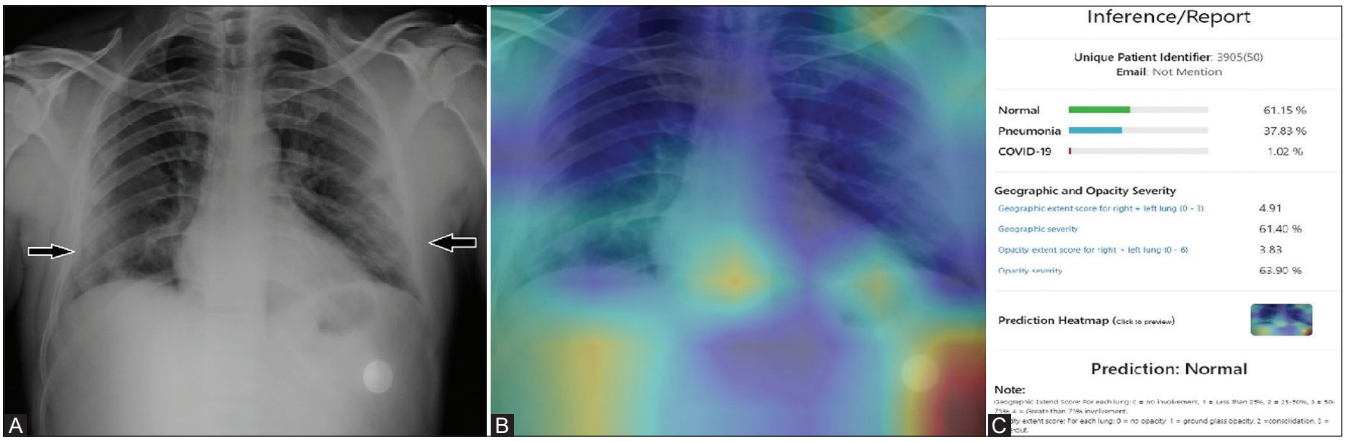

Figure 4 (A-C): (A): PA chest radiograph of a 50-year-old RT-PCR positive female patient with peripheral subpleural lung opacities (arrows) in both the lower zones which was classified as COVID positive by the radiologist. (B): The Heat Map image of the corresponding radiograph shows no areas of lung involvement. (C): The inference generated by the AI system predicts the radiograph as normal with a likelihood of 61% (false negative)

COVID negative (including no findings, findings consistent with non-COVID pneumonia and findings consistent with abnormalities other than pneumonia). The AI system correctly identified 123 images as COVID positive and 96 as negative (including normal and non-COVID pneumonia categories) [Figures 1 and 2]. 172 out of 295 RT-PCR proven positives were classified as negative by the AI (false negatives) and 66 out of 162 RT-PCR proven negatives were classified as positive (false positives) [Figures 3 and 4]. Sensitivity and specificity were calculated by applying standard methods using true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Sensitivity was calculated as TP/(TP+FN). Specificity was calculated as TN/(TN+FP). The sensitivity and specificity obtained were 44.1% and 92.6%, for the radiologist and 41.6% and 60%, for the AI system in correctly classifying COVID pneumonia on CXR images. In terms of sensitivity, the performance of AI system was comparable to the radiologist [Tables 1 and 2]. The ROC results for the reader and the AI using RT-PCR results as the gold standard were generated [Figures 5 and 6]. The radiologist and AI achieved an area under ROC curve of 0.68 and 0.48 respectively [Tables 3 and 4]. The AI system achieved a positive predictive value (PPV) of 65% and a negative predictive value (NPV) of 35.6%. This result was compared to the radiologist (PPV, 91.6%; NPV, 47.7%). The agreement between AI and gold standard was 0.08 (no agreement) [Figure 7].

Discussion

In a recent study, Keelin *et al*. compared the performance of a different AI system (CAD4COVID-Xray developed in Netherlands) to the results obtained after CXR interpretation by six independent radiologists and found that the performance of AI system was comparable or even better at high-sensitivity operating points. However, in our study, we used an independent test set to evaluate the radiological detection of COVID-19 pneumonia by a single radiologist and an AI system separately. We calculated

Table 1: The comparison of the sensitivities, specificities and area under the ROC curve with their *P* between the radiologist and AI

	Sensitivity (%)	Specificity (%)	Area under ROC curve	*P*
AI prediction	41.6	60	0.48	0.483
Radiologist's prediction	44.1	92.5	0.68	<0.001

Table 2: Difference between the AI and radiologist's detection of COVID-19 with a statistically significant *P*

	Sensitivity (%)	Specificity (%)	*P*
AI prediction	41.6	60	0.005
Radiologist's prediction	44.1	92.5	

Table 3: This table demonstrates the area under the ROC curve and *P* at 95% confidence interval for the radiologist

Area Under the Curve				
Test Result Variable(s): Expert				
Area	Std. Error	*P*	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.680	0.025	<0.001	0.631	0.728

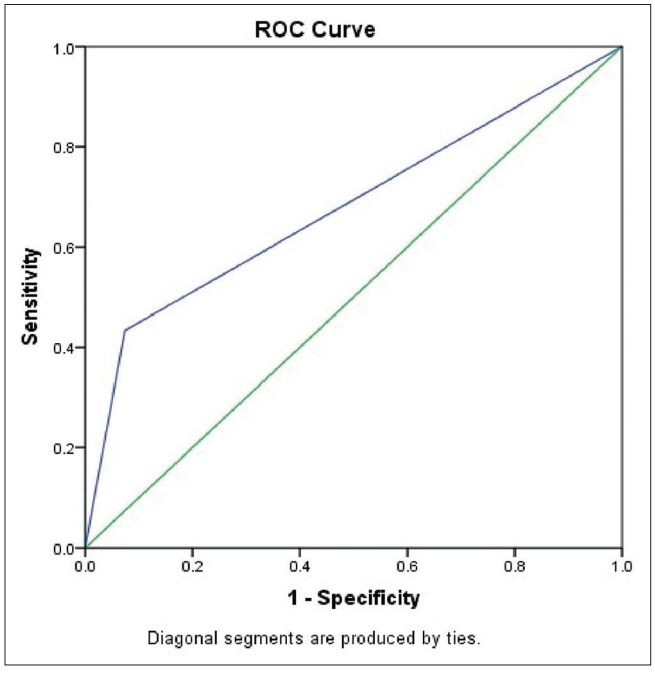

Figure 5: Diagnostic test comparison of radiologist with gold standard RT-PCR

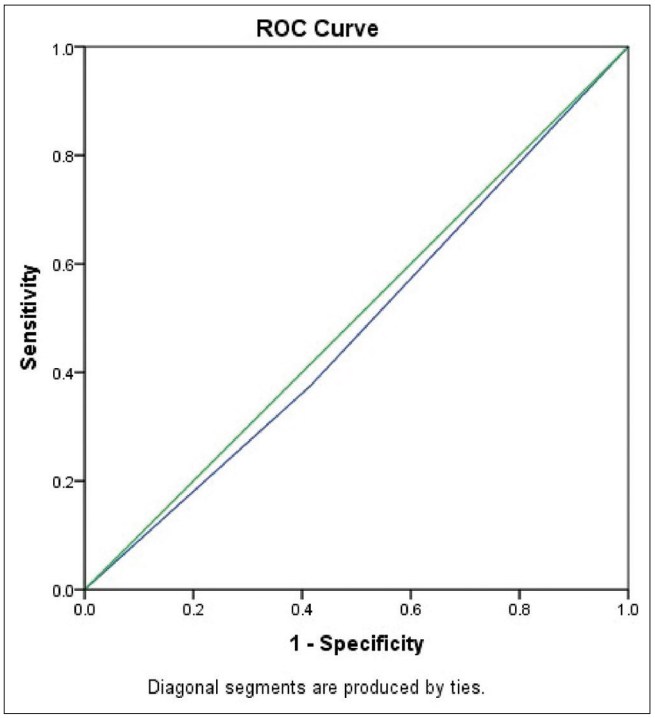

Figure 6: Diagnostic test comparison of AI with gold standard RT-PCR

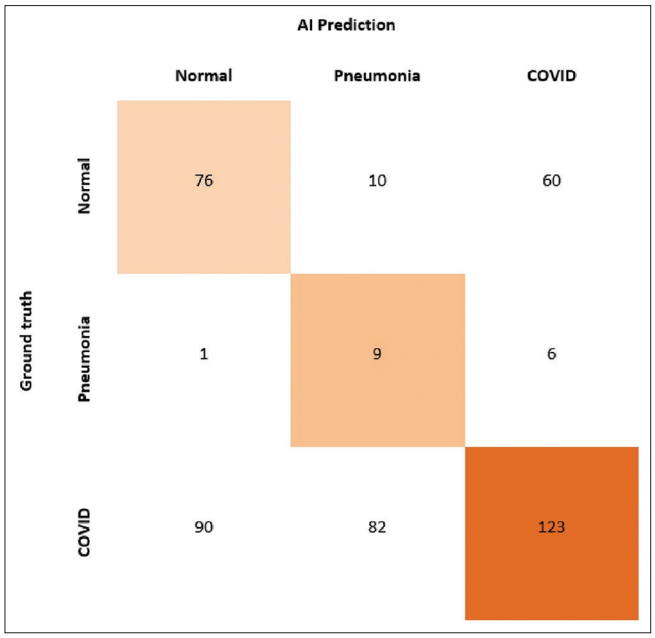

Figure 7: Shows the agreement between the gold standard and AI

Table 4: This table demonstrates the area under the ROC curve and *P* at 95% confidence interval for the AI system

Area Under the Curve				
Test Result Variable(s): Computer				
Area	Std. Error	*P*	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.480	0.029	0.483	0.424	0.536

the sensitivity, specificity, PPV and NPV for both the radiologist and the AI system and have compared the two. The radiologist's prediction was found to be superior to that of AI.

To improve the radiological detection of COVID-19 on CXR by AI system, a larger training set of CXR images is required. Additionally, the RT-PCR results, clinical and lab findings can also be included to accurately train the system. Also, the role of AI in triaging of patients during the COVID-19 pandemic should be further investigated by using a larger data set from different geographical locations and taking all relevant patient information and clinical findings into consideration. Greater sensitivity and specificity can be obtained when they are used in conjunction.

Our study had certain limitations. Our images were those of patients referred to only one institution, which may not represent the data for the whole population. Secondly, the sample size was small (457 CXR images). Also, the number of CXR of COVID-19 RT-PCR positive patients were more than the CXRs of non-COVID-19 patients because the study was performed during the peak of the pandemic. Also, the clinical data of the patients was not taken into the consideration. RT-PCR results were used as a gold standard which has a sensitivity of 60-70% in detection of COVID-19 infection.[13]

To summarise, the performance of the radiologist was superior to that of AI system in detection of changes of COVID-19 on chest radiographs. However, the AI system achieved a sensitivity which was comparable to that of the radiologist. This suggests that the AI system needs to be further trained with previously specified methods and its role to be used as a substitute in screening patients with suspected COVID pneumonia in the periphery or primary care centres be further investigated. The NeuraCovid web application is based on COVID-NET AI model to make it easily accessible and feasible for public use and surveillance in the COVID-19 triage process. In the interim, we would recommend tele-radiology as the way to assist early and accurate detection of COVID-19 pneumonia in the peripheral, low resource settings where there is scarcity of radiologists. With the advent of picture archiving and communication system (PACS) there is an added benefit of monitoring the progression or regression of the disease based on CXRs findings, which is a feature that the current AI model lacks.

Declaration of patient consent

The authors certify that they have obtained all appropriate patient consent forms. In the form the patient(s) has/have given his/her/their consent for his/her/their images and other clinical information to be reported in the journal. The patients understand that their names and initials will not be published and due efforts will be made to conceal their identity, but anonymity cannot be guaranteed.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Meo SA, Alhowikan AM, Al-Khlaiwi T, Meo IM, Halepoto DM, Iqbal M, *et al*. Novel coronavirus 2019-nCoV: prevalence, biological and clinical characteristics comparison with SARS-CoV and MERS-CoV. Eur Rev Med Pharmacol Sci 2020;24:2012-9.
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, *et al*. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497-506.
3. WHO's Certified [Internet]. Pneumonia of unknown cause–China', Emergencies preparedness, response. Disease outbreak news. Available from: https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/. [Cited on 2020 Sept 9].
4. Worldometers.info[Internet].COVID-19 coronavirus pandemic. Available at: https://www.worldometers.info/coronavirus/. [Updated on 2020 Sept 18; Cited on 2020 Sept 18].
5. Wu Z, McGoogan JM. Asymptomatic and pre-symptomatic COVID-19 in China. Infect Dis Poverty 2020;9:72.

6. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, *et al*. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. JAMA 2020;323:1061-9.
7. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, *et al*. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eur Surveill 2020;25:2000045.
8. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: Issues affecting the results. Expert Rev Mol Diagn 2020;20:453-4.
9. Watson J, Whiting PF, Brush JE. Interpreting a covid-19 test result. BMJ 2020;369:m1808.
10. Yang W, Sirajuddin A, Zhang X, Liu G, Teng Z, Zhao S, *et al*. The role of imaging in 2019 novel coronavirus pneumonia (COVID-19). Eur Radiol 2020;30:4874-82.
11. Jacobi A, Chung M, Bernheim A, Eber C. Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. Clin Imaging 2020;64:35-42.
12. Zu ZY, Jiang MD, Xu PP, Chen W, Ni QQ, Lu GM, *et al*. Coronavirus disease 2019 (COVID-19): A perspective from China. Radiology 2020;296:E15-25.
13. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, *et al*. Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. Radiology 2020;296:E115-7.
14. Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, *et al*. The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the Fleischner Society. Chest 2020;158:106-16.
15. Durrani M, Inam ul Haq UK, Yousaf A. Chest X-rays findings in COVID 19 patients at a University Teaching Hospital-A descriptive study. Pak J Med Sci 2020;36:S22-6.
16. Ng MY, Lee EY, Yang J, Yang F, Li X, Wang H, *et al*. Imaging profile of the COVID-19 infection: Radiologic findings and literature review. Radiology: Cardiothoracic Imaging 2020;2:e200034.
17. Antonio GE, Ooi CG, Wong KT, Tsui ELH, Wong JSW, Sy ANL, *et al*. Radiographic-clinical correlation in severe acute respiratory syndrome: Study of 1373 patients in Hong Kong. Radiology 2005;237:1081-90.
18. Toussie D, Voutsinas N, Finkelstein M, Cedillo MA, Manna S, Maron SZ, *et al*. Clinical and chest radiography features determine patient outcomes in young and middle age adults with COVID-19. [published online ahead of print, 2020 May 14]. Radiology 2020;29:E197-206.
19. Wang L, Wong A, Lin Z A. COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images. arXiv preprint 2020;arXiv: 2003:09871.
20. Murphy K, Smits H, Knoops AJ, Korst MB, Samson T, Scholten ET, *et al*. COVID-19 on the chest radiograph: A multi-reader evaluation of an ai system. Radiology 2020;296:E166-72.
21. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 2020;121:103792.