

1 **Validation of a Deep Learning Model to aid in COVID-19 Detection from**
2 **Digital Chest Radiographs**

3

4 Pranav Ajmera¹, Amit Kharat^{1,2}, Sanjay Khaladkar¹, Tanveer Gupte², Richa Pant^{2*}, Viraj
5 Kulkarni², Vinay Duddalwar³, Deepak Patkar⁴, Mona Bhatia⁵, Purnachandra Lamghare¹

6 ¹: Department of Radiodiagnosis, Dr DY Patil Medical College, Hospital and Research Center,
7 DY Patil Vidyapeeth, DPU, Pune, India

8 ²: DeepTek Medical Imaging Pvt. Ltd., Pune, India

9 ³: Department of Radiology and Biomedical Imaging, Keck School of Medicine, University of
10 Southern California, Los Angeles, CA, USA

11 ⁴: Department of Radiodiagnosis, Nanavati Super-speciality Hospital, Mumbai, India

12 ⁵: Department of Radiodiagnosis, Fortis Escorts Hospital, New Delhi, India

13

14 *Corresponding author: richa.pant@deeptek.ai

15

16

17

18

19

1 **Abstract**

2

3 **Introduction:** Using artificial intelligence in imaging practice helps ensure study list
4 reprioritization, prompt attention to urgent studies, and reduces the reporting turn-around time.

5

6 **Purpose:** We tested a deep learning-based artificial intelligence model that can detect COVID-
7 19 pneumonia patterns from digital chest radiographs.

8

9 **Material and Methods:** The deep learning model was built using the enhanced U-Net
10 architecture with Spatial Attention Gate and Xception Encoder. The model was named
11 DxCOVID and was tested on an external clinical dataset. The dataset included 2247 chest
12 radiographs comprising CXRs from 1046 COVID-19 positive patients (positive on RT-PCR) and
13 1201 COVID-19 negative patients.

14

15 **Results:** We compared the performance of the model with three different radiologists by
16 adjusting the model's sensitivity as per the individual radiologist. The area under the curve
17 (AUC) on the receiver operating characteristic (ROC) of the model was 0.87 [95% CI: 0.85,
18 0.89].

19

20 **Conclusion:** When compared to the performance of three expert readers, DxCOVID matched
21 the output of two of the three readers. Disease-specific deep learning models using current

1 technology are mature enough to match radiologists' performance and can be a suitable tool to be
2 incorporated into imaging workflows.

3

4 **Keywords:** COVID-19; Coronavirus disease; Artificial Intelligence; Deep Learning; Chest
5 Radiography

6

7 **Introduction**

8 The world has seen multiple pandemics over the previous two millennia, but most of them
9 occurred in the last two centuries. These included diseases like plague, Spanish flu, Asian Flu,
10 Hong Kong flu, HIV, Swine flu, and Severe Acute Respiratory Syndrome (SARS) (1). Like most
11 previous pandemics, Coronavirus Disease 2019 (COVID-19) primarily spreads via the
12 respiratory route. The defining characteristics of a pandemic include widespread geographical
13 extension, movement across territories, minimal pre-existing population immunity, and the
14 novelty of genetic makeup associated with a high attack rate (2). While most of the factors have
15 changed only a little over centuries, movement across continents has become very easy in the last
16 century, allowing the disease to spread across the globe before being detected. According to
17 World Health Organization (WHO) data from 2018, the availability of doctors in a developed
18 country like Germany was approximately 43 per 10,000 people, whereas in the United States of
19 America it was 23 per 10,000 people. The corresponding figures for developing nations like
20 India (9.3 doctors per 10000 people) were significantly lower. In both cases, the numbers are
21 insufficient to treat the general population, let alone provide adequate health care in the middle
22 of a pandemic (3). In 2011, radiologists comprised only 3.7% of the total physician workforce in
23 a developed nation like the USA, while in 2016 the number of radiographs performed globally

1 had steadily increased to approximately 3.6 billion per annum. COVID-19 has only caused a
2 further rise in the numbers (4,5).

3 The overall demand for chest radiography and Computed Tomography (CT) has increased as
4 these imaging modalities have become a part of the COVID-19 diagnosis and management
5 guidelines, as well as WHO recommendations because they allow the triage of the disease. While
6 chest radiography is less sensitive as compared to CT, it is also less resource-intensive, and
7 patients are exposed to lower doses of radiation. Moreover, in intensive care units, chest
8 radiography is a valuable diagnostic tool to monitor disease progression. As of February 09,
9 2022, the total number of COVID-19 cases during the pandemic stood at over 401 million, with a
10 mortality rate of nearly 5.7 million people globally (6). With the explosive increase in the
11 population affected by COVID-19 and the number of radiographic studies performed, the turn-
12 around time for radiology reports can be a bottleneck in the quick triage of studies that require
13 urgent attention. Using artificial intelligence to perform quick triage can ensure study list
14 reprioritization, ensuring prompt attention to urgent cases.

15 In the age of Computed Tomography (CT), our focus has been on developing a tool to assist
16 detection of COVID-19 from chest radiographs because of the volume of chest radiographs
17 performed daily and the emphasis placed on them for screening purposes. Even in Western
18 nations with easy access to CT, chest radiography accounts for 25% of the total diagnostic
19 procedures (7). Globally, these figures are even higher since chest radiography account for
20 approximately 40% of all diagnostic procedures (3.6 billion) performed annually (8). The insight
21 gained from an accurate interpretation of chest radiographs in COVID-19 can significantly alter
22 the treatment plan advised by the attending physician and often obviate the need for CT. Even in
23 Western nations, a chest radiograph is the first procedure to be advised for the purpose of

1 screening. According to UNSCEAR, CT accounts for 8% of all radiological examinations and
2 contributes to 47% of the total radiation dose in Level-I countries (9).

3 In this study, we present DxCOVID, a deep learning model for detecting COVID-19 from chest
4 radiographs. We also compare the performance of DxCOVID to the interpretations by three
5 board-certified radiologists.

6

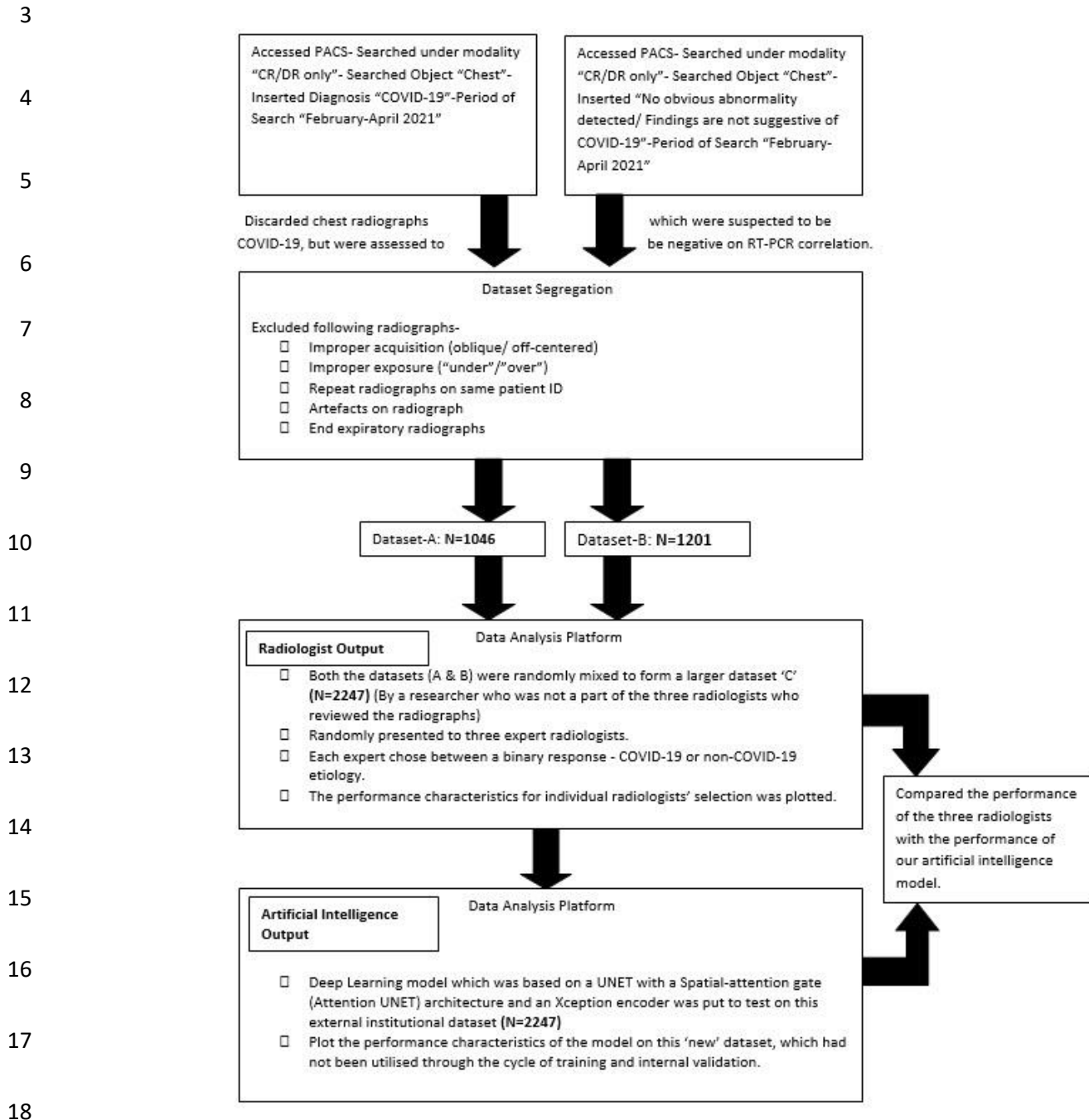
7 **Material and Methods**

8

9 *Data acquisition:*

10 This retrospective study was approved by the Institutional Review Board of our hospital. The
11 data was anonymized, and explicit written informed consent was waived in view of the
12 retrospective nature of the data collection and usage of completely anonymized patient datasets.
13 The data was a mixed dataset of inpatient and outpatients collected from February 2021 to April
14 2021. This included 2247 patients, out of which 1046 were COVID-19 positive on RT-PCR, and
15 the rest 1201 tested negative for COVID-19. The radiographs with negative COVID-19 RT-PCR
16 reports were either normal or had other non-COVID-19 pathologies. Some patients required
17 multiple radiograph examinations as their disease progressed. However, we avoided using
18 multiple radiographs of the same patient as there is uncertainty regarding the performance
19 evaluation of the deep learning model in such a case. To remove any bias in such a situation, a
20 single radiograph per person was selected randomly. Hence, each radiograph corresponded to a
21 single patient. De-identified chest radiographs that were well-exposed, well-centered, and
22 covered the complete field of interest were included in the study. Poor quality radiographs that
23 were end-expiratory, oblique, underexposed, overexposed, or exhibited motion blur and

1 extraneous artifacts were excluded from the study. The methodology utilized for data collection,
2 segregation, and analysis has been depicted in Fig 1.



19 **Figure 1:** The detailed methodology used to acquire the external validation dataset has been
20 elaborated. The method encompassed entering the appropriate query to the search system and

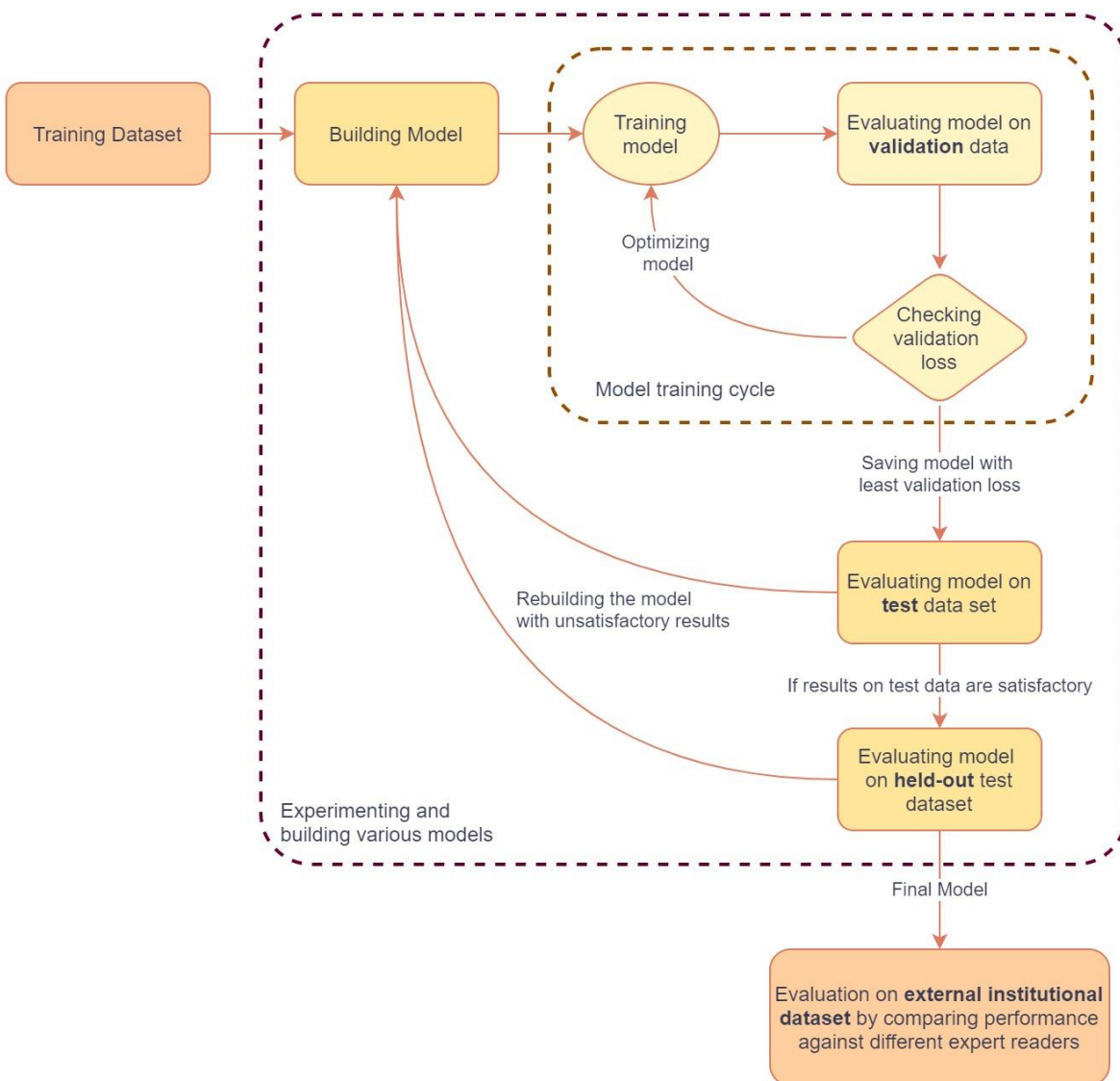
1 extracting the resultant data from PACS (the hospital stores and accesses data using MedSynapse
2 PACS). The data was segregated and finally uploaded onto a central data-analysis platform
3 developed by DeepTek Medical Imaging Pvt Ltd in a Health Insurance Portability and
4 Accountability Act (HIPAA) compliant manner. PACS: Picture Archiving and Communication
5 System, CR: Computed Radiography, DR: Digital Radiography.

6

7 *Artificial Intelligence (AI) algorithm:*

8 We used a multistep process involving cycles of training and retraining before further validating
9 or testing the AI model. The process involved data curation, anonymization, and pre-processing.
10 The architecture used to train the AI model was enhanced UNET with a Spatial-attention gate
11 (Attention UNET) and an Xception encoder. The processing step entailed checking for multiple
12 parameters like validation loss and optimizing it based on the feedback. As a result, each cycle
13 resulted in the development of a model that was trained and then modified based on the feedback
14 from validation. Post model optimization, the model was evaluated on the external institutional
15 hold-out dataset. This was done to minimize the overfitting of data. An illustration of the entire
16 pipeline is summarised in Fig 2.

17



1
2
3
4
5
6
7
8

Figure 2: A workflow depicting the systematic development of the COVID-19 detection algorithm. It starts from the training dataset and subsequently goes through a rigorous and repetitive process of training and validation. The optimized model was evaluated on the external institutional dataset.

1 *Panel of radiologists for comparison:*

2 A panel of three board-certified radiologists read the entire dataset and labeled each image as
3 COVID-19 positive or negative, individually. The panel included general radiologists with over
4 20-years of experience (R1), 23-years of experience (R2), and 27-years of experience (R3). The
5 heterogeneous group of readers was chosen to ensure that the models' output was tested against
6 results from radiologists with different durations of work experience. Radiologists provided a
7 binary output for each radiograph i.e., the radiograph was either marked positive for the presence
8 of COVID-19 pneumonia pattern or negative for the given set of findings (focal consolidation,
9 multifocal consolidations in a pattern and distribution typical for COVID-19).

10

11 *Statistical analysis:*

12 The 95% Confidence Interval (CI) was calculated using an empirical bootstrapping method. To
13 evaluate the diagnostic performance of DxCOVID, the specificity, sensitivity, and accuracy were
14 calculated in the entire test cohort of 2246 radiographs. The area under the receiver operating
15 characteristic curve was calculated to assess the performance of the model.

16

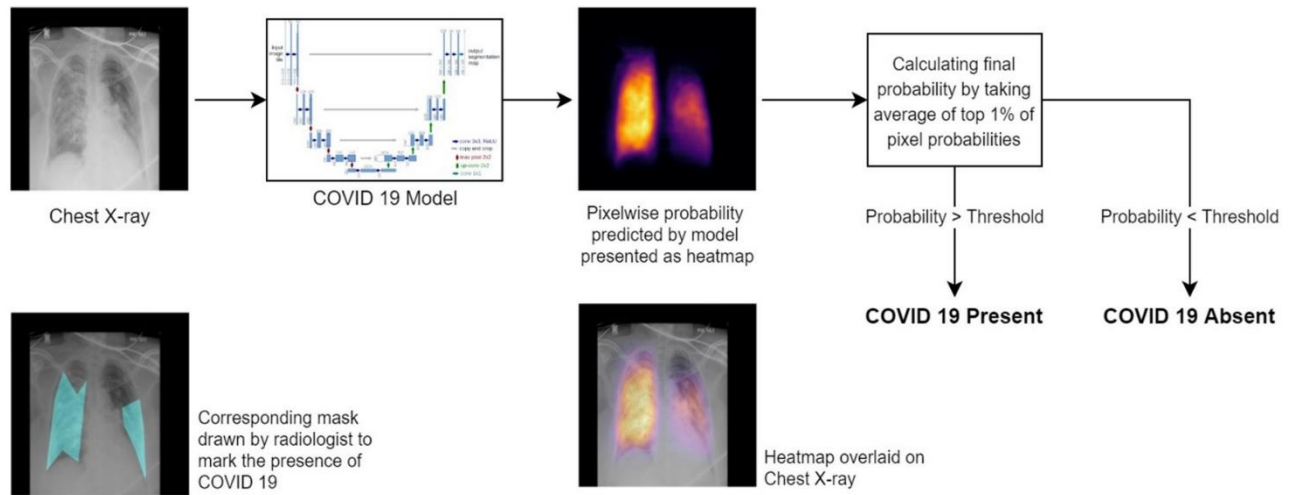
17 **Results**

18

19 *Model output:*

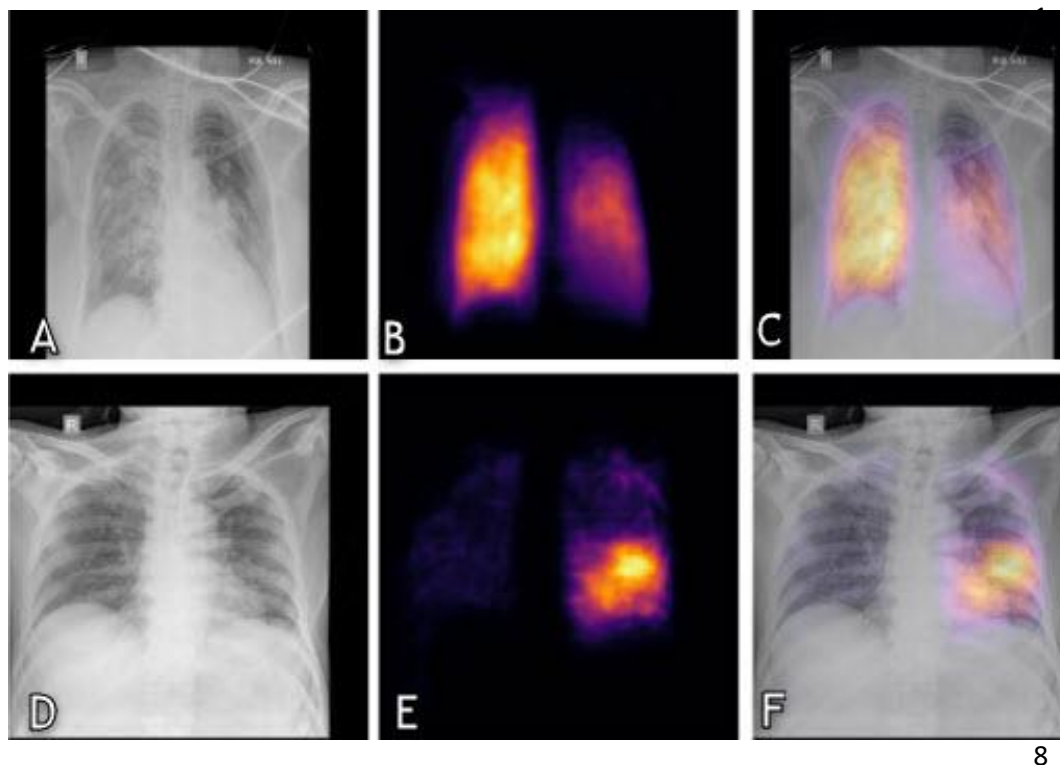
20 The output of the model contained a pixel-wise probability for the input image. The top 1% of
21 predicted pixels were averaged out to obtain an overall probability of COVID-19. These

1 probability scores were used to calculate the performance of the model. An illustration of the
2 same is represented in Fig 3.



3
4 **Figure 3:** An illustration to depict the intrinsic working of the COVID-19 algorithm. The chest
5 radiograph passes through multiple layers of the convolutional neural network and the output of
6 the model is depicted using a heatmap. To simplify the process of assessment of the chest
7 radiograph, the heatmap is superimposed on the chest radiograph. The final binary output of the
8 presence or absence of COVID-19 was obtained by averaging out the top 1% of the pixel
9 probabilities.

10
11 The possible localization of the manifestation of COVID-19 pneumonia is represented as a heat
12 map by the model. The intense yellow color represents the high confidence score from which the
13 model takes the decision to classify the area as COVID-19 positive or negative. To simplify the
14 process of assessment of the chest radiograph, heatmaps were superimposed on chest
15 radiographs (Fig 4).



8

9 **Figure 4:** The panel of images comprises input chest radiographs, AP view (A, D). Once these
10 chest radiographs were analyzed by the U-Net-based AI system, the system generated the
11 heatmaps as depicted in corresponding images B, E. The final output (C, F) is also shown as an
12 overlay of the AI model localizing the pathology on the chest radiograph, which helps the
13 clinician in identifying suspicious regions affected with COVID-19 pneumonia.

14

15 *Model performance:*

16 The trained AI model was evaluated on the test dataset. Table 1 summarises the performance of
17 DxCOVID (sensitivity, specificity, and accuracy) with its sensitivity adjusted to individual
18 expert readers. The receiver operating characteristic (ROC) curve was used to evaluate the
19 diagnostic accuracy of the model (Fig 5).

1 *Comparison of AI model with radiologists:*

2 The study design consisted of a multiparity assessment involving three board-certified
3 radiologists (R1, R2, and R3). All readers were blinded to patient information and were
4 instructed to read the radiographs independently and label them as positive or negative for
5 COVID-19. Radiologists were trained to flag studies as suspicious for COVID-19 if the
6 radiographs showed indications of either focal consolidation, multifocal consolidations in a
7 pattern, and distribution typical for COVID-19, i.e., ground-glass opacity/consolidation in a
8 predominantly peripheral distribution.

9 We compared the performance of DxCOVID with all three radiologists. As shown in Table 1,
10 the expert reader (R1) exhibited a sensitivity of 0.75 [95% CI: 0.72, 0.78], specificity of 0.81
11 [95% CI: 0.79, 0.83], and an accuracy of 0.79 [95% CI: 0.78, 0.81]. The second reader (R2) had
12 low sensitivity, i.e., 0.49 [95% CI: 0.45, 0.52], but extremely high specificity i.e., 0.96 [95% CI:
13 0.95, 0.97] for detecting COVID-19 pneumonia. The third reader (R3) had highest sensitivity
14 i.e., 0.77 [95% CI: 0.74, 0.8], and accuracy i.e., 0.83 [95% CI: 0.82, 0.85] in detecting COVID-
15 19 pneumonia. When compared to the performance of the three expert readers, the DxCOVID
16 model matched the output of two of the three readers (R1 and R2 with the difference being
17 statistically insignificant). The ROC results for all three expert readers and DxCOVID model are
18 depicted in Figure 5. The model achieved the area under the curve of 0.87 [95% CI: 0.85, 0.89].

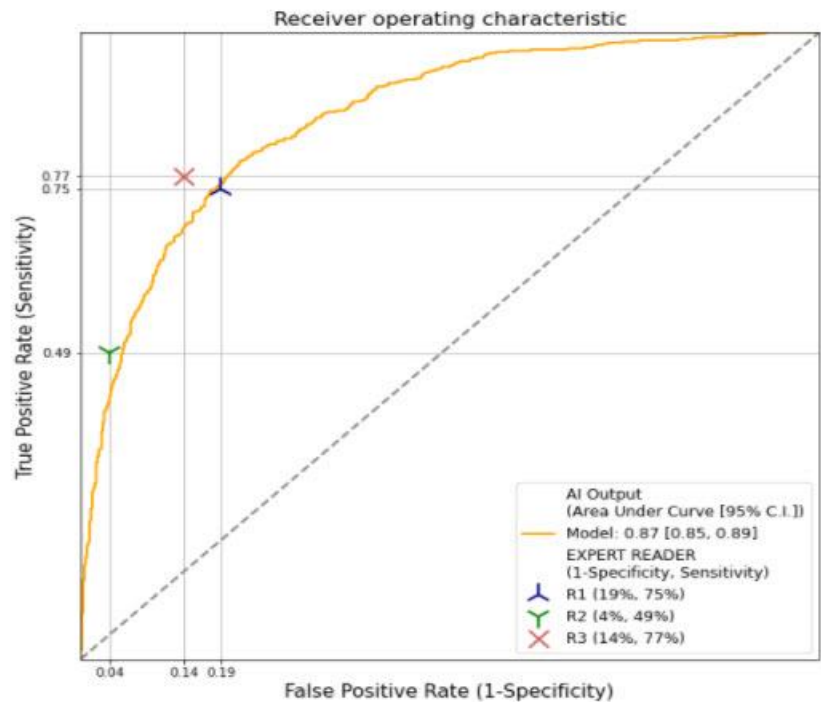
19
20
21
22
23

	Expert Reader R1	DxCOVID
Sensitivity	0.75 [0.72, 0.78]	0.75 [0.72, 0.78]
Specificity	0.81 [0.79, 0.83]	0.82 [0.80, 0.84]
Accuracy	0.79 [0.78, 0.81]	0.80 [0.78, 0.81]
Expert Reader R2		
	Expert Reader R2	DxCOVID
Sensitivity	0.49 [0.45, 0.52]	0.49 [0.45, 0.52]
Specificity	0.96 [0.95, 0.97]	0.94 [0.93, 0.95]
Accuracy	0.81 [0.8, 0.83]	0.80 [0.79, 0.82]
Expert Reader R3		
	Expert Reader R3	DxCOVID
Sensitivity	0.77 [0.74, 0.8]	0.77 [0.74, 0.8]
Specificity	0.86 [0.84, 0.88]	0.80 [0.78, 0.82]
Accuracy	0.83 [0.82, 0.85]	0.79 [0.77, 0.81]

1
2 **Table 1:** Performance of the COVID-19 detection model with sensitivity adjusted to match that
3 of the expert radiologists.

4
5
6

1
2
3
4
5
6
7
8
9



10
11
12
13
14

Figure 5: The receiver operator curve plots the output (AUC) of the three reviewing radiologists alongside the output of the externally validated AI model.

15 Discussion:

16 The Coronavirus Disease-2019 was declared as a Public Health Emergency of International
17 Concern on 30 January 2020. A month later, on 11 March 2020, the WHO officially elevated its
18 severity from an outbreak to a global pandemic, as multiple cases were being reported
19 simultaneously from countries spanning multiple continents (10,11). Epidemiologists,
20 researchers, physicians, and pharmaceutical companies backed by the government and/or
21 privately funded organizations pooled resources to assess and develop tools to diagnose the
22 disease and subsequently concentrate their efforts on the treatment aspect. The outcome was the

1 development of reverse transcriptase-polymerase chain reaction (RT-PCR) kits to diagnose
2 COVID-19. An array of blood panels including total leucocyte counts, ferritin, erythrocyte
3 sedimentation rate, C-reactive protein (CRP), lactate dehydrogenase (LDH), and D-dimer were
4 performed in all these cases, and while they could be linked to disease severity to varying
5 extents, utilizing them for diagnosing the disease proved difficult (12,13). Researchers explored
6 the usage of time-tested imaging modalities, such as chest radiography and computed
7 tomography (CT) for diagnostic purposes and subsequently documented the specific signs on
8 imaging for COVID-19 (14,15).

9 A detailed framework for the role of imaging in diagnosing COVID-19 was published by the
10 Fleischner Society (16). The document included elaborate algorithms depicting when and how to
11 utilize imaging techniques in COVID-19 and interpret them concurrently with other
12 investigations. With lockdown imposed in most of the countries alongside a shortage of beds and
13 medical equipment, patients with mild symptoms, particularly in the developing nations,
14 received treatment at home, while patients with more advanced symptoms were rushed to
15 emergency departments. In this subset of patients, radiography had the advantage of being
16 portable, quick, low in radiation exposure, and an adjunct tool allowing the monitoring of disease
17 progression (17).

18 While pneumonia and bronchopneumonia have characteristic imaging features, COVID-19
19 pneumonia has a wide range of imaging manifestations. COVID-19 radiographs have proven to
20 be difficult to read, particularly in patients with a mild form of the disease. This is because the
21 COVID-19 pneumonia pattern is similar to the pattern manifested in other infectious diseases,
22 such as pulmonary thromboembolism and pulmonary hemorrhage (18).

1 Extensive research in artificial intelligence and radiology has resulted in some actual tested and
2 working algorithms capable of detecting COVID-19. A study conducted by researchers from the
3 North-Western Memorial Health Care Center on a dataset of 5853 patients resulted in the
4 development of the DeepCOVID-XR algorithm. While their training and validation sets were
5 among the largest to date, the actual subset of data where comparisons to board-certified
6 radiologists were made was only 300 (19). In our case, the results of all 2247 patient datasets
7 were used to test the model were compare its performance with three board-certified
8 radiologists.

9 Zhang et al. developed a deep learning model called CV-19 that utilized a dataset of 2020
10 COVID-19 and 3148 non-COVID-19 patients. They utilized the test dataset of 500 radiographs
11 to ascertain the system's threshold that produced an AUC of 0.96, tested against the outputs of
12 three radiologists²⁰. Herein, we tested the model DxCOVID in a similar rigorous manner, and the
13 output of the AI model was plotted against the output of three experts. In our study, we avoided
14 utilizing multiple radiographs of the same patient and instead selected a single study for each
15 patient so that 2247 chest radiographs equated to 2247 patient datasets. By doing so, we avoided
16 the debate of whether utilizing multiple radiographs of the same patient could lead to a change in
17 performance evaluation, a concern voiced by Zhang et al. (20).

18 In a recent study by Roberts et al. (21), the researchers have evaluated the suitability of existing
19 models, their clinical relevance, the pitfalls of the various algorithms, and their findings. The
20 researchers concluded that while various AI models are available, none of the studies produced a
21 clinically relevant model, and therefore, the results are not generalizable to the masses. The
22 document reveals that, while the focus is on developing COVID-19 AI models in response to the
23 pandemic, the actual utility is still highly debatable in real-world practices. To that end, they

1 have identified several potential causes and have also recommended guidelines to be followed
2 while developing and testing the algorithms for enhanced comparisons. We took cues from this
3 study, and our model outputs were calibrated based on the individual radiologists' sensitivity,
4 rather than a separate ground truth or RT-PCR reports.

5 Our study has some limitations. First, the external test set was sourced from a single institution.
6 It is critical to test the AI algorithm across multiple datasets distributed across geographical
7 regions to ensure that the model's results are generalizable across cohorts and geographies.
8 Second, we utilized RT-PCR as the gold standard for the diagnosis of COVID-19 infections.
9 However, RT-PCR has a limited sensitivity of approximately 71%, so there may be cases where
10 the person is COVID-19 positive on chest radiographs but negative on RT-PCR results. Third,
11 our study does not incorporate clinical parameters and does not attempt at categorising patients
12 based on COVID-19 severity scores. We avoided providing COVID-19 scores based on chest
13 radiographs as unlike chest CT scans, there is often interobserver disagreement on the extent of
14 lung involvement in radiography for reasons encompassing different acquisition protocols, image
15 quality, and radiologist opinion. The development of an AI system based on the consensus
16 scoring of a few radiologists from an isolated geographical location may not represent the
17 consensus of radiologists globally. However, some researchers did attempt to build such a
18 system, like the one developed by Borghesi & Maroldi (22) on a small dataset of 100 patients.
19 Attempts were also made by Monaco et al. (23) with the dataset of 295 patients and Orsi et al.
20 (24) with the dataset of 155 patients to produce scoring systems for chest radiographs and link
21 them with certain key clinical and laboratory parameters. However, their results were limited by
22 factors from variable interobserver disagreements ranging from moderate to severe, a limited

1 dataset, datasets confined to narrow geographical locations, and the consensus scoring systems
2 were developed by a very limited number of radiologists.

3 In conclusion, we developed a DL model (DxCOVID) for detecting COVID-19 patterns from
4 digital chest radiographs. This model achieved an area under the curve (AUC) of 0.87 [95% CI:
5 0.85, 0.89] when tested on the institutional dataset of 2247 radiographs from a single institution.
6 When compared to the performance of the three expert readers, the DxCOVID model matched
7 the output of two of the three readers (R1 and R2). AI could serve as a reliable second reader,
8 augmenting the result interpretation by experts. Directions for future research include the
9 development of a successfully deployable model that incorporates various physical and clinical
10 parameters in addition to the imaging findings from multi-institutional datasets and provide a
11 combined output that can adequately prognosticate the patient.

12

13

14

15

16

17

18

19

20

21

22

1 **References:**

- 2 **1.** Huremović D. Brief history of pandemics (pandemics throughout history). In Psychiatry of
3 pandemics 2019 (7-35). Springer, Cham.
- 4 **2.** Morens DM, Folkers GK, & Fauci AS. What is a pandemic?. J Infect Dis 2009; 200; 1018–
5 1021.
- 6 **3.** [www.who.int/data/gho/data/indicators/indicator-details/GHO/medical-doctors-\(per-10-000-](http://www.who.int/data/gho/data/indicators/indicator-details/GHO/medical-doctors-(per-10-000-)
7 [population\)](http://www.who.int/data/gho/data/indicators/indicator-details/GHO/medical-doctors-(per-10-000-) [Accessed on 1st September, 2021]
- 8 **4.** Rosenkrantz AB, Hughes DR, Duszak Jr R. The US radiologist workforce: an analysis of
9 temporal and geographic variation by using large national datasets. Radiology; 2016; 279; 175-
10 184.
- 11 **5.** <https://www.who.int/news-room/feature-stories/detail/to-x-ray-or-not-to-x-ray-> [Accessed on
12 1st September, 2021]
- 13 **6.** <https://www.worldometers.info/coronavirus/> [Accessed on February 9, 2022]
- 14 **7.** Speets AM, van der Graaf Y, Hoes AW, et al. Chest radiography in general practice:
15 indications, diagnostic yield and consequences for patient management. Br J Gen Pract 2006; 56;
16 574-578.
- 17 **8.** Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates
18 of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;
19 68; 394-424. Erratum in: CA Cancer J Clin 2020; 70; 313.
- 20 **9.** Sun Z, Ng KH. Use of radiation in medicine in the Asia-Pacific region. Singapore Med J.
21 2012; 53; 784-788.
- 22 **10.** Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed 2020; 91; 157.

- 1 11. Bernheim A, Mei X, Huang M, et al. Chest CT findings in coronavirus disease-19 (COVID-
2 19): relationship to duration of infection. *Radiology* 2020; 20; 200463.
- 3 12. <https://www.who.int/teams/blueprint/covid-19> [Accessed on 10th August, 2021]
- 4 13. Stegeman I, Ochodo EA, Guleid F, et al. Cochrane COVID-19 Diagnostic Test Accuracy
5 Group. Routine laboratory testing to determine if a patient has COVID-19. *Cochrane Database*
6 *Syst Rev* 2020; 11: CD013787.
- 7 14. Fang Y, Zhang H, Xie J, et al. Sensitivity of chest CT for COVID-19: comparison to RT-
8 PCR. *Radiology* 2020; 296; e115-117.
- 9 15. Lyu P, Liu X, Zhang R, et al. The performance of chest CT in evaluating the clinical severity
10 of COVID-19 pneumonia: identifying critical cases based on CT characteristics. *Invest Radiol*
11 2020; 55; 412-421.
- 12 16. Rubin GD, Ryerson CJ, Haramati LB, et al. The role of chest imaging in patient management
13 during the COVID-19 pandemic: a multinational consensus statement from the Fleischner
14 Society. *Radiology* 2020; 296; 172-180.
- 15 17. Eisenhuber E, Schaefer-Prokop CM, Prosch H, et al. Bedside chest radiography. *Respir Care*
16 2012; 57; 427-443.
- 17 18. Duzgun SA, Durhan G, Demirkazik FB, et al. COVID-19 pneumonia: the great radiological
18 mimicker. *Insights Imaging* 2020; 11; 1-5.
- 19 19. Wehbe RM, Sheng J, Dutta S, et al. DeepCOVID-XR: an artificial intelligence algorithm to
20 detect COVID-19 on chest radiographs trained and tested on a large US clinical data set.
21 *Radiology* 2021; 299; e167-176.
- 22 20. Zhang R, Tie X, Qi Z, et al. Diagnosis of coronavirus disease 2019 pneumonia by using chest
23 radiography: Value of artificial intelligence. *Radiology* 2021; 298 ; e88-97.

- 1 21. Roberts M, Driggs D, Thorpe M, et al. Common pitfalls and recommendations for using
2 machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT
3 scans. *Nat Mach Intell* 2021; 3; 199-217.
- 4 22. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring
5 system for quantifying and monitoring disease progression. *Radiol Med* 2020; 125; 509-513.
- 6 23. Monaco CG, Zaottini F, Schiaffino S, et al. Chest x-ray severity score in COVID-19 patients
7 on emergency department admission: a two-centre study. *Eur. Radiol. Exp* 2020; 4; 1-7.
- 8 24. Orsi MA, Oliva G, Toluian T, Pittino CV, Panzeri M, Cellina M. Feasibility, reproducibility,
9 and clinical validity of a quantitative chest X-ray assessment for COVID-19.
10 *Am J Trop Med Hyg* 2020; 103; 822.

11

12

13

14

15

16

17

18

19