Henry Ford Health
Henry Ford Health Scholarly Commons

Diagnostic Radiology Articles | Diagnostic Radiology

9-24-2020

Diagnosis of COVID-19 Pneumonia Using Chest Radiography: Value of Artificial Intelligence

Ran Zhang

Xin Tie

Zhihua Qi
Henry Ford Health, zqi1@hfhs.org

Nicholas Bevins
Henry Ford Health, nickb@rad.hfh.edu

Chengzhu Zhang

See next page for additional authors

Follow this and additional works at: https://scholarlycommons.henryford.com/radiology_articles

Recommended Citation

Zhang R, Tie X, Qi Z, Bevins NB, Zhang C, Griner D, Song TK, Nadig JD, Schiebler ML, Garrett JW, Li K, Reeder SB, and Chen GH. Diagnosis of COVID-19 Pneumonia Using Chest Radiography: Value of Artificial Intelligence. Radiology 2020.

This Article is brought to you for free and open access by the Diagnostic Radiology at Henry Ford Health Scholarly Commons. It has been accepted for inclusion in Diagnostic Radiology Articles by an authorized administrator of Henry Ford Health Scholarly Commons.

Authors

Ran Zhang, Xin Tie, Zhihua Qi, Nicholas Bevins, Chengzhu Zhang, Dalton Griner, Thomas K. Song, Jeffrey D. Nadig, Mark L. Schiebler, John W. Garrett, Ke Li, Scott B. Reeder, and Guang-Hong Chen

This article is available at Henry Ford Health Scholarly Commons: https://scholarlycommons.henryford.com/radiology_articles/223

Radiology

Diagnosis of COVID-19 Pneumonia Using Chest Radiography: Value of Artificial Intelligence

Ran Zhang[1] PhD, Xin Tie[1] BS, Zhihua Qi[2] PhD, Nicholas B. Bevins[2] PhD, Chengzhu Zhang[1] BS, Dalton Griner[1] BS, Thomas K. Song[2] MD, Jeffrey D. Nadig[2], MD, Mark L. Schiebler[3] MD, John W. Garrett[3,1] PhD, Ke Li[1,3] PhD, Scott B. Reeder[3,1] MD/PhD, and Guang-Hong Chen[1,3] PhD

1. Department of Medical Physics, School of Medicine and Public Health, University of Wisconsin in Madison, Madison, WI 53705

2. Department of Radiology, Henry Ford Health System, Detroit, MI 48202

3. Department of Radiology, School of Medicine and Public Health, University of Wisconsin in Madison, Madison, WI 53792

Address correspondence to: Guang-Hong Chen, PhD

Department of Medical Physics, School of Medicine and Public Health

University of Wisconsin in Madison, Madison, WI 53705

Email: gchen7@wisc.edu

Funding: This work is partially supported by a grant from Wisconsin Partnership Program and a supplement grant from the National Institute of Biomedical Imaging and Bioengineering 3U01EB021183W4.

Manuscript type: Original research

In Press

This copy is for personal use only. To order printed copies, contact *reprints@rsna.org*

Summary Statement: An artificial intelligence algorithm differentiated between COVID-19 pneumonia and non-COVID-19 pneumonia in chest x-ray radiographs with high sensitivity and specificity.

Key Results:

- The overall performance of artificial intelligence (AI) algorithm achieved an area under the curve of 0.92 on the test dataset of 5869 chest x-ray radiographs (CXRs) from 2193 patients (acquired from multiple hospitals and multiple vendors)
- Over a set of 500 randomly selected test CXRs, the AI algorithm achieved an AUC of 0.94, compared to an AUC of 0.85 from three experienced thoracic radiologists.

Abbreviations:

AUC = area under the receiver operating characteristic curve, COVID-19 = coronavirus disease 2019, CXR = chest x-ray radiograph, ROC = receiver operating characteristic, RT-PCR = reverse transcriptase polymerase chain reaction, SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2

In Press

Abstract

Background

Radiologists are proficient in differentiating between chest x-ray radiographs (CXRs) with and without symptoms of pneumonia, but have found it more challenging to differentiate CXRs with COVID-19 pneumonia symptoms from those without.

Purpose

To develop an artificial intelligence algorithm to differentiate COVID-19 pneumonia from other causes of CXR abnormalities.

Materials and Methods

In this retrospective study, a deep neural network, CV19-Net, was trained, validated, and tested on CXRs from patients with and without COVID-19 pneumonia. For the COVID-19 positive CXRs, patients with reverse transcriptase polymerase chain reaction positive results for severe acute respiratory syndrome coronavirus 2 with positive pneumonia findings between February 1, 2020 and May 30, 2020 were included. For the non-COVID-19 CXRs, patients with pneumonia who underwent CXR between October 1, 2019 and December 31, 2019 were included. Area under the receiver operating characteristic curve (AUC), sensitivity, and specificity were calculated to characterize diagnostic performance. To benchmark the performance of CV19-Net, a randomly sampled test dataset containing 500 CXRs from 500 patients was evaluated by both the CV19-Net and three experienced thoracic radiologists.

Results

A total of 2060 patients (5806 CXRs; mean age 62 ± 16, 1059 men) with COVID-19 pneumonia and 3148 patients (5300 CXRs; mean age 64 ± 18, 1578 men) with non-COVID-19 pneumonia were included and split into training + validation and test datasets. For the test set, CV19-Net

In Press

achieved an AUC of 0.92 (95% confidence interval [CI]: 0.91, 0.93) corresponding to a sensitivity of 88% (95% CI: 87%, 89%) and a specificity of 79% (95% CI: 77%, 80%) using a high sensitivity operating threshold, or a sensitivity of 78% (95% CI: 77%, 79%) and a specificity of 89% (95% CI: 88%, 90%) using a high specificity operating threshold. For the 500 sampled CXRs, CV19-Net achieved an AUC of 0.94 (95% CI: 0.93, 0.96) compared to a 0.85 AUC (95% CI: 0.81, 0.88) of radiologists.

Conclusion

CV19-Net was able to differentiate COVID-19 related pneumonia from other types of pneumonia with performance exceeding that of experienced thoracic radiologists.

In Press

Introduction

The outbreak of coronavirus disease 2019 (COVID-19) (1) began with the initial diagnosis of an unknown viral pneumonia in late 2019 in Wuhan, China and subsequently spread around the globe as a pandemic. Ribonucleic acid sequencing of respiratory samples identified a novel coronavirus (called severe acute respiratory syndrome coronavirus 2 or SARS-CoV-2) as the underlying cause of COVID-19. Patients with COVID-19 present with symptoms that are similar to other viral illnesses, including influenza, as well as other coronaviruses such as severe acute respiratory syndrome (2,3) and Middle East respiratory syndrome (4). Symptoms are nonspecific and include fever, cough, fatigue, dyspnea, diarrhea, and even anosmia (5,6). The radiographic signs are also nonspecific and can be observed in patients with other viral illnesses, drug reactions, or aspiration (5,7,8).

The similarities in clinical presentation across other reactions and illnesses creates challenges towards establishing a clinical diagnosis. Currently, reverse transcriptase polymerase chain reaction (RT-PCR) is the reference standard method to identify patients with COVID-19 infection (9). In addition to the RT-PCR test, CT has also been widely used in China, and occasionally in other countries, to provide additional means in COVID-19 diagnosis and treatment response monitoring process (5,10,11). However due to concerns of contamination of CT imaging facilities and exposure to health care workers, healthcare professional organizations (12-14) do not recommend CT imaging as a general diagnostic imaging tool for patients with COVID-19.

Rather, major medical societies recommend the use of chest x-ray radiography (CXR) as part of the workup for persons under investigation for COVID-19 due to its unique advantages: almost all clinics, emergency rooms, urgent care facilities, and hospitals are equipped with stationary

In Press

and mobile radiography units, including both urban and rural medical facilities. These units can be easily protected from exposure or disinfected after use and can be directly used in a contained clinical environment without moving patients. However, the major challenge with the use of CXR in COVID-19 diagnosis is its low sensitivity and specificity in current radiological practice. A recent study found that the sensitivity of CXRs was poor for COVID-19 diagnosis (11). To some extent, this poor diagnostic performance can be attributed to the fact that many radiologists are seeing COVID-19 induced pneumonia cases for the very first time and radiologists need to read more cases to learn both the common and unique imaging features of this disease.

In this regard, machine learning, particularly deep learning (15,16) methods, have unique advantages in quick and tireless learning to differentiate COVID-19 pneumonia from other types of pneumonia using CXR images. The purpose of this study was to train and validate a deep learning method to differentiate COVID-19 pneumonia from other causes of CXR abnormalities and test its performance against thoracic radiologists.

Materials and Methods

This retrospective, Health Insurance Portability and Accountability Act -compliant study was approved by the Institutional Review Board at both Henry Ford Health System, Detroit, MI and the University of Wisconsin-Madison, Madison, WI. Written informed consent was waived because of the retrospective nature of the data collection and the use of de-identified images.

Patient Datasets

For algorithm development, we included CXRs from patients with and without COVID-19 (COVID-19 positive and non-COVID-19) pneumonia from Henry Ford Health System, which includes five hospitals and more than 30 clinics. The pneumonia findings for both COVID-19

In Press

and non-COVID-19 pneumonia were found using a commercial natural language processing tool (InSight, Softek Illuminate) that searched radiologist reports for positive pneumonia findings. Searches were performed over all radiologist reports at the institution over the COVID-19 and non-COVID-19 timeframes. The patients with non-COVID-19 pneumonia were selected based solely on positive pneumonia findings in the report and the date of study (October-December 2019). The patients with pneumonia from the COVID-19 timeframe were cross-referenced with the list of patients positive for COVID-19 to find the list of patients that had both positive pneumonia and positive COVID-19.

The inclusion criteria for the non-COVID-19 pneumonia were patients that underwent frontal view CXR, had pneumonia diagnosis, and imaging was performed between October 1, 2019 and December 31, 2019 (before the first COVID-19 positive patient in the United States was confirmed on January 19, 2020 in Seattle, WA [17]). Since these CXRs predate the first confirmed COVID-19 cases in the United States, we consider these CXRs to be positive for non-COVID-19 pneumonia. Patients under the age of 18 were excluded.

The inclusion criteria for the COVID-19 positive group were patients that underwent frontal view CXR, with RT-PCR positive test for SARS-CoV-2 with a diagnosis of pneumonia between February 1, 2020 and May 31, 2020. Patients were excluded if CXR was performed more than 5 days prior or 14 days after RT-PCR confirmation.

The resulting datasets consisted of 5805 CXRs with RT-PCR confirmed COVID-19 pneumonia from 2060 patients and 5300 CXRs with non-COVID-19 pneumonia from 3148 patients for use in this study (Figure 1 and 2).

In Press

CXR Acquisition

These CXRs were from six different vendors: Carestream Health (DRX-1, DRX-Revolution), GE Healthcare (Optima-XR220, Geode Platform), Konica Minolta (CS-7), Agfa (DXD40, DXD30, DX-G), Siemens Healthineers (Fluorospot Compact FD), and Kodak (Classic CR).

Training, Validation, and Test Datasets

It is important to consider any variables from CXR acquisition (such as x-ray tube potential [kVp values] and x-ray exposure levels) to mitigate any biases in algorithm training (for additional details see Appendix E1). Since our overarching objective was to develop a deep learning algorithm that could be successfully applied broadly to CXRs taken at different hospitals and clinics where CXR imaging systems from different vendors are used, our strategy was to train the deep learning method using a dataset with images from different vendor systems. CXRs were randomly selected from the four major vendors (Carestream Health, GE Healthcare, Konica Minolta, and Agfa) of the dataset and these vendors were randomly anonymized as V1, V2, V3 and V4. The curated CXRs were first grouped by vendors and a total of 5236 CXRs (2582 CXRs from the COVID-19 cohort and 2654 CXRs from the non-COVID-19 pneumonia cohort) were used as training and validation to develop our deep learning algorithm, which is referred to as CV19-Net.

The remaining data were used for performance evaluation of the developed CV19-Net algorithm, including 3223 positive COVID-19 CXRs from 1007 patients and 2646 non-COVID pneumonia CXRs from 1186 patients. A patient-based data partition scheme was used to ensure that CXRs of any particular patient will only appear in either the training dataset or test dataset, but not both. See Table 1 for details of the data partition.

In Press

Image Preprocessing in Machine Learning

The Digital Imaging and Communications in Medicine files of the collected CXRs were resized to 1024 x 1024 pixels and saved as 8-bit Portable Network Graphics grayscale images. Before being fed into the CV19-Net, images were further downscaled to 224 x 224 pixel, converted to red-green-blue images and normalized based on the mean and standard deviation of images in the ImageNet dataset (18). (See Appendix E2)

Neural Network Architecture and Training Strategy

The CV19-Net used in this work is an ensemble of 20 individually trained deep neural networks. Each deep neural network consists of four modules of the well-known DenseNet (19) architecture, with a binary classifier to differentiate COVID-19 pneumonia from other types of pneumonia. A three-stage transfer learning approach was used to train the 20 individual deep learning neural networks of the same architecture. After the CV19-Net was trained, an input CXR was fed into the CV19-Net to produce 20 individual probability scores, then a final score was generated by performing a quadratic mean. This process is similar to the group diagnosis protocol used in difficult clinical decision-making processes in that 20 individual "experts" are asked to evaluate the same input image, and then a final group score is generated by a voting scheme. This final probability score was then compared with a chosen decision-making threshold value to classify the input CXR images as COVID-19 or non-COVID-19 pneumonia (For details of the network architecture and the training process, see Appendix E3. The code is available at https://github.com/uw-ctgroup/CV19-Net).

Human Radiologists Reader Study to Generate Performance Reference

To benchmark the performance of the developed CV19-Net, three experienced thoracic radiologists (JDN, TKS, and MLS with more than 9, 14 and 34 years of experience, respectively)

In Press

performed binary classification (COVID-19 positive or COVID-19 negative) reader study using a randomly selected subset of the test images (Figure 1): 500 CXRs from 500 different patients (250 COVID-19 pneumonia and 250 non-COVID-19 pneumonia). All three readers have recent experience with COVID-19 CXR interpretation. The three readers were blinded to any clinical information and read all exams independently between June 1, 2020 and June 15, 2020. The three readers dictated each CXR as either COVID-19 positive or COVID-19 negative pneumonia using a picture archiving communication systems workstation under standard reading conditions. To compare the performance between CV19-Net and the three readers on the same test data set, the threshold of CV19-Net was adjusted to match the corresponding specificity of the radiologist and then the diagnostic sensitivity was compared between each radiologist and CV19-Net.

Statistical Analysis

To evaluate the diagnostic performance of the trained CV19-Net, the area under the receiver-operating-characteristic curve (AUC), sensitivity, and specificity were calculated over the entire test cohort of 5869 CXRs from 2193 patients. The 95% confidence intervals (CI) for the performance metrics were calculated using the statistical software R (version 4.0.0) with the pROC package (20). The CI for AUC was calculated using DeLong's nonparametric method (21); CIs for sensitivity and specificity were calculated using the bootstrap method (22) with 2000 bootstrap replicates. The McNemar test was performed to compare the sensitivity of CV19-Net to the three radiologists. *P*-value hypothesis testing method was used for each comparison (For details see Appendix E5). $P < .05$ was considered to indicate a statistically significant difference.

Results

In Press

Patient Overview

A total of 3507 (5672 CXRs) patients with non-COVID-19 pneumonia met the inclusion criteria. There were 359 patients (372 CXRs) that were under 18 years of age that were excluded. A total of 2086 patients (6650 CXRs) with COVID-19 pneumonia met the inclusion criteria and 340 patients (845 CXRs) were excluded for having CXRs performed outside of the preferred time window of RT-PCR (-5 to +14 days since positive test).

The resulting datasets that were used for the development (training + validation and testing) consisted of 5805 CXRs with RT-PCR confirmed COVID-19 pneumonia from 2060 patients (mean age, 62 ± 16 years; 1059 men) and 5300 CXRs with non-COVID-19 pneumonia from 3148 patients (mean age, 64 ± 18; 1578 men).

The data was randomized and partitioned based on data acquired on CXR equipment from different vendors. A total of 2654 CXRs (1962 patients) with non-COVID-19 pneumonia and 2582 CXRs (1053 patients) with RT-PCR confirmed COVID-19 were used for training and validation. A total of 2646 CXRs (1186 patients) with non-COVID-19 pneumonia and 3223 CXRs (1007 patients) with RT-PCR confirmed COVID-19 were used for CV19-Net testing, resulting in 5869 CXR images from 2193 patients (mean age 63 ± 16 years, 1131 men) within the test dataset (Figure 1).

Overall Performance of CV19-Net

The performance of the CV19-Net achieved an AUC of 0.92 (95% confidence interval [CI]: 0.91, 0.93) for the overall test dataset. As shown in Figure 3A and Table 2, for a high sensitivity operating threshold, this method showed a sensitivity of 88% (95% CI: 87%, 89%) and a specificity of 79% (95% CI: 77%, 80%); for a high specificity operating threshold, it showed a sensitivity of 78% (95% CI: 77%, 79%) and a specificity of 89% (95% CI: 88%, 90%).

In Press

The performance of CV19-Net for four major vendors and five major hospitals is presented in Figure 3C.

The three radiologists' interpretation results from the subset of 500 test images were summarized by sensitivities of 42%, 68%, and 90%, respectively, and specificities of 96%, 85%, and 55%, respectively. Using the interpretation results of the same image from three readers, an averaged receiver operating characteristic (ROC) curve with an AUC of 0.85 (95% CI: 0.81, 0.88) was generated for radiologists. As a comparison, when the CV19-Net was applied to the same sub-set of test images, it yielded an AUC of 0.94 and sensitivities of 71%, 87%, and 98%, respectively, and specificities of 96%, 85%, and 55%, respectively, when choosing a matched specificity to the performance of each radiologist (Figure 3B). All *P*-values were < .001, indicating CV19-Net had better sensitivity than human radiologists at all matched specificity levels. Figure 4 shows two example images in the reader study test dataset, which were correctly labeled by CV19-Net, but incorrectly labeled by all three radiologists. The heatmaps generated by CV19-Net are also shown in Figure 4. See Appendix E4 for details on the heatmap generation.

Performance by Age Group and Sex

The performance of CV19-Net is presented for patients with different age groups in Table 3 and for the two sexes in Table 4. There was no difference in CV19-Net performance between sex *(P* = .17). However, results showed a difference in performance between well-separated age groups (eg, age group of 18-30 years is different from age groups of 45-60 years [*P* = .02], 60-75 years [*P* = .002], and 75-90 years [*P* < .001]) while no difference in neighboring age groups (eg age groups 18-30 years compared to 30-45 years; *P* = .31) was found. See Table E1 for details.

In Press

Performance vs Training Sample Size

The relationship between the achievable AUC of CV19-Net vs the needed training sample sizes was systematically investigated to determine the training sample size used in this paper (See Figure E5). The results demonstrated that more than 3000 training samples (1500 positive COVID-19 cases and 1500 non-COVID-19) are needed to achieve an AUC better than 0.90. After the training sample size goes beyond 3000 the performance gain is diminished with the increase of training samples.

Discussion

It has been a routine clinical practice for radiologists to interpret chest x-ray radiographs with and without symptoms of pneumonia. However, it has been much more challenging to differentiate CXRs with COVID-19 pneumonia symptoms from those without due to the lack of the training in reading in this pandemic. In this work, we have demonstrated that an artificial intelligence algorithm can be trained and used to differentiate coronavirus disease 2019 (COVID-19) related pneumonia from non-COVID-19 related pneumonia using CXR images, with excellent performance on the same test image data set in terms of AUC of 0.94 (95% CI: 0.93, 0.96) compared to a 0.85 AUC (95% CI: 0.81, 0.88) of three thoracic radiologists.

Intensive efforts have been made globally through 2020 to seek fast and reliable machine learning solutions to help diagnose patients with COVID-19 and triage patients for proper allocation of rather limited resources in combating this global pandemic (See Table E2 for a summary of related studies). Most related studies used small datasets with fewer than 200 COVID-19 CXRs collected from various sources including cropped images from published journals or from authors' access to other image databases. Further, evaluations of these neural

In Press

networks were only performed over the same small data cohort. Due to the non-uniformity of image quality in these small datasets, the apparent test performances were often biased (23).

In contrast, two recent studies (24,25) reported their results using relatively larger data sets from clinical centers (one from Brazil with a total of 558 COVID-19 positive CXRs and the other from the Netherlands with a total of 980 COVID-19 positive CXR images used in both training and testing). Schwab et al (24) trained a small number of conventional machine learning algorithms from their dataset and reported an area under the curve (AUC) of 0.66 (95% confidence interval [CI]: 0.63, 0.70). In Murphy et al (25), a deep learning model was trained using 512 COVID-19 positive CXRs combined with 482 COVID-19 negative CXRs and reported a performance of AUC = 0.81 on CXRs from 454 patients. The potential variance of the reported AUC performance values remains unclear since there was no 95% CI reported. Their results were compared with that of six human radiologists, showing that the performance of their deep learning model is comparable with radiologists.

In our study, we systematically studied the performance of the trained deep learning model and how it changes with an increase of the training dataset size (For details, see Figure E5). With a training sample size of 1000 (500 positive and 500 negative cases), the achievable AUC was found to be 0.86, similar to what was reported (0.81) in Murphy et al (25). The slightly higher performance of our network may be attributable to differences in data curation strategies, as we included CXRs obtained contemporaneously with RT-PCR, within a narrow window (-5 to +14 days).

This study has several limitations. First, we only considered the binary classification task: COVID-19 pneumonia versus other types of pneumonia. Therefore, at this stage, the developed algorithm should be used in adjunction to radiologist's findings of pneumonia image features in

In Press

CXRs. For an automated artificial intelligence-assisted diagnostic system, it would be ideal to have finer classification categories such as “Normal”, “Bacterial”, “Non-COVID-19 viral”, and “COVID-19”. With global efforts in collecting CXRs with the above four labels, the work presented here may be further enhanced in future work. Second, the data collection of data from patients with COVID-19 pneumonia was conducted in the first peak of the COVID-19 pandemic. As a result, the collected data may not reflect the true prevalence of the disease. We also included multiple CXRs from the same patient since some patients took multiple exams as their diseases progress. One may question whether the use of multiple CXRs changes the performance evaluation, to address this question, a single CXR image was randomly selected from multiple CXRS per patient to participate in the overall test performance evaluation, and the overall AUC did not change from 0.92. Third, although the method was tested over multiple hospitals and clinics, the test sites need to be further expanded to determine whether the developed artificial intelligence algorithm in this work is generalizable to even broader population distributions over different regions and continents. Finally, in radiologist reader studies, only the averaged receiver operating characteristic (ROC) curve and the corresponding AUC was calculated based upon the diagnosis of each CXR from three readers. Thus, the reported ROC curve and AUC are averaged results from three independent readers. Ideally, each reader should have been asked to report their degree of confidence level in their diagnosis for each CXR and individual ROC and AUC for each reader can then be calculated and reported.

In conclusion, the combination of chest radiography with the proposed CV19-Net deep learning algorithm has the potential as an accurate method to improve the accuracy and timeliness of the radiological interpretation of COVID-19 pneumonia.

In Press

References

1. World Health Organization. Coronavirus disease (COVID-2019) situation reports. 2020. Available on: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200221-sitrep-32-covid-19.pdf.

2. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med. 2003 May 15;348(20):1953–66.

3. Kuiken T, Fouchier RAM, Schutten M, Rimmelzwaan GF, Van Amerongen G, Van Riel D, et al. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. Lancet. 2003 Jul 26;362(9380):263–70.

4. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, et al. Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group. J Virol. 2013 Jul 15;87(14):7790–2.

5. 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 Feb 15;395(10223):497–506.

6. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020 Apr 30;382(18):1708–20.

7. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT imaging features of 2019 novel coronavirus (2019-NCoV). Radiology. 2020 Feb 4;295(1):202–7.

8. Shi H, Han X, Zheng C. Evolution of CT Manifestations in a Patient Recovered from 2019 Novel Coronavirus (2019-nCoV) Pneumonia in Wuhan, China. Radiology. 2020 Apr 7;295(1):20–20.

In Press

9. US Food and Drug Administration. Accelerated emergency use authorization (EUA) summary COVID-19 RT-PCR test (Laboratory Corporation of America). 2020.

10. Caruso D, Zerunian M, Polici M, Pucciarelli F, Polidori T, Rucci C, et al. Chest CT Features of COVID-19 in Rome, Italy. Radiology. 2020 Apr 3;201237.

11. Yoon SH, Lee KH, Kim JY, Lee YK, Ko H, Kim KH, et al. Chest radiographic and ct findings of the 2019 novel coronavirus disease (Covid-19): Analysis of nine patients treated in korea. Korean J Radiol. 2020 Apr 1;21(4):498–504.

12. American College of Radiology. ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection. ACR website. https://www.acr.org/Advocacy-and Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection. Updated March. 2020 Mar 22.

13. Kim HW, Capaccione KM, Li G, Luk L, Widemon RS, Rahman O, et al. The role of initial chest X-ray in triaging patients with suspected COVID-19 during the pandemic. Emerg Radiol. 2020;1.

14. Simpson S, Kay FU, Abbara S, Bhalla S, Chung JH, Chung M, et al. Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. Radiol Cardiothorac Imaging. 2020;2(2):e200152.

15. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.

16. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. Cambridge: MIT press; 2016 Nov 18.

In Press

17. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med. 2020 Mar 5;382(10):929–36.

18. Deng J, Dong W, Socher R, Li L-J, Kai Li, Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In Institute of Electrical and Electronics Engineers (IEEE); 2010. p. 248–55.

19. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017. 2016 Aug 24;2017-Janua:2261–9.

20. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011 Mar 17;12(1):77.

21. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988 Sep;44(3):837.

22. Efron B. Nonparametric standard errors and confidence intervals. Can J Stat. 1981;9(2):139–58.

23. Maguolo G, Nanni L. A critic evaluation of methods for covid-19 automatic detection from x-ray images. arXiv preprint arXiv:2004.12823. 2020 Apr 27.

24. Schwab P, Schütte AD, Dietz B, Bauer S. predCOVID-19: A Systematic Study of Clinical Predictive Models for Coronavirus Disease 2019. arXiv preprint arXiv:2005.08302. 2020 May 17.

25. Murphy K, Smits H, Knoops AJG, Korst MBJM, Samson T, Scholten ET, et al. COVID-19 on the Chest Radiograph: A Multi-Reader Evaluation of an AI System. Radiology. 2020 May 8;201874.

In Press

Figures

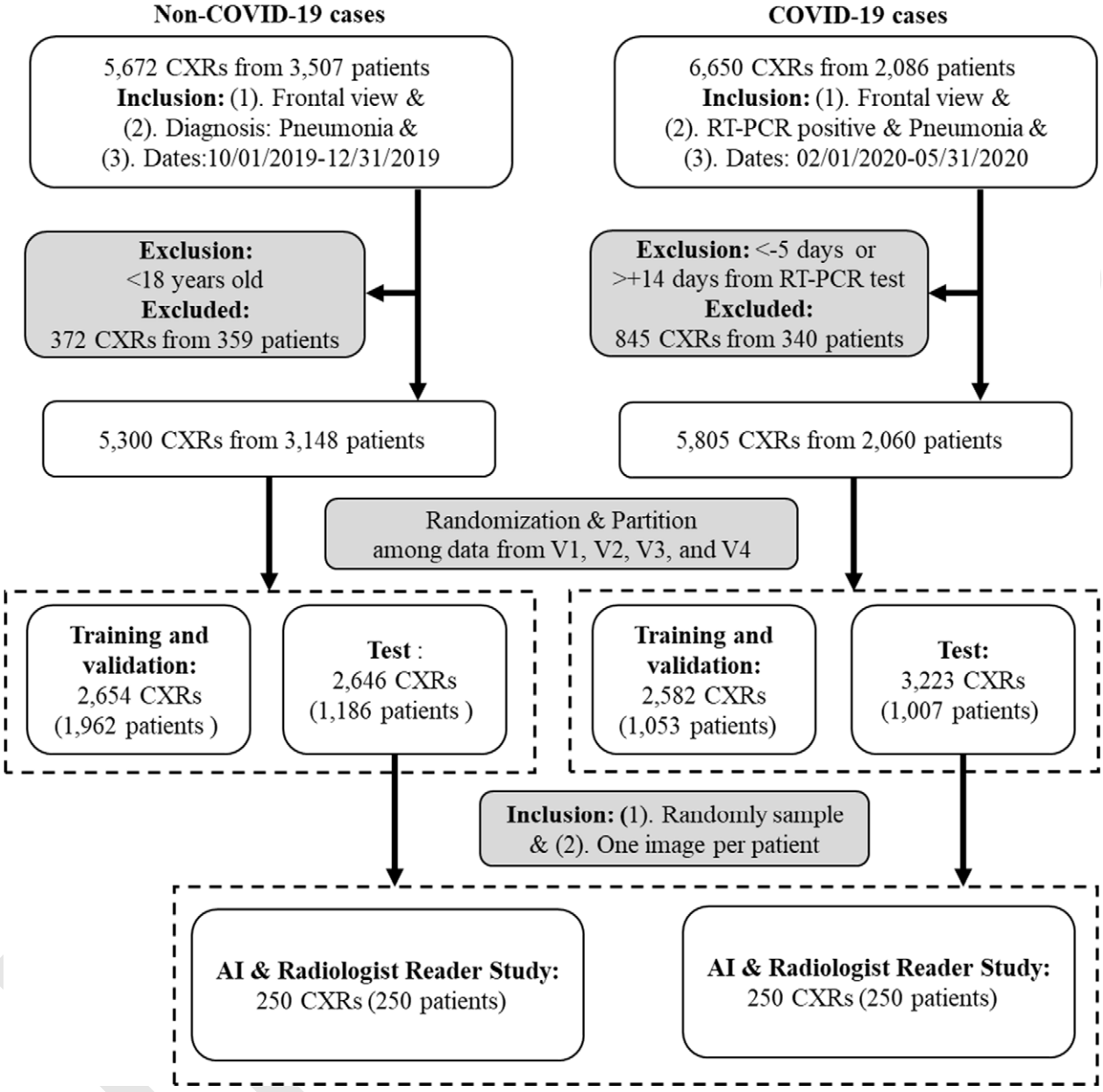
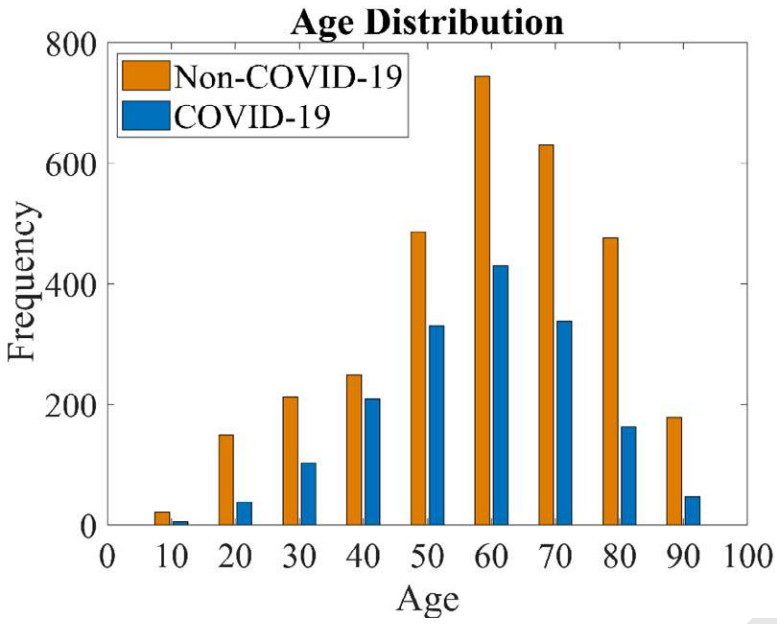
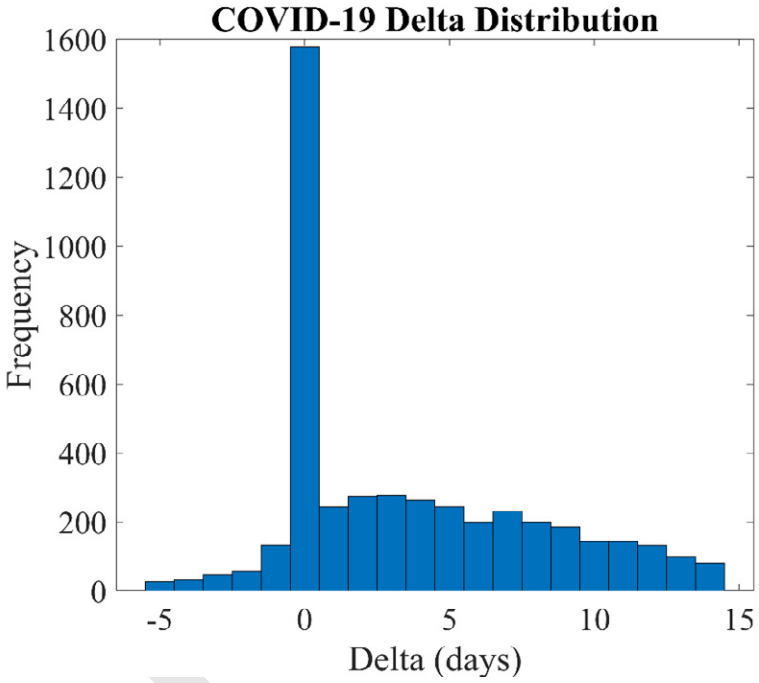


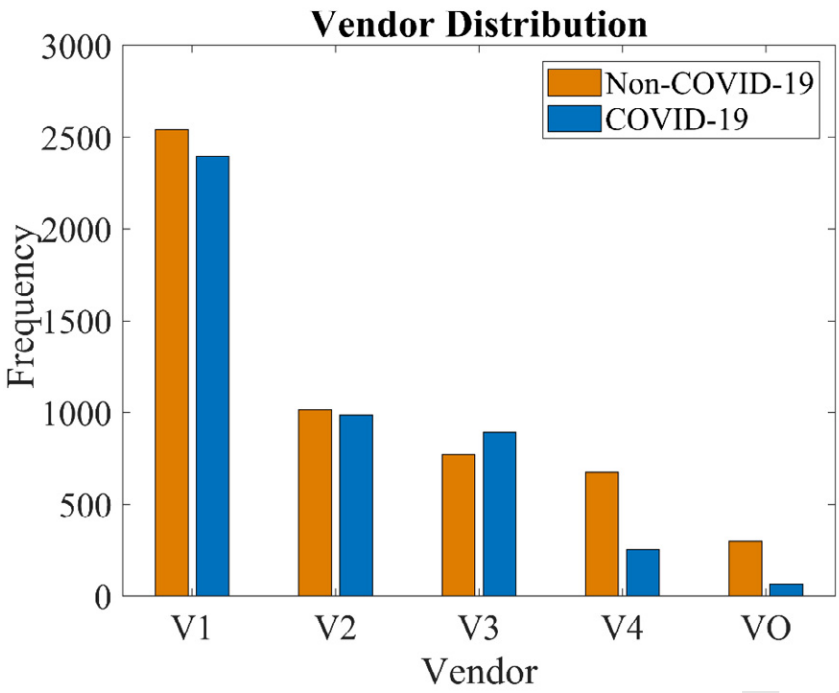
Figure 1: Study flowchart for data curation and data partition. Vendors 1-4 (V1-V4) are four major vendors of the acquired chest x-ray radiographs (CXR) in the dataset. AI = artificial intelligence, RT-PCR = reverse transcriptase polymerase chain reaction.


Age Distribution
Non-COVID-19
COVID-19
Frequency
Age
0
200
400
600
800
0
10
20
30
40
50
60
70
80
90
100

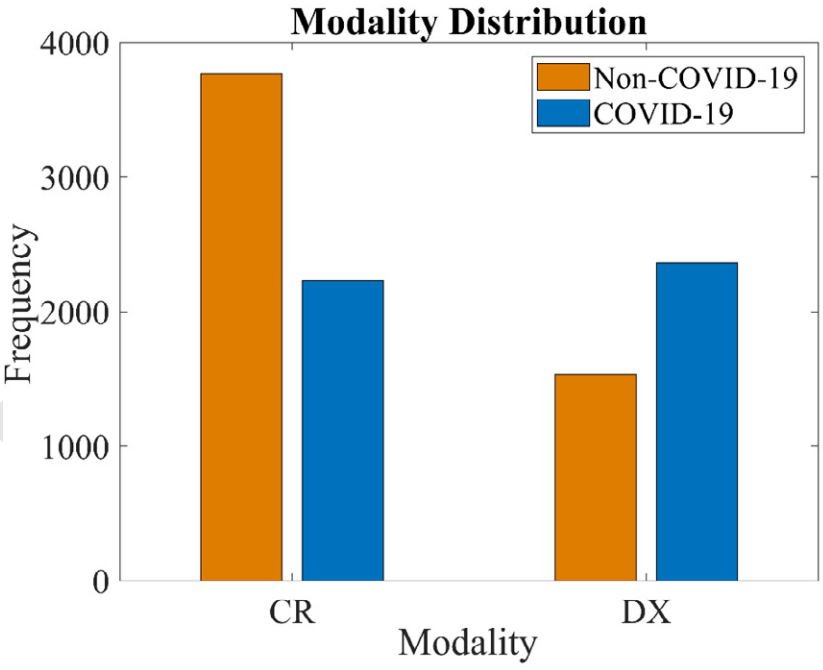
A


COVID-19 Delta Distribution
Frequency
Delta (days)
0
200
400
600
800
1000
1200
1400
1600
-5
0
5
10
15

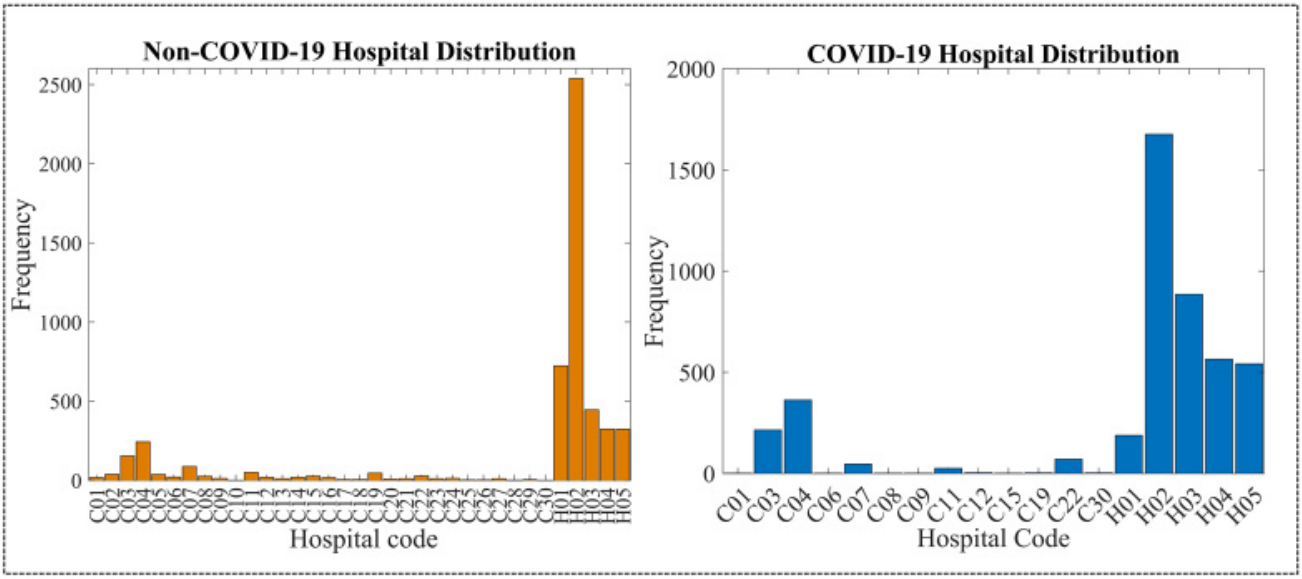
B


Vendor Distribution
Non-COVID-19
COVID-19
Frequency
3000
2500
2000
1500
1000
500
0
V1
V2
V3
V4
VO
Vendor

C

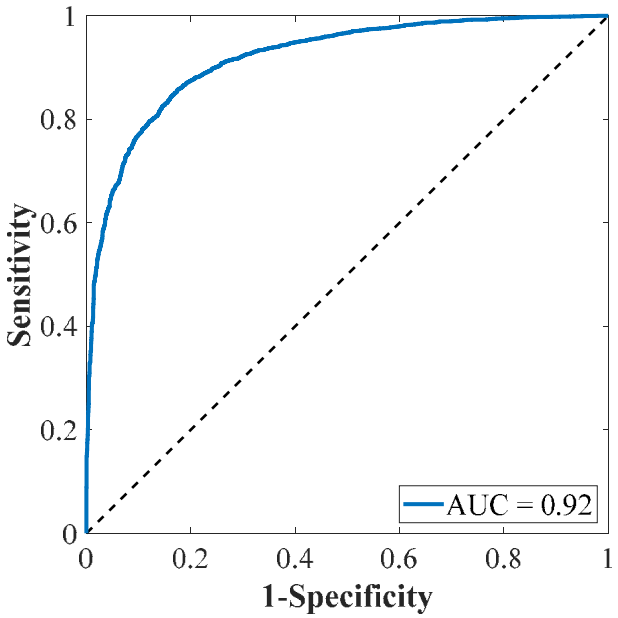

Modality Distribution
Non-COVID-19
COVID-19
Frequency
4000
3000
2000
1000
0
CR
DX
Modality

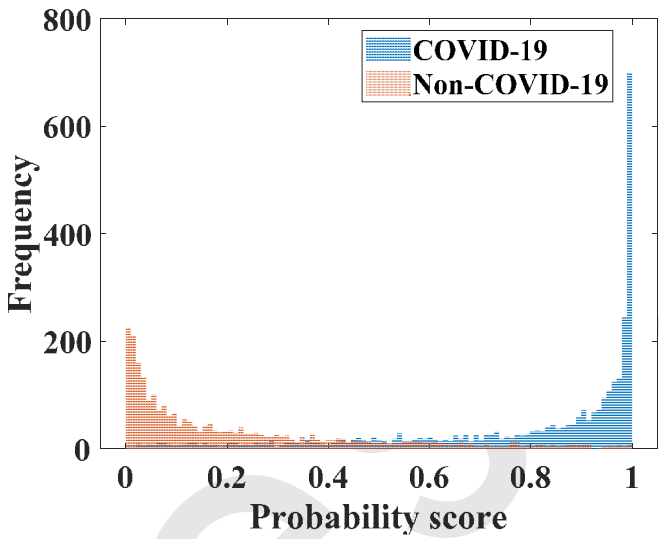
D



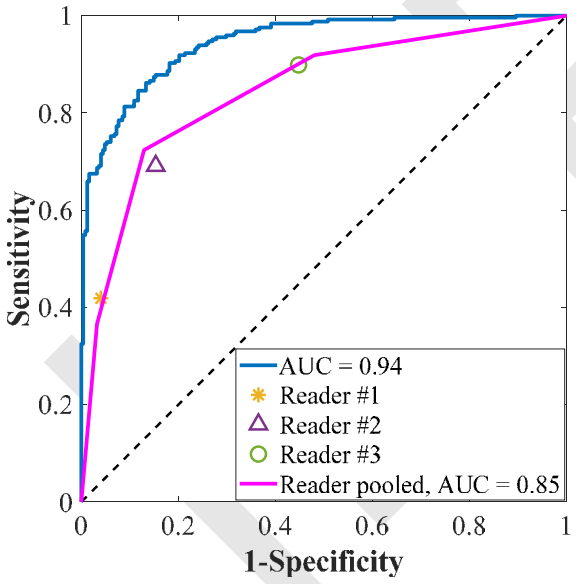
E

Figure 2: Detailed data characteristics. *A,* Age distribution of included patients. *B,* Distribution of the delta (time between the positive reverse transcriptase polymerase chain reaction [RT-PCR] test and the chest x-ray examination) for the positive cohort. A positive delta value indicates that the chest x-ray examination was performed after the RT-PCR test. *C,* Distribution of the x-ray radiograph vendors. *D,* Distribution of the use of computed radiography (CR) or digital radiography (DX). *E*, Distribution of data from different hospitals (H01-H05 indicates the five different hospitals and C01 to C30 indicate the 30 different clinics).



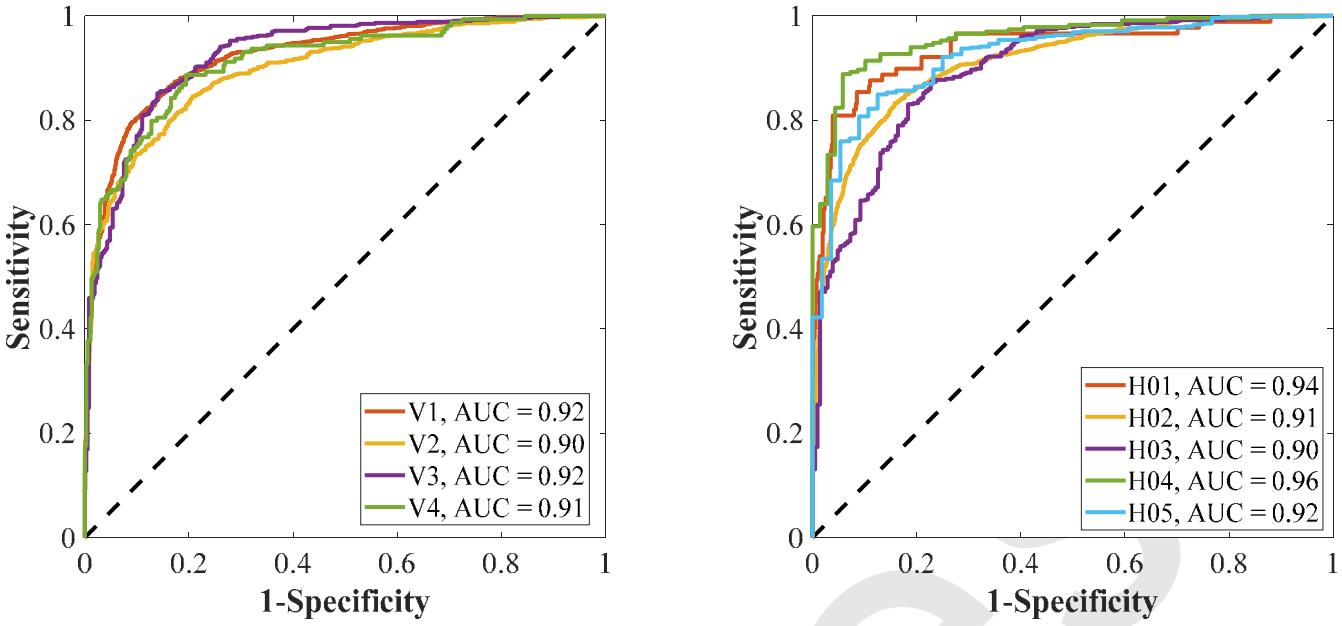


A



B

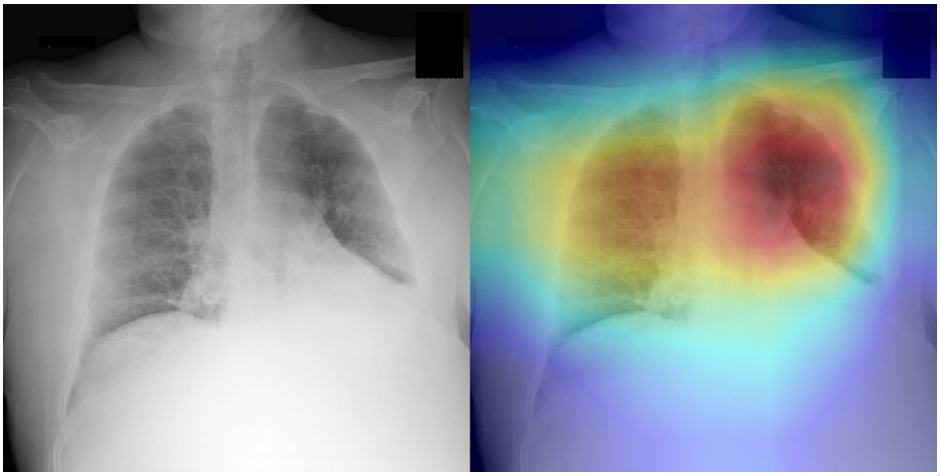
CV19-Net vs. Radiologists			
	Reader #1	Reader #2	Reader #3
Sensitivity/ Specificity	42%/96%	69%/85%	90%/55%
CV19-Net@ matched specificity	71%/96%	87%/85%	98%/55%
CV19-Net@ matched sensitivity	42%/99%	69%/96%	90%/82%



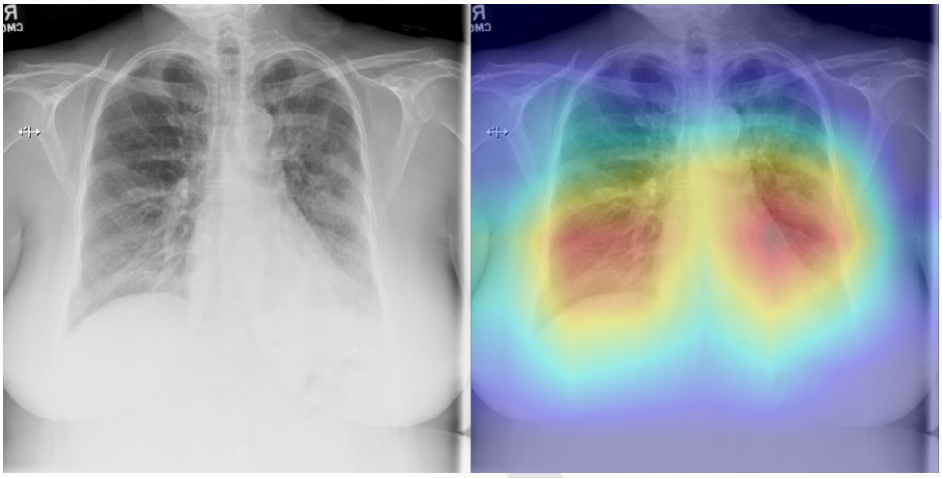
C

Figure 3: Performance of CV19-Net. *A,* Receiver operating characteristic (ROC) curve of the total test dataset (left) with 5869 CXRs and the probability score distribution (right), T1 and T2 denote high sensitivity operating point and high specificity operating point, respectively. *B,* Pooled performance of the three chest radiologists compared with CV19-Net for the 500 test cases. *C,* ROC curves of CV19-Net for different vendors (V1-V4) and hospitals (H01-H05) in the test dataset.

In Press



A



B

Figure 4: Examples of CXRs and the network generated heatmaps from the reader study test set. *A*, Left: a COVID-19 pneumonia case (64-year-old, male) that was classified correctly by CV19-Net but incorrectly by all three radiologists. Right: the heatmap generated by CV19-Net overlaid on the original image. The red coloring highlights the anatomical regions that contribute most to the CV19-Net prediction. *B*, Left: a non-COVID-19 pneumonia case (58-year-old, female) which was classified correctly by CV19-Net but incorrectly by all three radiologists. Right: the heatmap highlights the anatomical regions that contribute most to the CV19-Net prediction.

Tables

Table 1. Training, Validation, and Test Datasets

	Training and validation				Test			
	COVID-19		Non-COVID-19		COVID-19		Non-COVID-19	
	Patients	CXR	Patients	CXR	Patients	CXR	Patients	CXR
V1	623	1399	1122	1497	743	1757	417	1042
V2	269	458	332	457	424	715	289	556
V3	108	544	308	400	106	527	300	373
V4	53	181	200	300	80	159	280	375
VO	NA	NA	NA	NA	56	65	269	300
Total	1053	2582	1962	2654	1007	3223	1186	2646

Note.—Number of patients and CXRs in each dataset are shown. V1-V4 denotes Carestream Health, GE Healthcare, Konica Minolta and Agfa, respectively. VO denotes Siemens Healthineers and Kodak. CXR = chest x-ray radiography, COVID-19 = coronavirus disease 2019

Table 2. Test Performance of CV19-Net for Different Vendors

	Vendors				
Parameter	V1	V2	V3	V4	Overall
A. Performance					
No. images	2799	1271	900	534	5869
No. patients	1160	713	405	360	2193
AUC	0.92 (0.91-0.93)	0.90 (0.88-0.92)	0.93 (0.91-0.94)	0.91 (0.88-0.94)	0.92 (0.91-0.93)
B. High sensitivity operating point*					
Sensitivity	90% (88%-91%)	86% (83%-88%)	87% (84%-90%)	89% (84%-93%)	88% (87%-89%)
Specificity	78% (76%-81%)	77% (73%-80%)	82% (78%-85%)	78% (73%-82%)	79% (77%-80%)
C. High specificity operating point†					
Sensitivity	80% (78%-82%)	75% (72%-78%)	77% (73%-81%)	77% (70%-83%)	78% (77%-79%)
Specificity	90% (88%-92%)	88% (85%-91%)	90% (87%-93%)	88% (85%-92%)	89% (88%-90%)

Note.—Values in parenthesis are 95% confidence intervals. AUC = area under the receiver operating characteristic curve.

* Threshold of $T_1 = 0.4$.
† Threshold of $T_2 = 0.6$.

In Press

Table 3. Test Performance of CV19-Net for Different Age Groups

	Age group				
Parameter	18-30	30-45	45-60	60-75	≥75
A. Performance					
No. images	211	532	1519	2259	1348
No. patients	93	218	509	800	573
AUC	0.96 (0.94-0.98)	0.94 (0.93-0.96)	0.93 (0.91-0.94)	0.92 (0.91-0.93)	0.89 (0.88-0.91)
B. High sensitivity operating point*					
Sensitivity	90% (84%-96%)	91% (87%-94%)	92% (90%-94%)	88% (86%-90%)	82% (79%-85%)
Specificity	89% (84%-95%)	83% (79%-88%)	73% (70%-77%)	79% (76%-81%)	80% (77%-83%)
C. High specificity operating point†					
Sensitivity	78% (70%-87%)	85% (80%-89%)	84% (81%-86%)	78% (75%-80%)	69% (66%-72%)
Specificity	94% (89%-98%)	91% (88%-94%)	85% (83%-88%)	89% (88%-91%)	90% (88%-92%)

Note.—Values in parenthesis are 95% confidence intervals. AUC = area under the receiver operating characteristic curve.

* Threshold of $T_1 = 0.4$.
† Threshold of $T_2 = 0.6$.

In Press

Table 4. Test Performance of CV19-Net for Men and Women

	Sex	
Parameter	Men	Women
A. Performance		
No. images	3521	2348
No. patients	1131	1062
AUC	0.92 (0.91-0.93)	0.91 (0.90-0.92)
B. High sensitivity operating point*		
Sensitivity	88% (87%-89%)	89% (87%-90%)
Specificity	79% (77%-81%)	78% (75%-80%)
C. High specificity operating point†		
Sensitivity	78% (76%-80%)	79% (76%-81%)
Specificity	90% (88%-91%)	89% (87%-91%)

Note.—Values in parenthesis are 95% confidence intervals. AUC = area under the receiver operating characteristic curve.

* Threshold of $T_1 = 0.4$.
† Threshold of $T_2 = 0.6$.

In Press

Appendix E1: Bias Mitigation

In image classification tasks, machine learning methods use image intensity values to extract digital image features and then use these image features to compute, for example, the COVID-19 probability score of an input CXR. Therefore, the actual digital values in each CXR image determine the final machine learning classification decision of the input CXR. However, different vendors of CXR imaging systems use different proprietary postprocessing algorithms to process digital CXRs (ie each vendor will adjust their final digital values differently for desired image contrast for interpretation). Further, many hospitals and clinics often use multiple CXR imaging systems from different vendors. Additionally, different clinics and different technologists may choose different imaging parameters such as x-ray tube potential (kVp values) and x-ray exposure levels (mAs values) to acquire the CXR. As a result, similar pneumonia findings may have very different digital image representations in retrospectively collected digital CXRs. Without taking these variables into account, machine learning algorithms may produce biased results.

Appendix E2: Image Preprocessing

The DICOM files were resized to 1024×1024 and saved as 8-bit PNG grayscale images. The image intensity value was adjusted based on the window level and window width attributes in the DICOM file. Contrast inversion is applied for images with DICOM attribute MONOCHROME1. See Figure E1 for the flow chart of the image preprocessing step.

Before being fed into the network, PNG images were further downscaled to 224×224, converted to red (R)- green (G) -blue (B) images and normalized based on the mean and standard deviation of images in the ImageNet dataset:

In Press

$$R = (R - 0.485)/0.229$$

$$G = (G - 0.456)/0.224$$

$$B = (B - 0.406)/0.225$$

Appendix E3: Network Architecture and Training

The DenseNet-121[1] architecture with 50 convolutional neural network (CNN) layers was used as the image feature extraction module of CV19-Net. Followed by the last convolutional layer of DenseNet-121 (layer 120) is a fully connected (FC) classifier with a Softmax activation function to combine the extracted feature vector for the final predicted probability score.

A three-stage transfer learning process was used in model training (see Figure E2):

1. Stage one: The image feature extraction module was trained on ImageNet images with 14 million images to differentiate between 1,000 image classes.
2. Stage two: The network was initialized with weights trained in stage one and was further trained using the NIH data set with 112,120 chest x-ray images from 30,805 unique patients to classify chest x-ray images into 14 different disease classes. A similar design was used in CheXNet.[2] This step allows the network to learn CXR-specific image features.
3. Stage three: The network was initialized using weights obtained from stage two and trained using our training dataset consists of 5,236 CXRs (2,582 CXRs from the COVID-19 cohort and 2,654 CXRs from the non-COVID-19 pneumonia cohort) to train the network to perform the final binary classification: COVID-19 and non-COVID-19 pneumonia classification.

CV19-Net (Figure E3) was developed using the PyTorch framework. The network was trained to minimize the binary cross entropy loss. Adam optimizer was used with an initial learning rate $=6.0 \times 10^{-5}$ for all convolutional layers and 1.0×10^{-4} for the FC classifier. The minibatch size

In Press

was empirically selected to be 50. Data augmentation techniques including rotation (30-degree range) and horizontal flipping were used. To prevent model overfitting, an early stopping strategy was adopted by monitoring the training loss on the validation set. The validation set was randomly sampled from the total training dataset (25% of the training samples). The model with the lowest validation loss was taken as the final model for prediction. To reduce fluctuations of prediction results, the well-known ensemble averaging technique common in machine learning was introduced in this work. The prediction scores of input images from N=20 individually trained networks with identical training parameters, but different random seeds in model initialization and different randomly sampled validation sets. A quadratic mean of the prediction probability scores was taken to generate the final predication probability score:

$$S = \left[\frac{1}{N}\sum_{i=1}^{N} S^2(i)\right]^{1/2} \text{ (N=20).}$$

This final probability score was compared with the selected threshold values in decision making to perform binary classification.

Appendix E4: Class Activation Maps

To help visualize which part of the input images contributed most to CV19-Net's decision used to produce the final probability score, a heat map employing the gradient-weighted Class Activation Mapping (Grad-CAM)[3] was used to highlight those key image pixels in the CXR image primarily responsible for COVID-19 pneumonia. The paired heatmap of COVID-19 image features and the original CXR images are presented in Figure E4 to help aid human eyes to identify the key morphological and contextual features in CXR images.

In Press

Appendix E5: Additional Statistical Analyses

(a) Test performance difference on men and women

The following statistical hypothesis testing was performed:

$$H_0: \mathrm{AUC}(\text{male}) = \mathrm{AUC}(\text{female}) \text{ vs.}$$

$$H_1: \mathrm{AUC}(\text{male}) \neq \mathrm{AUC}(\text{female})$$

P-value method was used in hypothesis testing with a rejection p-value of .05. Result shows P=.17, therefore no statistically significant difference between two groups.

(b) Test performance difference on patients of different age groups

The following statistical hypothesis testing was performed:

$$H_0: \mathrm{AUC}(\text{Group-i}) = \mathrm{AUC}(\text{Group-j}) \text{ vs.}$$

$$H_1: \mathrm{AUC}(\text{Group-i}) \neq \mathrm{AUC}(\text{Group-j})$$

Results are shown in Table E1.

In Press

Supplementary Figures

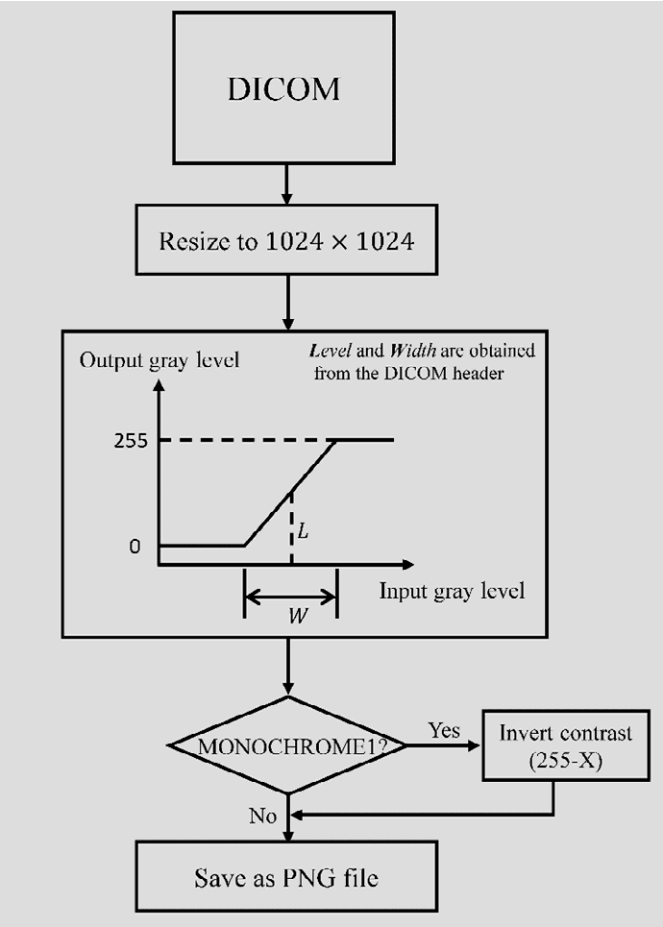


Figure E1: Flow chart showing the image preprocessing method.

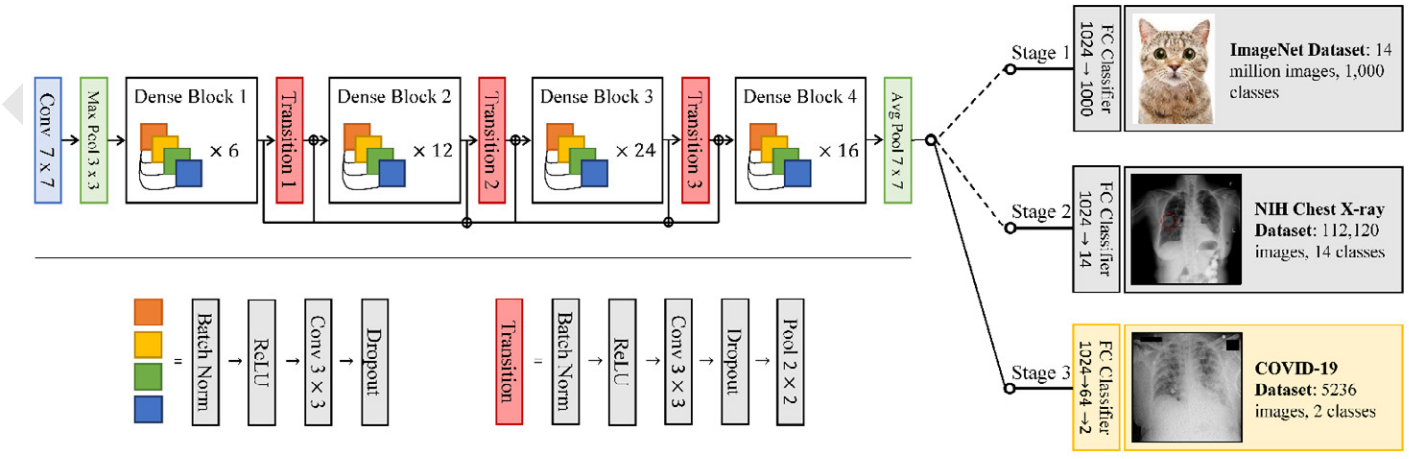


Figure E2: Individual network architecture and training process in CV19-Net.

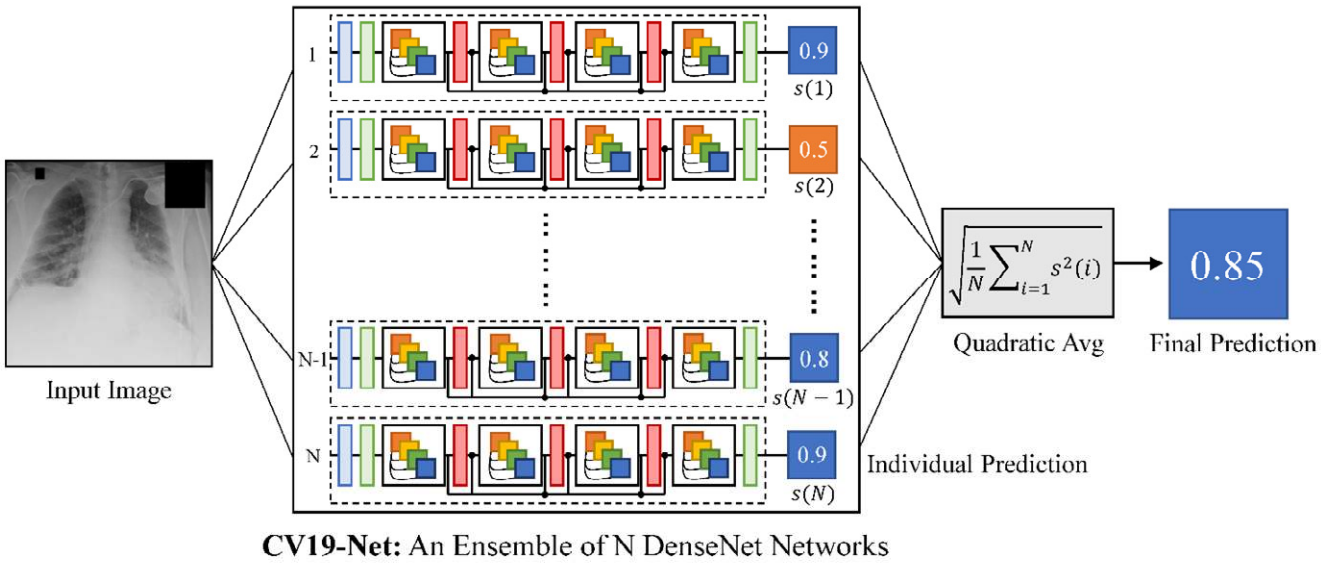


Figure E3: CV19-Net: An ensemble of individually trained deep neural network models to perform ensemble prediction of an input image.

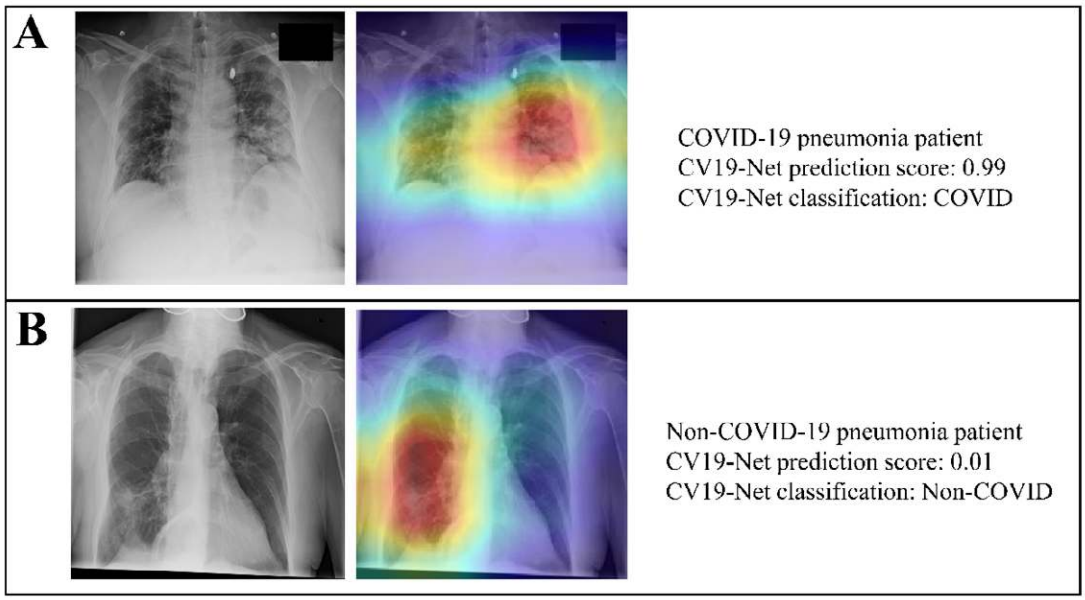


Figure E4: Class activation map examples.

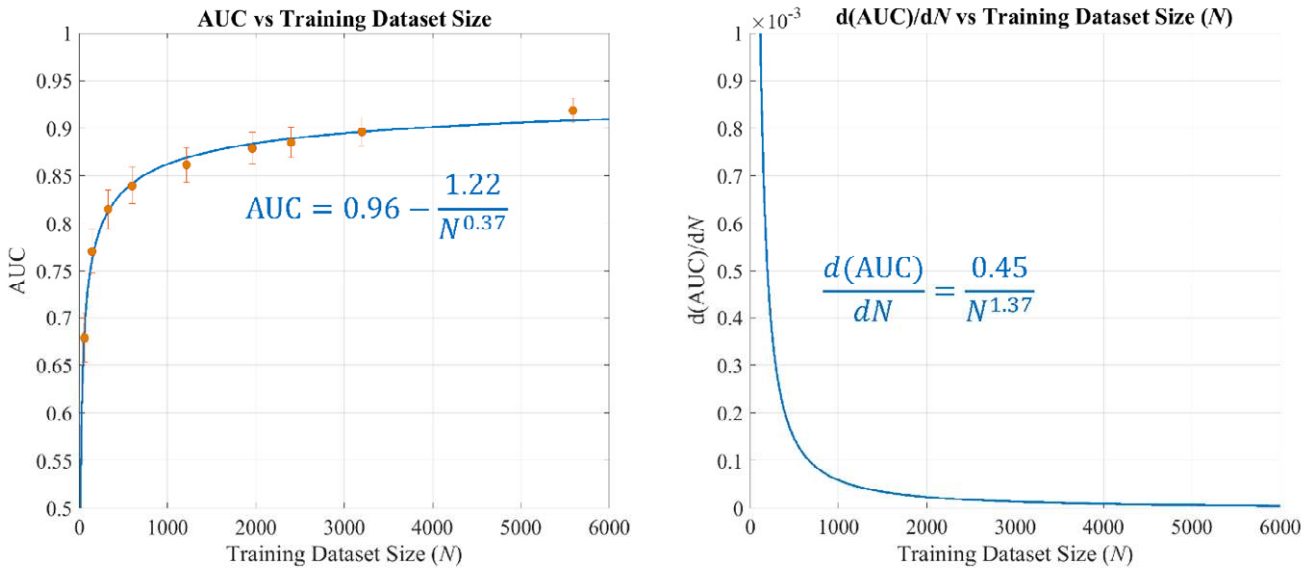


Figure E5: AUC vs. training sample size and the increment of AUC vs increment of training sample size. After the training sample size goes beyond 3000 (1500 positive and 1500 negative cases), the performance gain with the increase of training sample is diminished.

In Pre

Supplementary Tables

Table E1. Paired AUC Difference between Different Age Groups

Table E1. Age group AUC difference, P-value

	18-30	30-45	45-60	60-75	75-90
18-30		.31	.02	.002	<.001
30-45			.13	.01	<.001
45-60				.25	.002
60-75					.02
75-90					

In Press

Table E2. Related Works

Reference	Number of positive CXRs in training/validation	Number of positive CXRs in testing	Data type
Pereira et al.[4]	63	27	Cohen[5]
Rahimzadeh et al.[6]	149	31	Cohen
Zokaeinikoo et al.[7]	267		Cohen
Ozturk et al.[8]	127		Cohen
Kishore et al.[9]	150		Cohen
Narin et al.[10]	269		Cohen
Gil et al.[11]	288		Cohen
Horry et al.[12]	100		Cohen
Khan et al.[13]	284		Cohen
Elasnaoui et al.[14]	231		Cohen
Afshar et al.[15]	Not clear		Cohen
Karim et al.[16]	149	31	Cohen
Oh et al.[17]	144	36	Cohen
Wang et al.[18]	358		Wang
Luz et al.[19]	152	31	Wang
Ucar et al.[20]	66	10	Wang
Farooq et al.[21]	68		Wang
Shibly et al.[22]	232	51	Wang
Majeed et al.[23]	111	73	Cohen, Kaggle
Zhang et al.[24]	258	60	Cohen, Kaggle
Kumar et al.[25]	42	20	SIRM
Tahir et al.[26]	338	85	Cohen, SIRM, Radiopaedia,
Yeh et al.[27]	415	95	Local hospital
Schwab et al.[28]	391	167	Local hospital

In Press

Murphy et al.[29]	512	468	Local hospital

SIRM: https://www.sirm.org/category/senza-categoria/covid-19
Cohen: https://github.com/ieee8023/covid-chestxray-dataset
Kaggle: https://www.kaggle.com/andrewmvd/covid19-X-rays
Wang: https://github.com/lindawangg/COVID-Net
Radiopaedia: https://radiopaedia.org/playlists/25975?

In Press

References for Supplementary Materials

1. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017 2016;2017-Janua:2261–9.

2. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018;15(11):e1002686.

3. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision 2017 (pp. 618-626).

4. Pereira RM, Bertolini D, Teixeira LO, et al. COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios. Computer Methods and Programs in Biomedicine. 2020 May 8:105532.

5. Cohen JP, Morrison P, Dao L. COVID-19 image data collection. arXiv preprint arXiv:2003.11597. 2020 Mar 25.

6. Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Informatics in Medicine Unlocked. 2020 May 26:100360.

7. Zokaeinikoo M, Kazemian P, Mitra P, Kumara S. AIDCOV: An Interpretable Artificial Intelligence Model for Detection of COVID-19 from Chest Radiography Images. medRxiv. 2020.

8. Ozturk T, Talo M, Yildirim EA, et al. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in Biology and Medicine. 2020 Apr 28:103792.

In Press

9. Jamil M, Hussain I. Automatic Detection of COVID-19 Infection from Chest X-ray using Deep Learning. medRxiv. 2020.

10. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849. 2020 Mar 24.

11. Gil D, Díaz-Chito K, Sánchez C, Hernández-Sabaté A. Early Screening of SARS-CoV-2 by Intelligent Analysis of X-Ray Images. arXiv preprint arXiv:2005.13928. 2020 May 28.

12. Horry MJ, Paul M, Ulhaq A, Pradhan B, Saha M, Shukla N. X-Ray Image based COVID-19 Detection using Pre-trained Deep Learning Models. https://engrxiv.org/wx89s/

13. Khan AI, Shah JL, Bhat MM. Coronet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Computer Methods and Programs in Biomedicine. 2020 Jun 5:105581.

14. Elasnaoui K, Chawki Y. Using x-ray images and deep learning for automated detection of coronavirus disease. Journal of Biomolecular Structure and Dynamics. 2020 May 9(just-accepted):1-22.

15. Afshar P, Heidarian S, Naderkhani F, et al. Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. arXiv preprint arXiv:2004.02696. 2020 Apr 6.

16. Karim M, Döhmen T, Rebholz-Schuhmann D, et al. Deepcovidexplainer: Explainable covid-19 predictions based on chest x-ray images. arXiv preprint arXiv:2004.04582. 2020 Apr 9.

17. Oh Y, Park S, Ye JC. Deep learning covid-19 features on cxr using limited training data sets. IEEE Transactions on Medical Imaging. 2020 May 8.

In Press

18. Wang L, Wong A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. arXiv preprint arXiv:2003.09871. 2020 Mar 22.

19. Luz E, Silva PL, Silva R, et al. Towards an efficient deep learning model for covid-19 patterns detection in x-ray images. arXiv preprint arXiv:2004.05717. 2020 Apr 12.

20. Ucar F, Korkmaz D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based Diagnostic of the Coronavirus Disease 2019 (COVID-19) from X-Ray Images. Medical Hypotheses. 2020 Apr 23:109761.

21. Farooq M, Hafeez A. Covid-resnet: A deep learning framework for screening of covid19 from radiographs. arXiv preprint arXiv:2003.14395. 2020 Mar 31.

22. Shibly KH, Dey SK, Islam MT, et al. COVID Faster R-CNN: A Novel Framework to Diagnose Novel Coronavirus Disease (COVID-19) in X-Ray Images. medRxiv. 2020

23. Majeed T, Rashid R, Ali D, et al. Covid-19 detection using CNN transfer learning from X-ray Images. medRxiv. 202.

24. Zhang Y, Niu S, Qiu Z, et al. COVID-DA: Deep Domain Adaptation from Typical Pneumonia to COVID-19. arXiv preprint arXiv:2005.01577. 2020 Apr 30.

25. Kumar R, Arora R, Bansal V, et al. Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers. medRxiv. 2020.

26. Tahir A, Qiblawey Y, Khandakar A, et al. Coronavirus: Comparing COVID-19, SARS and MERS in the eyes of AI. arXiv preprint arXiv:2005.11524. 2020 May 23.

27. Yeh CF, Cheng HT, Wei A, et al. A Cascaded Learning Strategy for Robust COVID-19 Pneumonia Chest X-Ray Screening. arXiv preprint arXiv:2004.12786. 2020 Apr 24.

In Press

28. Schwab P, Schütte AD, Dietz B, et al. predCOVID-19: A Systematic Study of Clinical Predictive Models for Coronavirus Disease 2019. arXiv preprint arXiv:2005.08302. 2020 May 17.

29. Murphy K, Smits H, Knoops AJG, et al. COVID-19 on the Chest Radiograph: A Multi-Reader Evaluation of an AI System. Radiology 2020;201874.

In Press