COVID-Net CXR-S: Deep Convolutional Neural Network for Severity Assessment of COVID-19 Cases from Chest X-ray Images

Hossein Aboutalebi (haboutal@uwaterloo.ca)
University of Waterloo
Maya Pavlova
University of Waterloo
Mohammad Javad Shafiee
University of Waterloo
Ali Sabri
McMaster University
Amer Alaref
Thunder Bay Regional Health Sciences Centre
Alexander Wong
University of Waterloo

Article

Keywords: CXR, SARS-CoV-2, COVID-Net CXR-S, radiology, RSNA RICORD

Posted Date: June 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-580218/v1

Version of Record: A version of this preprint was published at Diagnostics on December 23rd, 2021. See the published version at https://doi.org/10.3390/diagnostics12010025.

COVID-Net CXR-S: Deep Convolutional Neural Network for Severity Assessment of COVID-19 Cases from Chest X-ray Images

Hossein Aboutalebi[1,3], Maya Pavlova[2], Mohammad Javad Shafiee[2,3,4], Ali Sabri[5], Amer Alaref[6,7], and Alexander Wong[2,3,4]

[1]Department of Computer Science, University of Waterloo, Canada
[2]Department of Systems Design Engineering, University of Waterloo, Canada
[3]Waterloo Artificial Intelligence Institute, University of Waterloo, Canada
[4]DarwinAI Corp., Canada
[5]Department of Radiology, Niagara Health, McMaster University, Canada
[6]Department of Diagnostic Radiology, Thunder Bay Regional Health Sciences Centre, Canada
[7]Department of Diagnostic Imaging, Northern Ontario School of Medicine, Canada

ABSTRACT

The world is still struggling in controlling and containing the spread of the COVID-19 pandemic caused by the SARS-CoV-2 virus. The medical conditions associated with SARS-CoV-2 infections have resulted in a surge in the number of patients at clinics and hospitals, leading to a significantly increased strain on healthcare resources. As such, an important part of managing and handling patients with SARS-CoV-2 infections within the clinical workflow is severity assessment, which is often conducted with the use of chest x-ray (CXR) images. In this work, we introduce COVID-Net CXR-S, a convolutional neural network for predicting the airspace severity of a SARS-CoV-2 positive patient based on a CXR image of the patient's chest. More specifically, we leveraged transfer learning to transfer representational knowledge gained from over 16,000 CXR images from a multinational cohort of over 15,000 patient cases into a custom network architecture for severity assessment. Experimental results with a multi-national patient cohort curated by the Radiological Society of North America (RSNA) RICORD initiative showed that the proposed COVID-Net CXR-S has potential to be a powerful tool for computer-aided severity assessment of CXR images of COVID-19 positive patients. Furthermore, radiologist validation on select cases by two board-certified radiologists with over 10 and 19 years of experience, respectively, showed consistency between radiologist interpretation and critical factors leveraged by COVID-Net CXR-S for severity assessment. While not a production-ready solution, the ultimate goal for the open source release of COVID-Net CXR-S is to act as a catalyst for clinical scientists, machine learning researchers, as well as citizen scientists to develop innovative new clinical decision support solutions for helping clinicians around the world manage the continuing pandemic.

1 Introduction

The impact of the coronavirus disease 2019 (COVID-19) pandemic on the health and economy has been unprecedented. While more than one year has been passed since the declaration of the global pandemic by the World Health Organization[1], countries are still struggling with controlling the spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus causing the pandemic. In this regard, the global healthcare system has suffered a devastating impact from this pandemic, with hospitals and clinics overwhelmed by the surge of patients and thus not all patients can have access to intensive care units for further treatment and care[2]. Furthermore, there have been shortages in personal protective equipment (PPE), ventilators, and other medical supplies due to the increasing demand on healthcare resources[2,3].

This significant strain on healthcare resources caused by the COVID-19 pandemic from both a personnel and supplies perspective that necessitates improved clinical decision support tools for aiding clinicians and front-line healthcare workers with more efficient and effective clinical resource allocation. A critical part of clinical resource allocation during this pandemic has been severity assessment[4–6], which is often conducted with the assistance of chest x-ray (CXR) images[4–12]. More specifically, radiology indicators within the patient's lungs such as ground-glass opacities can provide critical information for determining whether the condition of a SARS-CoV-2 positive patients warrants advanced care such as ICU admission and ventilator administration. For example, Wong *et al.*[4] proposed a scoring strategy for SARS-CoV-2 severity assessment based on the Radiographic Assessment of Lung Edema (RALE) score introduced by Warren *et al.*[5]. Toussie *et al.*[6] proposed a scoring strategy where each lung was divided into three zones (for a total of six zones), with each zone assigned a binary score based on

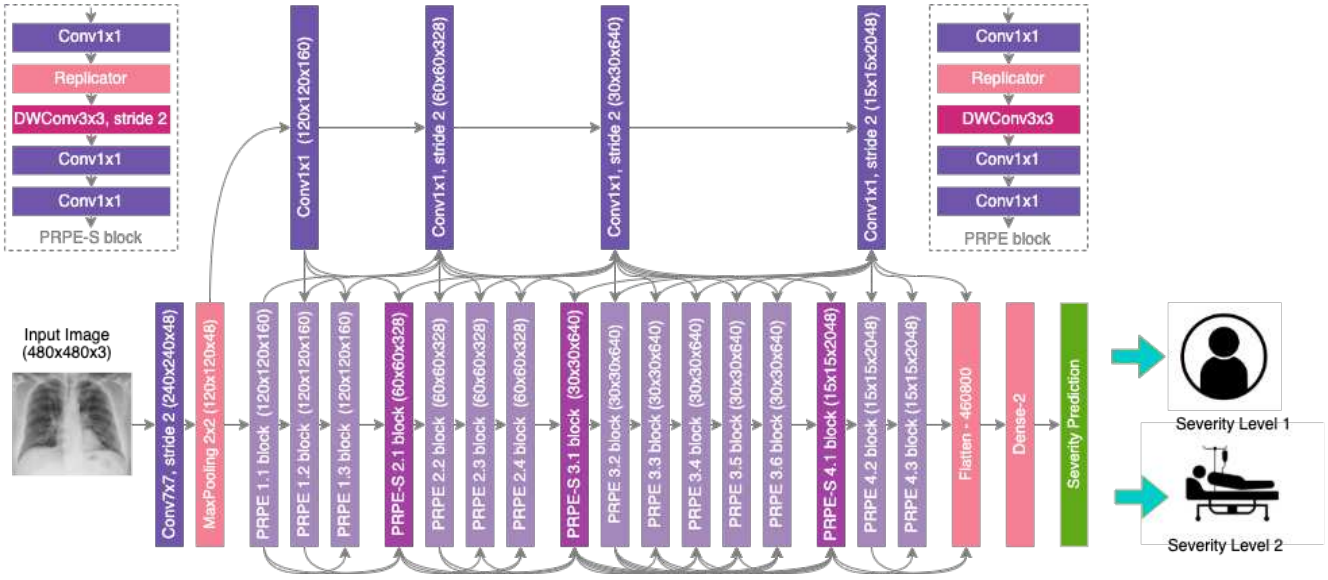

Figure 1. COVID-Net CXR-S network design. The COVID-Net backbone design exhibits high architectural diversity and sparse long-range connectivity, with macroarchitecture and microarchitecture designs tailored specifically for the detection of COVID-19 from chest X-ray images. The network design leverages light-weight design patterns in the form of projection-expansion-projection-expansion (PEPE) patterns to provide enhanced representational capabilities while maintaining low architectural and computational complexities.

opacity and the final severity score computed as the aggregate of the scores from the different zones. Borghesi and Maroldi[13] proposed a scoring strategy where each lung was divided into three zones, as with that of Toussie *et al.*[6], but each zone was assigned a score from 0 to 3 based on interstitial and alveolar infiltrates. Tsai *et al.*[14], as part of the Radiological Society of North America (RSNA) RICORD initiative, proposed an airspace disease grading strategy where each lung is split into three zones (for a total of six zones), as with that of Toussie *et al.*[6] and Borghesi *et al.*[13], but instead quantified the level of severity based on the number of lung zones with opacities.

Given that severity assessment using CXR images can be quite challenging for front-line healthcare workers without expertise in radiology, providing computer-aided clinical support for this task can greatly benefit the hospitals in determining patients' conditions and respond more quickly to those who may require more advanced treatments or intensive care. While much of research literature have focused on computer-aided COVID-19 detection from CXR images[15,16] and computed tomography (CT) scans[17–21], the area of computer-aided severity assessment is significantly less well explored. Some of the notable works in this area are COVID-Net S, a tailored deep convolutional neural network proposed by Wong *et al.*[22] to predict the extent scores from CXR images, as well as the study by Cohen *et al.*[23] where the same extent scores are predicted.

Motivated to extend upon this area in the direction of airspace disease grading, we introduce COVID-Net CXR-S, a convolutional neural network for predicting the airspace severity of a SARS-CoV-2 positive patient based on a CXR image of the patient's chest, as part of the COVID-Net open source initiative[15,18,19,22,24,25]. The paper is organized as follows. Section 2 describes the methodology behind the design and construction of the proposed COVID-Net CXR-S, a demographic and protocol analysis of the multi-national patient cohort used, as well as radiologist validation. Section 3 presents and discusses the quantitative and qualitative results obtained from the experiments evaluating the efficacy of the proposed COVID-Net CXR-S.

2 Methods

In this work, we introduce COVID-Net CXR-S, a convolutional neural network tailored for the prediction of airspace severity of a SARS-CoV-2 positive patient based on chest X-ray images. To train COVID-Net CXR-S, we transferred representational knowledge from CXR images of a large multi-national patient cohort, and then leveraged CXR data grouped based on airspace severity levels. We further validated the behaviour of COVID-Net CXR-S in a transparent and responsible manner via explainability-driven performance validation, as well as conducted radiologist validation on select cases by two expert board-certified radiologists. The details between network design, data preparation, explainability-driven performance validation, and radiologist validation are described below.

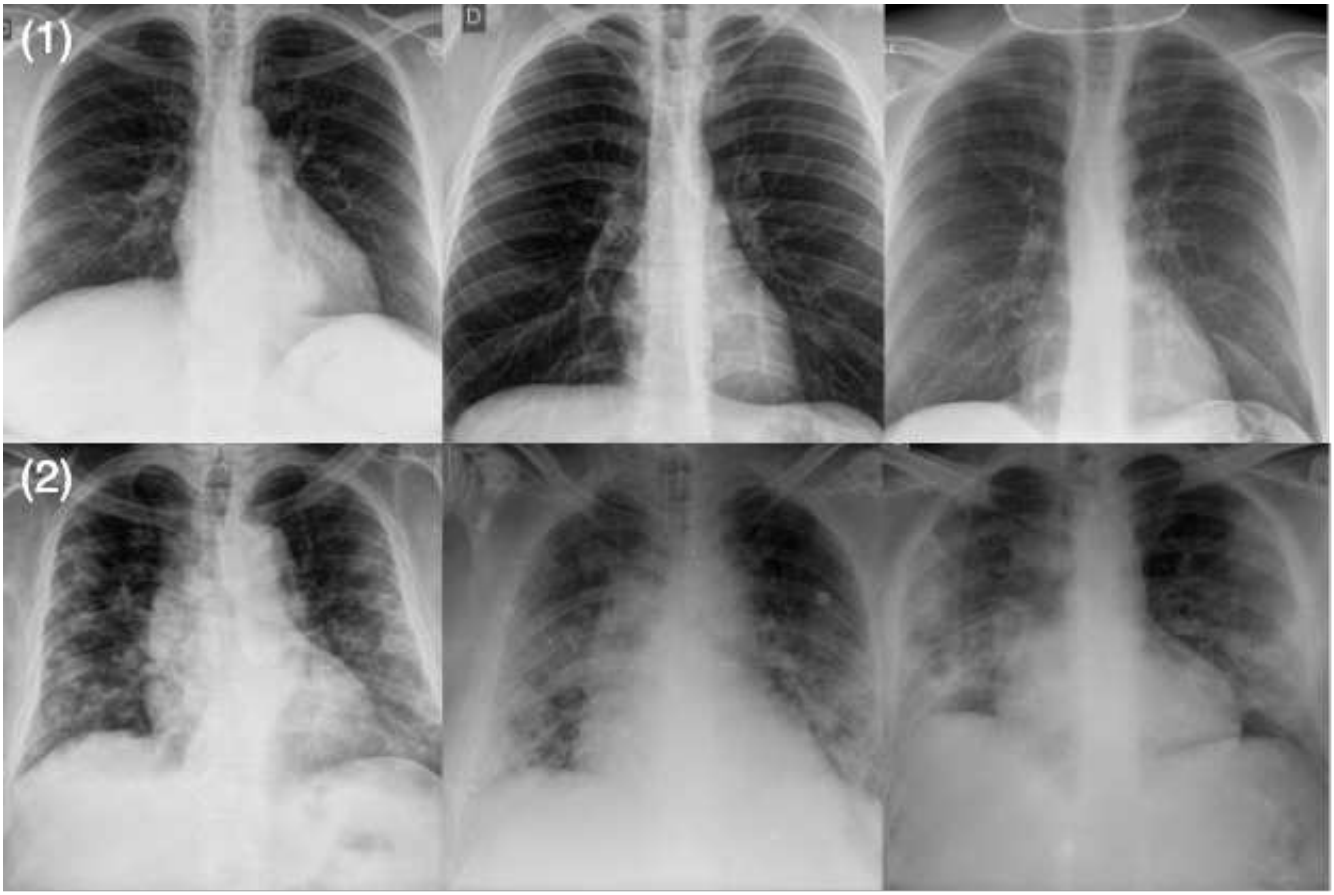

Figure 2. Example chest X-ray images from the multi-national patient cohort curated by the RSNA RICORD initiative: (1) Level 1 airspace severity: opacities in 1-2 lung zones and (2) Level 2 airspace severity: opacities in 3 or more lung zones.

2.1 Network design

The proposed COVID-Net CXR-S architecture is depicted in Figure 1. More specifically, we leveraged a machine-driven design exploration strategy[26] to construct a backbone architecture tailored for a strong balance between accuracy and efficiency[27]. The constructed backbone architecture exhibits light-weight macro-architecture and micro-architecture designs comprised primarily of depthwise and pointwise convolutions with selective long-range connectivity. We then leveraged transfer learning to transfer representational knowledge gained from over 16,000 CXR images from a large multi-national cohort of over 15,000 patient cases[14,28–32] via the constructed backbone architecture into a custom network architecture for severity assessment, where a combination of a dense layer and a severity prediction layer is used to predict between two levels of airspace severity. All of the model development was conducted using Python and the Keras deep learning library with a TensorFlow backend.

The COVID-Net CXR-S network and associated scripts are available in an open source manner at `http://www.covid-net.ml`.

2.2 Data preparation

In this study, we leveraged the multi-national patient cohort curated by the RSNA RICORD initiative[33] for the severity scoring to train COVID-Net CXR-S after transfer learning. The multi-national patient cohort was curated by the Radiological Society of North America (RSNA) as part of a global initiative to assemble an international task force of scientists and radiologists to create a multi-institutional, multinational, expert-annotated COVID-19 imaging data collection. More specifically, we leveraged the airspace disease grading provided by RSNA RICORD, where each lung is split into 3 separate zones (for a total of 6 zones) and opacity is studied for each zone. We group the patient cases into two airspace severity level groups: 1) Level 1: opacities in 1-2 lung zones, and 2) Level 2: opacities in 3 or more lung zones. Example CXR images for the different airspace severity level groups from the multi-national patient cohort are shown in Figure 2. This severity level designation was chosen given clinical similarities between patient cases within each airspace severity level groups in terms of the treatment regimen, and thus facilitates clearer guidelines for course of action.

Table 1. Summary of demographic variables and imaging protocol variables of CXR data from the multi-national patient cohort used in this study. Age and sex statistics are expressed on a patient level, while imaging view statistics are expressed on an image level.

Age	mean ± std	59.11 ± 16.19
	<20	2 (0.8%)
	20-29	7 (2.7%)
	30-39	26 (10.1%)
	40-49	37 (14.3%)
	50-59	58 (22.5%)
	60-69	58 (22.5%)
	70-79	42 (16.3%)
	80-89	22 (8.5%)
	90+	6 (2.3%)
Sex		
	Male	161 (62.4%)
	Female	97 (37.6%)
Imaging view		
	AP	505 (55.6%)
	PA	5 (0.6%)
	Unknown	399 (43.9%)

Given this airspace severity level grouping scheme, the multi-national patient cohort used in this study consists of 909 CXR images from 258 patients, with 227 images from 129 patients in Level 1 and 682 images from 184 patients in Level 2. We used 150 randomly selected CXR images for the test set, ensuring no patient overlap between the test and train data. Table 1 summarizes the demographic variables and imaging protocol variables of the CXR data in the multi-national patient cohort used in this study. It can be observed that the patient cases in the cohort used in the study are distributed across the different age groups, with the mean age being 59.11 and the highest number of patients in the cohort are between the ages of 50-69.

The data preparation scripts are available in an open source manner at `http://www.covid-net.ml`.

2.3 Network training

Training on the COVID-Net CXR-S architecture after knowledge transfer is conducted on the training portion of the aforementioned patient cohort using the Adam optimizer with a learning rate of 0.0001 for 137 epochs with a batch size of 50. During the training, we account for the imbalance in the number of patient cases between the airspace severity levels by performing batch balancing, where at each epoch we randomly sampled an equal number of CXR images from each severity level for each batch of data. As a pre-processing step, the CXR images were cropped (top 8% of the image) prior to the training process to better mitigate the influence of commonly-found embedded textual information, and resampled to 480×480 for training purposes. In addition, we leveraged data augmentation during the training process with the following augmentation types: translation (± 10% in x and y directions), rotation (± 10°), horizontal flip, zoom (± 15%), and intensity shift (± 10%).

The scripts for the aforementioned process are available in an open source manner at `http://www.covid-net.ml`.

2.4 Explainability-driven performance validation

To study the decision-making behaviour of the proposed COVID-Net CXR-S network in a transparent and responsible manner, we conducted explainability-driven performance validation using GSInquire[34], which was shown to provide state-of-the-art explanations. More specifically, GSInquire leverages the concept of generative synthesis[26] from the machine-driven design exploration process via an inquisitor $\mathscr{I}$ within a generator-inquisitor pair $\{\mathscr{G}, \mathscr{I}\}$ to generate quantitative interpretations of the decision-making process of COVID-Net CXR-S on a given CXR image. In this case, the generator $\mathscr{G}$ in the generator-inquisitor pair is the optimal generator that was leveraged to construct the backbone architecture during the network design process. The details pertaining to GSInquire for explaining the decision-making behaviour of deep neural networks on CXR images can be found in Wang *et al.*[15]. An interesting property of GSInquire that also makes it well-suited for explainability-driven performance validation is that it is capable of producing explanations identifying specific critical factors within an image that quantitatively impacts the decisions made by a deep neural network, thus making it more readily interpretable and more quantitative for validation purposes than the types of relative importance variations visualized by other methods.

This explainability-driven performance validation process enables the identification of anomalies in decision-making behaviour or potential erroneous indicators leading to invalidate decisions or biases, as well as the validation of whether

Table 2. Architectural and computational complexity of the proposed COVID-Net CXR-S, CheXNet[35], and ResNet-50[36]. Best numbers are highlighted in **bold**.

Network	Parameters (M)	FLOPs (G)
CheXNet[35]	**8.1**	26.0
ResNet-50[36]	23.6	35.5
COVID-Net CXR-S	8.8	**11.1**

Table 3. Sensitivity, positive predictive value (PPV), and accuracy of the proposed COVID-Net CXR-S, CheXNet[35], and ResNet-50[36]. Best numbers are highlighted in **bold**.

[width=7em]NetworkMetric	Sensitivity (Level 1)	Sensitivity (Level 2)	PPV (Level 1)	PPV (Level 2)	Accuracy
CheXNet[35]	**93.88%**	63.46%	82.88%	84.62%	83.33%
ResNet-50[36]	91.84%	78.85%	**89.11%**	83.67%	87.33%
COVID-Net CXR-S	92.3%	**92.85%**	87.27%	**95.78%**	**92.66%**

Table 4. Confusion Matrix of COVID-Net CXR-S.

Severity Level	Level 1	Level 2
Level 1	48	4
Level 2	7	91

clinically relevant indicators are leveraged.

2.5 Radiologist validation

The results obtained for COVID-Net CXR-S during the explainability-driven performance validation process for selected patient cases are further reviewed and reported on by two board-certified radiologists (A.S. and A.A.). The first radiologist (A.S.) has over 10 years of experience, and the second radiologist (A.A.) has over 19 years of radiology experience.

3 Results and Discussion

To evaluate the efficacy of the proposed COVID-Net CXR-S for the purpose of severity assessment of COVID-19 cases from CXR images, we conducted both quantitative performance evaluation as well as qualitative explainability-driven performance validation. The quantitative and qualitative results are presented and discussed below.

3.1 Quantitative results

The overall quantitative performance assessment results of the proposed COVID-Net CXR-S network on the multi-national patient cohort from the RSNA RICORD initiative can be seen in Table 3. For comparison purposes, quantitative performance assessment was also conducted on the ResNet-50[36] network architecture as well as on CheXNet[35], a state-of-the-art deep neural network architecture that has been shown to outperform other network architectures for CXR image analysis tasks. The architectural and computational complexity of COVID-Net CXR-S in comparison to CheXNet[35] and ResNet-50[36] is also shown in Table 2.

A number of observations can be made from the quantitative results. More specifically, it can be observed that the COVID-Net CXR-S network can achieve a high level of accuracy at 92.66%, which is 9.33% and 5.33% higher than that achieved by CheXNet and ResNet-50 network architectures, respectively. The COVID-Net CXR-S network achieved higher accuracy at a significantly lower computational complexity of ∼57% and ∼69% (at ∼11.1G FLOPs) in comparison to the CheXNet and ResNet-50 networks, respectively, as well as a ∼63% lower architectural complexity (at ∼8.8M parameters)

Figure 3. Examples of Level 2 severity patient cases and the associated critical factors (highlighted in red) as identified by GSInquire[34] during explainability-driven performance validation as what drove the decision-making behaviour of COVID-Net CXR-S. (left) Case 1, (middle) Case 2, (right) Case 3. Radiologist validation showed that several of the critical factors identified are consistent with radiologist interpretation.

than the ResNet-50 network, and achieving similar architectural complexity when compared to CheXNet. Achieving high architectural efficiency and computational efficiency is important for operation in resource-limited clinical scenarios where low-cost computing devices are desired.

It can also be observed that the COVID-Net CXR-S network achieved a sensitivity of 92.3% on the Level 2 cases and sensitivity of 92.85% on the Level 1 cases. Furthermore, it can also be seen that the proposed COVID-Net CXR-S network can achieve high positive predictive value (PPV) of 95.78% and 87.27% for Level 2 cases and Level 1 cases, respectively. This high PPV for Level 2 cases ensures that fewer false positives for more severe cases are reported by COVID-Net CXR-S, which is important since patients with severe conditions require advanced treatment and management and thus high false positive rate can lead to significant burden on the healthcare system where resources are limited. Finally, Table 4 provides a more detailed picture of the performance of COVID-Net CXR-S via the confusion matrix.

3.2 Qualitative results

Results from the conducted explainability-driven performance validation[34] (see Figure 3) show that clinically relevant visual indicators in the lungs were leveraged in the decision-making process. This validation is very important for auditing COVID-Net CXR-S in a transparent and responsible manner to ensure that not only is it leveraging the right indicators for driving the severity assessment process, but also that it is not primarily leveraging erroneous visual indicators (e.g., embedded markers, motion artifacts, imaging artifacts, etc.) to make 'right decisions for the wrong reasons'. Furthermore, this validation have the potential to help in the discovery of additional visual indicators to assist a clinician in their severity assessment as well as improve the trust that clinicians may have during operational use.

3.3 Radiologist analysis

The expert radiologist findings and observations for select patient cases with respect to the identified critical factors during explainability-driven performance validation as shown in Figure 3 are as follows. In all three cases, COVID-Net CXR-S correctly detected them to be patients with Level 2 airspace severity, which were clinically confirmed.

Case 1. According to radiologist findings, it was observed by both radiologists that there is patchy airspace opacity in the lower left lung lobe, which is consistent with one of the identified critical factors leveraged by COVID-Net CXR-S.

Case 2. According to radiologist findings, it was observed by both radiologists that there are patchy airspace opacities in the left and right midlung regions that coincide with the identified critical factors leveraged by COVID-Net CXR-S in that region. It was further observed by one of the radiologists that there are additional lower lobe opacities in both lungs.

Case 3. According to radiologist findings, it was observed by one of the radiologists that there are hilar opacities on the right lung that coincide with the identified critical factors leveraged by COVID-Net CXR-S. It was also observed that there are opacities in the left lower lobe, with the superior aspect of the opacities being leveraged by COVID-Net CXR-S. The second radiologist observed patchy airspace opacities in the right lung that overlap with the critical factors leveraged by COVID-Net CXR-S.

As such, based on the radiologist findings and observations on the three patient cases, it was shown that although some opacities were not identified by GSInquire as critical factors driving the decision-making behaviour of COVID-Net CXR-2,

several other abnormalities identified as critical factors were consistent with radiologist interpretations. Therefore, based on the critical factors identified by GSInquire as critical factors driving the decision-making behaviour of COVID-Net CXR-S, the network was able to differentiate between the airspace severity levels but not necessarily leverage all regions of concern in making its severity assessment decisions.

4 Conclusion

In this study, we introduced COVID-Net CXR-S, a convolutional neural network for the prediction of airspace severity of a SARS-CoV-2 positive patient based on a CXR image of the patient's chest. Leveraging transfer learning, we transferred representational knowledge gained from over 16,000 CXR images from a multinational cohort of over 15,000 patient cases into a custom network architecture for severity assessment. The promising quantitative and qualitative results obtained from the conducted experiments demonstrate that the proposed COVID-Net CXR-S, while not a production-ready solution, can be a potentially become a powerful tool for aiding clinicians and front-line healthcare workers via computer-aided severity assessment of CXR images of COVID-19 positive patients. The ultimate goal for the open source release of COVID-Net CXR-S (http://www.covid-net.ml) is to act as a catalyst for clinical scientists, machine learning researchers, as well as citizen scientists to develop innovative new clinical decision support solutions for helping clinicians around the world manage the continuing pandemic.

Acknowledgment

We thank the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chairs program, the Canadian Institute for Advanced Research (CIFAR), DarwinAI Corp., and the organizations and initiatives from around the world collecting valuable COVID-19 data to advance science and knowledge.

Author contributions statement

H.A., M.P., M.S., and A.W. conceived the experiments, H.A. and M.P. conducted the experiments, all authors analysed the results, A.A. and A.S. reviewed and reported on select patient cases and corresponding explainability results illustrating model's decision-making behaviour, and all authors reviewed the manuscript.

Declaration of interests

M.S. and A.W. are affiliated with DarwinAI Corp.

Ethics approval

The study has received ethics clearance from the University of Waterloo (42235). All experimental protocols were approved by University of Waterloo. All methods were carried out in accordance with University of Waterloo ethics guidelines and regulations. Informed consent was obtained from all participants.

References

1. Pak, A. *et al.* Economic consequences of the covid-19 outbreak: the need for epidemic preparedness. *Front. public health* **8** (2020).
2. Azoulay, E. *et al.* Admission decisions to intensive care units in the context of the major covid-19 outbreak: local guidance from the covid-19 paris-region area. *Critical Care* **24**, 1–6 (2020).
3. Tyrrell, C. S. *et al.* Managing intensive care admissions when there are not enough beds during the covid-19 pandemic: a systematic review. *Thorax* **76**, 302–312 (2021).
4. Wong, H. *et al.* Frequency and distribution of chest radiographic findings in COVID-19 positive patients. *Radiology* (2020).
5. Warren, M. A. *et al.* Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. *Thorax* **73**, 840–846 (2018).
6. Toussie, D. *et al.* Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. *Radiology* (2020).
7. Rubin, G. D. *et al.* The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the fleischner society. *Radiology* (2020).

8. Jacobi, A., Chung, M., Bernheim, A. & Eber, C. Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. *Clin. Imaging* (2020).

9. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**, 497–506 (2020).

10. Guan, W.-j. *et al.* Clinical characteristics of coronavirus disease 2019 in china. *New Engl. J. Medicine* **382**, 1708–1720, DOI: 10.1056/NEJMoa2002032 (2020). https://doi.org/10.1056/NEJMoa2002032.

11. Mao, B., Liu, Y., Chai, Y. *et al.* Assessing risk factors for SARS-CoV-2 infection in patients presenting with symptoms in Shanghai, China: a multicentre observational cohort study. *Lancet Digit. Heal.* (2020).

12. Wu, G. & Li, X. Mobile X-rays are highly valuable for critically ill COVID patients. *Eur. Radiol.* (2020).

13. Borghesi, A. & Maroldi, R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. *La radiologia medica* **125**, 509–51 (2020).

14. Tsai, E., Simpson, S. *et al.* The RSNA international covid-19 open annotated radiology database (RICORD). *Radiology* 203957 (2021).

15. Wang, L., Lin, Z. Q. & Wong, A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Sci. Reports* (2020).

16. Jain, R., Gupta, M., Taneja, S. & Hemanth, D. J. Deep learning based detection and analysis of covid-19 on chest x-ray images. *Appl. Intell.* 1–11 (2020).

17. Sahlol, A. T. *et al.* Covid-19 image classification using deep features and fractional-order marine predators algorithm. *Sci. Reports* **10**, 1–15 (2020).

18. Gunraj, H., Sabri, A., Koff, D. & Wong, A. Covid-net ct-2: Enhanced deep neural networks for detection of covid-19 from chest ct images through bigger, more diverse learning (2021). 2101.07433.

19. Gunraj, H., Wang, L. & Wong, A. COVIDNet-CT: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images. *Front. Medicine* DOI: 10.3389/fmed.2020.608525 (2020).

20. He, X. *et al.* Sample-efficient deep learning for covid-19 diagnosis based on ct scans. *MedRxiv* (2020).

21. Zhang, K. *et al.* Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography. *Cell* **181**, 1423–1433 (2020).

22. Wong, A. *et al.* Covid-net s: Towards computer-aided severity assessment via training and validation of deep neural networks for geographic extent and opacity extent scoring of chest x-rays for sars-cov-2 lung disease severity (2020). 2005.12855.

23. Cohen, J. P. *et al.* Predicting covid-19 pneumonia severity on chest x-ray with deep learning. *Cureus* **12** (2020).

24. Ebadi, A. *et al.* Covidx-us – an open-access benchmark dataset of ultrasound imaging data for ai-driven covid-19 analytics (2021). 2103.10003.

25. Wong, A., Lee, J. R. H., Rahmat-Khah, H., Sabri, A. & Alaref, A. Tb-net: A tailored, self-attention deep convolutional neural network design for detection of tuberculosis cases from chest x-ray images (2021). 2104.03165.

26. Wong, A., Shafiee, M. J., Chwyl, B. & Li, F. Ferminets: Learning generative machines to generate efficient neural networks via generative synthesis (2018). arXiv:1809.05989.

27. Wong, A. Netscore: Towards universal metrics for large-scale performance analysis of deep neural networks for practical on-device edge usage (2018). 1806.05512.

28. of North America, R. S. RSNA pneumonia detection challenge. *https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data* (2019).

29. of North America, R. S. COVID-19 radiography database. *https://www.kaggle.com/tawsifurrahman/covid19-radiography-database* (2019).

30. Chung, A. Figure 1 COVID-19 chest x-ray data initiative. *https://github.com/agchung/Figure1-COVID-chestxray-dataset* (2020).

31. Chung, A. Actualmed COVID-19 chest x-ray data initiative. *https://github.com/agchung/Actualmed-COVID-chestxray-dataset* (2020).

32. Cohen, J. P., Morrison, P. & Dao, L. COVID-19 image data collection. *arXiv 2003.11597* (2020).

33. Tsai, E. B. *et al.* The rsna international covid-19 open annotated radiology database (ricord). *Radiology* 203957 (2021).

34. Lin, Z. Q. *et al.* Do explanations reflect decisions? a machine-centric strategy to quantify the performance of explainability algorithms (2019). arXiv:1910.07387.

35. Rajpurkar, P. *et al.* Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. (2017). 1711.05225.

36. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).