



COVID-Net CXR-2: An Enhanced Deep Convolutional Neural Network Design for Detection of COVID-19 Cases From Chest X-ray Images

Maya Pavlova^{1*}, Naomi Terhijan¹, Audrey G. Chung^{2,3}, Andy Zhao¹, Siddharth Surana⁴, Hossein Aboutaleb^{2,4}, Hayden Gunraj¹, Ali Sabri^{5,6}, Amer Alaref^{7,8} and Alexander Wong^{1,2,3}

¹ Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada, ² Waterloo AI Institute, University of Waterloo, Waterloo, ON, Canada, ³ DarwinAI Corp., Waterloo, ON, Canada, ⁴ Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada, ⁵ Department of Radiology, McMaster University, Hamilton, ON, Canada, ⁶ Niagara Health System, St. Catharines, ON, Canada, ⁷ Department of Diagnostic Imaging, Northern Ontario School of Medicine, Thunder Bay, ON, Canada, ⁸ Department of Diagnostic Radiology, Thunder Bay Regional Health Sciences Centre, Thunder Bay, ON, Canada

OPEN ACCESS

Edited by:

Peter Mandl,
Medical University of Vienna, Austria

Reviewed by:

Sumeet Saurav,
Central Electronics Engineering
Research Institute (CSIR), India
KC Santosh,
University of South Dakota,
United States

*Correspondence:

Maya Pavlova
mspavlova@uwaterloo.ca

Specialty section:

This article was submitted to
Pulmonary Medicine,
a section of the journal
Frontiers in Medicine

Received: 26 January 2022

Accepted: 12 May 2022

Published: 10 June 2022

Citation:

Pavlova M, Terhijan N, Chung AG, Zhao A, Surana S, Aboutaleb H, Gunraj H, Sabri A, Alaref A and Wong A (2022) COVID-Net CXR-2: An Enhanced Deep Convolutional Neural Network Design for Detection of COVID-19 Cases From Chest X-ray Images. *Front. Med.* 9:861680. doi: 10.3389/fmed.2022.861680

As the COVID-19 pandemic devastates globally, the use of chest X-ray (CXR) imaging as a complimentary screening strategy to RT-PCR testing continues to grow given its routine clinical use for respiratory complaint. As part of the COVID-Net open source initiative, we introduce COVID-Net CXR-2, an enhanced deep convolutional neural network design for COVID-19 detection from CXR images built using a greater quantity and diversity of patients than the original COVID-Net. We also introduce a new benchmark dataset composed of 19,203 CXR images from a multinational cohort of 16,656 patients from at least 51 countries, making it the largest, most diverse COVID-19 CXR dataset in open access form. The COVID-Net CXR-2 network achieves sensitivity and positive predictive value of 95.5 and 97.0%, respectively, and was audited in a transparent and responsible manner. Explainability-driven performance validation was used during auditing to gain deeper insights in its decision-making behavior and to ensure clinically relevant factors are leveraged for improving trust in its usage. Radiologist validation was also conducted, where select cases were reviewed and reported on by two board-certified radiologists with over 10 and 19 years of experience, respectively, and showed that the critical factors leveraged by COVID-Net CXR-2 are consistent with radiologist interpretations.

Keywords: COVID-19, chest X-ray, computer aided diagnosis, computer vision, deep neural networks

1. INTRODUCTION

As the global devastation of the coronavirus disease 2019 (COVID-19) pandemic continues, the need for effective screening methods has grown. A crucial step in the containment of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus causing the COVID-19 pandemic is effective screening of patients in order to provide immediate treatment, care, and isolation precautions. While the main screening method is reverse transcription polymerase chain reaction

(RT-PCR) testing (1), recent studies have shown that the sensitivity of such tests can be relatively low and highly variable depending on how or when the specimen was collected (2–5).

Chest X-ray (CXR) radiography screenings are a complimentary screening method to RT-PCR that has seen growing interest and increased usage in clinical institutes around the world. Studies have shown characteristic pulmonary abnormalities in SARS-CoV-2 positive cases such as ground-glass opacities, bilateral abnormalities, and interstitial abnormalities (6–10). Compared to other imaging modalities, CXR equipment is readily available in many healthcare facilities, is relatively easy to decontaminate, and can be used in isolation rooms to reduce transmission risk (11) due to the availability of portable CXR imaging systems (12). More importantly, CXR imaging is a routine clinical procedure for respiratory complaint (13), and thus is frequently conducted in parallel with viral testing to reduce patient volume.

Despite the growing interest and usage of CXR radiography screenings in the COVID-19 clinical workflow, a challenge faced by clinicians and radiologists during CXR screenings is differentiating between SARS-CoV-2 positive and negative infections. More specifically, it has been found that potential indicators for SARS-CoV-2 infections may also be present in non-SARS-CoV-2 infections, and the differences in how they present can also be quite subtle. As such, computer-aided screening systems are highly desired for assisting front-line healthcare workers to streamline the COVID-19 clinical workflow by more rapidly and accurately interpreting CXR images to screen for COVID-19 cases.

Motivated by this, we launched the COVID-Net open source initiative at <http://www.covid-net.ml> (14–21) for accelerating the advancement and adoption of deep learning for tackling this pandemic. While the initiative has been successful and leveraged globally, the continuously evolving nature of the pandemic and the increasing quantity of available CXR data from multinational cohorts has led to a growing demand for ever-improving computer-aided diagnostic solutions as part of the initiative. Since the launch of the COVID-Net open source initiative, there have been many studies in the area of COVID-19 case detection using CXR images (22–29) emphasizing appropriate data curation and training regimes (30–32), with many leveraging the open access datasets and open source deep neural networks made publicly available through this initiative (33–49).

In this work, we introduce COVID-Net CXR-2, an enhanced deep convolutional neural network design for COVID-19 chest X-ray detection built on a greater quantity and diversity of patients than the original COVID-Net network design (14). To facilitate this, we introduce a benchmark dataset that is, to the best of the authors' knowledge, the largest, most diverse open access COVID-19 CXR cohort, with patients from at least 51 countries. We leverage explainability-driven performance validation to audit COVID-Net CXR-2 in a transparent and responsible manner to ensure the decision-making behavior is based on relevant visual indicators for improving trust in its usage. Furthermore, radiologist validation was conducted where select cases were reviewed and reported on by two board-certified radiologists with over 10 and 19 years of experience, respectively.

While not a production-ready solution, we hope the open-source, open-access release of COVID-Net CXR-2 and the respective CXR benchmark dataset will help encourage researchers, clinical scientists, and citizen scientists to accelerate advancements and innovations in the fight against the pandemic.

The paper is organized as follows. Section 2 describes the underlying methodology behind the construction of the proposed COVID-Net CXR-2 as well as the preparation of the benchmark dataset. Section 3 presents and discusses the efficacy and decision-making behavior of COVID-Net CXR-2 from both a quantitative perspective as well as a qualitative perspective. Finally, conclusions are drawn in Section 4.

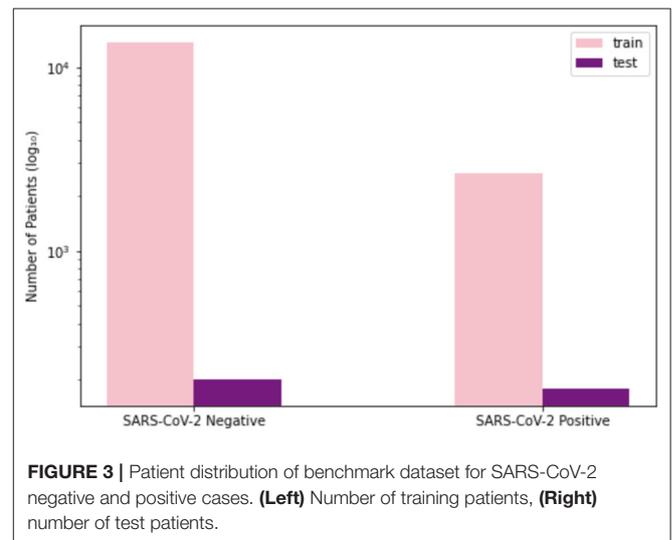
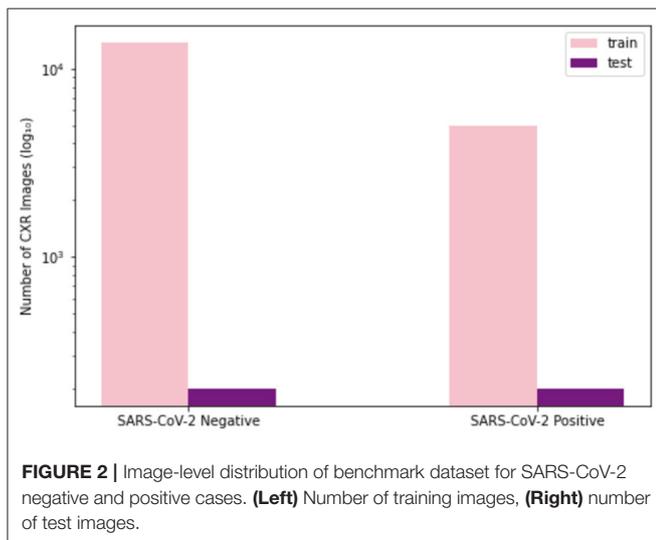
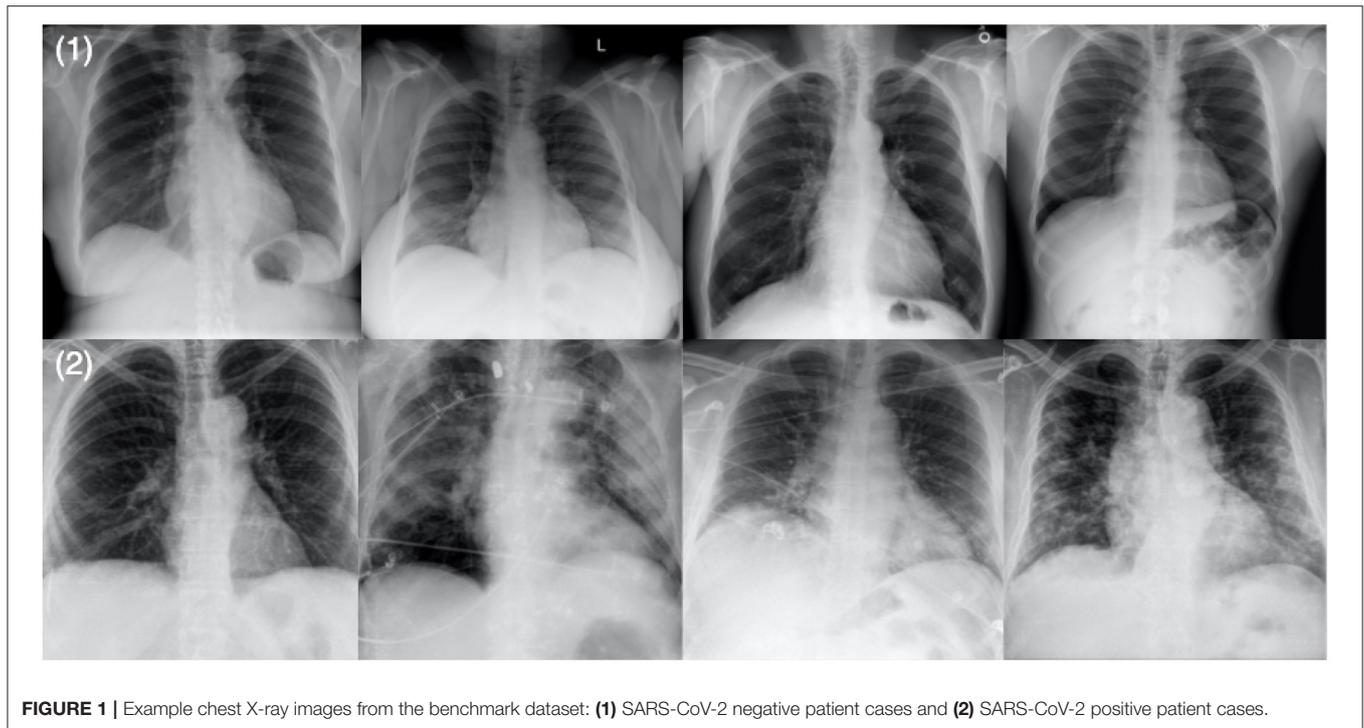
2. METHODOLOGY

In this study, we introduce COVID-Net CXR-2, an enhanced deep convolutional neural network design for detection of COVID-19 from chest X-ray images. To train and test the network, we further introduce a new CXR benchmark dataset which represents the largest, most diverse open access COVID-19 CXR dataset available, spanning a multinational patient cohort from at least 51 countries. All methods and experimental protocols in this study were carried out in accordance with the Tri-Council Policy Statement (TCPS2) and the University of Waterloo Research Integrity guidelines. The study has received ethics clearance from the University of Waterloo (42235). Data used in this study was curated by several organizations and initiatives from around the world with their own respective ethics clearance and informed consent. The details regarding data preparation, network design, and explainability-driven performance validation are described below.

2.1. Benchmark Dataset Preparation

To train and evaluate COVID-Net CXR-2, we first created a new CXR benchmark dataset with example images shown in **Figure 1**, unifying patient cohorts from several organizations and initiatives from around the world (50–56). The new CXR benchmark dataset comprises 19,203 CXR images from a multinational cohort of 16,656 patients from at least 51 countries, making it the largest, most diverse COVID-19 CXR dataset available in open access form to the best of the authors' knowledge. In terms of data and patient distribution, there are a total of 5,210 images from 2,815 SARS-CoV-2 positive patients and 13,993 images from 13,851 SARS-CoV-2 negative patients. The negative patient cases comprise of both no pneumonia and non-SARS-CoV-2 pneumonia patient cases, with 8,418 no pneumonia images from 8,300 patients and 5,575 non-SARS-CoV-2 pneumonia images from 5,551 patients. The distribution of CXR images in the benchmark dataset for SARS-CoV-2 negative and positive cases is shown in **Figure 2**, with respective patient distribution shown in **Figure 3**. Select patient cases from the benchmark dataset were reviewed and reported on by two board-certified radiologists with 10 and 19 years of experience, respectively.

The COVID-Net CXR-2 network is evaluated on a balanced test set of 200 SARS-CoV-2 positive images from 178 patients



and 200 SARS-CoV-2 negative images from 100 no pneumonia and 100 non-SARS-CoV-2 pneumonia patient cases. The test images were randomly selected from international patient cohorts curated by the Radiological Society of North America (RSNA) (50, 51), with the cohorts collected and expertly annotated by an international group of scientists and radiologists from different institutes around the world. The test set was selected in such a way to ensure no patient overlap between training and test sets.

Table 1 summarizes the demographic variables and imaging protocol variables of the CXR data in the benchmark dataset. It

can be observed that the patient cases in the cohort used in the benchmark dataset are distributed across the different age groups, with the mean age being 46.89 and the highest number of patients in the cohort being between the ages of 50–59.

The benchmark dataset, along with all data generation and preparation scripts, are available in an open source manner at <http://www.covid-net.ml>.

2.2. Network Design and Learning

Leveraging the aforementioned benchmark dataset, we built COVID-Net CXR-2 to be tailored for COVID-19 case detection

TABLE 1 | Summary of demographic variables and imaging protocol variables of CXR data in the benchmark dataset. Age and sex statistics are expressed on a patient level, while imaging view statistics are expressed on an image level.

Age		
Mean ± Std		46.89 ± 17.65
<20		1,026 (6.2%)
20–29		1,821 (10.9%)
30–39		2,268 (13.6%)
40–49		2,908 (17.5%)
50–59		3,486 (20.9%)
60–69		2,358 (14.2%)
70–79		1,010 (6.1%)
80–89		300 (1.8%)
90+		86 (0.5%)
Unknown		1,393 (8.4%)
Sex		
Male		8,774 (52.7%)
Female		6,768 (40.6%)
Unknown		1,114 (6.7%)
Imaging view		
PA		9,321 (48.5%)
AP		7,307 (38.1%)
Unknown		2,575 (13.4%)

from CXR images using machine-driven design. The machine-driven design exploration strategy automatically discovers highly customized and unique macroarchitecture and microarchitecture designs to optimize the trade-off between accuracy and efficiency. More specifically, the concept of generative synthesis (57) was leveraged, where the macroarchitecture and microarchitecture designs of a tailored deep neural network architecture are determined by an optimal generator \mathcal{G} whose generated deep neural network architectures $\{N_s | s \in S\}$ maximize a universal performance function \mathcal{U} [e.g., (58)], with constraints imposed on a set of operational requirements as defined by an indicator function $1_r(\cdot)$,

$$\mathcal{G} = \max_{\mathcal{G}} \mathcal{U}(\mathcal{G}(s)) \text{ subject to } 1_r(\mathcal{G}(s)) = 1, \forall s \in S. \quad (1)$$

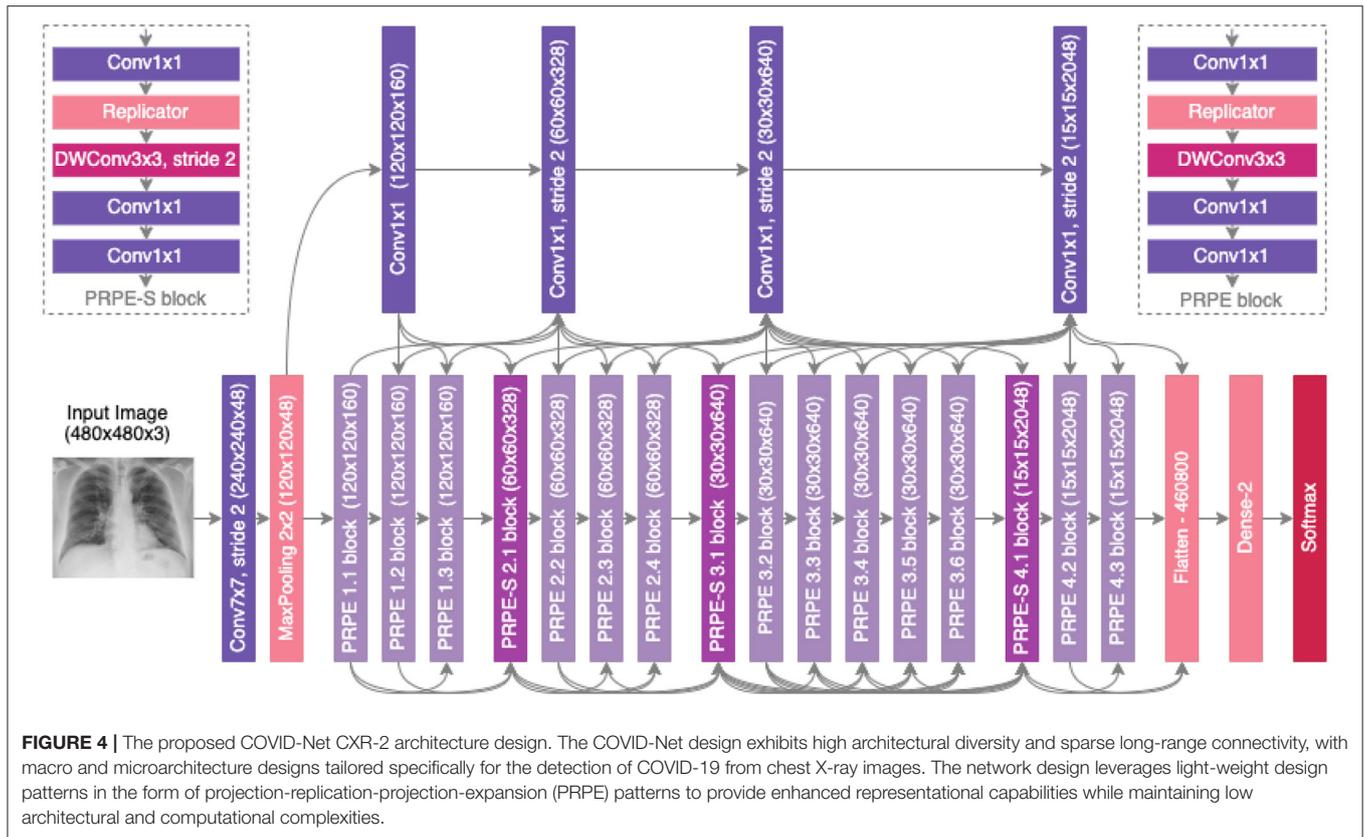
where S denotes a set of seeds to the generator. For the purpose of building COVID-Net CXR-2, the set of constraints imposed *via* indicator function $1_r(\cdot)$ were: (1) sensitivity $\geq 95\%$, and (2) positive predictive value (PPV) $\geq 95\%$.

A number of observations can be made about the proposed COVID-Net CXR-2 deep convolutional neural network architecture design shown in **Figure 4**. It can be observed that the proposed COVID-Net CXR-2 deep convolutional neural network possesses a light-weight network architecture design that exhibits notable diversity from both a macroarchitecture and microarchitecture design perspective. More specifically, the COVID-Net CXR-2 architecture design possesses a diverse mix of point-wise and depth-wise convolutions, and a very sparing use of conventional convolutions at the input stage of

the architecture. In particular, the network design leverages light-weight design patterns in the form of project-replication-projection-expansion (PRPE) patterns to provide enhanced representational capabilities while maintaining low architectural and computational complexities. More specifically, the PRPE design pattern replicates the input feature representations and disentangles these learned features through the use of depthwise convolutions before mixing them *via* a series of pointwise convolutions. This allows for more efficient representational learning for high efficiency while maintaining high representational performance. The key difference between PRPE and PRPE-S blocks is that PRPE-S blocks further reduces spatial dimensionality through the introduction of a strided depthwise convolution to further improve computational efficiency at appropriate positions in the network architecture. Furthermore, sparse use of long-range connections can also be observed in the network architecture design to strike a good balance between architectural and computational efficiency and representational capacity. The strong balance between efficiency and accuracy achieved by the proposed network highlights the utility of machine-driven design exploration for tailored architectures beyond the capabilities of manual, human designs. It is important to note that compared to the COVID-Net network architecture (14), the COVID-Net CXR-2 architecture design has very different macroarchitecture designs and microarchitecture designs, ranging from the use of replicators within a PRPE design pattern to greatly reduce computational complexity, leveraging two different forms of design patterns (PRPE and PRPE-S) as opposed to the more limited design patterns in COVID-Net, and much fewer parameters at each stage of the network architecture for greater computational and architectural efficiency. Finally, it is important to note that the network design leverages a more complex flatten layer as opposed to a less complex global average pooling layer, which illustrates how the machine-driven design exploration strategy takes into account different factors when optimizing for overall trade-offs between accuracy and efficiency.

Training was conducted using a binary cross-entropy loss and Adam optimization with learning rate of $1e-5$ on a batch size of 8 for 40 epochs. The final model was selected by tracking the validation accuracy throughout training and employing early stopping. All construction, training, and evaluation are conducted in the TensorFlow deep learning framework. As a pre-processing step, the CXR images were cropped (top 8% of the image) prior to training and testing to better mitigate commonly-found embedded textual information. The CXR images were then resampled to 480×480 and normalized to the range $[0, 1]$ *via* division by 255. Furthermore, data augmentation was leveraged during training with the following augmentation types: translation ($\pm 10\%$ in x and y directions), rotation ($\pm 10^\circ$), horizontal flip, zoom ($\pm 15\%$), and intensity shift ($\pm 10\%$). A batch re-balancing strategy was introduced to promote better distribution of SARS-CoV-2 positive cases and SARS-CoV-2 negative cases at a batch level.

The COVID-Net CXR-2 network and associated scripts are available in an open source manner at <http://www.covid-net.ml>.



2.3. Explainability-Driven Performance Validation

The trained COVID-Net CXR-2 network was audited to gain deeper insights into its decision-making behavior and ensure that it is driven by clinically relevant indicators rather than erroneous cues such as imaging artifacts and embedded metadata. We leveraged GSInquire (59) to conduct explainability-driven performance validation as it was shown to provide state-of-the-art explanations compared to other methods in literature, including gradient-based explainability methods such as Expected Gradients which has been shown to be superior for explainability than methods like Grad-CAM (60). In this work, we define explainability as the ability to obtain an explanation on the key factors from the input data that the model relied on to produce an output prediction and decision, presented in a way that a human can understand and interpret the results. More specifically, GSInquire takes advantage of the generative synthesis (57) strategy leveraged during machine-driven design exploration to identify and visualize the critical factors that COVID-Net CXR-2 uses to make predictions. Insights are gained through an inquisitor \mathcal{I} within a generator-inquisitor pair $\{\mathcal{G}, \mathcal{I}\}$, where the generator \mathcal{G} is the optimal generator used to generate COVID-Net CXR-2 as shown in Equation (1). More specifically, the inquisitor function is defined as $\mathcal{I}(\mathcal{G}; \theta_{\mathcal{I}})$, parameterized by $\theta_{\mathcal{I}}$ that given the generator \mathcal{G} , produces a set of parameter changes denoted by $\Delta\theta_{\mathcal{G}} = \mathcal{I}(\mathcal{G})$. The insights gained by the inquisitor \mathcal{I} are not only used to improve the generated deep

neural networks but can also be leveraged to interpret decisions made by the generated network.

Compared to other explainability methods in literature such as Grad-CAM (60) that produce relative heat maps that visualize variations in potential importance within an image, GSInquire has a unique capability of surfacing specific critical factors within an image that quantitatively impact the decisions made by the deep neural network. This makes the explanations easier to interpret objectively and better reflects the decision-making process of the deep neural network for validation purposes.

Explainability-driven performance validation is crucial for improved transparency and trust, particularly in healthcare applications such as clinical decision support. It can also help clinicians to uncover new insights into key visual indicators associated with COVID-19 to improve screening accuracy. To further audit the results for COVID-Net CXR-2, select patient cases from the explainability-driven performance validation were 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, while the second radiologist (A.A.) has over 19 years of radiology experience.

3. RESULTS AND DISCUSSION

To explore and evaluate the efficacy of the proposed COVID-Net CXR-2 deep convolutional neural network design for detecting COVID-19 cases from CXR images, we conducted a

TABLE 2 | Architectural and computational complexity of COVID-Net CXR-2 network in comparison to COVID-Net (14) and other state-of-the-art computer vision architectures. Best results highlighted in bold.

Architecture	Parameters (M)	MACs (G)
InceptionResNetV2 (61)	54.34	35.4
ResNet-50 (62)	24.97	17.75
InceptionV3 (63)	21.81	15.32
DenseNet201 (64)	18.33	19.82
COVID-Net (14)	11.8	7.5
COVID-Net CXR-2	8.8	5.6

TABLE 3 | Quantitative analysis. Sensitivity, positive predictive value (PPV), area under receiver operator curve (AUC), and accuracy of COVID-Net CXR-2 on the test data from the CXR benchmark dataset in comparison to COVID-Net (14) and other state-of-the-art computer vision architectures. Best results highlighted in bold.

Architecture	Sensitivity (%)	PPV (%)	AUC	Accuracy (%)
InceptionResNetV2 (61)	90.5	86.2	94.3	88.0
ResNet-50 (62)	85.3	95.4	96.0	89.8
InceptionV3 (63)	89.5	94.2	96.2	92.0
DenseNet201 (64)	92.0	88.9	94.7	90.3
COVID-Net (14)	93.5	100	99.2	94.0
COVID-Net CXR-2	95.5	97.0	99.4	96.3

TABLE 4 | Confusion matrix of COVID-Net CXR-2 network.

SARS-CoV-2	Negative	Positive
Negative	194	6
Positive	9	191

quantitative performance analysis to assess its architectural and computational complexity as well as its detection performance on the benchmark dataset. We further explored its decision-making behavior using an explainability-driven performance validation approach to audit COVID-Net CXR-2 in a transparent and responsible manner. The quantitative and qualitative results are presented and discussed in detail below.

3.1. Quantitative Analysis

Let us first explore the quantitative performance and underlying complexity of the proposed COVID-Net CXR-2 deep neural network architecture tailored for the detection of COVID-19 cases from CXR images. For comparison purposes, we also provide quantitative results on the test data for COVID-Net (14), which was shown to provide state-of-the-art performance for COVID-19 detection when compared to other methods in research literature, and other state-of-the-art architectures commonly leveraged in computer vision including InceptionResNetV2 (61), ResNet-50 (62), InceptionV3 (63), and DenseNet201 (64). The COVID-Net CXR-2 network builds upon the originally proposed COVID-Net architecture by offering a tailored network for SARS-CoV-2 detection of lower complexity

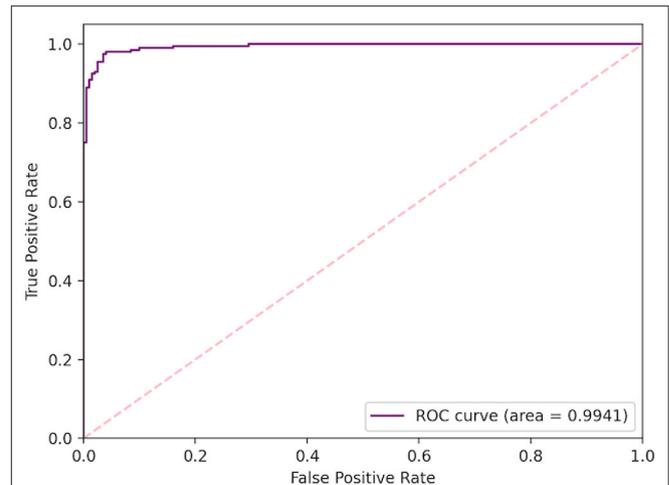


FIGURE 5 | Quantitative analysis. Receiver operating characteristic (ROC) curve for COVID-Net CXR-2 SARS-CoV-2 positive and negative CXR classification, with a computed area under curve (AUC) of 99.41%.

and higher detection performance as a result of training on a larger, more diversified dataset. In addition, the COVID-Net CXR-2 network is leveraged for binary SARS-CoV-2 positive and negative detection in regard to physician priorities, while the original COVID-Net network is utilized for normal, pneumonia, and SARS-CoV-2 multi-class classification. The state-of-the-art deep neural networks referenced in this study were trained using the same proposed CXR benchmark dataset, with the same hyperparameters including binary cross-entropy loss and Adam optimizer tuned to a learning rate of $1e-5$ and batch size of 8 for 40 epochs for optimal performance. The architectural and computational complexity of COVID-Net CXR-2 in comparison is shown in **Table 2**, with quantitative performance results shown in **Table 3**. It can be observed from the results that the COVID-Net CXR-2 network achieves overall the highest SARS-CoV-2 sensitivity, area under ROC curve (AUC), and accuracy in comparison to other state-of-the-art architectures while maintaining significantly lower network complexity. Specifically, the COVID-Net CXR-2 network achieved an architectural complexity of 8.8M parameters and computational complexity of 5.6G MACs that is ~ 25 and $\sim 84\%$ lower than the least and most complex comparison architectures of the COVID-Net and InceptionResNetV2. In addition, the proposed COVID-Net CXR-2 architecture achieved the highest test accuracy of 96.3%, highest area under ROC curve (AUC) of 99.4%, and highest SARS-CoV-2 sensitivity of 95.5%. In comparison to the other networks, the COVID-Net CXR-2 achieved 2% higher sensitivity than the next performing COVI

D-Net architecture, and 10.2% higher than the ResNet-50 deep network. In respect to positive predictive value (PPV), the COVID-Net CXR-2 network achieved a lower performance than the COVID-Net network at 97.0% test PPV, but was still able to outperform the other comparison architectures by a 1.6% minimum. This trade-off of higher sensitivity gained by

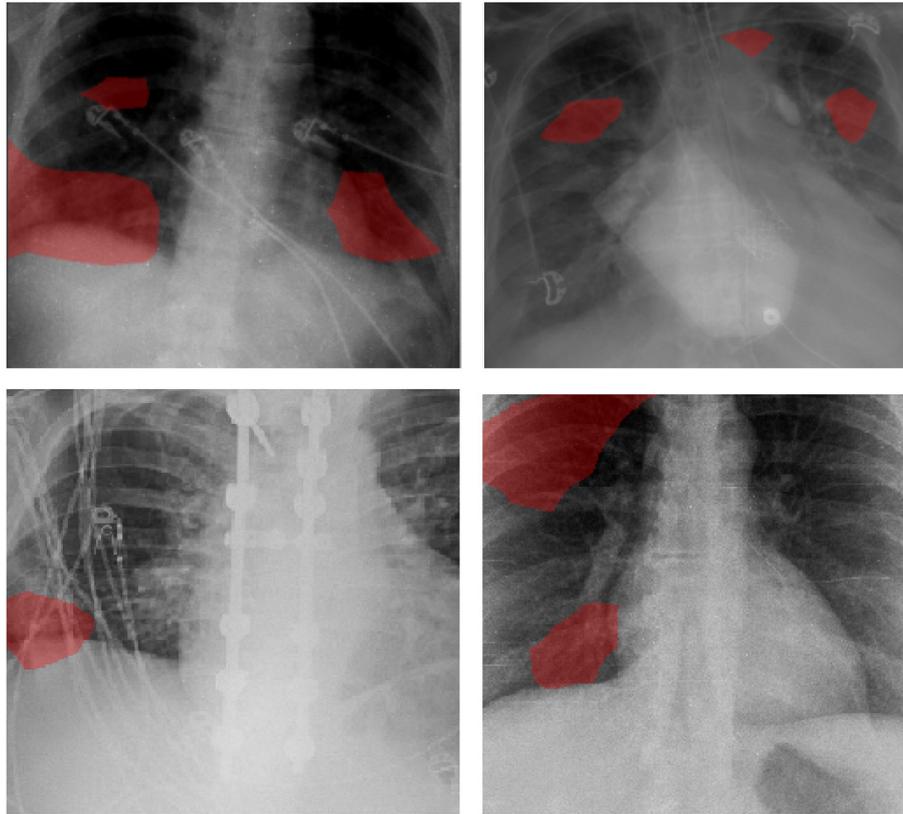


FIGURE 6 | Examples of patient cases and the associated critical factors (highlighted in red) as identified by GSInquire (59) during explainability-driven performance validation as what drove the decision-making behavior of COVID-Net CXR-2. Radiologist analysis was conducted on (top-left) Case 1 and (top-right) Case 2. Radiologist validation showed that several of the critical factors identified are consistent with radiologist interpretation.

COVID-Net CXR-2 compared to COVID-Net in exchange for a decrease in PPV (which is still quite high for COVID-Net CXR-2 at 97.0%) is a reasonable one given that a higher sensitivity results in fewer missed SARS-CoV-2 positive patient cases during the screening process. This is very important from a clinical perspective in controlling the spread of the SARS-CoV-2 virus during the on-going COVID-19 pandemic in light of the new highly infectious variants. Finally, **Table 4** and **Figure 5** provides a more detailed picture of the performance of COVID-Net CXR-2 *via* the confusion matrix and receiver operator characteristic (ROC) curve.

3.2. Qualitative Analysis

Examples of patient cases and the associated critical factors identified by GSInquire as the driving factors behind the decision-making behavior of COVID-Net CXR-2 are shown in **Figure 6**. It can be observed that the network primarily leverages areas in the lungs of the CXR images and is not relying on incorrect factors such as artifacts outside of the body, motion artifacts, and embedded markup symbols. From further investigation into the correctly detected COVID-19 cases, the critical factors typically identified correspond to clinically relevant visual cues such as ground-glass opacities, bilateral abnormalities, and interstitial abnormalities. These observations

indicate that the network's decision-making process is generally consistent with clinical interpretation.

This explainability-driven performance validation process is important for a number of important reasons from the perspectives of transparency, dependability, and trust. First of all, this process enabled us to audit and validate that COVID-Net CXR-2 exhibits dependable decision-making behavior since it is not only guided by clinically relevant visual indicators, but more importantly it is not dependent on erroneous visual indicators such as imaging artifacts, embedded markup symbols, and embedded text in the CXR images. This ensures the network does not make the right decisions for the wrong reasons. Second, this validation process allows for the discovery and identification of potential new insights into what types of clinically relevant visual indicators are particularly useful for differentiating between SARS-CoV-2 infections and non-SARS-CoV-2 infections. Such discoveries could be useful information for aiding clinicians and radiologists in better detecting SARS-CoV-2 infection cases during the clinical decision process. Finally, by validating the behavior of COVID-Net CXR-2 in a transparent and responsible manner, one can provide greater transparency and garner greater trust for clinicians and radiologists during usage in their screening process to make faster yet accurate assessments.

These quantitative and qualitative results show that COVID-Net CXR-2 not only provides strong COVID-19 detection performance, but also exhibits clinically relevant decision-making behavior.

3.3. Radiologist Analysis

The expert radiologist findings and observations with regards to the critical factors identified by GSInquire for select patient cases shown in **Figure 6** are as follows. In both cases, COVID-Net CXR-2 detected them to be patients with SARS-CoV-2 infection, which were clinically confirmed.

Case 1. According to radiologist findings, it was observed by both radiologists that there is an opacity at the right lung base, which is consistent with one of the identified critical factors leveraged by COVID-Net CXR-2. Additional imaging would be recommended by both radiologists.

Case 2. According to radiologist findings, it was observed by both radiologists that there are opacities in the right midlung and left paratracheal region that coincide with the identified critical factors leveraged by COVID-Net CXR-2 in that region. Additional imaging would be recommended by one of the radiologists.

As such, based on the radiologist findings and observations on the two patient cases, it was demonstrated that several of the identified critical factors leveraged by COVID-Net CXR-2 are consistent with radiologist interpretation.

4. CONCLUSION

In this study, we introduced COVID-Net CXR-2, an enhanced deep convolutional neural network design tailored for COVID-19 detection from CXR images that is built based on a greater quantity and diversity of patient cases than the original COVID-Net. A new benchmark dataset of CXR images representing a multinational cohort of 16,656 patients from at least 51 countries was also introduced, which is the largest, most diverse COVID-19 CXR dataset in open access form to the best of the authors' knowledge. Experimental results demonstrate that COVID-Net CXR-2 can not only achieve strong COVID-19 detection performance in terms of accuracy, sensitivity, and PPV, but also exhibit behavior consistent with clinical interpretation during an explainability-driven performance validation process, which was further validated based on radiologist interpretation. The hope is that the release of COVID-Net CXR-2 and its respective benchmark dataset in an open source manner will help encourage researchers, clinical scientists, and citizen scientists to accelerate advancements and innovations in the fight against the pandemic. Several potential limitations with the proposed

work include demographic imbalances that can affect how the network may make decisions for particularly patient groups, and limited data quantity in the current benchmark dataset that may lead to potential biases in the network's decision-making process. Further work involves the continued improvement of the benchmark dataset as well as architecture design, as well as exploration into other clinical workflow tasks (e.g., severity assessment, treatment planning, resource allocation, etc.) as well as other imaging modalities (e.g., computed tomography, point-of-care ultrasound, etc.). Furthermore, we aim to conduct more comprehensive auditing of both the benchmark dataset as well as the deep neural network to identify potential decision-making biases and potential gaps in the trustworthiness in the decision-making process, as well as conduct cross-validation experiments on a larger benchmark dataset as it becomes available.

DATA AVAILABILITY STATEMENT

The benchmark dataset, along with all data generation and preparation scripts, are available in an open source manner at: <http://www.covid-net.ml>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Waterloo (42235). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MP, NT, AC, and AW conceived the experiments. MP, NT, AZ, SS, HA, and HG conducted the experiments. AA and AS reviewed and reported on select patient cases and corresponding explainability results illustrating model's decision-making behavior. All authors analyzed the results. All authors reviewed the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

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.

REFERENCES

1. Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA*. (2020) 323:1843–44. doi: 10.1001/jama.2020.3786
2. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*. (2020) 296:115–7. doi: 10.1148/radiol.2020200432
3. Yang Y, Yang M, Shen C, Wang F, Yuan J, Li J, et al. Laboratory diagnosis and monitoring the viral shedding of SARS-CoV-2 infection. *Innovation*. (2020) 1:10061. doi: 10.1016/j.xinn.2020.100061

4. Li Y, Yao L, Li J, Chen L, Song Y, Cai Z, et al. Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19. *J Med Virol.* (2020) 92:903–8. doi: 10.1002/jmv.25786
5. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology.* (2020) 296:32–40. doi: 10.1148/radiol.2020200642
6. Wong H, Lam H, Fong A, Leung S, Chin T, Lo C, et al. Frequency and distribution of chest radiographic findings in COVID-19 positive patients. *Radiology.* (2020) 296:72–8. doi: 10.1148/radiol.2020201160
7. Warren MA, Zhao Z, Koyama T, Bastarache JA, Shaver CM, Semler MW, et al. Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. *Thorax.* (2018) 73:840–6. doi: 10.1136/thoraxjnl-2017-211280
8. Toussie D, Voutsinas N, Finkelstein M, Cedillo MA, Manna S, Maron SZ, et al. Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. *Radiology.* (2020).
9. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* (2020) 395:497–506. doi: 10.1016/S0140-6736(20)30183-5
10. Guan W, Ni Z, Hu Y, Liang J, Ou C, He J, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* (2020) 382:1708–20. doi: 10.1056/NEJMoa2002032
11. Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raouf S, 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) 296:172–80. doi: 10.1148/radiol.2020201365
12. Jacobi A, Chung M, Bernheim A, Eber C. Portable chest X-ray in coronavirus disease-19 (COVID-19): a pictorial review. *Clin Imaging.* (2020) 64:35–42. doi: 10.1016/j.clinimag.2020.04.001
13. Nair A, Rodrigues J, Hare S, Edey A, Devaraj A, Jacob J, et al. A British society of thoracic Imaging statement: considerations in designing local imaging diagnostic algorithms for the COVID-19 pandemic. *Clin Radiol.* (2020) 75:637. doi: 10.1016/j.crad.2020.05.007
14. Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci Rep.* (2020) 10:19549. doi: 10.1038/s41598-020-76550-z
15. Wong A, Lin ZQ, Wang L, Chung AG, Shen B, Abbasi A, et al. Towards computer-aided severity assessment via deep neural networks for geographic and opacity extent scoring of SARS-CoV-2 chest X-rays. *Sci Rep.* (2021) 11:9315. doi: 10.1038/s41598-021-88538-4
16. 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 Med.* (2020) 7:608525. doi: 10.3389/fmed.2020.608525
17. 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. *Front Med.* (2022) 8:729287. doi: 10.3389/fmed.2021.729287
18. Ebadi A, Xi P, MacLean A, Tremblay S, Kohli S, Wong A. COVIDx-US-An open-access benchmark dataset of ultrasound imaging data for AI-driven COVID-19 analytics. *arXiv:2103.10003.* (2021). doi: 10.48550/arXiv.2103.10003
19. Aboutalebi H, Pavlova M, Shafiee MJ, Sabri A, Alaref A, Wong A. COVID-Net CXR-S: Deep convolutional neural network for severity assessment of COVID-19 cases from chest x-ray images. *Diagnostics.* (2021) 12:25. doi: 10.3390/diagnostics12010025
20. Wong A, Lee JRH, 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. *arXiv:2104.03165.* (2021). doi: 10.48550/arxiv.2104.03165
21. Zhang J, Xi P, Ebadi A, Azimi H, Tremblay S, Wong A. COVID-19 detection from chest x-ray images using imprinted weights approach. *arXiv:2105.01710.* (2021). doi: 10.48550/arXiv.2105.01710
22. Das D, Santosh KC, Pal U. Truncated inception net: COVID-19 outbreak screening using chest X-rays. *Phys Eng Sci Med.* (2020) 43:915–25. doi: 10.1007/s13246-020-00888-x
23. Mukherjee H, Ghosh S, Dhar A, Obaidullah SM, Santosh KC, Roy K. Deep neural network to detect COVID-19: one architecture for both CT Scans and Chest X-rays. *Appl Intell.* (2021) 51:2777–89. doi: 10.1007/s10489-020-01943-6
24. Mukherjee H, Ghosh S, Dhar A, Obaidullah SM, Santosh KC, Roy K. Shallow convolutional neural network for COVID-19 outbreak screening using chest X-rays. *Cogn Comput.* (2021). doi: 10.1007/s12559-020-09775-9
25. Mahbub MK, Biswas M, Gaur L, Alenezi F, Santosh K. Deep features to detect pulmonary abnormalities in chest X-rays due to infectious disease: COVID-19, pneumonia, and tuberculosis. *Inform Sci.* (2022) 592:389–401. doi: 10.1016/j.ins.2022.01.062
26. Kamal MS, Chowdhury L, Dey N, Fong SJ, Santosh K. Explainable AI to analyze outcomes of spike neural network in COVID-19 chest X-rays. In: *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC).* (2021). p. 3408–15. doi: 10.1109/SMC52423.2021.9658745
27. Santosh KC, Ghosh S, GhoshRoy D. Deep learning for COVID-19 screening using chest X-rays in 2020: a systematic review. *Int J Pattern Recogn Artif Intell.* (2022) 36:2252010. doi: 10.1142/S0218001422520103
28. Gidde PS, Prasad SS, Singh AP, Bhatheja N, Prakash S, Singh P, et al. Validation of expert system enhanced deep learning algorithm for automated screening for COVID-Pneumonia on chest X-rays. *Sci Rep.* (2020) 11:23210. doi: 10.1038/s41598-021-02003-w
29. Aboutalebi H, Pavlova M, Gunraj H, Shafiee MJ, Sabri A, Alaref A, et al. MEDUSA: multi-scale encoder-decoder self-attention deep neural network architecture for medical image analysis. *Front Med.* (2022) 8:821120. doi: 10.3389/fmed.2021.821120
30. Santosh KC, Ghosh S. COVID-19 imaging tools: how big data is big? *J Med Syst.* (2021) 45:71. doi: 10.1007/s10916-021-01747-2
31. Santosh KC. COVID-19 prediction models and unexploited data. *J Med Syst.* (2020) 44:170. doi: 10.1007/s10916-020-01645-z
32. Santosh KC. AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data. *J Med Syst.* (2020) 44:93. doi: 10.1007/s10916-020-01562-1
33. Li X, Li C, Zhu D. COVID-MobileXpert: on-device COVID-19 screening using snapshots of chest X-ray. *arXiv:200403042.* (2020). doi: 10.1109/BIBM49941.2020.9313217
34. Minaee S, Kafieh R, Sonka M, Yazdani S, Soufi GJ. Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. *arXiv:200409363.* (2020). doi: 10.1016/j.media.2020.101794
35. Afshar P, Heidarian S, Naderkhani F, Oikonomou A, Plataniotis KN, Mohammadi A. COVID-CAPS: a capsule network-based framework for identification of COVID-19 cases from X-ray images. *arXiv:200402696.* (2020). doi: 10.3389/frai.2021.598932
36. Luz E, Silva PL, Silva R, Silva L, Moreira G, Menotti D. Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images. *arXiv:200405717.* (2020). doi: 10.1007/s42600-021-00151-6
37. Khobahi S, Agarwal C, Soltanalian M. CoroNet: a deep network architecture for semi-supervised task-based identification of COVID-19 from Chest X-ray images. *medRxiv.* (2020). doi: 10.1101/2020.04.14.20065722
38. Ucar F, Korkmaz D. COVIDDiagnosis-net: deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Med Hypotheses.* (2020) 140:109761. doi: 10.1016/j.mehy.2020.109761
39. Tartaglione E, Barbano CA, Berzovini C, Calandri M, Grangetto M. Unveiling COVID-19 from Chest X-ray with deep learning: a hurdles race with small data. *arXiv:200405405.* (2020). doi: 10.3390/ijerph17186933
40. Yeh CF, Cheng HT, Wei A, Chen HM, Kuo PC, Liu KC, et al. A cascaded learning strategy for robust COVID-19 pneumonia chest X-ray screening. *arXiv:200412786.* (2020). doi: 10.48550/arXiv.2004.12786
41. Zhang Y, Niu S, Qiu Z, Wei Y, Zhao P, Yao J, et al. COVID-DA: deep domain adaptation from typical pneumonia to COVID-19. *arXiv:200501577.* (2020). doi: 10.48550/arXiv.2005.01577
42. Karim MR, Döhmen T, Reholz-Schuhmann D, Decker S, Cochez M, Beyan O. DeepCOVIDexplainer: explainable COVID-19 predictions based on chest x-ray images (2020). doi: 10.1109/BIBM49941.2020.9313304
43. Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med.* (2020) 43:635–40. doi: 10.1007/s13246-020-00865-4
44. Farooq M, Hafeez A. COVID-ResNet: a deep learning framework for screening of COVID19 from radiographs (2020).
45. Majeed T, Rashid R, Ali D, Asaad A. Issues associated with deploying CNN transfer learning to detect COVID-19 from chest X-rays. *Phys Eng Sci Med.* (2020) 43:1289–303. doi: 10.1007/s13246-020-00934-8

46. Monshi MMA, Poon J, Chung V, Monshi FM. CovidXrayNet: optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR. *Comput Biol Med.* (2021) 133:104375. doi: 10.1016/j.combiomed.2021.104375
47. Jia G, Lam HK, Xu Y. Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method. *Comput Biol Med.* (2021) 134:104425. doi: 10.1016/j.combiomed.2021.104425
48. Ismael AM, Sengur A. Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Syst Appl.* (2021) 164:1114054. doi: 10.1016/j.eswa.2020.114054
49. Abbas A, Abdelsamea MM, Gaber MM. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Appl Intell.* (2021) 51:854–64. doi: 10.1007/s10489-020-01829-7
50. Radiological Society of North America. *RSNA Pneumonia Detection Challenge.* (2019). Available online at: <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data>
51. Tsai EB, Simpson S, Lungren MP, Hershman M, Roshkovan L, Colak E, et al. The RSNA International COVID-19 open annotated radiology database (RICORD). *Radiology.* (2021) 2021:203957. doi: 10.1148/radiol.2021203957
52. Radiological Society of North America. *COVID-19 Radiography Database.* (2019). Available online at: <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>
53. Chung A. *Figure 1 COVID-19 Chest X-ray Data Initiative.* (2020). Available online at: <https://github.com/agchung/Figure1-COVID-chestxray-dataset>
54. de la Iglesia Vayá M, Saborit JM, Montell JA, Pertusa A, Bustos A, Cazorla M, et al. *BIMCV COVID-19+: A Large Annotated Dataset of RX and CT Images From COVID-19 Patients.* (2020).
55. Chung A. *ActualMed COVID-19 Chest X-ray Data Initiative.* (2020). Available online at: <https://github.com/agchung/Actualmed-COVID-chestxray-dataset>
56. Cohen JP, Morrison P, Dao L. COVID-19 image data collection. *arXiv 200311597.* (2020). Available online at: <https://github.com/ieee8023/covid-chestxray-dataset>
57. Wong A, Shafiee MJ, Chwyl B, Li F. GenSynth: a generative synthesis approach to learning generative machines for generate efficient neural networks. *Electron Lett.* (2019) 55:986–9. doi: 10.1049/el.2019.1719
58. Wong A. NetScore: towards universal metrics for large-scale performance analysis of deep neural networks for practical on-device edge usage. In: *Image Analysis and Recognition: 16th International Conference.* Waterloo, ON (2018). doi: 10.1007/978-3-030-27272-2_2
59. Lin ZQ, Shafiee MJ, Bochkarev S, Jules MS, Wang XY, Wong A. Do explanations reflect decisions? A machine-centric strategy to quantify the performance of explainability algorithms. *arXiv:1910.07387.* (2019). doi: 10.48550/arXiv.1910.07387
60. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *2017 IEEE International Conference on Computer Vision (ICCV).* (2017). p. 618–26. doi: 10.1109/ICCV.2017.74
61. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv:1602.07261.* (2016). doi: 10.48550/arXiv.1602.07261
62. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: Leibe B, Matas J, Sebe N, Welling M, editors. *Computer Vision - ECCV 2016.* Cham: Springer International Publishing (2016). p. 630–45. doi: 10.1007/978-3-319-46493-0_38
63. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision *arxiv.org/abs/1512.00567.* (2015). doi: 10.1109/CVPR.2016.308
64. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected Convolutional Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Honolulu, HI (2017). p. 2261–9. doi: 10.1109/CVPR.2017.243

Conflict of Interest: AC was employed by and AW serves on the executive board of DarwinAI Corp.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Pavlova, Terhjan, Chung, Zhao, Surana, Aboutalebi, Gunraj, Sabri, Alaref and Wong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.