



OPEN

## Validation of expert system enhanced deep learning algorithm for automated screening for COVID-Pneumonia on chest X-rays

Prashant Sadashiv **Gidde**<sup>1,8</sup>, Shyam Sunder **Prasad**<sup>1,6,8</sup>, Ajay Pratap **Singh**<sup>2,6,8</sup>, Nitin **Bhatheja**<sup>2</sup>, Satyartha **Prakash**<sup>2</sup>, Prateek **Singh**<sup>2,6</sup>, Aakash **Saboo**<sup>3</sup>, Rohit **Takhar**<sup>3</sup>, Salil **Gupta**<sup>3</sup>, Sumeet **Saurav**<sup>1,6</sup>, Raghunandan **M. V.**<sup>2,6</sup>, Amritpal **Singh**<sup>7</sup>, Viren **Sardana**<sup>2,6</sup>, Harsh **Mahajan**<sup>3</sup>, Arjun **Kalyanpur**<sup>5</sup>, Atanendu Shekhar **Mandal**<sup>1,6</sup>, Vidur **Mahajan**<sup>3</sup>, Anurag **Agrawal**<sup>2,6</sup>, Anjali **Agrawal**<sup>4</sup>✉, Vasantha Kumar **Venugopal**<sup>3</sup>✉, Sanjay **Singh**<sup>1,6</sup>✉ & Debasis **Dash**<sup>2,6</sup>✉

SARS-CoV2 pandemic exposed the limitations of artificial intelligence based medical imaging systems. Earlier in the pandemic, the absence of sufficient training data prevented effective deep learning (DL) solutions for the diagnosis of COVID-19 based on X-Ray data. Here, addressing the lacunae in existing literature and algorithms with the paucity of initial training data; we describe CovBaseAI, an explainable tool using an ensemble of three DL models and an expert decision system (EDS) for COVID-Pneumonia diagnosis, trained entirely on pre-COVID-19 datasets. The performance and explainability of CovBaseAI was primarily validated on two independent datasets. Firstly, 1401 randomly selected CxR from an Indian quarantine center to assess effectiveness in excluding radiological COVID-Pneumonia requiring higher care. Second, curated dataset; 434 RT-PCR positive cases and 471 non-COVID/Normal historical scans, to assess performance in advanced medical settings. CovBaseAI had an accuracy of 87% with a negative predictive value of 98% in the quarantine-center data. However, sensitivity was 0.66–0.90 taking RT-PCR/radiologist opinion as ground truth. This work provides new insights on the usage of EDS with DL methods and the ability of algorithms to confidently predict COVID-Pneumonia while reinforcing the established learning; that benchmarking based on RT-PCR may not serve as reliable ground truth in radiological diagnosis. Such tools can pave the path for multi-modal high throughput detection of COVID-Pneumonia in screening and referral.

SARS-CoV-2 pandemic created a global disruption on an unprecedented scale<sup>1</sup>. Tracing, testing, isolation, treatment is the universally agreed-upon strategy along with the laid standard operating procedures for limiting the spread. But this sequence often faced a bottleneck at the testing stage. Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR), the gold standard, is a specialized test often not available in a timely manner<sup>2</sup> due to being time-consuming and rapid antigen test having a high false-negative rate. Further, infected patients need an assessment of severity which qRT-PCR does not provide. This led to advocacy of using ubiquitously available resources like Chest X-rays for identifying likely infections while awaiting qRT-PCR results and determining the severity of COVID pneumonia. The concept of causative diagnosis from X-rays might be

<sup>1</sup>CSIR-Central Electronics Engineering Research Institute, Pilani, Rajasthan 333031, India. <sup>2</sup>CSIR-Institute of Genomics and Integrative Biology, Mathura Road, New Delhi 110025, India. <sup>3</sup>Centre for Advanced Research in Imaging, Neurosciences Genomics (CARING), New Delhi, India. <sup>4</sup>Teleradiology Solutions, 12B Sriram Road, Civil Lines, Delhi 110054, India. <sup>5</sup>Teleradiology Solutions, 7G, Opposite Graphite India, Whitefield, Bangalore, Karnataka 560048, India. <sup>6</sup>Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India. <sup>7</sup>Maulana Azad Medical College (MAMC), New Delhi, India. <sup>8</sup>These authors contributed equally: Prashant Sadashiv Gidde, Shyam Sunder Prasad and Ajay Pratap Singh. ✉email: anjali.agrawal@telradsol.com; drvasanth@mahajanimaging.com; sanjay@ceeri.res.in; ddash@igib.res.in

concerning because the conventional approach of understanding pathologies is based on causality. Most deep learning algorithms on the other hand learn patterns of association by analyzing large volumes of data thereby offering an opportunity to identify associations hitherto unknown or unexplored by medical practitioners. Several studies have demonstrated that deep learning can make causative diagnoses like tuberculosis from X-rays as good as radiologists<sup>3</sup>.

COVID-19 pneumonia (CoV-Pneum) was seen to have characteristic findings on Chest X-rays (CxR); Bilateral peripheral hazy or consolidative opacities were seen; usually without pleural effusion or mediastinal lymphadenopathy<sup>4–7</sup>. These findings overlap with presentations of other atypical or viral pneumonia but are different from the typical bacterial pneumonia with large asymmetric consolidation, air bronchograms, and pleural effusion. Experienced radiologists use such distinctions to determine whether an abnormal CxR represents a high likelihood of CoV-Pneum. Non-pneumonia forms of COVID-19 cannot be diagnosed on CxR, at least by the human eye. Artificial Intelligence (AI) algorithms for COVID-19 prediction on CxR have claimed high performance in determining not only CoV-Pneum but also SARS-CoV-2 infection without any obvious pneumonia on CxR<sup>8–10</sup>. Most of these algorithms are deep learning (DL)-based<sup>11–15</sup> trained on a limited amount of COVID-19 data that is available, and without clear explainability. This led to three concerns; First, robustness or generalizability is typically low for unexplainable AI, especially when DL systems are trained on smaller datasets. Second, in small datasets when a subset is set aside for validation, apparent algorithm performance is artificially boosted. Last, many datasets do not make clear distinctions between the RT-PCR diagnosis of SARS-CoV-2 infection and the radiological or clinical one of CoV-Pneum. It is *a priori* unlikely that AI tools can detect SARS-CoV-2 infection that has not manifested in the lung, on a CxR. Thus, it seems reasonable to develop explainable AI tools that merge established DL capacity in detecting routine CxR pathology, with a logic-based determination of CoV-Pneum likelihood from nature and spatial distribution of the pathology. Whether such an approach would perform as well as pure DL-based methods is not known.

Here, we report an AI classifier of CoV-Pneum that is composed of an ensemble model consisting of three DL modules and an expert decision system. The DL modules are; pathology classification, lung segmentation, and an opacity detection module which are explainable to the extent of an activation map output. The expert decision system is a rule-based classification system that classifies the X-ray into one of three classes, namely COVID-unlikely, indeterminate, and COVID-likely and is fully explainable as well as modifiable as required based on the pathology in question. We tested this classifier on two test sets that had not been seen by the algorithm. First, a general population dataset of randomly selected CxR from an Indian quarantine center, where SARS-CoV-2 infection status was not known, was used as a testing dataset. To evaluate our proposed method on these CxRs, COVID likelihood marked by a single radiologist was compared against. Second, a curated data set for which SARS-CoV-2 infections status was clearly known and four independent radiologists had classified the probability of CoV-Pneum and marked the lesions. Other than concordance with RT-PCR and radiological diagnosis, we also defined concordance between the bounding boxes of lesions identified by the radiologist and the CovBaseAI algorithm. We found that our approach matches the performance of published DL systems while providing a high degree of explainability. Given the social and medical consequences of being declared “COVID positive”, this appeared to be a better approach. One of the reasons we chose to have an additional expert decision system on top of the three independent DL models was to ensure that no individual model controls the outcome<sup>16</sup>.

## Related work

The advent of deep learning techniques and their applications in the tasks involving classification<sup>17,18</sup> and object detection<sup>19,20</sup> has already proven its efficacy which led to the escalation in efforts for the development of deep learning based methods for the screening of Chest X-rays<sup>21</sup> for SARS-CoV-2. Multiple studies have proposed detection of SARS-CoV-2 based on CT X-ray data<sup>11–15,22,23</sup>. Leveraging CxR imaging for screening of COVID-19 pandemic has several advantages such as portability and accessibility, specifically in areas having inadequacy of resources and areas which are declared as hot spots for viral infection and thus can aid in triaging of affected patients.

As one of the early efforts<sup>22</sup> proposed a convolutional neural network based approach for screening of COVID-19 patients. They termed it COVID-Net and trained it using the open-source COVIDx dataset<sup>24</sup>. COVID-Net incorporated features such as architectural diversity, long-range connectivity, and a lightweight design pattern for classification among normal, COVID-19, and pneumonia CxRs. COVID-Net reported an accuracy of 93.3% on the COVIDx dataset, a sensitivity of 91% for COVID-19, and a high positive predictive value (PPV) reflecting few false positives and also incorporated explainability features by leveraging GSInquire<sup>25</sup>. In another study proposed by Ozturk et al., DarkCOVIDNet<sup>11</sup> architecture was presented which was inspired from DarkNet<sup>26</sup>, a proven deep learning architecture for high-speed object detection. DarkCOVIDNet was trained on a dataset of 1125 images comprising of COVID-19(+), Pneumonia, and No-Findings. It resulted in an accuracy of 98.08% and 87.02% for binary and three classes, respectively. Performance of both the architectures<sup>11,22</sup> discussed above can be significantly improved provided they are trained on a larger dataset and can result in increased generalization. However to overcome the issue of data scarcity Afshar et al.<sup>12</sup> proposed COVID-CAPS based upon capsule networks<sup>27</sup> as conventional convolutional neural networks (CNNs) can lose spatial information between image specimens and need substantially larger datasets to train for better performance. COVID-CAPS on the other hand is capable of handling small datasets. COVID-CAPS was trained on a dataset having four labels i.e. Normal, Bacterial, Non-COVID Viral, and COVID-19, and obtained an accuracy of 95.7% and specificity of 95.8% which is further enhanced after incorporating pre-training and transfer learning based on an additional dataset of X-ray images. Many of these models just reported cross-validation accuracy and did not reflect upon the need to test the developed algorithms on independent blind datasets based on both the RT-PCR tests and the radiological findings. Maguolo and Nanni<sup>28</sup> who have reviewed the recently published literature on CxR

based COVID-19 detection and included most of the recently published datasets which had been utilized for deep learning based COVID-19 detection; demonstrated that models could identify the source correctly even after the lung fields are masked with black boxes thereby concluding that most models might be biased and learn to predict features that depend more on the source dataset than the relevant medical condition. A review by Michael Roberts et al.<sup>16</sup> described the common pitfalls when developing an AI-based method for COVID-19 detection and prognosis. Out of the 2212 papers, they found in the literature, they filtered these papers based on a pre-set criteria to finally review 62 papers. While reviewing these papers, it was observed that none of the methods published are suitable for adoption in a clinical setting due to flaws in their methodology and the inclusion of certain underlying biases. In view of the aforementioned literature, we have tried to comply with most of the recommendations and address their lacunae; such as using standard and curated datasets, not using RT-PCR as the only ground truth, and various methodological considerations along with a robust rule-based system devised by experts themselves, etc.

Arias-Londoño et al.<sup>29</sup> described how pre-processing improves performance and explainability during detection of COVID-19 from CxRs. They found that segmenting the lungs automatically before detecting COVID-19 improves explainability performance considerably and is useful for deployment in a clinical setting. BS-Net<sup>30</sup> published by Alberto Signoroni et al. demonstrated an end-to-end framework for a multi-regional score to assess the degree of lung compromise on CxRs. They have curated an annotated dataset of over 4000 CxRs in-house and have built a Deep Learning based method for segmentation, alignment, and COVID-19 severity score calculation. They have divided the lung into six regions and the severity score was calculated for each region. The final output is provided along with super-pixel-based explainability maps. CovBaseAI follows a similar approach by integrating the diagnostic and detection information from deep learning modules with an Expert Decision System to calculate COVID-19 likelihood. For tools to be useful in screening and triaging; explainability and grading of disease is needed. Additionally, there have been several other investigations on COVID-19 detection employing deep learning based methods utilizing CT scans and chest X-ray images<sup>31–33</sup>.

Although several tools and models have been developed and published that claim to detect COVID-19 infection from CxRs, they lack from a perspective of testing outcomes on images from different geographical entities, image quality effects, and performance in comparison to RT-PCR as a gold standard. Hence, in the context of the aforementioned limitations, we went with a non-conventional approach to detect COVID pneumonia and not infection utilizing a set of pre-trained algorithms. Further, we developed an expert derived decision-making system which could harness the power of machine learning along with intellectual learning to be of use in real case scenarios and be utilized for as a solution for clinical triaging in resource-constrained environments. To address this, we propose a robust novel framework comprising coalescence of deep learning architectures and rules devised by radiologists for screening of COVID-19 disease. In the following sections, the proposed methodology is described in detail followed by datasets utilized for training evaluation and the results obtained.

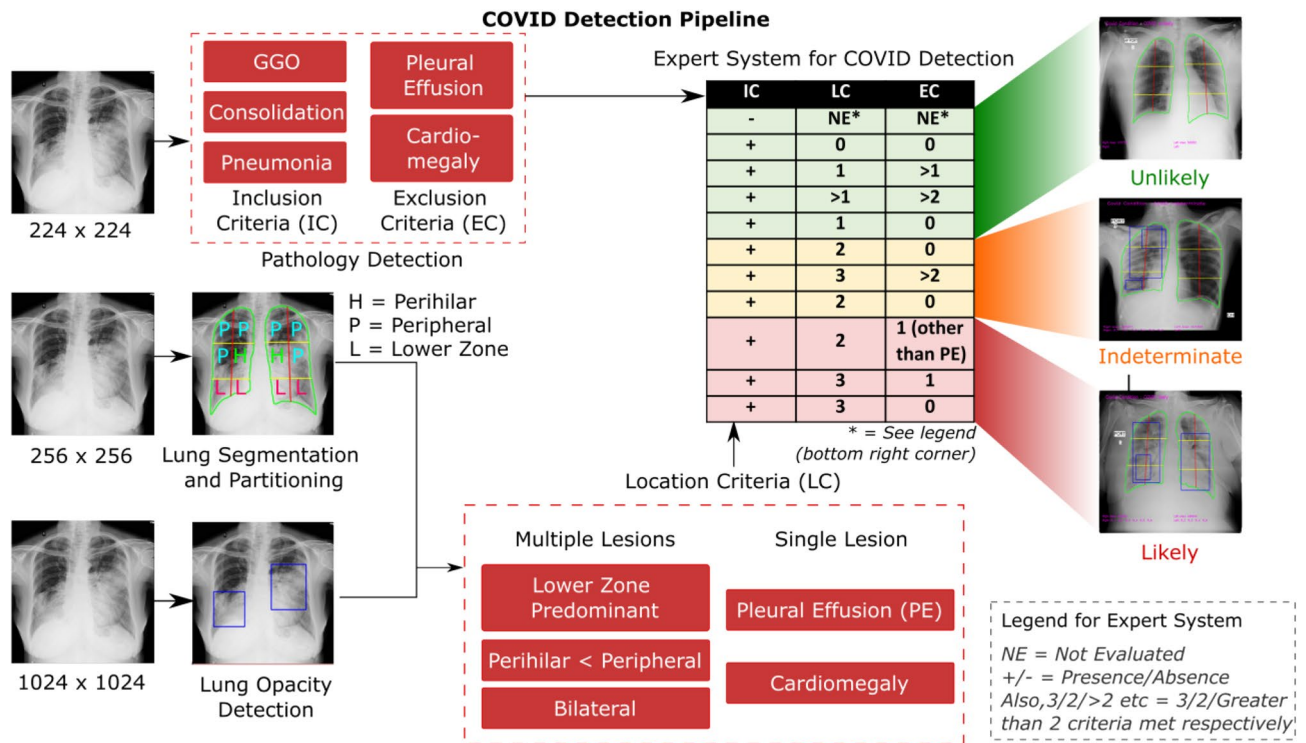
## Methodology

The architecture of CovBaseAI is depicted in Fig. 1. It comprises of three deep learning modules responsible for lung segmentation, lung opacity detection, and chest X-ray pathology detection. Each deep learning module is operating independently of the other and has a different input size. The input X-ray image is resized to  $256 \times 256$ ,  $1024 \times 1024$ , and  $224 \times 224$  for lung segmentation, lung opacity detection, and chest X-ray pathology classification module respectively. Lung fields identified as an output of the lung segmentation module are partitioned into different zones and the bounding boxes that are obtained from the opacity detection module are projected onto the original chest X-ray image. The pathology detection module provides the probabilities associated with pathology labels mentioned in the CheXpert dataset<sup>34</sup>. Finally, the outputs of all the deep learning modules are fed into the expert system to assign the COVID likelihood.

**Ethical approvals.** The study was duly approved by the Council of Scientific and Industrial Research (CSIR)—Institute of Genomics and Integrative Biology (IGIB) Human Ethics Committee and approval number CSIR-IGIB/IHEC/2020-21/02. The research was performed in accordance with the Indian Council of Medical Research (ICMR) Guidelines on Biomedical and Health Research with Human Participants 2017. Data from IIT Alumni Association and Mahajan Imaging was provided under MoU's with reference number AA/MOU/RS/080720 and CEERI/IGIB/TRS/MI/052020/01 respectively for research purposes in early times of COVID-19 and was anonymized for all purposes. Due to the retrospective nature of the study, informed consent was waived by the Council of Scientific and Industrial Research (CSIR)—Institute of Genomics and Integrative Biology (IGIB) Human Ethics Committee.

**Lung segmentation.** The lung segmentation module is a fully connected network adapted as a modification of U-Net<sup>35</sup>; a proven convolutional network for biomedical image segmentation. The CxR image was resized to  $256 \times 256$  and provided as input to the lung segmentation module to obtain the output lung mask. U-Net being a convolutional neural network is open to layer-wise modifications and blocks of various other neural networks to enhance the segmentation output. In the U-Net architecture, we replaced the encoder part with the VGG16<sup>36</sup> network which is pre-trained on the ImageNet<sup>37</sup> dataset and thus substantially reduces training weight convergence time and aids to prevent over-fitting<sup>38</sup> as has been demonstrated previously with Chest X-rays<sup>39</sup>. For training the segmentation module, a batch size of 4 was used, the learning rate was kept at 0.001, the loss function used was BCEWithLogitsLoss, Adam was used as the optimizer and it was trained for 30 epochs.

**Lung opacity detection.** Lung opacity detection was performed using a Faster R-CNN<sup>19</sup> architecture utilizing the VGG16 network as the backbone for feature extraction in Faster R-CNN. It is a two-stage object



**Figure 1.** CovBaseAI model architecture for COVID likelihood detection.

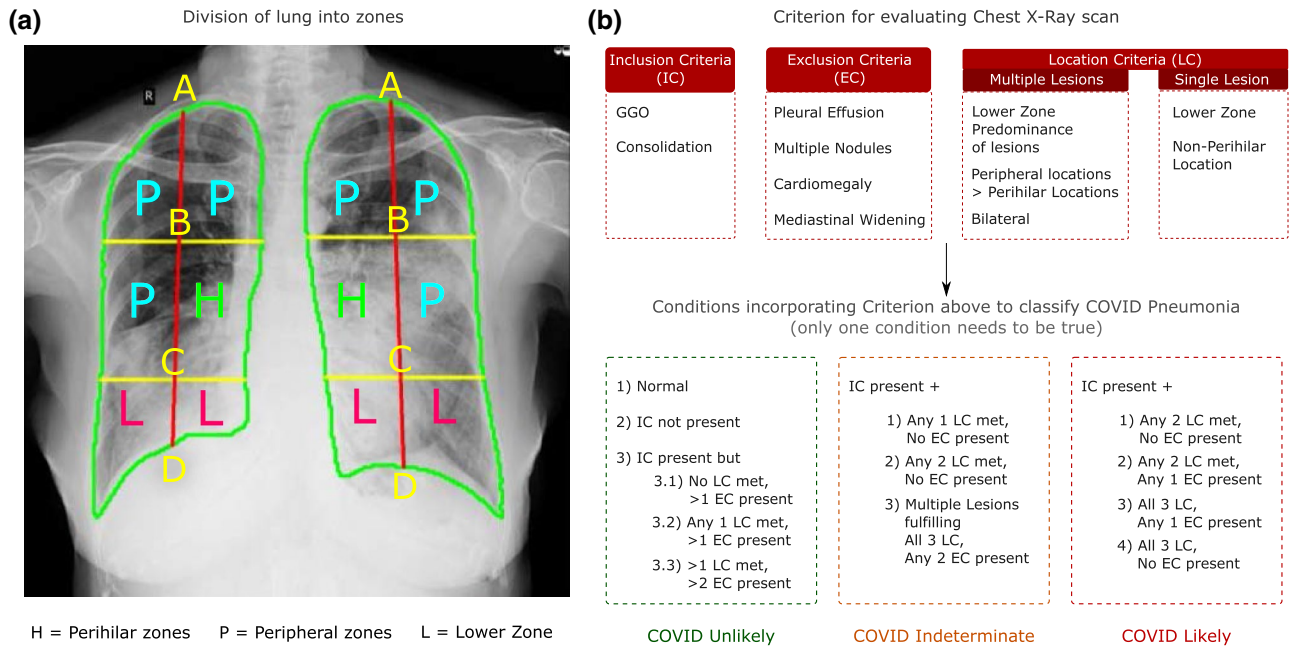
detection network that offers high accuracy as compared to a single-stage network. It is a merger of the region proposal network (RPN) and the Fast R-CNN<sup>40</sup> method. In the first stage; object-like proposals are generated, the second stage focuses on recognition of these proposals. Input to the Lung opacity module is the CxR image which is resized to 1024 × 1024 and the bounding boxes surrounding opacity regions are obtained as output. The hyper-parameters for opacity detection module were: for the two components in RPN, softmax classifier was used for determining the loss in score generation of each region predicted by the classification network. Smooth L1 loss was deployed to compute the loss in the regression layer, batch size was kept as 8, Adam was used as the optimizer, the learning rate was set to 0.001 and it was trained for 100 epochs.

**Pathology detection on chest X-rays.** Pathology detection on chest X-rays was performed using DenseNet-201<sup>17</sup> as a backbone network. The input CxR image was resized to 224 × 224 using the linear interpolation<sup>41</sup> technique and was initialized with pre-trained weights of the ImageNet dataset using a transfer learning approach similar to the technique used in Pham et al.<sup>42</sup>. The pathology detection module outputs the probability corresponding to each pathology which is taken into consideration by the rule-based COVID-19 likelihood detector to give the final prediction. In the CheXpert dataset, three labels are given corresponding to every image for each pathology i.e. positive, negative, and uncertain. While training the pathology detection module; uncertain labelled images were considered as negative according to the “U-Zero approach” mentioned in the CheXpert documentations<sup>34</sup> and only frontal CxRs were taken into account. The CheXpert validation dataset<sup>34</sup> does not contain uncertain labels, only positive and negative labels. In our work, uncertain labels were taken as negative and utilized because; when the U-Zero approach is followed as per literature<sup>34</sup>; the best AUC is achieved for consolidation specifically. Consolidation is one of the most important markers for the detection of pneumonia and is a major manifested anomaly. Hyper-parameters for training the pathology detection module were: a batch size of 32 was used, the learning rate was kept at 0.001, Adam was used as an optimizer, the Sigmoid classifier was employed and it was trained for 20 epochs.

**Lung opacity detection with lung partition.** Lung mask coordinates obtained from the lung segmentation module aids us in identifying the left lung and right lung on the input CxR which are then partitioned into six different regions to localize the perihilar, peripheral, middle, upper, and lower zone segments of the two lungs. This zonal classification roughly corresponds to the locations of anterior ribs<sup>43</sup>. The lung parenchyma overlying the upper two anterior ribs corresponds to the upper zone, the next two ribs to the mid zone, and the lower anterior ribs correspond to the lower zone. Partitioned lung along with the bounding box coordinates obtained from the opacity detection stage was developed to assist the rule-based expert system for COVID-19 classification to predict COVID-19 likelihood.

As shown in Fig. 2a after the lungs segmentation module detects the left lung and right lung separately, the uppermost (A) and lowermost (D) coordinate of each lung are connected. We then divided the line A-D into two equal partitions to obtain points B and C and drew a horizontal line on B and C thus portioning each lung





**Figure 2.** (a) Lung partitioning framework. (b) Rule-based expert decision system devised by radiologists.

into six zones. This was devised in view of the observed distribution of lesions in COVID pneumonia and was not meant to alter the known anatomy and lobular classification.

**Rule-based expert decision system for COVID classification.** The rules for classification were framed by a consensus among four radiologists with more than ten years of experience in thoracic radiology. The frequency of findings on chest imaging studies including Chest X-rays and CT scans reported by several researchers were considered in establishing the consensus rules<sup>4,5,44</sup>.

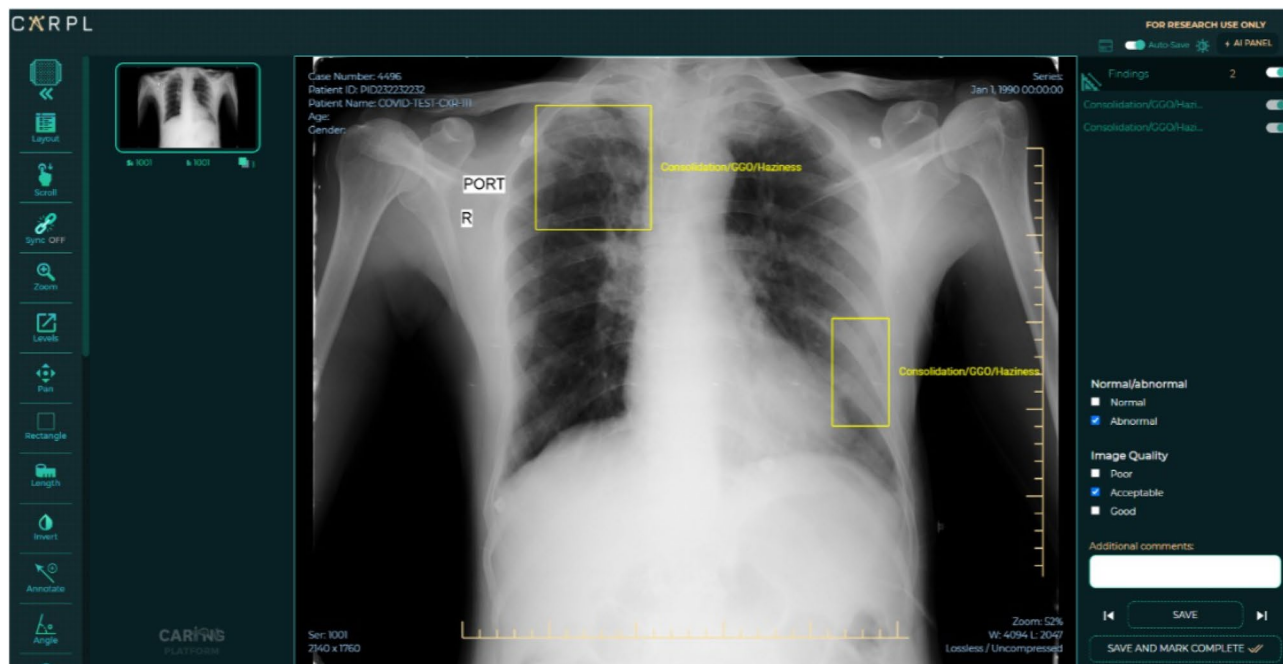
The rules were framed based on three broad sets of criteria: Lesion inclusion criteria, Lesion exclusion criteria, and Location criteria for lesions and pathology identified. Figure 2b delineates the rule-based inference logic utilized for defining COVID-19 likelihood. In the available literature, the algorithms have been trained to identify subtle features by neural networks which could help differentiate COVID images from Non-COVID images. We have here demonstrated an amalgamation of learning networks with the experience of radiologists to have expert systems which help define the COVID likelihood.

**Validation.** To evaluate the performance of the CovBaseAI model, we used two datasets. First, 1401 randomly selected CxR from an Indian quarantine center to assess effectiveness in excluding radiologic Cov-Pneum that may require higher care. Second, a curated data set with 434 RT-PCR positive cases of varying levels of severity and 471 historical scans containing normal studies and non-COVID pathologies, to assess performance in advanced medical settings. We compared the outputs of the model against the ground truth. Since radiological and RT-PCR ground truth vary and each has a significant error associated with it, this was done in multiple different ways as described in the dataset section and results. Generally, we prioritized radiological ground truth since that is the most relevant use case scenario at the current state of technology.

### Datasets

**Training and validation datasets. Lung segmentation.** For training and validation of lung segmentation module, a subset of RSNA Pneumonia Detection Challenge dataset<sup>45</sup> was used which is similar to <https://github.com/limingwu8/Lung-Segmentation>. It consists of 1000 CxR images along with their corresponding binary mask. Out of these 1000 CxRs, 900 CxRs were used for training and 100 CxRs were used for validation. The images and their corresponding binary mask were originally of size 1024 × 1024 pixels but they were resized to 256 × 256 pixels as training the model using the original size required high RAM consumption.

**Lung opacity detection.** Lung Opacity detection discussed in the methodology section was performed using RSNA pneumonia detection challenge dataset<sup>45</sup> to identify and localize pneumonia. This dataset consists of 26,684 images of three different classes; Normal, Lung Opacity, and No Lung Opacity/Not Normal. The Lung Opacity class consists of 6012 images. The Normal Class consists of 8851 images and the No Lung Opacity/Not Normal class consists of 11,821 images in DICOM format. For Lung Opacity Class, coordinates in the form of X-min, Y-min, height, and width for lung opacity are provided. The size of each image is 1024 × 1024 pixels and the Lung Opacity class contains 9555 coordinates of lung opacities in 6012 images. The lung opacity detection module was validated on 1012 test X-ray images from the RSNA pneumonia detection challenge dataset.



**Figure 3.** A snippet of the CARPL platform used for annotation and validation. (Image courtesy: CARING, India).

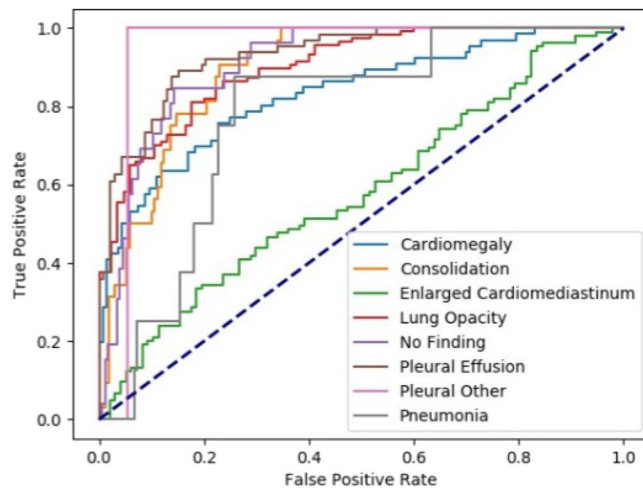
**Pathology detection on chest X-rays.** Pathology detection on chest X-rays was carried out using the CheXpert<sup>34</sup> dataset which was released by the Stanford ML group in 2019 as a competition for automated chest X-ray interpretation. CheXpert dataset consists of the following pathologies: Lung Lesion, Edema, Consolidation, No Findings, Atelectasis, Pneumothorax, Pleural Effusion, Pneumonia, Pleural Others, Cardiomegaly, Enlarged Cardio-mediastinum, Lung Opacity, Fracture, and Support Devices. CheXpert dataset originally consisted of 234 (AP View/ PA View/ Side View) validation X-ray images out of which only 202 (AP View/ PA View) frontal X-rays were selected as all the Deep Learning modules were trained only on frontal X-rays. The pathology detection module requires  $224 \times 224$  pixels as input size so we resized the original image to  $224 \times 224$  pixels. In the CheXpert dataset three labels were given corresponding to every image for each pathology i.e. positive, negative, and uncertain. The labels of the CheXpert dataset were used as mentioned in Pathology detection under Methodology section.

**Independent testing datasets.** As mentioned in “Introduction” section, to avoid bias based on datasets and preventing the algorithm from learning any biased classification, we carefully choose to test the algorithm on three different types of datasets which were standard datasets developed from pre-COVID times, COVID times, and based on RT-PCR results<sup>16</sup>.

To determine usefulness as a screening and triage tool, we used 1401 CxRs (IITAC1.4K) randomly selected from over 8500 CxRs acquired from the NSCI Dome initiative at Mumbai (an Indian quarantine center). SARS-CoV-2 infection status was not known for this dataset, but this reflects an actual mix of cases likely to present to quarantine centers where a determination has to be made regarding patients that can be safely kept in general isolation and those that may need medical care. These 1401 CxRs were carefully annotated by an experienced radiologist for the COVID likelihood. This dataset is composed of 135 COVID likely and 1266 COVID unlikely CxRs. (available at <http://covbase4all.igib.res.in>).

Furthermore, the CovBaseAI model can also be used as a diagnostic tool for SARS-CoV-2 infection or confirmed COVID pneumonia, we used a testing dataset which comprised of a total of 905 CxRs acquired from Mahajan Imaging, Delhi. These 905 CxRs comprised of 434 RT-PCR positive scans and 471 historical scans that were acquired before the worldwide outbreak of COVID-19. The X-rays were annotated with the consensus of four senior radiologists on the CARPL platform provided by the Centre for Advanced Research in Imaging, Neuroscience, and Genomics (CARING), India for the presence or absence of consolidation at the study level. A snippet of the CARPL platform used for annotation and testing is depicted in Fig. 3. The presence or absence of Pleural Effusion, Atelectasis, Cardiomegaly, Fibrosis, Mediastinal Widening, Nodule, Pleural Effusion, and Pneumothorax were also recorded. The CovBaseAI model was tested on this set of 905 CxRs using the following three combinations of test scans:

1. *Pneumonia Detector (PD1K)*: Prediction output of the CovBaseAI model was compared against pneumonia/consolidation label annotated by radiologists instead of their RT-PCR status. PD1K consists of 484 pneumonia/consolidation + ve and 421 pneumonia/consolidation -ve labels.



**Figure 4.** ROC curves of our model for different pathologies on CheXpert validation dataset.

Datasets	Sensitivity	Specificity	PPV	NPV	F1 score	Accuracy	MCC	AUC
IITAC1.4K	0.90	0.86	0.41	0.98	0.57	0.87	0.56	0.88
PD1K	0.84	0.81	0.83	0.81	0.84	0.82	0.65	0.89
CID1K	0.66	0.57	0.59	0.64	0.62	0.61	0.23	0.63
CPD600	0.83	0.77	0.79	0.81	0.81	0.80	0.60	0.83
COVIDx1K	0.78	0.97	0.79	0.97	0.78	0.95	0.76	0.89

**Table 1.** Performance metrics of CovBaseAI model on independent testing datasets. *PPV* positive predictive value, *NPV* negative predictive value, *MCC* Matthews correlation coefficient, *PD1K* pneumonia detector, *CID1K* COVID infection detector, *CPD600* COVID pneumonia detector.

2. *COVID Infection Detector (CID1K)*: Prediction output of the CovBaseAI model was compared against the RT-PCR results for testing its ability to detect COVID infection. CID1K consists of 434 RT-PCR +ve and 471 RT-PCR -ve labels.
3. *COVID Pneumonia Detector (CPD600)*: All cases with RT-PCR positive and radiologist annotated pneumonia/consolidation positive cases were taken as a positive class [336 cases] and all RT-PCR negative and consolidation negative cases [323 cases] were taken as negative class.

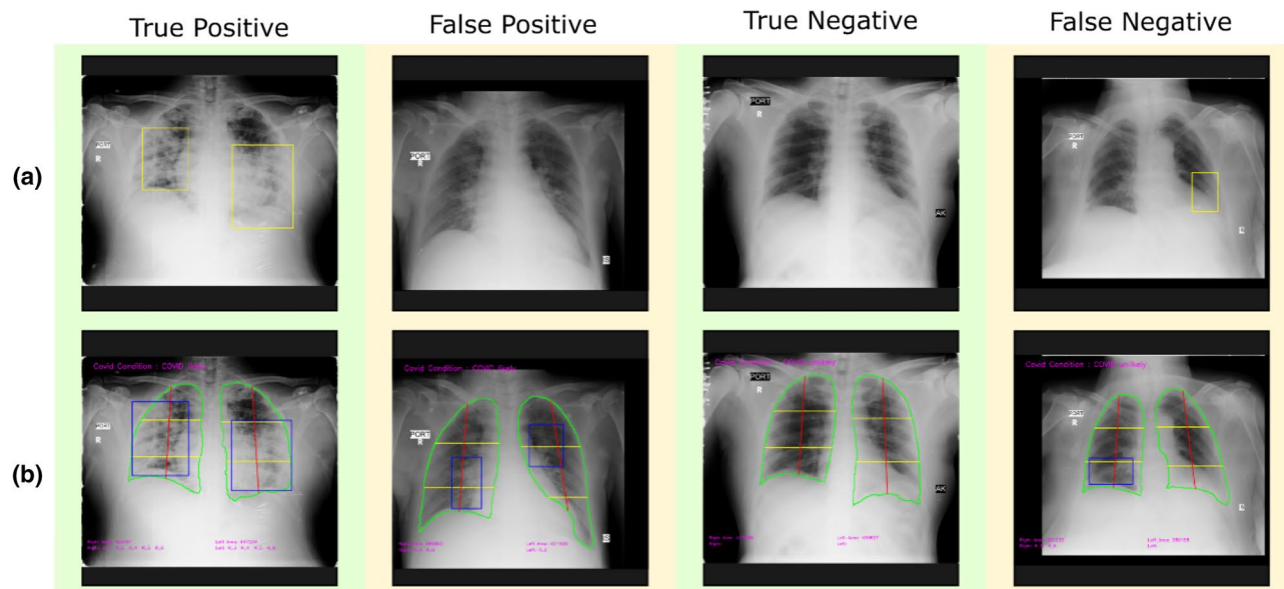
Additionally, the CovBaseAI model was also tested on a test dataset as mentioned in Wang et al.<sup>22</sup>. The aforementioned test dataset is referred to as COVIDx1K in the manuscript. The COVIDx1K dataset can be found at (<https://github.com/ddlab-igib/COVID-Net/blob/master/docs/COVIDx.md>). COVIDx1K comprises of 885 normal and 100 COVID-19 CxRs (test images with pneumonia labels were not considered in the COVIDx1K dataset).

## Results

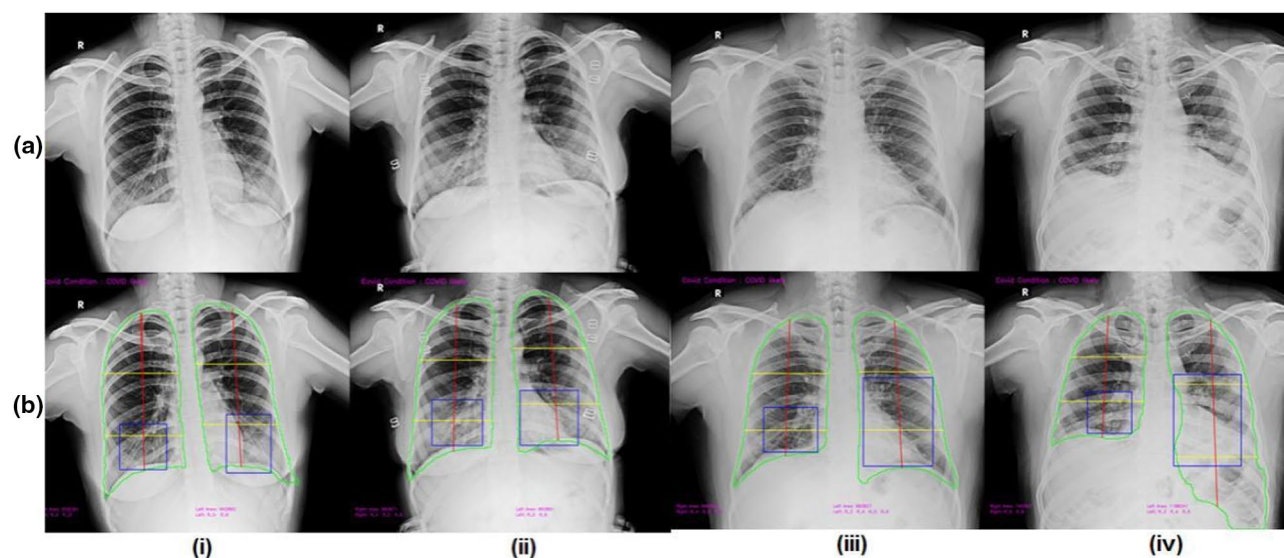
The segmentation algorithm of lung mask detection was validated using fivefold cross-validation on 100 chest X-rays, from the pool of 1000 X-rays from the RSNA Pneumonia Detection Challenge dataset. We obtained 0.90 as the average Jaccard similarity index<sup>46</sup>.

The lung opacity detection module was validated on 1012 test X-ray images from the RSNA Kaggle dataset. The exactness of object detection is usually well determined by mAP (Mean Average Precision)<sup>47</sup>, for opacity detection mAP of 0.34 is achieved. Based upon the pathology detection module's classification probability scores corresponding to pathologies included in the CheXpert dataset, the AUC's (Area under the ROC Curve) for Cardiomegaly, Consolidation, Lung Opacity, No Findings, Pleural Effusion and Pneumonia was 0.83, 0.90, 0.90, 0.91, 0.93, 0.78 respectively. The ROC curves for the pathology detection module for different pathologies on the CheXpert validation set have been depicted in Fig. 4.

Table 1 depicts the performance metrics of the CovBaseAI model on different independent validation datasets. Since the ground truth available for the validation dataset was either positive or negative (i.e. RT-PCR +ve/-ve or consolidation +ve/-ve) and the output of the CovBaseAI model is given in three classes (COVID likely, COVID indeterminate and COVID unlikely), the COVID indeterminate class was merged with COVID likely class for calculating performance metrics such as Sensitivity, Specificity, Accuracy, Negative Predictive Value (NPV), etc. Figure 5 depicts the sample of True positive, False positive, True Negative, and False Negative obtained by the CovBaseAI model. Validation studies corresponding to the CovBaseAI model were done using CARPL. Figure 6



**Figure 5.** Samples of true positive, false positive, true negative, and false negative from independent validation set of 905 CxRs (a) Ground truth CxR (b) CovBaseAI inferring result.



**Figure 6.** Samples of false-positive from IITAC1.4K dataset (a) Ground truth CxR (b) CovBaseAI inferring result.

shows bounding boxes of representative false-positive images from IITAC1.4K data that were read as normal by the radiologist. On review, the findings inside these bounding boxes were prominent bronchovascular markings, a common finding in the Indian subcontinent, with no clinical significance. Table 2 shows the concordance of bounding boxes (mAP) between lesions identified by AI and radiologists. In 905 CxRs (434 RT-PCR+ve and 471 historical scans) from Independent Validation datasets, read by four radiologists, intersections between AI-human pairs are like human-human pairs. Further, in the case of the IITAC1.4k dataset, which is read by a single radiologist, the majority of the time the bounding box of AI and radiologist had an intersection. Thus, the determination of COVID-19 pneumonia in our model is based on the same parts of the CxR as marked by the expert radiologist. Therefore, the explainability can be considered to be high.

To understand the relative performance of our model versus a publicly available COVID-19 CxR algorithm, COVID-CAPS that has reported high-performance characteristics on a subset of data used in Wang et al.<sup>22</sup>, we deployed COVID-CAPS on the Indian dataset. The sensitivity and specificity were only 33% and 67% on IITAC1.4K data that represents an actual use case scenario. The accuracy was below chance (50%) on the CPD600 dataset, which is more challenging. This illustrates the difficulty in the portability of DL solutions to other regions.



		Prediction box				
		CovBaseAI (mAP)	RAD 1 (mAP)	RAD 2 (mAP)	RAD 3 (mAP)	RAD 4 (mAP)
Ground truth	RAD 1	0.78	1	0.79	0.54	0.79
	RAD 2	0.73	0.80	1	0.51	0.75
	RAD 3	0.43	0.54	0.50	1	0.46
	RAD 4	0.81	0.81	0.75	0.47	1

**Table 2.** Bounding box analysis on 905 CxRs (434 RT-PCR +ve and 471 historical scans) from independent validation datasets. RAD radiologist, mAP mean average precision.

In contrast, CovBaseAI performed well across datasets, with higher specificity on a western dataset (COVIDx1K) than Indian ones, which reflects the fact that the DL components had never seen Indian data.

## Discussion

There is a great need for AI solutions that have explainability and an assurance of a minimum performance when exposed to data they have not been trained on. While COVID-19 has brought out this need globally, not discriminating between the developed and developing world, the fact of the matter is that this problem has been around for a long time. Here, we have addressed the problem of diagnosing COVID-19 pneumonia in Indian CxR without training on either COVID-19 or Indian CxR data, by using an explainable system driven by logic and expert opinion. While the solution has useful performance characteristics (see Table 1) across Indian (IITAC1.4K, PD1K, CPD600) and western (COVIDx1K) datasets, with high explainability, the insights lie in the differences. In particular, 97% specificity on a western dataset, but 86% specificity on a comparable Indian dataset is a notable comparison. Therefore, our algorithm provides a viable method that combines a radiologist-created EDS fed by three Deep Learning algorithms—the Lung Segmentation module, the Lung Opacity detection module, and the Pathology detection module using Chest X-Rays.

To avoid any bias due to non-relevant areas of the Chest X-Rays, we overlapped the segmented lung lobes prior to EDS similar to Arias-Londono et al.<sup>29</sup>. Using this methodology, we were able to remove diagnostically non-significant features from the deep learning models. We also divided the lungs to provide zone-based inputs to the EDS similar to BS-Net<sup>30</sup> which also divides the lungs into the zones before feeding into the prediction algorithms. While our lung zones can correspond to lung anatomical structures according to radiologists, our divisions were not anatomically inspired, unlike BS-Net. A pure DL algorithm that reportedly performed well on a subset of the western data, had only 67% specificity on the quarantine center dataset, which makes it unsuited for clinical use in India. COVID-CAPS<sup>12</sup> used publicly available Western datasets to train DL models for discriminating between bacterial, normal, non-viral COVID-19 and COVID-19. Using the same model on Indian datasets, we found a significant drop in performance compared to published results (33% sensitivity and 67% specificity on IITAC1.4K). This highlights the significant issue of “dirty lungs” in the Indian subcontinent compared to Western datasets thereby needing specialized training datasets. Therefore, similar to recommendations by Roberts et al.<sup>16</sup>, we utilized curated standardized datasets independent from the training datasets for testing the performance of the algorithm. We found that RT-PCR may not be a reliable method for benchmarking COVID-Pneumonia as suggested by Roberts et al.<sup>16</sup>. Given this learning, we took radiologist labels as ground truth for the development of testing datasets for our model.

Most AI solutions, large datasets for training, quality annotations, have come from the developed world. Yet, it is well known that even something as simple as CxRs looks different in regions with high ambient pollution, exposure to dirty fuel, endemic tuberculosis, low body mass, and other differences in lifestyle<sup>48</sup>. It is well known in the radiology community that this well-described “dirty lungs” appearance with prominent bronchovascular markings can easily be confused with abnormal. AI systems, however, are likely to give false positives when exposed to such images, if not previously exposed during training (see Fig. 6 for examples). Such false positives have been part of abstracts and anecdotal experiences<sup>49</sup> but publication bias against failures has led to a situation where AI solutions for interpreting CxR are working very well in published manuscripts, without similar enthusiasm in actual practice.

The only long-term solution is to have high-quality annotated images from Low-to-Middle-Income countries (LMIC) regions, via global initiatives. In the short term, the use of human logic filters can help in keeping the specificity high enough for clinical use. In our results, the specificity was 97% on a western dataset and 86% for India quarantine center data. Notably, the human logic filters did not substantially compromise sensitivity, which was in the 80–90% zone for COVID-pneumonia, falling to 66% for COVID infection. While it does not seem possible, using our approach, to detect SARS CoV2 infection, without pneumonia, this likely reflects a fundamental lack of sensitivity of CxR rather than a flaw in the algorithm and is consistent with WHO guidelines that do not recommend imaging for infection status determination<sup>50</sup>.

This is one of the limitations of our algorithm and is coherent with the existing literature for not utilizing X-ray based imaging as a diagnostic tool for SARS-CoV2 infection. Another limitation of our work was that training was performed using pre-COVID datasets. This could be argued in favour of the fact that without using a conventional approach, we went with a non-conventional expert-derived rule-based system to achieve a specificity of 86% on Indian datasets. However, COVID datasets could potentially be utilized in the future for training with rule-based systems to enhance accuracy in outcomes. Yet another limitation of the said algorithm is outcome

performance on geographically different datasets. This is primarily attributable to the training part with limited data and could be addressed with the use of annotated data from different geographical regions in the future.

The performance characteristics across all the COVID pneumonia datasets are in the clinically useful range and AUC was well above 0.8, the usual threshold at which diagnostic tools are considered useful. Further, the explainability was high (Table 2) and led to an understanding of the underlying issues (Fig. 6), which makes further improvements possible. Such work is now underway and will hopefully lead to more usable AI systems for the Indian subcontinent and similar global regions.

## Conclusion

Several Deep learning architectures have been proposed for COVID-19 detection from CxRs. Most of these methods used limited COVID-19 datasets for training purposes. Models trained on these limited datasets are not ready for real-world usage without training on a large variety of regional datasets to accommodate local variations of data acquisition hardware, population-specific differences, and temporal disease manifestation. For physicians to be able to use AI solutions for screening or triaging, the tools should provide robust detection results, across global data, with explainable findings and grading of severity, but requiring minimal amounts of new data so that solutions can be rapidly developed and deployed. Such an all-inclusive tool is yet to be developed for COVID-19. However, CovBaseAI is a useful step in that direction. We have developed and demonstrated an algorithm as an ensemble of three DL-based models under an umbrella of an expert derived rule-based decision system to detect COVID Pneumonia on chest X-Ray images. The algorithm while addressing fallacies in the available literature and having potential usage in screening and triaging has reinforced the learning that RT-PCR could not reliably be used as ground truth for such algorithms and there is a need for global curated and structured datasets which could be utilized for training of these algorithms to have a wide and universal applicability in clinical scenarios.

## Data availability

IITAC1.4K is available at <http://covbase4all.igib.res.in>. COVIDx1K is available at <https://github.com/ddlab-igib/COVID-Net/blob/master/docs/COVIDx.md>. The other test datasets generated (PD1K, CID1K, CPD600) for this study may be available for research purposes from the corresponding author upon reasonable request. Public datasets used for training and validation in the current study are available at their respective source locations (CheXpert<sup>34</sup>, RSNA Pneumonia Dataset<sup>45</sup>).

Received: 16 July 2021; Accepted: 1 November 2021

Published online: 01 December 2021

## References

1. WHO. *Who Characterizes COVID-19 as a Pandemic* (2020).
2. Wang, W. *et al.* Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* **323**, 1843–1844. <https://doi.org/10.1001/jama.2020.3786> (2020).
3. Qin, Z. Z. *et al.* Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Sci. Rep.* **9**, 1–10 (2019).
4. Venugopal, V. K. *et al.* A systematic meta-analysis of CT features of COVID-19: Lessons from radiology. *medRxiv* <https://doi.org/10.1101/2020.04.04.20052241> (2020).
5. Wong, H. Y. F. *et al.* Frequency and distribution of chest radiographic findings in patients positive for COVID-19. *Radiology* **296**, E72–E78. <https://doi.org/10.1148/radiol.2020201160> (2020).
6. Jacobi, A., Chung, M., Bernheim, A. & Eber, C. Portable chest x-ray in coronavirus disease-19 (COVID-19): A pictorial review. *Clin. Imaging* **64**, 35–42 (2020).
7. Ali, T. F., Tawab, M. & ElHariri, M. A. Ct chest of COVID-19 patients: what should a radiologist know?. *Egypt. J. Radiol. Nucl. Med.* **51**, 1–6 (2020).
8. Brunese, L., Mercaldo, F., Reginelli, A. & Santone, A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from x-rays. *Comput. Methods Programs Biomed.* **196**, 105608. <https://doi.org/10.1016/j.cmpb.2020.105608> (2020).
9. Panwar, H. *et al.* A deep learning and grad-cam based color visualization approach for fast detection of COVID-19 cases using chest x-ray and CT-scan images. *Chaos Solitons Fractals* **140**, 110190. <https://doi.org/10.1016/j.chaos.2020.110190> (2020).
10. Haghani, A., Molahasani Majdabadi, M., Choi, Y., Deivalakshmi, S. & Ko, S. COVID-CXNet: Detecting COVID-19 in frontal chest X-ray images using deep learning. *arXiv e-prints* arXiv: <http://arxiv.org/abs/2006.13807> (2020)
11. Ozturk, T. *et al.* Automated detection of COVID-19 cases using deep neural networks with x-ray images. *Comput. Biol. Med.* **121**, 103792. <https://doi.org/10.1016/j.compbiomed.2020.103792> (2020).
12. Afshar, P. *et al.* COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray Images. *arXiv e-prints* arXiv: <http://arxiv.org/abs/2004.02696> (2020)
13. Mei, X. *et al.* Artificial intelligence-enabled rapid diagnosis of COVID-19 patients. *MedRxiv* <https://doi.org/10.1101/2020.04.12.20062661> (2020).
14. Karim, M. R. *et al.* DeepCOVIDexplainer: Explainable COVID-19 diagnosis based on chest x-ray images. *arXiv: Image Video Process.* (2020).
15. Chowdhury, M. E. H. *et al.* Can ai help in screening viral and COVID-19 pneumonia?. *IEEE Access* **8**, 132665–132676. <https://doi.org/10.1109/access.2020.3010287> (2020).
16. Roberts, M. *et al.* Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat. Mach. Intell.* **3**, 199–217. <https://doi.org/10.1038/s42256-021-00307-0> (2021).
17. Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).
18. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778 (2016).
19. Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2017).
20. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788 (2016).

21. Çalli, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K. G., & Murphy, K. Deep learning for chest X-ray analysis: A survey. *Med. Image Anal.* 102125 (2021).
22. 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 e-prints* arXiv: <http://arxiv.org/abs/2003.09871> (2020)
23. Rajpurkar, P. *et al.* Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *ArXiv* <http://arxiv.org/abs/abs/1711.05225> (2017).
24. Cohen, J. P., Morrison, P. & Dao, L. COVID-19 image data collection. *arXiv* <http://arxiv.org/abs/2003.11597> (2020).
25. Lin, Z. *et al.* Do explanations reflect decisions? a machine-centric strategy to quantify the performance of explainability algorithms. *ArXiv* <http://arxiv.org/abs/abs/1910.07387> (2019).
26. Redmon, J. & Farhadi, A. Yolo9000: Better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6517–6525 (2017).
27. Sabour, S., Frosst, N. & Hinton, G. E. Dynamic routing between capsules. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, 3859–3869 (Curran Associates Inc., 2017).
28. Maguolo, G. & Nanni, L. A critic evaluation of methods for COVID-19 automatic detection from X-ray images. *arXiv e-prints* arXiv: <http://arxiv.org/abs/2004.12823> (2020)
29. Arias-Londoño, J. D., Gómez-García, J. A., Moro-Velázquez, L. & Godino-Llorente, J. I. Artificial intelligence applied to chest x-ray images for the automatic detection of COVID-19. A thoughtful evaluation approach. *IEEE Access* 8, 226811–226827. <https://doi.org/10.1109/ACCESS.2020.3044858> (2020).
30. Signoroni, A. *et al.* Bs-net: Learning COVID-19 pneumonia severity on a large chest x-ray dataset. *Med. Image Anal.* 71, 102046. <https://doi.org/10.1016/j.media.2021.102046> (2021).
31. Mahmud, T., Rahman, M. A. & Fattah, S. A. Covxnet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization. *Comput. Biol. Med.* 122, 103869. <https://doi.org/10.1016/j.cmpbiomed.2020.103869> (2020).
32. Khan, A. I., Shah, J. L. & Bhat, M. M. Coronet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput. Methods Programs Biomed.* 196, 105581. <https://doi.org/10.1016/j.cmpb.2020.105581> (2020).
33. Hussain, L. *et al.* Machine-learning classification of texture features of portable chest x-ray accurately classifies COVID-19 lung infection. *BioMed. Eng. OnLine* 19, 1–18 (2020).
34. Irvin, J. *et al.* Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *AAAI* (2019).
35. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. 9351 of *LNCS*, 234–241 (Springer, 2015) (available on arXiv: <http://arxiv.org/abs/1505.04597> [cs.CV]).
36. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations* (2015).
37. Deng, J. *et al.* ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09* (2009).
38. Iglovikov, V., & Shvets, A. A. (2018). TernaNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. *ArXiv* <http://arxiv.org/abs/1801.05746>.
39. Frid-Adar, M., Ben-Cohen, A., Amer, R. & Greenspan, H. Improving the Segmentation of Anatomical Structures in Chest Radiographs Using U-Net with an ImageNet Pre-trained Encoder: Third International Workshop, RAMBO 2018, Fourth International Workshop, BIA 2018, and First International Workshop, TIA 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings. [https://doi.org/10.1007/978-3-030-00946-5\\_17](https://doi.org/10.1007/978-3-030-00946-5_17) (2018).
40. Girshick, R. Fast r-cnn. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV, Vol. 15, 1440–1448. <https://doi.org/10.1109/ICCV.2015.169> (IEEE Computer Society, 2015).
41. Parker, J. A., Kenyon, R. V. & Troxel, D. E. Comparison of interpolating methods for image resampling. *IEEE Trans. Med. Imaging* 2, 31–39 (1983).
42. Pham, H. H., Le, T. T., Tran, D. Q., Ngo, D. T. & Nguyen, H. Q. Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels. *arXiv e-prints* arXiv: <http://arxiv.org/abs/1911.06475> (2019).
43. De Lacey, G., Morley, S. & Berman, L. *The Chest X-ray, a Survival Guide* (Saunders, 2008).
44. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395(10223), 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5) (2020).
45. Rsn pneumonia detection challenge. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge> (2018).
46. Eelbode, T. *et al.* Optimization for medical image segmentation: Theory and practice when evaluating with dice score or Jaccard index. *IEEE Trans. Med. Imaging* 1–1 (2020).
47. Liu, L. & Özsu, M. T. (eds) *Mean Average Precision 1703–1703* (Springer US, 2009).
48. Calderón-Garcidueñas, L. *et al.* Lung radiology and pulmonary function of children chronically exposed to air pollution. *Environ. Health Perspect.* 114, 1432–1437 (2006).
49. Khan, A. *et al.* Detection of chest x-ray abnormalities and tuberculosis using computer-aided detection vs interpretation by radiologists and a clinical officer. In *45th World Conf. on Lung Heal.* (2014).
50. World Health Organization. Use of Chest Imaging in COVID-19: A Rapid Advice Guide: Web Annex A: Imagingfor COVID-19: A Rapid Review. Technical documents, World Health Organization 76, <https://apps.who.int/iris/handle/10665/332326> (2020).

## Acknowledgements

All the authors duly acknowledge the IIT Alumni Council for their contribution of 8500 CxR data and Drs. Swarn Kanta and Virender Kumar Jain for radiologist and physician inputs on the 1401 CxRs as a subset of the above data. CARPL is a commercial product owned by CARING (Mahajan Imaging) and was utilized for this work as a part of R&D MoU between CSIR-IGIB and Mahajan Imaging. Availability of CARPL is duly acknowledged.

## Author contributions

P.S.G., S.S.P., A.P.S., S.P., A.S., S.S., S.Singh, and A.S.M. contributed to the model development. P.S., N.B., V.S., V.M., V.V., and D.D. contributed to model evaluation and validation. R.M.V., R.T., and S.G. contributed to model-building platform support. V.V., AnjaliA., A.K., and H.M. contributed as radiologists for the ground truth creation and rule-based logic for EDS. D.D., V.M., and AnuragA. conceptualized the idea and coordinated the research activity. P.S.G., S.S.P., A.P.S., A.Singh, P.S., N.B., V.S., V.V., AnuragA., AnjaliA., and D.D. contributed to the writing and editing of the manuscript. All authors reviewed and agreed with the final version of the manuscript.

## Funding

This activity was partially funded by project MLP-2002 (CSIR-IGIB), GAP-211 (CSIR-IGIB) and CSIR-Intelligent Systems Mission Mode Project HCP-0013 of Council of Scientific and Industrial Research (CSIR), India.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to A.A., V.K.V., S.S. or D.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021