



# CoVSeverity-Net: an efficient deep learning model for COVID-19 severity estimation from Chest X-Ray images

Sagar Deep Deb<sup>1</sup> · Rajib Kumar Jha<sup>1</sup> · Rajnish Kumar<sup>2</sup> · Prem S. Tripathi<sup>3</sup> · Yash Talera<sup>3</sup> · Manish Kumar<sup>4</sup>

Received: 14 June 2022 / Accepted: 8 December 2022 / Published online: 6 February 2023  
© The Author(s), under exclusive licence to The Brazilian Society of Biomedical Engineering 2023

## Abstract

**Purpose** COVID-19 is not going anywhere and is slowly becoming a part of our life. The World Health Organization declared it a pandemic in 2020, and it has affected all of us in many ways. Several deep learning techniques have been developed to detect COVID-19 from Chest X-Ray images. COVID-19 infection severity scoring can aid in establishing the optimum course of treatment and care for a positive patient, as all COVID-19 positive patients do not require special medical attention. Still, very few works are reported to estimate the severity of the disease from the Chest X-Ray images. The unavailability of the large-scale dataset might be a reason.

**Methods** We aim to propose CoVSeverity-Net, a deep learning-based architecture for predicting the severity of COVID-19 from Chest X-ray images. CoVSeverity-Net is trained on a public COVID-19 dataset, curated by experienced radiologists for severity estimation. For that, a large publicly available dataset is collected and divided into three levels of severity, namely Mild, Moderate, and Severe.

**Results** An accuracy of 85.71% is reported. Conducting 5-fold cross-validation, we have obtained an accuracy of  $87.82 \pm 6.25\%$ . Similarly, conducting 10-fold cross-validation we obtained accuracy of  $91.26 \pm 3.42$ . The results were better when compared with other state-of-the-art architectures.

**Conclusion** We strongly believe that this study has a high chance of reducing the workload of overworked front-line radiologists, speeding up patient diagnosis and treatment, and easing pandemic control. Future work would be to train a novel deep learning-based architecture on a larger dataset for severity estimation.

**Keywords** COVID-19 · Severity estimation · Chest X-Ray · CoVSeverity-Net

## Introduction

The novel Corona Virus Disease was first detected in Wuhan, Hubei Province, China, in December 2019. Acute Respiratory Syndrome with Severity Coronavirus2 is a virus that belongs to the “Coronavirus” (CoV) family and is a rare strain (SARS-CoV-2). In February 2020, the World Health Organization designated it as COVID-19. According to the most recent statistics, this virus has infected 22.6 million people as of today, with 792,000 people dying from it (Roser

et al. 2020). COVID-19 was declared a pandemic by the World Health Organization in March 2020. The number of new instances has climbed dramatically since then. The virus has infected over 210 countries, with the worst-affected countries being the USA, Brazil, Europe, and India. Specially countries like India, where the population density is extremely high needs to get a testing mechanism where the results can be delivered within fraction of seconds. But the testing mechanisms available with us, RT-PCR and the antigen tests are very much slow. So, for faster detection of the virus, developing faster testing mechanism is the need of the hour.

With the rapid spread of the global coronavirus disease 2019 (COVID-19) pandemic, radiology has become more crucial than ever in providing clinical insights to aid in diagnosis, treatment, and management. When the virus was first spreading rapidly in China, most of the early research focused on imaging features retrieved from COVID-19

---

Rajnish Kumar, Prem S. Tripathi, Yash Talera and Manish Kumar contributed equally to this work.

✉ Sagar Deep Deb  
sagardeepdeb@gmail.com

Extended author information available on the last page of the article.

infected people's computed tomography (CT) scans (Zhou et al. 2020; Chung et al. 2020; Ai et al. 2020). CT scanners, on the other hand, are not widely available in many parts of the world due to high costs, a high risk of virus transmission during patient transfer to and from CT imaging suites, and long decontamination intervals between scans, all of which have limited their use. Furthermore, CXR imaging devices are more widely available around the world than CT scanners due to their lower cost and faster sanitizing times; additionally, the availability of portable Chest X-Ray (CXR) equipment allows imaging to take place within an isolation chamber, significantly reducing the risk of transmission (Rubin et al. 2020; Jacobi et al. 2020).

According to Kucharski et al. (2020), isolation, testing, contact tracking, and physical separation all have a role in preventing the spread of SARS-CoV-2. However, the paucity of testing kits makes large-scale testing difficult, making contact tracing difficult. Furthermore, the RT-PCR test, which is widely regarded as the gold standard for COVID-19 detection, is not real-time. Most notably, in nations with larger population densities, a low-cost and quick testing system is required. For the detection of COVID-19 from CXR pictures, several researchers have developed deep learning-based systems. To varying degrees, they were successful.

It is worth mentioning that all patients suffering from COVID-19 do not need medical attention. COVID-19 infection severity scoring can aid in establishing the optimum course of treatment and care for a COVID-19 positive patient. With successful development of severity estimator, a patient can know if (s)he can be treated at home, need oxygen therapy, or needs ventilator support. This automatic severity estimation will also help to reduce the overburden on the radiologists and can also help to eliminate unnecessary rush to hospitals. This will significantly help in flattening the curve as shown in Fig. 1. The blue curve shows the number of patients infected by COVID-19 if no proper COVID-19 appropriate behaviour like wearing mask and

following social distancing is followed and the yellow curve shows the number of patients infected with COVID-19 if all the protocols are to be followed. The dotted horizontal line shows the number of people that can be treated at hospital. If a successful severity estimator can be developed COVID-19 will become a normal flu that can be treated easily.

Since the inception of COVID-19, several researchers have utilized various deep and machine learning techniques to diagnose the disease from different image modalities. The literature review is divided into three sections; the first subsection focuses on COVID-19 detection using CT-Scans, as initially, radiologists employed CT-Scans for COVID-19 detection. The second subsection focuses on COVID-19 detection from CXR images, while the third subsection discusses severity estimation from CT and CXR images using various image processing algorithms.

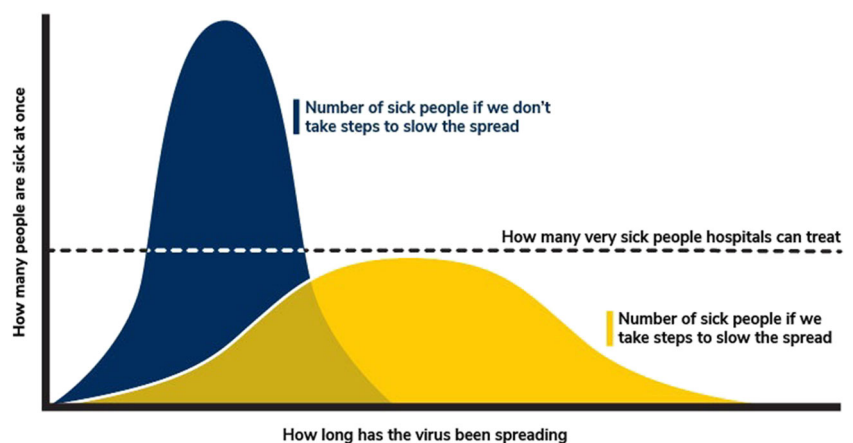
### Detection of COVID-19 from Chest CT images

Researchers in image processing and biomedical engineering are working tirelessly to combat the pandemic. Since the inception of this disease, several models have come up for automatic detection of COVID-19 from Chest CT images. Initially, the models were based on various image processing techniques, and later on, deep learning techniques came up.

With the development of Deep Learning and, mainly, CNNs, nearly all academicians have expressed an increased interest in employing Deep Learning to solve the challenge. CNN was first used to recognize hand digits (Fukushima et al. 1983) and was then applied to pattern recognition tasks in the medical profession (Ranjan et al. 2021, 2022). Deep Learning does not rely on handcrafted features, which is why DCNN is preferred. DCNN has been demonstrated to be the best image classification algorithm.

CT-Scans can easily be misclassified, especially when many causes of pneumonia are present at the same time. Pure ground-glass opacities (GGO) are the most common

**Fig. 1** Curve showing the progress of the virus



chest CT findings. However, additional abnormalities such as consolidations with or without vascular enlargement, interlobular septal thickening, and air bronchogram can also be seen (Li and Xia 2020). Polsinelli et al. (2020) extracted low-level features from Chest CT images and trained a light CNN for COVID-19 detection. The network proposed by them was based on SqueezeNet and has reported an accuracy of about 85.03%.

Dibag et al. (Singh et al. 2020) proposed a DCNN which is based on multi-objective differential evolution (MODE) for identification of COVID-19, which considers CT-Scans as an input image. They report a descent accuracy of 93%. Maghdid et al. (Maghded et al. 2020) created an AI-powered mobile app that analyses smartphone sensor signal data and examines the input CT-Scans to assess the extent of pneumonia in an individual. Wang et al. (2021) proposed an Inception migration-based learning model to detect COVID-19. A private dataset of 453 CT images was used for training and testing the proposed model. They have achieved an accuracy of 73.1%. Li et al. (2020a) reported an AUC of 0.96.

### COVID-19 detection from CXR images

Wang et al. (2020) proposed a deep learning-based architecture, named, COVID-Net, which classifies the input CXR images into four categories: normal, COVID-19, bacterial pneumonia, and viral pneumonia. For classification, they suggest a customized network. They claimed an accuracy rate of over 84.5%. Hemdan et al. (2020) proposed another deep learning-based ensemble model called COVIDX-net, which uses seven different pre-trained DCNN structures to classify CXR images into COVID and non-COVID classes. Sethy et al. (Sethy and Behera 2020) proposed ResNet50 model followed by SVM for high level feature extraction from the input CXR images. They report an accuracy of 95.38%. Another model based on the Xception network, proposed by Asif et al. (Khan et al. 2020) obtained an accuracy of 89.6%. Deb et al. (Deb and Jha 2020) proposed an ensemble model based on three different DCNN structures for the classification of CXR images into three classes. For COVID-19 identification, they then employed an ensemble model based on different pre-trained DCNN architectures. Using a publicly available dataset, Deb et al. (2022) reported accuracy of 88.98%. They performed a classification between three different classes, namely, Pneumonia, Normal and COVID-19. The accuracy of the performance was 93.48% when tested on a private dataset obtained from a local hospital. It is worth mentioning that roughly 50% of the patients with COVID-19 infection have a normal CT scan if examined during the incubation period (Kanne et al. 2020). Furthermore, compared to X-ray imaging, the radiation dose, which is a significant health risk, is substantially

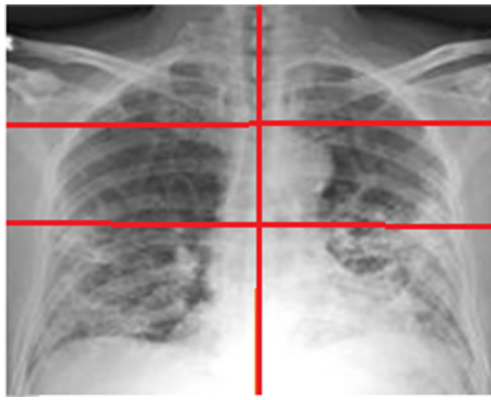
higher for CT scans. This is particularly critical for pregnant women and children, who are more likely to be exposed to high doses of radiation (Kim et al. 2016). Considering all of these factors, we decided to conduct our research using CXR images.

### Severity estimation

As the epidemic grows, automatic severity assessment systems that use deep learning to identify COVID-19 infected patients that require significant clinical treatment are becoming increasingly relevant. If illness progression, diagnosis time, and fatality rates are to be reduced, it is now critical to assess COVID-19 patients as soon as possible (Irmak 2021). Very little research was conducted for severity assessment. One primary reason may be the unavailability of large dataset. Tang et al. (2020) proposed a technique to identify COVID-19 patients based on severity, which is based on several machine learning approaches based on handmade features extracted from CT lung scans. The overall accuracy for this binary classification was 87.5%. The major drawback is that their experiment was conducted only using 176 CT images of COVID-19 patients. He et al. (2021) proposed a multi-task multi-instance learning approach to classify the input Chest CT images into two levels severity. Using 666 chest CT images, they have obtained an accuracy of 98.5%. Zhu et al. (2020) proposed transfer learning to classify COVID-19 patients into four different levels of severity but again their dataset was too small. Their dataset contains CXR images from 84 different patients and it contains a total of 131 images. Li et al. (2020b) used an automated deep learning method for estimating the severity from 531 CT Scan images. Covid-NetS, developed by Wong et al. (2021), is a regression model used to predict the severity score of a COVID-19-positive patient by looking at their Chest X-Ray (CXR) images. The severity score is based on geographic extent and opacity extent. Geographic extent is based on the consolidation of the lung image, and opacity extent calculates the degree of opacity.

As mentioned above, several researchers have proposed or published various models based on deep learning algorithms for the detection of COVID-19, but very few works have been done on the estimation of the severity of the disease. Most of the reported works on severity estimation were done on very small number of images. The main motivation behind developing automatic severity assessment approaches is to identify COVID-19 patients who require extensive clinical care. The contributions are listed below-

- A deep learning-based severity estimator called **CoVSeverity-Net** is proposed. **CoVSeverity-Net** is an ensemble network, which extracts low-level features from the input CXR images of COVID-19 positive



**Fig. 2** Subdivision of a sample CXR image, with three parts for each lung: superior zone, middle zone, and inferior zone

patients. It lets the patients know the severity level (mild, moderate, and severe) of the COVID-19 infection within a fraction of second.

- We have introduced a large curated dataset for severity estimation. The large dataset from Kaggle is curated by a group of expert radiologists into mild, moderate, and severe categories. The dataset used in our research consists of almost 3000 images.
- Finally, to comprehend the idea of transfer learning and conduct a comparative comparison of various COVID-19 severity estimating models present in the literature.

## Methods

### Dataset description

The dataset that we have used for training, testing, and validating our model is collected from Kaggle (Dataset 2021a). The mentioned dataset is compiled from seven different sources (Dataset 2021b; Winther et al. 2020; of Medical and Database 2021; Haghanifar et al. 2020;

Chowdhury et al. 2020; Cohen et al. 2020; Resources 2021). This COVID-19 dataset is being made available in two different stages. It contains a total of 3616 COVID-19 CXR images. The dataset can be obtained from Dataset (2021a). The two most important reasons for using the above dataset are mentioned below:

- Deep learning models are data hungry. This repository contains about 3616 COVID-19 positive CXR images, one of the largest public dataset available.
- The above dataset is made by combining images from seven different sources. So basically the model trained using the above dataset cannot be biased towards a particular type of images.

### Dataset preparation

After collecting the dataset from Kaggle (Dataset 2021a), we have asked our experts at Netaji Subhas Medical College and Hospital (NSMCH), Bihta, Patna Medical College and Hospital (PMCH), Patna and Mahatma Gandhi Memorial Medical College (MGM), Indore to manually classify the dataset into three levels of severity, namely **Mild, Moderate, and Severe**. The experts at all the Hospitals have more than 10 years of experience in Radiology. All anonymized and randomized CXRs collected from Kaggle were reviewed by all the teams independently and blindly. The readers used a semiquantitative severity score to rate pulmonary parenchymal involvement. For that the experts divided each lung into three different parts as shown in Fig. 2, upper part (from the apex of the lung to the aortic arch profile), mid part (from the aortic arch profile to the lower margin of the left pulmonary hilum), and finally the lower part (from the aortic arch profile to the lower margin of the left pulmonary hilum). The experts then gave a score on a scale of zero to three in one-point increments: 0 when they found normal lung parenchyma; 1 when they found interstitial involvement only; 2 when the experts found the presence of radiopacity for less than 50% of visible



(a)



(b)



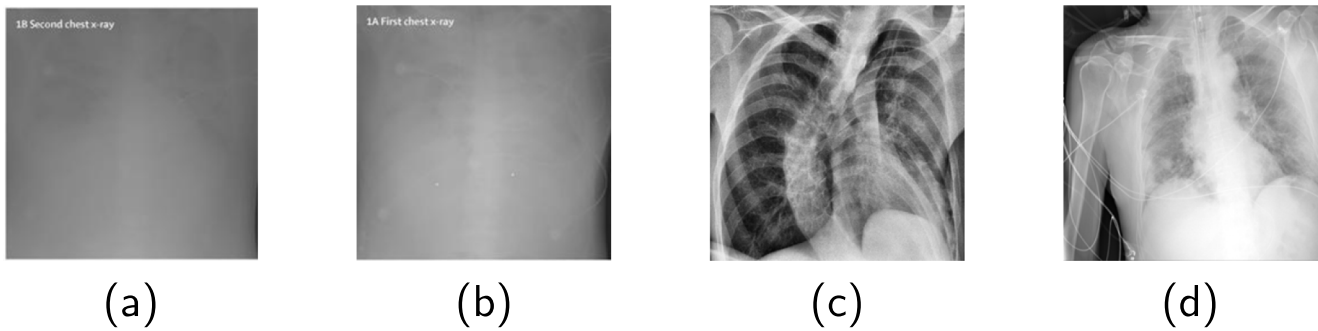
(c)



(d)

**Fig. 3** (a) Nothing visible, difficult to comment (COVID-698.png) (b) Nothing visible, difficult to comment (COVID-1523.png) (c) Rotated film, difficult to comment (COVID-1512.png) (d) Poor exposure,

bony markings and underlying tissues not clearly visible, difficult to comment (COVID-1613.png)



**Fig. 4** (a) Looks normal with no signs of COVID-19 infection, difficult to comment (COVID-698.png) (b) Looks normal but bilateral Cardiophrenic angle is not visualize in x-ray (COVID-1523.png) (c)

Looks normal with no signs of COVID-19 infection, difficult to comment (COVID-1512.png) (d) Confusion over inclusion in mild or normal (objective assessment bias) (COVID-1613.png)

lung parenchyma; and finally 3 when the experts found the presence of radiopacity for 50% or more than 50% of visible lung parenchyma (Monaco et al. 2020). Interestingly the experts at NSMCH could not comment on few images because of unclear CXR images or poor exposure. Figure 3 shows few images, along with their filenames which could not be classified into Mild, Moderate or Severe category. Similarly the experts at MGM could not comment on few images citing various reasons as mentioned in Fig. 4. For preparing the final dataset, we have discarded all the ambiguous images.

The experts at NSMCH and PMCH classified almost 2800 images into mild, moderate, and severe and similarly the experts at MGM classified almost 3200 images into the said classes. Finally the images which were classified as mild by all the experts were considered Mild. Similarly the images that were classified as moderate and severe by all the experts were considered as Moderate and Severe. Table 1 shows the total images taken for training, testing and validation of the proposed model. Figures 5, 6, and 7 show few sample images along with their filename considered in Mild, Moderate and Severe category respectively.

**Data pre-processing**

As mentioned in the “Dataset description” section, the dataset used for training, testing and validating our proposed model is made up by combining COVID-19 positive CXR images from seven different sources. Different sources

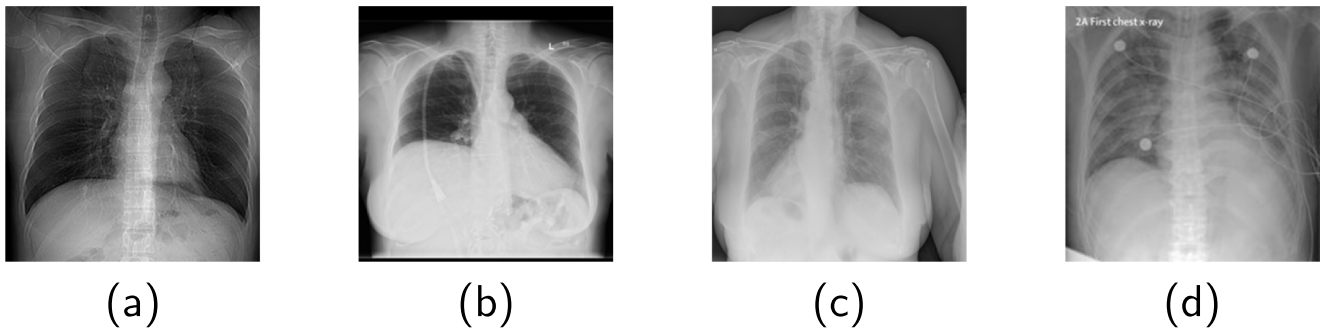
indicate different Hospitals from where the dataset is collected. There will be significant difference between the images collected from two different sources as the CXR images have been taken using different Chest X-Ray machine, which in turn have different specifications. So, to remove all those differences Zero mean normalisation is applied. Zero mean Normalization converges the model at a faster rate as it makes all the pixel values of the input images between 0 and 1 (Li et al. 2019; Deb et al. 2022). A sample image before and after image pre-processing is shown in Fig. 8.

**Transfer learning**

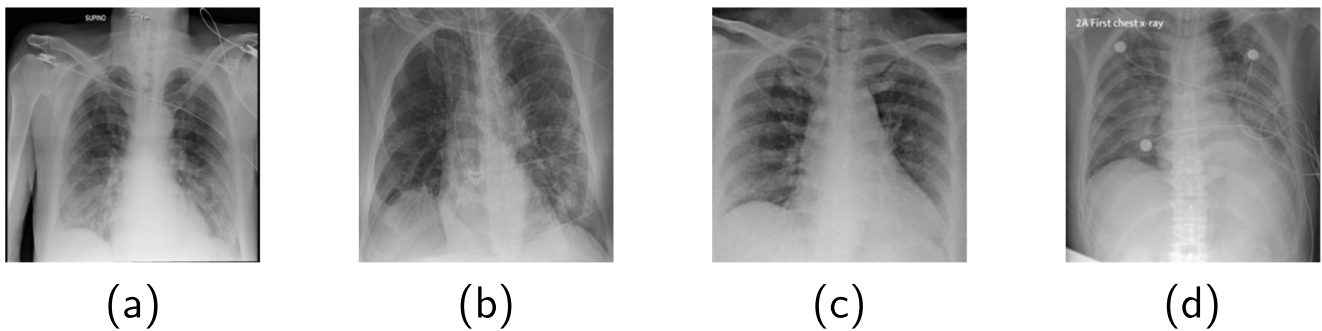
As shown in Table 1, the total number of images in the dataset is 2295, and out of them 1580 images are used for training. With merely 1580 images left for training, Transfer Learning Scheme for feature extraction is explored. Transfer Learning is the practice of reusing a network trained for a completely different purpose on a much larger dataset than we need. As seen in the diagram, Fig. 9, the proposed CoVSeverity-Net is an ensemble of two pre-trained DCNN structures, namely VGG-19 (Simonyan and Zisserman 2014) and MobileNet (Howard et al. 2017). Both the networks are trained on ImageNet dataset (Deng et al. 2009). The model proposed is precisely shown in **Block A** of Fig. 9, the details about the same are given in Fig. 10. And the final classification layer (**Block B**) is given in Fig. 11. Both Figs. 10 and 11 are generated using Netron App (NETRON 2021).

**Table 1** The number samples used for training, testing and validation

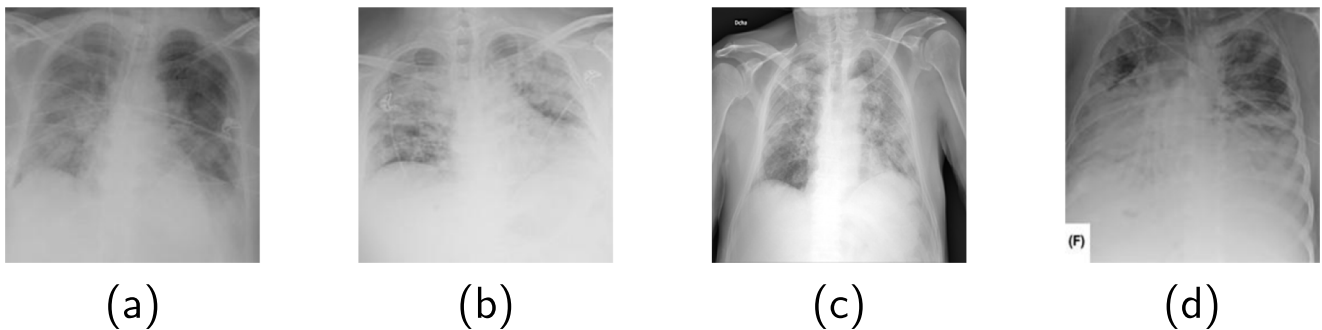
Classes	Number of images				
	# of images	Total	Training Set	Validation Set	Test Set
Mild	535	2295	1580	400	315
Moderate	920				
Severe	840				



**Fig. 5** Few images which are classified as **Mild** by both the experts independently (a) COVID-118.png (b) COVID-467.png (c) COVID-578.png (d) COVID-783.png

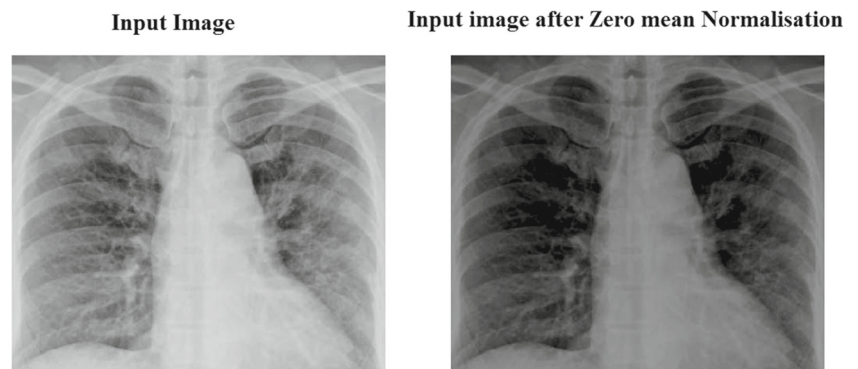


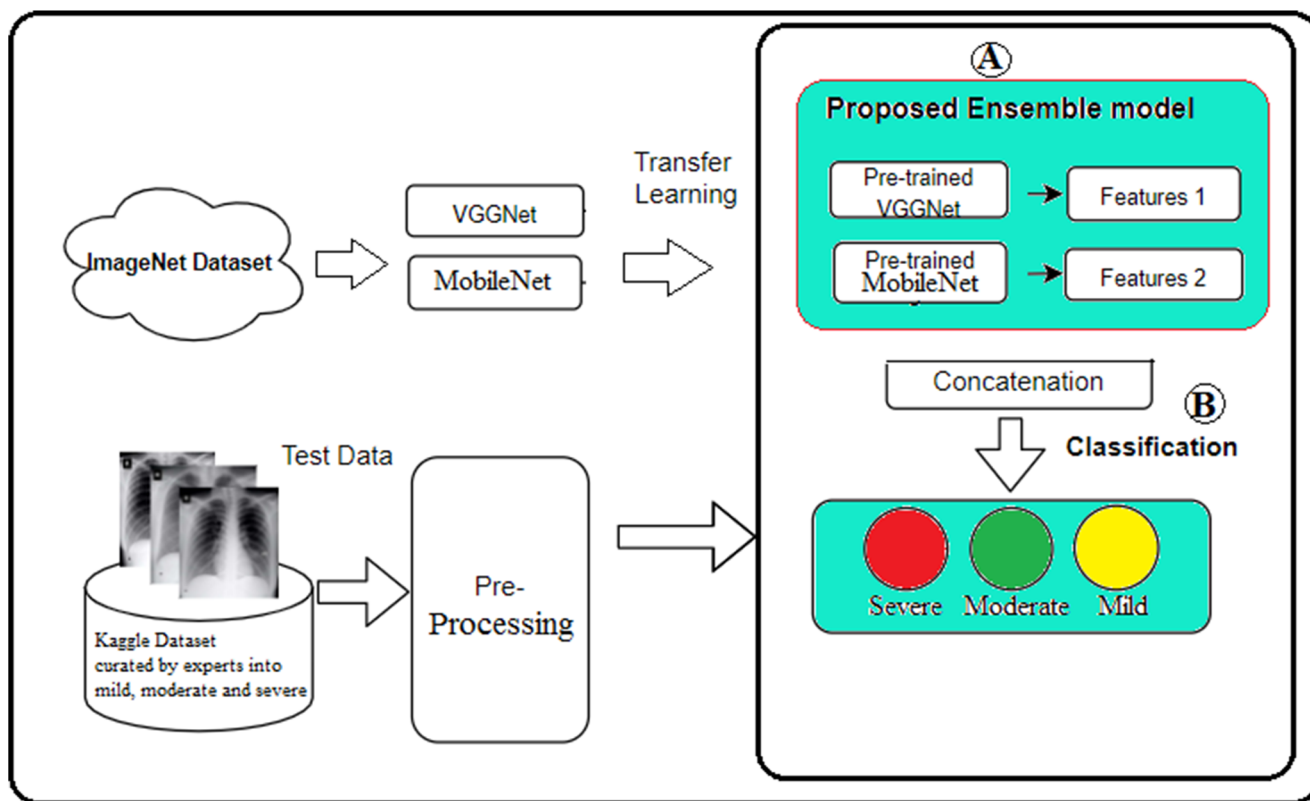
**Fig. 6** Few images which are classified as **Moderate** by both the experts independently (a) COVID-25.png (b) COVID-253.png (c) COVID-372.png (d) COVID-376.png



**Fig. 7** Few images which are classified as **Severe** by both the experts independently (a) COVID-414.png (b) COVID-216.png (c) COVID-359.png (d) COVID-452.png

**Fig. 8** A sample COVID-19 positive CXR image from the Kaggle dataset before and after pre-processing





**Fig. 9** Block diagram of the work flow. Low-level features from the input CXR images extracted using CoVSeverity-Net, which is an ensemble of two pre-trained networks, namely VGGNet and

MobileNet. The features are concatenated before final classification. Details about block (A) is given in Fig. 10 and details about block (B) is given in Fig. 11

### CoVSeverity-Net: Its architecture and development

The block diagram of the proposed CoVSeverity-Net is shown in Fig. 9. The network proposed is named CoVSeverity-Net, as it estimates the severity of the COVID-19 patients from the input Chest X-ray images and is based on Convolutional Neural Net.

As shown, in Fig. 9, low-level features are extracted using an ensemble of two pre-trained networks from the input CXR images. After independently extracting the features, they are passed through a *Global Average Pooling layer*. This is done to significantly reduce the length of the features extracted from the ensemble model (Lin et al. 2013). The length of the feature vectors collected from the pre-trained VGG-19 and MobileNet models were 512 and 1024, respectively. For both, the pre-trained networks, the final 1000-class classification layers designed for ImageNet classification were chopped off, and a classification layer, as shown in Fig. 11, is attached. As shown in the block diagram, the input size of the classification layer, which has dense connections of 256 nodes, is 1536 (512 + 1024). Finally, before the softmax layer, it has a ReLU activation function. A dropout with a probability of 0.5 is used. Dropout is a technique for preventing overfitting in any deep

learning model (Srivastava et al. 2014). Details about the length of the feature are given in Table 2. Later on, the features are concatenated before classification. The TSNE-plot of the features extracted is shown in Fig. 12. The classification layer is shown in Fig. 11.

### Implementing and training

The entire COVID-19 positive CXR image dataset is obtained from Kaggle (Dataset 2021a), which itself is formed by combining seven different datasets (Dataset 2021b; Winther et al. 2020; of Medical and Database 2021; Haghanifar et al. 2020; Chowdhury et al. 2020; Cohen et al. 2020; Resources 2021). The dataset is independently curated into mild, moderate, and severe categories by two expert groups from NSMCH, PMCH, and MGM Hospitals. The dataset is partitioned into train, test and validation as shown in Table 1. All the images were resized to 224 X 224 X 3 (Table 3). The entire experiment was carried out using the Google Colab platform, which is equipped with a Tesla K80 graphics card, and was built in Python using the Keras package with Tensor-flow as a backend. The studies were carried out utilizing the Adam optimizer and a categorical smooth loss function with a learning rate of  $10^{-5}$ . All the

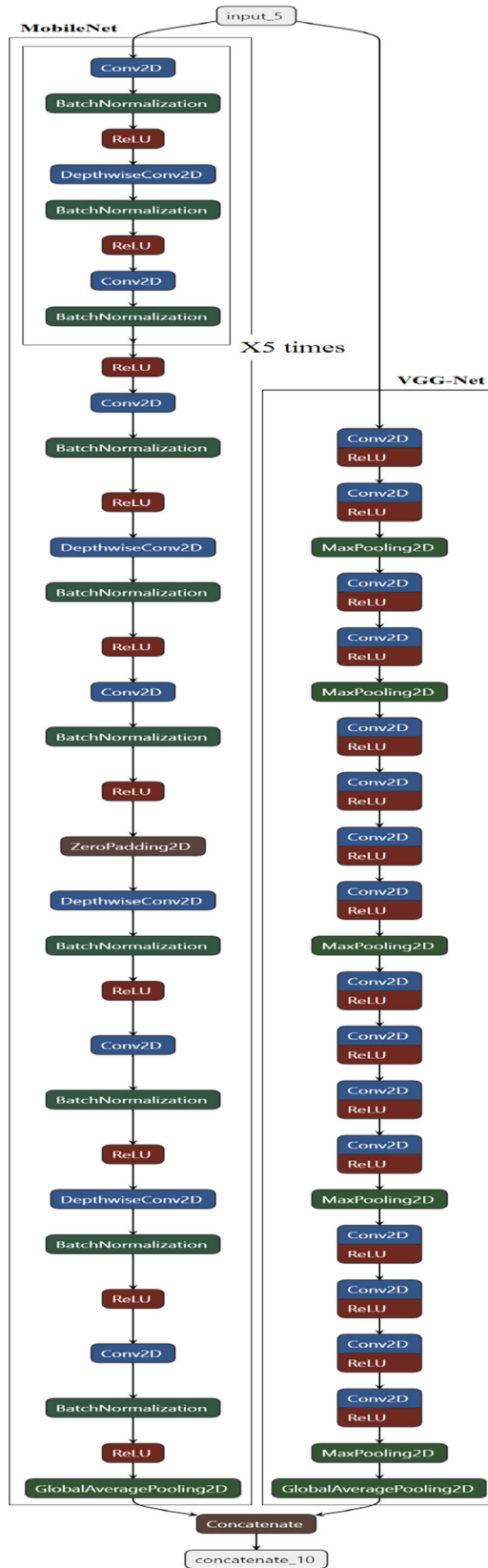


Fig. 10 A schematic diagram of CoVSeverity-Net, the model extracts low-level features from input CXR images using MobileNet (left) and VGG-19 net (right)

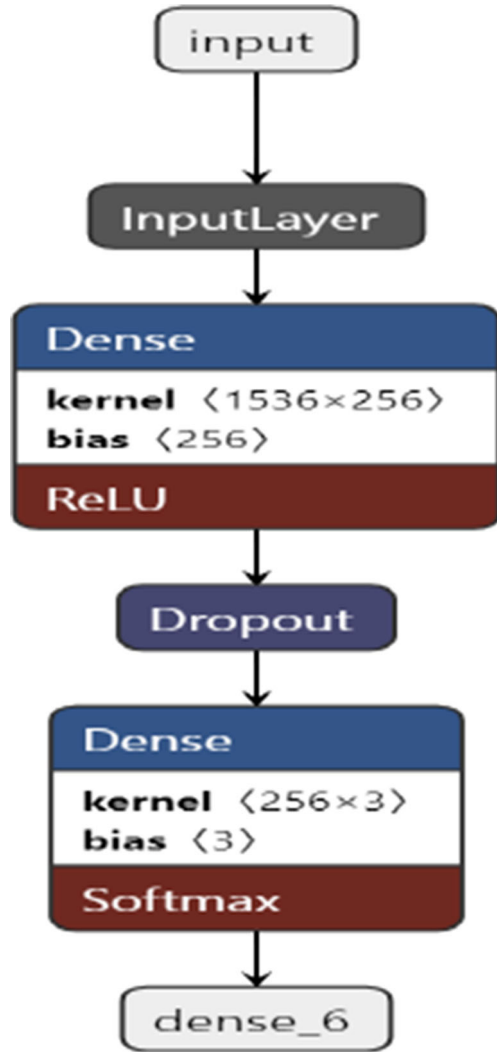


Fig. 11 The final Classification Layer

experiments were executed up to 100 epochs. A batch size of 32 is considered for performing the experiments.

### Evaluation matrices

Accuracy, Precision, Recall, and F1 score are used to evaluate classification performance. The formula is as follows, with True Positive, False Positive, False Negative, and True Negative having their typical meanings.

$$Accuracy = \frac{T.P + T.N}{T.P + T.N + F.P + F.N} \quad (1)$$

$$Recall = \frac{T.P}{T.P + F.N} \quad (2)$$

$$Precision = \frac{T.P}{T.P + F.P} \quad (3)$$

$$F1Score = \frac{2T.P}{2T.P + F.N + F.P} \quad (4)$$



**Table 2** Details of the proposed CoVSeverity-Net

Layer (type)	Shape of the Output	Number of parameter
VGG (DCNNmodel)	512	20,035,963
MobileNet (DCNN model)	1024	398,688
concatenate_6 (Concatenate)	1536	0
dense(Dense)	256	1,417,216
ReLU	256	0
Dropout	256	0
dense_6(Dense)	3	768
Softmax(Classification)		
Total Parameters		21,842,595
Parameters trained		1,417,985

## Results

The proposed **CoVSeverity-Net** achieves an overall accuracy of 85.23% as given in Table 4. In the same table, accuracy comparisons are made between individual DCNNs and other available pre-trained networks. As can be seen, VGG-19 provides the best accuracy, followed by MobileNet. An accuracy of 83.26% and 82.42% is obtained using VGG-19 and MobileNet, respectively. Having VGG-19 and MobileNet as an ensemble is because the individual performance is the best amongst all the networks considered for comparison. Table 4 also indicates that using an ensemble approach improves the classification performance. To demonstrate the importance of data pre-processing, the proposed model's performance is compared on the curated dataset with and without pre-processing processes (as mentioned in section “Data pre-processing”). Without the pre-processing processes, the accuracy is reported to be 78.71% (Table 5). In terms of accuracy, zero-mean-normalization shows a considerable increase of 8.89%.

Precision, Recall, and F1 scores are considered the top metrics for evaluating any classification performance. The measures for the *Mild*, *Moderate*, and *Severe* classes are presented in Table 6 and finally the training progress is presented in Fig. 13a. As shown in Fig. 13a, the model is executed up to 100 epochs. However, if carefully observed, the model saturates after the 65th epoch. Figure 13b shows the confusion matrix on the test data. Figure 13c and d show the AU-ROC curve and its zoomed version of the multi-class classification. In the case of multi-class classification, the AU-ROC curve is drawn based on the one-vs-all technique. As shown in Fig. 13c, the AU-ROC value of class severe is maximum, followed by moderate and mild. For severe class, the value is 0.97. As shown in Table 6, for Mild class, both precision and recall of 90% are reported. That is, 90% of the images classified as Severe class was correctly

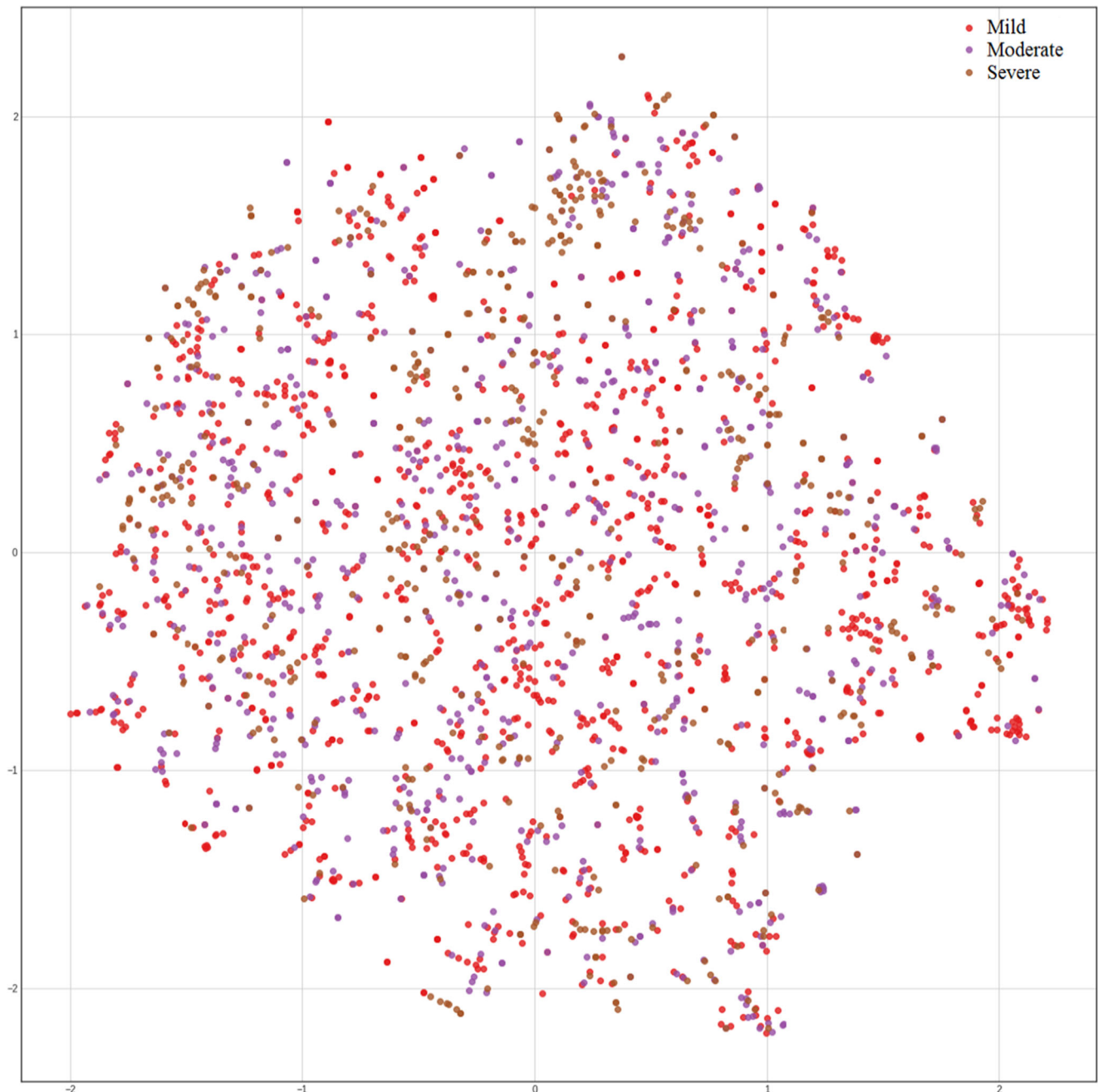
classified, and similarly, 90% of all the severe class images in the test set were correctly classified. For the Moderate class, the precision and recall obtained were 83% and 86%, respectively. Similarly, the precision and recall of 84% and 78% were obtained for the Severe class. Moreover, as shown in Fig. 13c, the class-wise AU-ROC of 0.95, 0.90, and 0.97 were obtained for mild, moderate, and severe classes, respectively.

## Discussions

After minutely studying the literature, we found that several researchers have proposed or published various models based on deep learning algorithms for detecting COVID-19 from CXR images. However, very few works have been done on estimating the severity of the disease from the CXR images; this might be because of the lack of a large-scale dataset. Covid-NetS, developed by Wong et al. (2021), is a regression model used to predict the severity score of a COVID-19-positive patient by looking at their Chest X-Ray (CXR) images. The severity score is based on geographic extent and opacity extent. Geographic extent is based on the consolidation of the lung image, and opacity extent calculates the degree of opacity. CoVSeverity-Net, proposed by us, is a classification model that classifies the input CXR images into three categories of severity. The advantages of the CoVSeverity-Net over Covid-NetS are as follows.

- CoVSeverity-Net is an entirely automated system, whereas Covid-NetS need manual intervention as the artifacts and the patient metadata from the CXRs are to be removed (Wong et al. 2021).
- CoVSeverity-Net is trained on almost 3000 CXR images, whereas Covid-NetS is trained using 396 images. Moreover, the dataset used to train the proposed CoVSeverity-Net is collected from seven different sources. On the contrary, the dataset used to train Covid-NetS is collected from a single source. Thus we can say that the model proposed by us is robust as compared to the one proposed by Wong et al.
- CoVSeverity-Net reported a classification accuracy of 85.71%, whereas the best-performing Covid-NetS reported an R2 score of 0.739 and 0.741 for geographic and opacity extent, respectively.

In this manuscript, we have collected a publicly available large-scale dataset and asked expert radiologists to curate the dataset based on three levels of severity. The experts divided the lung area into six parts and, after carefully studying them, divided them into mild, moderate, and severe. Later, we proposed a deep learning-based architecture called CoVSeverity-Net for classifying the images into three levels of severity. The dataset's



**Fig. 12** TSNE-plot of the features extracted from the input CXR images using CoVSeverity-Net

details are given in the “[Dataset description](#)” section. Few images were excluded and removed from the final dataset. Valid reasons for the same are provided in the same subsection. Deep learning algorithms are generally data hungry and thus require much data for training. The dataset that we have created is not that large. So the idea of training a Convolutional Neural Network from scratch is not considered. Instead, we have used pre-trained networks for feature extraction. To finalize the feature extractors, we evaluated the performance of all the available pre-trained

networks on the dataset we prepared. The details about the same are given in Table 4. The proposed CoVSeverity-Net is an ensemble of two pre-trained networks, namely VGGNet and MobileNet. The ensemble is considered for feature extraction as ensembling is always better than single architecture, and secondly, extracting features using pre-trained networks does not require any additional parameters to be trained. The proposed CoVSeverity-Net achieved an accuracy of 85.71%. Cross-validation is a re-sampling procedure used to evaluate machine learning models on a

**Table 3** Table containing hyperparameters of the proposed CoVSeverity-Net

Hyperparameters	Values
Input Shape	224 X 224 X 3
Input to the classifier	1536
Number of layers in the Classifier	2
Activation Function	ReLU, Softmax (Final Classification Layer)
Optimizer	Adam
Adam Hyper parameters	$\beta_1 = 0.7, \beta_2 = 0.8$
Learning rate	0.00001
Epochs	100
Batch-size	32

**Table 4** Performance of the severity estimator with the individual DCNN models

Architecture	Accuracy
VGGNet	83.26%
MobileNet	82.42%
InceptionV3	80.33%
DenseNet169	77.54%
NASNetSmall	75.56%
ResNet	65.98%
CoVSeverity-Net	85.71%

**Table 5** Performance Comparison With/Without pre-processing

Architecture (With/w/o Pre-processing)	Accuracy
CoVSeverity-Net (w/o Pre-processing)	78.71%
CoVSeverity-Net (With Pre-processing)	85.71%

**Table 6** Class-wise performance measures

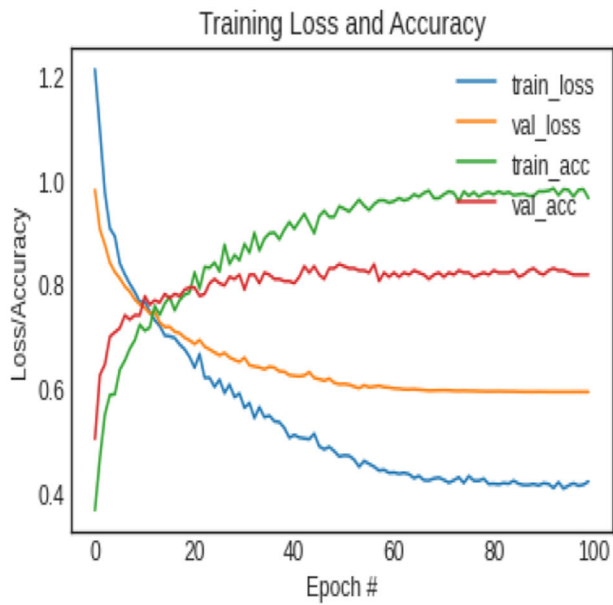
Class	Precision (%)	Recall (%)	F1-Score (%)
Mild	90	90	90
Moderate	83	86	84
Severe	84	78	81

limited data sample. The procedure has a single parameter called *k* that refers to the number of groups that a given data sample is to be split into. Conducting a 5-fold cross-validation accuracy of  $87.82 \pm 6.25$  is reported. Similarly, conducting 10-fold cross-validation we obtained accuracy of  $91.26 \pm 3.42$ . The class-wise AUC is given in Fig. 13c. As shown in the figure, AU-ROC for mild class is 0.95, for moderate it is 0.90 and finally for severe class an AU-ROC of 0.97 is obtained. The limitation of the work presented is that the proposed algorithm is not very accurate. The accuracy achieved by the CoVSeverity-Net is a little above 85%. This might be because the pre-trained networks we have used are trained on the ImageNet dataset, which only contains natural images.

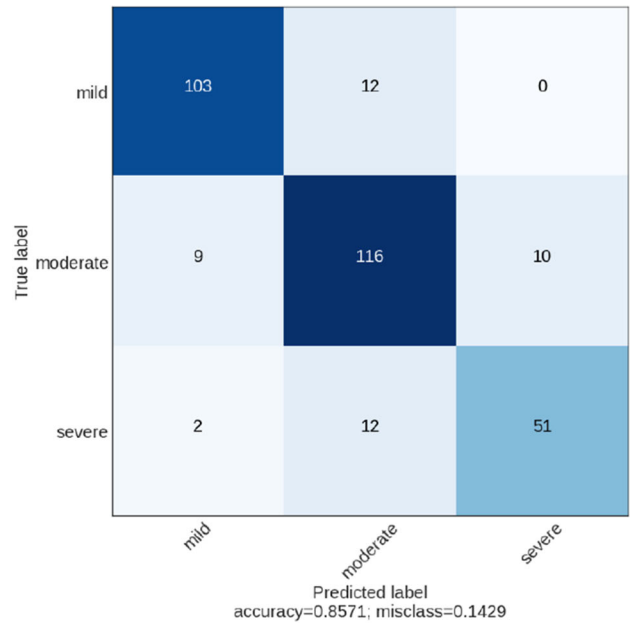
However, considering the rapid research in biomedical image processing using deep learning, we firmly believe that researchers will create a large dataset. An extensively annotated dataset can be used to train a Convolutional Neural Network from scratch.

## Conclusions

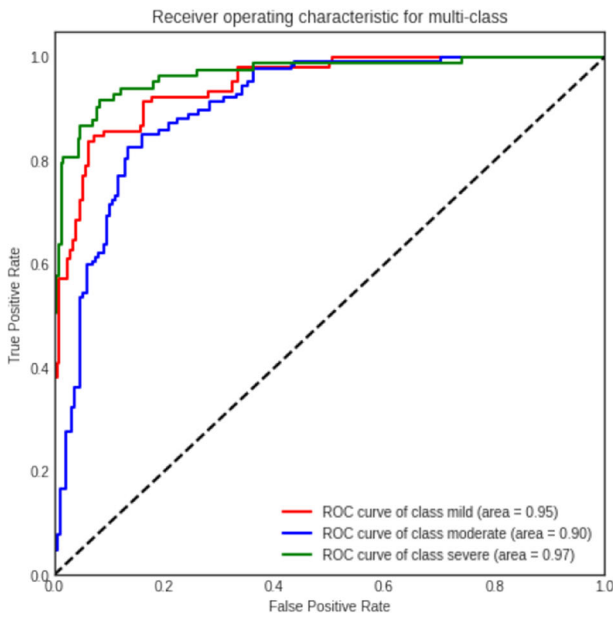
Much research has been reported to detect COVID-19 from various image modalities like CXR images or CT-Scans, but very few algorithms have been proposed for severity estimation. Not all COVID-19 positive patients require intensive care or hospitalization. Severity estimation, if developed properly, can help patients determine their future course of action. For that we have proposed CoVSeverity-Net, a deep learning-based severity estimator. CXR images are selected as X-Ray machines are available at almost every primary hospital, and an X-Ray can be done at a very cheap rate. The large COVID-19 positive CXR from Kaggle is selected as it contains images from seven different sources. Later on, the large dataset of almost 3500 images was curated independently by two expert groups. The images that were of low quality and is having some discrepancies were removed from the dataset. Proper explanation for the same were also included in the manuscript. CoVSeverity-Net uses two pre-trained DCNN structures, namely VGG-19 and MobileNet, for feature extraction. The final features extracted were classified into three levels of severity, namely, Mild, Moderate and Severe. The model achieved an accuracy of 85.71%. Conducting a 5-fold cross-validation accuracy of  $87.82 \pm 6.25$  is reported. Similarly, conducting 10-fold cross-validation we obtained accuracy of  $91.26 \pm 3.42$ . We strongly believe that this study has a high chance of reducing the workload of overworked front-line radiologists, speeding up patient diagnosis and treatment, and easing pandemic control.



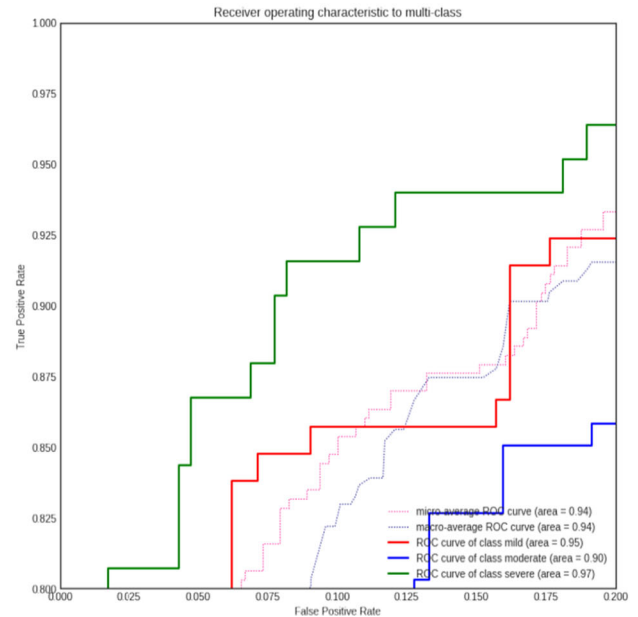
(a) Training progress



(b) Confusion Matrix



(c) Class-Wise AU-ROC



(d) Left corner of AU-ROC Curve

Fig. 13 The figure showing the training progress, confusion matrix and the AUROC curve

Future work would be to train a novel deep learning-based architecture on a larger dataset for severity estimation and also to collect images from local hospitals to test our model on.

We have posted the Dataset and the algorithm on our github repositories. The same can be obtained from [https://github.com/sagardeepdeb/covid\\_severity](https://github.com/sagardeepdeb/covid_severity)

## Declarations

The authors declare no competing interests.

## References

- Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*. 2020;296(2):E32–40.
- Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR, Khan MS, Iqbal A, Al Emadi N, et al. Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access*. 2020;8:132665–76.
- Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, Cui J, Xu W, Yang Y, Fayad ZA, et al. CT imaging features of 2019 novel Coronavirus (2019-nCoV). *Radiology*. 2020;295(1):202–7.
- Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. COVID-19 image data collection: Prospective predictions are the future. *arXiv:2006.11988*. 2020.
- Dataset KR. Kaggle radiography dataset. 2021.
- Dataset TC-C. Twitter COVID-19 CXR dataset. 2021.
- Deb SD, Jha RK. COVID-19 detection from chest X-ray images using ensemble of CNN models. In: 2020 international conference on power, instrumentation, control and computing (PICCC). IEEE; 2020. p. 1–5.
- Deb SD, Jha RK, Jha K, Tripathi PS. A multi model ensemble based deep convolution neural network structure for detection of COVID19. *Biomed Sig Process Control*. 2022;71:103126.
- Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE; 2009. p. 248–55.
- Fukushima K, Miyake S, Ito T. Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE Trans Syst Man Cybern*. 1983;5:826–34.
- Haghanifar A, Majdabadi MM, Choi Y, Deivalakshmi S, Ko S. COVID-CXNET: Detecting COVID-19 in frontal chest X-ray images using deep learning. *arXiv:2006.13807*. 2020.
- He K, Zhao W, Xie X, Ji W, Liu M, Tang Z, Shi Y, Shi F, Gao Y, Liu J, et al. Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images. *Pattern recognit*. 2021;113:107828.
- Hemdan EE-D, Shouman MA, Karar ME. COVIDX-NET: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images. *arXiv:2003.11055*. 2020.
- Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv:1704.04861*. 2017.
- Irmak E. COVID-19 disease severity assessment using CNN model. *IET Image Process*. 2021;15(8):1814–24.
- 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.
- Kanne JP, Little BP, Chung JH, Elicker BM, Ketaj LH. Essentials for radiologists on COVID-19: An update—radiology scientific expert panel. 2020.
- Khan AI, Shah JL, Bhat MM. Coronet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images. *Comput Methods Programs Biomed*. 2020;196:105581.
- Kim YY, Shin HJ, Kim M-J, Lee M-J. Comparison of effective radiation doses from X-ray, CT, and PET/CT in pediatric patients with neuroblastoma using a dose monitoring program. *Diagn Interv Radiol*. 2016;22(4):390.
- Kucharski AJ, Klepac P, Conlan AJ, Kissler SM, Tang ML, Fry H, Gog JR, Edmunds WJ, Emery JC, Medley G, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *Lancet Infect Dis*. 2020;20(10):1151–60.
- Li H, Zhuang S, Li D-A, Zhao J, Ma Y. Benign and malignant classification of mammogram images based on deep learning. *Biomed Sig Process Control*. 2019;51:347–54.
- Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*. 2020a.
- Li Y, Xia L. Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management. *Am J Roentgenol*. 2020;214(6):1280–6.
- Li Z, Zhong Z, Li Y, Zhang T, Gao L, Jin D, Sun Y, Ye X, Yu L, Hu Z, et al. From community-acquired pneumonia to COVID-19: a deep learning-based method for quantitative analysis of COVID-19 on thick-section CT scans. *Eur Radiol*. 2020b;30(12):6828–37.
- Lin M, Chen Q, Yan S. Network in network. *arXiv:1312.4400*. 2013.
- Maghded HS, Ghafoor KZ, Sadiq AS, Curran K, Rawat DB, Rabie K. A novel AI-enabled framework to diagnose coronavirus COVID-19 using smartphone embedded sensors: design study. In: 2020 IEEE 21st international conference on information reuse and integration for data science (IRI). IEEE; 2020. p. 180–7.
- Monaco CG, Zaottini F, Schiaffino S, Villa A, Della Pepa G, Carbonaro LA, Menicagli L, Cozzi A, Carriero S, Arpaia F, et al. Chest X-ray severity score in COVID-19 patients on emergency department admission: a two-centre study. *Eur Radiol Exp*. 2020;4(1):1–7.
- NETRON. Netron app. 2021.
- of Medical IS, Database IRC. Italian society of medical and interventional radiology COVID-19 database. 2021.
- Polsinelli M, Cinque L, Placidi G. A light CNN for detecting COVID-19 from CT scans of the chest. *Pattern Recognit Lett*. 2020;140:95–100.
- Ranjan A, Lalwani D, Misra R. GAN for synthesizing CT from T2-weighted MRI data towards MR-guided radiation treatment. *Magn Resonan Mater Phys Biol Med*, 1–9. 2021.
- Ranjan A, Shukla S, Datta D, Misra R. Generating novel molecule for target protein (SARS-CoV-2) using drug–target interaction based on graph neural network. *Netw Model Anal Health Inform Bioinforma*. 2022;11(1):1–11.
- Resources C. COVID-19 resources. 2021.
- Roser M, Ritchie H, Ortiz-Ospina E, Hasell J. Coronavirus pandemic (COVID-19). Our world in data. 2020.
- Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, Schluger NW, Volpi A, Yim J-J, Martin IB, 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(1):172–80.
- Sethy PK, Behera SK. Detection of coronavirus disease (COVID-19) based on deep features. 2020.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*. 2014.
- Singh D, Kumar V, Kaur M, et al. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks. *Eur J Clin Microbiol Infect Dis*. 2020;39(7):1379–89.

- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15(1):1929–58.
- Tang Z, Zhao W, Xie X, Zhong Z, Shi F, Liu J, Shen D. Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images. *arXiv:2003.11988.* 2020.
- 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(1):1–12.
- Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Cai M, Yang J, Li Y, Meng X, et al. A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *Eur Radiol.* 1–9. 2021.
- Winther H, Laser H, Gerbel S, Maschke S, Hinrichs J, Vogel-Claussen J, Wacker F, Höper M, Meyer B. COVID-19 image repository. *Figshare (Dataset).* 2020.
- Wong A, Lin Z, Wang L, Chung A, Shen B, Abbasi A, Hoshmand-Kochi M, Duong T. 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(1):1–8.
- Zhou S, Wang Y, Zhu T, Xia L. CT features of coronavirus disease 2019 (COVID-19) pneumonia in 62 patients in Wuhan, China. *Am J Roentgenol.* 2020;214(6):1287–94.
- Zhu J, Shen B, Abbasi A, Hoshmand-Kochi M, Li H, Duong TQ. Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. *PLoS ONE.* 2020;15(7):e0236621.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Affiliations

Sagar Deep Deb<sup>1</sup>  · Rajib Kumar Jha<sup>1</sup> · Rajnish Kumar<sup>2</sup> · Prem S. Tripathi<sup>3</sup> · Yash Talera<sup>3</sup> · Manish Kumar<sup>4</sup>

Rajib Kumar Jha  
jharajib@iitp.ac.in

Rajnish Kumar  
doctor.rajnish.kumar@gmail.com

Prem S. Tripathi  
drpstripathi@gmail.com

Yash Talera  
yashktalera@gmail.com

Manish Kumar  
medico.manish@gmail.com

- <sup>1</sup> Department of Electrical Engineering, Indian Institute of Technology Patna, Patna, 801103, India
- <sup>2</sup> Department of Paediatrics, Netaji Subhas Medical College & Hospital, Patna, 801106, India
- <sup>3</sup> Department of Radiodiagnosis, Mahatma Gandhi Memorial Government Medical College, Indore, 452001, India
- <sup>4</sup> Patna Medical College and Hospital, Bihar, 800001, India