

# Quantifying prognosis severity of COVID-19 patients from deep learning based analysis of CT chest images

Ashish Rana<sup>1</sup> • Harpreet Singh<sup>1</sup> <sup>(D)</sup> • Ravimohan Mavuduru<sup>2</sup> • Smita Pattanaik<sup>2</sup> • Prashant Singh Rana<sup>1</sup>

Received: 17 February 2021 / Revised: 4 January 2022 / Accepted: 10 January 2022 / Published online: 8 March 2022 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

# Abstract

The COVID-19 pandemic has affected all the countries in the world with its droplet spread mode. The colossal amount of cases has strained all the healthcare systems due to the serious nature of infections especially for people with comorbidities. A very high specificity Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) test is the principal technique in use for diagnosing the COVID-19 patients. Also, CT scans have helped medical professionals in patient severity estimation & progression tracking of COVID-19 virus. In study we present our own extensible COVID-19 viral infection tracking prognosis technique. It uses annotated dataset of CT chest scan slice images created with the help of medical professionals. The annotated dataset contains bounding box coordinates of different features for COVID-19 detection like ground glass opacities, crazy paving pattern, consolidations, lesions etc. We qualitatively identify the severity of the patient for later prognosis stages in our study to assist medical staff for patient prioritization. First we detected COVID-19 positive patients with pre-trained Siamese Neural Network (SNN) which obtained 87.6% accuracy, 87.1% F1-Score & 95.1% AUC scores. These metrics were achieved after removal of 40% quantitatively highly similar images from the COVID-CT dataset. This reduced dataset was further medically annotated with COVID-19 features for bounding box detection. After this we assigned severity scores to detected COVID-19 features and calculated the cumulative severity score for COVID-19 patients. For qualitative patient prioritization with prognosis clinical assistance information, we finally converted this score into a multi-classification problem which obtained 47% weighted-average F1-score.

Keywords COVID-19 detection  $\cdot$  COVID-19 prognosis  $\cdot$  Single shot detection network  $\cdot$  Siamese neural network

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

Harpreet Singh akalharpreet@gmail.com

#### 1 Introduction

There are a handful of coronavirus strains known that cause infections in humans [41]. For example, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) coronavirus strains. The former originated in China and resulted in 774 deaths out of 8,098 infections and the latter originated in Saudi Arabia liable for 848 deaths among 2,458 infected individuals [2, 22]. Severe Acute Respiratory SyndromeCoronavirus 2 (SARS-Cov-2) virus is responsible for causing COVID-19. This virus originated in China and in the matter of months was declared a global pandemic by the World Health Organization (WHO). Respiratory transmission and physical contact were the prime reason for its widespread infection base. Also, the fact that 60% of total infected subjects demonstrated either mild or no symptoms. Because of which they ended up infecting others by acting as carriers and made this disease even harder to control [3].

Fever, dry cough and dyspnoea are symptoms of the most common symptoms of the infection whereas in more critical cases it can lead to septic shock, organ failure, pneumonia and even death [14, 17, 40]. With a huge number of cases raking in at such a short period of time, there is no doubt that healthcare of all countries are strained up to the maximum limit if not broken. RT-PCR is used for diagnosing the patients that are suffering from COVID-19. With its improved sensitivity and very high specificity this is the principal technique of choice [36]. Along with that CT and X-ray images have served as a supplement tool for prognosis with vast consensus regarding the features of COVID-19 as the virus progresses in on a given patient [4, 11, 30]. CT chest scans clearly overpowers X-ray scans with their high end image quality and capabilities to detect COVID-19 related features relatively early.

These COVID-19 lung disorder features definitely serve as an indicator for analyzing severity of the patient. With ground glass opacity, a type of opacity that appears in initial stages of contraction of COVID-19 by a patient. Later, it mostly transforms into chunks of consolidation patches or crazy paving patterns even accompanied by lesions causing more serious respiration issues and probable requirement for breathing aid. But, these imaging methods do not have reliable specificity as the features of COVID-19 do overlap with a large number of respiratory diseases like influenza A & B, pneumonia & HSV pneumonia, atypical bacterial pneumonia, interstitial lung disease, pulmonary edema etc. This indicates that diagnosis with reliable certainty may be lacking for COVID-19 from radiological findings only. With this argument we focus our efforts on evaluating the severity of a COVID-19 patient after the diagnosis of the disease is done with RT-PCR.

This paper is aimed at expounding on a severity estimation approach of COVID-19 patients which leverages the feature detection and localization in CT chest scan slice images. Based on the proposed severity metric we can assist medical professionals to prioritize the patients for the required vital resources. Which would in turn help to allocate resources with maximum utilization and cater to the maximum number of patients. We specifically focus on detection and localization of ground glass opacities, crazy paving pattern, reverse halo effect, vascular dilatation, subpleural band, bronchiectasis, traction bronchiectasis, consolidations and lesions. Based on these features we classify the severity of the patient. An earlier clinical study also uses a severity estimation approach [38]. It is done by manually accumulating scores over the different segment regions of lungs in a CT scan image slice instead of directly focusing on detecting COVID-19 features. But, this manual analysis from medical experts consumes vital human resources which is not good for medical facilities under stress while handling large numbers of patients.

We also tackle the issue of availability of quality dataset annotations for both CT chest scan images and provide medically validated annotations for CT chest scan dataset [39].

The annotations provided in the dataset are reviewed by medical professionals for creating good quality annotation dataset. Our data is biased towards few COVID-19 features because of their frequent recurrence with other COVID-19 features in the dataset. Hence, our model performance is also skewed towards predicting better results for those features. Also, we acknowledge the imbalanced & biased nature of data and the negative effects on the predictions caused by it [6]. We try to limit these effects by using few shot learning techniques with hyperparameter optimizations and data augmentation that deliver best results out of smaller datasets. The CT chest scan dataset contains images of the same patient multiple times with different slices only reiterating similar COVID-19 features for training. Hence, we take a subset of training dataset images that captures most variance amongst the whole dataset and performs best for our test dataset. This in turn minimizes our manual annotation efforts for labelling the images on the reduced version of the original dataset. Our image variance capture approach combined with extensive data augmentation techniques helps us maximize prediction throughput from our annotated data of chest CT scan images.

For calculating patient extremity estimation, we develop a two step deep learning pipeline where the first step is for detecting positive COVID-19 cases with the N-shot learning approach of SNNs. This step filters out negative COVID-19 images and feeds only COVID-19 positive images for our second feature localization and detection model, as shown in Fig. 1. Also, COVID-19 negative images constitute multiple lung diseases except COVID-19 having similar lung disorders and regular lung CT scan slice images that are disease free. *First step*, in this figure demonstrates the preprocessing pipeline for selecting images with high variance features eliminating quantitatively similar images. *Second step*,



Fig. 1 Data reduction preprocessing process followed by detection and prognosis deep learning pipeline

defines our deep learning pipeline to detect COVID-19 positive images and provide prognosis severity analysis accordingly. In this figure,  $C_{Vi} \& NC_{Vi}$  represents high variance subset of image datasets  $C_i \& NC_i$  for COVID-19 positive and negative cases respectively. First detection step is also complementary to our annotation effort for designing a better quality dataset where we are specifically focused on labelling false positives and negative slice images from our detection step. Hence, we reorganized our labelling efforts more around the miss-classified images from the first step for creating a better severity analysis model and annotated dataset.

We utilized the data reduction step to minimize our annotation efforts without compromising much on prediction performances of both models in our deep learning pipeline. This preprocessing step reduces the volume of quantitatively similar data by comparing the images with cosine similarity metric instead of qualitative features like blur, brightness, sharpness etc. We deployed a VGG16 [34] like similarity measurement neural network that measures variance of an image with respect to all dataset images for both COVID-19 positive and negative dataset split respectively. The high variance metric qualitatively serves as a measure for capturing maximum numbers peculiar features of an image with respect to other images in the dataset. These features can be COVID-19 features in COVID-19 positive dataset and some other distinctive features in case of COVID-19 negative images dataset. We were able to maintain our accuracy for our SNN model with minimum training images. As even after removal of 41.35 % highly similar images from the training dataset the performance of the model is not compromised beyond 1.5 %. Because their contribution is limited to model accuracy improvement and is majorly compensated with data augmentation techniques. Also, this variance maximization approach allowed us to capture a significant number of different instances per class for our localization model training.

Our contributions in this paper are summarized in this paragraph. In this research paper we provide a newly annotated dataset that captures features of COVID-19 in CT chest scans image slices. The annotations are presented in *Pascal VOC* and *csv* format in two different image resolutions of 300x300 and 500x500 pixels. Second, we introduce image similarity based data reduction methodology that omits unnecessary images that are not valuable to model accuracy improvement. Third, we propose a two step deep learning pipeline that is tuned to work well even on smaller datasets. Both steps are independent of one another and can be utilized individually with first being employed for diagnosis and second one for prognosis. Here, we have utilized the diagnosis and data reduction steps to minimize our labelling effort for the prognosis dataset. But, these efforts can be skipped for calculating severity metric if bounding box annotations are available. The clinical motivation behind this study is to propose a prognosis deep learning pipeline that is robust & extensible even on a small volume of data. This approach would also assist medical professionals to prioritize resources for more severe patients and would give assisting metric data to informatively supplement their decision.

#### 2 Related work

The detection and isolation of SARS-Cov-2 was done via bronchoalveolar-lavage samples from Wuhan Jinyintan Hospital with negative-stained electron micrographs of SARS-Cov-2 particles. The structure of the virus was analyzed with its morphology and cytopathic effects [41]. Multiple studies in the past have analyzed and collated their findings of COVID-19 CT chest scans features of patients from offset to recovery phase [4, 11]. From these studies it has been agreed upon that ground glass opacities, vascular dilation, crazy

paving pattern pattern, reverse halo and consolidation opacities are amongst the most common features when diagnosed with COVID-19. Most bilateral ground glass opacities appear in the initial stages of infection and that later transforms into consolidation blots [30]. Increase in ground glass lesions is associated with increased severity of patient's condition and future oxygen aid requirement [11]. Also, occurrence of these lung disorders is quite common amongst many lung related diseases which results in difficulty while determining the root cause as SARS-CoV-2.

A baseline CT chest severity score study [38] divided both lungs into 20 regions which were then subjectively evaluated and given scores based on their parenchymal opacification. The final score was then defined over summation of individual scores across these 20 regions. Our study is inspired from this similar concept of detection and localization of COVID-19 features or lung disorders. But, instead of focusing on calculating segmented lung region severity, we estimate the severity of the whole CT chest scan by detecting and localizing COVID-19 features only. In our study we developed an annotated dataset containing multiple COVID-19 related lung disorder features, labelled completely onto in focused sliced CT chest scans images as shown in Fig. 2 for detection. In this figure, the first row demonstrates some sample images of COVID-19 positive patients with different severity from the dataset. While in the second row their corresponding annotated images demonstrate labelled COVID-19 features in increasing order of severity from yellow to red colored bounding boxes. The green bounding boxes highlight negative samples or background class for hard negative mining by the model. The feature or lung disorder list for COVID-19 identification in the annotated dataset is decided after collation of information from multiple clinical studies [23, 25].

There are several deep learning architectures already that explore diagnostic capabilities of Convolutional Neural Network(CNNs) with X-ray and CT images [5, 7, 27, 32]. X-ray scan image analysis with data augmentation and transfer learning for COVID-19 detection from multiple CNN architectures have also demonstrated the efficacy in COVID-19 diagnosis predictions [29]. For dataset imbalance and noise removal, Generative Adversarial Networks (GANs) and Principal Component Analysis (PCA) combined with transfer learning modelling approaches have also demonstrated promising results [15]. Even though X-ray modalities have more availability as compared to CT modalities, their sensitivity and specificity is found rather low medically while diagnosing earlier stages of COVID-19.

Studies using Fourier-Bessel series expansion-based decomposition (FBSED) method for CT scan & X-ray images and Fourier-Bessel series expansion-based dyadic



Fig. 2 Sample images demonstrating different COVID-19 lung disorders with different associated colors from bounding box dataset

decomposition (FBD) method for X-ray images have also shown promising diagnosis results for COVID-19 [8, 9]. These methods obtain respective sub-band images (SBIs) from the X-ray and CT-scan images either in separated or merged state, post which these SBIs are fed into the main CNN network for final softmax based classification of COVID-19 images. Also, multiple studies have used COVID-CT as their dataset for diagnosis analysis because of its high patient count of CT chest scan slice images. And these studies have used transfer learning, capsule networks, AutoML, self-supervision and novel cross validation approaches to achieve better results in comparison to baseline solutions [12, 16, 26, 28, 33]. But, most of the above stated studies haven't employed extensive techniques to remove data related biases in predictions and also haven't analyzed limitations of the data quality. With the primary goal of these studies being improving accuracy and sensitivity of the models. The limited focus on achieving high specificity and no efforts on providing interpretability of their disease diagnostics for clinical validity the results from these studies are rather convoluted [6].

The base dataset under consideration in this paper is biased in nature with most CT chest scans belonging to mid-30s & mid-60s age group patients and approximately 60% of patients under consideration being biologically male. With no specifically detailed metadata, prior comorbidities information and quality control for given CT scan images slices in the dataset, the results would be tampered and unreliable in nature. Hence, we rely on our methodology that minimizes above stated biases in the dataset and captures maximum variation in the provided dataset to give fair generalizations. SNN exploits a unique architecture to rank similarity between input images and achieves good performance even on smaller datasets [21]. With fine-tuning from Bayesian optimization technique [35], good discriminative prediction performance is achieved from this model architecture. Also, triplet loss based methods have shown good performance in ranking similarity amongst the images [1, 37]. These latter stated approaches are good for determining the similarity between images but not for classifying them unlike SNN. Hence, we use the first above stated approach to evaluate and classify the COVID-19 positive & negative images. And employ the second similarity approach for data reduction during the model training step.

For localization and detection of an instance, SSDs are often used considering their simplicity and high frames-per-second (FPS) processing speed [24]. We limit our study to 2D CT chest scan sliced representation of images in this paper. But, also while scanning across 3D CT chest scan volumes in video like formats, these architectures would yield very high performance as well. Hence, it would have been even better model architecture to localize and detect dynamic bounding boxes per frame. SSDs fundamental idea is to use bounding boxes of different scales and aspect ratios per feature map for localization and remove region proposal network steps like Fast RCNN [13]. At prediction time, the network combines predictions from multiple feature maps to handle different size objects and generates a presence score based on intersection over union (IoU) method to create adjustments for matching the box better to instance shape. In our study we utilized the SSD approach for detecting COVID-19 lung disorders in a more granular manner for better medical interpretability.

#### **3** Datasets for experiments

In this section we first describe the properties associated with our base COVID-CT dataset on which we have built our annotation dataset. And we also describe the methodology that was used for collection of these CT chest scan slice images. Second, we describe properties associated with our annotated dataset for COVID-19 positive and negative patients. And we





(b) COVID-19 Negative Samples

Fig. 3 First and second row demonstrates samples of CT chest scan of COVID-19 positive and negative patients respectively

also describe methodology that was used for annotating our dataset with the bounding box coordinates.

# 3.1 Base dataset description

The COVID-19 dataset images collected are resized into standard sizes of 300x300 and 500x500 pixels [39]. This dataset is originally created from medRxiv and bioRxiv available preprint images on COVID-19 patients. Metadata information like gender, age, location, offset time period, severity of COVID-19 and radiology is also provided. Whereas negative COVID-19 images constituting of healthy & non-COVID-19 lung disorders CT scan slices are added from sources like MedPix, LUNA, Radiopaedia and PubMed Central. The patient data used from these sources is completely anonymized as per the privacy regulations and compliance procedures. Multiple images of a single patient are available that are visually very similar in nature are also present in this dataset. This dataset is already split into 349 images of 216 COVID-19 positive patients and 397 images of negative COVID-19 samples. Also, the COVID-19 patient dataset is skewed in nature with relatively more cases of male gender and older age groups. Specifically, 62.77% of the COVID-19 positive patients are biologically male in this dataset. And the patient data contains a relatively large number of samples for middle-aged (32 to 51 years) & older age groups (62 to 81 years). The Fig. 3 highlights some sample COVID-19 positive and negative images for a facile review of COVID-19 opacities and lung disorders. This figure demonstrates the similarity of features in both COVID-19 positive and negative samples. Demonstrating that without specific COVID-19 related domain knowledge of Chest CT scan it is a very hard task to classify these images and provide diagnosis information.<sup>1</sup>

<sup>&</sup>lt;sup>1</sup>Base dataset repository: https://github.com/UCSD-AI4H/COVID-CT

#### 3.2 Annotated dataset description

After completion of the data reduction preprocessing step we created annotations of our reduced dataset of 175 COVID-19 positive and 175 COVID-19 negative images. The criteria for elimination is high cosine similarity of image with respect to other images. And second, the systematic removal of such images doesn't cause a greater than 1.5 % decrease in model detection performance. This new dataset is again skewed, containing a larger number of CT scan slices belonging to male gender and older infected age groups. Each image on an average is having approximately around 15 bounding box annotations in this dataset. We annotated these training images with the listed COVID-19 disorders stated in Table 1. Also, we annotated the test dataset images to evaluate model performance in terms of mAP, average recall and precision. The annotations contain labels for specifically nine classes and one negative sampling background class. These labels are stated in Table 1 for reference and we have also associated a specific severity score & color code for a given class of COVID-19 feature manually. The color code's warmness expresses the severity of that corresponding COVID-19 lung disorder.

For creating bounding boxes we used an open source tool called *labelImg* which saves the annotations in *Pascal VOC* format. We also converted these annotations into *CSV* format as well with Python3 scripts. With this GUI tool we have created bounding boxes in CT chest scan image slices for all the classes corresponding to COVID-19 features. These annotations are added by two experienced radiologists from PGIMER, Chandigarh with the same professional exposure in COVID-19 diagnosis domain expertise. In any given annotation group the bounding box corresponding to an instance of each class completely captures the COVID-19 features in whole. Also, different features bounding boxes do partially overlap in annotations because of close vicinity of these features but complete overlap is generally avoided at all costs. For creating a bounding box, we made sure that the bounding box contained important traits associated with a feature instead of creating a complete bounding box over the whole feature. The reasoning for adopting this labelling technique is to avoid overlaps amongst multiple features and background class. We got these annotations medically reviewed for qualitative validation from practicing medical professionals for establishing standardized quality control over our dataset.

We created two variants of the annotation dataset that are subsets of base dataset having image resolution as 300x300 and 500x500 pixels respectively for SSD300 and SSD512

Cls.	COVID-19 Disorder	Sever.	BGR Code
1	ground glass opacity	1	(65, 255, 238)
2	vascular dilatation	1	(118, 241, 255)
3	subpleural band	1	(59, 235, 255)
4	bronchiectasis	2	(0, 255, 255)
5	traction bronchiectasis	2	(77, 183, 255)
6	crazy paving pattern	2	(0, 145, 255)
7	reverse halo	2	(0, 61, 255)
8	consolidation	3	(68, 23, 255)
9	lesion	3	(0, 0, 255)
0	Negative samples	0	(0, 255, 0)

 Table 1
 COVID-19 lung disorder classes description

18137

models respectively. In Fig. 4, different classes of annotations are highlighted in isolation in given sample images with their respective color code. We specifically present only a single class per image for more concise comprehension of our annotation methodology. For maintaining the annotation standard across multiple images we also cross-referenced the metadata associated with each image while annotating each feature. Along with that we have associated severity metric data for each COVID-19 positive image manually from the metadata file with review feedback from medical professionals. This was done to finally evaluate severity calculation performance of the model in the form of a multi-class classification problem categorizing COVID-19 patients into three categories (*Initial, Intermediate & Severe*) based on their severity score.

#### 4 Experimentation methodology

Our experimentation methodology broadly focuses on two tasks: first detection of COVID-19 patients and second is measuring the severity of the COVID-19 positive personnel. In our preprocessing step, we first measured the variance of each image with respect to every other image on the dataset. It assisted in eliminating similar images of probably the same patients that aren't assisting in improving accuracy of the diagnostic SNN model of COVID-19 on the test dataset. This reduction in dataset size without compromise in diagnostic model's accuracy helped us in saving time for creating annotation dataset that contained labels of all important images. Our base CT chest scan dataset was divided into standard 70:30 train-test split having equal number of COVID-19 positive and negative images for evaluation of our diagnostic model.

First, we created a 70:30 split for each data subset for COVID-19 positive and negative patients from our base dataset that contains these two classes separately already. Then we independently passed these images in different MobileNet [19] models for determining the similarities amongst different COVID-19 positive and negative images separately. As shown in Fig. 5's *Step 1: Variance Calculation*, each MobileNet similarity network evaluated the similarity of each image under consideration with respect to all the other images. The tuples  $(W_{i..x}, b_{i..x})$  and  $(W'_{i..x}, b'_{i..x})$  represents the weight values of the corresponding COVID-19 positive and negative image similarity of each



Fig. 4 Isolated single class annotated labelled examples in sample images from the bounding box annotated dataset



Fig. 5 Illustrative visualization of methodology stated in the form of algorithm 1

image with respect to every other image as a cosine similarity value ranging from [-1, 1] and stored all the values in 2D matrix. These stored values in 2D matrix are represented in the Fig. 5 with  $C_+\sigma^2$  and  $C_{\cdot}\sigma^2$  respectively for COVID-19 positive and negative image data subsets. For evaluating cosine similarity we transformed output of MobileNet from keras package with *GlobalAveragePooling2D()* operation. After that we evaluated the respective mean and variance value of each image from (1) and (2) with respect to the whole dataset from the cosine similarity values obtained from similarity calculator model predictions. The variance was calculated for both COVID-19 positive and negative dataset splits separately with two different similarity calculator models.

In the current context, variance ( $\sigma^2$ ) measures the spread of features or deviations in an image with respect to all the other images. It measures how far the features of the current image are from every other image in the dataset. Here, *N* represents the total number of images under consideration in the dataset from which the distance of the current image is evaluated. And  $x_i$  represents the cosine similarity score of a given image with respect to other images taken from the dataset.

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{1}$$

$$\sigma^{2} = \frac{\sum_{i=1}^{N} (x_{i} - \mu)^{2}}{N}$$
(2)

Second, we trained our diagnostic SNN model for separating COVID-19 positive images from negative ones. We used Bayesian optimization for optimizing hyperparameters like learning rate, momentum, momentum slope, convolutional layer multipliers and L2 regularization. We also executed another experiment after hyperparameter optimization in which we kept on removing images having least variance in random batch sizes of [3,4,5] most similar images having almost equal variance values. It was iteratively executed till the point when sharp decrease in accuracy arrived and then we halted the experiment at optimum dataset size.

The hyperparameter optimization and dataset reduction experiments were executed separately for reducing computation footprint. Methodology steps are stated in the form of algorithm 1 for reference and visual illustration of the algorithm 1 is presented in Fig. 5 for better comprehension. In Fig. 5's *Step 3: Data Reduction*, **D**<sub>R</sub> represents the final reduced training dataset for the diagnostic SNN model with ( $W_{i..y}$ ,  $b_{i..y}$ ) parameters which predicts with **Acc**<sub>i</sub> which is slightly smaller than **Acc**<sub>max</sub> by a number  $\varphi$  (1.5 %). **C**<sub>pred+</sub> and **C**<sub>pred-</sub> represent image subsets of dataset containing COVID-19 positive and negative images that are predicted by the diagnostic model. For final performance evaluation of the diagnostic SNN model we carried out 5-fold cross-validation in compliance with the imbalance dataset validation protocols in order to avoid overoptimistic and overfitted results [31].

Thereafter, we added annotations for the final data subset of COVID-19 positive and negative patient images which were derived from data reduction experimentation in the diagnostic network's performance capping step. We labelled the different COVID-19 features as described in Table 1 for all the 175 COVID-19 positive and negative images in our final data subset. We also labelled the distinctive features in the negative images with necessary domain knowledge for hard negative mining in SSD models. After that, we created a 60:15:25 split for training, validation and test set from our total 350 annotated images obtained after the data reduction step. The background class probability used for negative hard mining was not expressed independently for an image containing only background classes. Hence, our model raises an exception when such images containing only background classes are under analysis to calculate intersection of union (IoU) values. We handled these exceptions in model predictions by nullifying all predictions and don't proffer insights on COVID-19 negative images containing background class only. Also, we avoided interfering and overlapping the labelled feature classes with the background class. These bounding boxes are just omitted out of the process to avoid unnecessary noise in the model. After that with the same base hyperparameters, we trained our SSD models namely SSD7, SSD300 and SSD512 on the training dataset for qualitative and quantitative analysis on test dataset which contained annotated test images of 30 COVID-19 positive patients for severity metric analysis. These three models were created to get a qualitative sense of computational resources required to be spent for achieving the relative gain in accuracy of COVID-19 feature detection and localization.

We reviewed performance of our two-step deep learning pipeline both in terms of discrimination abilities of COVID-19 positive patients from negative ones and determining severity of the patient from bounding boxes annotations. For both the stated tasks performance evaluation metrics are independent of one another. For determining the performance of the diagnostic step, we only used the accuracy of the SNN model on the test dataset. For avoiding overoptimistic performance results we used 5-fold cross-validation with three different evaluation strategies: *Random, Slices & Voting*. First approach splitted the data randomly for training and test sets. The "Slice" strategy evaluated performance by creating a dataset test-train split at patient level, meaning that model performance was always evaluated on unknown individuals. The "Voting" cross-validation methodology considered all patient images for arriving at a diagnosis decision instead of relying on a single slice of CT scan.

Algorithm 1 Diagnosis modelling experiment pipeline: COVID-19 detection, hyperparameter optimization and data reduction process.

**Inputs:** Dataset C<sub>+</sub> and C<sub>-</sub> representing COVID-19 positive and negative dataset splits.

Deep Learning model graph declaration for siamese\_neural\_network and mobile\_net architectures with their respective model parameters.

#### **Pre-requisites:**

**70-30, Test-Train split:**  $T_{0.7}$ ,  $t_{0.3}$  = test\_train\_split(merge(C<sub>+</sub>, C<sub>-</sub>)), Random Split: fixed seed(16)

Similarity analysis network graph declaration:  $pooling\_average\_2d(mobile\_net(I_1, I_2))$ , Outputs values in between [-1, 1].

**Detection network graph declaration:** siamese\_neural\_network(hyperparameters,  $T_{0,7}$ ,  $t_{0,3}$ ), Outputs a similarity score in range [0,1]

**Result**: C<sub>pred+</sub>, C<sub>pred-</sub>, D<sub>R</sub>

*Where*,  $C_{pred+}$  and  $C_{pred-}$  represents COVID-19 positive and negative predictions on final reduced dataset  $D_R$ .

#### for I<sub>i</sub> in I<sub>total</sub> do

for  $I_j$  in  $I_{total}$  do  $C_+ \sigma_{ixj}^2 = \text{pooling\_average\_2d(mobile\_net(I_i, I_j))}$ end

end

```
C_{+}\sigma^{2} = \text{var\_calc}(C_{+}\sigma^{2}) / *2D \text{ variance matrix. */}
for I_{i} in I_{total} do
for I_{j} in I_{total} do
| C_{-}\sigma^{2}{}_{ixj} = \text{pooling\_average\_2d}(\text{mobile\_net}(I_{i}, I_{j}))
end
```

end

```
C_{\sigma}^{2} = \text{var_calc}(C_{\sigma}^{2}) / *2D \text{ variance matrix. */ opt_hyperparameters} = bayes_optimization(hyperparameters, T_{0.7}, t_{0.3}, siamese_neural_network) max_accuracy = siamese_neural_network(opt_hyperparameters, T_{0.7}, t_{0.3}) sort_increase(C_{\sigma}^{2}) / * Removing least variance images. */ sort_increase(C_{+}\sigma^{2}) / * Removing least variance images. */
```

while *round(max\_accuracy)* - *round(curr\_accuracy)*  $\geq \phi$  do

And with respect to performance evaluation of SSD models we evaluated mean average precision (mAP), average recall and average precision. To determine the exact severity parametric value, we analyzed the top 36 feature class predictions from the SSD models that have IoU value greater than 0.36. We calculated and determined the affected area from the summation of predicted bounding box coordinates for each class feature. After that we multiplied the affected area score with the severity weight associated with class features of COVID-19. And carried out a summation over all predicted features by the model with IoU value greater than 0.36. We normalized these calculated scores in the end by dividing the whole quantity by total image dimensions as shown in (3).

$$\Lambda = \frac{\sum_{y=1}^{Y} (A_f * W_f)}{A_i} \tag{3}$$

Y denotes the number of features predicted by the SSD model under consideration. A signifies the severity score calculated over top Y features detected for different classes in a COVID-19 positive case. A<sub>f</sub> is the area of that feature predicted which is multiplied by W<sub>f</sub> which sums over most predominant Y features. Finally, A<sub>i</sub> is the normalizing factor which is the total dimensional area of the image under analysis.

#### 5 Diagnostic experiment and results

First step of this experiment was to evaluate the similarity amongst the images from the base CT chest scan dataset for COVID-19 positive and negative patients independently. For calculating this similarity metric of images, we used the *MobileNet* model implementation in keras. After that, we pipelined model output to *GlobalAveragePooling2D* layer and finally evaluated cosine similarity of the two input images. All these cosine similarity values were stored in a 2D matrix of N by N size, where N is the size of COVID-19 positive and negative image dataset respectively. Finally, we evaluated the variance of each image with respect to the dataset by using (2). The histogram distribution plot of the actual and normalized variance values is presented in Fig. 6. We then normalized the value distribution of variance metric in images with square root normalization function and finally employed standard deviation as final metric to remove low variance images from the dataset.

We also used the base dataset as input to our diagnosis SNN model. This twin network with tied weights to ensure the correct mapping of images in feature space. For implementation of this network in keras, 3 Conv-RELU-MaxPool blocks were used which were then connected to Conv-RELU block feeding the fully connected sigmoid layer. The feature vectors from this layer were fused with the L1 weighted layer and its output was fed to the final layer that makes predictions. L2 regularization was used for every layer with stochastic gradient descent with specific earlier momentum of 0.5. We also used Bayesian optimization to assess the best architecture hyperparameters like learning rate, momentum, momentum slope, convolutional layer multiplier and layer penalization for regularization of the network. Table 2 shows the model hyperparameters and their respective optimization intervals. Specifically, we divided these intervals into four equal width bins including the end interval values to obtain five unique values for each optimized hyperparameter. We used GPyOpt library to optimize on these set of parameters as our domain. Our sample model training function trained these prototype models upto 100,000 iterations with accuracy and F1-score evaluations for performance comparison. The Table 2's highlights the best hyperparameters for the SNN model that were used to train the final similarity detection diagnostic network.

The dataset for training was split into a 80:20 train and validation split. In this implementation we decayed the learning rate by 0.01% after 500 iterations and Glorot initialization



II. Image count and Normalized Variance Distribution Plot

Fig. 6 Distribution analysis of variance values in image features of both COVID-19 positive and negative patients

was used for default weight value initialization for the network. We also employed callback keras method to early stop the model after achieving maximum accuracy with a patience parameter of 5 epochs. First, we trained our network on the base dataset to find the optimum hyperparameters for the model. Our initial model performance measurement was our baseline diagnosis accuracy and it was used for determining the best hyperparameters on the given base dataset. After that, we iteratively removed the subset of images having least variance in a random group size of 3 to 5 from the base training dataset of both positive and negative COVID-19 patients as stated in the algorithm 1. We executed this process iteratively until the accuracy started decreasing at a rapid rate and surpassed a specific threshold  $\varphi$  of 1.5 % relatively.

We observed the beginning of sharp decline after 130<sup>th</sup> iteration at an accuracy value of 87.6% from the accuracy-iteration data reduction graph in the Fig. 7. Hence, we created a snapshot of the dataset at that particular iteration which contained a list of all the COVID-19 positive and negative images to reproduce that diagnostic performance for given hyperparameters. Table 3 compares the global maximum diagnostic performance with previous studies and we can see that proposed approach outperforms the existing approaches. Table 4 provides the cross-validated evaluation outputs on the test dataset. This also demonstrates the importance of using 'Slice' and 'Voting' strategies for more qualitatively accurate performance representation by our model.

Further, the reduced version of the dataset was used as the original training and validation dataset for the SSD models. We plotted the iterative process in the Fig. 7 to observe this steep decline which only happened after large chunks of similar images were removed CNINI hruno

Sinn hyperparameters	
Learning Rate	10e-4
Learning Rate Decay	0.01/epoch
Batch Size	64
Conv. Layer Momentum	0.5
Conv. Layer Weights	$\mu$ =0 & $\sigma$ =0.01
Conv. Layer Biases	$\mu$ =0.5 & $\sigma$ =0.01
Full Conn. Layer Weights	$\mu$ =0 & $\sigma$ =0.2
Full Conn. Biases	$\mu$ =0.5 & $\sigma$ =0.01
Conv. Layer L2 Penalty	0.005
Full Conn. Layer L2 Penalty	0.005
Momentum Slope	0.0198
Momentum <sub>j</sub> Threshold	0.9
Bayesian Opt. Intervals	
Learning Rate	[10e-7, 10e-2]
Learning Rate Decay	[0.0, 0.05]
Layer-wise Momentum	[0.0, 1.0]
Momentum Slope	[0.001, 0.1]
Momentum Threshold	[0.5, 1.0]
Layer-wise L2 Penalty	[0.00, 0.025]
Layer-wise Learning Rate Multiplier	[0.1, 1]

Table 2	SNN hyperparameter s	pecifications and Bayesian	Optimization	parameter intervals

from our N-shot learning experiment. From this experiment we observed that the diagnostic performance of the model is not altered when low variance images are removed from the training dataset. Hence from our results, we confirm that even after removal of 40%



Average Accuracy Evaluation with Dataset Image Reduction Iterations

Fig. 7 Average accuracy plot during the data reduction process

Table 3Model performancesummary on COVID-CT dataset	Approach	Acc(%)	F1(%)	AUC(%)
	Yang et al. [38]	85.0	85.9	92.8
	Silva et al. [33]	87.6	86.19	90.5
	Mobiny et al. [28]	87.6.0	87.1	96.1
	He et al. [16]	86.0	85.0	94.0
	Ewen et al. [12]	86.2	87.0	86.1
	Mantas et al. [26]	88.3	85.0	94.0
	Proposed Approach	89.3	88.7	95.6

of the training dataset the performance of the network is only slightly affected in terms of diagnostic power.

We have reduced our training data by a total of 41.35 % percent by removing low variance images which in turn reduced our labelling efforts for the next experiment. After that we fed the resultant training and test dataset to our next prognostic model on which we carried out the severity analysis experiment.

# 6 Prognosis experiment and results

We used the training set images from the detection experiment as input images for training the severity estimation SSD networks. From these images COVID-19 features were detected and localized in the first step of the network. We pipelined multiple data augmentation techniques for performance improvements in our SSD models, as shown in Fig. 8. *Red* color represents the background class whereas *blue* and *green* color represents COVID-19 feature classes in the given image demonstrating three sample data augmentation transformations. The predictions for different feature classes and bounding box coordinates on COVID-19 positive images were made by serving batches of these data augmented images as input to our SSD networks. From which effective area of each predicted attribute class was calculated and we selected the top 36 predicted classes excluding the background class. Which was then fed into (3) for severity estimate value calculation. Finally, we divided the derived severity estimates from test images into three classes known as *initial, intermediate* & severe. From multiclass categories derived on these three severity types, we measure the prognosis capabilities of our best SSD network.

In SSDs each set of feature layers produces a fixed set of predictions with convolutional filters. Bounding boxes of different aspect ratios and sizes are associated with each feature map of the network. Which are relatively fixed to a corresponding cell over which they are tiled. Hence, for each feature map cell offset relative to the default box shape is predicted for the given cell. And scores specific to a given class marks the presence of class instances in that box. The objective loss function is the weighted sum of localization & confidence

Table 4         5-fold cross-validation           result summarization	Approach	Acc(%)	F1(%)	AUC(%)
	Random	88.88±0.42	88.02±0.68	94.17±1.43
	Slicing	$77.62 \pm 7.51$	$76.19 \pm 8.43$	$85.52 \pm 4.3$
	Voting	$80.60 {\pm} 4.39$	79.11±6.57	$88.74 {\pm} 5.92$



a. Original Image

b. Resized Image

c. Translated Image

d. Stretched Image

Fig. 8 Three sample transformations from data augmentation with corresponding resized bounding boxes

loss combined together. Where localization loss is smooth L1 loss regressing predicted and ground truth boxes. And confidence loss is softmax loss over multiple classes' confidence values.

Our base network was VGG16 based and the fc6 & fc7 were converted to convolutional layers. After that, sub-sampling of parameters was done from these two layers. Changes



Fig. 9 Abstract representations of SSD models trained in this experiment with their respective architecture details

Table 5       Processing training and inference time with Quadro         P6000 GPU on 500x500 image input annotation dataset	SSD Models	Training	Inference (1 image)
input annotation dataset	SSD7	2h 47min	0.0021s
	SSD300	5h 47min	0.0049s
	SSD500	6h 38min	0.0067s

were also made in pool5 layer from 2x2-s2 to 3x3-s1 convolutional operation and "holes" were filled with *a trous* algorithm [18].

In our training experiment, we trained six model variants at two different resolutions of 300x300 and 500x500 pixels. First we trained a small and light model of SSD7 with four predictor and seven convolutional layers network on both the stated resolutions. We primarily utilized this simple architecture of SSD7 for prototyping models to determine optimum value of hyperparameters like aspect ratios, scale values, IoU threshold for non-maximum suppression etc. This was done to get a qualitative sense of the baseline mAP performance, estimating basic hyperparameters and also qualitatively comparing training time of this relatively small network with heavier SSD networks. Also, we trained baseline SSD300 and SSD512 networks on the above two resolutions respectively [24] and based our hyperparameters from prototyping information derived with the SSD7 network modelling. The architectural differences qualifying different convolutional and prediction layers used by each network are stated in Fig. 9. The baseline SSD models are more than *100* times larger than SSD7 in terms of trainable parameters resulting in a greater than 3 times larger computational training footprint. Table 5 demonstrates the computational and inference time for respective architectures under analysis.

We trained our models with Adam optimizer having learning rate = 0.001,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 10^{-08}$  and decay = 0.0001. We employed techniques like plateau learning rate reduction and early stopping with a patience parameter of 2 epochs for achieving best model results in our experimentation. In our data augmentation pipeline we randomly added brightness, contrast, saturation, hue, translation, rotation and shear scaling, as shown in Fig. 8. For inference, we used non-maximum suppression technique with threshold 0.01 to filter out majority negative boxes. After that, we applied Jaccard similarity overlap i.e. intersection-of-union (IoU) of 0.36 per class and kept the top 36 detections per image.

The Table 6 states the instance detection and localization related results for COVID-19 feature classes from SSD models. These results propound the fact that SSD models have demonstrated reasonably well mAP values on a relatively small dataset with the help of extensive data augmentation. Also, with increasing model size the results do show significant improvements. The precision-recall plots in Fig. 10 demonstrate different averaged

THE COURD 10 C						
detection results	SSD Models	mAP	R <sub>avg</sub>	Pavg		
	SSD7-300x300	0.48	0.43	0.46		
	SSD7-500x500	0.53	0.57	0.49		
	SSD300-300x300	0.62	0.63	0.55		
	SSD300-500x500	0.65	0.67	0.59		
	SSD512-300x300	0.67	0.71	0.63		
	SSD512-500x500	0.69	0.73	0.66		



I. Precision-Recall Plot: [500x500] pixels CT images.



II. Precision-Recall Plot: [300x300] pixels CT images.

Fig. 10 Precision and recall plots for object detection problems for evaluating mAP

out mAP values for different SSD models derived from evaluating the Area Under Curve (AUC). The SSD512 model with 500x500 pixel size images as input performs better in every aspect of model performance evaluation as shown in Table 6. Therefore, we use it to determine severity estimates on the test dataset of COVID-19 patients. Also, it had relatively higher recall values to quantify the models ability to correctly detect COVID-19 features provided the image is of COVID-19 positive patient.

For severity analysis estimates, we used (3) to determine COVID-19 patient's condition. In our test dataset, we had 30 images of COVID-19 positive patients and an almost equivalent number of negative patients. Our diagnostic model eliminated the COVID-19 negative images containing background class only and the prognosis model didn't provide any bounding box results for those images. Hence, we determined severity estimates only for the 30 COVID-19 positive images under consideration in our test dataset. The results from SSD models are fed into the severity estimation (3) and the final numerical values from this equation were normalized into the range of [0, 1] values only.

The Table 7 provides the complete result report of severity analysis estimates for the COVID-19 patients predicted by our model. The estimated severity numeric value instead of being presented as standard [0, 1] values were transformed into categorical *initial, interme-diate and severe* states based on the equal division of [0, 1] numeric space. The actual patient severity categories were determined manually from the analysis of metadata associated with the test set images. Hence, the problem was transformed into a multi-class classification analysis for determining quantitative model performance. The confusion matrix corresponding to this problem is presented in Fig. 11 and results are summarized in result Table 7. Also, the total accuracy of the prognosis model is **47** % as derived from the confusion matrix.

Teble 7 COVID 10						
estimation analysis results	Labels	Pr.	Rcl.	Spec.	F1-Scr.	Supp.
	Initial	0.86	0.50	0.74	0.63	12
	Intermediate	0.30	0.27	0.44	0.29	11
	Severe	0.38	0.71	0.6	0.50	7
	F1-Types	Pr.	Rcl.	Spec.	F1-Scr.	Supp.
	Micro Avg.	0.47	0.47	0.59	0.47	30
	Macro Avg.	0.51	1.00	0.59	1.00	30
	Weighted Avg.	0.54	0.47	0.60	0.47	30

Accuracy: It determines the fraction of all samples that are correctly predicted by the classifying model.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$
(4)

Precision: It determines the fraction of all positive predictions that were actually positive.

$$\mathbf{Precision} = \frac{TP}{TP + FP} \tag{5}$$

Recall: It determines the fraction of all positive samples that were correctly predicted positive by the classifier i.e. True Positive Rate (TPR).

$$\mathbf{Recall} = \frac{TP}{TP + FN} \tag{6}$$

Specificity: It determines the fraction of all negative samples that were correctly predicted negative by the classifier i.e. True Negative Rate (TNR).

$$\mathbf{Recall} = \frac{TN}{TP + FN} \tag{7}$$



Fig. 11 Confusion matrix for severity analysis prediction results

 F1-Score: It is a harmonic mean of precision and recall which delivers more insights on biased data. It is used in this case as False Negatives and False Positives both are crucial to our analysis as well.

$$F1 - Score = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$
(8)

- Micro Average: The micro-averaged parameters do not consider each class individually, instead all metrics are calculated by summation of quantities over all classes. The formula to evaluate all metrics is the same as above, only its scope is different which takes into account global effects.
- Macro Average: The macro-averaged parameters are calculated by taking the mean of each metric calculated for single classes evaluated in isolation. The formula to evaluate all metrics is the same as above, only its scope is different which takes into account averaging effects.
- Weighted Average: The weighted-averaged parameters are calculated by taking the weighted mean of each metric calculated for single classes evaluated in isolation. The weight is assigned to each class based on the Support each class is having for representation, where Support is the number of true samples present for each class.

The results in Table 7 shows that our model performs best when the feature type in the images are *Initial* ones in nature. This can be justified with an abundance of annotations being available in the annotation dataset for *Vascular Dilation* and *Ground Glass Opacity* classes. These classes have the lowest severity amongst all COVID-19 features. Also, our model is having good specificity value for *Severe* class as well. This is especially helpful in order to prioritize severe patients more accurately. Also, our test results demonstrate 47% accuracy while determining the exact category of the COVID-19 positive patient severity. And reports even higher specificity of 60% to avoid wrong prognosis analysis.

Also in Fig. 12, we qualitatively observe our model predictions for both correctly and incorrectly predicted images. From the first incorrectly predicted image in the second row, we observe that sometimes the model might consider consolidation patches appearing near to background features like lung walls and might miss-identify them. From Fig. 12, we also observe that our SSD models are very good at capturing vascular dilation and ground glass opacity features as these are the most abundant annotations present in the dataset. Also, we observe that the model does identify the lesions and bronchiectasis present in the images



b. Incorrect Severity Classification

Fig. 12 *First* row demonstrates sample COVID-19 positive images with correctly classified severity of patients by SSD512 detection network and the *second* row demonstrates sample images that are miss-classified by the severity analysis model

quite correctly even if their annotations are not present in large amounts for this class in particular. This can be attributed to their simpler circle-like shape and almost the same size in every instance.

The overall performance in the severity estimation task turns out to be acceptable as we combine all the COVID-19 features for analysis. And then effectively calculate the severity metric based on the sum total of all the features as a whole for an image. Also, fixating on strong generalizations regarding the mAP trends from quantitative tests and qualitative discussions is not entirely possible with this biased dataset. But, with increased size of the dataset capturing a variety of COVID-19 features having a very high number of annotations for each feature class, valuable generalization trends can be derived for every metric discussed in this paper. In the next two sections we present model pipeline limitations and our conclusion to summarize insights from all the experiments. We also comment on future steps and possible scope of such deep learning pipelines for smaller datasets.

# 7 Limitation analysis

To better have an understanding about the errors made by the model pipeline, we conducted automated and manual analysis on test set predictions for both steps in our model pipeline. For step one, baseline SNN model accuracy was slightly overestimated and only after "Slice" and "Voting" evaluations did we arrive at correct predictions. We believe this happened because we didn't factored in the multiple instance learning (MIL) nature of this problem.

Also, the first diagnostic step does not provide explainable and interpretable information which is required for medically compliant procedures. Our SNN pipeline's performance on unseen data depends heavily on the variance of COVID-19 positive and negative images fed during training. But with the current data size for testing, validation and training our pipeline is vulnerable to a lot of unaccounted biases. Because of these two specific reasons we used the first step combined with image similarity removal approach to get statistically different quality images for annotation labelling process.

For the second step, the detection performance of our model was highly biased towards *Initial* and *Severe* classes. The performance for *Initial* class can be attributed to abundantly available annotations for initial stage COVID-19 features like ground glass opacities and vascular dilation. And the performance for *Severe* class can be attributed to abundant availability of lesion annotations and distinctive visual features of consolidation formation in lungs. Hence, the limited annotation availability for different defined classes in SSD models restricted the detection capabilities of our model.

# 8 Conclusion

Our deep learning pipeline demonstrates that it can quantify severity of COVID-19 patients from annotated CT chest scan slice images which would further assist in the prognosis process. This novel approach instead of using segmentation or regression based modelling techniques for diagnosis adopts a bounding box estimation approach for COVID-19 feature detection and affected area estimation. It provides the prognosis information by evaluating the sum of areas covered by different COVID-19 features for severity estimations. This analysis quantifies the COVID-19 feature affected areas in lungs which would serve as deescalation or escalation assisting factors for admitting patients in intensive care units (ICU)

based on the calculated severity score. The analytical information from this system provides valuable information for resource management in medical facilities.

In the stated methodology we also acknowledge the bias present in given dataset samples of CT scan images. And we also attempted to minimize its impact with removal of quantitatively highly similar images. We also acknowledge the limited annotation effort hours at our disposal for creation of bounding box dataset. Therefore, adapted and tuned our deep learning prognosis pipeline to mitigate this issue. In this paper we only restricted our functionally two-step deep learning pipeline to supervised approaches for both diagnosis and prognosis steps. As part of future work our direction would be more oriented towards hybrid unsupervised and supervised approaches for reducing the supervised labelled data preparation effort. A study on meticulous inspection of Scanning Electron Microscope (SEM) images of the electrospun nanofiber with hybrid unsupervised and supervised machine learning pipeline has already delivered promising results for anomaly detection task in nanofibers [20]. Another approach, although in aspect-level sentiment analysis task outper-formed baselines on SemEval-16 dataset with two-step mixed approach. First, it used rule based linguistic pattern identification for aspect extraction and then reused the extracted aspects as labels for aspect-term extraction with attention-based Bi-LSTM model [10].

From an extensibility standpoint, our approach would achieve more insightful and reliable results with more annotated samples in a more comprehensive dataset on better standardized CT Chest scan images. The severity score would be more meaningful, extensible and expressive with more standardized high-quality dataset with statistically diverse enough samples in terms of gender, age, comorbidities etc. Additionally, basic structure of this proposed methodology and pipeline can be used to get prognosis information on new diseases by creating a new minimal volume annotated dataset for that disease. In conclusion, we proposed an automated and extensible robust methodology that provides useful prognosis information for medical assistance to prioritize COVID-19 patients during their treatment.

Acknowledgments We also gratefully acknowledge the support of NVIDIA Corporation with donation of the Quadro P6000 GPU used for this research.

#### Declarations

**Conflict of Interests** We certify that there is no actual or potential conflict of interest in relation to this article.

#### References

- 1. Appalaraju S, Chaoji V (2017) Image similarity using deep cnn and curriculum learning, arXiv:1709.08761
- Azhar EI, Hui DS, Memish ZA, Drosten C, Zumla A (2019) The middle east respiratory syndrome (mers). Infect Dis Clin 33(4):891–905
- Bai Y, Yao L, Wei T, Tian F, Jin D.-Y., Chen L, Wang M (2020) Presumed asymptomatic carrier transmission of covid-19. Jama 323(14):1406–1407
- Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, Diao K, Lin B, Zhu X, Li K et al (2020) Chest ct findings in coronavirus disease-19 (covid-19): relationship to duration of infection. Radiology:200463
- Bressem KK, Adams LC, Erxleben C, Hamm B, Niehues SM, Vahldiek JL (2020) Comparing different deep learning architectures for classification of chest radiographs. Sci Rep 10(1):1–16

- Burlacu A, Crisan-Dabija R, Popa IV, Artene B, Birzu V, Pricop M, Plesoianu C, Generali D (2020) Curbing the ai-induced enthusiasm in diagnosing covid-19 on chest x-rays: the present and the nearfuture. medRxiv
- Butt C, Gill J, Chun D, Babu BA (2020) Deep learning system to screen coronavirus disease 2019 pneumonia. Appl Intell:1
- Chaudhary PK (2020) Automatic diagnosis of covid-19 and pneumonia using fbd method. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, pp 2257–2263
- Chaudhary PK, Pachori RB (2021) Fbsed based automatic diagnosis of covid-19 using x-ray and ct images. Comput Biol Med 134:104454
- Chauhan GS, Meena YK, Gopalani D, Nahta R (2020) A two-step hybrid unsupervised model with attention mechanism for aspect extraction. Expert Syst Appl 161:113673
- Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, Cui J, Xu W, Yang Y, Fayad ZA et al (2020) Ct imaging features of 2019 novel coronavirus (2019-ncov). Radiology 295(1):202–207
- Ewen N, Khan N (2021) Targeted self supervision for classification on a small covid-19 ct scan dataset. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, pp 1481–1485
- Girshick R (2015) Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 1440–1448
- 14. Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, Wade R (2020) The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (sars-cov-2; covid-19): A systematic review and meta-analysis of 148 studies from 9 countries. Available at SSRN 3582819
- Gupta RK, Sahu Y, Kunhare N, Gupta A, Prakash D (2021) Deep learning-based mathematical model for feature extraction to detect corona virus disease using chest x-ray images: International Journal of Uncertainty. Fuzziness and Knowledge-Based Systems
- He X, Yang X, Zhang S, Zhao J, Zhang Y, Xing E, Xie P (2020) Sample-efficient deep learning for covid-19 diagnosis based on ct scans. MedRxiv
- Hoang A, Chorath K, Moreira A, Evans M, Burmeister-Morton F, Burmeister F, Naqvi R, Petershack M, Moreira A (2020) Covid-19 in 7780 pediatric patients: A systematic review. EClinicalmedicine:100433
- Holschneider M, Kronland-Martinet R, Morlet J, Tchamitchian P (1990) A real-time algorithm for signal analysis with the help of the wavelet transform. In: Wavelets. Springer, pp 286–297
- Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv:1704.04861
- Ieracitano C, Paviglianiti A, Campolo M, Hussain A, Pasero E, Morabito FC (2020) A novel automatic classification system based on hybrid unsupervised and supervised machine learning for electrospun nanofibers. IEEE/CAA J Autom Sin 8(1):64–76
- 21. Koch G, Zemel R, Salakhutdinov R (2015) Siamese neural networks for one-shot image recognition. In: ICML Deep learning workshop, vol 2. Lille
- Lam CW, Chan MH, Wong CK (2004) Severe acute respiratory syndrome: clinical and laboratory manifestations. Clin Biochem Rev 25(2):121
- Li Y, Xia L (2020) Coronavirus disease 2019 (covid-19): role of chest ct in diagnosis and management. Am J Roentgenol 214(6):1280–1286
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC (2016) Ssd: Single shot multibox detector. In: European conference on computer vision. Springer, pp 21–37
- 25. Liu H, Liu F, Li J, Zhang T, Wang D, Lan W (2020) Clinical and ct imaging features of the covid-19 pneumonia: Focus on pregnant women and children. Journal of infection
- 26. Mantas J et al (2020) Setting up an easy-to-use machine learning pipeline for medical decision support: A case study for covid-19 diagnosis based on deep learning with ct scans. importance health inform. Public Health Dur Pandemic 272:13
- 27. Mei X, Lee H-C, Diao K-y, Huang M, Lin B, Liu C, Xie Z, Ma Y, Robson PM, Chung M et al (2020) Artificial intelligence–enabled rapid diagnosis of patients with covid-19. Nat Med:1–5
- Mobiny A, Cicalese PA, Zare S, Yuan P, Abavisani M, Wu CC, Ahuja J, de Groot PM, Van Nguyen H (2020) Radiologist-level covid-19 detection using ct scans with detail-oriented capsule networks, arXiv:2004.07407
- Nayak SR, Nayak DR, Sinha U, Arora V, Pachori RB (2021) Application of deep learning techniques for detection of covid-19 cases using chest x-ray images: A comprehensive study. Biomed Signal Process Control 64:102365
- Salehi S, Abedi A, Balakrishnan S, Gholamrezanezhad A (2020) Coronavirus disease 2019 (covid-19): a systematic review of imaging findings in 919 patients. Am J Roentgenol:1–7
- Santos MS, Soares JP, Abreu PH, Araujo H, Santos J (2018) Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier], vol 13

- 32. Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, He K, Shi Y, Shen D (2020) Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. IEEE Rev Biomed Eng
- Silva P, Luz E, Silva G, Moreira G, Silva R, Lucio D, Menotti D (2020) Covid-19 detection in ct images with deep learning: A voting-based scheme and cross-datasets analysis. Inf Med Unlocked 20:100427
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556
- Snoek J, Larochelle H, Adams RP (2012) Practical bayesian optimization of machine learning algorithms. In: Advances in neural information processing systems, pp 2951–2959
- Tahamtan A, Ardebili A (2020) Real-time rt-pcr in covid-19 detection: issues affecting the results. Expert Rev Mol Diagnostics 20(5):453–454
- Wang J, Song Y, Leung T, Rosenberg C, Wang J, Philbin J, Chen B, Wu Y (2014) Learning fine-grained image similarity with deep ranking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1386–1393
- Yang R, Li X, Liu H, Zhen Y, Zhang X, Xiong Q, Luo Y, Gao C, Zeng W (2020) Chest ct severity score: An imaging tool for assessing severe covid-19. Radiol Cardiothoracic Imaging 2(2):e200047
- 39. Zhao J, Zhang Y, He X, Xie P (2020) Covid-ct-dataset: a ct scan dataset about covid-19, arXiv:2003.13865
- 40. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X et al (2020) Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study the lancet
- 41. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R et al (2020) A novel coronavirus from patients with pneumonia in china. 2019 New England Journal of Medicine

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

# Affiliations

# Ashish Rana<sup>1</sup> · Harpreet Singh<sup>1</sup> · Ravimohan Mavuduru<sup>2</sup> · Smita Pattanaik<sup>2</sup> · Prashant Singh Rana<sup>1</sup>

Ashish Rana arana\_be15@thapar.edu

Ravimohan Mavuduru ravismi2003@yahoo.com

Smita Pattanaik drs\_pattanaik@yahoo.com

Prashant Singh Rana psrana@gmail.com

- <sup>1</sup> Department of Computer Science and Engineering, TIET, Patiala, Punjab, India
- <sup>2</sup> Department of Urology and Pharmacology, PGIMER, Chandigarh, India