



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Coronavirus disease (COVID-19) detection using X-ray images and enhanced DenseNet

Saleh Albahli^{a,*}, Nasir Ayub^b, Muhammad Shiraz^b

^a Department of Information Technology, Qassim University, Buraydah, Saudi Arabia

^b Department of Computer Science, Federal Urdu University, Islamabad, 44000, Pakistan

ARTICLE INFO

Article history:

Received 26 December 2020
Received in revised form 18 June 2021
Accepted 23 June 2021
Available online 25 June 2021

Keywords:

Deep learning
COVID-19
Biomedical imaging
DenseNet
ResNet
Convolutional neural network

ABSTRACT

The 2019 novel coronavirus (COVID-19) originating from China, has spread rapidly among people living in other countries. According to the World Health Organization (WHO), by the end of January, more than 104 million people have been affected by COVID-19, including more than 2 million deaths. The number of COVID-19 test kits available in hospitals is reduced due to the increase in regular cases. Therefore, an automatic detection system should be introduced as a fast, alternative diagnostic to prevent COVID-19 from spreading among humans. For this purpose, three different BiT models: DenseNet, InceptionV3, and Inception-ResNetV4 have been proposed in this analysis for the diagnosis of patients infected with coronavirus pneumonia using X-ray radiographs in the chest. These three models give and examine Receiver Operating Characteristic (ROC) analyses and uncertainty matrices, using 5-fold cross-validation. We have performed the simulations which have visualized that the pre-trained DenseNet model has the best classification efficiency with 92% among two other models proposed (83.47% accuracy for inception V3 and 85.57% accuracy for Inception-ResNetV4).

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

In October of 2019, the novel coronavirus (COVID-19) pandemic emerged in Wuhan, China, and has become a major public health issue worldwide [1]. The virus that caused the COVID-19 pandemic disease was also called coronavirus 2 severe acute respiratory syndromes, also known as SARS-CoV-2 [2]. Coronaviruses (CoV) are a family of viruses that cause cold-related illnesses like Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV). Coronavirus disease (COVID-19) is a new species that was discovered in 2019 and had not been identified in humans prior. Coronaviruses are zoonotic due to animal-to-human contamination [3]. Studies have shown that the SARS-CoV virus is transmitted to humans from musk animals, and the MERS-CoV virus is infected to humans from dromedary [4]. The COVID-19 virus is believed to be spread to humans by bats [5]. The infection was widely spread due to airborne transmission of the disease from person to person.

Although COVID-19 cases cause milder symptoms as per 82% of total cases and these are considered as serious or critical [6]. The cumulative number of Coronavirus cases by the end of January is around 104 million, of which 2 million died and 75 million recovered. Although the disease marginally survives 99%

of the number of affected patients, 1% have a severe or vital condition [7].

Infection signs include worsening of cough, fever and dyspnea. The infection will cause pneumonia, extreme acute respiratory syndrome, septic shock, multi-organ failure, and death in more serious cases [4–7]. It was determined that men are more afflicted than women, and that children from 0–9 years of age do not die [8]. Respiratory concentrations of COVID-19 pneumonia cases are higher as compared to stable individuals [9]. In many developing nations, the health sector has failed to control the reduction of cases and providing more intensive care units. The number of infected patients is increased more than the available resources. Intensive care facilities are packed with patients of worsening COVID-19 pneumonia symptoms.

According to the guidance shared by the Chinese government, the diagnosis of COVID-19 should be confirmed by gene sequencing of respiratory samples or blood samples as a primary predictor of Reverse Transcription Polymerase Chain Reaction (RT-PCR) or hospitalization. In the latest public health emergency, the low sensitivity of RT-PCR means that many COVID-19 cases may not be easily detected and will not be treated properly. They also run the risk of infecting a wider population, considering the extremely contagious nature of the virus [25,26]. Instead of testing people for a positive virus scan through the older method, treatments now require chest X-ray scan images to identify the COVID-19 virus. These approaches would help hospitals to identify and treat patients quicker. Even if COVID-19

* Corresponding author.

E-mail addresses: salbahli@qu.edu.sa (S. Albahli), nasir.ayubse@gmail.com (N. Ayub), muhhammad.shiraz@fuust.edu.pk (M. Shiraz).

Table 1

Overview of previous approaches, their strengths, gaps and improvement of proposed DenseNet model over them.

Source: Table was derived from [10].

Reference	Proposed approach	Main features
Apostolopoulos et al. [11]	Used transfer learning models based on MobileNetV2, VGG19, Xception, Inception and ResNetV2	Used X-ray and achieved a maximum of 93% accuracy
Wang and Wong [12]	Introduced their own model named COVID-Net, the first open-source COVID-19 detection system	Used X-ray and achieved a maximum of 92% accuracy
Song et al. [13]	Proposed their own model DRE-Net and compared its performance with DenseNet, VGG-16 and ResNet	Used CT and achieved a maximum of 86% accuracy
Sarker et al. [14]	Used transfer learning on DenseNet-121 network	Used X-rays and achieved 85.35% accuracy with 3 classes
Hasan et al. [15]	Used DenseNet-121 network	Used CT and achieved 92% accuracy
Zheng et al. [16]	Proposed their own model DeCoVNet for classification	Used CT and achieved 97% accuracy
Xu et al. [17]	Proposed modifies version of ResNet-18 based CNN network	Used X-ray and achieved 86% accuracy
Minaee et al. [18]	Used ResNet50, ResNet18, DenseNet-121, and SqueezeNet network	Used X-ray and achieved 97.6% AUC
Ozturk et al. [19]	Proposed their own model DarkCovidNet for classification of COVID-19	Used X-ray and achieved 97% accuracy
Ardakani et al. [20]	Used 10 CNN networks including AlexNet and ResNet-101 for classification of 2 classes	Used CT and achieved a maximum of 99% accuracy
Li et al. [21]	Proposed their own model COV-Net for classifying 3 classes	Used CT and achieved 96% accuracy
Yang et al. [22]	Used ResNet-50 and DenseNet-169 network for classification.	Used CT and achieved 79.5% accuracy with DenseNet-16
Abbas et al. [23]	Proposed their own model DeTrac-ResNet18 CNN that uses Decompose, Transfer, and Compose architecture	Used X-ray and achieved a maximum of 95.12% accuracy
Chen et al. [24]	Used UNet++ along with Keras for segmentation of CT images and detection of COVID-19	Used CT and achieved 95.24% accuracy

does not cause death, some patients still recover with permanent lung damage. According to the World Health Organization (WHO), COVID-19 can open holes in the lungs such as Severe Acute Respiratory Syndrome (SARS), giving lungs a “honeycomb-like appearance” [27].

One of the techniques used to detect pneumonia is a computed chest tomography (CT) scan. Artificial Intelligence (AI) based automated CT image analysis systems have been developed to identify, measure, and track coronavirus, and also differentiate patients with coronavirus from disease-free [28]. They developed a deep learning-based method for automated segmentation of all lungs and infection sites using CT [29], in a study by Fei et al. Xiaowei et al. aimed at developing an early screening model for the differentiation between COVID-19 and Influenza-A viral pneumonia and stable cases using pulmonary CT images and deep learning techniques [30]. Butt et al. [31] used ResNet-18 on CT scan images to achieve an accuracy of 86%.

Depending on the COVID-19 radiographic changes from CT images, Shuai et al. research developed a deep learning algorithm that can extract the graphical characteristics of COVID-19 to provide a clinical diagnosis before pathogenic testing, and thus saves the valuable time for diagnosis of the disease [32].

COVID-19 is demonstrated as the cousin of MERS-CoV and SARS-CoV. The diagnosis of MERS-CoV and SARS-CoV involves scientific publications using X-ray images from the chest. Ahmet Hamimi’s study of MERS CoV shows that there are features in the X-ray and CT of the chest which is similar to pneumonia manifestations [33]. Data mining techniques were used by Xu-anyang et al. to differentiate SARS and typical pneumonia based on X-ray images [34]. X-ray devices are used to scan the bones and infections in the body, such as breaks, bone dislocations, lung diseases, pneumonia, and cancers. CT scanning is a kind of advanced X-ray machine that examines the very soft structure of the active part of the body and clearer images of the internal soft tissues and organs [35]. Narin et al. [36] aimed at classifying 4 classes using 3 binary classification models. For achieving this task, they used pre-trained models. Using X-ray images, they

achieved an accuracy of 98%. Brunese et al. [37] defined 2 models based on VGG16. They proposed a 2 level system, the first level to classify affected images from normal images and the second to classify COVID-19 images from affected images. Using an X-ray is a method that is quicker, easier, cheaper, and less harmful than a CT scan. Failure to recognize and treat COVID-19 pneumonia quickly could lead to increased mortality.

A brief review of previous research done for the detection of COVID-19 is shown in Table 1. It contains what method has been used and on what kind of datasets i.e, CT or X-ray along with what accuracy they achieved. The literature states that most of the techniques used pre-trained models without any changes except for the output layer. Moreover, the research done on DenseNet achieved positive accuracy on CT images, but on X-ray images accuracy was low. In addition to that, it was noticed that when compared to DenseNet other models are more complex and have a large number of parameters that need to be tuned. This motivated the use of DenseNet as it is less complex, but to also increase accuracy for X-ray images as an additional dense layer with 512 perceptrons is proposed to be added in the already existing DenseNet model. In previous research, DenseNet models used classical structures and gave average accuracy (80% to 90%) on X-rays. We proposed a DenseNet model with an additional layer of 512 perceptrons, which is expected to give better accuracy.

In this study, we proposed an automatic COVID-19 prediction system using DenseNet (Using an additional Dense layer with 512 perceptrons), pre-trained BiT models, and chest X-ray images. To this end, we have used pre-trained BiT models $R50 \times 1$, $R50 \times 3$, $R101 \times 1$, $R101 \times 3$, and $R152 \times 4$ to obtain a higher prediction accuracy for this large X-ray dataset. DenseNet is more efficient than other models because of its structure, which is deeper, but has less number tuning parameters and low computational complexity than other models due to its bottleneck approach. With an addition of a single Dense layer of 512 perceptrons, it is observed to yield better accuracy on basis of our experiments. The novelty of this paper is summarized as follows: (i) The proposed

Table 2

Training dataset.

Pneumonia	COVID-19	Normal
5463	490	7966

Table 3

Test dataset.

Pneumonia	COVID-19	Normal
594	100	885

models have an end-to-end structure without the extraction and selection methods of manual features. (ii) We demonstrate that DenseNet is an effective pre-trained model amongst two other pre-trained models. (iii) Chest X-ray images are the best tool for COVID-19 detection. (iv) The pre-trained models were shown to yield very high performance on the data-set via simulations.

2. Methods and materials

In this section, the methods and methodology are described in detail.

2.1. Dataset

In this research, chest X-ray images were retrieved from the open-access github repository of 590 COVID-19 patients [38]. This list is comprised primarily of patients with Acute Respiratory Distress Syndrome (ARDS), COVID-19, Respiratory Syndrome of the Middle East (RSME), influenza, and extreme acute respiratory syndrome. In comparison, 6057 chest X-ray images affected by pneumonia were collected from Kaggle's collection named "Chest X-ray Images (Pneumonia)" [39] and 8851 normal X-ray images. The distribution of data is shown in Tables 2 and 3.

It is evident from Tables 2 and 3, that the data are more biased towards normal and pneumonia images. The "COVID-19" training examples are 16 times less likely than the normal Images, which seems a very unbalanced dataset. Without appropriate hyper-parameter tuning, getting positive accuracy on the COVID Images is hard. As without tuning the model, it would only learn to differentiate between "Pneumonia" and "Normal" images while ignoring the "COVID-19" completely.

Possible Solution

One solution is using image sampling, i.e., randomly sample images from the "COVID-19" block and then again put them in the "COVID-19" block. This technique provides positive results, however, not all the time. This technique is used to let the model adjust to the specific images.

Another solution is reducing the size of "Pneumonia" and "Normal" images to level with the "COVID-19" Images. However, it would not be a beneficial choice in this case. There would be no issue if we set a larger number of "COVID-19" images, maybe around 2000. However, in this case, there are only 590 images in total. This would not be a great option because we could lose data.

In this paper, the model learns the features of X-ray images by the "Pneumonia" and "Normal" images, and then uses weighted loss at the end to force the model to better classify the "COVID-19" images. This method of using weighted loss is very promising and achieved positive results. The weighted loss is discussed in detail in Section 2.2.

In Figs. 1 and 2, representative chest X-ray images of normal and COVID-19 patients are given.

Data Preprocessing and Augmentation

All the images were resized to a fixed size of 480×480 pixels. We used different fine-tuned models to identify the most

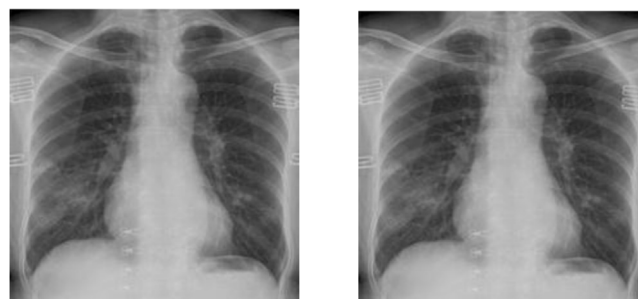
**Fig. 1.** Normal chest X-ray.**Fig. 2.** Covid-19 affected chest X-ray.**Table 4**

Image augmentation parameter details.

Parameter	Value
Samplewise center	True
Samplewise std normalization	True
Horizontal flip	True
Vertical flip	False
Height shift range	0.05
Width shift range	0.1
Rotation range	10
Shear range	0.1
Fill mode	Nearest
Zoom range	0.15

accurate model in them. Later, the images are resized to their original size (higher resolution) and given as input to the identified fine-tuned model to reduce time spent viewing. Besides this, we used *samplewise center* and *sample-wise std normalization* methods for data preprocessing and augmentation. Due to an increase in the training loss, no other data preprocessing or augmentation techniques are applied to the dataset. The training loss of a model trained using image augmentation technique, their parameters and values are shown in Table 4. The history of training loss generated by training the same model on two augmentation techniques is shown in Figs. 3 and 4, respectively.

Fig. 3 shows that when the model is trained on images that are augmented through *samplewise center*, it does not learn, and training loss of the model keeps fluctuating. While Fig. 4 shows the loss curve when the model is trained on images augmented through *samplewise std normalization*, the loss curve keeps on decreasing and fluctuation is also very low. The accuracy obtained by the first method in Fig. 3 was also far less than the second method in Fig. 4. This shows *samplewise std normalization* is better in our experimental setup.

2.1.1. Image augmentation

Image augmentation is the most important technique to increase the samples of the dataset and to allow the model to better generalize to the real-world, data but sometimes it decreases

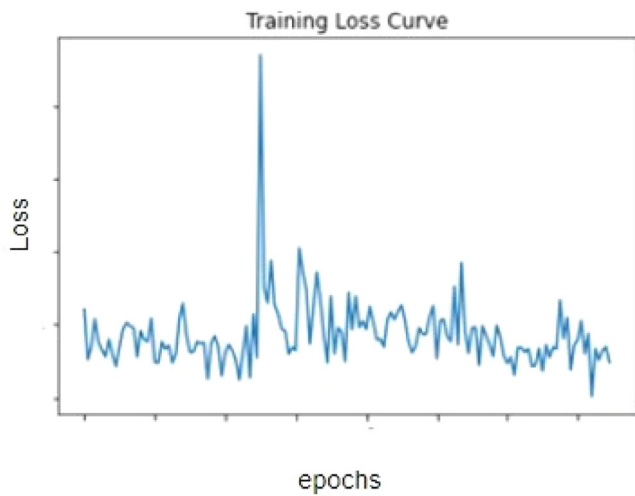


Fig. 3. Training loss curve when *samplewise_center* augmentation is applied (without model learning).

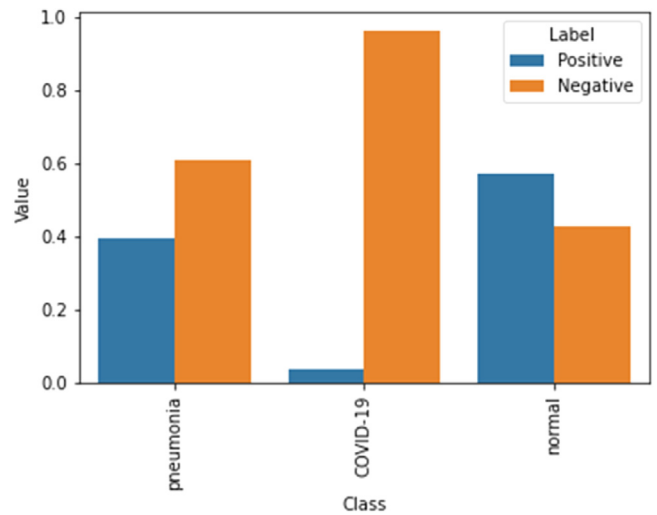


Fig. 5. Contribution before introducing weights for each class.

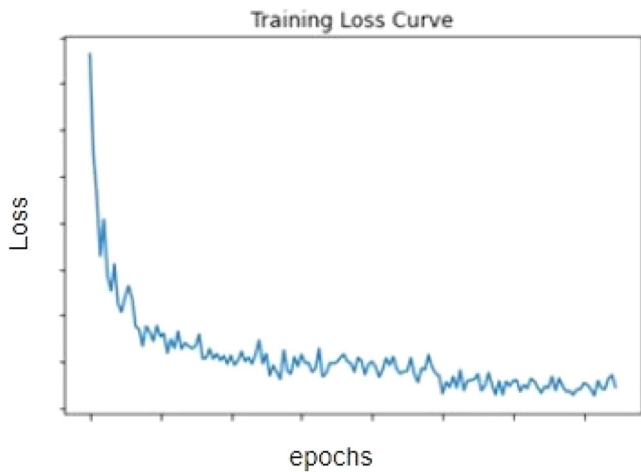


Fig. 4. Training loss curve when *samplewise_std_normalization* augmentation is applied (with model learning).

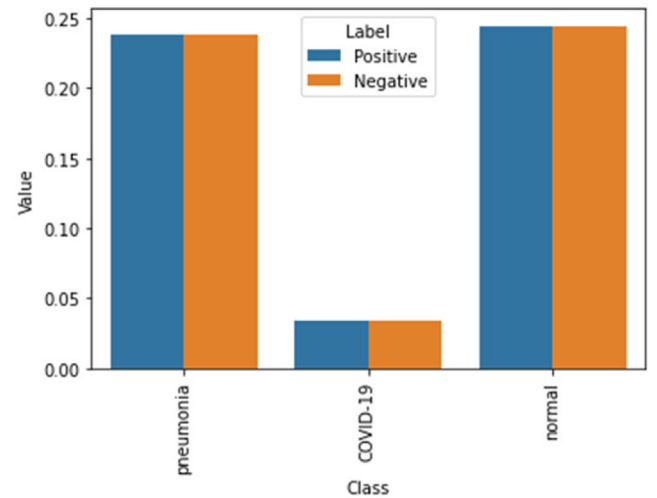


Fig. 6. Contribution after introducing weights for each class.

the model's performance. In our case, we used the Chest X-ray images of different patients. The details in a Chest X-ray are very sensitive, for example the placement of the heart, the relative size of the lungs, the size of the area of the infected part of the lungs. When we apply the augmentation technique as used in the first method, we are unintentionally making changes to the structure of our body. For example, flipping the image of the heart moves it to the other side, that results in a case called Dextrocardia, which is not that common and is useless to teach the model to understand. Moreover, by rotating the image, we are rotating/displaying every major part of our chest. These changes might not be affected when dealing with other types of datasets. However, in this case, it greatly affects the data. Moreover, image augmentation is done to match with the real data. However, the real-world data will also be similar to one that we are using and there will be no rotation, shear, or random zooming. The images will be perfect because we are dealing with a very sensitive type of data.

2.2. Weighted loss

The weighted loss can be described as $N = \text{total number of Images/Samples}$ Positive weights (Pw). = No. of negative examples for each class/ N

Negative weights (Nw) = No. of positive examples for each class/ N

$$L\bar{\sigma} = - (Pw * y * \log(p) + Nw * (1 - y) * \log * (1 - p)) \quad (2.1)$$

It is used to balance out the contribution of positive and negative cases of each class. As seen from Fig. 5, each class has a very uneven distribution of positive and negative cases in terms of percentage. To deal with this problem the contribution of weights are leveled for positive and negative class by introducing weights, as can be seen in Fig. 6. However, there is still a problem that "COVID-19" is not contributing as much as other classes. Weighted loss solves the problem of the balance between positive and negative samples for each class, which is useful and gives positive results; however, there will be a bias towards the other two classes. To deal with this, we introduced a new hyperparameter "c", which will try to balance the contribution of each class. After including this hyperparameter, our loss function becomes:

$N = \text{total number of Images/Samples}$

Table 5

Value of parameter c vs. Variation in accuracy value.

Value of C	Accuracy of the COVID-19 class
Less than 0	Decreases
0-5	Increases
Greater than 5	No learning, Accuracy fluctuates

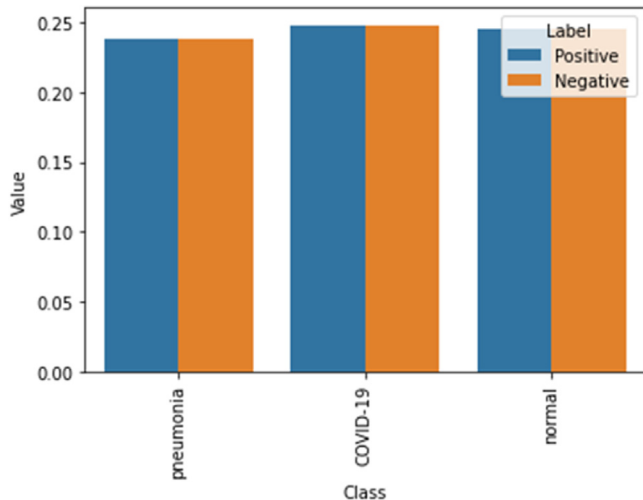


Fig. 7. Labeling after choosing the right value of parameter c .

Negative weights = $Nw = \text{No. of positive examples for each class} / N$
 Updated pos weights = $Pw [i] + c$, where i is the index of the covid class

$$L\bar{\sigma} = -(Pw * y * \log(p) + Nw * (1 - y) * \log * (1 - p)) \quad (2.2)$$

It will add additional weight to the concerned class, in the class of COVID-19. Dealing with this parameter is quite complex. If the value of c is too low then there will be no effect, if it is too high then the loss will be beyond infinity and there will be no learning.

One example of using the right value of c results in Fig. 7. It can be seen that all classes are almost at the same level and contributes equally to training the model.

This might not work in all cases because this may cause the loss to lead to infinity. As in the COVID-19 case, it causes the loss to reach infinity and thus no learning occurs. Many values of c are tested on the dataset along with the base model. The output in the variation of accuracy by the value of c is shown in Table 5. The optimal value of c lies between 0 to 5, at this point distribution of data is such that the overall accuracy of classification increases. Below and above this value distribution is uneven and accuracy decreases or the model does not learn.

In our case, the output of optimal value of c resulted in Fig. 8.

Fig. 8 shows the tuned results of parameter c , which leads to higher accuracy in labeling the data. The addition of weighted loss keeps the percentage of each class from 20% to 25% while training.

2.3. Proposed model and simulations

We have tested many models of architectures and tried out many hyperparameters, some of them are discussed below.

2.3.1. BiT model

BiT also stands for “Big Transfer” is a new CNN-based model released by Google. The main purpose of developing such a model is highlighted by Google itself, “A common method to mitigate the lack of labeled data for computer vision problems is to use

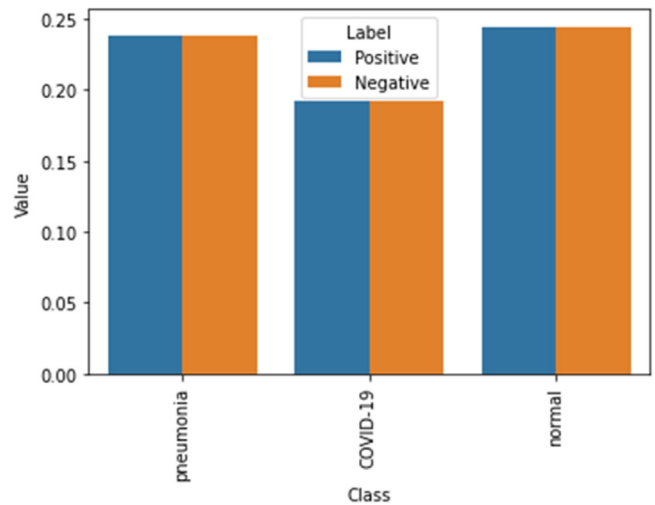


Fig. 8. Different classes after optimal value of parameter c .

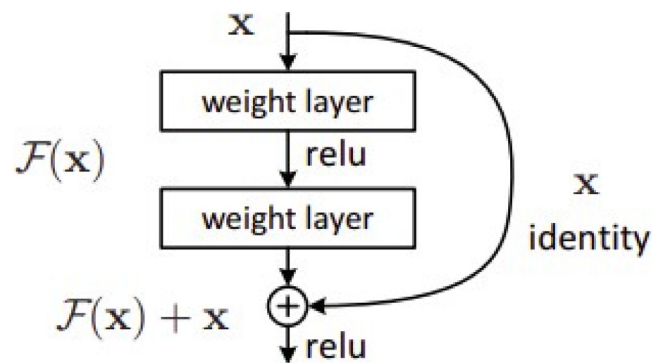


Fig. 9. Residual Layer: Building block of ResNet [40].

methods that have been pre-trained on standard data (e.g., ImageNet). The concept is that visual features learned on standard data can be re-used for the task of interest. Although this pre-training starts working reasonably well in practice, the ability to quickly grasp new concepts still fall short of both”.

To summarize, this model was developed to fine-tune different datasets for achieving state-of-the-art performance on different tasks. We took the original ResNet and increased the dimensions of each layer to create five different types of models, ranked in order of their size “R50 × 1, R50 × 3, R101 × 1, R101 × 3, R152 × 4”. For example, we took R152 and widened each layer by 4. All of the BiT models used in this paper are from the BiT-M category trained on ImageNet-21K.

A ResNet model consists of a residual layer as shown in Fig. 9. The residual layer takes input from previous layers and does not let the model overfit. ResNet takes the residual layer as the building block, the convolutional layer with average pooling layers in the end. The architecture of ResNet is shown in Fig. 10.

For fine-tuning of our data set, we started with the biggest available model, “R152 × 4” trained on ImageNet-21K and added ahead of the size equal to the number of labels at the end. The last layer of this model is only trained. Training the last layer of such a big model is quite a difficult task. When enabled 25 GB RAM and GPU support on Google Collab, it could only process a batch size of length 4. Any Batch size bigger than 4 would result in “OOM”, which cannot allocate the TensorFlow model in memory. However, the model is unable to give any positive results. Our proposed model achieved positive performance results in comparison with the state-of-the-art techniques.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
		$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
		1×1				
FLOPs		1.8×10 ⁹	3.6×10 ⁹	3.8×10 ⁹	7.6×10 ⁹	11.3×10 ⁹

Fig. 10. Architecture of different ResNet models [40].

Table 6

Training overview (Epoch vs. Learning rate).

Epoch	Learning rate
1–20	(0.003 * Batch Size/512)
20–30	(0.003 * Batch Size/512)*0.1
30–40	(0.003 * Batch Size/512)*0.01
40–50	(0.003 * Batch Size/512)*0.001
After 50	(0.003 * Batch Size/512)*0.0001

There are many hyperparameters (learning rate) of the proposed technique involved in the training. However, they are not considered in tuning. Due to the lot of usage of RAM, the learning rate of scheduler usually occurs OOM. So, the proposed model is based on the epoch number for tuning the learning rate. Most of the hyper-parameters, except the loss and the epochs, were initialized using the BiT Hyper rule. The hyper rule is a set of rules that the BiT model has defined for fine-tuning. The overview of the training cycle is described in Table 6. It can be seen that as the number of epochs increases learning rate decreases. If the learning rate is high throughout the training, then the model might cross the optimal value or if the rate is too low, the model will take a large number of epochs to reach the optimal value of the loss. As the model moves towards optimal value, a decrease in learning rate ensures that the model does not cross the optimal value, and the number of epochs until convergence is also not high as the learning rate decreases gradually. At first, we fine-tuned the “R101 × 1” model and skipped the “R101 × 3” model to see a pattern in the relation between data and the model. The model “R101 × 3” was eventually trained as well. However, in the second phase, this model is not large in size as compared to the previous model. So, the batch size is increased to 8 and no OOM occurred during the training. We trained the “R101 × 3” model for a certain time and the accuracy stopped increasing after 80 epochs and reached its minima. Nevertheless, a boost in accuracy is achieved and tried by adding one extra layer at the end before the output layer,. However, it remains the same. Additionally, we tried using dropout and ‘Batch Normalization’, but the training accuracy decreased instead of increasing.

After all, we fine-tuned the base model, “R101 × 1” on our dataset by keeping the learning rate very low and despite a very small dataset. This technique allowed us to achieve the best results. This technique also introduced a small amount of conflict with the table we created in the section for weighted loss. Also, we observed that with the increase in accuracy for the COVID class, the overall performance of the model decreased by a slight decline of about 1%.

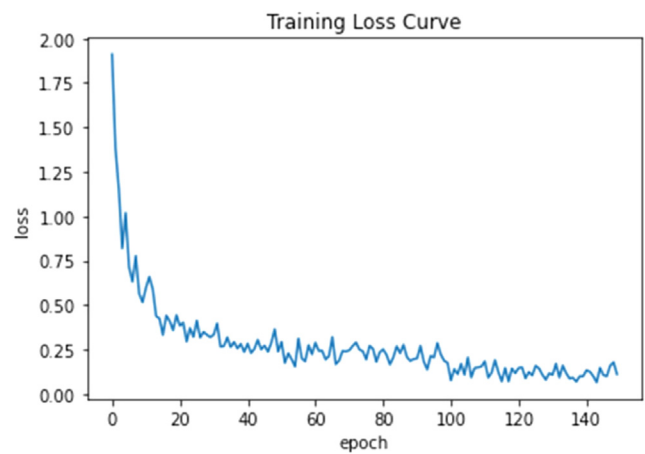


Fig. 11. Training loss vs. Epoch.

After that, we fine-tuned the “R101 × 3” model using the above technique and also fine-tune the whole model. The pattern and this model are unable to outperform the “R101 × 1” model but certainly outperformed the “R152 × 4” model. Furthermore, we fine-tuned both “R50 × 3” and “R50 × 1” models, but none could outperform “R101 × 1”. If we increase the model size from “R101 × 1”, accuracy starts to decrease, and if the model size is decreased accuracy further decreases (less than “R101 × 3”).

2.3.2. DenseNet and inception V3

The main focus of this paper is to get the most out of the BiT model. Furthermore, it is quite useful to test other state-of-the-art models. We then fine-tuned a DenseNet model trained on the ImageNet dataset. We added a fully connected layer with 512 units at the end, and after that added the output layer which is equal to the size of labels. We first fine-tuned the model by freezing the top layers of the model for 100 epochs. After that, we fine-tuned the whole model, including the base model, with a comparatively low learning rate for 150 epochs. By using the weighted loss and proper augmentation, the overall accuracy of the model is greater than our previous best model. However, the accuracy of the COVID class is far less than the best accuracy. Due to the potential of the model and increase of the value of the hyperparameter “c” in the weighted loss to 5.5 and further trained it for 100 epochs to force the model to learn to better classify COVID images. This model outperformed the previous model, however, it was unable to achieve accuracy in detecting

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

Fig. 12. Basic architecture of DenseNet models [41].

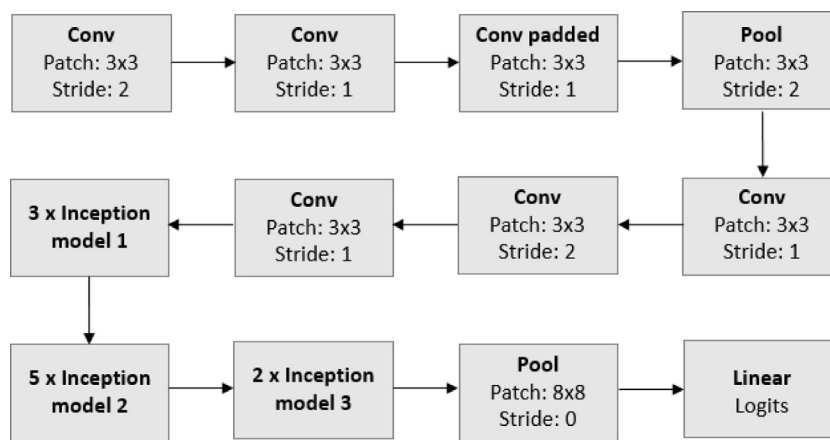


Fig. 13. Inception-V3 architecture.

COVID-19. After the implementation of DenseNet, we tried to tune the Inceptionv3 model. The result is the same because model size, dataset size, and the nature of the dataset impact a number of data. Even though Inceptionv3 is a much bigger model as compared to DenseNet. The architecture of different DenseNet models and InceptionV3 is shown in Figs. 12 and 13 respectively.

The accuracy of the top 4 models is compared and described in Table 7. The top 2 models, in this case, are DenseNet and enhanced R101 × 1 with over-all training and testing accuracy greater than 90%. Other models have accuracy smaller than 90%. In terms of COVID-19 detection, R101 × 1 surpasses all the other models.

Inceptionv3 accuracy is much less than other models. We have further compared the results of the top two models more closely in Table 8. DenseNet and R101 × 1 are very close to each other in terms of everything but in terms of detecting COVID-19 R101 × 1 is best. In terms of other classes, R101 × 1 shows 91.08% prediction accuracy for pneumonia and 91.18% prediction accuracy for normal class. While DenseNet shows 93.26% and 92.43% prediction accuracy for pneumonia and normal class respectively.

Fig. 11 shows the loss curve of the DenseNet model, which can be seen as steady in the end. This shows that the DenseNet model has converged successfully and in a better way. In the overall scenario, the DenseNet model performs better and have the same Receiver Operating Characteristic Curve(ROC) Area Under Curve(AUC) score as compared to BiT model as can be seen from Figs. 14 and 15. However, in the case of COVID-19, the BiT model outperforms. Figs. 14 and 15 shows ROC curves and AUC values for detection of three classes under study for R101 × 1 and DenseNet respectively.

2.3.3. Insights from the model training

Five different models (50 × 1, R50 × 3, R101 × 1, R101 × 3, R152 × 4) were trained for the classification of COVID-19, normal, and pneumonia X-ray images. Among those models R152 × 4, R101 × 3, and R101 × 1 showed promising results and gave test accuracy of 85.8%, 83.47%, and 91.13%. The first insight from testing these models is that fine-tuning big models do not always help. Consider the first case, when we fine-tuned the largest available BiT Model, i.e., “R152 × 4”, which was pre-trained on the ImageNet-21K dataset. However, it did not perform well on our dataset. The reason is too biased towards the distribution of the Images in the ImageNet-21K dataset. Also, the chest X-ray dataset is not very different when it comes to feature extraction but the distribution of the data, in general, is a lot different.

On the other hand, models below a certain range of size also do not help. In the case of “R50 × 3”, the model accuracy further drops. However, when model size is reduced further i.e. in the case of R101 × 1, the model accuracy increases.

Despite an increase in the accuracy, the model remains complex and there are many parameters to be tuned in the case of R101 × 1. To reduce the model complexity and save resources, DenseNet model was used but there are many pieces of research in literature that used DenseNet and showed that they are less complex but do not gives state-of-the-art accuracy. Taking this into consideration, we modified DenseNet by adding a dense layer of 512 perceptrons and saw a boost in the accuracy. Our modified DenseNet showed a test accuracy of 92% along with the advantage that the model is not complex and costs less.

Table 7
Comparison of models training and testing accuracy.

Model	Average training accuracy	Test accuracy	Accuracy of the COVID class
R152 × 4 (N^)	86%	85.8%	~75%
R101 × 3	87%	83.47%	90%
R101 × 1	93%	91.13%	91%
DenseNet (Using an additional Dense layer with 512 perceptrons)	93.5%	92%	85%

Table 8
Comparison of models ROC and AUC.

Model	Sensitivity and specificity (percentage)	Confusion matrix	Accuracy of the COVID class (percentage)	Accuracy of the pneumonia class (percentage)	Accuracy of the normal class (percentage)	ROC AUC Score (percentage)
R101 × 1	91 and 96	541, 16, 37, 5, 91, 4, 49, 29, 807	91	91.08	91.19	98
DenseNet (Using an additional Dense layer with 512 perceptrons)	85 and 99	554, 2, 38, 12, 85, 3, 63, 4, 818	85	93.27	92.43	98

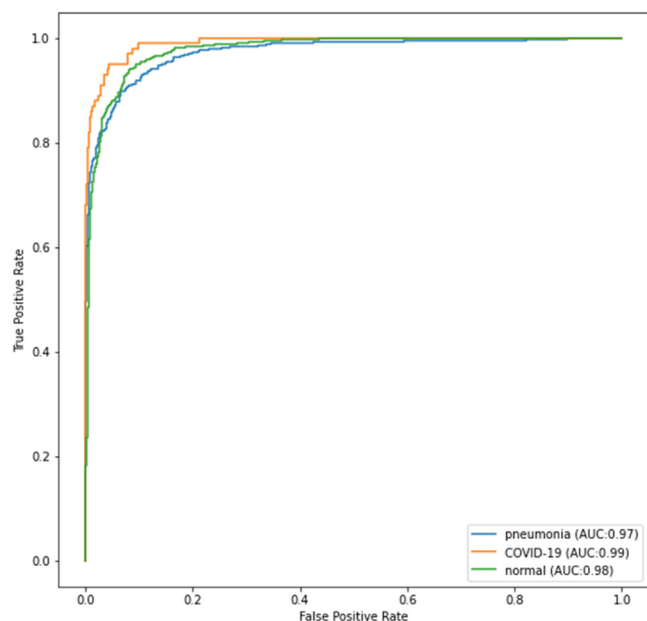


Fig. 14. R101 × 1 TPR vs. FPR.

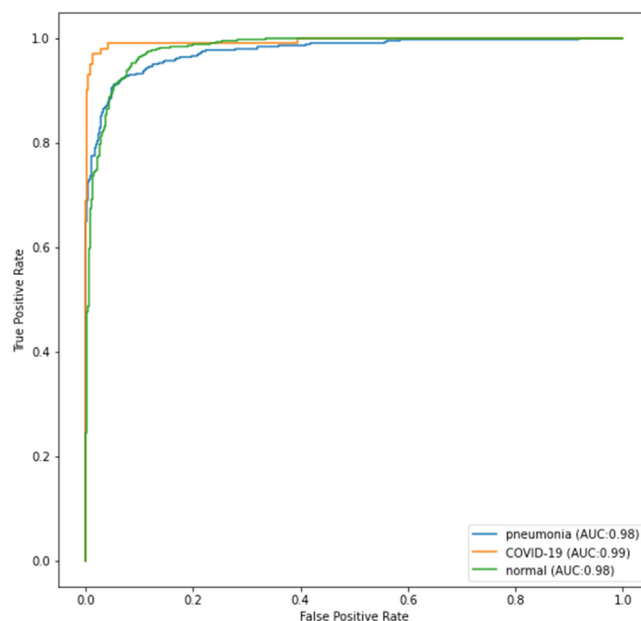


Fig. 15. DenseNet TPR vs. FPR.

In the end, two models were compared more thoroughly from all the models as they gave comparatively higher training and testing accuracy. R101 × 1 and modified DenseNet are those two models. For COVID-19, R101 × 1 gave higher test accuracy, but DenseNet showed higher accuracy for the other two classes (normal and pneumonia). The explained comparison on basis of accuracy and ROC AUC score can be seen in Table 8.

2.3.4. Key points

- The fine-tuning of the BiT model is easy on most data sets rather than other models and also performs better.
- Fine-tuning of BiT models also does not take a large amount of time. It could get the maximum possible accuracy after 40 epochs.
- Fine-tuning BiT models is easy as compared to fine-tuning other models, such as DenseNet. All the hyperparameters can be initialized using the official BiT hyper rule.

- Fine-tuning models other than the BiT are sometimes complex. Additional layers may need to be added at the end and have to be carefully tested at different learning rates to make sure the number of epochs is enough. Sometimes, training for more epochs results in less accuracy.

All of the results in this research study have been obtained after a great amount of hyperparameter tuning.

2.3.5. Comparison with other models

A great amount of models trained in the literature showed promising training and testing accuracy for COVID-19 classification as shown in Table 1. Most models that were used gave promising accuracy for COVID-classification were either trained on CT-Scan images or were acting as binary classifiers. Binary classifiers have fewer classes to predict and thus show better accuracy than models trained on more classes. Other problems with previous models were that if they achieved 90%+ accuracy

then they were more complex than other models, thus they were computationally costly.

We have obtained superior performance as compared to many other studies in the literature with more classes, less COVID-19 dataset, and a less complex model. Previous research was also at a disadvantage because COVID-19 data is not in abundance and those research models likely used less data than we used. The size of COVID-19 datasets is increasing day by day and research in this domain is enhancing every day, thus it is not fair to compare those research models with ours when it is compared to those with fewer data. However, with an AUC ROC score of 98% and accuracy of 91% on the COVID class, our results out-performed the state-of-the-art as shown in Table 1 using $R101 \times 1$.

3. Conclusion

Early diagnosis of COVID-19 patients is important for preventing the disease from spreading to others. In this study, we proposed a deep transfer learning method based on chest X-ray images that were collected from COVID-19 patients. We trained our proposed system and tested it on test data. Performance results show that the DenseNet pre-trained model has yielded the highest accuracy up to 92% between the three models. Due to the high results, it is assumed in the light of our findings that it would help doctors to make decisions in clinical practice. This study provides insight into how deep transfer learning approaches can be used to detect COVID-19 at an early stage. The classification efficiency of various BiT models can be evaluated in subsequent experiments by the number of images in the dataset.

CRedit authorship contribution statement

Saleh Albahli: Conceptualization, Methodology, Software, Writing - original draft. **Nasir Ayub:** Data curation, Writing - original draft. **Muhammad Shiraz:** Validation, Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J.M. Hyman, P. Yan, G. Chowell, Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020, *Infect. Dis. Model.* 5 (2020) 256–263.
- [2] L. Yan, H.T. Zhang, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, S. Li, M. Zhang, Y. Xiao, H. Cao, Y. Chen, T. Ren, J. Jin, F. Wang, Y. Xiao, S. Huang, X. Tan, N. Huang, B. Jiao, Y. Zhang, A. Luo, Z. Cao, H. Xu, Y. Yuan, Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in wuhan, 2020, medRxiv 2020.02.27.20028027.
- [3] S.B. Stoecklin, P. Rolland, Y. Silue, A. Mailles, C. Campese, A. Simondon, M. Mechain, L. Meurice, M. Nguyen, C. Bassi, E. Yamani, S. Behillil, S. Ismael, D. Nguyen, D. Malvy, F.X. Lescure, S. Georges, C. Lazarus, A. Taba'i, M. Stempfelet, V. Enouf, B. Coignard, D. Levy-Bruhl, and Team, I., First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, 2020, *Eurosurveillance* 25 (6) (2020) 2000094.
- [4] G. Gabutti, E. d'Anchera, F. Sandri, M. Savio, A. Stefanati, Coronavirus: update related to the current outbreak of COVID-19, *Infect. Dis. Ther.* (2020) 1–13.
- [5] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet* 395 (10223) (2020) 497–506.
- [6] A. McKeever, Here's what coronavirus does to the body, 2020, National Geographic: <https://www.nationalgeographic.com/science/2020/02/here-is-what-coronavirus-does-to-the-body/>.
- [7] Covid-19 coronavirus pandemic, 2020, worldometer: <https://www.worldometers.info/coronavirus/>.
- [8] E. Mahase, Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate, *BMJ* 368 (2020) m641, <http://dx.doi.org/10.1136/bmj.m641>.
- [9] Y. Wang, M. Hu, Q. Li, X.P. Zhang, G. Zhai, N. Yao, Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner, 2020, arXiv:2002.05534.
- [10] S. Albahli, G.N.A.H. Yar, Fast and accurate COVID-19 detection along with 14 other chest pathology using: Multi-level classification, *J. Med. Internet Res.* (2021) <http://dx.doi.org/10.2196/23693>, (forthcoming).
- [11] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks, *Phys. Eng. Sci. Med.* 1 (2020).
- [12] L. Wang, A. Wong, Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images, 2020, arXiv preprint arXiv:2003.09871.
- [13] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, Y. Jie, R. Wang, et al., Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images, 2020, medRxiv.
- [14] L. Sarker, M.M. Islam, T. Hannan, Z. Ahmed, Covid-densenet: A deep learning architecture to detect covid-19 from chest radiology images, 2021.
- [15] N. Hasan, Y. Bao, A. Shawon, DenseNet convolutional neural networks application for predicting COVID-19 using CT image, 2020.
- [16] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, X. Wang, Deep learning-based detection for covid-19 from chest ct using weak label, 2020, medRxiv.
- [17] X. Xu, et al., A deep learning system to screen novel coronavirus disease 2019 pneumonia, *Engineering* 6 (10) (2020) 1122–1129.
- [18] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G.J. Soufi, Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning, *Med. Image Anal.* 65 (2020) 101794.
- [19] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of covid-19 cases using deep neural networks with X-ray images, *Comput. Biol. Med.* (2020) 103792.
- [20] A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Moham madi, Application of deep learning technique to manage covid-19 in routine clinical practice using ct images: Results of 10 convolutional neural networks, *Comput. Biol. Med.* (2020) 103795.
- [21] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al., Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct, *Radiology* (2020) 200905.
- [22] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, P. Xie, COVID-CT-dataset: A CT image dataset about COVID-19, 2020, arXiv preprint arXiv:2003.13865.
- [23] A. Abbas, M.M. Abdelsamea, M.M. Gaber, Classification of covid-19 in chest X-ray images using detrac deep convolutional neural network, 2020, arXiv preprint arXiv:2003.13815.
- [24] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng, et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study, 2020, MedRxiv.
- [25] Covid-19 situation update worldwide, as of 10 July 2020, 2020, European Centre for Disease Prevention and Control: <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>.
- [26] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, X. Tao, Z. Sun, L. Xia, Correlation of Chest CT and RT-PCR testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 cases, *Radiology* 200642 (2020) 1–23, <http://dx.doi.org/10.1148/radiol.2020200642>.
- [27] W. Zhang, Imaging changes of severe COVID-19 pneumonia in advanced stage, *Intensive Care Med.* (2020) 1–3, <http://dx.doi.org/10.1007/s00134-020-05990-y>.
- [28] O. Gozes, M. Frid-Adar, H. Greenspan, P.D. Browning, H. Zhang, W. Ji, A. Bernheim, E. Siegel, Rapid AI development Cycle for the coronavirus (COVID-19) Pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis, 2020, arXiv:2003.05037.
- [29] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, Y. Shi, Lung infection quantification of COVID-19 in CT images with deep learning, 2020, arXiv:2003.04655.
- [30] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Y. Chen, J. Su, G. Lang, Y. Li, H. Zhao, K. Xu, L. Ruan, W. Wu, Deep learning system to Screen Coronavirus Disease 2019 Pneumonia, 2019, arXiv:2002.09334.
- [31] C. Butt, J. Gill, D. Chun, B.A. Babu, Deep learning system to screen coronavirus disease 2019 pneumonia, *Appl. Intell.* (2020) 1.
- [32] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, B. Xu, A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19), 2020, medRxiv 2020.02.14.20023028.
- [33] A. Hamimi, MERS-CoV: Middle East respiratory syndrome corona virus: Can radiology be of help? Initial single center experience, *Egypt. J. Radiol. Nucl. Med.* 47 (1) (2016) 95–106.

- [34] X. Xie, X. Li, S. Wan, Y. Gong, in: *Graham J. Williams, Simeon J. Simoff (Eds.), Mining X-Ray Images of SARS Patients. Data Mining: Theory, Methodology, Techniques, and Applications*, Springer-Verlag, Berlin, Heidelberg, ISBN: 3540325476, 2006, pp. 282–294.
- [35] A. Tahamtan, A. Ardebili, Real-time RT-PCR in COVID- 19 detection: issues affecting the results, 2020.
- [36] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of Coronavirus disease (covid-19) using X-ray images and deep convolutional neural networks, 2020, arXiv preprint [arXiv:2003.10849](https://arxiv.org/abs/2003.10849).
- [37] L. Brunese, F. Mercaldo, A. Reginelli, A. Santone, Explainable deep learning for pulmonary disease and Coronavirus COVID- 19 detection from X-rays, *Comput. Methods Programs Biomed.* 196 (2020) 105608.
- [38] M.M. Ahsan, K.D. Gupta, M.M. Islam, S. Sen, M. Rahman, M.S. Hossain, Study of different deep learning approach with explainable ai for screening patients with COVID- 19 symptoms: Using ct scan and chest x-ray image dataset, 2020, arXiv preprint [arXiv:2007.12525](https://arxiv.org/abs/2007.12525).
- [39] M.S. Boudrioua, COVID-19 detection from chest X-ray images using CNNs models: Further evidence from Deep Transfer Learning, 2020, Available at SSRN 3630150.
- [40] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [41] G. Huang, Z. Liu, L. VanDerMaaten, K.Q. Weinberger, Densely connected convolutional networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.