



# Deep learning based detection of COVID-19 from chest X-ray images

Sarra Guefrechi<sup>1</sup> · Marwa Ben Jabra<sup>2,3</sup> · Adel Ammar<sup>4</sup> · Anis Koubaa<sup>4,5,6</sup> · Habib Hamam<sup>1</sup>

Received: 12 October 2020 / Revised: 19 May 2021 / Accepted: 24 June 2021 /  
Published online: 19 July 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

The whole world is facing a health crisis, that is unique in its kind, due to the COVID-19 pandemic. As the coronavirus continues spreading, researchers are concerned by providing or help provide solutions to save lives and to stop the pandemic outbreak. Among others, artificial intelligence (AI) has been adapted to address the challenges caused by pandemic. In this article, we design a deep learning system to extract features and detect COVID-19 from chest X-ray images. Three powerful networks, namely ResNet50, InceptionV3, and VGG16, have been fine-tuned on an enhanced dataset, which was constructed by collecting COVID-19 and normal chest X-ray images from different public databases. We applied data augmentation techniques to artificially generate a large number of chest X-ray images: Random Rotation with an angle between  $-10$  and  $10$  degrees, random noise, and horizontal flips. Experimental results are encouraging: the proposed models reached an accuracy of 97.20 % for Resnet50, 98.10 % for InceptionV3, and 98.30 % for VGG16 in classifying chest X-ray images as Normal or COVID-19. The results show that transfer learning is proven to be effective, showing strong performance and easy-to-deploy COVID-19 detection methods. This enables automatizing the process of analyzing X-ray images with high accuracy and it can also be used in cases where the materials and RT-PCR tests are limited.

**Keywords** Deep learning · COVID-19 · Convolution Neural Network · CNN · Chest X-ray

## 1 Introduction

The novel coronavirus disease first appeared in the city of Wuhan, China, in late 2019. While China's government took a number of precautionary measures, including a lockdown of the city, to limit the spread of COVID-19 [14], it had, by then, been too late to contain the virus.

---

✉ Sarra Guefrechi  
guefrechisarra@gmail.com

By early spring 2020, the virus had been reported in most countries, and by the end of March 2020, the World Health Organization (WHO) officially declared the new virus a pandemic [8]. According to data released by the WHO, by April 30, 2021, more than 157 million people have been affected with the virus and more than 3 million had lost their lives to the disease [31].

While coronaviruses are not novel, SARS-CoV-2 is not standard [28]. This virus is very probably to have originated from an animal reservoir [32]. Patients with COVID-19 are assessed and cared for using different treatments than those applied for other coronavirus transmissions [33]. But our knowledge of the disease remains limited and is expanding simultaneously with the pandemic [24]. Commonly reported symptoms of COVID-19 include fever, coughing, tiredness, a sore throat and body aches, while numerous accounts of loss of taste or scent have been reported across the globe. In rarer but typically more severe cases, patients can experience difficulty breathing, a high fever, chills, fatigue, muscle or body aches or even death [34].

The standard COVID-19 test is known as the polymerase chain reaction (PCR) test and is used to detect the existence of infection antibodies. Unfortunately, these tests demand high precision and are time-consuming with a non-negligible possibility of false negatives [4]. It goes without saying that an erroneous conclusion of the absence of the virus can lead to drastic results and is counterintuitive to governments' efforts to restrict the spread of the virus.

Moreover, many countries lack the adequate resources to implement COVID-19 tests and testing sites on a large scale. To bypass such issues, radiography chest image analysis is considered an alternative method to the PCR test.

Artificial Intelligence (AI) models can be an apt solution [15]. Thanks to its great accuracy, the deep learning approach has been widely welcomed and successful for medical image classification applications. Many recent works based on deep learning technology have promoted the development of intelligent diagnostic systems, which can help human experts make better decisions about patients' health. Lopez et al. try to raise the problem of skin lesion classification in their paper [23], especially the identification of early melanoma, and propose a deep learning method to solve the problem of image classification and identification of skin lesions whether they are malignant or benign.

A survey from the latest UN Global Pulse assessment of the application of AI to COVID-19-related needs [10] shows that, compared with standard testing, AI has potential for human accuracy and may significantly save radiologists' time and effort. Thus, the potential for a cheaper and more timely diagnosis cannot be overlooked [6]. Both computed tomography (CT) and X-ray can be used [1].

Due to considerable sustained achievement in machine learning, especially in statistical learning that integrates big data [12] and the important interest of interpretable AI in medicine [13], AI and deep learning can improve COVID-19: discovery and identification. The main challenge is to detect COVID-19 with an accurate and low-cost detection methods. Convolutional Neural Networks (CNN) have been demonstrated to be very powerful in including extraction and learning, so they are generally adopted by researchers [17]. The purpose of this work is to establish a fully automated system for the classification of COVID-19 and non-COVID-19 pneumonia. We have trained three popular convolutional networks on the elaborated dataset. These networks (VGG16, ResNet50, and InceptionV3) have achieved compelling results in some tasks in recent years. We fine-tune them for the purpose of the detection of COVID-19. As of today, only a limited set of X-ray images relating to COVID-19 inquiries is available for public use. Thus, we could not train these models from scratch. In this work, we adopt two strategies to solve the COVID-19 image shortage problem:

- we apply data augmentation in order to build a converted version of the COVID-19 image (e.g., flip, small rotation, small distortion, etc.) to triple the set of samples.
- we fine-tuned the last layer of the model; thus, we can use less labeled samples per category for the training process.

The originality of our work may be summarized as follows. First, it consists in developing an augmented dataset by implementing three augmentation strategies: rotation, random noise and horizontal flips that are appropriate for deep learning. Second, we fine-tuned the last layer of three commonly used powerful algorithms to detect the virus from noisy chest images efficiently. It is worth mentioning that VGGnet is a robust, flexible architecture for benchmarking on a particular task. In contrast, the Inception and Resnet models perform superior accuracy and higher efficiency than existing CNNs [16]. Third, we made a loop over all layers of the network and suspended their update during the first training process (to avoid the destruction of the information contained during future training rounds). Fourth, our technique is of nature to be used as an authentic tool for clinical decision support.

The remainder of the paper is ordered as follows. After the present Introduction, Section 2 will overview related work. Section 3 analyzes how the dataset is developed and describes the overall proposed framework. We provide experimental studies as well as comparisons with previous work in Section 4. Then, the article is concluded in Section 5.

## 2 Related work

Some of the latest research on diagnosing COVID-19 requires the application of various deep learning methods. The novelty of COVID-19 and the consequent unavailability of large data sets have forced most researchers to use transfer learning. Our goal is to reduce false positives and false negatives by implementing three enhancement strategies on the chest X-ray images we collected, and by using the transfer learning process of three Convolutional Neural Network (CNN) on the enhanced data set.

There are several works that use transfer learning on chest X-ray images to recognize patients suffered from COVID-19. Here we limit our attention to those closely related to our proposal.

Ioannis et al. [3] assess the performance of the latest CNN architectures used in recent years for medical image classification, namely VGG19, MobileNet v2, Xception, Inception-ResNet-v2 and Inception. The author uses transfer learning because it performs well in detecting various abnormalities in small medical image data sets [9]. They used 1,442 X-ray data sets from patients, including 714 cases of bacterial pneumonia and viral pneumonia, and 224 cases of confirmed COVID-19 disease and 504 healthy cases. The results show that among the remaining CNNs, MobileNet-v2 and VGG19 offer the best classification in terms of accuracy. While VGG19's outperforms the other techniques in terms of accuracy (reaching 98.75%), MobileNet-v2 shows better performance regarding the sensitivity and the specificity (reaching 99.10 and 97.09%, respectively).

Narin and others solved the problem of the limited supply of COVID-19 test kits accessible in public hospitals. [20] proposes to apply an automatic detection system as another rapid diagnosis recourse to avert the spread of COVID-19 and its pressure on medical institutions. In order to detect patients infected with coronavirus pneumonia, the author proposed three CNN-based models (Inception-ResNetV2, InceptionV3 and ResNetV2), using totally 100 chest X-

ray images (50 health images and 50 COVID-19 images). In view of the high-performance results obtained, we can conclude that compared with the other two proposed models, the pre-trained ResNet50 model achieves 98 % accuracy (InceptionV3 reach an accuracy of 97 % and Inception-ResNetV2 reach only an accuracy of 87 %).

For the sake of accurately detect COVID-19 and help overcome the lack of specialist doctors in remote villages, Ozturk et al. [21] propose a new model that uses raw radiographic images to automatically detect COVID-19. Among the proposed models, the DarkNet model is utilized as the classifier of the real-time object detection system (YOLO), conceived of 17 convolutional layers. The authors proceed by different filtering at the level of each layer. The purpose of this algorithm is to offer an accurate diagnosis for two-class classification (COVID/no-findings) and a multi-class classification (COVID/no-findings/pneumonia). The classification accuracy for the binary classification is 98.08 %, and for the multi-classification is 87.02 %.

Sethy and Behra [25] propose a method founded on deep features as well as support vector machine (SVM) to detect patients with a coronavirus infection based on X-ray images. Instead of deep learning, they use SVM based classifiers for classification. They extract the deep features from the fully connected layers of the CNN model. Then they enter them into SVM for classification. SVM classifies X-ray images affected by coronavirus. The method includes three types of X-ray images, namely COVID-19, pneumonia and normal X-ray images. The author evaluates SVM to detect COVID-19 by using the deep functions of 13 different CNN models. By using the deep features of ResNet50, SVM can produce the best results. The highest accuracy reached by ResNet50 and SVM is 98.66 %.

Wang et al. [30] propose a CNN model (COVID-Net) to diagnose the coronavirus cases from chest X-rays. The author used the COVIDx chest X-ray data set, which contains only 76 X-ray photos of COVID-19 cases, 8066 normal and 5526 non-COVID-19 pneumonia patients. The COVID-Net utilizes the lightweight residual projection extension-projection extension (PEPX) design pattern. By achieving a test accuracy of 92.4 %, and only 2.26 billion MAC operations are required to perform case prediction, a good compromise is achieved between accuracy and computational effort and complexity.

There is other several works that use CT scan rather than chest x-ray images in identify COVID-19.

For instance, Xu et al. [35] find that the computer tomography (CT) imaging characteristics of this new virus are not the same as the other types of viral pneumonia. They use various CNN models to classify computed tomography images, calculate the possibility of COVID-19 infection, and assist in the early identification of COVID-19. They collect a total of 618 CT images: 175 healthy people, 219 of patients infected with COVID-19 and 224 CT samples of influenza A virus pneumonia. They use a 3D CNN model based on the classic ResNet-18 network structure to segment many candidate images cubes. The author uses a 3D image classification model to classify all image blocks. This classification model aims to understand the relative position information of plaques on lung images. In addition, they use Noisy or Bayesian functions to calculate the type of infection (COVID-19, influenza A viral pneumonia, or no infection found) and the total confidence score of CT cases. The overall classification accuracy of the proposed model for the three groups reached 86.7 %.

Shan et al. [26] suggest to develop a deep learning system, named “VB Net”, to automatically segment and quantify the coronavirus infected zones in in computer tomography images. The “VB Net” model is a combination between the V-Net model and the bottleneck model. V-Net uses down-sampling and convolution to extract global image features, while the bottleneck model uses up-sampling and convolution to merge fine-grained image features. The system

was trained using data from 249 patients infected with COVID-19 and validated in 300 new patients affected with the COVID 19 virus. In order to speed up the process of depicting COVID-19 CT images for training, which is very time-consuming, the author proposes a “in loop (HITL)” strategy, which aims to generate training samples in an iterative manner. After three iterations of the model update, the recommended manual loop strategy reduces the drawing time to 4 min. The Dice correspondence coefficient is used to evaluate the subdivision accuracy of the deep learning model in the entire 300 verification set which reaches 91.6 %.

### 3 Method

In this study, we train, evaluate and test three well-known pre-trained deep learning architectures to classify chest radiography images for a two-class classification: COVID-19 and non-COVID-19 chest X-ray. VGG16 [27], ResNet50 [29] and InceptionV3 [11] are mainly used as deep learning models.

Figure 1 provides a general overview of the methodology of this study that describes the deep learning proposed method based on a simple standard pipeline, that is, chest image preprocessing, and then the classification model obtained through transfer learning. After preprocessing the data, a deep model is trained. For instance, we will do a fine-tuning which includes an unfreezing of some of the top layers of the frozen model that is utilized for feature extraction, and then training both the freshly added part of the model (in our experiment, the fully-connected classifier) and the unfreezing top layers.

Freezing means that we don’t want to update the weights of these layers when we train the model on new data for new tasks. We want all these weights to remain the same as the weights after training on the original task. We only want to update the new layer or modify the weights in the layer.

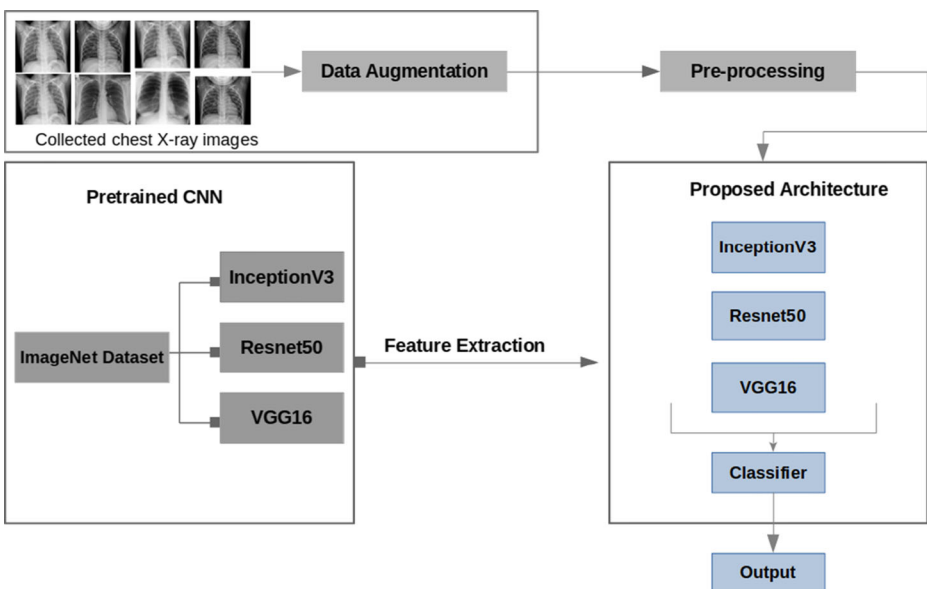


Fig. 1 Overview for COVID-19 and non-COVID-19 Chest X-ray images classification

After this is done, all that remains is to train the model on the new data. Likewise, during this training process, the weights of all layers we retained from the original model will remain unchanged, and only the weights in the new layer will be updated.

### 3.1 Dataset

#### 3.1.1 Creation

We start by preparing the dataset, as it is the first step to apply deep learning. As COVID-19 addresses the epithelial cells lining our airways, we use a chest X-ray image in order to analyze the health of a patient's lung.

In this article, we adopt chest X-ray images instead of computer tomography scans to fine-tune the three proposed classification models. Compared with higher radiation exposure, time-consuming CT scans, and expensive, X-rays are a lot cheaper, faster, lower doses for the patient and more available. In addition, portable X-ray machines can be tested in isolation wards, thereby reducing the risk of hospital infections and reducing the number of personal protective equipment used.

Furthermore, chest X-ray image analysis is a practical alternative to the PCR method. They can provide a variety of assistance from the discovery of the disease to the selection of high-risk patients for isolation and prioritization, as well as selective testing to identify false-negative PCR cases, they can provide a variety of help. However, because most cases of viral pneumonia are similar and overlapping, it is hard for radiologists and doctors to distinguish adequate details visually, and it is very time-consuming. Using the deep learning models can be an accurate solution.

In our experiment, we focus on reducing false positives and false negatives by using the transfer learning process with 3 Convolutional Neural Network (CNN) on an augmented dataset by implementing three augmentation strategies on our collected chest X-ray images.

The constructed dataset for this work contains a total of 5000 images:

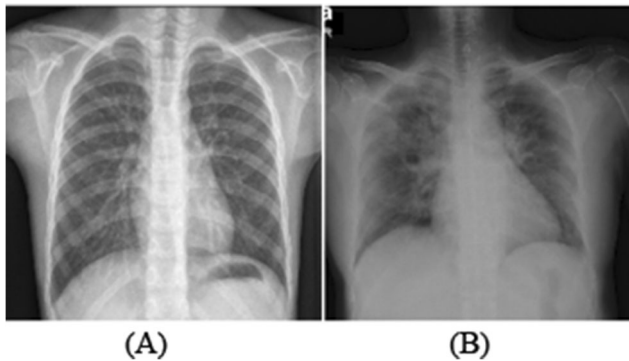
- 3000 images for normal chest X-ray were selected from different public image databases: Kaggle repositories “Chest X-Ray Images (Pneumonia)” [7] as well as “Covid-19 Radiography Dataset” [19].
- 623 chest X-ray COVID-19 images were collected from the GitHub repository [22], Covid-19 Radiography Data Set [19]. Therefore, we used image augmentation to expand the size of the total number to 2000 images.

Table 1 below presents the content of the prepared dataset, that was divided into two folders namely Normal and COVID-19.

Figure 2 illustrates two examples of chest X-ray images taken from our prepared dataset.

**Table 1** Content of our prepared dataset

	Without data augmentation	With data augmentation
COVID-19	623	2000
Normal	3000	3000
Total of images	3623	5000



**Fig. 2** (A): Chest X-ray image of a healthy person. (B): COVID-19 chest X-ray image

### 3.1.2 Data augmentation

Deep learning models usually demand a significant volume of training data. Image augmentation technology has been generally used in computer vision and has attracted attention since the advent of deep learning [5].

The more the data, the better the model's performance, and because COVID-19 is still a newly emerging disease, so far, no appropriate dataset is publicly available. Therefore, we have to use data augmentation as it is a very powerful technique to artificially generate a large dataset.

We apply three augmentation strategies: Random Rotation with an angle between  $-10$  and  $10$  degrees, random noise, and horizontal flips (means reversing the columns of pixels).

In fact, image noising is a significant way that permits our model to figure out how to isolate signal from noise in an image. This makes the model more powerful to changes in the information.

### 3.1.3 Data preprocessing

During data preprocessing, it is possible to resize the X-ray images. It's due to the fact that the various algorithms require different image inputs. The images should be normalized according to the given model standards.

The input images were in different original size, then they were all processed and they were made uniform by changing the dimensions to  $224 \times 224$  pixels.

## 3.2 The proposed framework

Due to the insufficient number of free COVID-19 radiography images, it could not be possible to develop a CNN model from scratch to automatically identify COVID-19 from X-ray images. In order to control this problem, we adopt a famous method called "transfer learning" and fine-tune three well-known pre-trained models on the prepared data set.

### 3.2.1 Transfer learning approach

Most deep learning applications use a transfer learning method, which requires fine-tuning of the pre-trained framework. We start with an existing network and enter new data containing

previously unknown classes. After making some modifications to the network, we can immediately perform a new task.

There exist two major fashions to use pre-trained models for multiple tasks. The first method is to use the already pre-trained model as a feature extractor, in other words, the weights of the pre-trained model are not suitable for new tasks, since the extracted features are then run through a new classifier, which is trained from scratch. This process will use the convolutional basis of the previously trained network, run new data through it, and train a new classifier on the output.

In the second method, the network is fine-tuned for new tasks. Thus, the weight of the pre-trained model is regarded as the first value of the new task and is updated while training the network. In our case, due to the limited resources used to obtain COVID-19 images, we only fine-tune the last layer of the CNN and use the pre-trained model as the feature extractor.

### 3.2.2 COVID-19 Detection using VGG16:

Figure 3 shows the proposed architecture used for VGG16 and highlights the frozen and trainable layers.

The VGGnet architecture was proposed in 2014 by Simonyan et al. and referred to as “Very Deep Convolutional Networks for Large-scale Image Recognition” [27]. The characteristics of VGG series networks are the  $3 \times 3$  convolutional layers that are one on top of the other, and the depth is getting larger and larger. Reducing the volume size is treated by a maximum pooling.

The VGG16 architecture is composed as following:

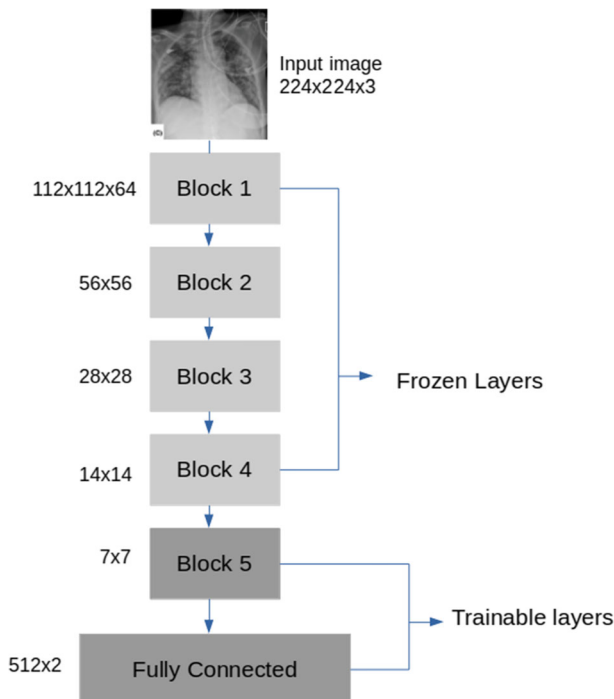


Fig. 3 Proposed VGG16 architecture



- Two Convolutional layers with 64 filters followed by Max pooling layer
- Two Convolutional layers with 128 filters followed by Max pooling layer
- Three Convolutional layers with 256 filters followed by Max pooling layer
- Two stack each with 3 convolutional layers with 512 filters and separated by a max pooling layer
- A final Max pooling layer
- Two fully connected layers with 4096 channels
- Softmax output layer with 1000 classes

### 3.2.3 COVID-19 detection using ResNet50

Figure 4 shows the proposed architecture used for Resnet50 and highlights the frozen and trainable layers.

ResNet-50 is a CNN contains 50 layers; it is deeper than VGG16. Since a global average pool is used instead of a fully connected layer, the size of the model is actually much smaller, which reduces the model size of ResNet50 to 102 MB [29]. The special part of ResNet is residual block learning. This means that each layer should feed into the next layer as well as directly into the layers about 2–3 hops away. Its architecture is composed as follow:

- A convolutional layer with 64 filters and kernel size of  $7 \times 7$ . This is followed by a max pooling layer with a stride size of 2.
- Then, a convolutional layer with 64 filters and a kernel size of  $1 * 1$ , followed by a second convolutional layer with 64 filters and a kernel size of  $3 * 3$ . Then, we have another convolutional layer with 256 filters and a kernel size of  $1 * 1$ . These 3 layers are replicated in total 3 time and 9 layers are obtained at this stage.
- Next, 3 convolutional layers, the first one is with 128 filters and a kernel size of  $1 * 1$ , the second one is with 128 filters and kernel size of  $3 * 3$ , and the third one is with 512 filters and a kernel size of  $1 * 1$ . These layers are replicated 4 time to give us 12 layers at this stage.
- Afterwards, we have convolutional layer with 256 filters and a kernel size of  $1 * 1$ , and two others with 256, 1024 filters and a kernel size of  $3 * 3$ ,  $1 * 1$ . This is replicated 6 time to give us totally 18 layers.
- Then, we have a convolutional layer with 512 filters and a kernel size of  $1 * 1$ , with two others with 512, 2048 and a kernel size of  $1 * 1$ ,  $3 * 3$ . This is replicated 3 times to give us totally 9 layers.

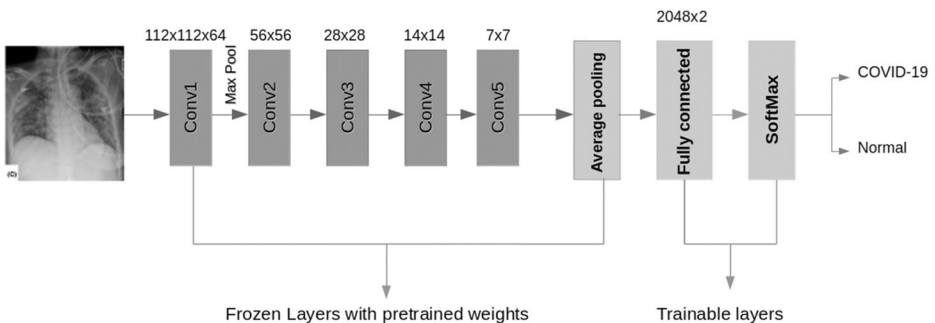


Fig. 4 Proposed Resnet50 architecture

- Finally, we apply an average pooling and finish it with a fully connected layer (with 1000 nodes) and then a softmax function to give us 1 layer as a final stage.

### 3.2.4 COVID-19 detection using inceptionV3

Figure 5 shows the proposed architecture used for InceptionV3 and highlights the frozen and trainable layers.

Szegedy et al. presented the “Inception” module (and the resulting Inception architecture). They published the paper “Convolution Deeper” in 2014 [11]. The purpose of the initial module is to serve as a “multi-level feature extractor” by calculating  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$  convolutions in the same module of the network. Then, the output and channel size of these filters are then fed back to the followed layer.

Inception architectures are less demanding than VGGnet and Resnet in terms of computational efforts (i.e., less RAM is needed to use this framework). Nevertheless, it turned out to show high performance.

## 4 Experimental results

The hidden layers are all activated by the Rectified Linear Unit activation function. The input shape is  $[224, 224, 3]$ . We fine-tune all the models for 25 epochs. We set the batch size to 32 and the learning rate to 0.0001, and we use ADAM as a loss function. We chose to train all the models with a cross-entropy loss function.

We divide the dataset, described in Sec. 3.1, into two groups, for training (80 %) and for validation (20 %), where the first is used for the training process and the second is used for testing the final evaluation. Six performance criteria are utilized to measure the performance: Sensitivity, Accuracy, Specificity, Recall, Precision, and finally F1 score. Here are the obtained results.

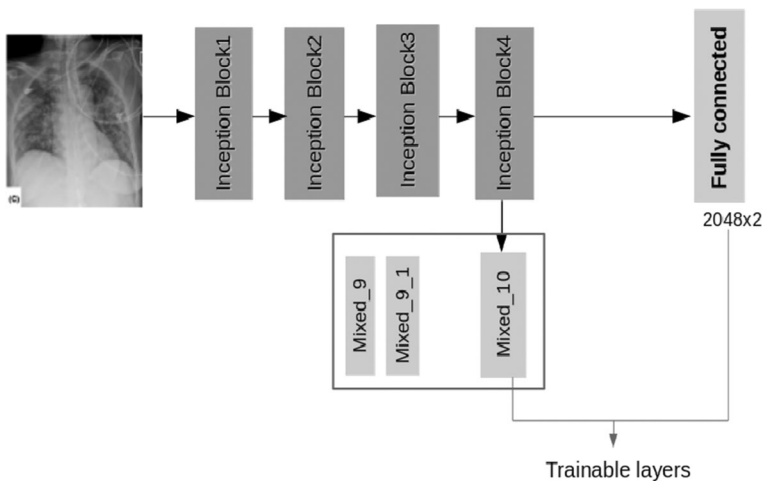


Fig. 5 Proposed InceptionV3 architecture

## 4.1 Training results

The Training results of the proposed models are recorded and presented in the form of plots in Figs. 6, 7, and 8. The orange curve is for validation and the blue curve is according to training.

The results for the three models show that the training accuracy rate is as high as 97 %, and the training loss is reduced to 0.1, which is highlighted in each figure. This can be seen as a good sign for good classification results, especially in the field of medical diagnosis.

## 4.2 Performance criteria

There are a multiple of performance criteria that we can use to assess the performance of a classification model, by using the aforementioned criteria: sensitivity, accuracy, specificity, recall, precision and finally F1 score.

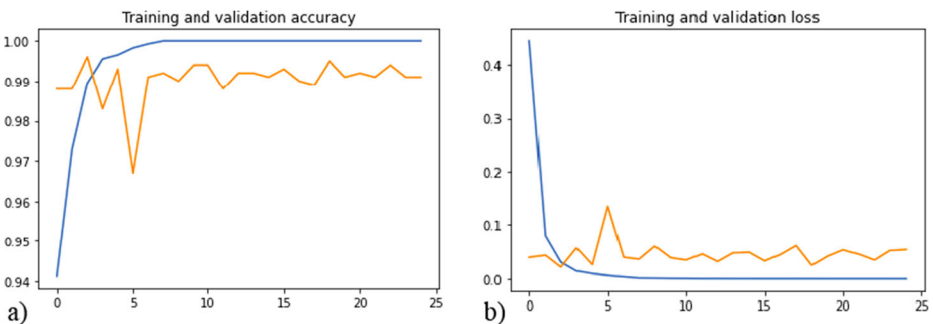
The two criteria of specificity and sensitivity could be used to evaluate a model. They are indeed widely used in the domain of health [18].

### 4.2.1 Definition of the Terms

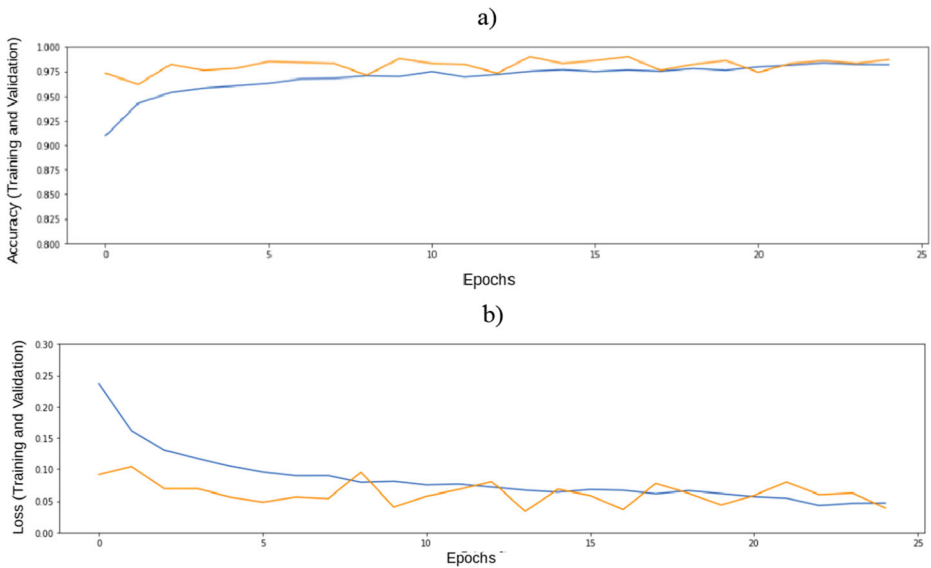
To assess the performance of this classifier, we should distinguish four types of elements that are classified for the desired class: TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative).

- TP: It's when the model correctly predicts the positive class. Here, positive class refers to a patient suffering from COVID 19.
- TN: It's when the model correctly predicts the negative class. Here, negative class refers to a patient NOT suffering from COVID 19.
- FP (Type 1 Error): It's when the model incorrectly predicts the positive class. Predicted that a patient suffering from COVID-19 but it's wrong.
- FN (Type 2 Error): It's when the model incorrectly predicts the negative class. Predicted that a patient NOT suffering from COVID-19 but it's wrong.

Let us first define the performance criteria used to evaluate the performance of the pretrained models we used.

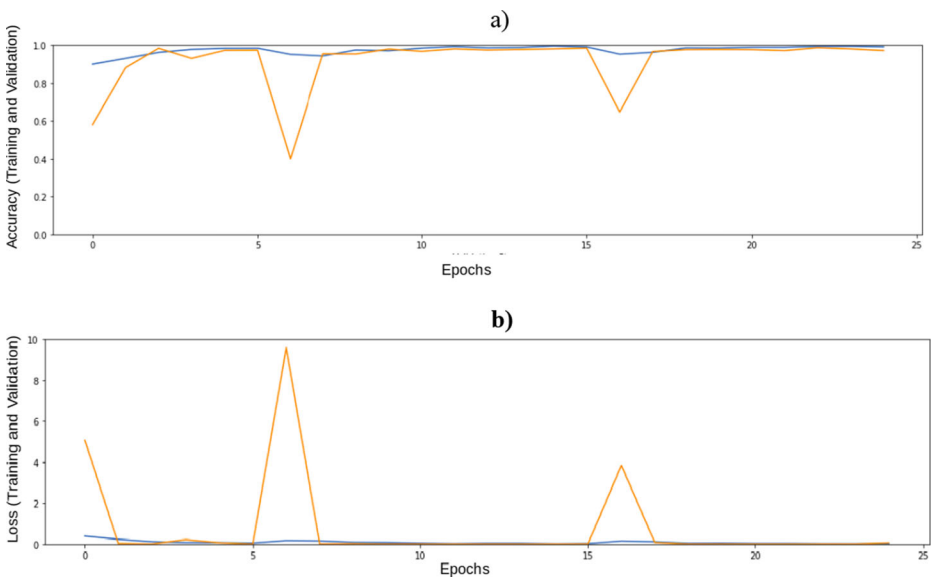


**Fig. 6** Plots of (a) Training and validation accuracy and (b) Training and validation loss by using training epochs-InceptionV3



**Fig. 7** Plots of (a) Training and validation accuracy and (b) Training and validation loss by using training epochs-VGG16

- Classification accuracy =  $\frac{TP + TN}{TP + TN + FP + FN}$ : The accuracy is defined as the rate of correctly classified images.
- Sensitivity =  $\frac{TP}{FN + TP}$ : Measures how the model detects events in the positive category. Therefore, given that COVID-19 is a positive category, sensitivity can quantify how much X-ray images are correctly predicted as COVID-19.



**Fig. 8** Plots of (a) Training and validation accuracy and (b) Training and validation loss by using training epochs-Resnet50

- Specificity =  $TN / (FP + TN)$ : Specificity determines the proportion of actual negatives which are correctly detected.
- Precision =  $TP / (TP + FP)$ : It is the proportion of the number of correctly classified positive categories to the number of predicted positive categories. In other words, precision is the response of the question: among all patients predicted as positive how many are really infected by COVID-19. The Precision should be high.
- Recall =  $TP / (FN + TP)$ : The recall rate is the proportion of the number of correctly classified positive subjects to the number of positive subjects. The aim is to have it as high as possible.
- F1 score =  $2 * (precision * recall) / (precision + recall)$ : To compare two models with high recall but low precision, or with low recall but high precision, is not an easy task. F1-Score is generally used, to make this comparison feasible. It enables the measuring of precision and recall at the same time. In practice, we replace the Arithmetic Mean by the Harmonic Mean. The result is that we further penalize the extreme values.

#### 4.2.2 Results

Table 2 shows the sensitivity, accuracy, specificity, recall, precision as well as F1 score of Resnet50, VGG16, and InceptionV3. It points out that the two proposed fine-tuned versions of VGG16 and InceptionV3 outperform the proposed fine-tuned version of Resnet50 with respect to all the six performance criteria. These two fine-tuned versions have exactly the same performance with respect to three performance criteria, namely recall, precision, and F1 score, as depicted in the last three columns of the last two rows of Table 2.

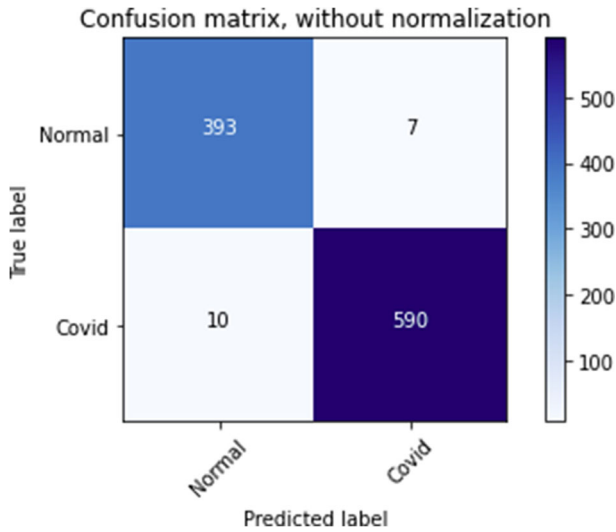
The modified version of the VGG16 model shows slightly better results than the fine-tuned InceptionV3 with respect to accuracy and specificity. However, the fine-tuned InceptionV3 shows better results than VGG16 in terms of sensitivity. We can then conclude that overall, the fine-tuned InceptionV3 is the choice that we recommend.

We provide the confusion matrix for the three performed models. Figures 9, 10, and 11 show the confusion matrix of the fine-tuned ResNet50, InceptionV3, and VGG16 models on 1000 test image sets.

For VGG16 model, the confusion matrix shows only 10 out of 600 COVID-19 images are involved in the normal class (false negative), and only 7 out of 400 normal images (non-COVID-19) are classified as COVID-19 class (false positive). As for Resnet50, we find 21 false negative and 7 false positive cases. The confusion matrix of the final model, InceptionV3, shows also 16 false negative and only 3 false positive cases.

**Table 2** Classification report for Resnet50, InceptionV3 and VGG16

Modified version of	Accuracy	sensitivity	Specificity	Precision	Recall	F1 Score
Resnet50	97.20 %	98.25 %	97.00 %	97.00 %	96.00 %	97.00 %
InceptionV3	98.10 %	99.25 %	98.00 %	98.00 %	98.00 %	98.00 %
VGG16	98.30 %	98.25 %	98.33 %	98.00 %	98.00 %	98.00 %



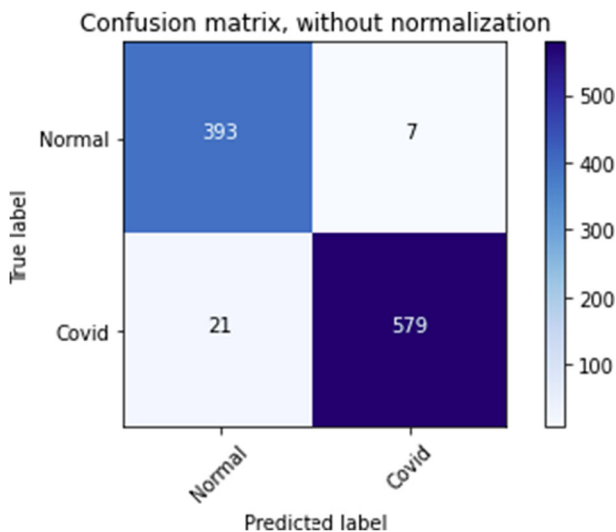
**Fig. 9** The confusion matrix of the proposed VGG16 model

### 4.3 Discussion

All three models have obtained very promising results, with a sensitivity and a specificity rate of around 98 %. The performance of InceptionV3 is slightly better than Resnet50 and VGG16.

Our results show that by using a combination of transfer learning and data augmentation methods with VGG16, Resnet50, InceptionV3, an accurate CNN model can be constructed.

It can be seen from the results that the positive observation is the accuracy and recall rate of COVID-19 cases. A higher recall value means a lower number of FN cases. It is worth mentioning that FN prediction is very dangerous for the patients themselves and the society,



**Fig. 10** The confusion matrix of the proposed Resnet50 model

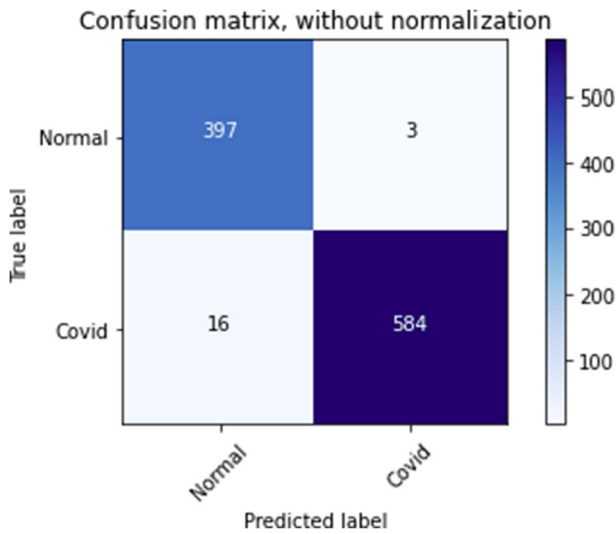


Fig. 11 The confusion matrix of the proposed InceptionV3 model

since infected patients are declared healthy and therefore, they lead a normal life without taking any measures for themselves and for public health. This turns out to be important since the main purpose of this research is to minimize FN COVID-19 cases in order to better support the clinical decision.

Table 3 summarizes discoveries of automatic diagnosis of COVID-19 based on chest X-ray images and compares it with the proposed model.

We consider in Table 3 the commonly used models involving the same range of number of parameters (around 25 Million). It shows clearly that our 3 fine-tuned models outperform the rest of the models in Table 3 with respect to accuracy. We do not have at our disposal data for the other five performance criteria. This will be the subject of a future work.

It’s worth noting that the VGG19 model uses 143 Million parameters [3], which is about 7 times the number of parameters used in our fine-tuned models and in those of Table 3. This model provides slightly better accuracy (98.75 %) at the cost of a larger computational effort. It’s not worth expanding our fine-tuned models to use a larger number of parameters since the difference between the accuracies of the VGG19 model and the fine-tuned VGG16 is less than 0.5 %.

Table 3 Comparison of the most commonly used automatic diagnosis of COVID-19 that are based on chest X-ray images to our fine-tuned models

Study	Architecture	Accuracy	Number of parameters in Million
Sethy and Behra [25]	Resnet50	95.38 %	36
Narin et al. [20]	InceptionV3	97 %	26
Ioannis et al. [3]	Xception	85.57 %	33
Ozturk et al. [21]	DarkNet	98.08 %	1.1
Ioannis et al. [3]	VGG19	98.75 %	143
Fine-tuned Resnet50	Resnet50	97.20 %	23
Fine-tuned VGG16	VGG16	98.30 %	15
Fine-tuned InceptionV3	InceptionV3	98.10 %	21

The encouraging results of the deep learning model for detecting COVID-19 in radiographic image detection indicate that in the near future, deep learning will play a greater clinical support role for fighting against this epidemic. Some of the limitations of this study can be overcome by performing a more analysis when we get more available data (symptomatic and asymptomatic patients).

## 5 Conclusions

Early diagnosis of the novel coronavirus is extremely important to avoid further spread of the virus to others. Along this work, we design a method based on deep transfer learning that uses chest X-ray images related to patients affected with COVID-19 and patients without COVID-19 to automatically detect the disease. The suggested classification model for detecting COVID-19 can reach an accuracy of more than 98 %. According to our research results, due to its high overall performance, we believe it is of nature to help doctors and health experts make clinical decisions. To discover COVID-19 as early as possible, this study has an in-depth understanding of how to use deep transfer learning approaches.

COVID-19 presents a threat to the world's healthcare community and kills millions of people. Due to the large number of patients seen outdoors or in emergencies, doctors have limited time, and computer-aided analysis could rescue lives through early screening as well as appropriate care.

By efficiently training through a relatively small image set, our fine-tuned models show high performance in the classification of COVID-19 pneumonia. Our conviction is that the proposed computer-aided diagnosis mechanism could outstandingly improve the diagnosis of COVID-19 cases.

This is very helpful in a pandemic, especially when the available health resources do not match the burden of disease as well as the need for preventive measures to be taken.

Research in deep learning always strives to build better representations of reality and to create models capable of learning these representations from non-labeled data on a large scale. Some of these representations are based on the latest developments in several areas. For example, Ahmad Ali et al. [2], use the algorithms of deep learning in their research in order to explore temporal relations and spatial relations. They suggest a dynamic deep hybrid spatio-temporal neural network to predict the traffic flow in each area of the city with high accuracy.

As a future perspective work, we are in the point of thinking to combine the three proposed models in this work and to train all the layers as a new approach to provide a better result.

## References

1. Ai T et al (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*: 200642
2. Ali A, Zhu Y, Zakarya M (2021) A data aggregation-based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-020-10486-4>
3. Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 43(2):635–640
4. Axell-House DB, Lavingia R, Rafferty M, Clark E, Amirian ES, Chiao EY (2020) The estimation of diagnostic accuracy of tests for COVID-19: A scoping review. *J Infect* 81(5):681–697



5. Bloice MD, Roth PM, Holzinger A (2019) Biomedical image augmentation using Augmentor. *Bioinformatics* 35(21):4522–4524
6. Cohen J, Paul et al (2020) Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988
7. covid-chestxray-dataset. <https://github.com/ieee8023/COVID-chestxray-dataset>. Accessed 25 Mar 2020
8. Eurosurveillance Editorial Team (2020) Note from the editors: World Health Organization declares novel coronavirus (2019-nCoV) sixth public health emergency of international concern. *Eurosurveillance* 25(5): 200131e
9. Gazzah S, Bencharef O (2020) A Survey on how computer vision can response to urgent need to contribute in COVID-19 pandemics. 2020 International Conference on Intelligent Systems and Vision C (ISCV). IEEE, New York
10. Globalpulse, Need for greater cooperation between practitioners and the AI community. <https://www.unglobalpulse.org/2020/05/need-for-greater-cooperation-between-practitioners-and-the-ai-community/>. Accessed 27 May 2020
11. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
12. Holzinger A et al (2018) Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI. International Cross-Domain Conference for Machine Learning and Knowledge Extraction. Springer, Cham, 2018
13. Holzinger A et al (2019) Causability and explainability of artificial intelligence in medicine. *Wiley Interdisc Rev Data Min Knowl Discov* 9(4):e1312
14. Isa A. Computational intelligence methods in medical image-based diagnosis of COVID-19 infections. *Computational Intelligence Methods in COVID-19: Surveillance, Prevention, Prediction and Diagnosis*. Springer, Singapore, pp 251–270
15. Kallianos K et al (2019) How far have we come? Artificial intelligence for chest radiograph interpretation. *Clin Radiol* 74(5):338–345
16. Khan A, Sohail A, Zahoor U, Qureshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. *Artif Intell Rev* 53(8):5455–5516
17. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. *Commun ACM* 60(6):84–90
18. Lalkhen AG, McCluskey A (2008) Clinical tests: sensitivity and specificity. *Contin Educ Anaesth Crit Care Pain* 8(6):221–223
19. Mooney P (2018) Chest x-ray images (pneumonia). Online. <https://www.Kaggle.com/paultimothymooney/chest-xray-pneumonia>, tanggal akses
20. Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849
21. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR (2020) Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput Biol Med* 121:103792
22. Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam MT (2020) Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* 8:132665–132676
23. Romero Lopez A, G-i-Nieto X, Burdick J, Marques O (2017) Skin lesion classification from dermoscopic images using deep learning techniques. (2017) 13th IASTED International Conference on Biomedical Engineering (BioMed), pp 49–54. <https://doi.org/10.2316/P.2017.852-053>
24. Bergman SJ, Cennimo DJ, Miller MM, Olsen KM (2020) Treatment of coronavirus disease 2019 (COVID-19): investigational drugs and other therapies. Medscape. 2020
25. Sethy PK, Behera SK (2020) Detection of coronavirus disease (covid-19) based on deep features
26. Shan F et al (2020) Lung infection quantification of covid-19 in ct images with deep learning. arXiv preprint arXiv:2003.04655
27. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
28. Stoecklin SB et al (2020) First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020. *Eurosurveillance* 25(6):2000094
29. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
30. Wang L, Lin ZQ, Wong A (2020) Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Sci Rep* 10(1):1–12
31. World health organization (2020) Director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—6-may-2020. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19>. Accessed 6 May 2020

32. World health organization. How WHO is working to track down the animal reservoir of the SARS-CoV-2 virus. <https://www.who.int/news-room/feature-stories/detail/how-who-is-working-to-track-down-the-animal-reservoir-of-the-sars-cov-2-virus>. Accessed 6 Nov 2020
33. World health organization. Coronavirus. [https://www.who.int/health-topics/coronavirus#tab=tab\\_3](https://www.who.int/health-topics/coronavirus#tab=tab_3)
34. World health organization. Coronavirus disease (COVID-19) advice for the public. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public>
35. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, Li L (2020) A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering* 6(10):1122–1129

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

## Affiliations

Sarra Guefrechi<sup>1</sup> · Marwa Ben Jabra<sup>2,3</sup> · Adel Ammar<sup>4</sup> · Anis Koubaa<sup>4,5,6</sup> · Habib Hamam<sup>1</sup>

<sup>1</sup> Faculty of Engineering, University of Moncton, Moncton, NB, Canada

<sup>2</sup> Charisma University, British Overseas Territories, Englewood, UK

<sup>3</sup> Robotics and Internet- of-Things Unit (RIoT) Lab, Riyadh, Saudi Arabia

<sup>4</sup> Prince Sultan University, Riyadh, Saudi Arabia

<sup>5</sup> Gaitech Robotics, Shanghai, China

<sup>6</sup> INESC- TEC, ISEP, Polytechnic Institute of Porto, Porto, Portugal