



COVID-19 detection in CT and CXR images using deep learning models

Ines Chouat · Amira Echtioui · Rafik Khemakhem · Wassim Zouch · Mohamed Ghorbel · Ahmed Ben Hamida

Received: 17 August 2021 / Accepted: 2 December 2021 / Published online: 22 January 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract Infectious diseases pose a threat to human life and could affect the whole world in a very short time. Corona-2019 virus disease (COVID-19) is an example of such harmful diseases. COVID-19 is a pandemic of an emerging infectious disease, called coronavirus disease 2019 or COVID-19, caused by the coronavirus SARS-CoV-2, which first appeared in December 2019 in Wuhan, China, before spreading around the world on a very large scale. The continued rise in the number of positive COVID-19 cases has disrupted the health care system in many countries, creating a lot of stress for governing bodies around the world, hence the need for a rapid way to identify cases of this disease. Medical imaging is a widely accepted technique for early detection and diagnosis of the disease which includes different techniques such as Chest X-ray (CXR), Computed Tomography (CT) scan,

etc. In this paper, we propose a methodology to investigate the potential of deep transfer learning in building a classifier to detect COVID-19 positive patients using CT scan and CXR images. Data augmentation technique is used to increase the size of the training dataset in order to solve overfitting and enhance generalization ability of the model. Our contribution consists of a comprehensive evaluation of a series of pre-trained deep neural networks: ResNet50, InceptionV3, VGGNet-19, and Xception, using data augmentation technique. The findings proved that deep learning is effective at detecting COVID-19 cases. From the results of the experiments it was found that by considering each modality separately, the VGGNet-19 model outperforms the other three models proposed by using the CT image dataset where it achieved 88.5% precision, 86% recall, 86.5% F1-score, and 87% accuracy while the refined Xception version gave the highest precision, recall, F1-score, and accuracy values which equal 98% using CXR images dataset. On the other hand, and by applying the average of the two modalities X-ray and CT, VGG-19 presents the best score which is 90.5% for the accuracy and the F1-score, 90.3% for the recall while the precision is 91.5%. These results enables to automatize the process of analyzing chest CT scans and X-ray images with high accuracy and can be used in cases where RT-PCR testing and materials are limited.

I. Chouat · A. Echtioui (✉) · R. Khemakhem · M. Ghorbel · A. B. Hamida
ATMS Lab, Advanced Technologies for Medicine and Signals, ENIS, Sfax University, Sfax, Tunisia
e-mail: amira.echtioui@isims.usf.tn

I. Chouat
e-mail: ineschouat10@gmail.com

I. Chouat · R. Khemakhem
Higher Institute of Management of Gabès, Gabès University, Gabès, Tunisia

W. Zouch
King Abdulaziz University (KAU), Jeddah, Saudi Arabia

Keywords COVID-19 · CT image · X-ray · Medical imaging · Deep learning

Introduction

Coronaviruses (nicknamed CoV) are a family of viruses of varying severity: depending on the government website, they can cause simple colds or more serious conditions such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). These are common RNA viruses of the Coronaviridae family that cause digestive and respiratory infections in humans and animals.

The World Health Organization (WHO) has found COVID-19, a novel coronavirus illness, in Wuhan, Hubei, China ([https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it)). Previous research has linked the virus's origin to a fish market in Wuhan, and it has also been suggested that the virus was spread to humans by bats. This virus is the most contagious and rapidly spreading member of the Coronavirus family, and it has become a public health emergency as cases have grown despite a lack of health facilities and resources.

The novel coronavirus is a disease that is transmissible from human to human via cough or sneeze droplets from an infected person (Chen et al. 2020). It is spread by coming into contact with an infected person, touching something that has the virus on its surface, and then touching their mouth, eyes, ears, or nose. High fever, cough, sore throat, difficulty breathing, and diarrhea are all common signs of coronavirus infection. Fever is the first symptom, and it might proceed to pneumonia. Shortness of breath, chest tightness, and other dangerous infections are all possible complications.

The elderly and those with pre-existing disorders have a higher chance of death (Bansal and Sridhar 2020). Cardiovascular disease, asthma, diabetes, and hypertension are examples of high-risk people. Only a few cases have been reported in children (Zunyou and McGoogan 2020).

The WHO has announced a strategy to limit the spread of the virus from infected people to healthy people by detecting them and isolating them from others.

On the other hand, laboratory tests were used as reference tests to detect COVID-19 patients such as PCR, antigen, and antibodies. However, these tests have certain limitations regarding the accuracy of the

results which can often produce false alarms, the availability of laboratories that can serve the test, the time required to obtain the results, the cost of the test, the phase of infection.

All of these limitations risk a high spread of the disease among other peoples by patients (Elpeltagy and Sallam 2021).

Given the low sensitivity of the RT-PCR test, designing other automated and reliable methods to screen COVID-19 patients is currently a major challenge for researchers.

As a solution to these problems and since the coronavirus attacks our epithelial cells and damages our lungs, medical imaging techniques, such as chest computed tomography and chest X-ray offer a non-invasive alternative to identify COVID-19 and diagnose infected lungs. While selectively testing false negative PCR cases.

The majority of medical images of pneumonia are complicated, difficult to comprehend, and take a long time for radiologists to see the smallest features by vision.

In this context, artificial intelligence models intervene, which represents a quick solution for this type of problem. More specifically, the deep learning (DL) approach has been very popular and successfully used in the classification of medical images due to its high accuracy.

The major goal of this research is to use pre-trained deep learning architectures to refine and develop an automated tool for detecting and diagnosing COVID-19 in chest X-ray and Computed Tomography images (Elpeltagy and Sallam 2021).

The following is a breakdown of the rest of the article: Sect. 2 will focus on related work after the introduction. Section 3 covers our proposed framework as well as how the dataset was created. In Sect. 4, we propose experiments. Section 5 concludes the article.

Related works

The diagnosis and early detection of diseases with high infection is a major challenge for health officials and researchers in order to reduce the suffering of patients.

Recently, several scientific papers have been published on the application of deep learning and CNN

approaches in the field of detection of COVID-19 cases using different medical imaging modalities mainly the CXR and CT images.

This section will discuss and analyze approaches to some of the important work done by other researchers. This analysis helps overcome the problems encountered by others and opens a new approach in the design and development of a precise solution to fight against the COVID-19 pandemic and limit its transmission.

The authors of Xu et al. (2020) created a predictive algorithm to differentiate COVID-19 pneumonia from influenza. Deep learning algorithms were used to create viral pneumonia. For such predictions, the CNN model was used. The prediction model's highest level of precision was 86.7%.

Apostolopoulos and Mpesiana (2020) proposed to evaluate the performance of convolutional neural network architectures developed in recent years (VGGNet-19, MobileNet-v2, Inception, Xception and Inception ResNet-v2) using medical image classification techniques as a tool for automatic detection of coronavirus disease. The authors adopted a procedure called "Transfer Learning" because it achieves good performance for the detection of various anomalies in small datasets of medical images. They used a dataset of 1,442 X-ray images of patients: 714 images with bacterial pneumonia and viral pneumonia, 224 images with confirmed COVID-19 disease, and 504 images of normal incidents. The results show that VGGNet-19 and MobileNet-v2 achieve the best classification accuracy.

They claimed to have an overall accuracy rate of 83.5%. The non-COVID-19 class had the lowest positive predictive value (67.0%), whereas the normal class had the highest (95.1%).

For the categorization of COVID-19 chest radiographs into COVID-19 and normal classes, Ai et al. (2020) built a CNN based on the InceptionV2, Inception-ResNetV3, and ResNet50 models. They discovered a strong link between the CT imaging results and the PCR method.

InceptionNet was used by Wang et al. (2020) to detect anomalies related to COVID-19 in lung CT scan images. On 1065 CT images, the InspectionNet model was evaluated, and 325 contaminated people were identified with an accuracy of 85.20%.

Methodology

Recently, Deep Learning (DL) methods have demonstrated great talent with peak performance in the field of image processing and computer vision (segmentation, detection, and classification).

In this study, we have undertaken the task of classifying CT scans and CXR images into one of two classes: COVID-19 positive or normal.

We suggested four DL algorithms to achieve this classification: VGGNet-19, ResNet50, Xception, and InceptionV3. The proposed methodology for COVID-19 screening using CT scans is provided in this section. The suggested technique is depicted in Fig. 1.

Dataset

Finding an appropriate dataset is one of the criteria for image recognition and computer vision research. This study employed a database containing 3,000 Chest CT scans normal images and 623 Chest CT scans images of COVID-19 patients. These images were received from Dr. Jkooy's open-source GitHub repository (<https://github.com/UCSD-AI4H/COVID-CT/tree/master/Images-processed>). We have used 408 Chest CT scans images of normal patients and 325 Chest CT scans images of COVID-19 patients. We have used also 500 Chest X-ray images of normal patients and 500 Chest X-ray images of COVID-19 patients from Kaggle (<https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>).

Figure 2 shows some examples of chest CT scans and X-ray images from the prepared dataset.

Pre-processing and data augmentation

The data contains a lot of noise due to the increasing forms of intrusion in the imaging and data collection procedure (Sakib et al. 2020).

The preprocessing techniques are the useful solution to remove unwanted noise and prepare the input data to be compatible with the requirements of our model.

In this study, the images were standardized by changing the dimensions (resizing) using images of height 224 pixels, width 224 pixels, the number of channels 3 (224*224*3), and the format (JPEG).

Our original images include RGB coefficients ranging from 0 to 255, but these values are too high

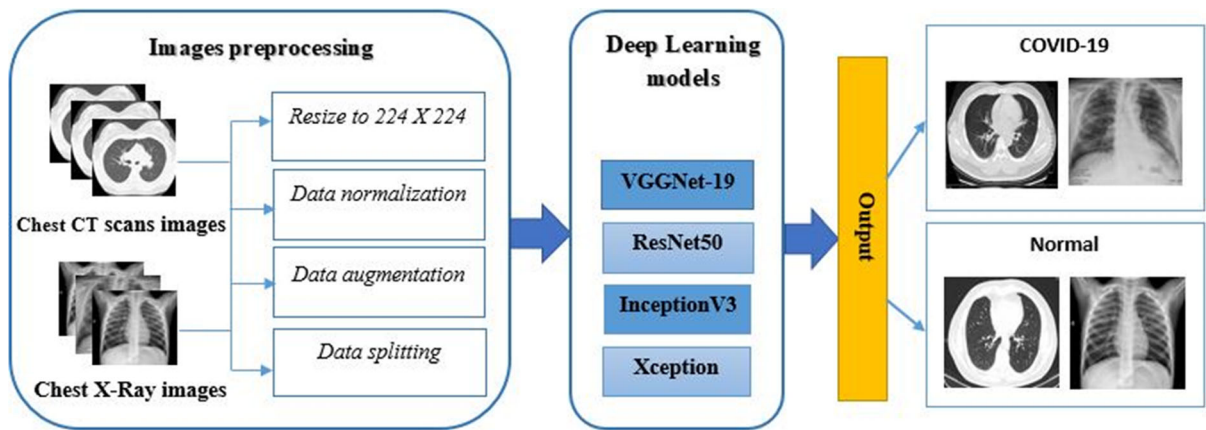


Fig. 1 Overview of proposed method

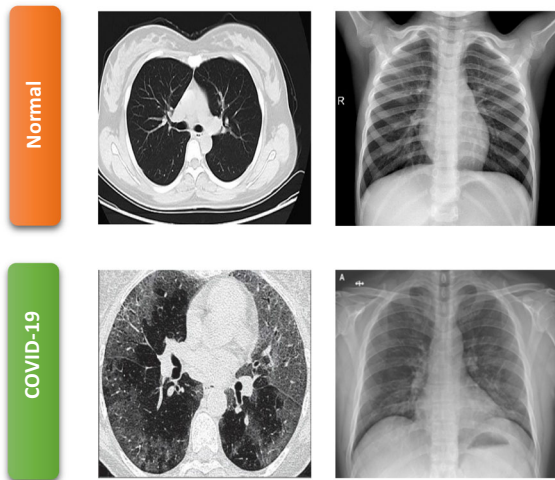


Fig. 2 Sample chest CT scans and X-ray images dataset for normal cases (first row) and COVID-19 patients (second row)

for our models to comprehend (given a typical learning rate), so we aim for values between 0 and 1 by scaling them to $1/255$.

Normal images were tagged as 1 and COVID-19 images were tagged as 0.

We use data augmentation techniques to artificially boost the amount of our training data because our data collection is rather small. The increase in data is an often applied DL method that generates the required number of samples. It also improves network efficiency for a small database by optimizing it.

Rotation, shifting, flipping, zooming, and transformation are all examples of traditional data augmentation procedures. We used “*Keras ImageDataGenerator*” to apply image augmentations

during training in this investigation. The geometric transform used in this study such as scaling, rotating, shifts, and flips are shown in Table 1:

Dataset splitting

The dataset was separated into two independent data sets, with 80% and 20% used for training and testing, respectively as shown in Fig. 3.

Proposed methods

The mission of recognizing patients infected with the coronavirus using CT and CXR was performed using four alternative models based on a deep convolutional neural network (VGGNet-19, ResNet50, InceptionV3, and Xception).

In order to develop our models, we implemented a transfer learning technique based on the weight of the pre-trained models (ImageNet (Russakovsky et al. 2014)).

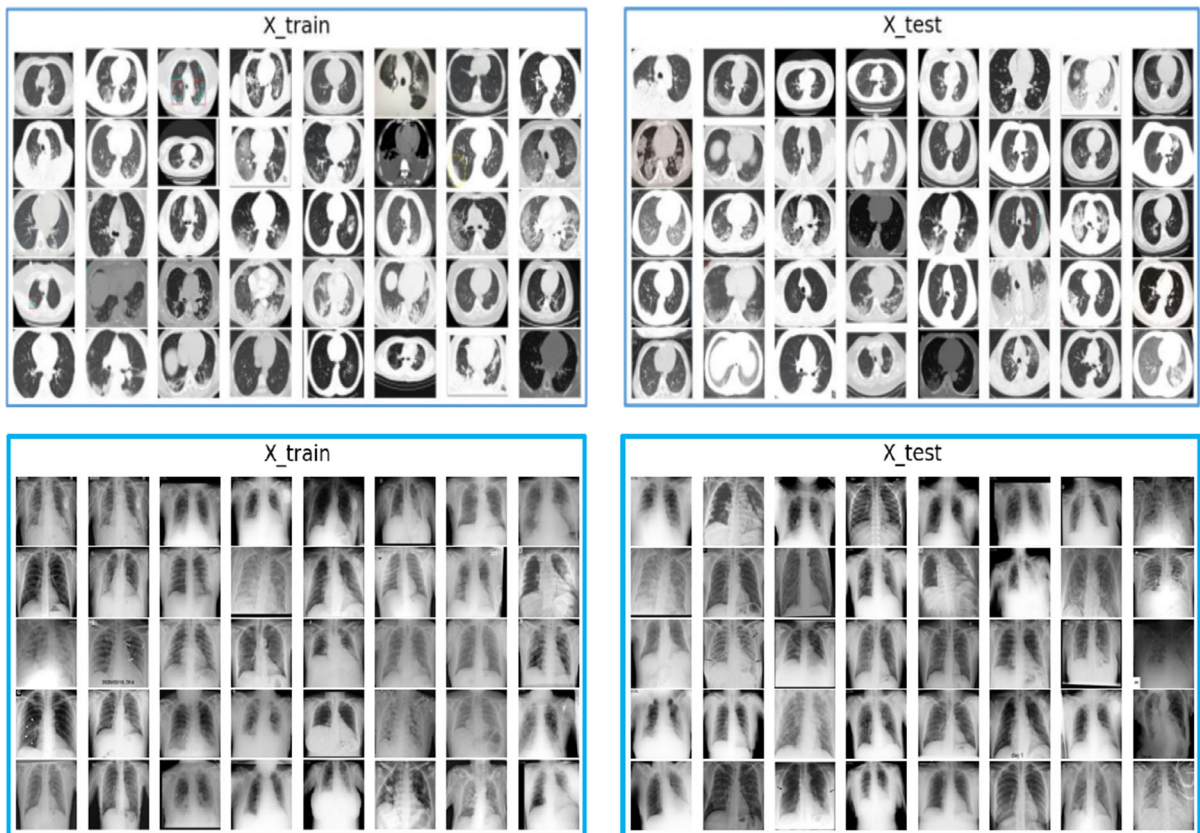
The basic architecture of the model contains three components: a pre-trained network, a set of 3 layers modified according to our needs, and a prediction classifier.

To achieve efficient prediction performance, some changes are made to the architecture of each model. To be more specific, we modified each model’s architecture by adding three extra layers at the end of its architecture to fit our categorization purpose.

The following subsections explain the details of each of the four developed models.

Table 1 Data augmentation used

Argument	Parameter value	Description
Rotation range	20	Apply a specified angle or degree of rotation to the image (https://dataaspirant.com/data-augmentation-techniques-deep-learning/#t-1598844713773)
Height shift range	0.2	The height shift range function shifts pixels vertically to the top or bottom at random (https://dataaspirant.com/data-augmentation-techniques-deep-learning/#t-1598844713773)
Width Shift range	0.2	The width shift range parameter adjusts pixels horizontally to the left or right at random (https://dataaspirant.com/data-augmentation-techniques-deep-learning/#t-1598844713773)
Horizontal flip	True	When a given input can be flipped horizontally during the training process is controlled

**Fig. 3** Example of training and testing images

Detection of COVID-19 using the VGGNet model

Karen Simonyan and Andrew Zisserman of the Oxford Institute of Robotics developed the Visual Geometry Group Network (VGG) based on the convolutional neural network architecture (Simonyan and Zisserman 2015). During the 2014 Large Scale Visual

Recognition Challenge, it was addressed (ILSVRC2014). On the ImageNet dataset, VGGNet performed admirably. The VGGNet uses three convolutional layers with several filters stacked on top of each other to improve image extraction functionality, and the depth gets more and bigger (He et al. 2016).

VGGNet-16 and VGGNet-19 are two alternative variations of this deep network architecture, with varying depths and layers. VGGNet-16 and VGGNet-19 differ in that VGG-19 adds one additional layer in each of the three convolutional blocks, making VGGNet-19 deeper than VGGNet-16 (Simonyan and Zisserman 2015).

Figure 4 shows the proposed method for the detection of COVID-19 using the VGGNet model. In this paper, VGGNet consists of 5 blocks, the first two blocks contain 2 convolutional layers followed by Max pooling while the last 3 contain 4 convolutional layers followed by Max pooling. These blocks are followed by flatten, dropout, and two dense layers.

Detection of COVID-19 using the ResNet50 model

ResNET, short for Residual Networks, is a deep neural network used for many computer vision tasks. Its development took place in 2015 when it won the ImageNet competition (He et al. 2016). It is seen as a continuation of deep networks that revolutionized the CNN architectural race by introducing the concept of residual learning into CNNs and put developed an effective methodology for the training of networks.

The proposed method for detecting COVID-19 using the ResNet50 model is shown in Fig. 5. It is primarily made up of residual blocks. The hidden layers of shallow neural networks are linked to each other. However, there exist connections between the residual blocks in the ResNet architecture. This means that each layer should flow into the next, with a space of around 2 to 3 swings between them.

The main advantage of residual connections in the ResNet architecture is that the connections preserve the knowledge gained during training and speed up model training time by increasing network capacity.

In this study, we used ResNet50 as the base model architecture and refined it for the purposes of our classification problem.

Detection of COVID-19 using the InceptionV3 model

The “Inception” module is introduced by Szegedy et al. (and the resulting Inception architecture). In 2014, they published the article “Deeper Convolution”. The first module’s goal is to operate as a “multi-level feature extractor” by performing 1×1 , 3×3 , and 5×5 convolutions within the same network module. These filter’s output is stacked with the

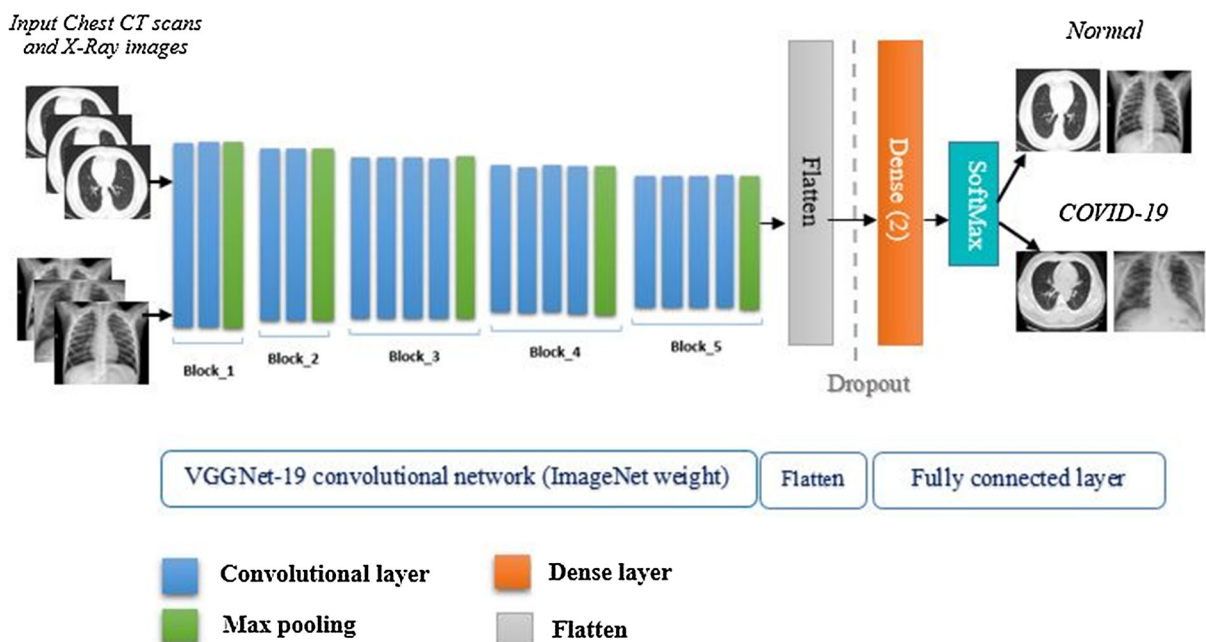


Fig. 4 Proposed method for the detection of COVID-19 using the VGGNet-19 model

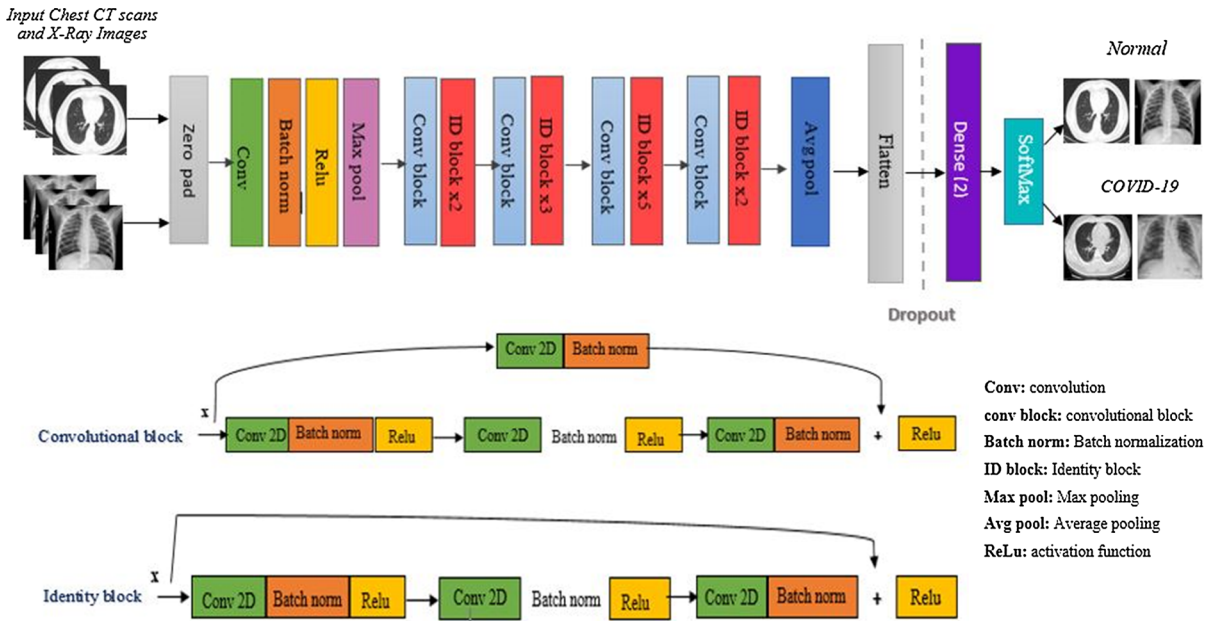


Fig. 5 Proposed method for the detection of COVID-19 using the ResNet50 model

channel size before being fed into the network’s next layer.

In terms of computational effort, Inception architectures are less demanding than VGGNet and ResNet (i.e. less RAM is needed to use this framework). Despite this, it turned out to be a high-performing system (Guefrechi et al. 2021).

The proposed method for the detection of COVID-19 using the Inception model is shown in Fig. 6.

Detection of COVID-19 using the Xception model

Xception is a new structure proposed by François Chollet in 2017. Xception is considered as an extension of the Inception architecture, where the Inception modules are replaced by convolutions separable in depth.

Except for the first and last modules, the network comprises 36 convolutional layers that serve as the foundation for extracting the network’s features. The network is arranged into 14 modules, each of which has linear residual connections surrounding it.

In the Xception architecture, using residual connections will result in faster convergence and better overall performance.

The proposed model details are illustrated in Fig. 7.

Presentation of the used optimizers

To minimize the error of our predictions (the loss function) and to make our predictions as correct and optimized as possible, we tried to test one optimizer:

- **RMSprop** (Root Mean Squared Propagation): In neural network training, RMSprop is a gradient-based optimization strategy. This normalization equalizes the step size (momentum), lowering it for high gradients to avoid exploding and raising it for minor gradients to avoid vanishing (<https://medium.com/analytics-vidhya/a-complete-guide-to-adam-and-rmsprop-optimizer-75f4502d83be>).

Performance criteria

Definition of terms

The accuracy, sensitivity or recall, precision, and F1 score were used to assess the utility and productivity of each deep learning model tested. A confusion matrix is introduced for each model as a result.

True positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are the four

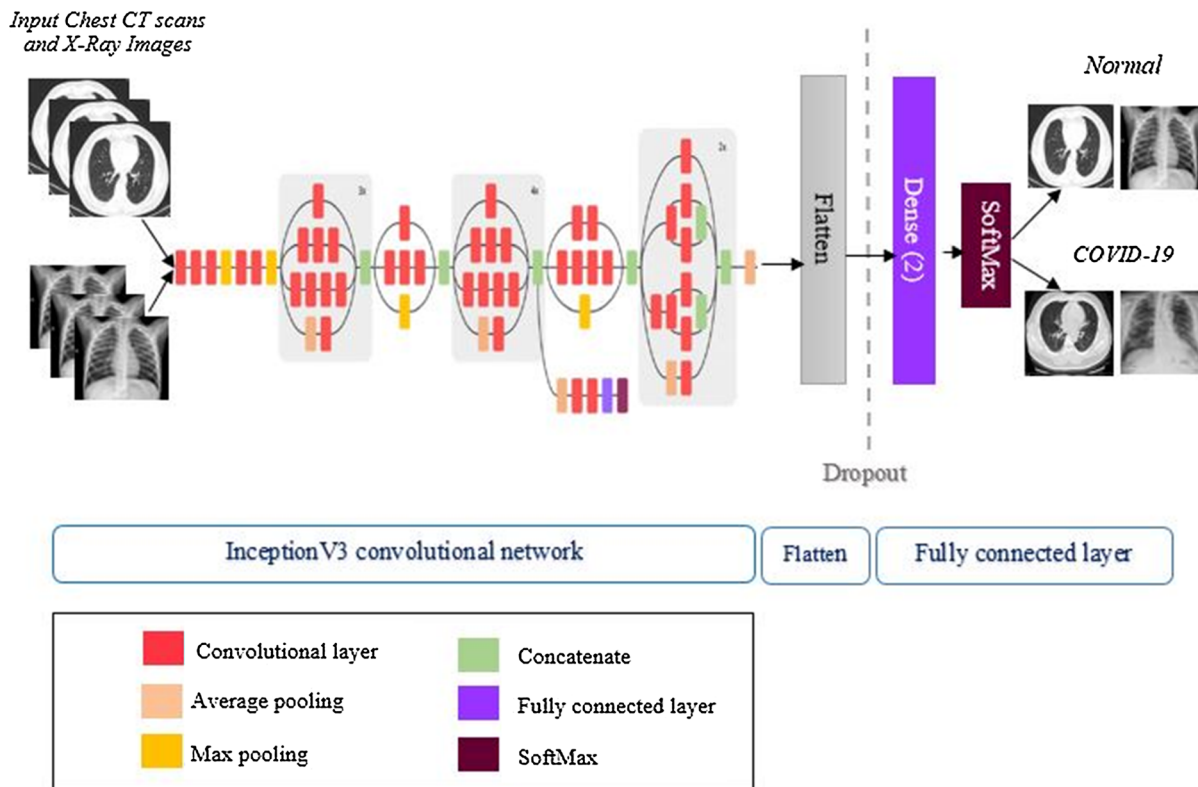


Fig. 6 Proposed method for the detection of COVID-19 using the InceptionV3 model

expected results of the confusion matrix, which are required to construct the assessment criteria.

- **TP:** has been classified as COVID-19 infected, which is correct.
- **TN:** stands for “not infected with COVID-19,” which is correct.
- **FP:** classed as COVID-19 infected, although this is incorrect.
- **FN:** incorrectly classed as NOT infected with COVID-19.

The following performance measures were calculated after calculating the values of the confusion matrix in order to assess the performance of the various pre-trained models used.

- **Accuracy:** As shown in Eq. (1), accuracy is equal to the sum of true positives and true negatives divided by the total values of the confusion matrix components.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

- **Precision:** Eq. (2) shows precision, which is the difference between the number of samples that are truly positive and the total number of samples that are projected to be positive.

$$Precision = \frac{TP}{FP + TP} \quad (2)$$

- **Recall:** The ratio of the total number of correctly categorized positive patients to the total number of positive patients is described as sensitivity or recall, which is provided in Eq. (3).

$$Recall = \frac{TP}{FN + TP} \quad (3)$$

- **F1-score:** Is the harmonic mean of precision and recall. It is shown in Eq. (4).

$$F1\text{-Score} = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

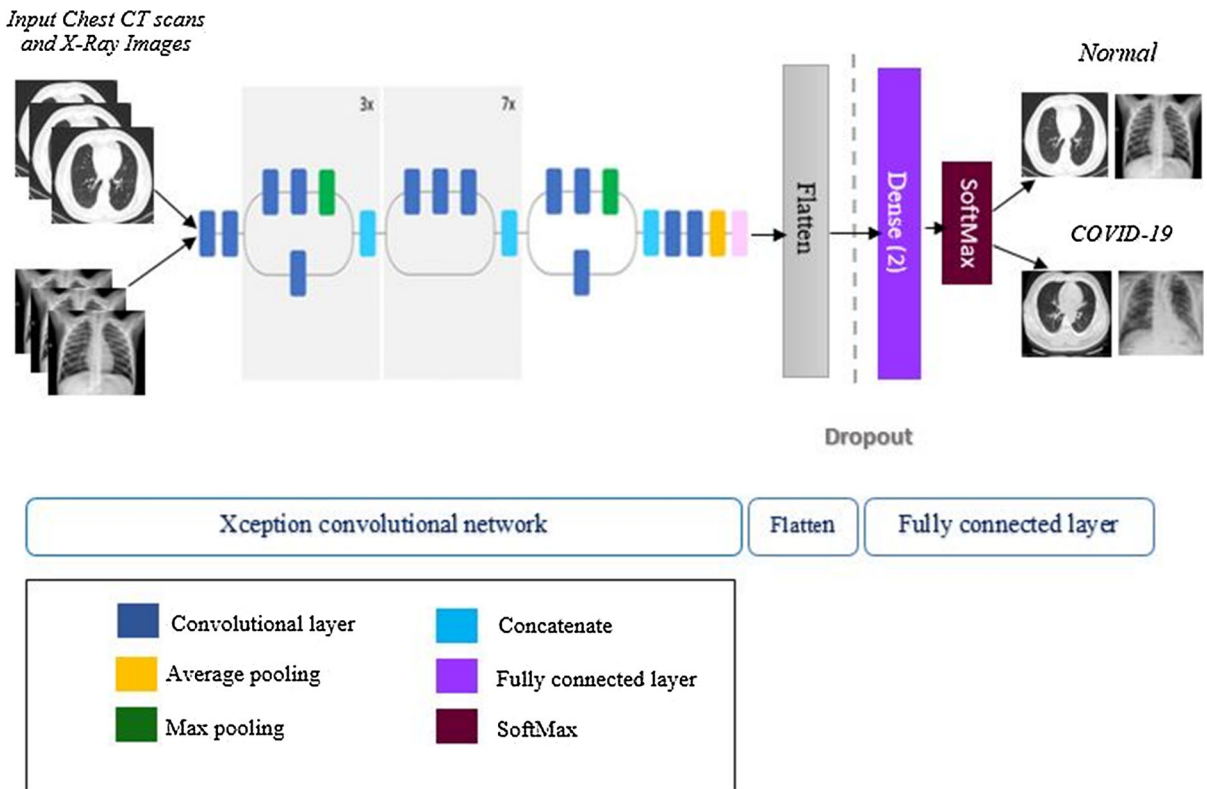


Fig. 7 Proposed method for the detection of COVID-19 using the Xception model

Experimental results

To achieve our goal of automatic CT and CXR images classification and after resizing the images in the dataset to 224×224 pixels and use the input image augmentation methods using a Keras Augmentor API.

Our proposed deep learning models were fed the resulting image.

We used the batch size of 32 to train our models. These models were tuned across 100, 200, and 300 epochs, respectively.

Python 3 and the Keras framework were used to create the models. These were developed using Google Colab.

Results obtained by using CT images

The confusion matrices obtained by the VGGNet-19, ResNet, Inception V3 and Xception models using the CT images for the different numbers of epochs are shown in Figs. 8, 9, 10 and 11.

Figures 12, 13, 14, and 15 shows the loss and accuracy between the training and validation phases

with the different numbers of epochs. As can be seen, the training loss and accuracy continued to improve (low training error compared to high accuracy).

Results obtained by using CXR images

The models are evaluated with precision, recall, accuracy and f1-score metrics. This allows us to compare the models’ performances within a dataset, and the effectiveness of the CXR and CT techniques in the detection of COVID-19 and non-COVID-19 cases.

The confusion matrix obtained by the four models using the CXR images for the different numbers of epochs are shown in Figs. 16, 17, 18 and 19.

The loss and accuracy between the training and validation phases with the different numbers of epochs are shown in Figs. 19, 20, 21, and 22.

Based on the confusion matrices, we can summarize the performances of the different models already studied according to the CT and CXR images.

VGGNet-19 confusions matrices are illustrated in Fig. 8. By classifying CT images, we have:

- for a number of epochs of 100:

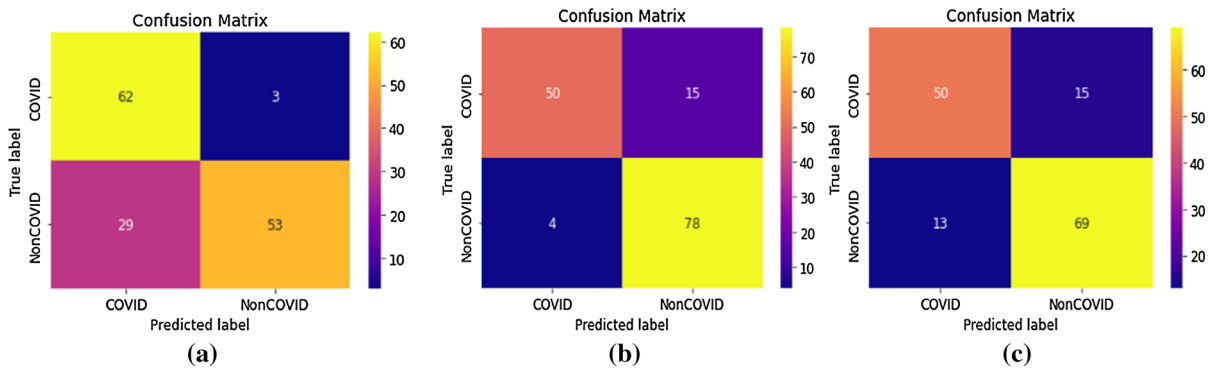


Fig. 8 Confusion matrix obtained by the VGGNet-19 model after **a** 100, **b** 200 and **c** 300 epochs

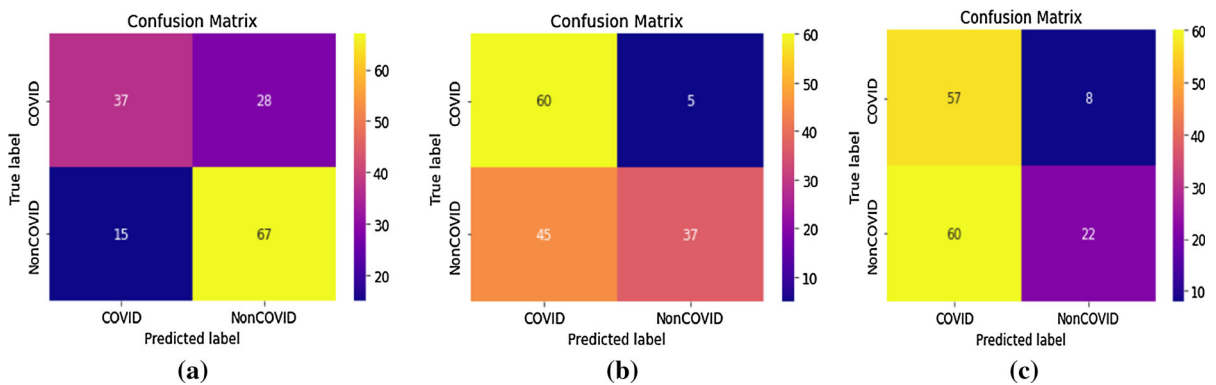


Fig. 9 Confusion matrix obtained by the ResNet50 model after **a** 100, **b** 200 and **c** 300 epochs

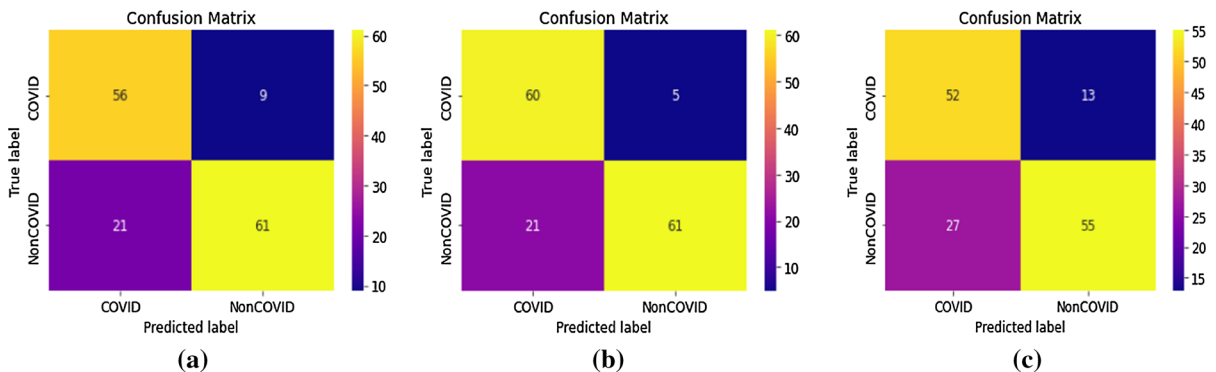


Fig. 10 Confusion matrix obtained by the InceptionV3 model after **a** 100, **b** 200 and **c** 300 epochs

115 images out of 147 are considered to be well classified while 32 images are misclassified

- 128 images against 19 badly classified for a number of epochs of 200 and 119 against 28 for the 300 epochs.

As for the CXR images classification (Fig. 16), the model has achieved:

- 190 images were correctly classified, and 10 are considered wrongly classified for 100 epochs.
- For 200 and 300 epochs, 11 are considered wrongly classified and 189 images were correctly classified.

The **ResNet-50** confusions matrices of CT classification (Fig. 9) are:

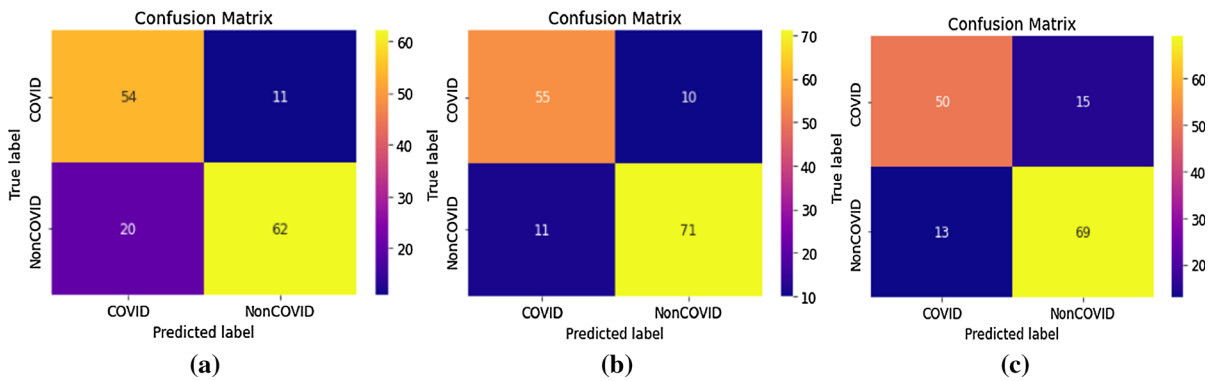


Fig. 11 Confusion matrix obtained by the Xception model after **a** 100, **b** 200 and **c** 300 epochs

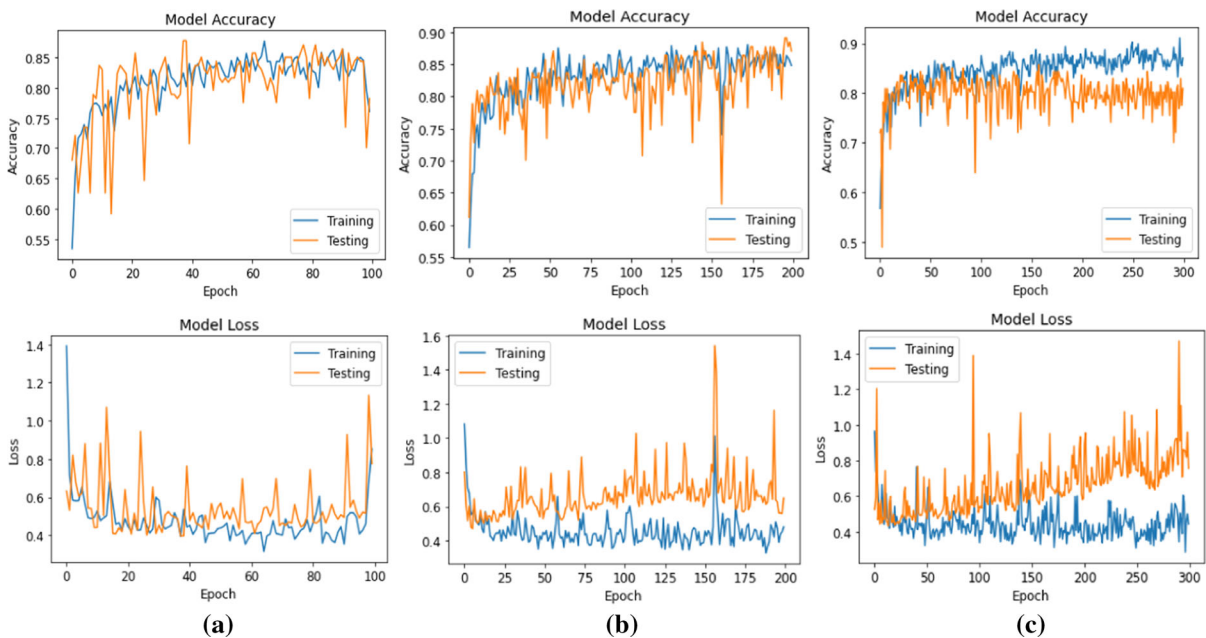


Fig. 12 Precision and loss plots on training and validation sets for the VGGNet-19 model after **a** 100 epochs, **b** 200 epochs, **c** 300 epochs

- The total number of patients reached by the COVID-19 for a number of epochs of 100 is equal to 52 including 37 TP and 15 FN. The number of healthy patients is 95 divided into 28 FP and 67 TN.
- While for a number of epochs of 200 and 300 this model classifies a high number of FN respectively 45 and 60, which is an indicator of a misclassification.

By classifying CXR images (Fig. 17), we obtain:

- An improvement of the classification where the number of misclassified images decreased for the 3 numbers of epochs: 21 images, 15 images and 18 images for respectively 100, 200 and 300 model epochs classified.
- The high number of well-classified images was obtained with a number of epochs of 200 which is 185 images.

From the resulted confusion matrices of the **InceptionV3** model (Figs. 10 and 18).

For the CT images we have:

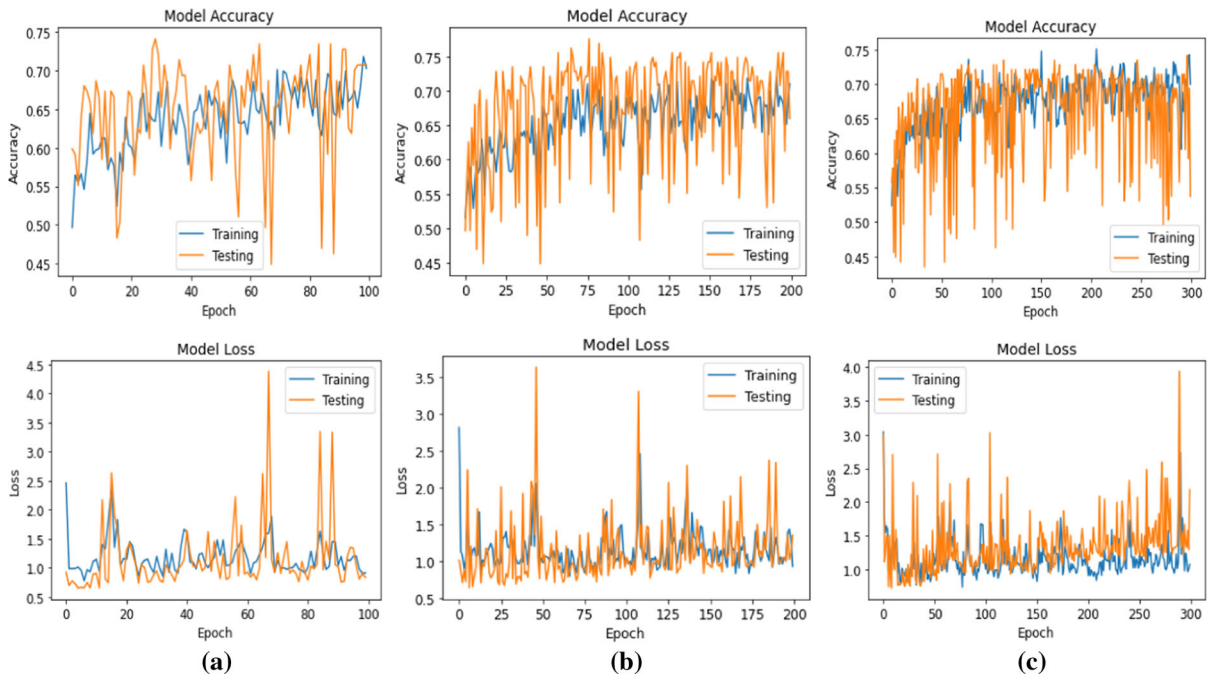


Fig. 13 Precision and loss plots on training and validation sets for the ResNet50 model after **a** 100 epochs, **b** 200 epochs, **c** 300 epochs

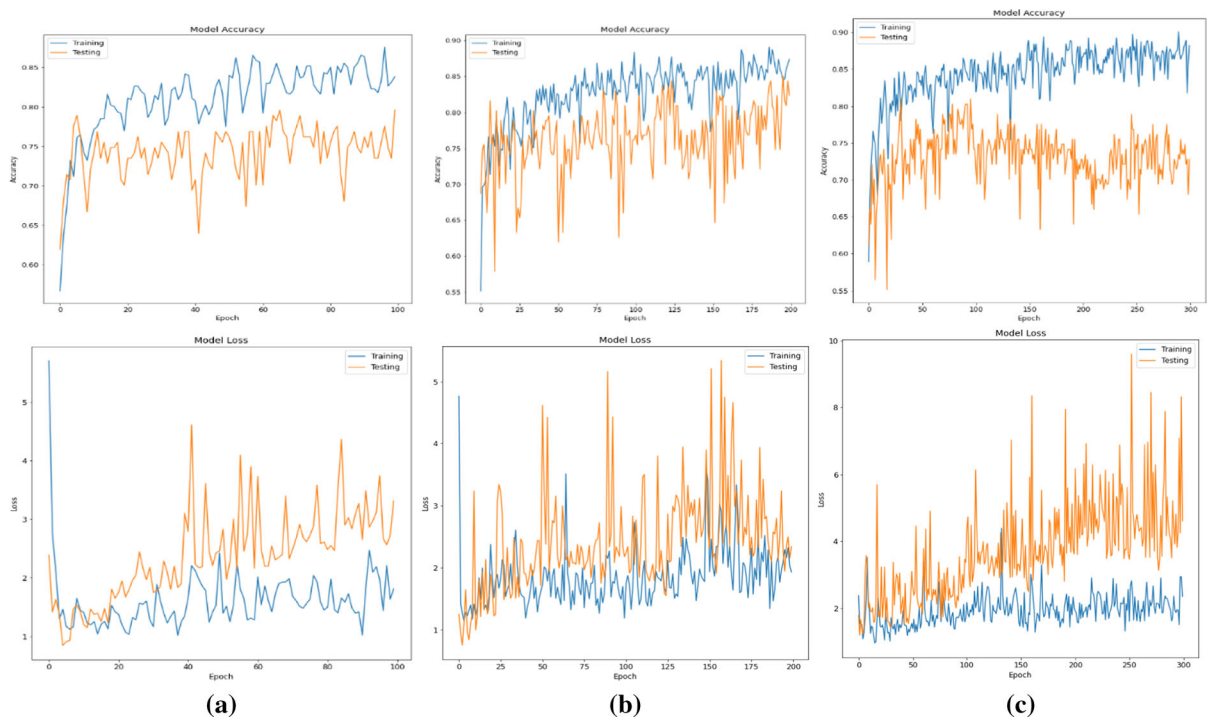


Fig. 14 Precision and loss plots on training and validation sets for the InceptionV3 model after **a** 100 epochs, **b** 200 epochs, **c** 300 epochs

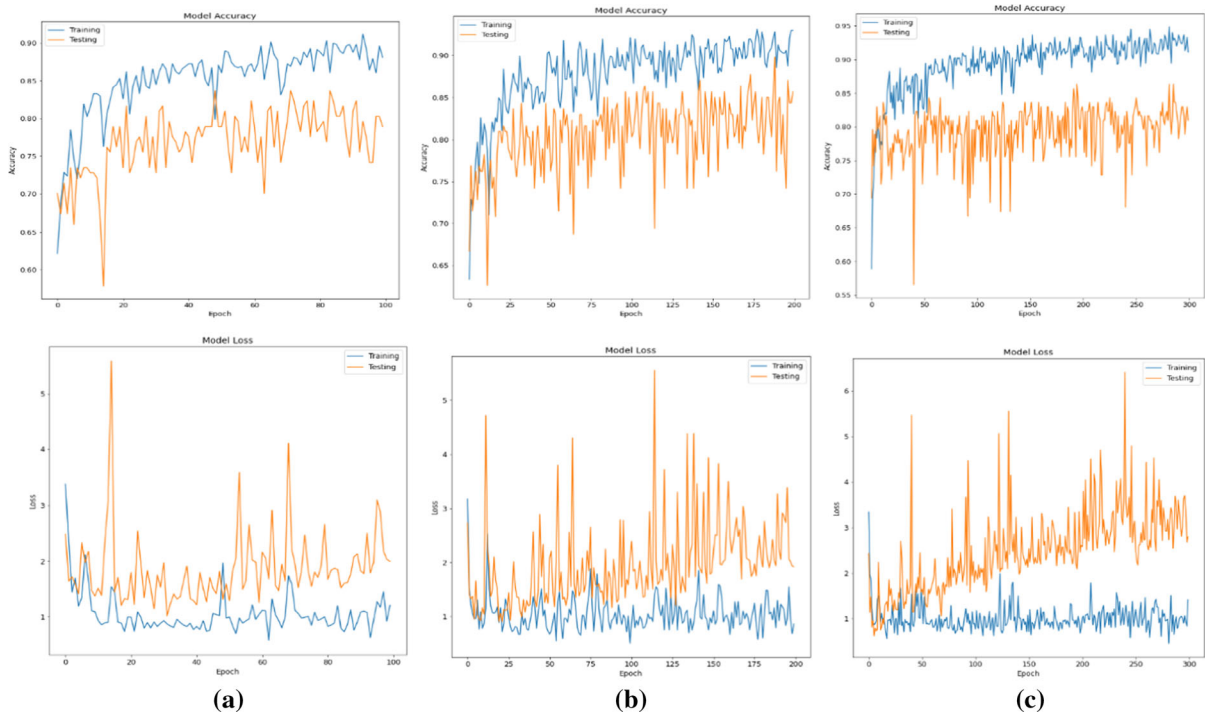


Fig. 15 Precision and loss plots on training and validation sets for the Xception model after **a** 100 epochs, **b** 200 epochs, **c** 300 epochs

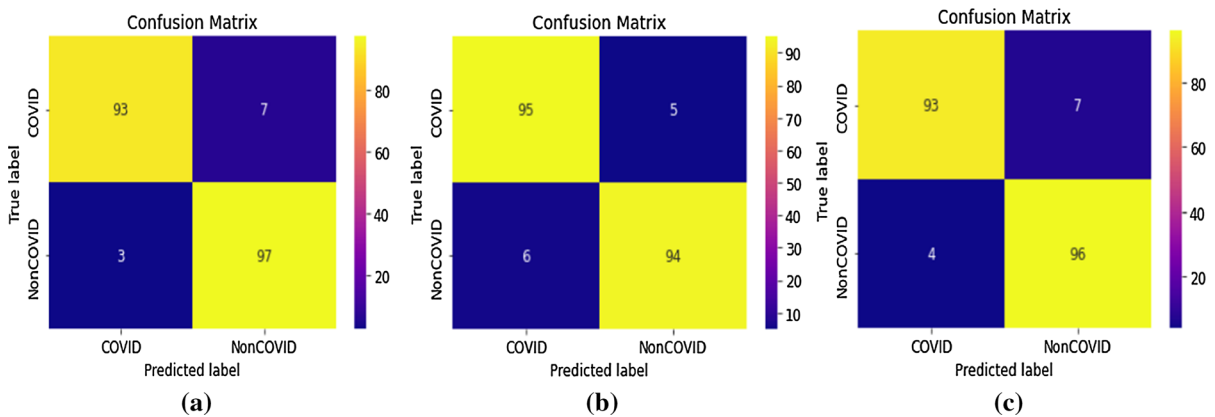


Fig. 16 Confusion matrix obtained by the VGGNet-19 model after **a** 100, **b** 200 and **c** 300 epochs

- 117 images were correctly classified, and 30 are considered wrongly classified for 100 epochs.
- As for the 200 epochs, 121 were correctly detected, and the remaining 26 were wrongly classified, 5 as FP and 21 as FN cases.
- Using 300 epochs, 107 were correctly classified, 27 cases were diagnosed as FN, and 13 cases as FP.

As for the CXR images classification, the model has achieved:

- For 100 and 300 epochs, 186 were correctly classified, and 14 were wrongly classified.
- Behind, using 200 epochs only 9 images were wrongly classified and 191 images were correctly classified.

Figures 11 and 23 shows the **Xception** confusions matrices. The results of CT classification are:

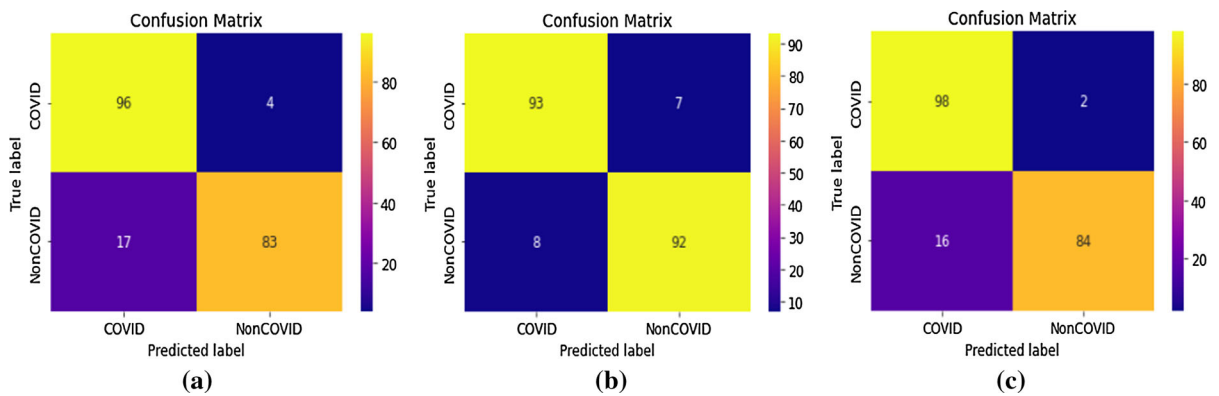


Fig. 17 Confusion matrix obtained by the ResNet50 model after **a** 100, **b** 200 and **c** 300 epochs

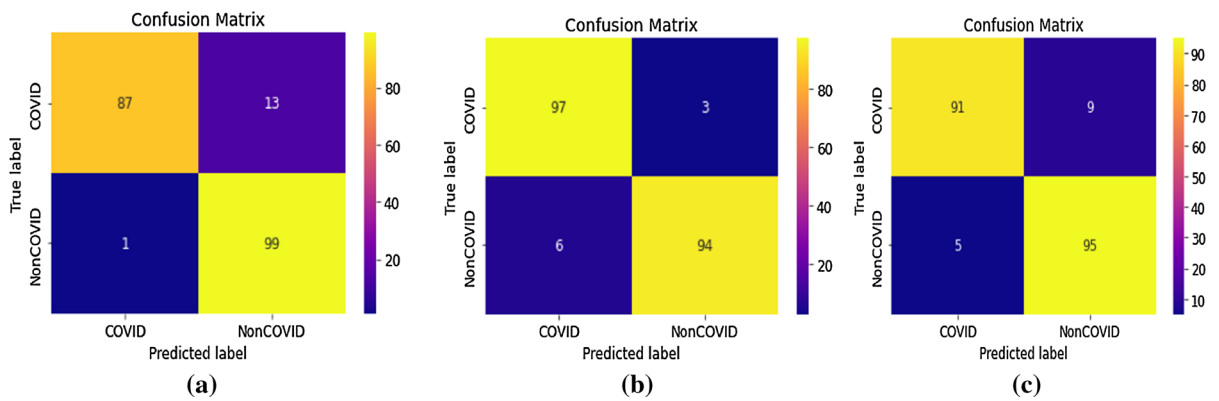


Fig. 18 Confusion matrix obtained by the InceptionV3 model after **a** 100, **b** 200 and **c** 300 epochs

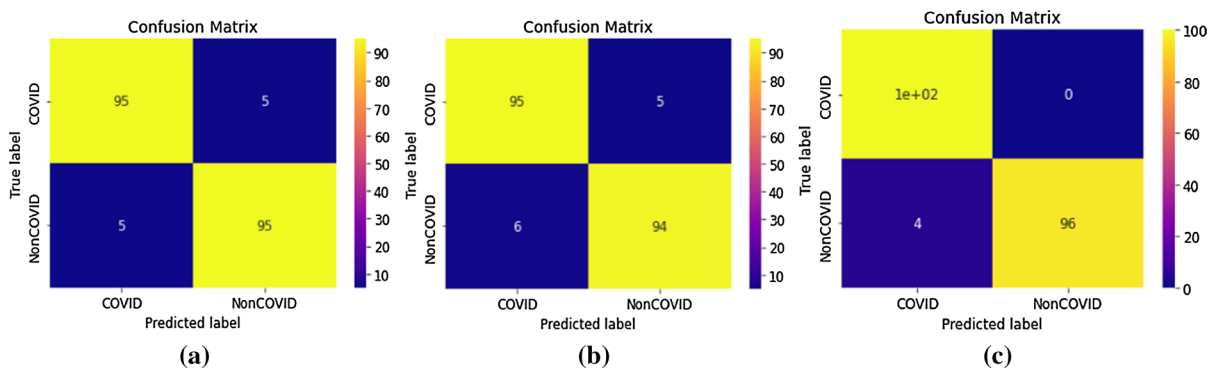


Fig. 19 Confusion matrix obtained by the Xception model after **a** 100, **b** 200 and **c** 300 epochs

- For 100 epochs 116 images out of 147 are considered to be well classified while 31 images are misclassified
- For 200 epochs, 126 images were correctly classified, and 21 were wrongly classified.
- While using 300 epochs the model classified 119 images correctly and the rest is considered misclassified 28.

The CXR images classification demonstrate that:

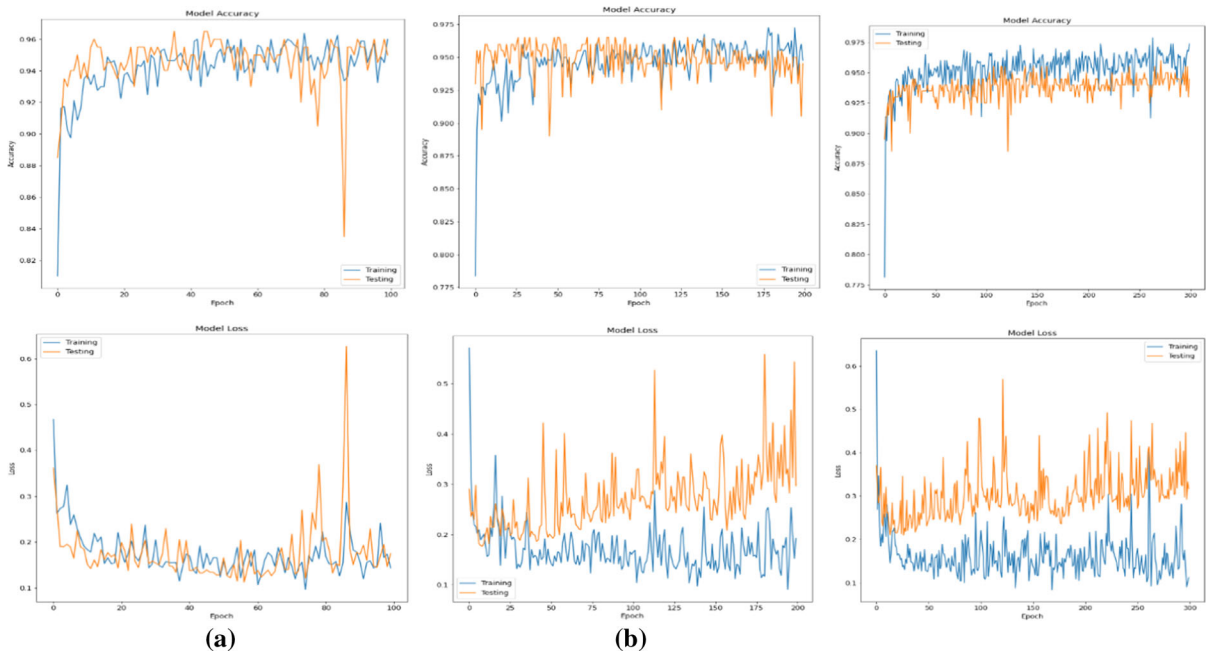


Fig. 20 Precision and loss plots on training and validation sets for the VGGNet-19 model after **a** 100, **b** 200 and **c** 300 epochs

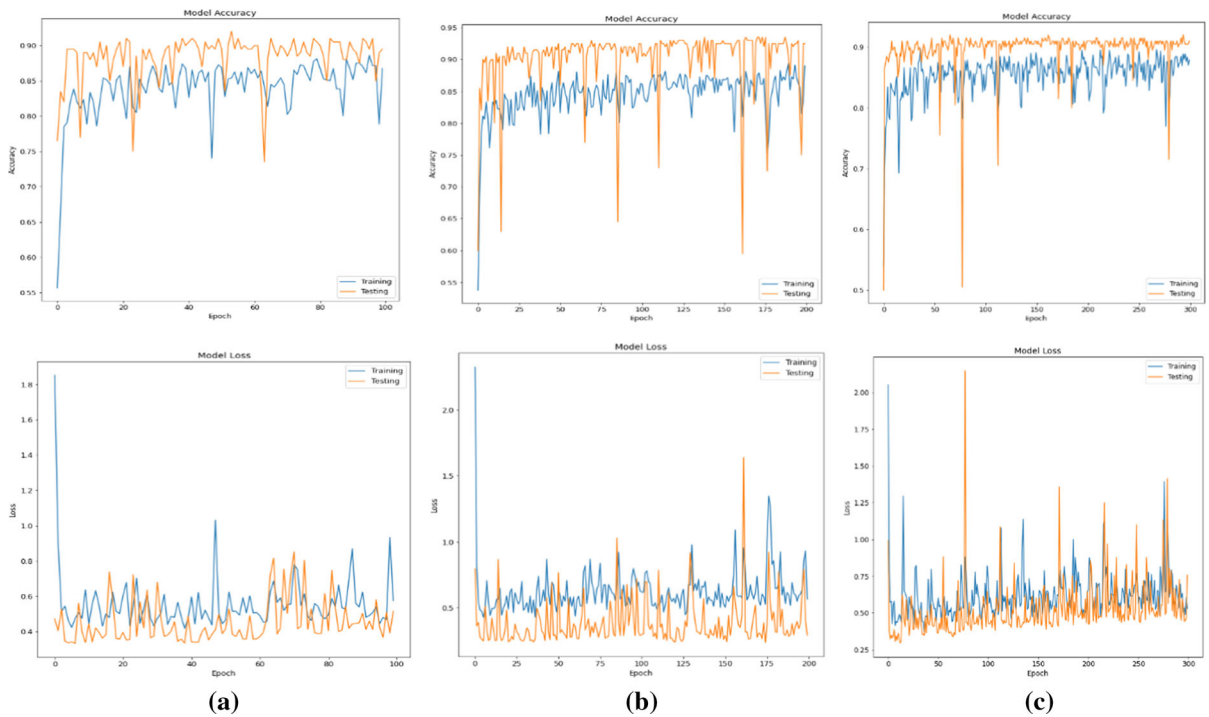


Fig. 21 Precision and loss plots on training and validation sets for the ResNet50 model after **a** 100, **b** 200 and **c** 300 epochs

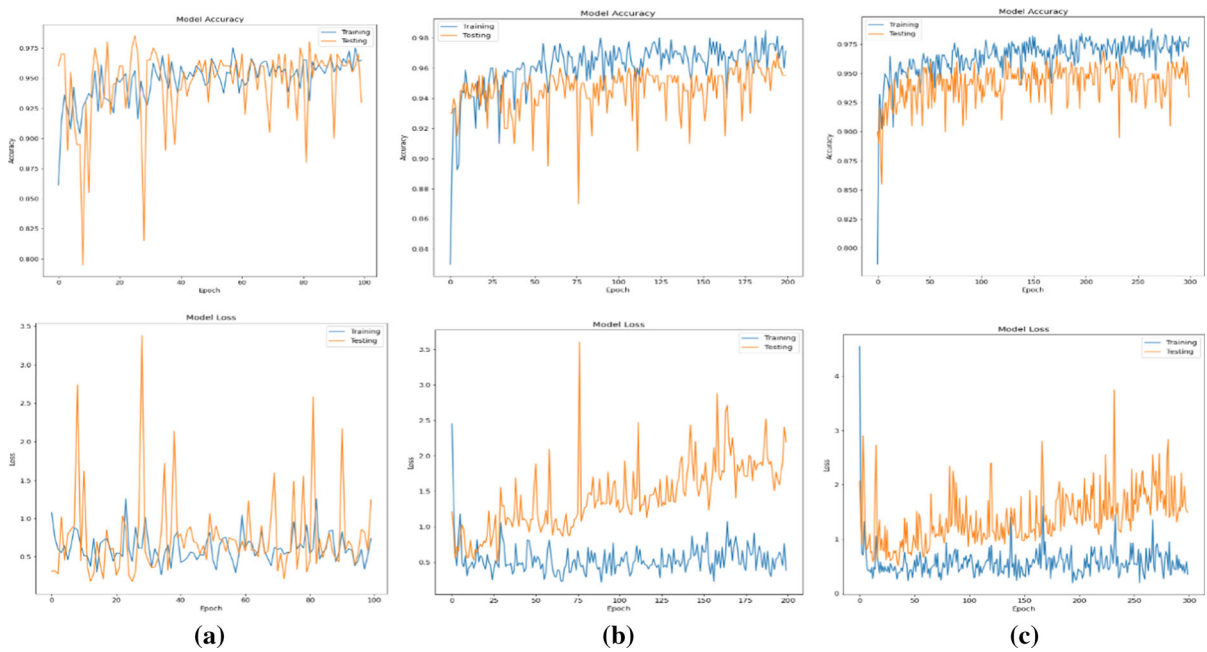


Fig. 22 Precision and loss plots on training and validation sets for the InceptionV3 model after **a** 100, **b** 200 and **c** 300 epochs

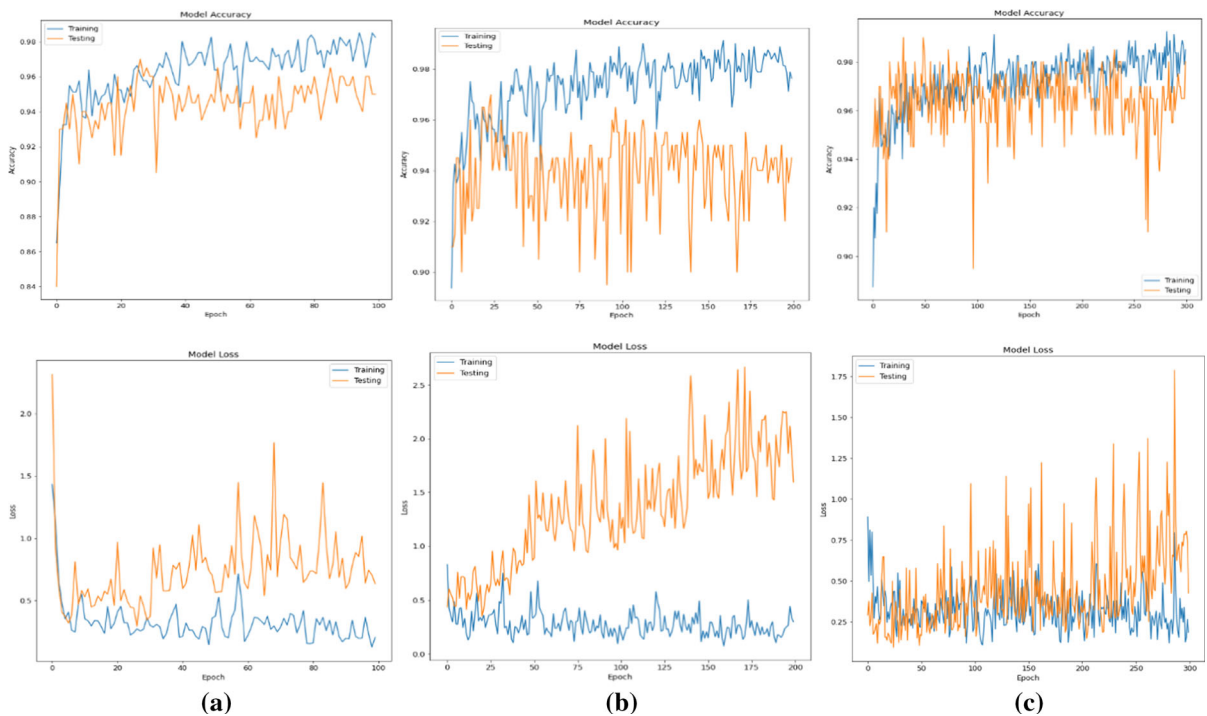


Fig. 23 Precision and loss plots on training and validation sets for the Xception model after **a** 100, **b** 200 and **c** 300 epochs

- 190 images were correctly classified, and 10 were wrongly classified were 5 considered as FP, and 5 as FN for 100 epochs.
- For 200 epochs 189 were correctly classified than 11 are misclassified.
- The highest number of well classifying images was obtained using 300 epochs where the model has classified 196 correctly and 4 images were wrongly classified.

Based on the results provided in the confusion matrices presented previously, we notice that the prediction of FN is very dangerous for the patients themselves and the society, since the infected patients are declared healthy and therefore, they lead a life normal without taking any action for themselves and for public health. This is important since the main goal of this research is to minimize the cases of FN COVID-19 in order to better support the clinical decision.

From the figures that represent the graphical performance evaluation of the deep learning models tested during the learning and validation stage, we notice that the accuracy of the Xception model training is more raised than others and the loss gradually decreases.

To compare models' performances within a dataset, and show the CXR and CT imaging techniques' effectiveness, we calculated the overall precision, recall, F1-score and accuracy.

A summary of the performance metrics of the 4 models taking into account the CXR and CT datasets is are illustrated in Table 3.

Discussion

Four deep learning architectures for COVID-19 detection in the human pulmonary system are proposed in this paper. These architectures are used to classify patients' COVID-19 and normal status.

In the above, we have used CT and CXR images for screening and diagnosis of COVID-19 disease.

On these images, the four modified classification models based on VGGNet-19, ResNet50, InceptionV3 and Xception were trained and tested. The training process was performed in the 100, 200, and 300 epochs for all pre-trained models. Performance of

which has been evaluated using the performance metrics as discussed in the previous section.

Of all the models offered, Xception achieves 98% accuracy depending on the CXR dataset applied using 300 epochs and 81% using CT scan images.

The table highlights that the two proposed refined versions of VGGNet-19 and InceptionV3 have a similar accuracy rate (accuracy) equal to 95% and very similar performance (precision, recall, F1-score); for the CXR dataset. Whereas for CT images we had 87% and 82% (accuracy) in favor of VGGNet-19.

For the ResNet50 model, the lowest performance value were achieved with an accuracy of 71% for the CT images, and 90% using CXR images. Low recall value means more false negative (FN) cases. Thus, the low accuracy rate indicates a high number of false positive (FP) results. This implies that this model is a classifier of low skills and which classifies a high number of FP and FN.

The Xception model performed best in terms of precision 98%, recall 98%, accuracy 98%, and F1-score 98% for the CXR dataset using 300 epochs. These high values imply that the model was successful in classifying the majority of images and it is then considered to be a competent classifier which is most suited to our problem.

The InceptionV3 model demonstrates superior performance using 200 epochs on the CXR dataset by achieving an accuracy of 95% which surpasses the method based on ResNet50 in term of recall 95,5%, precision 95,5% and F1-score 95,5%. Using CT scan images performances are: 82% for both accuracy and F1-score while precision and recall are equal to 83% for InceptionV3 model.

VGGNet-19 has a precision rate equal to that of InceptionV3 for CXR image knowing that in terms of recall, precision and F1-score, the InceptionV3 model surpasses VGGNet-19 by 0.5%.

On the other hand, using CT scan images VGGNet-19 was considered as the best model adopted due to the high rates of precision, recall, F1-score and accuracy which is equal to 88,5%, 86%, 86,5% and 87%.

Finally, we provide the F1-score, a metric that takes precision and recall into account, Xception achieved the highest F1-score of 98%, using the CXR dataset, while InceptionV3 classifying images has 95.5%. These values are 95% and 92.5%

Table 3 Classification report of deep learning models

Classifier	Nb epochs	Precision %			Recall %			F1-score %			Accuracy %		
		CT scans	CXR	Mean	CT scans	CXR	Mean	CT scans	CXR	Mean	CT scans	CXR	Mean
VGGNet-19	100	81,5	95	88,3	80	95	87,5	78	95	86,5	78	95	86,5
	200	88,5	94,5	91,5	86	94,5	90,3	86,5	94,5	90,5	87	94	90,5
	300	80,5	94,5	87,5	80,5	94,5	87,5	80,5	94,5	80,5	81	94	87,5
ResNet50	100	71	90	80,5	89,5	69,5	79,5	69,5	89,5	79,5	71	90	80,5
	200	72,5	92,5	82,5	68,5	92,5	80,5	65,5	92,5	79,0	66	93	79,5
	300	61	92	76,5	57,5	91	74,3	51	91	71,0	54	91	72,5
InceptionV3	100	80	93,5	86,8	80	93	86,5	79,5	93	86,3	80	93	86,5
	200	83	95,5	89,3	83	95,5	89,3	82	95,5	88,3	82	95	88,5
	300	73,5	93	83,3	73,5	93	83,3	73,5	93	83,3	73	93	83,0
Xception	100	79	95	87,0	79,5	95	87,3	79	95	87,0	79	95	87,0
	200	80,5	94,5	87,5	80,5	94,5	87,5	80,5	94,5	87,5	81	94	87,5
	300	80,5	98	89,3	80,5	98	89,3	80,5	98	89,3	81	98	89,5

respectively for VGGNet-19 and ResNet50 (with the same CXR images).

The CT scan images give low values in terms of 4 performance criteria for the four models offered. The highest precision is reached by VGGNet-19 after 200 epochs and which had an F1-score of 86.5%.

Comparing the CT scan and CXR images, the CXR showed better performance in detecting positive cases using the same models. This imaging technique provides better contrast and creates detailed quality images on CT, which helps models extract relevant information from the images.

By adding the average of the two modalities CXR and CT for each model. According to the table and considering the average of the scores of the 4 models, we can see that the VGGNet-19 model has the best scores, namely 91.5%, 90.3%, 90.5% and 90.5% respectively for Precision, Recall, F1-score and Accuracy using 200 epochs.

While Xception is the 2nd place model which gave us a recall, precision, F1-score equal to 89,3% and an accuracy equal to 89,5%.

ResNet50 model has the lowest performance values which were achieved with an accuracy of 80,5%, a recall and F1-score of 79,5% while the precision is equal to 80,5% for a number of epochs of 100.

So taking into account the average of the two modalities CXR and CT, we notice that VGGNet-19 presents the best score. If we consider each modality

separately, it can be noted that the VGGNet-19 model overcomes all the other models for the CT modality while the Xception model is the best for the CXR modality.

Table 4 shows a summary of the findings on COVID-19 diagnosis using CT and CXR images, as well as a comparison to our proposed method. Our proposed model is performing better or comparable to most of the already available works in this field. It can be observed that in terms of precision value, the two proposed approaches based on VGGNet-19 and Xception exceed previous methods.

In the first reference (Mohamed Loey et al. 2020), five alternative deep convolutional neural network models (AlexNet, VGGNet16, VGGNet19, GoogleNet, and ResNet50) were chosen in order to detect the patient infected with the coronavirus using digital

Table 4 Comparison of automatic detection methods for COVID-19

References	Used Model	Accuracy
Loey et al. (2020)	ResNet50	82,91%
Xu et al. (2020)	ResNet	86,7%
Wang and Wong (2020)	COVID-Net	83,5%
Proposed Method	VGGNet-19	90,5%
	Xception	89,5%

X-ray pictures. The results reveal that ResNet50 is the best model for detecting COVID-19 from a small chest CT data set utilizing traditional data augmentation, with an 82.91% test accuracy.

The authors looked at a dataset called COVIDx and COVID-Net that was designed to detect COVID-19 from chest X-ray images in Wang and Wong (2020). Their proposed method achieves an accuracy value of 83.5%. The dataset comprises four types of chest X-rays: uninfected cases, bacterial X-rays, COVID-19 viral X-rays, and COVID-19 pneumonia non-positive X-rays.

In the reference (Xiaowei Xu et al. 2020), authors established a new method for automatically COVID-19 screening, using deep learning approaches. They demonstrated that models with the attention to localization mechanism are capable of effectively classifying COVID-19 on CXR, with an overall accuracy rate of roughly 86.7%.

Conclusion

A binary-classification deep learning model for detecting COVID-19 from CT scans was developed and tested in this study. We tested four deep learning techniques such as VGGNet-19, ResNet50, InceptionV3 and Xception to verify the validity of the proposed system, as it allows us to know the advantages and disadvantages of each technique.

The results obtained showed that if we consider each modality separately, we can note that the VGGNet-19 model outperforms the designed versions of all models for the CT modality while the refined Xception version is the best for the CXR modality. The CXR images have better performance in detecting COVID-19 cases.

On the other hand, taking into account the average of the two modalities CXR and CT, VGGNet-19 gives us the highest score.

In order to show the feasibility of our approach, we compared the results obtained by our system with other state-of-the-art approaches. Based on the evaluations conducted in this brief, we have shown that the Xception and VGGNet-19 models perform well and provide 89,5% and 90,5% accuracy.

Based on the research findings of our study, we can conclude that our approach will be useful in assisting doctors and health professionals in making clinical

decisions in order to locate COVID-19 as soon as feasible and accurately.

Indeed, if this system is applied in hospitals, the number of treatments will be reduced and it will contribute to reducing the intervention of the doctor, especially in certain cases which do not require his intervention at all, which will reduce the medical material and those affected.

Even though experiments have shown promising results, we argue that there are still some limitations and recommendations for future work.

References

- Ai T, Yang Z, Hou H et al (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease (COVID19) in China: a report of 1014 cases. *Radiology* 296(2):E32–E40. <https://doi.org/10.1148/radiol.2020200642>
- Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 43:635–640. <https://doi.org/10.1007/s13246-020-00865-4>
- Bansal N, Sridhar S (2020) Classification of X-ray images for detecting Covid-19 using deep transfer learning. *Res Square*. <https://doi.org/10.21203/rs.3.rs-32247/v1>
- Chen N, Zhou M, Dong X et al (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 395(10223):507–513. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)
- Elpeltagy M, Sallam H (2021) Automatic prediction of COVID-19 from chest images using modified ResNet50. *Multimed Tools Appl* 80:26451–26463. <https://doi.org/10.1007/s11042-021-10783-6>
- Guefrechi S, Ben Jabra M, Ammar A, Koubaa A, Hamam H (2021) Deep learning based detection of COVID-19 from chest X-ray image. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-021-11192-5>
- 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 (CVPR)*, pp 770–787. <https://doi.org/10.1109/CVPR.2016.90>
- Loey M, Manogaran G, Khalifa NEM (2020) A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images. *Neural Comput Appl*. <https://doi.org/10.1007/S00521-020-05437-X>
- Russakovsky O, Deng J, Su H et al (2014) ImageNet large scale visual recognition challenge. <http://arxiv.org/abs/1409.0575>
- Sakib S et al (2020) Detection of COVID-19 disease from chest X-ray images: a deep transfer learning framework. *MedRxiv*. <https://doi.org/10.1101/2020.11.08.20227819>
- Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: *The 3rd*

- international conference on learning representations (ICLR2015). <https://arxiv.org/abs/1409.1556>
- Wang L, Wong A (2020) COVID-Net: a tailored deep convolutional neural network design for detection of COVID19 cases from chest radiography images. ArXiv200309871. <http://arxiv.org/abs/2003.09871>
- Wang S, Kang B, Ma J et al (2020) A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *Eur Radiol* 31:6096–6104. <https://doi.org/10.1007/s00330-021-07715-1>
- Xu X, Jiang X, Ma C et al (2020) Deep learning system to screen coronavirus disease 2019 pneumonia. *Engineering* 6:1122–1129. <https://doi.org/10.1016/j.eng.2020.04.010>
- Zunyou W, McGoogan JM (2020) Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 323(13):1239–1242. <https://doi.org/10.1001/jama.2020.2648>

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