Check for updates

Recognizing COVID-19 from chest X-ray images for people in rural and remote areas based on deep transfer learning model

Mamoun Qjidaa · Anass Ben-Fares · Hicham Amakdouf ·
Mostafa El Mallahi · Badre-eddine Alami ·
Mustapha Maaroufi, et al. *[full author details at the end of the article]*

Abstract

In this article, we propose Deep Transfer Learning (DTL) Model for recognizing covid-19 from chest x-ray images. The latter is less expensive, easily accessible to populations in rural and remote areas. In addition, the device for acquiring these images is easy to disinfect, clean and maintain. The main challenge is the lack of labeled training data needed to train convolutional neural networks. To overcome this issue, we propose to leverage Deep Transfer Learning architecture pre-trained on ImageNet dataset and trained Fine-Tuning on a dataset prepared by collecting normal, COVID-19, and other chest pneumonia X-ray images from different available databases. We take the weights of the layers of each network already pre-trained to our model and we only train the last layers of the network on our collected COVID-19 image dataset. In this way, we will ensure a fast and precise convergence of our model despite the small number of COVID-19 images collected. In addition, for improving the accuracy of our global model will only predict at the output the prediction having obtained a maximum score among the predictions of the seven pre-trained CNNs. The proposed model will address a three-class classification problem: COVID-19 class, pneumonia class, and normal class. To show the location of the important regions of the image which strongly participated in the prediction of the considered class, we will use the Gradient Weighted Class Activation Mapping (Grad-CAM) approach. A comparative study was carried out to show the robustness of the prediction of our model compared to the visual prediction of radiologists. The proposed model is more efficient with a test accuracy of 98%, an f1 score of 98.33%, an accuracy of 98.66% and a sensitivity of 98.33% at the time when the prediction by renowned radiologists could not exceed an accuracy of 63.34% with a sensitivity of 70% and an f1 score of 66.67%.

1 Introduction

In February 2019, transmissible pneumonia in humans was first localize in China. This particular pneumonia causes a severe respiratory gene caused by the corona-2 virus called COVID-19. As a result, individuals of 16 to 21% according to the WHO fell seriously ill and consequently 2 to 3% died from the consequences of this highly contagious disease whose rate of transmissibility was around 3.77% [43]. Thus, to minimize the spread, it is therefore urgent to quick directly individuals infected with the virus to quarantine them and to start their treatment.

The diagnosis of COVID-19 based on well-defined criteria such as clinical symptoms, epidemiological history, positive CT and X-ray images, and positive pathogen tests based on real-time RT-PCR and virus nucleic acid sequencing [1, 9, 12]. To confirm these last two tests, it has been repeat several times in many cases and this is a serious limitation on their accuracy. To solve this problem, we are going to use chest X-ray images as a means to detect this virus. Indeed, the signature of atypical pneumonia due to COVID-19 will be manifest in an image by the presence of opacities in ground glass at the first stage of the disease, which manifests as pulmonary consolidation in the last stage of the disease [10, 25, 28]. Unfortunately, the features of atypical pneumonia caused by COVID-19 are confused with the features of other typical inflammatory pneumonia, making it difficult for the radiologist to detect COVID-19. As solution to this problem, the researchers used artificial intelligence and have already compiled a list of features extracted from images in order to detect the presence of the COVID-19 virus [22]. In literature [4, 7], radiologists have been able to develop clinically several characteristics of viral pathogens at the level of chest radiographic images.

In fact, the researchers noted that the presence of an uneven distribution of bilateral shadows and opacity of the ground glass is a strong indication of the presence of a COVID19 infection. Therefore, we propose to train a convolutional neural network (CNN) using these characteristics extracted from radiographic images to accurately predict infection by COVID19.

In this work, we will rely on our research in the field of artificial intelligence for the service of health [2, 3, 13, 15–20, 29–32, 34, 44]. We plan to offer a new approach for the early detection of COVID19 that will be make available to structures in rural and remote areas with a simple chest X-ray machine. To achieve this goal, we propose a transfer learning-based model to reduce learning step and speed up the whole training process. The transfer model offers a promising alternative to refine an already pre-trained CNN on a large dataset like ImageNet using its weights. In this way, it will help us to speed up the process of convergence during the training stage. Transfer learning is a very useful technique and has also obtained significant results in computer vision and other fields of application.

Moreover, our proposed model uses seven already pre-trained CNN extracted from the ImageNet database. The use of these pre-trained models, avoids the passage through the long process of learning models from scratch. We have chosen seven different models pre-trained on the ImageNet database: InceptionResNetV2 [6], DenseNet121 [24], MobileNet [21], InceptionV3 [40], Xception [8], and VGG16 and VGG19 [23]. Although these models have already been trained on the ImageNet dataset, we will also train them on our base of chest x-ray images that we have taken. The architecture of the proposed model begins with a pre-processing of the input images followed by an increase in quality of the data. Then, the model begins a step to extract the characteristics followed by the learning step. Finally, the model begins a classification and prediction process with a fully connected network made up of

several classifiers. To show the location of the important regions of the image which strongly participated in the prediction of the considered class, we will use the Gradient Weighted Class Activation Mapping (Grad-CAM) approach [37]. A comparative study was carried out to show the robustness of the prediction of our model compared to the visual prediction of radiologists.

The rest of the paper is organized as follows. In section II, we will present the development methodology of our classification model, which is composed of four stages: the collection of images to form the database, the pre-processing and the increase of the collected data, transfer learning model and presentation of the pre-trained networks of our model. In section III, we present our simulation results. In section 4, we present the regions of interest which participated enormously in the classification. In section 5, we present some prediction results performed by a renowned radiologist. We end our article with a discussion followed by a general conclusion.

2 Related work and materials and methods

Recently, the scientists in COVID-19 classification has numerous methods with different datasets used in some study. Furthermore, many attempts are proposed some ready network with some changes in some cases to enhance the performance of the classification. Myriad works on COVID-19 classification are available, and we will briefly discuss the most relevant works. The American College of Radiology (ACR) advised against the use of CTs and X-rays as a first-line diagnostic or screen tool for COVID-19 diagnosis (the American College of Radiology 2020, March 22). Where they indicated that the images could only show signs of infection that may be due to other factors. However, there have been plenty of studies where artificial intelligence was applied to test COVID-19 based on chest X-ray images [23, 26, 27, 33, 35, 38, 42]. In [33] Narin et al. experimented with three different CNN models (ResNet50, InceptionV3, and Inception-ResNetV2), and ResNet50 achieved the best accuracy of 98% for 2-class classification. Since they did not include pneumonia cases in their experiment, it is unknown how well their model would distinguish between COVID-19 and other pneumonia cases. Also The authores in Hemdan et al. in [23] used deep learning models in X-ray images to diagnose COVID-19 and suggested a COVIDX-Net model consisting of seven CNN models. Wang and Wong [42] presented a deep residual architecture called COVID-Net .it is one of the early works that has been done on COVID-19, which uses a deep neural network to classify chest X-ray images into three categories (COVID-19, Healthy, Non-COVID-19). In this study, we propose Deep Transfer Learning (DTL) model for recognizing covid-19 from chest x-ray images. The latter are less expensive, easily accessible to populations in rural and remote areas. In addition, the device for acquiring these images is easy to disinfect, clean and maintain. we will design a global transfer model composed of seven pre-trained convolutional neural networks (CNNs): DenseNet121, VGG16, VGG19, InceptionResNetV2, Xception, MobileNet and InceptionV3. The proposed model will address a three-class classification problem: COVID-19 class, pneumonia class, and normal class. To show the location of the important regions of the image which strongly participated in the prediction of the considered class, we will use the Gradient Weighted Class Activation Mapping (Grad-CAM) approach. A comparative study was carried out to show the robustness of the prediction of our model compared to the visual prediction of radiologists.

Our experiments were based on a chest X-ray image dataset collected the COVID-19 radiographic image database which was developed by Cohen JP [11] and the ChestX-ray14 database provided by Wang et al. [41].

CNN models were developed using Tensorflow, with a wrapper library Keras in Python. The experiments were performed on a Lenovo Thinkpad p51 with Intel® Core™ i7-7820HQ, 2.90 GHz, NVIDIA Quadro M2200 8GB GPU and 8GB of RAM. Adam optimization algorithm was used for hyperparameter optimization for training the CNN models, and cross-entropy as loss function. The learning rate is started from the value 0.001 and is reduced after 4 epochs if the loss value didn't improve with the help of callbacks function. The models were configured to train for 60 iterations.

The proposed methodology includes four steps. The first step deals with collecting the images to form the database. The second step deals with pre-processing and augmentation of the collected data using range rotation, width shift range, height shift range, zoom range, horizontal flip, and vertical flip. After these changes, it was possible to balance the Dataset from both COVID and NON-COVID classification on the testing and training sets. This database augmentation happens during run time, when chest X-ray images are presented as an input for the input image. The third step deals with presenting our proposed Convolutional Neural Networks (CNN) and the use of Deep Transfer Learning (DTL). The fourth step deals with pre-training CNN. We will present these steps in the following paragraphs.

3 Dataset

In this work, radiographic images were obtained from two different sources that were collected to form the lung radiographic image database. The first source concerns the COVID-19 radiographic image database, which was developed by Cohen J.P. [11]. This database contains images from various open-access sources. It is constantly updated with images shared by researchers from different regions. To date, 230 radiographic images have been diagnosed with COVID-19 in the database. We randomly selected 146 images of cases declared positive to COVID19. The second source concerns the ChestX-ray14 database provided by Wang et al. [41]. This database contains 5863 JPEG radiographic pneumonia (viral pneumonia / bacterial pneumonia) images. We randomly selected 210 images of cases with pneumonia and 200 images randomly selected from normal cases. Table 1 below shows the summary of the prepared dataset. Figure 1 below shows some samples of chest X-ray images from the prepared dataset. Thus, we collected a total of 556 radiological images classified into 3 classes: a class of COVID-19 type, a Class of Pneumonia type, and a class of Normal type. In the experimental analysis, 70% of the data set was used as training and 10% for validating data and 20% was used as the test set.

Table 1 Dataset summary

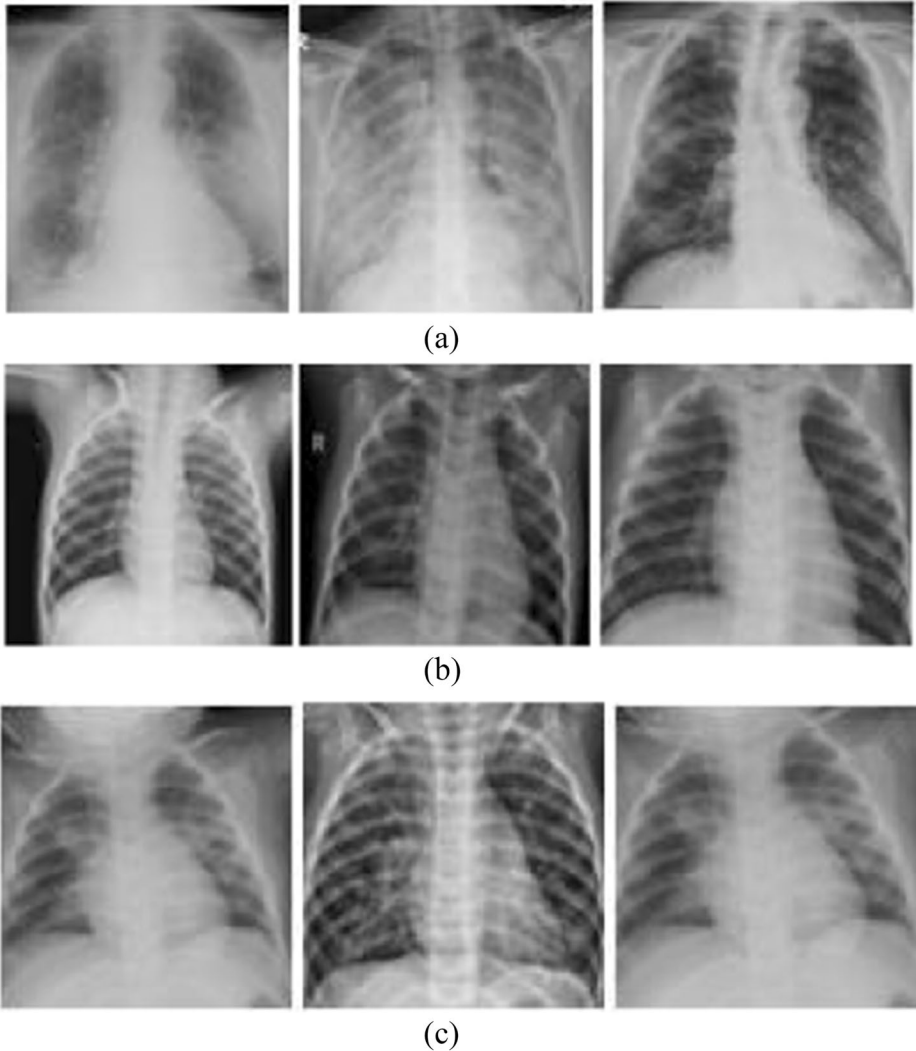| Disease	Number of images	Training set (70%)	Validation set (10% of training dataset)	Testing set (20%)
Normal	200	140	20	40
Pneumonia (viral / bacterial)	210	147	21	42
COVID-19	146	102	14	30

Fig. 1　X-ray Images of the three classes: **a** COVID-19, **b** Normal **c** Pneumonia

4 Data pre-processing and augmentation

Not all of the data available on the internet has been subject to the same pre-processing, most of our positive COVID-19 data have all the X-rays occupy most of the screen, with little or no black bars on the sides. This becomes an issue, as our models may later learn that it is enough to look at the black bars on the side, to decide whether the sample is the case of COVID-19, normal or pneumonia. After manual inspection of the data set, it became apparent that 10 to 20 COVID-19 images, 90 to 95 normal images, and 90 to 95 images of pneumonia have black bars. Therefore, to solve this issue, we create a script that removes these black pixels from the samples, an example of this process is illustrated in Figs. 1 and 2. To combine images for the training from different dataset, first we focus on the pixel scaling techniques for different dataset (Pixel Normalization, Pixel Centering, Pixel Standardization) to have the same
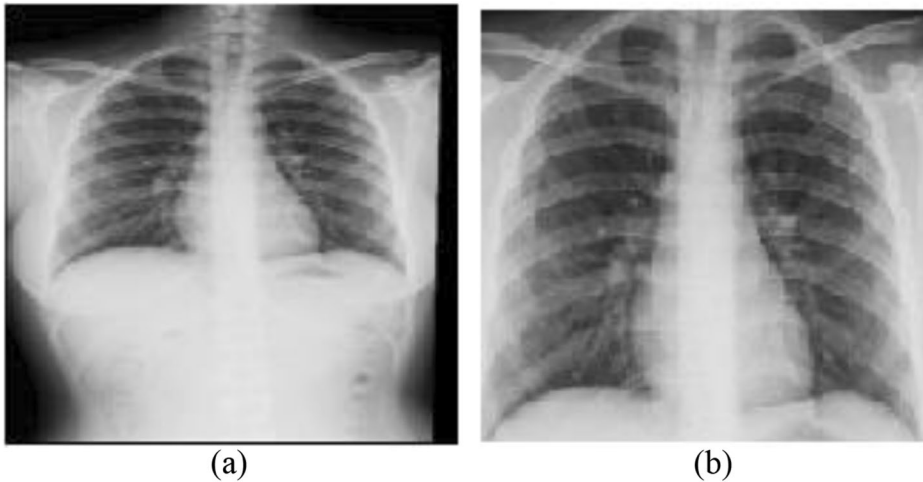
Fig. 2 Elimination of black bars on the sides of the image. **a** Original image, **b** Pre-processed image

distribution for the different datasets, In order to overcome the unbalanced data problem, we used resampling technique called random under-sampling which involves randomly deleting examples from the majority class until the dataset becomes balanced. We then resized all the images to the dimension of 224 × 224 pixels with a resolution of 72 dpi.

5 Convolutional neural networks using the deep transfer learning

In a traditional CNN, the input image usually goes through a sequence of convolution and pooling to squeeze the width and height dimension white, increase the feature depth. The purpose of this stacked convolution and pooling layer is to learn the representation of the image or features. The learned featured will be then the input to the fully connected layers for classification. In much previous architecture, researchers have tried to stack increasingly convolution and pooling layers together, hence the term Deep Learning, with an assumption that deeper networks will learn better representation of the object of interest. However, in practice, this design suffers from vanishing gradient issue and the performance degraded as the networks gets deeper. Often, forming a CNN from scratch is generally difficult as this process required large training data as well as significant expertise to select appropriate model architecture for appropriate convergence.

As a solution to this issue, we used Deep transfer learning model. This later have been used for different types of applications [14, 36]. This model offers a promising alternative to refine an already pre-trained CNN on a large dataset like ImageNet using its weights. This helps us to speed up the convergence process during training. Transfer learning is a very useful technique and has also achieved significant results in computer vision and image processing areas.

The proposed model is composed of two modes. A pre-trained start mode which transforms the input images into descriptor vectors. Another mode consists of several classifiers strongly connected together where each classifier will give its output its own prediction. In this way, the prediction which obtained a maximum score is that which will be retained by the global system at its exit. More precisely, the proposed model consists of three main processes as shown in

Fig. 3. The first process used is that of image pre-processing followed by an increase in data. The pre-treatment consisted in eliminating all that outside the zones, delimiting the two lungs while keeping an image of 224 × 224. Therefore, the increase in the number of images was achieved through horizontal feedback, random cropping and varying the intensity of each image.

The second process used is that of feature extraction and learning. At this step, the process uses seven already pre-trained CNN extracted from the ImageNet database. Furthermore, the use of these pre-trained models avoids the passage through the long process of learning models from scratch. Although these models have already been trained on the ImageNet dataset, we will also train them on our base of chest x-ray images that we have taken.

The third process used is that of classification and prediction through a fully connected CNN network made up of several classifiers. At the output of the network, a prediction vector is obtained. However, the final prediction retained by the global system is that which obtained a maximum score.
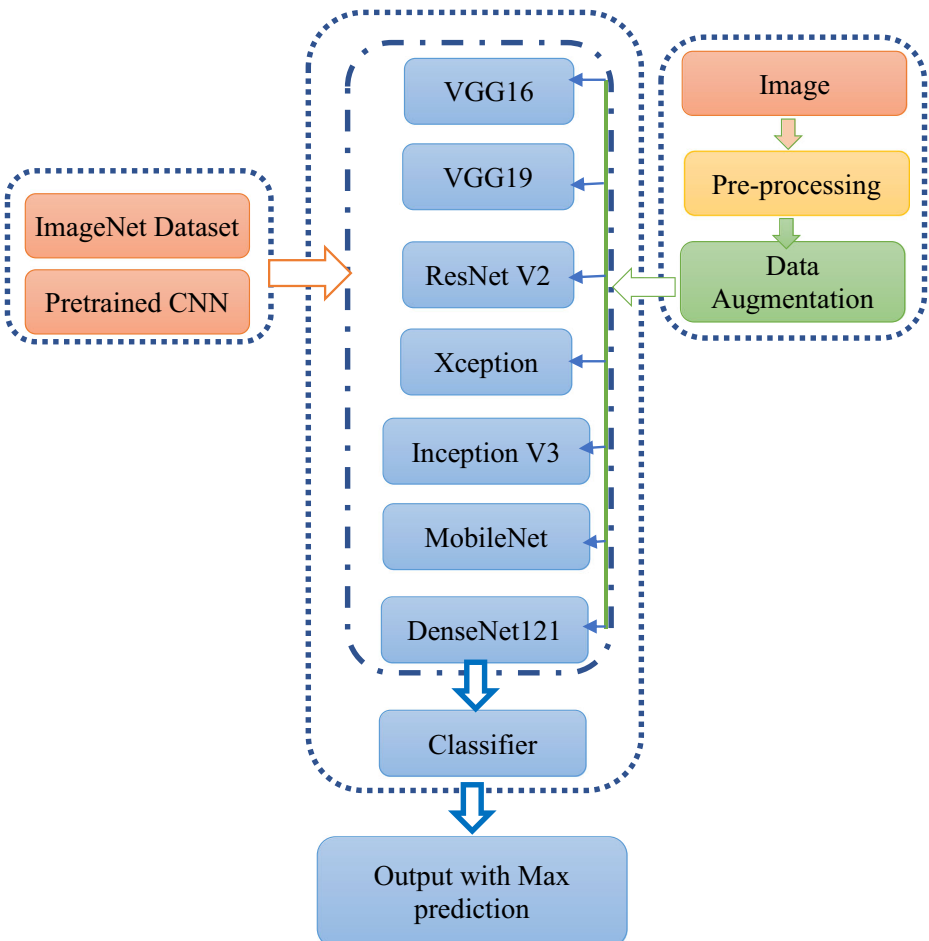

Fig. 3 Proposed model process

6 Pre-trained convolutional neural networks

The main advantage of proposed model is avoiding the passage through the long process of learning models from scratch through the use of the weights of the layers of already preformed models extracted from the ImageNet database. In a transfer model, these already predefined weights will be assigned to the first layers placed just after entering the model, while the weights of the last layers placed at the exit of the model will be trained on the images of the new domain. In the following sections, we will present different architecture models based on a convolutional neural network with different layers of depth.

6.1 VGG16 architecture

In this paper, we will use model with trainable and frozen layers based on a convolutional neural network model proposed by K. Simonyan [39]. We used the layers from CONV1–1 up to CONV4–3 as pre-trained layers with weights that are already fixed. Hence, the remaining layers were used for training on our dataset of radiographic lung images already in place as is shown in Fig. 4.

6.2 VGG19 architecture

The VGG-19 model is a trained Convolutional Neural Network based on VGG-16 architecture. While the number 19 stands for the number of layers with trainable weights. In total, we have 16 convolutional layers and 3 fully connected layers. In the proposed model, we used the layers from block1_conv1 to block5_conv1 as pre-trained layers with weights that are already fixed. The remaining layers were used for training on our dataset.

6.3 MobileNet architecture

MobileNet architecture was proposed by Google. In our proposed model, we used the layers from 1 to 75 as pre-trained layers with weights that are already fixed. Therefore, as shown in Fig. 5, the remaining layers were used for training on our dataset of radiographic lung images already in place.
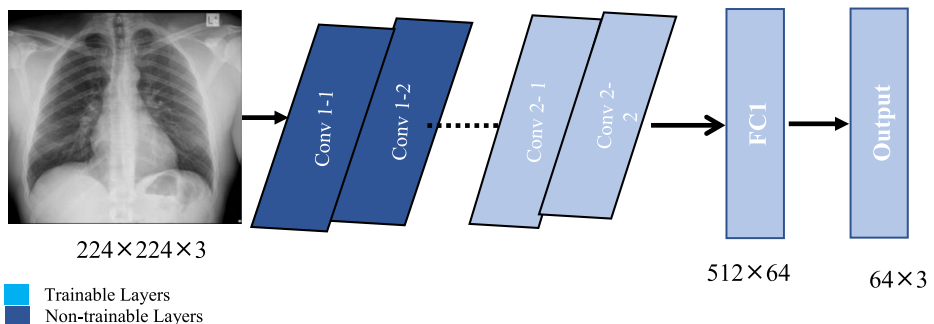

224×224×3

Conv 1-1 Conv 1-2   Conv 2-1 Conv 2-2   FC1   Output

512×64   64×3

Trainable Layers
Non-trainable Layers

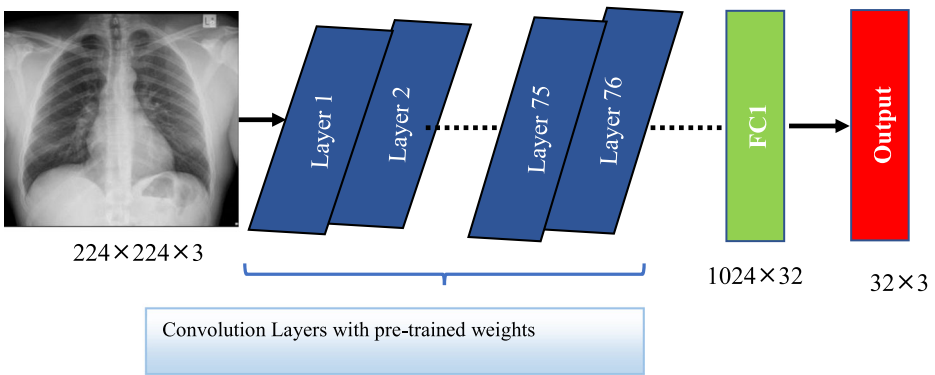**Fig. 4** VGG16 Model with trainable and frozen layers

Fig. 5 MobileNet model with trainable and frozen Layers

6.4 InceptionV3 architecture

InceptionV3 architecture is the third version introduced by Google to improve the inception convolutional neural network and it is composed by 48 deep layers. In our proposed model we used the layers from the first to mixed_10b layers as pre-trained layers with weights that are already fixed. Therefore, the remaining layers were used for training on our dataset of radiographic lung images already in place as is shown in Fig. 6.

6.5 Xception architecture

On another side, Xception architecture is a convolutional neural network that is 71 layers deep proposed by Francois Chollet [VGG]. In our proposed model, we used all the layers for training on our dataset of radiographic lung images with using pre-trained weights from ImageNet as a starting point, as is shown in Fig. 7.
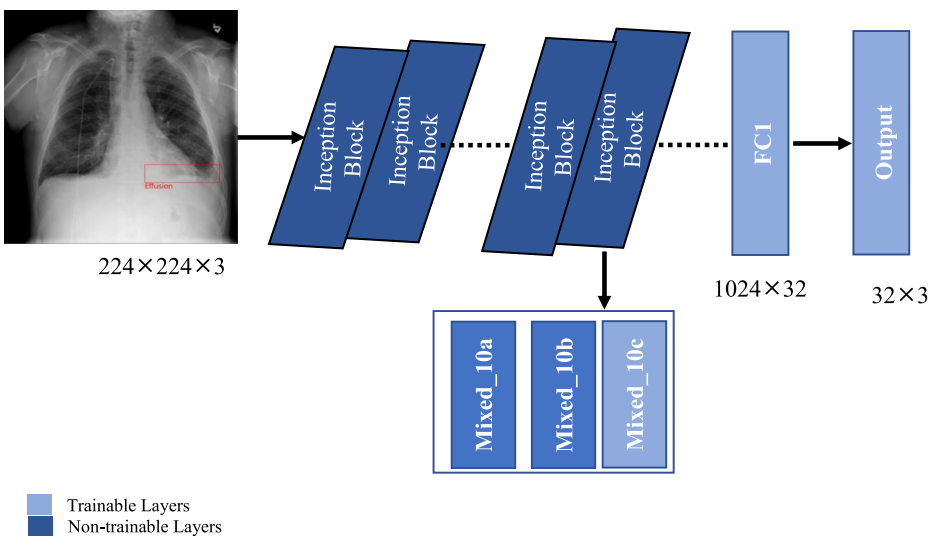

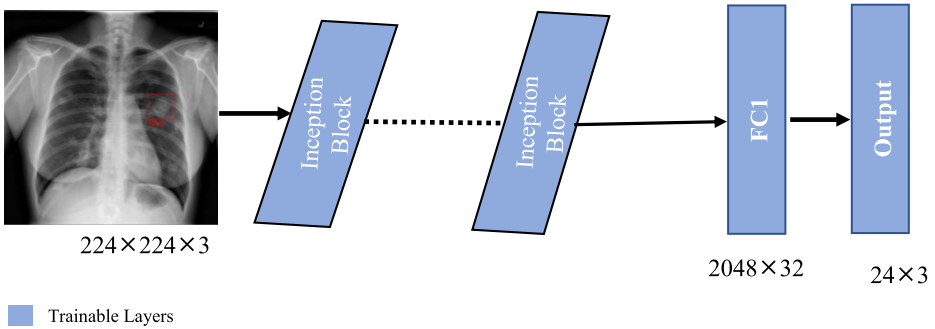**Fig. 6** InceptionV3 architecture with trainable and frozen Layers

Fig. 7 Xception architecture with trainable Layers

6.6 InceptionResNetV2 architecture

While InceptionResNetV2 architecture is obtained following modifications made by researchers on the third CNN inception version. The researchers drew on residual connections from the Microsoft ResNet network to offer a deeper, simple, and meaningful version of the Inception architecture. In our proposed model, we used the layers from the first to conv2D-58 layers as pre-trained layers with weights that are already fixed. In this way the remaining layers were used for training on our dataset of radiographic lung images already in place as is shown in Fig. 8.

6.7 DenseNet121 architecture

Finally, in DenseNet121 architecture, each layer has as input all the outputs of the layers that precede it, making the network architecture very dense, allowing deep supervision. In our proposed model, we used the layers from the first to conv3-block10–2-conv layers as pre-trained layers with weights that are already fixed. While, the remaining layers were used for training on our dataset of radiographic lung images already in place as is shown in Fig. 9.

6.8 Stacking model

In the remaining section we will present our new Stacking model that we developed. The latter is based on the neural network sub-models are integrated into an overall stacking model. Our
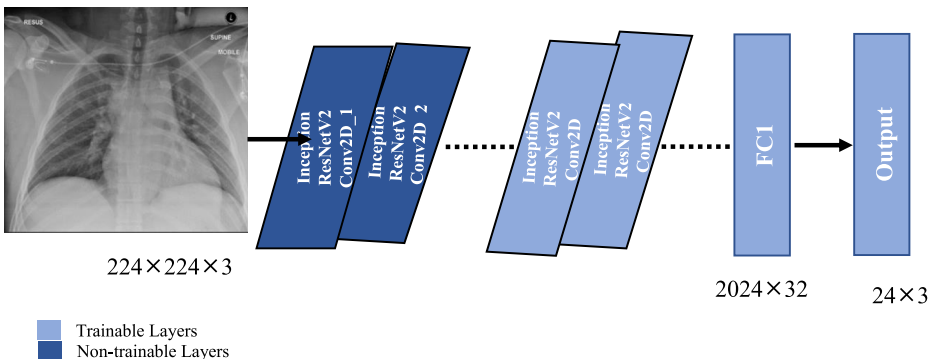

Fig. 8 Inception-ResNetV2 architecture with trainable and frozen Layers

224×224×3

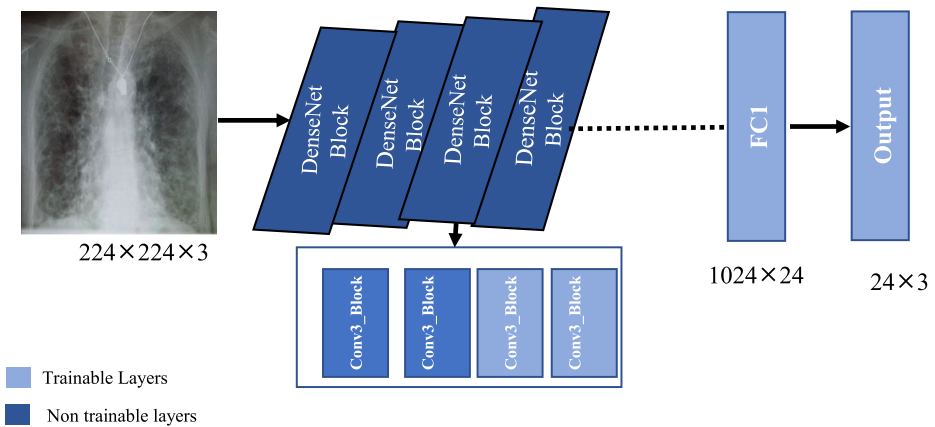1024×24   24×3

Trainable Layers

Non trainable layers

Fig. 9 DenseNet121 architecture with trainable and frozen Layers

new model allowed us to find the best way to combine the predictions of several existing models already pre-trained. Consequently, we will develop a stacking model using 7 neural networks as a sub-model and a scikit-learn classifier as a meta-learner. The basic idea of this approach is to consider the prediction of each network by assigning it a score. Once the seven networks have finished their prediction, the score obtained by each prediction is counted. Only the prediction with the highest score is selected as output from the model, as is shown in Fig. 10.

7 Performance metrics evaluation

The following metrics were used to validate the CNN system:
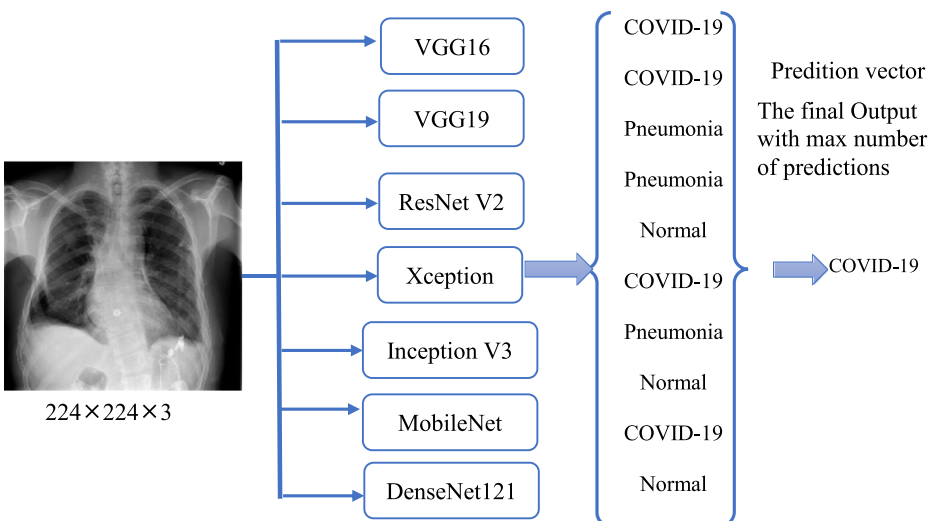

224×224×3

Fig. 10 Stacking model using seven different pre-trained architecture and majority voting

- Accuracy (ACC): accurate classification rate as per the total number of elements.
- Precision (P): ratio of true positive classifieds.
- Recall/Sensitivity (SE): true positive rate.
- Specificity (SP): true negative rate.
- F1-Score: relationship between precision and recall/sensitivity.

They are commonly used to assess the performance of classification algorithms. There is a standard way to show the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) to be more visual. This method is called the confusion matrix.

The confusion matrix allows us to determine the following metrics

$$Accurency : ACC = \frac{No.of\ images\ correctly\ classified}{Total\ no.of\ images}$$

$$Precesion : P = \frac{Sum\ of\ all\ True\ Positives\ (TP)}{Sum\ of\ all\ True\ Positives\ (TP) + All\ False\ Positives(FP)}$$

$$Recall/Sensitivity : SE = \frac{Sum\ of\ all\ True\ Positives\ (TP)}{Sum\ of\ all\ True\ Positives\ (TP) + Al\ l\ False\ Negatives(FN)}$$

$$Specificity : SP = \frac{TN}{TN + FP}$$

$$F1\ Score : S1{-}Score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}$$

8 Simulation results

We used the Keras library of the platform Python and Tensorflow to develop our architectures of the CNN. We used a 5-fold cross-validation approach to assess the performance of our main 3-class model. The training set was randomly divided into 5 equal sets. Four of five sets were used to train the CNN model while the remaining set was used for validation. This strategy was repeated five times by shifting the validation and training sets. The final performance of the model was reported by averaging values obtained from each fold.

We also used the Adam algorithm to optimize our CNN architectures, and we have defined the loss function through the notion of cross-entropy. The learning rate is started from the value 0.001 and is reduced after 4 epochs if the loss value didn't improve with the help of callbacks function. The models were configured to train for 60 iterations. We split our dataset into training set and test set using stratify parameter to preserve the proportion of target as in original dataset for better prediction and reproducibility of results. Performance of each class prediction results obtained by our model during the validation of the learning phase is presented in Table 2. We found that the stacking model has allowed improving considerably the result obtained by each model taken individually.

As we can see in the Table 3, VGG16 and VGG19 achieved the best test accuracy 96.88 and 95.31, respectively. The result for each neural network model is shown in Table 3. After analyzing the obtaining of the predictions of all models, we combined the results of 7 models: VGG16, VGG19, IceptionV3, Xception, DenseNet121, InceptionResNetV2, and MobileNet. Hence, we combine the predicted class of each model in a vector and we take the class which was most frequently predicted by all models. By using this stacking model, we were able to make our final classifier achieved the best performance with test accuracy of 98%, f1-score 98.33%, precision of 98.66% and sensitivity 98.33% as is presented in Table 3.

Table 2 The prediction results for each class using the model set which combines the seven models (0: COVID-19, 1: normal, 2: pneumonia)

Model	Precision (%)			Sensitivity (%)			F1-score (%)		
	0	1	2	0	1	2	0	1	2
VGG16	99.99	82.69	97.14	93.1	99.99	80.95	96.43	90.53	88.31
VGG19	99.99	82.69	94.59	86.21	99.99	83.33	92.59	90.53	88.61
Incepption ResNetV2	99.98	93.18	95.12	99.97	95.35	92.86	99.99	94.25	93.98
Xception	93.55	92.11	84.44	99.98	81.40	90.48	96.67	86.42	87.36
InceptionV3	99.98	95.45	97.62	96.55	97.67	97.62	98.25	96.55	97.62
MobileNet	99.99	93.18	95.24	96.55	95.35	95.24	98.25	94.25	95.24
DenseNet121	99.99	89.58	99.99	99.98	99.99	88.10	99.99	94.51	93.67
stacking method	99.99	96.40	99.98	99.99	99.99	95.50	99.99	97.73	97.56

For simulation results and performance analysis, we present in Fig. 11 various plots of sensitivity against specificity for each model and the stacking model on the testing set. We notice that all models achieve a good performance.

Figure 12 illustrates and displays the plots of accuracy and cross-entropy versus epoch for all trained models. The plot (a) presents accuracy versus epoch and plot (b) presents cross-entropy versus epoch. Also Table 4 presents performance comparison of existing deep learning methods and our proposed Stacking model on 3-class classification task.

9 Localization of the pulmonary lesions

To locate lesions in Chest X-ray images caused by the disease we highlight the points of interest that have strongly contributed to the classification of our model. For this we first perform a linear combination of the gradient maps of all the layers of the model. The region of interest of dimension $n \times m$ for a class (i) is obtained by first calculating the gradient of the probability of presence P^i of each class for a given layer of the model C^k. The global average of these gradients will fix the weighting of the neurons relative to class (i).

$$\beta_k^i = \frac{1}{Z} \sum_r^m \sum_s^n \frac{\partial P^i}{\partial B_{rs}^k} \tag{1}$$

Table 3 Comparative results for each model on the test set

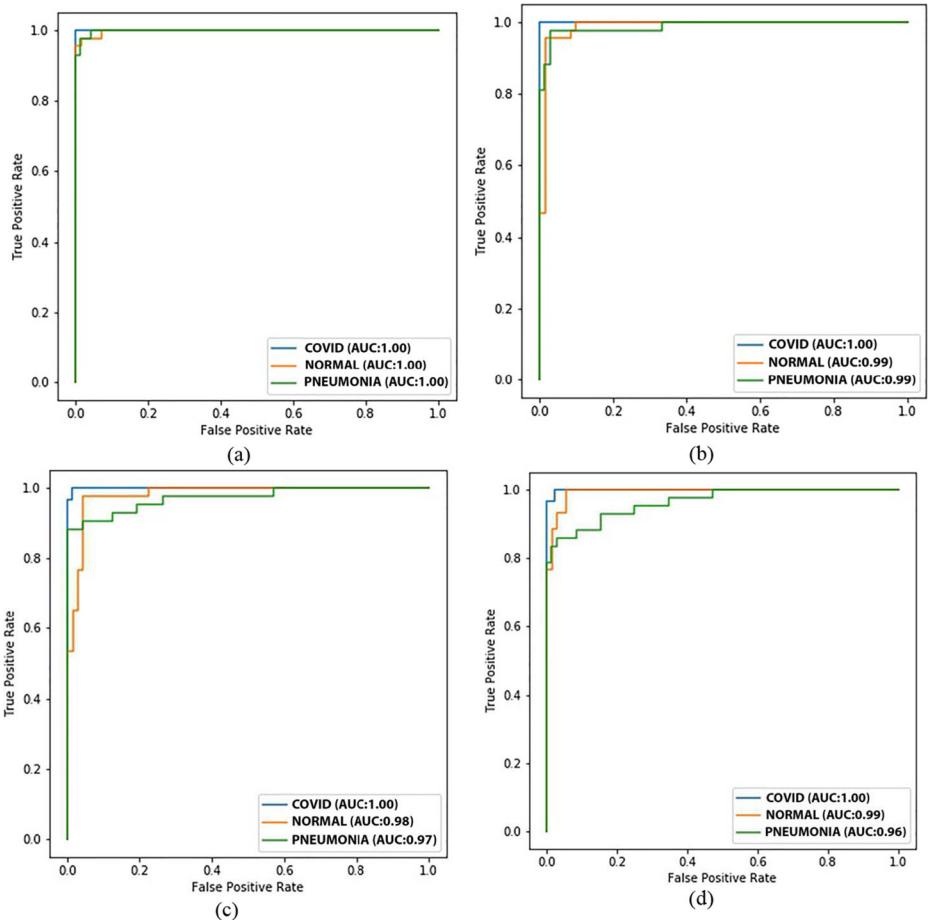| Model	Precision (%)	Sensitivity (%)	F1-score (%)	Test Accuracy (%)
VGG16	91.66	90	90	96.88
VGG19	96.60	97	96	95.31
Incepption ResNetV2	96	95.66	95.66	89.06
Xception	95.66	95	95.33	95.31
InceptionV3	92.33	93	92.66	92.66
MobileNet	92	89.66	90.33	89.06
DenseNet121	97	97.21	97.66	92
Stacking method	**98.66**	**98.33**	**98.30**	**98**

Fig. 11 Receiver Operating Characteristic (ROC) curves of a our proposed model: **a** denseNet121 CNN, **b** InceptionResNetV2 CNN, **c** Inception V3 CNN, **d** MobileNet, **e** VGG16 CNN, **f** VGG19, **g** Xception, **h** Stacking model

Then we apply the ReLU function:

$$L^i = ReLU\left(\sum_k B_k^i B^k\right) \tag{2}$$

The regions of interest obtained by our algorithm correspond to the three types of classes: Normal, Pneumonia and COVID-19 are shown in Fig. 5. We represented with yellow the pixels which presented a strong gradient and which contributed enormously to the classification of our model. The blue color was assigned to the pixels which presented a weak gradient and which did not contribute to the classification of our model. Note that in the case of the images of patients declared positive for COVID-19, our algorithm focused on the region showing the opacity of the frosted glass, which indicates a clinical pathological character of COVID-19. For Pneumonia class images our algorithm showed a pulmonary inflammation
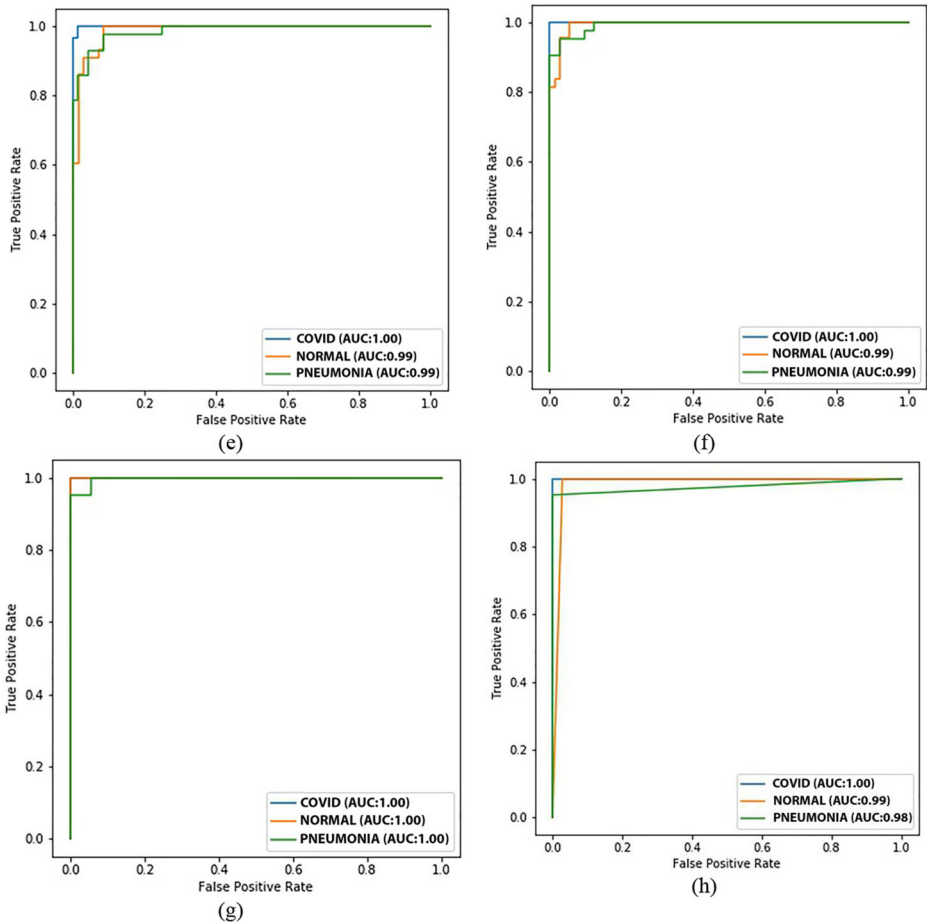
Fig. 11 (continued)

characterizing a typical pneumonia. For the images of the normal class, no region showed a strong variation of the gradient and consequently no region was detected by our algorithm.

In addition, we gave six test images to our partner team from the radiology department of the Fez University Hospital, made up of renowned radiologists. The latter surrounded with a marker in red lesions of the pathology present on the image, as is shown in the images above of Fig. 13. Then we compared their location with that indicated by our algorithm, as is shown by the images below of the Fig. 13. The comparison showed that there was a perfect agreement with the location of the radiologists and that of our algorithm.

10 Comparison between radiologist and machine prediction

To be able to compare the performance of our proposed model with that of radiologists, we chose a sample composed of 10 images of COVID-19 type, 10 images of Normal type and 10
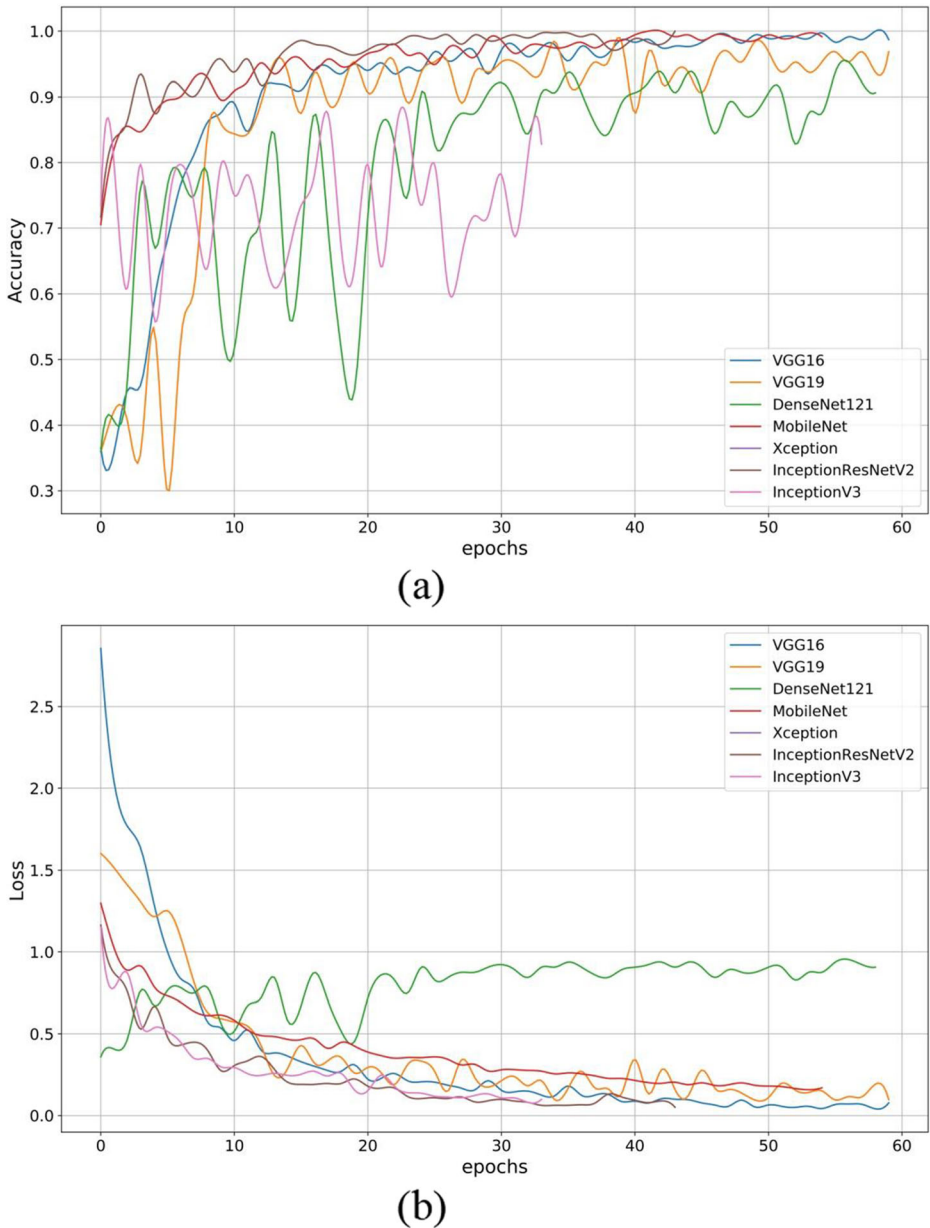
Fig. 12 Accuracy and cross-entropy versus epoch for all trained models (**a**) Accuracy versus epoch, (**b**) cross-entropy versus epoch

images of Pneumonia type. Then we asked a renowned radiologist to visually reclassify his 30 images. As shown in Table 5, the radiologist 1 could not exceed an accuracy of 63.64% with a sensitivity of 70% and a F1-score of 66.67%. Similarly, radiologist 2 could not exceed an accuracy of 37.5% with a sensitivity of 30% and a F1-score of 34%.

Table 4 Comparison of the proposed Stacking with other existing deep learning methods

Study	Architecture	Accuracy 3-class (%)	#Params (in million)
Ioannis et al. [5]	VGG19	93.48	143
Ioannis et al. [5]	Xception	92.85	33
Wang and Wong [42]	Covid-Net (Residual Arch)	NA	116
Sethy and Behra [38]	ResNet-50	NA	36
Hemdan et al. [23]	VGG19	NA	143
Narin et al. [33]	ResNet-50	NA	36
Narin et al. [33]	InceptionV3	NA	26
Ozturk et al. [35]	DarkNet	87.02	1.1
Khan et al. [27]	CoroNet (Xception)	89.6	33
Proposed Stacking method	Stacking method	98	245

11 Discussion

Given the unavailability of PCR tests as well as their exorbitant price, it was useful to make available to healthcare personnel a method based on artificial intelligence to predict COVID-19 quickly and precisely. In this article, we have proposed an intelligent clinical decision support system for the early detection of COVID 19 from chest x-rays, which are cheaper, easily accessible to rural populations, and whose acquisition device is easily disinfected, easily cleaned and easily maintained. This model uses deep learning through seven Convolutional Neurons Networks extracted from the ImageNet database already pre-trained. Each network will predict a class as the input image. The advantage of our model is that it will assign a score to each class prediction and only send out the prediction which obtains the highest score. In this way, we were able to obtain an overall precision of 98.6% at the time when the other models taken individually give a precision lower than this. In addition, our model achieved better performance with a test accuracy of 98%, an f1 score of 98.33%, an accuracy of 98.66% and a sensitivity of 98.33% at the time when the prediction of renowned radiologists could not
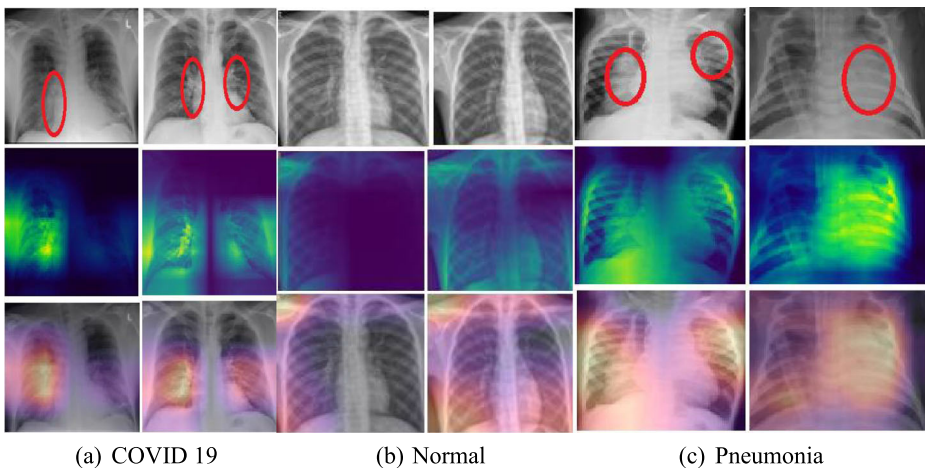

(a) COVID 19   (b) Normal   (c) Pneumonia

Fig. 13 Visual results obtained by Grad-CAM on different disease classes: **a** COVID19 case Top to down: Initial image and its Grad-CAM transform, **b** Normal case: Top to down: Initial image and its Grad-CAM transform, **c** Pneumonia case: Top to down: Initial image and its Grad-CAM transform

Table 5 The proposed model performances compared to prognostic radiologists

Model	Our Model	Radiologist 1	Radiologist 2
Precision (%)	98.66	63.64	37.5
Sensitivity (%)	98.33	70	30
F1-score (%)	98.30	66.67	34

exceed an accuracy of 63.34% with a sensitivity of 70% and f1 score of 66.67%. With these results, it can be confirmed that the proposed model can be used to detect COVID-19 precisely and quickly, especially since it succeeds in discriminating between pneumonia typical of that of COVID-19. The comparison of the prediction of our model with the prediction of the radiologist showed a much higher precision and sensitivity.

The results obtained by our proposed model are superior compared to other studies in the literature. Table 4 presents a summary of studies conducted in the automated diagnosis of COVID-19 from chest X-ray images and their comparison with our proposed model stacking.

These results demonstrate that an accurate and reduced diagnosis can be made with in-depth learning, especially in isolated rural areas where CT and Polymerase Chain Reaction (PCR) tests are rare. Among the limitations of this model is the non-availability of COVID-19 image data, which prevents complete learning and thereby achieving higher precision. Indeed, learning deep neural networks with limited data available leads to over-adjustment and prevents good generalization. To remedy this issue, researchers should look into the notion of increasing image data to further improve accuracy and avoid over-fitting.

12 Conclusion

The objective of this research was to propose an intelligent clinical decision support system for the early detection of COVID 19 from chest x-rays. Given the lack of PCR and CT tests in rural and isolated regions of the underdeveloped countries, we have developed in this paper a system for the early detection of COVID-19 from chest X-ray images which are accessible and inexpensive in these regions and whose acquisition equipment is easily disinfectable, easily accessible and easily maintained. For this, we have used Deep Transfer Learning models to classify three classes: COVID-19, Pneumonia and normal using Transfer Learning concept. We have used the pre-trained architectures such as DenseNet121, VGG16, VGG19, InceptionResNetV2, Xception, MobileNet, and InceptionV3. We were able to determine characteristic features from chest X-ray images, and we take the advantages of the seven models to build a stacking model that outperformed all other models. The proposed model has achieved the best performance with a test accuracy of 98%, an f1 score of 98.33%, an accuracy of 98.66% and a sensitivity of 98.33% when the prediction of renowned radiologists could not exceed an accuracy of 63.34% with a sensitivity of 70% and an f1 score of 66.67%. We observed that the performance could be improved further by increasing dataset size with the using of data augmentation in the future. In addition to that, we used the concept of Gradient Class Activation Map (Grad-CAM) to visualize the areas which are the most responsible for the final prediction as a heat map. Our study supports the notion of Deep Transfer learning (DTL) methods can be used to simplify the diagnostic process and improve disease management. It will act as assistive intelligence to medical practitioners and potentially help to reduce the burden on the healthcare system. The initial results show promise, but with the increase in

the number of X-ray images of COVID-19 and more diverse data, we will be able to make our models more accurate and generalized.

Data availability The data used to support the findings of this study are available from the corresponding author upon request.

Declarations

Conflict of interest I, the undersigned Pr. Mostafa EL MALLAHI, author of this article, declares that there is no conflict of interest regarding the publication of this paper.

References

1. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W et al (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases [published online ahead of print, 2020 Feb 26]. 450 radiology
2. Amakdouf H, El Mallahi M, Zouhri A, Qjidaa H (2018) Classification and recognition of 3D image of charlier moments using a multilayer perceptron architecture. Procedia Comput Sci 127:226–235
3. Amakdouf H, Zouhri A, El Mallahi M et al (2020) Color image analysis of quaternion discrete radial Krawtchouk moments. Multimed Tools Appl 79:26571–26586. https://doi.org/10.1007/s11042-020-09120-0
4. Amakdouf H, Zouhri A, El Mallahi M et al (2021) Artificial intelligent classification of biomedical color image using quaternion discrete radial Tchebichef moments. Multimed Tools Appl 80:3173–3192. https://doi.org/10.1007/s11042-020-09781-x
5. Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. In: COVID -19: automatic detection from X-ray images utilizing transfer learning with convolutional neural network. Physical and Engineering Sciences in Medicine, March. https://doi.org/10.1007/s13246-020-00865-4
6. Y. Bhatia, A. Bajpayee, D. Raghuvanshi and H. Mittal, (2019). Image captioning using Google's inception-resnet-v2 and recurrent neural network, 2019 twelfth international conference on contemporary computing (IC3), Noida, India, 2019, pp. 1–6, https://doi.org/10.1109/IC3.2019.8844921.
7. Choe J, Lee SM, Do KH, Lee G, Lee JG, Lee SM, Seo JB (2019) (2019). Deep learning-based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. Radiology. 292(2):365–373
8. F. Chollet,. (2017). Xception deep learning with depthwise separable convolutions. 2017 IEEE conference on computer vision and pattern recognition (CVPR).
9. Chu DKW, Pan Y, Cheng SMS, Hui KPY, Krishnan P, Liu Y, Ng DYM, Wan CKC, Yang P, Wang Q, Peiris M, Poon LLM (2020) Molecular diagnosis of a novel coronavirus (2019-nCoV) causing an outbreak of pneumonia. (2020). Clin Chem 66(4):549–555. https://doi.org/10.1093/clinchem/hvaa029
10. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, Cui J, Xu W, Yang Y, Fayad ZA, Jacobi A, Li K, Li S, Shan H (2020) CT imaging features of 2019 novel coronavirus (2019-nCoV). Radiology. 2020: 200230
11. Cohen J.P. (2020). COVID-19 Image Data Collection. https://github.com/ieee8023/COVID-chestxray-dataset.
12. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK et al (2020, 2020) Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill 25(3)
13. Coronavirus disease 2019 (COVID-19): situation report – 78. World Health Organization. 2020 Apr 07. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200407-sitrep-78-covid-19.pdf.
14. Dai, W.; Chen, Y.; Xue, G.-R.; Yang, Q.; Yu, Y.; Koller, D.; Schuurmans, D.; Bengio, Y.; Bottou, L. (2008). Translated learning: transfer learning across different feature spaces. In advances in neural information processing systems 21, proceedings of the neural information processing systems 2008, Vancouver,

BC, Canada, 8–10 December 2008; neural information processing Systems Foundation, Inc. (NIPS): Vancouver, BC, Canada, 2008; pp. 353–360.

15. El Mallahi M, Zouhri A, El-Mekkaoui J et al (2017) Three dimensional radial Tchebichef moment invariants for volumetric image recognition. Pattern Recognit Image Anal 27:810–824. https://doi.org/10.1134/S1054661817040113

16. El Mallahi M et al (2017) Radial Meixner moments for rotational invariant pattern recognition. In: 2017 intelligent systems and computer vision. ISCV 2017.

17. El Mallahi M et al (2017) Radial Charlier moment invariants for 2D object/image recognition. In: International conference on multimedia computing and systems -proceedings

18. El Mallahi M, Zouhri A, El Affar A et al (2018) Radial Hahn moment invariants for 2D and 3D image recognition. Int J Autom Comput 15:277–289. https://doi.org/10.1007/s11633-017-1071-1

19. El Mallahi M, Zouhri A, Mesbah A et al (2018) 3D radial invariant of dual Hahn moments. Neural Comput & Applic 30:2283–2294. https://doi.org/10.1007/s00521-016-2782-x

20. M. El Mallahi, A. Zouhri, A. Mesbah, I. El Affar and H. Qjidaa. (2018) "Radial invariant of 2D and 3D Racah moments " Springer, Multimedia Tools and Applications An International Journal, Vol 77, Issue 6, March 2018, Pages 6583–6604 .

21. Gavai, N. R. et al. (2018). MobileNets for flower classification using TensorFlow', 2017 international conference on big data, IoT and data science, BID 2017, 2018-Janua, pp. 154–158. https://doi.org/10.1109/BID.2017.8336590

22. Gomez P, Semmler M, Schutzenberger A, Bohr C, Dollinger M (2019) Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. Med Biol Eng Comput 57(7):1451–1463

23. Hemdan, E.E.D.; Shouman, M.A.; Karar, M.E (2020) Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv preprint arXiv:2003.11055

24. G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger. (2017). Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

25. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223):497–506. https://doi.org/10.1016/s0140-6736(20)30183-5

26. Asif Iqbal Khan, Junaid Latief Shah, Mohammad Mudasir Bhat (2020) CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images, Computer Methods and Programs in Biomedicine, Computer Methods and Programs in Biomedicine 196: 105581, https://doi.org/10.1016/j.cmpb.2020.105581

27. Khan M, Sohrabi C, Alsafi Z, O'Neill N, Kerwan A, Al-Jabir A, Iosifidis C, Agha R (2020) World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg 76:71–76

28. Kumarm R, Zhang N, Wang L, Deng X, Liang R, Su M, He C et al (2020) Recent advances in the detection of respiratory virus infection in humans. J Med Virol 92(4):408–417. https://doi.org/10.1002/jmv.25674

29. A. Machhour, M. El Mallahi, Z. Lakhliai, A. Tahiri, D. Chenouni, 2019 Image Classification Using Legendre–Fourier Moments and Artificial Neural Network, Proceedings of ESAI 2019, Fez, Morocco, pp 305–315

30. Machhour A, Zouhri A, El Mallahi M, Lakhliai Z, Tahiri A, Chenouni D (2020) Plants Classification Using Neural Shifted Legendre-Fourier Moments. Advances in Smart Technologies Applications and Case Studies. SmartICT 2019. Lecture notes in electrical engineering, vol 684. Springer, Cham. http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-030-53187-4_18

31. Machhour A, El Mallahi M, Lakhliai Z, Tahiri A, Chenouni D (2020) Image classification using Legendre–Fourier moments and artificial neural network. In: Bhateja V, Satapathy S, Satori H (eds) Embedded systems and artificial intelligence. Advances in intelligent systems and computing, vol 1076. Springer, Singapore. https://doi.org/10.1007/978-981-15-0947-6_29

32. Mahmood S, Hasan K, Colder Carras M, Labrique A (2020) Global preparedness against COVID-19: we must leverage the power of digital health. JMIR Public Health Surveill 6(2):e18980

33. 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

34. Note: This data is mandatory. Please provide..

35. 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

36. Ravishankar, H.; Sudhakar, P.; Venkataramani, R.; Thiruvenkadam, S.; Annangi, P.; Babu, N.; Vaidya, V. (2016). Understanding the mechanisms of deep transfer learning for medical images. In deep learning and data labeling for medical applications;' DLMIA 2016, LABELS 2016; Carneiro, G., Ed.; Springer: Cham, Switzerland, 2016; volume 10008.Springer: Cham, Switzerland, 2016; volume 10008.

37. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. (2018). Grad-CAM: visual explanations from deep networks via gradient-based localization. In Proc. of the IEEE international conference on computer vision, pp. 618–626, 2017.
38. Sethy PK, Behera SK (2020) Detection of coronavirus disease (covid-19) based on deep features. Preprints 2020:2020030300
39. Karen Simonyan, Andrew Zisserman. (2014). Very deep convolutional networks for large-scale image recognition. Computer Vision and Pattern Recognition, arXiv:1409.1556
40. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. In AAAI, 2017, pp 4278–4284
41. Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. (2017). Chestxray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 2097–2106.
42. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al (2020) A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). :2020.02.14.20023028. https://doi.org/10.1101/2020.02.14.20023028
43. Yang Y, Lu Q, Liu M, Wang Y, Zhang A, Jalali N, et al (2020) Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. medRxiv [Preprint]. medRxiv [posted February 21], Available from: https://www.medrxiv.org/content/10.1101/2020.02.10.20021675v2. https://doi.org/10.1101/2020.02.10.20021675
44. Zouhri A, Amakdouf H, El Mallahi M et al (2020) Invariant Gaussian–Hermite moments based neural networks for 3D object classification. Pattern Recognit Image Anal 30:87–96. https://doi.org/10.1134/S1054661820010186

Affiliations

Mamoun Qjidaa [1] · **Anass Ben-Fares** [2] · **Hicham Amakdouf** [2] · **Mostafa El Mallahi** [3] 📧 · **Badre-eddine Alami** [1] · **Mustapha Maaroufi** [1] · **Ahmed Lakhssassi** [4] · **Hassan Qjidaa** [2]

✉ Mostafa El Mallahi
mostafa.elmallahi@hotmail.com

1　Faculty of Medicine and Pharmacy, Department of Radiology, Sidi Mohamed Ben Abdellah University, Fez, Morocco

2　Faculty of Sciences Dhar El Mahraz, Laboratory of Computer Science Signals Automation and Cognitivism, Sidi Mohamed Ben Abdellah University, Fez, Morocco

3　High Normal School, Mathematics and Computer Sciences Department, Laboratory of Computer Science and Interdisciplinary Physics, Sidi Mohamed Ben Abdellah University, Fez, Morocco

4　Laboratory of Engineering and Advanced Micro Systems, University of Quebec in Outaouais, Gatineau, Canada