Gezira J. of Engineering and Applied Sciences
ISSN Linking: 1858-5698
http://journals.uofg.edu.sd/index.php/gjeas/

Classification of COVID-19 from Chest X-ray Images Using Transfer Learning

Lamyaa Alser Mohammed [1], Murtada Khalafallah Elbashir [2], Gais Alhadi Babikir [3*]

1. Department of Computer Sciences, University of Gezira, Wad Madani, Sudan.
   lamyaaelser35@gmail.com
2. Department of Computer Sciences, University of Gezira, Wad Madani, Sudan.
   murtadabashir@yahoo.com
3. Department of Computer Sciences, University of Gezira, Wad Madani, Sudan.
   Gais.alhadi@uofg.edu.sd

INFORMATION:

Submission: **22/07/2022**

Accepted: 17/04/2022

Publication: 03/06/2026

ABSTRACT

Covid-19 descended from a virus strain called the corona, or coronavirus, in December 2019. It originated in Wuhan city of China, spread across the world, and became a pandemic. It should be noted that to prevent the further spread of this epidemic and to treat those infected with it quickly, researchers have been trying to develop effective methods to classify COVID2-19. One of these methods is the analysis of chest X-ray (CXR) images by radiologists, but this diagnosis has many drawbacks: it can take a long time, those radiologists are not available all the times, especially in remote areas, and the new features of COVID-19 are unknown to some radiologists. Therefore, it is necessary to implement an automatic classification system that helps doctors to classify COVID-19 in the early stages and provides a quick alternative diagnosis option to prevent the pandemic from spreading among people. This paper aimed to analyze and classify COVID-19-infected patients as infected (+ve) or not (−ve) from CXR images. It should be noted that the images used in this paper were collected from two sources as follows: 440 positive COVID-19 CXR images were collected from the GitHub repository, and 440 normal CXR images were obtained from the Kaggle repository. Then, data preprocessing techniques were applied, specifically resizing and normalization, to suit the classification process. Subsequently, three different transfer learning models (Xception, InceptionV3, and MobileNetV2) were proposed for the classification of coronavirus pneumonia-infected patients using chest CXR radiographs. Moreover, the experimental results obtained from the pre-trained Xception model have provided the highest classification performance with 99.43% accuracy, 99% precision, 99% recall, and 99% F1-score. It should be noted that, to the best of our knowledge, the best state-of-the-art model achieves an accuracy of 97%, precision of 99%, recall of 94 %, and F1-score of 91%. So, the results of the proposed model (Xception) are the best achieved so far. Therefore, the pre-trained Xception model could contribute great importance to the automatic classification of COVID-19 from CXR images.

KEYWORDS

COVID-19; chest X-ray radiographs; convolutional neural network; deep transfer learning

1. INTRODUCTION

Covid-19 descended from a virus strain called the corona, or coronavirus. On December 31, 2019, reports were made on unknown causes of pneumonia in Wuhan, Hubei Province, China, and the disease became a rapid epidemic. On January 30, 2020, the World Health Organization (WHO) declared this pandemic to be a Public Health Emergency of International Concern. By January 5, 2021, more than 86 million people worldwide had been infected with COVID-19, of whom more than 1,870,000 people had died worldwide [1]. Medical staff from both developed and undeveloped countries, all united to fight against this severe virus in spite of inappropriate facilities, insufficient healthcare systems, and improper diagnostic approaches [2]. The majority of tests being used to diagnose COVID-19 have been genetic tests known as Reverse Transcription Polymerase Chain Reaction (RT-PCR) [3]. Such tests have limited aspects with certain features that make them difficult to diagnose the disease. RT-PCR requires an experienced laboratory technician to take samples from the mucous membrane of the nose and throat, which is a painful method; and that is why many people refuse to undergo nasal biopsies. The drawback of this method is that is a very time-consuming, complex, costly, and manual process. Additionally, it requires a type of laboratory kit, the provision of which is difficult or even impossible for many countries during crises and epidemics. More importantly, many studies indicated the low sensitivity of the RT-PCR test; several studies have reported that the sensitivity of this diagnostic method is about 30% to 60%, which indicates a decrease in the accuracy of identifying COVID-19 in many cases. Some studies have also pointed out to its false-negative rate and contradictory results [1]. This false-negative rate can lead to the existence of infected people among healthy ones and that can accelerate the infection rate. Therefore, RT-PCR and quarantine are not enough to control this pandemic, so the researchers have been trying to develop alternative solutions. They have focused on analyzing radiographs to diagnose COVID-19. In this case, the analysis of chest x-ray images can play an important role. Chest x-ray (CXR) and computer tomography (CT) can be considered as an alternative to PCR if they are properly used for diagnosis by an expert physician. Notably, researchers have discovered that COVID-19-infected chest x-ray images have some distinct features such as vague darkened points or ground-glass opacity, which can help in detecting and classifying COVID-19 ([14],[15], [16]). CXR scanning is a kind of advanced CXR machine that examines the very soft structure of the active body part, showing clearer images of the inner soft tissues and organs. Using CXR is an easier, faster, available, and less harmful method than computer tomography (CT). Radiologists can analyze these images and help doctors to detect COVID-19 from the early stages. However, in a pandemic, a more updated, more available, and quicker system is needed. Manual analysis of images can take a long time, radiologists may not be available all the time, especially in remote areas, and new features of covid-19 are unknown to some radiologists [4].It should be noted that more sophisticated classification systems for COVID-19 are needed for diagnosis while reducing excessive pressure on radiologists' services. Deep learning has already shown its power in classifying images with high accuracy. In addition, the field of medical image processing is vividly exploring deep learning. However, one of the major problems in the medical field is the availability of large data sets with reliable ground truth annotation. Therefore, transfer learning approaches are often considered to overcome the problem of small datasets [5]. Transfer learning is the method of taking the weights of a pre-trained model and using the previously learned features to make a decision on a new class label. A model is used in transfer learning that is pre-trained in the ImageNet dataset, and this model learns to identify high-level features of images in the initial layers. For transfer learning, a few dense layers are added at the end of the pre-trained model. Then, the model learns what combinations of features will help identify the features in new data collection without transfer learning, and it requires more computational power to train the huge models on large datasets and too much time, which can be up to weeks, is needed to train the model [6].

In this paper, we have developed an automated method to classify the COVID-19 disease from CXR images. The study took into consideration the deep learning algorithms that can recognize the features of the images and process them. To classify covid-19 from these images three transfer learning models were used: Xception, InceptionV3, and MobileNetV2.

The rest of this paper is organized as follows: the next part, Section 2 presents related works. In Section 3, we provide the materials and methodology. Section 4, presents the main results and discussion. Finally, Section 5 concludes this paper.

2. RELATED WORKS

Given the aim of this work, several studies have been presented to classify COVID-19, and improve performance via deep learning techniques to achieve high accuracy in COVID-19 classification data. For example, Goldstein et al. [6] used a pre-trained model (ReNet50) to detect COVID-19 from CXR images. The authors used a dataset consisting of 360 COVID-19 and 1024 normal CXR images. They also showed that the proposed model obtained an accuracy of 89.7% and a sensitivity of 87.1%. Asif et al. [7] proposed an automatic detection of COVID-19 using an Inception V3 model with transfer learning from CXR images. They used a dataset consisting of 864 COVID-19, 1345 viral pneumonia, and 1341 normal chest x-ray images, and obtained a classification accuracy of 98%. A new CVOIDX-Net framework was proposed in [8], to automatically identify COVID-19 based on several pre-trained models. The authors showed that the proposed results of COVIDX-Net verified that the best performance scores of deep learning classifiers were for the VGG19 and DenseNet201models. They also obtained a 90% accuracy rate using 25 COVID-19 positive and 25 normal images. Narin, et al. [9] proposed an automatic detection of COVID-19 using a ResNet50, InceptionV3 and Inception-ResNetV2 pre-trained models, and a relatively small dataset (50 COVID-19 vs. 50 Normal) of CXR images. The authors showed that among the pre-trained models, the ResNet50 model provided the highest classification performance with an accuracy of 98%, 97% accuracy for the InceptionV3 model and 87% accuracy for the Inception-ResNetV2 model. Mohammad et al. [10] presented a concatenated neural network based on Xception and ResNet50V2 networks for classifying the CXR images into three categories of normal pneumonia, and COVID-19. They used two open-source datasets containing 180 and 6054 images of patients with COVID-19 and pneumonia, respectively, and 8851 images of normal people. The authors showed that the Xception achieved an accuracy of 91.31%, ResNet50V2 achieved an accuracy of 89.79% and an overall accuracy of 91.4%.

3. MATERIALS AND METHODS

In this work, many steps were followed to achieve the objectives, including data collection, data pre-processing, splitting data into training and validation, applying pre-trained models, and doing performance analysis for all models. These steps are discussed as follows:

A. Dataset

The data used in this work were collected from two sources. In the first dataset, 440 positive **COVID**-19 CXR images were collected from a GitHub repository developed by Cohen et al. [10]. This repository contained CXR images of positive COVID-19, negative COVID-19, and pneumonia cases. Clearly, images of this repository were collected from open sources and hospitals for patients with an average age of about 55. In fact, this repository contained CXR images of positive COVID-19, negative **COVID**-19, and pneumonia cases. Images of this repository were gathered from open sources and hospitals, where the average age of the infected patients was around 55 years. It should be noted that Cohen's database contained 500 CXR images of positive COVID-19-infected patients, whereas the negative COVID-19 images included other viral and bacterial pneumonia (MERS, SARS, and ARDS). In the second dataset, we considered the COVID-19 disease only, 440 normal CXR images were obtained from the Kaggle repository, which contained 5930 normal images and other pneumonia [11]. Hence, the dataset in this paper contained a total of 880 images after balancing the data, 440 covid-19 positive images and 440 covid-19 negative images, and all CXR images were collected in one dataset. In Figure 1, we present a sample of CXR images dataset of normal cases and COVID-19 patients.

B. Data Pre-Processing

To make the data suitable for the classification process, data pre-processing techniques were applied. Since all the sample images were of different sizes, the images were resized to 224 × 224 pixels. Then, normalization was applied to the dataset to make it suitable for further processing within the deep learning pipeline. Thus, the dataset was ready to be fed into the CNN and for model training.

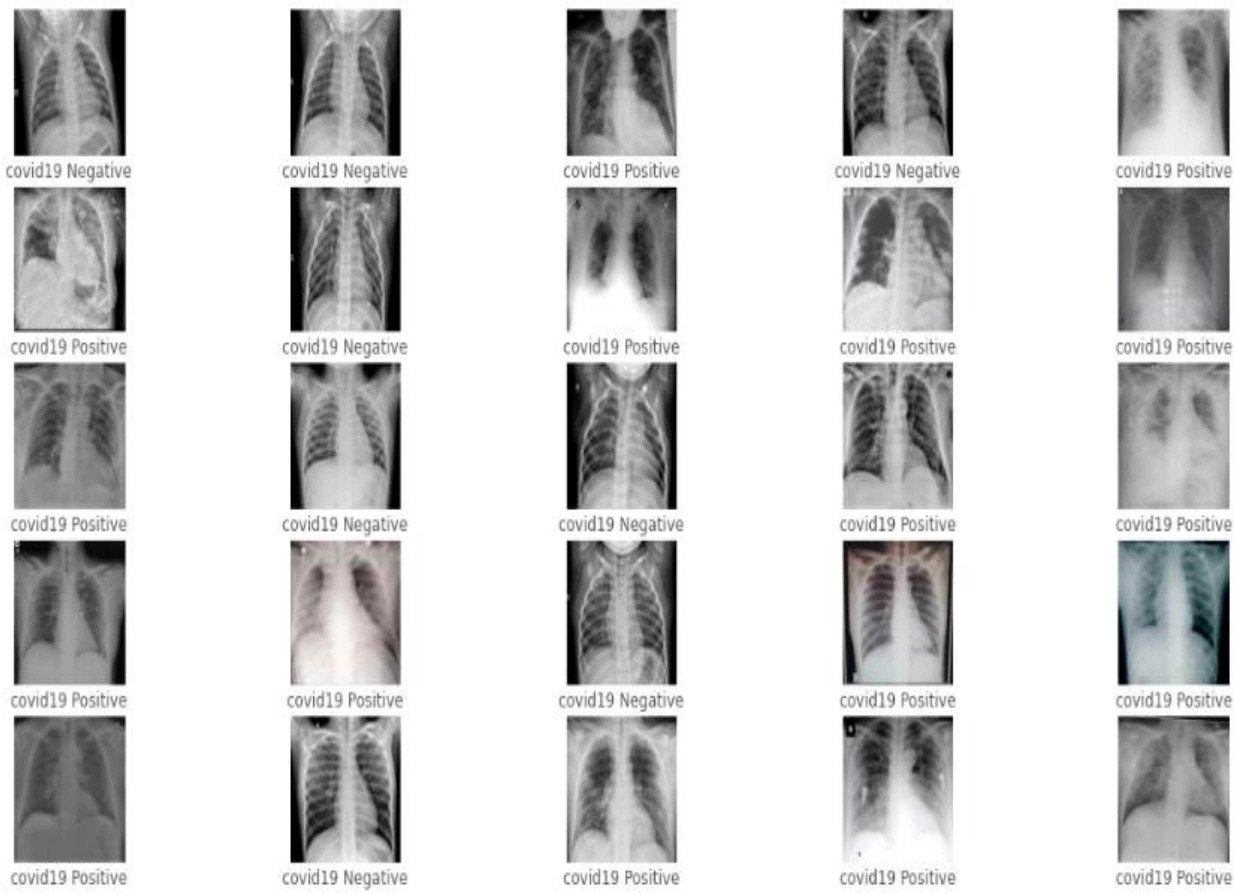


Figure 1. A sample of CXR images dataset for normal cases and COVID-19 patients

C. Experimental setup

To make two partitions of dataset one for training the model called training set and the other for testing the performance of the trained model called the test set, K-fold cross-validation approach was used. In this paper 5-fold cross-validation was used; one was used for testing, and the rest for training. It should be noted that cross-validation is a statistical process for evaluating and comparing learning algorithms by dividing the data into two parts: the first part is used to learn or train the model, and the second part is used to validate the model. The experiment was run on Google Collaboratory. In fact, Google Collaboratory is a Jupiter notebook-based cloud service for disseminating knowledge and working on machine learning, which offers fully optimized runtime for deep learning and free-of-charge access to a stable GPU.

D. Pre-trained models

After splitting the dataset into training and testing sets, the transfer learning technique achieved using the ImageNet dataset was applied to overcome insufficient data and training time. Then, to adapt the three transfer learning models to detect COVID-19, we replaced the last dense layer of these models with a new flatten layer and two dense layers (FCL), where the flatten layer converted a 2D feature map into a 1D feature vector, which was fed to an FCL. Clearly, the first dense layer with the rule activation function and the second dense layer matched the number of labels and added a sigmoid activation function for the classification process. After that, the whole models were trained, excluding the frozen layers. Furthermore, we used the Adam optimizer with a learning rate of 1e-5 to train our models. The models were trained for 50 epochs. In Figure 2, we describe the structures of the three transfer learning models.

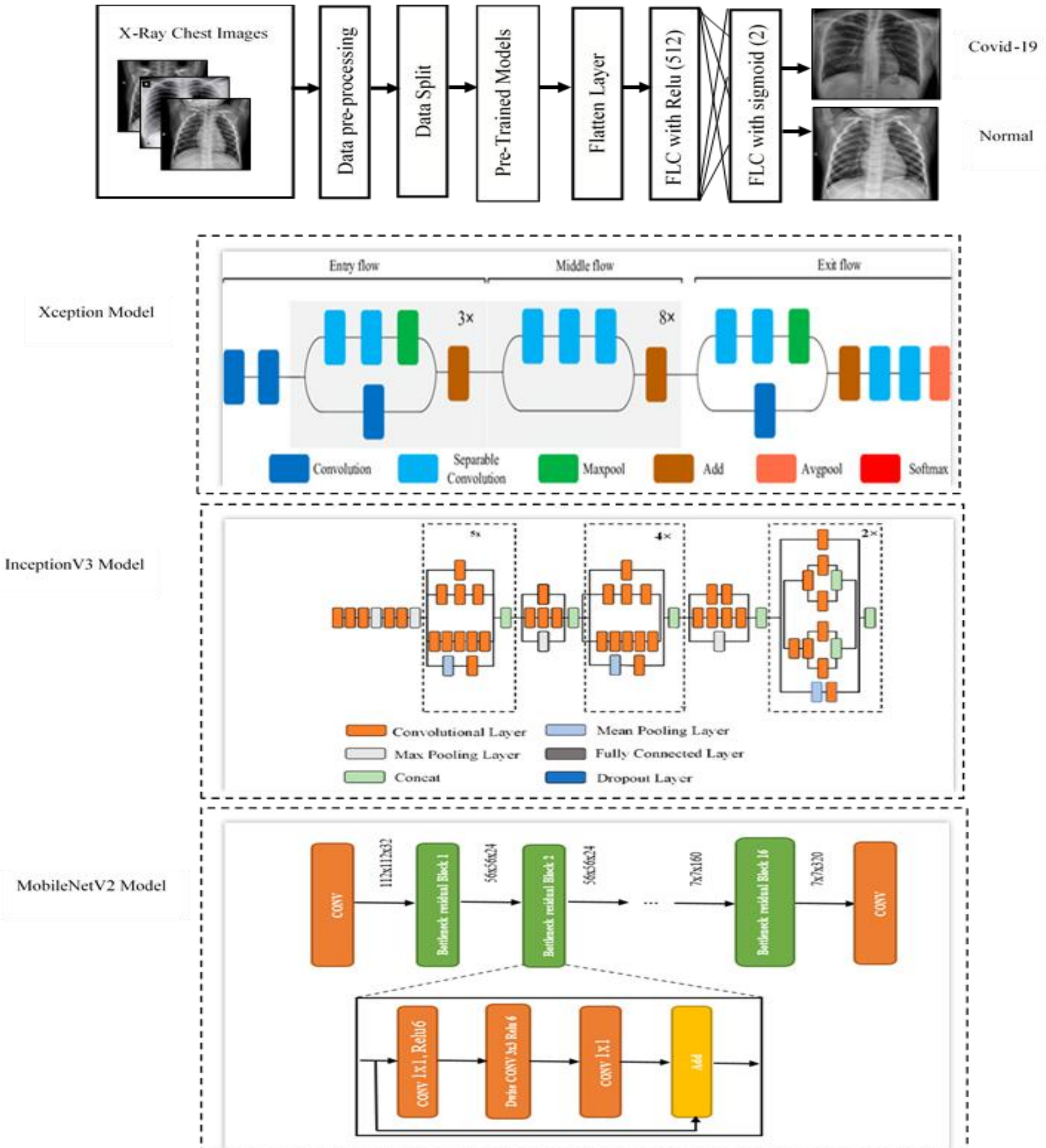


Figure 2. Research methodology

E. Performance analysis

In the final step of the proposed framework, the test data were fed to the tuned deep learning classifier to categorize all the image patches into two classes: confirmed positive COVID-19 or normal case (negative COVID-19), as shown in Figure 1. Finally, an analysis of the overall performance of each deep learning classifier was evaluated based on the metrics described in the rest of this section.

i. Confusion matrix

One of the most comprehensive ways to represent the result of evaluating binary classification used is the confusion matrix, which contains information about the actual and predicted classification system. In Table 1, we present a confusion matrix for a two-class classifier.

Table 1. Confusion matrix.

Actual	Predicted	
	Positive	Negative
Positive	TP	FN
Negative	FP	PN

It is worth noting that TP is the number of correct predictions that an instance is positive, FN is the number of incorrect predictions that an instance is negative, FP is the number of incorrect predictions that an instance is positive and TN is the number of correct predictions that an instance is negative.

ii. Recall/Sensitivity

Recall (R) is the proportion of positive cases that were correctly identified, and it can be calculated as follows:

$$R = \frac{TP}{TP + FN}$$

iii. Precision

Precision (P) is the proportion of the predicted positive cases that were correct, and it can be calculated as follow:

$$P = \frac{TP}{TP + FP}$$

iv. F-measure/F-score

Precision and Recall are two very important measures, but looking at just one of them will not provide the full picture. Thus, another measure will be needed which is the F-score or F-measure, and this measure is the harmonic mean of Precision and Recall, and it can be calculated as follow:

$$F = \frac{2 * Recall * Precision}{Recall + Precision}$$

v. Accuracy

The proportion of the total number of predictions that are correct is known as "accuracy" (AC). It shows the overall effectiveness of the classifier, and it can be calculated as follow:

$$\mathrm{AC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}}$$

vi. Receiver operating characteristics (ROC)

The ROC curve refers to the receiver operating characteristics curve, which is an overall indicator that reflects continuous variables of sensitivity and specificity and is a configuration method to reveal the relationship between sensitivity and specificity. By defining a number of different critical values of continuous variables, a series of sensitivities and properties are calculated, then sensitivity is plotted on the coordinate and (1-specificity) is plotted on the coordinate. On the ROC curve, the point closest to the top left of the graph is the critical value with higher sensitivity and specificity. A ROC curve is plotted with the true positive rate (TPR) along the x-axis, and a false-positive rate (FPR) is plotted along the y-axis. The formulas for obtaining TPR and FPR are shown in the following equations:

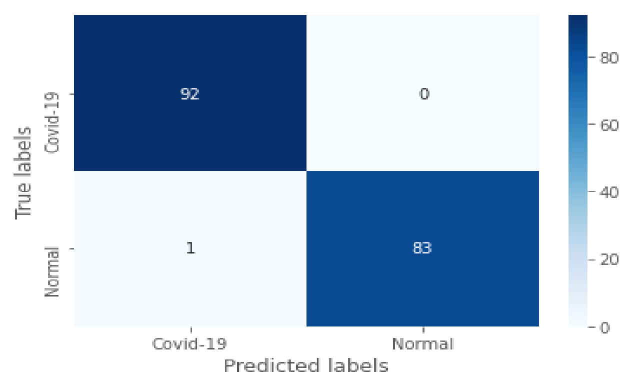
$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}$$

4. RESULTS AND DISCUSSION

Python programming language and Keras packages were used to train the proposed deep transfer learning models. All experiments were performed on Kaggle.

Figures 3, 4, and 5 show the confusion matrixes for the three pre-trained models which are Xception, InceptionV3, and MobileNetV2 for COVID-19 disease classification. It should be noted that among the 176 images, only one was misclassified by the Xception model. Meanwhile, only two images were misclassified by the InceptionV3. Eight images were also misclassified by the MobileNetV2. We found that the Xception model outperformed the others as it had better and consistent true positive and true negative values, and lesser false negative and false positive values. Therefore, the Xception model could efficiently classify COVID-19 cases.



.

Figure 3. The confusion matrix for Xception

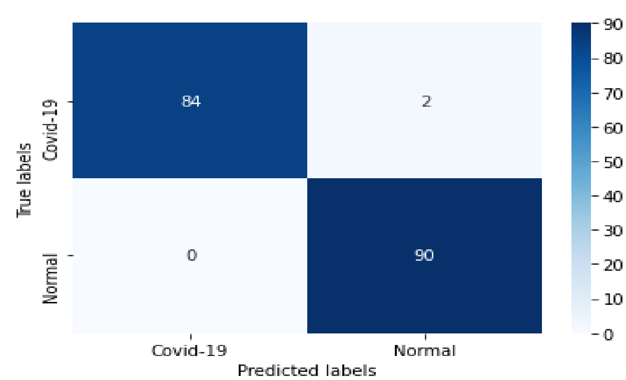


Figure 4. The confusion matrix for InceptionV3

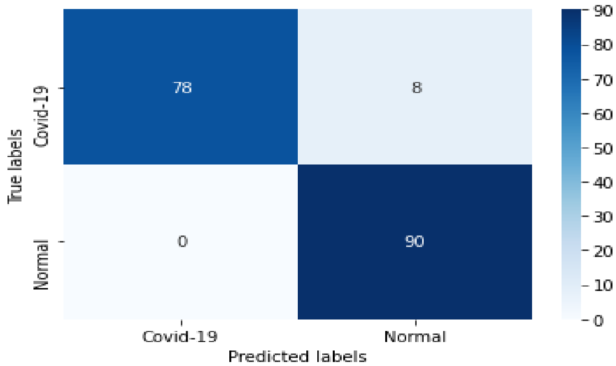


Figure 5. The confusion matrix for MobileNetV2

Table 2 provides comparisons between the three proposed models. The overall cccuracy, precision, recall, and F1-Score were presented for each case of the pre-trained models. Clearly, we can see that, for example, the Xception model achieved 100% accuracy, 99.0% precision, 100% recall, and 99.0% F1-Score for the COVID-19 negative cases. Whereas for COVID-19 positive cases, the Xception model achieved an accuracy of 98.9%, precision of 100%, recall of 99%, and 99% F1-Score. We can also see that the Xception pre-trained model got the best performance compared to the other models (i.e., InceptionV3 and MobileNetV2).

Table 2. The average of the precision, recall, F1-Score, and support for the pre-trained models

Pre-trained Models	Class/ Covid-19	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Xception	Negative	100	99.0	100	99.0
	Positive	98.9	100	99.0	99.0
InceptionV3	Negative	97.8	100	98.0	99.0
	Positive	100	98.0	100	99.0
MobileNetV2	Negative	91.8	100	91.0	95.0
	Positive	100	92.0	100	96.0

Table 3 provides a comparative performance evaluation of our three models (Xception, InceptionV3 and MobileNetV2) with other recent research for the classification of COVID-19. To the best of our knowledge, the best state-of-the-art model achieved an accuracy of 97%, while our Xception model had an accuracy of 97%. Therefore, our model achieved the best accuracy so far, which demonstrated that it might be an effective tool for the classification of COVID-19 from CXR images. Indeed, the number of images used in the training process were greater than those used in previous studies. In fact, even though we had the same dataset, we got better accuracy, as we added a flat layer for faster computation time and reduced overfitting.

Table3. Performance analysis of the study in compression with recent research

Study	Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Hemdan, et al. [8]	VGG19, DenseNet201	90.00	91.50	90.00	90.00
Narin, et al (Narin, 2020)	InceptionV3	**97.00**	99.00	94.00	91.00
Mohammad et al [9]	Xception	91.31	72.00	85.07	77.99
Our Model	**Xception**	**99.43**	**99.00**	**99.00**	**99.00**
Our Model	MobileNetV2	95.45	96.00	95.00	95.00
Our Model	InceptionV3	98.86	99.00	99.00	99.00

Furthermore, Figures 6, 7, and 8 show the training accuracy and validation accuracy with respect to epochs, and all models run up to 50 epochs. Clearly, Figure 6 depicts the performance evaluation of the Xception model; we can see in this figure that the maximum accuracy obtained is 99.43%. Similarly, Figure 7 shows the performance evaluation of the MobileNetV2 model with an accuracy of 95.45%. Figure 8 depicts the performance evaluation of the InceptionV3 model with an accuracy of 98.86%. The experimental results show that the Xception model outperformed the InceptionV3 and MobileNetV2 models. It is clear that better scores of training and validation accuracy were achieved with the Xception model compared to the others, and the state-of-the-art models (to the best of our knowledge). In fact, the Xception model learned well and could correctly classify COVID-19 versus normal cases.

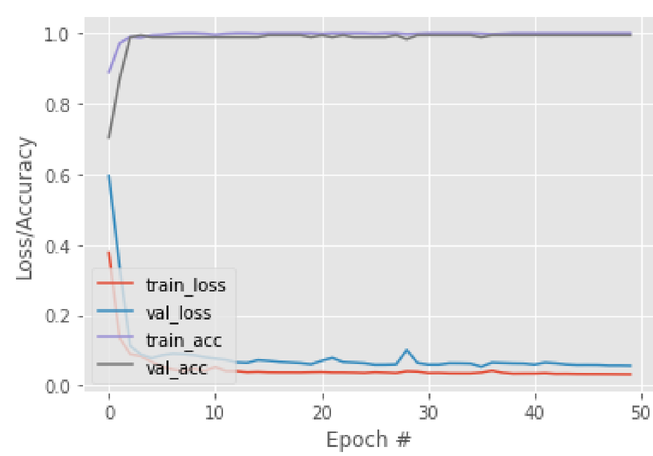


Figure 6. Training and validation loss and accuracy for Xception

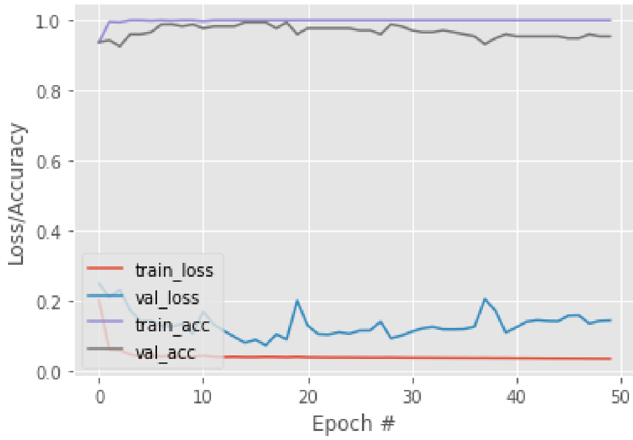


Figure 7. Training and Validation Loss and Accuracy for MobileNetV2

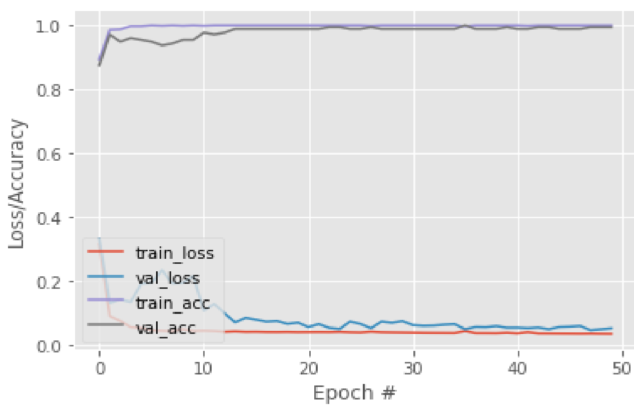


Figure 8. Training and validation loss and accuracy for InceptionV3

The ROC curves of our classifiers are plotted in Figure 9. It is clear that the higher the area under the curve (AUC) in the ROC, the more effective the medical and diagnostic models. For the Xception model, the AUC is 1.000, the InceptionV3 and MobileNetV2, it is 0.986 and 0.963, respectively. Thus, we can confirm that the Xception model can efficiently contribute to the classification of COVID-19 cases from chest X-ray images.

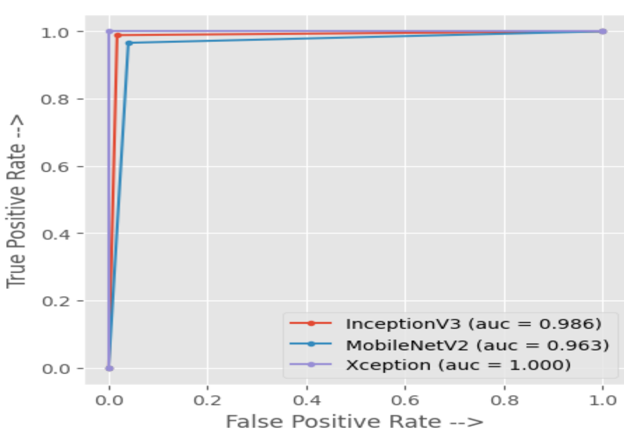


Figure 9. ROC curve for the classifiers

5. Conclusion

Early diagnosis of COVID-19 patients is vital to prevent the spread of the disease to other people. In this paper, we proposed a deep transfer learning-based approach using CXR images. The images used in this paper were collected from two sources: 440 positive COVID-19 CXR images were collected from the GitHub repository, whereas 440 Normal CXR images were obtained from the Kaggle repository. We applied data pre-processing techniques, specifically resizing and normalization. Clearly, three different transfer learning models (Xception, InceptionV3, and MobileNetV2) were proposed for the classification of coronavirus pneumonia infected patients using chest X-ray radiographs. The obtained results have shown that the pre-trained model Xception provides the highest classification performance with an Accuracy of 99.43%, a Precision of 99%, a Recall of 99%, and a 99% F1-score. Moreover, the pre-trained Xception model gives a better result compared to previous studies. Therefore, the pre-trained Xception model could potentially contribute to the classification of COVID-19 from CXR images. Hence, it will help doctors to make decisions in clinical practice due to the high performance.

REFERENCES

[1] M. Ghaderzadeh and A. Farkhondeh, "Deep Learning in the Detection and Diagnosis of COVID-19 Using Radiology Modalities: A Systematic Review," *Journal of Healthcare Engineering,* p. 10, 2021.

[2] T. Ozturk, M. Talo, E. ,. Yildirim, U. ,. Baloglu, O. Yildirim and R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," p. 11, 2020.

[3] A. K. Das, G. Sayantani, T. Samiruddin, D. Rohit, A. Sachin and C. Amlan, "Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network," *Pattern Analysis and Applications,* p. 14, 2021.

[4] P. Saha, M. ,. Sadi and M. M. Islam, "EMCNet: Automated COVID-19 diagnosis from X-ray images using convolutional neural network and ensemble of machine learning classifiers," *Informatics in Medicine Unlocked,* p. 12, 2021.

[5] I. Baltruschat, H. Nickisch, M. Grass, T. Knopp and A. Saalbach, "Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classifcation," nature, 2019.

[6] T. Krishna and K. ,. Hemantha, "Deep Learning and Transfer Learning Approaches for Image Classification," *International Journal of Recent Technology and Engineering (IJRTE),* p. 7, 2019.

[7] E. Goldstein, D. Keidar, D. Yaron, Y. Shachar, A. Blass, L. Charbinsky and I. Aharony, "COVID-19 Classification of X-ray Images Using Deep Neural Networks," p. 22, 2020.

[8] S. Asif, Y. Wenhui, H. Jin and S. Jinhai, "Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Networks," p. 7, 2020.

[9] E. E.-D. Hemdan, M. A. Shouman and M. E. Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," p. 14, 2020.

[10] A. Narin, K. Ceren and P. Ziynet, "Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks," p. 17, 2020.

[11] M. Rahimzadeh and A. Abolfazl, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xceptionand ResNet50V2," *Informatics in Medicine Unlocked,* p. 9, 2020.

[12] J. P. Cohen, P. Morrison and L. Dao, "COVID-19 image data collection," 2020. [Online]. Available: http://arxiv.org/abs/2006.11988.. [Accessed 12 4 2021].

[13] P. Mooney, "Chest X-ray images (pneumonia)," Kaggle Repository, 2018. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia, 2018. [Accessed 12 4 2021].

[14] A. M. Ismael and A. Şengür. Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Systems with Applications*, 164, 114054, 2021.

[15] A. Abbas, M. M. Abdelsamea and M. M. Gaber. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Applied Intelligence*, 51(2), 854-864, 2021.

[16] R. Abdalrhman, M. Elbashir and G. A.Babikir, "Classification of COVID-19 in Chest X-Ray Images using Deep Transfer Learning," *Gezira Journal of Engineering and Applied Sciences*, 16(1), 1-6, 2022.