Comparative Analysis of deep learning techniques for fast detection of COVID-19 using CXR images

Shahid Kamal [1], Mohammad Sarfraz [1], Sumaiya Fatma [1], Ifat Al Fatma [1], Shagufta Parween [1], Muhammad Al Maathidi [2]

[1] *Electrical Engineering Department, Aligarh Muslim University, Aligarh, India*

[2] *Computer Science Department, Bahrain University, Bahrain*

Email: kamalshahid20@gmail.com , msarfraz@zhcet.ac.in

Abstract

Abstract— The COVID-19 pandemic has led to significant public health impacts and overwhelmed healthcare systems, necessitating widespread testing to control its spread. The current gold standard for diagnosing COVID-19, as endorsed by the World Health Organization, is the Reverse Transcription - Polymerase Chain Reaction (RT-PCR) test using a nasal swab sample from the suspected individual. However, this method is resource-intensive and has limited accuracy, with results taking several hours to obtain at a substantial cost. Therefore, alternative methods of rapid detection and diagnosis of COVID-19 are needed. Chest X-Ray (CXR) images, widely accessible through computed tomography (CT) scans and chest x-rays, present an opportunity for early COVID-19 diagnosis. This research uses the COVID-QU-Ex dataset, consisting of 33,920 CXR images, including 11,956 COVID-19 positive, 10,701 normal, and 11,263 infected samples. To detect COVID-19 from CXR images, seven deep learning models were used: AlexNet, MobileNet, ResNet, DenseNet, ShuffleNet, InceptionV1, and XceptionNet. The performance of each model was evaluated using five parameters: accuracy, recall, precision, specificity, and F1-Score. Our experimental results demonstrate that the XceptionNet model had the highest accuracy of 98.86% in detecting COVID-19 from CXR images. This model can potentially aid radiologists in the early diagnosis of COVID-19.

Keywords—Deep learning, COVID-19, CXR imaging, Convolutional neural networks

1. Introduction

The novel coronavirus emerged as a tragic disease in Wuhan, a city in china, at the end of 2019. It was a severe acute respiratory syndrome. As per a recent record by WHO[1], it has caused 631 million causalities and 6.58 million deaths. Due to this pandemic, the whole world suffered socially, mentally, and economically. The transmission rate of the virus was very high, which led to the government's strict lockdowns in several countries to reduce the infection rate[2]. Mass vaccination was helpful in some countries, but on the other hand, several countries entered second and third waves. The prime symptoms of an infected person are fever, cough, pneumonia, muscular pain, and shortness of breath. For diagnosing COVID-19 earlier, Reverse Transcription-Polymerase Chain Reaction(RT-PCR) was used, which was expensive, time-consuming, and needed professionals[3].

In high-income nations, RT-PCR kits are widely accessible, whereas, in low-income countries like Bangladesh, the price of an X-ray ranges from 450 BDT to 1200 BDT, but on the other hand cost of RT-PCR is 3000BDT.

In India cost of an X-Ray varies from Rs.183 to Rs.1370, whereas the cost of RT-PCR varies from Rs.980 to Rs.1800. Due to the lack of healthcare facilities, medical professionals, and adequate supplies, RT- PCR kits need to be readily available[4]. Besides, these patients must travel to laboratories to get themselves tested, affecting people on the way. As it was time-consuming, many countries started using rapid antigen detection tests, but they gave less accurate results.

As a result, researchers have begun utilizing Deep Learning to automatically detect the virus using computed tomography (CT) scans and chest X-ray (CXR) images. [2]. CXR imaging systems are readily available in medical centers and are economical.

Also, CXRs are handy, making them easier to use in homes, thereby reducing transmission risk. Reliable coronavirus detection is crucial; therefore, a fully automated classification technique is needed[1]. When the dataset was rare, the researchers used image augmentation techniques for deep learning models[3].

Large datasets enhance performance and eradicate overfitting problems [4]. This paper is divided into different sections. Section 1 and 2 gives a brief introduction and related work. In section 3, we analyze the model's performance. Sections 4 and 5 cover the deep learning architecture and performance metrics, respectively. Results and discussions are covered in section 6. Section 7 contains the conclusion and future work.

2. Literature review

2.1 Related Work

The COVID QU Ex dataset, created by A.M. Tahir et al.[1], The largest CXR dataset and comprises 11,956 COVID-19 and 10,701 Normal CXR images. Its performance accuracy is 96.11%, and its sensitivity is higher than 99%. They showed sensitivity as the primary parameter to detect COVID-19.

Aysen Degerli et al.[5] experimented on the QaTa-COVID-19 dataset on the OsegNet model, with a precision of 98.09% for COVID-19 detection. Agata Gielczyk et al.[6] presented a paper on a Novel Lightweight supporting COVID-19 diagnosing approach based on the X-Ray image. This approach was fast and efficient, with an accuracy of 1.00 and a precision of 1.00.

Tawsifur Rahman et al.[2] presented a paper in which CNN models were used to identify COVID-19 in two symptomatic and asymptomatic suspects using cough and breath sound spectrogram images.. Karen Simonian et al.[7] presented a paper in which they demonstrated that their models could be used for various tasks and datasets. To overcome the problem of the traditional method of detecting COVID-19, Mei-Ling Huang et al.[8] used CNN for efficient and accurate detection of COVID-19 from CXR and CT images. They used seven convolutional neural networks: InceptionV3, DenseNet121, EfficientNet-130, Efficient Net V2, MobileNet V2, XceptionNet, and ResNet50. They have found that InceptionV3 gives the highest accuracy of 96.50% before fine-tuning, and EfficientNet V2 gives the highest accuracy of 97.73% after fine-tuning.

Shah Siddiqui et al.[4] compare deep learning models for detecting COVID-19 in CXR images, but they compared it on a very small dataset. This dataset is divided into COVID and Normal, with 579 CXR images belonging to the COVID class and 1773 CXR images belonging to the Normal class. They have taken three deep learning models, VGG 16, VGG-19, and InceptionV3, for comparison. VGG 16 performed best among all three models with an accuracy of 90%.

Our work aims to effectively detect COVID-19 cases from CXR images and acompare different deep-learning models for detecting COVID-19 from CXR images.

Furthermore, to achieve the highest accuracy in classifying COVID-19 from CXR image than the existing deep learning model. To train deep learning models of convolutional neural networks for detecting COVID-19 on the largest available dataset, as earlier researchers had trained neural network models on a minimal dataset.

3. Methodology

3.1 Dataset

To the best of our knowledge, the COVID QU Ex dataset, created by academics at Qatar University, is the largest dataset used in this work.
This dataset contains 33,920 CXR images, of which 11,956 are COVID-19, 11,263 are non-COVID, and 10,701 are normal (healthy) CXR images[1]. The quantity of image samples used for model training, validation, and testing is displayed in Table 1.
This dataset was compiled from various publicly available datasets such as the COVID-19 Chest X-Ray dataset, Chest X-Ray pneumonia dataset, Padchest dataset, and Montgomery and Shenzhen CXR lung mask.

3.2 Data Pre-processing

Data pre-processing was performed before training the model. All the images were resized to 256×256 before being used as input to the algorithm. The images were shuffled during training, which ensures that the model does not memorize the incoming data sequence.

3.3 Deep Learning Models

In this study, seven deep-learning models have been re-implemented and trained for our analysis. The models used are AlexNet, MobileNet, ResNet, DenseNet, InceptionV1, ShuffleNet, and XceptionNet. These models are based on convolutional neural networks[9], which work on images to detect the features in the images. Convolutional Neural Networks started with LeNet-5, simply a stack of convolution for feature extraction and max pooling operation for spatial sampling.

AlexNet: AlexNet was first introduced in 2012 in the ImageNet large-scale visual recognition challenge (ILSVRC)[10]. It won the ImageNet competition by beating the previous best models with 10 % accuracy. AlexNet was trained on a subset of the ImageNet dataset comprising 1.2 million images of 1000 classes. AlexNet has eight layers (5 convolutional and the remaining fully connected layers) with learnable parameters. The activation function being used is ReLU. To avoid overfitting, it also makes use of dropout regularisation. For our needs, the last layer has only two dense units instead of 1000.

InceptionV1: InceptionV1 was first introduced in the 2014 ILSVRC ImageNet competition[11]. It has a total of 22 deep-layer networks. This Network uses multiple sets of InceptionV1 modules, which use a network-in-network approach. The InceptionV1 modules heavily use 1×1 convolutions. It serves the purpose of a dimensionality reduction module to remove the computational bottlenecks and helps the model get an enormous depth and width without sacrificing performance. A 1×1 convolution is applied before every 3×3 and 5×5 convolutions. InceptionV1 is an architecture consisting of InceptionV1 modules stacked upon each other. The basic idea of InceptionV1 is that instead of us selecting the filter sizes or pooling layers, it lets the Network decide what parameters it wants to learn. In our case, we have used InceptionV1, which has 28 convolutional layers.

ResNet: He et al. introduced ResNet in the 2015 Large Scale Visual Recognition Challenge (ILSVRC) ImageNet competition. Deep neural networks often suffer from vanishing gradient problems as the gradients are back-propagated to earlier layers, and the gradients become infinitely smaller due to repeated multiplications. He et al.[12] proposed a deep residual learning framework for training deeper networks, incorporating the main idea of an "Identity shortcut connection "that skips one or more layers. They compared three different models, which are VGG-19, plain Network (34 parameter layer), and ResNet network with 34 layers; they showed that training error reduces with the residual connection. The models were trained on 1.28 million training images. In our case, we use ResNet50, which has a 50 convolution layer.

MobileNet: MobileNet was introduced in the 2017 Large Scale Visual Recognition Challenge (ILSVRC) ImageNet competition[13].MobileNet is the most computationally efficient Network for mobile and embedded devices. MobileNet is most widely used in real-world object detection and image classification applications since it is very lightweight and takes much less memory. MobileNet uses depth-wise separable convolutions instead of normal convolutions. Depth-wise separable convolutions involve two operations: A depth-wise convolution followed by a pointwise convolution. In depth-wise convolution, each filter is applied to each input channel. Then a pointwise convolution is needed to map the output of depth-wise convolution to the desired number of output channels. We have used MobileNet V1 in our paper, which contains 28 layers consisting of depthwise and pointwise convolutions.

DenseNet: Huang et al.[14] proposed DenseNet in computer vision and pattern recognition(CVPR) in 2017.DenseNet alleviates the problem of vanishing gradients and promotes feature reusability and parameter reduction. DenseNet uses dense connections for the information flow between layers. In DenseNet, each layer receives feature maps of all preceding layers as input and passes its feature maps to the subsequent successive layers. The DenseNet architecture is divided into dense blocks followed by a transition block. A transition block is used for down-sampling, consisting of batch normalization followed by a 1×1 convolution layer and a 2×2 average pooling layer. We have used DenseNet-121 in this paper, which has 121 convolutional layers.

XceptionNet: Xception was first introduced by F. Chollet et al.[15] in Computer Vision and Pattern Recognition (CVPR) in 2017. XceptionNet stands for Extreme InceptionV1. XceptionNet has outperformed InceptionV3 on the ImageNet dataset. It has the same number of trainable parameters as Inception V3. XceptionNet excessively uses Depth wise separable convolutions. In XceptionNet, channel-wise spatial convolutions are first applied, and then the channel compression is achieved using a 1×1 convolution, whereas the InceptionV1 does the Reverse of this operation. InceptionV1[11] uses ReLU(non-linearity) activation function after each convolution operation, while XceptionNet does not use non-linearity. XceptionNet has 36 convolutional layers that serve as feature extraction, and uses a linear residual connection. The number of learnable layers in XceptionNet is 71.

ShuffleNet: Zhang et al.[12] first introduced ShuffleNet in Computer Vision and Pattern Recognition (CVPR) in 2018. ShuffleNet is a highly efficient network for low computational power devices like drones, robots, and smartphones. ShuffleNet outperforms MobileNet on ImageNet and MS COCO Object detection tasks. It also achieves 13 × actual speed up over AlexNet. ShuflleNet is basically inspired from InceptionV1, Xception, and ResNet. Zhang et al. proposed a pointwise group convolution along with a residual connection to reduce the computational complexity while maintaining accuracy. ShuffleNet also performs channel shuffling for group convolutions. Channel shuffling leads to improved classification scores. ShuffleNet architecture contains a total of 50 convolutional layers.

We have taken Chest X-Ray images of two classes, COVID-19 and Normal, from the COVID QU Ex dataset[1]. The sample of COVID-19 and Normal images are shown in Fig. 1.

Table 1. Covid QU Ex dataset images in each category

Split Data	**COVID-19**	**Normal**
Training Set	7658	6849
Validation Set	1903	1712
Test Set	2395	2140

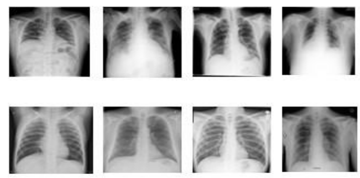


Fig. 1. Chest X-Ray sample images

4. Deep Learning Architecture

Deep learning techniques have shown remarkable results in image classification, detection, segmentation, and neural style transfer in recent years. The neural Network which is used in image classification is a Convolutional Neural Network (CNN). CNN transforms input images into matrix format, detecting useful features to classify any image. We often use multiple sets of CNN layers to detect the rich set of hierarchical features. CNN layers are used in conjunction with fully connected layers. The general procedure followed for the classification task of COVID-19 and Normal cases in this study is shown in Fig. 2.

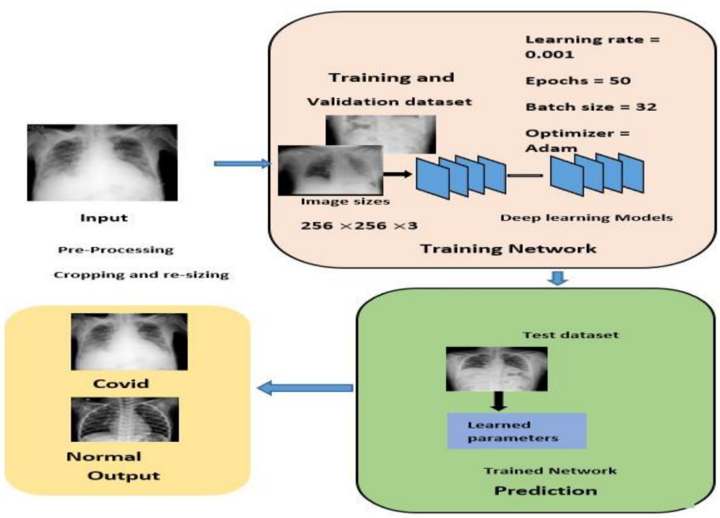


Fig. 2. Procedure for the classification of COVID-19 and Normal[3]

A brief description of CNN models are described below:

Input Layer: In this layer, pre-processed input images are fed, and all the images are resized (in our case) 256×256×3 (width, height, and channels).

Convolutional Layers: The primary layer in CNNs extracts features from input images; in this layer, a kernel (filter) performs a convolution operation by maintaining a spatial relationship.

Max-Pooling Layer: Max-pooling layers are applied to shrink the feature maps of CNN layers by achieving translational invariance.

Batch Normalisation Layer: It normalizes the features and provides training stability, making the training process faster.

Rectified Linear Unit (ReLU): Neurons learn non-linear decision boundaries through the activation function.

Fully Connected Layer: It is a feed-forward neural network used for feature classification after convolutional operations.

Output Layer: The final layer of the CNN has neurons equal to the number of classes. This layer is responsible for the final prediction of the classification task. In our case, the number of neurons in the output layers is two.

5. Performance Metrics

We have considered five metrics to evaluate and compare the models on the test dataset: accuracy, precision, recall, specificity, and F1 score. They are determined by calculating true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

1) Accuracy is the ratio of correct predictions out of total predictions.

$$Accuracy(\%) = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

2) Precision is the percentage of positive class Chest X-Ray samples correctly identified out of all the samples.

$$Precision(\%) = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

3) Recall (Sensitivity) is the percentage of correctly classified positive class chest X-Ray samples among all classified positive samples.

$$Sensitivity(\%) = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

4) *Specificity* is defined as the sensitivity of a negative class sample.

$$Specificity(\%) = \frac{TN}{TN+FP} \times 100\% \quad (4)$$

5) F1 score is the harmonic mean of sensitivity and precision

$$F1 = 2\,\frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (5)$$

6. Results and Discussion

In this research paper, we trained deep learning models on the COVID-19 X-ray dataset [1] for binary classification, whether COVID-19 positive or normal. Initially, a subset of 2913 X-ray images from the dataset showed poor performance on the test dataset. To enhance the results, we trained the model on the entire dataset. To avoid overfitting, we applied dropout regularisation, randomly turning off 20% of neurons during training. We implemented the models in TensorFlow, a highly efficient deep-learning library in Python. The softmax function served as the activation function in the output layer, and we used the categorical cross-entropy loss as the loss function. We employed the Adaptive Moment Estimation (Adam) optimizer with an initial learning rate (α) of 0.001 and gradually decreased it until reaching the maximum number of epochs. We kept the default momentum values, β1 set to 0.9 and β2 set to 0.999. The maximum number of epochs was set to 50 for training –a batch size of 32 for both training and testing. We randomly shuffled the images during training to prevent learning the training order and minimize bias.

Table 2 displays the model's complexity in terms of trainable parameters. AlexNet has the highest parameters of 71.99 million, while ShuffleNet has the least at 0.92 million paramters. The accuracy of the models heavily relies on the quality of the training dataset. This study used the COVID-19 X-ray dataset [1] consisting of high-quality X-ray images with distinct features. Together with utilizing suitable regularisation techniques, optimizers, and loss functions, the models accurately classified X-ray images as COVID-19 positive or normal.

Table 2. Model Complexity and Computational Time

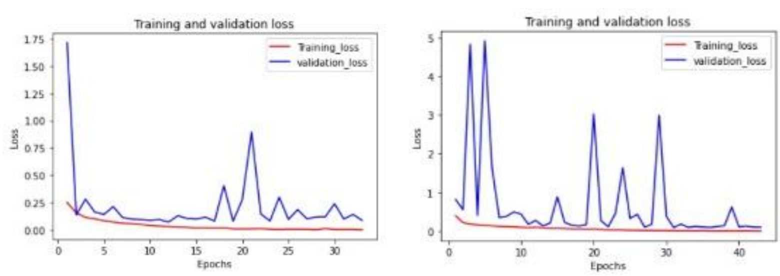
Model	**Parameters (Million)**	**Epochs**	**Training Time(Minute)**
AlexNet	71.99	46	174.56
MobileNet	3.24	43	197.43
ResNet	23.54	43	216.88
DenseNet	6.96	47	280.88
XceptionNet	**20.86**	**33**	**309.61**
InceptionNet	3.22	29	188.38
ShuffleNet	0.92	50	210.6

XceptionNet performs better because of the efficient implementation rather than because of the Network's capacity. All the Models were evaluated on the test dataset, and the evaluated metrics are presented in Table 3. The loss and accuracy plots of XceptionNet and ResNet are shown in fig. 3 and fig. 4, respectively. Table 3 shows that the XceptionNet and ResNet have achieved the highest accuracy, 98.86%, and 98.35%, respectively.

Table 3. Comparison of Performance Metrics of deep learning models

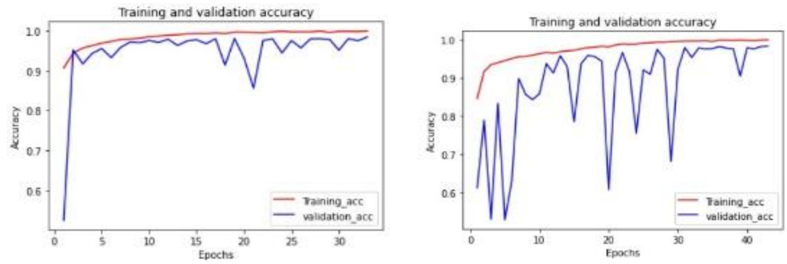
Model	**Accuracy (%)**	**Recall (%)**	**Precision (%)**	**Specificity (%)**	**F1 Score**
AlexNet	97.54	97.86	96.96	97.24	0.974
MobileNet	98.26	98.43	97.88	98.11	0.982
ResNet	98.35	98.36	98.14	98.33	0.983
DenseNet	92.92	87.30	97.46	97.95	0.921
ShuffleNet	97.21	97.53	96.59	96.91	0.971
InceptionNet	98.26	98.84	97.52	97.75	0.982
XceptionNet	**98.86**	**98.93**	**98.65**	**98.79**	**0.988**

The following two models whose accuracy is better after XceptionNet and ResNet are InceptionV1Net and MobileNet, having 98.26 % each. On the other hand, ShuffleNet and AlexNet have an accuracy of 97.21% and 97.54%, slightly less than MobileNet and ResNet. DenseNet has abysmal performance among all the models, with an accuracy of 92.92%. The test dataset consists of 4545 images of both COVID-19 and Normal cases. The confusion matrix of XceptionNet and ResNet is shown in fig. 5, and the confusion matrix of InceptionV1 and MobileNet is shown in fig. 6.



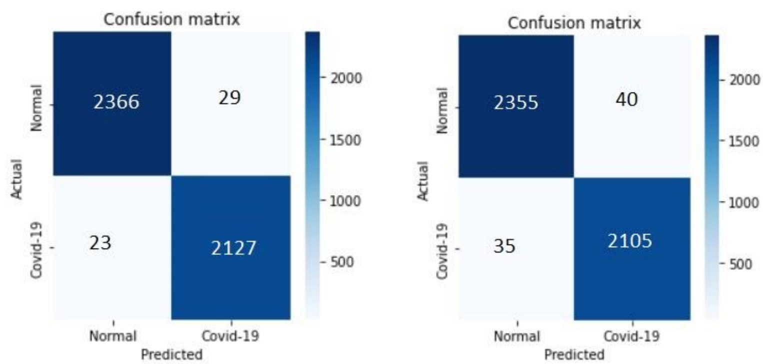
(a) (b)

Fig 3. (a) XceptionNet Loss and (b) ResNet Loss



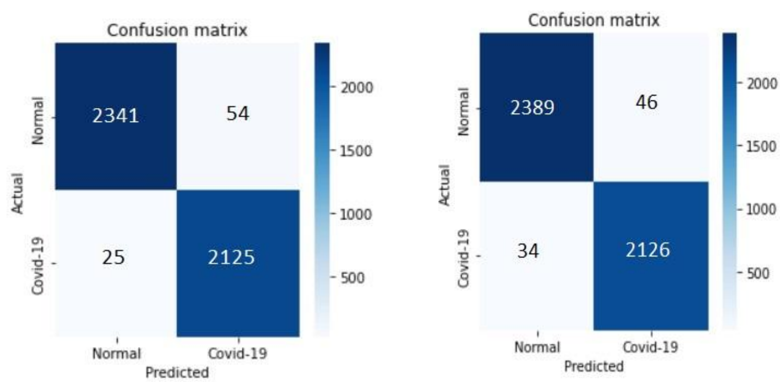
(a) (b)

Fig 4. (a) XceptionNet Accuracy and (b) ResNet Accuracy



(a) (b)

Fig 5. Confusion matrix plot of (a) XceptionNet and (b) ResNet



(a) (b)

Fig. 6. Confusion matrix plot of (a) InceptionV1Net and (b) MobileNet

In a disease detection task, we should prioritize increasing the recall or reducing the number of false negatives since we do not want a patient to be misclassified as negative.

In reality, the patient is positive that is having the disease. The confusion matrix of XceptionNet shows that it has the highest recall of 98.93%. It only misclassified 23 images as Normal cases and 29 as COVID-19 cases on the test dataset. While the second highest accurate model, ResNet, misclassified 35 images as Normal and 40 as COVID-19, shown in fig. 5 (b). Even though InceptionV1 has slightly lesser accuracy than ResNet, it wrongly predicted only 25 images as Normal cases. InceptionV1 and MobileNet achieved the same accuracy, which is 98.26%. MobileNet's recall is only 98.43%, while InceptionV1 has 98.84%. XceptionNet has the highest F1 score of 0.988, which suggests choosing this model for the covid-19.

The only limitation of our work is that binary classification has been used to detect COVID-19 from Chest X-Ray images. However, we plan to work on the multiclass classification in future work.

7. Conclusion

In this study, seven deep-learning models have been examined for the diagnosis of COVID-19 from the broadly available Chest X-Ray dataset (COVID QU Ex). The evaluation of the models were done on five performance metrics: Accuracy, Precision, Sensitivity, Specificity, and F1 score. The best-performing model was XceptionNet having an accuracy of 98.86%. Deep learning models often get overfitting problems when trained with fewer datasets. Most deep learning models were examined on datasets containing only a few hundred samples of Chest X-Ray in all the previous work. Therefore, they were more prone to overfitting.

We have found that the model predictions on the test dataset have improved due to training on a large dataset. The accuracy would have been further improved if the dataset had been more extensive than COVID QU Ex. In our future work, we plan to improve the accuracy further by hyperparameter tuning and extending it to a multi-class classification.

References

[1] Tahir, A. M., et al. (2021). COVID-19 infection localisation and severity grading from chest X-ray images. Computers in Biology and Medicine, 139, doi: 10.1016/J.COMPBIOMED.2021.105002.

[2] Rahman, T., et al. (2022). QUCoughScope: An Intelligent Application to Detect COVID-19 Patients Using Cough and Breath Sounds. Diagnostics, 12(4), doi: 10.3390/diagnostics12040920.

[3] Majeed, S. A., Darghaoth, A. M. H., Hamed, N. M. Z., Yahya, Y. A., Raed, S., & Dawood, Y. S. (2021). Detection of COVID-19 from X-Ray Images Using Transfer Learning Neural Networks. In Proceedings of the 2021 2nd Information Technology to Enhance E-Learning and other Applications Conference (IT-ELA 2021), 58–63. doi: 10.1109/IT-ELA52201.2021.9773657.

[4] Siddiqui, S., et al. (2022). A Comparative Study of Deep Learning Models for COVID-19 Diagnosis Based on X-Ray Images. In (pp. 163–174). doi: 10.1007/978-981-16-9101-0_12.

[5] Degerli, A., Kiranyaz, S., Chowdhury, M. E. H., & Gabbouj, M. (2022). OSegNet: Operational Segmentation Network for COVID-19 Detection using Chest X-ray Images. arXiv, vol. abs/2202.1, no. Cvdi, 5–9. [Online]. Available: http://arxiv.org/abs/2202.10185

[6] Piorkowski, A., et al. (2022). A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images. Journal of Clinical Medicine, 11(19), 5501, doi: 10.3390/JCM11195501.

[7] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations (ICLR 2015) - Conference Track Proceedings, 1–14.

[8] Huang, M. L., & Liao, Y. C. (2022). A lightweight CNN-based network on COVID-19 detection using X-ray and CT images. Computers in Biology and Medicine, 146, 105604, doi: 10.1016/J.COMPBIOMED.2022.105604.

[9] L., Y. et al. (1998). Learning algorithms for classification: a comparison on handwritten digit recognition. (pp. 261–276).

[10] Krizhevsky, B. A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90.

[11] Szegedy, C., et al. (2015). Going Deeper With Convolutions. (pp. 1–9).

[12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (pp. 770-778).

[13] Howard, A. G., Sandler, M., Zhu, M., Zhmoginov, A., & Chen, L. C. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861.

[14] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2261–2269). doi:10.1109/CVPR.2017.243

[15] Chollet, F. (2017). Xception: Deep Learning with Computer Vision and Pattern Recognition (CVPR) (pp. 1800–1807). doi:10.1109/CVPR.2017.195