International Journal of Advanced Technology and Engineering Exploration, Vol 10(103)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2022.10100044

Research Article

Transfer learning with fine-tuned deep CNN model for COVID-19 diagnosis from chest X-ray images

Mamta Patel[1*] and Mehul Shah[2]
Research Scholar, Gujarat Technological University, Gandhinagar Highway, Chandkheda, Ahmedabad – 382424, Gujarat, India[1]
Associate Professor, Vishwakarma Government Engineering College, Chandkheda, Ahmedabad – 382424, Gujarat, India[2]

Received: 22-August-2022; Revised: 21-June-2023; Accepted: 24-June-2023

Abstract
The COVID-19 pandemic has had a significant impact on people's lives, necessitating accurate detection and early diagnosis to control the dissemination of virus. Reverse transcription-polymerase chain reaction (RT-PCR) is the most prevalent diagnostic strategy, but its accuracy is influenced by various factors such as sample collection, timing, and processing. Deep Convolutional Neural Networks (DCNNs) have shown great promise in medical image analysis and are consequently being utilized for the diagnosis of COVID-19 from radiographic images. This study evaluates the effectiveness of different convolutional neural network (CNN) architectures with optimum hyperparameters for COVID-19 diagnosis using publicly available chest radiography datasets. The evaluated models included both CNN architectures built from scratch and pre-trained CNN architectures, such as residual network (ResNet-50), visual geometry group (VGG-16), VGG-19, Inception-V3, and MobileNet-V2. The experimental results demonstrate that MobileNet-V2 achieved 96% accuracy, precision, recall, F1 score, and area under the curve (AUC), making it a prospective and acceptable model for COVID-19 diagnosis. In contrast to existing models, the proposed model's evaluation also includes an assessment of network training time and memory consumption. The study also describes the web deployment of a deep CNN-based computer-aided diagnosis (CAD) system that can assist doctors in diagnosing COVID-19 faster, more accurately, and more consistently. This advancement leads to better patient outcomes and improved efficiency within the healthcare system.

Keywords
COVID-19, Chest radiography, Deep learning (DL), Convolutional neural network (CNN), Transfer learning (TL).

1.Introduction
The World Health Organisation (WHO) proclaimed the COVID-19 pandemic in response to the appearance of a novel coronavirus in Wuhan, China, in December 2019[1]. This public health crisis has caused global concern, significantly impacting the economy, and healthcare systems [2]. According to a medical research article [3], the rapid spread of the COVID-19 epidemic has posed a challenge for physicians in differentiating it from typical respiratory symptoms, leading to misdiagnosis among patients with pneumonia who exhibit similar symptoms. A critical step in combating this global epidemic is the prompt and reliable detection and treatment of COVID-19-infected patients [4].

The reverse transcription-polymerase chain reaction (RT-PCR) test is the most used for COVID-19 detection [5–6]. However, a study published in [7] indicates that RT-PCR can yield false negative (FN) results in cases where patients have a lower viral load at the initial stage, when the composition of the sample inhibits chemical action, or due to laboratory errors. Therefore, detecting COVID-19 solely with RT-PCR is not advisable. X-rays and computed tomography (CT) scans of the chest are most effective modalities in identifying the morphological patterns of lung lesions associated with COVID-19 compared to other diagnostic methods such as RT-PCR [8]. The study recommends the use of imaging techniques as an important tool for early and accurate COVID-19 diagnosis, particularly when RT-PCR results are inconclusive or negative. However, it is important to note that CT scans have a higher risk of spreading COVID-19, along with longer scan duration, high radiation doses, and higher costs, making chest

*Author for correspondence

radiography the most comprehensive and cost-effective diagnostic tool for COVID identification [9]. Nevertheless, manual interpretation of chest radiography is slow and prone to human interpretation errors. Since COVID-19, pneumonia, and other types of lung abnormalities are observed with same characteristics, their reliability is questionable [10]. Considering these limitations, there is a need for the healthcare system to develop alternative methods using computational intelligence tools for COVID-19 detection along with additional lung infections.

With significant advancements in the image processing area, it has garnered attention of researchers, leading to the development of cost-effective and technologically accepted architectures without compromising accuracy. Recently, deep convolutional neural network (DCNN) models have emerged as powerful tools, with various algorithms and architectures being developed [11–14]. The convolutional neural network (CNN) models facilitate the analysis of large amounts of data and utilization of advanced algorithms, offering solutions to complex clinical problems. The remarkable performance of CNNs in disease diagnosis provides a solid foundation for their utilization in the diagnosis of threatening COVID-19 diseases. This technology offers patients a unique opportunity to receive a prompt, safe, and cost-effective diagnosis.

The acceptance of deep CNNs and their promising results in COVID-19 detection have prompted significant interest. However, several challenges need to be addressed in order to maximize their potential. In a study [15], the authors discussed some of the challenges associated with utilizing DCNNs for COVID-19 detection. These challenges include scarcity of large and diverse datasets, the need for expert knowledge in designing and fine-tuning network architectures, and interpretability of a network's decision. Furthermore, the quality of the input images and the existence of confounding factors affect DCNN performance. Another challenge, as highlighted in the research article [16], is the issue of imbalanced data when using CNNs for COVID-19. The limited availability of positive cases compared to abundance of negative cases results in imbalanced datasets, which introduce bias in the learning process and affect the accuracy of the CNNs. To mitigate this, the authors discuss various techniques, such as oversampling, under-sampling, and cost-sensitive learning, to address the imbalanced data problem. In a case study [17], the authors point out another challenge related to the limited COVID-19 datasets, which is the risk of overfitting when using DCNNs. To overcome this issue, the authors propose the use of data augmentation techniques.

Selecting an optimal model from the multitude of options available in the domains of machine learning (ML) and deep learning (DL) poses a challenging and time-consuming task. Although extensive research has been conducted on several traditional methods in these fields, there is still potential for further improvement. However, enhancing these methods may present significant technical challenges [18].

Under the information technology enabled services policy of the Government of India, the Department of Science and Technology (DST) granted a high computing center at Information Technology Department of Vishwakarma Government Engineering College, Chandkheda. The experiments utilized the same resources for data processing. Our research aims to enhance the current performance metrics of DL while also incorporating local web deployment of the trained model. This will improve access to the computer-aided diagnosis (CAD) system from remote locations, resulting in a more effective and efficient disease detection system. To address the challenges associated with imbalanced data sets and other related issues, several techniques were employed. These techniques included data augmentation, transfer learning (TL) with tuning for overfitting, and resampling.

The proposed work aims to assess the efficacy of CNN in detecting COVID-19 using chest X-rays. The research focuses on assessing effectiveness of three different approaches: a scratch CNN architecture, pre-trained models with TL, and TL with fine-tuning. These approaches are utilized to categorize chest X-ray images into two distinct groups: COVID-19-infected and normal. Various performance metrics, including accuracy, F1 score, precision, recall, and area under the curve (AUC), are employed to analyse models. The results obtained from each approach are compared to determine their relative effectiveness.

Our work makes several contributions:
1. Development of an effective DL model for COVID-19 screening from chest X-rays
2. Comparison of CNN models built from scratch and various pre-trained models to identify the most efficient CNN model for predicting COVID-19 disease
3. Investigation of the impact of hyperparameters and dataset size on model performance

4. Optimization of model parameters to provide intelligent solutions for accurate COVID-19 detection

This research article is structured into different sections. Section 2 provides a literature review that examines previous research on DL methods for COVID-19 diagnosis. Section 3 covers the proposed method for developing a DL model specifically designed for COVID-19 screening. This section also outlines various approaches utilized to enhance performance. Section 4 is dedicated to discussing the results and discussion of the study. Finally, the conclusion summarizes the findings and presents potential future directions for further research.

2.Literature review

DL is commonly used in medicine for disease classification and CAD [19‒24]. CNNs capability of extracting features from images makes them most suitable for such tasks [24], and they have demonstrated comparable or superior performance to trained radiologists [22–24]. TL has also proven effective in DL applications as it leverages pre-trained networks from various domains, reducing the time and resources required to achieve high performance [25]. The emergence of COVID-19 and its global impact have spurred the development of DL based COVID-19 CAD using different imaging modalities.

Yang et al. [23] evaluated multiple pre-trained CNN models, including visual geometry group-16 (VGG-16), DenseNet-121, residual network-50 (ResNet-50), and ResNet-152, to create a COVID-19 binary classification system using CT scans. They utilized the fast ResNet framework to select the optimal model architecture, pre-processing techniques, and training settings. The results demonstrated 96% accuracy in binary classification, with an F1 score of 96%. However, it is important to note that the study is limited to the use of CT scans. Additionally, the dataset employed in the study is relatively small, consisting of only 618 images, which may impact the generalizability of the findings.

The study presented by Toraman et al. [24] utilizes capsule network (CapsNet) CNN for both binary and multiple category classification applications. The dataset used in the experiment consisted of 1050 x-ray images for each class. To evaluate the performance, a tenfold cross-validation technique was utilized, yielding an accuracy score of 97% and 84% for binary multi-class classification respectively. However, the study had certain limitations, including a relatively small dataset and the absence of an external validation dataset.

He et al. [25] presented sample-efficient DL techniques for accurate COVID-19 diagnosis using CT images. They introduced a novel training strategy that addresses data imbalances and optimizes model performance across various metrics, even with a limited number of training images. Specifically, they proposed a method called Self-Trans, which combines TL and contrastive self-supervised learning. Using this technique, 90.63 % accuracy score was obtained with a small number of CT images for training. However, one limitation of their study is that the dataset was collected from a single hospital in China, which restricts the generalizability of the model to other populations.

Kassania et al. [26] compared the performance of seven distinct pre-trained DL architectures using TL for the categorization of COVID-19 versus pneumonia using both CT and X-ray images. The images underwent preprocessing and augmentation before being fed to the network. Among all seven DL models, Inception-ResNetV2 achieved the highest accuracy of 92.18%. However, the study had several limitations. First, the dataset used in the study was relatively small, comprising only 349 CT scans and 331 X-rays. The small size of the dataset in the current research limits the generalizability of its results. Additionally, study lacks a detailed analysis of false positive (FP) and FN, which are crucial aspects of medical diagnosis.

To assess the impact of the convolutional layer count in a scratch CNN model, Basha et al. [27] developed three different CNN models with varying numbers of convolutional layers. They compared the performances of these models with pretrained architectures. Surprisingly, the model with only four convolutional layers achieved an impressive accuracy of 97.5 % surpassing the performances of the three and five-convolutional layer models.

Haque et al. [28] suggested an algorithm for classifying normal, COVID-19, and pneumonia using chest radiographs. They employed a fusion approach, combining ResNet-101 and ResNet-151, and improved the dynamic weight ratio to elevate this model's performance. During the testing phase, the model demonstrated a remarkable accuracy of 96.1 %.

A novel CNN model called COVID-Net was built by Wang et al. [29] for diagnosing COVID-19 using chest radiographs from the COVIDx dataset. The architecture of COVID-Net was designed using lightweight residual patterns and classified the images into three categories: COVID-19, pneumonia, and normal. To assess the performance of COVID-Net, the authors evaluated it on a separate test set consisting of 1,000. COVID-Net gained a sensitivity of 93.5% and a specificity of 95.7%, according to the data. These metrics indicate the model's ability to correctly identify positive and negative cases of COVID-19. Furthermore, COVID-Net demonstrated superior performance compared to the VGG-19 and ResNet-50 network architectures in terms of test accuracy.

Diaz-escobar et al. [30] utilized the POCUS dataset, which consists of 3326 lung ultrasounds from normal, COVID-19, and pneumonia patients, for training and fine-tuning the models. They conducted two sets of tests: binary classification and three-class classification. Several pre-trained DL architectures were employed in the experiments. Among the examined models, Inception-V3 had the best accuracy of 89.1%, according to the testing data. This indicates its effectiveness in accurately classifying lung ultrasounds into the appropriate categories.

Wang et al. [31] proposed a TL approach to develop a COVID-19 diagnosis using CT scans. They investigated impact of two image enhancement techniques, histogram equalization (HE) and contrast limited adaptive histogram equalization (CLAHE), on improving the quality of the images. Several pre-trained networks, including ResNet-101, DenseNet-201, VGG-19, EfficientNet-B4, and MobileNetV2 were utilized to train the model. The CT images were processed using the chosen enhancement technique, and then networks were fine-tuned with these processed data. The experimental results demonstrated that the VGG-19 model combined with CLAHE achieved the highest accuracy and recall rates. When tested on a SARS-CoV-2 dataset, the model had 95.75% accuracy and 97.13% recall.

Wang et al. [32] conducted experiments to classify and localize eight common thoracic diseases using ChestX-ray8 database. The experiments involved training a CNN network using weight parameters for VGG-16, AlexNet, ResNet-50, and GoogleNet architectures. According to findings, ResNet-50 outperformed the other architectures for multi class classification of lung diseases.

Kaur et al. [33] developed a COVID-19 detection model using DL techniques, including TL and scratch learning. support vector machine (SVM), k-nearest neighbor (KNN), and random forest (RF) were used as classifiers, and pre-trained models like VGG-16, ResNet-50, and MobileNet were used to extract features. The CNN scratch technique with an RF classifier achieved a high accuracy of 94.03%. However, the model's limitation was that it was trained using only 285 chest X-rays.

Bahuguna et al. [34] built a hybrid model for COVID-19 diagnosis from chest radiographs, utilizing a combination of a CNN and a KNN algorithm. The proposed model incorporated a VGG-16 for feature extraction and a KNN algorithm for classification. The proposed model achieved impressive performance metrics with a high accuracy of “94.84%,

Biswas et al. [35] proposed an algorithm for COVID-19 prediction using ensemble learning from chest radiographs. The researchers employed three widely used DL models, namely VGG-16, ResNet-50, and Xception. They proposed a mechanism to combine these pre-trained models to enhance the prediction capability. The resulting ensemble method gained a remarkable accuracy of 98.79%.

Hossain et al. [36] proposed model for classifying COVID-19 using a fine-tuned ResNet-50. The model was fine-tuned using ten pre-trained weights that were developed on large-scale datasets using multiple methodologies, including supervised and self-supervised learning. The proposed model, named $iNat2021\_Mini\_SwAV\_1k$, was pre-trained on the iNat2021 mini dataset using unsupervised contrastive learning. It outperformed other ResNet-50 TL models with impressive metrics. However, the authors did not explicitly mention crucial factors such as training time, memory consumption, and hardware requirements. These factors are essential for assessing the model's feasibility in real-world clinical settings.

Mercaldo et al. [37] published a report on automated COVID-19 detection using TL on CT medical images aiming for a faster and more automatic diagnosis. Their approach involved distinguishing between healthy patients, those with pulmonary disease, and those with COVID-19, while automatically identifying and highlighting areas of infection in affected patients. They proposed a modified VGG-16 architecture and attained an accuracy of 0.95 in their experiments. The average detection time for their system was approximately 8.9 seconds.

Duong et al. [38] proposed an algorithm for COVID-19 detection from chest X-ray and CT images. They employed TL using two newly developed deep neural network (DNN) architectures, EfficientNet and MixNet. To evaluate the algorithm's efficiency, the authors used a five-fold cross-validation on four real datasets. The experimental findings demonstrated that the suggested model obtained over 95% accuracy in all configurations.

The analysis of literature on COVID-19 detection using DL for radiographic images reveals promising results, indicating that CNNs are reliable and efficient methods for diagnosing COVID-19, even with limited training data. Researchers have utilized various techniques and architectures, such as, TL, self-supervised learning, and scratch learning to develop these models. The literature also highlights the importance of data balancing, data cleaning, and image pre-processing, along with selecting appropriate DL models and hyperparameters when constructing disease diagnosis models. While DL models offer significant potential for COVID-19 detection, further research is needed to ensure their robustness, reliability, and accessibility for deployment in real-world healthcare environments. Lighter-weight CNN architectures play a crucial role in optimizing memory consumption, enabling the deployment of these models on smart devices or cloud platforms. By employing DL models in COVID-19 detection, healthcare professionals gain the advantage of faster and more accurate diagnosis and treatment.

3.Materials and methods

Our objective is to build a DL supported binary classification model to detect COVID-19 using chest radiographs. The aim is to develop an effective DL model that achieves accurate predictions while maintaining a lightweight structure.

Figure 1 demonstrates the graphical representation of the proposed model. The workflow for the COVID detection process primarily consists of the following stages:

1. Selection of an appropriate COVID dataset.
2. Pre-processing of the images to enhance their quality and extract relevant features.
3. Selection of a suitable DL model architecture and customization of the network architecture and hyperparameter settings as necessary.
4. Evaluation of performance metrics to assess the effectiveness of the model.
5. Creation of a user interface to facilitate seamless interaction with the developed DL model.

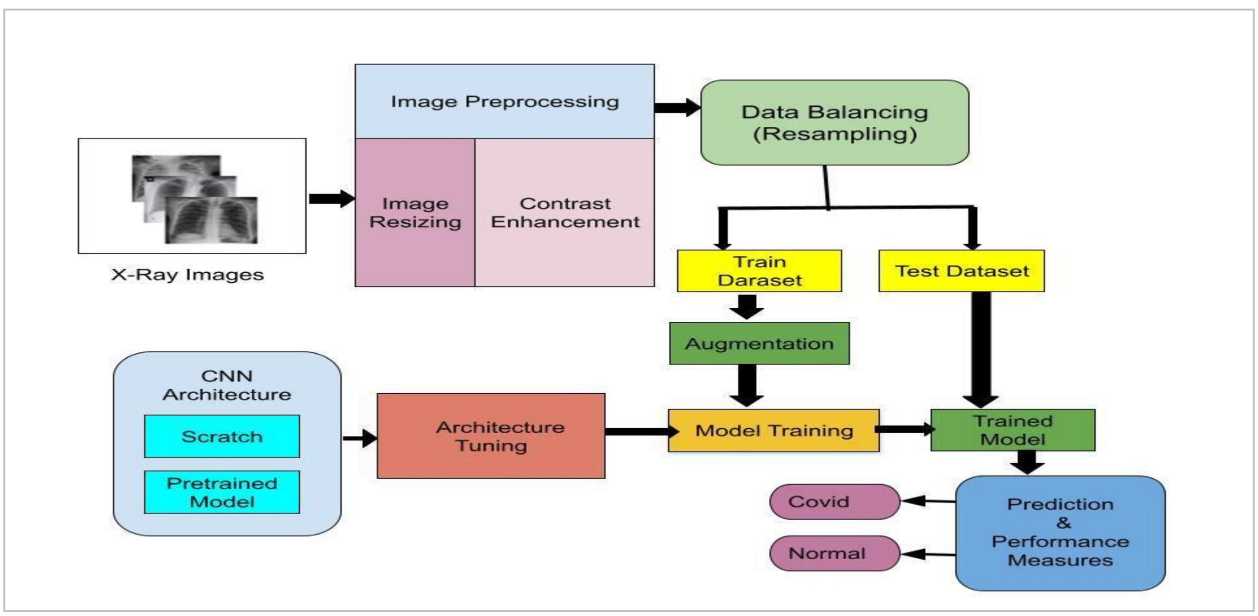
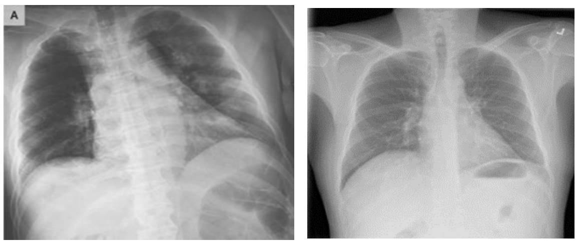


Figure 1 Graphical representation of proposed Covid-19 diagnostic approach

3.1Dataset

Chest X-ray is a cost-effective modality for lung diseases, offering the advantage of low radiation doses for patients [39]. The COVID-19 Radiography Database, obtained from Kaggle [40], is a publicly accessible dataset that consists of images of COVID-19 (3616), lung opacity (6012), healthy (7980), and viral pneumonia (1345). Due to class imbalance and poor contrast of the dataset necessitates image pre-processing to enhance network performance. *Figure 2* demonstrates the contrast variation observed in X-ray images.



COVID image Normal image

Figure 2 Original X-Ray images of COVID-19 and normal patient

3.2Image pre -processing

The dataset consisted of images collected from various sources, exhibiting variations in resolution, size, and light intensity. To address these variations, pre-processing steps were applied to the images, including resizing them to a standardized size and applying HE to enhance contrast. *Figure 3* visually represents the images after undergoing the pre-processing steps, showcasing the improved clarity and contrast achieved through these techniques.

Figure 3 Pre-processed X-Ray images for COVID and normal

The problem of an imbalanced dataset is eliminated with various techniques such as under-sampling or oversampling, especially when one class significantly outweighs the others. Under-sampling involves reducing the instance numbers in the majority class to achieve a balanced class distribution. This is accomplished through random sampling or specific methods like cluster centroids, Tomek links, or NearMiss. By reducing the number of instances in the majority class, the model becomes less biased towards that class, potentially leading to improved classification performance for the minority class. In our study, we opted to utilize random under-sampling as our chosen technique for addressing the class imbalance in our dataset.

Data augmentation was employed to increase the number of training samples while preserving image semantic content. Three transformations, namely rotation, horizontal flipping, and zooming, were applied to the training samples. These transformations help introduce variations in the data, enabling the model to generalize better. *Figure 4* illustrates an example of an augmented image generated through these transformations.

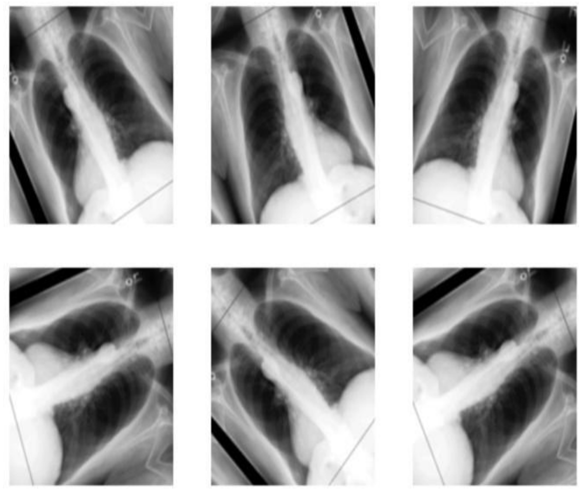


Figure 4 X-ray images after image augmentation

3.3Model selection

DL algorithms exhibit a hierarchical learning architecture that comprises three layers (input layer, hidden layer(s), and an output layer). The application complexity and the level of abstraction necessary to extract task-specific features determine the number of hidden layers to be included in network. DNNs have gained popularity in medical image analysis due to their capability to automatically extract complex and abstract features directly from images without the need for explicit programming. Furthermore, DL models often outperform manual interpretation approaches in terms of accuracy. DL offers various learning approaches, including supervised, semi-supervised, unsupervised, and reinforcement learning. Each approach caters to different learning scenarios and objectives.

CNN, inspired by the visual perception mechanisms of living beings, has emerged as one of the most successful DL algorithms. It possesses the capability to pre-process data and learns diverse image features using filters. CNNs are specifically suitable for image classification tasks owing to their ability to reduce the dimensionality of an image while preserving crucial abstract features before passing them to the classification stage. The architecture of a CNN is organized into layers, as depicted in *Figure 5*. The three essential layers are convolutional, pooling layer, and fully connected layer. Each layer serves a specific purpose in the learning process and contributes to the network's ability to extract meaningful features and make accurate predictions.

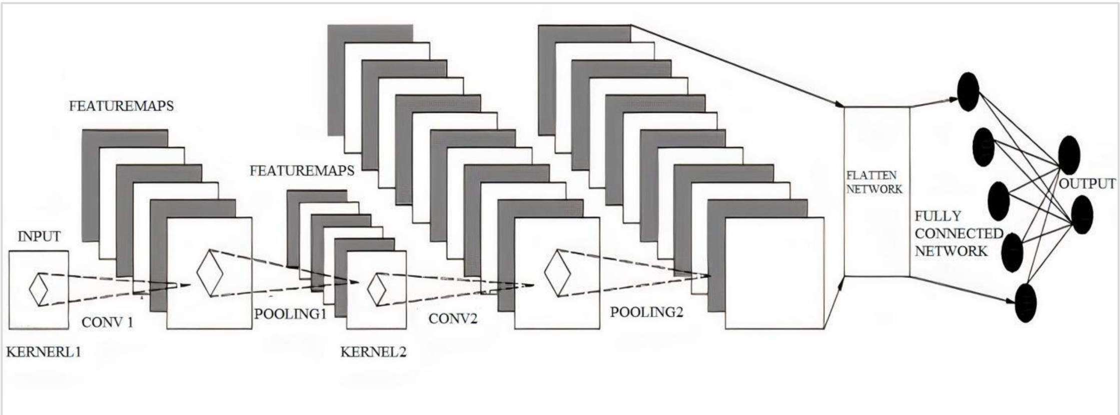


Figure 5 Layered construction of CNN

To determine the optimum CNN architecture, three distinct methods are utilized: CNN from scratch learning, TL, and TL with fine-tuning. However, in any of these methods, selection of hyperparameters plays an important role before training the network.

Hyperparameters are crucial in establishing and modifying the learning process of ML models. They include parameters like the learning rate, batch size, number of hidden layers, number of neurons per layer, activation function, optimizer, regularization parameter, and dropout rate. Tuning these hyperparameters is a crucial step in optimizing ML performance. It entails carefully selecting appropriate hyperparameter values prior to training the model. Typically, this tuning process involves methods such as trial and error or automated techniques like grid search, random search, or Bayesian optimization, aiming to find the optimal combination of hyperparameter values.

3.3.1 CNN from scratch learning

In this setup, a CNN model is constructed from the ground up, comprising convolutional, pooling, and fully connected layers. Subsequently, this model is trained using a dataset, and performance metrics are computed to assess the model's effectiveness.

3.3.2 Transfer learning

This configuration incorporates two TL approaches: TL and TL with fine-tuning. TL uses a previously learned model as a basis for training on a similar but distinct task [41, 42]. The classification layers are replaced and trained on the new dataset. *Figure 6* provides a visual depiction of TL and the fine-tuning process.

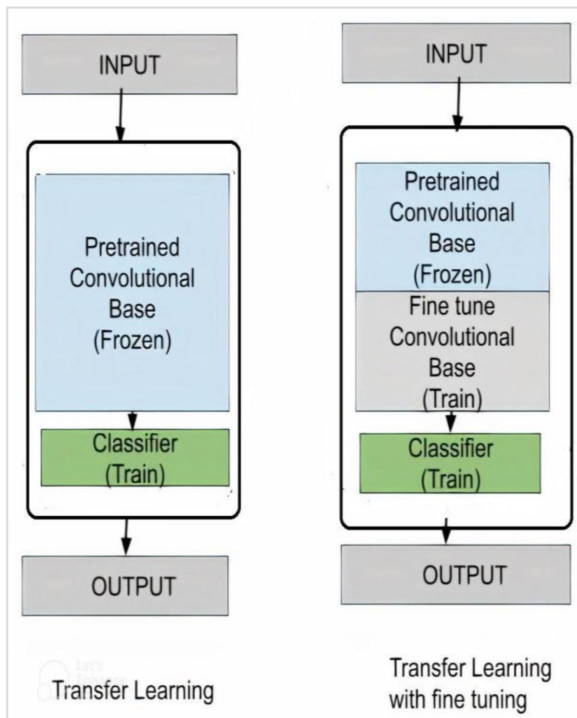


Figure 6 TL and TL with fine tuning approach

The most popular pre-trained CNN architectures are ResNet [43], VGG [44], Inception [45], AlexNet [46], LeNet, and MobileNet [47]. For our experiment, we utilized VGG-16, VGG-19, ResNet50, InceptionV3, and MobileNetV2 for the experiments.

ResNet

A DNN architecture, utilizes skip connections, also known as residual connections, to bypass multiple layers in the network. This technique allows for a smoother flow of information throughout the network, even in cases where the network is extremely deep. In each block of layers, ResNet incorporates a residual connection that adds the output of the previous block directly to the output of the current block. This creates a shortcut and facilitates faster and more efficient training. With the inclusion of skip connections, ResNet effectively addresses issues like vanishing gradients and helps prevent overfitting, which is a common challenge when training DNNs. Notable ResNet structures include ResNet18, ResNet50, and ResNet101.

VGG

It is a CNN architecture that was introduced by Karen and Andrew in 2014. It is made up of several convolutional layers, each followed by a max pooling layer. Rather than using filters of large sizes, VGGNet employs multiple cascaded 3x3 kernel-sized filters, resulting in significant performance gains over the AlexNet network. VGG-16 and VGG-19 are the most used VGGNet variants.

InceptionNet

Inception is a CNN architecture developed by Google researchers in 2014 with the goal of boosting deep network performance while reducing computing complexity. The architecture comprises a module called the Inception module, which replaces the traditional approach of adding more layers to the model to deepen it. The Inception module applies multiple filter sizes in parallel to the input image, and the resulting outputs are concatenated and fed to the next Inception module. The Inception family embodies several variations, including InceptionV1 (GoogleNet), InceptionV2, InceptionV3, InceptionV4, and Inception-ResNet, among others.

MobileNet

It is a CNN architecture that was introduced by Google researchers in 2017, aimed at achieving efficient computation and optimization for mobile devices, making it a suitable choice for applications with limited resources. The MobileNet design includes depth wise separable convolutions, which divide the standard convolution operation into depthwise and pointwise convolutions, resulting in a substantial reduction in the number of parameters and computations needed while maintaining performance. MobileNetV1 and MobileNetV2 are the primary variations of the architecture, with MobileNetV2 featuring improvements such as residual connections and linear bottlenecks. It is a lightweight architecture that offers a practical solution for deploying DL models on smart portable devices.

3.3.3 TL with fine tuning

In a DL model, the lower layer is responsible for capturing general features, while the higher layer focuses on extracting more complex and task-specific features [48]. To improve the efficiency of pretrained networks, some layers of the convolutional base are often frozen while others are trainable. Additionally, the classification layer is customized based on the specific number of classes required for the task at hand. In this research, our emphasis was on fine-tuning VGG-16 and MobileNetV2, as they demonstrated superior performance compared to other pretrained architectures for the classification of COVID-19. To improve the effectiveness of the VGG-16 network, modifications were made to its classification layer, the number of trainable layers was adjusted, and additional layers were introduced. This approach aimed to enable the network to learn more specific features that were relevant to our classification task. The modified VGG-16 architecture, shown in *Figure 7*, illustrates the customization of the classification layers, the inclusion of two extra layers, and the preservation of the remaining layers by keeping them frozen. To fine-tune the MobileNetV2 model, we made changes to its classification layer and added additional trainable layers. The modified structure of MobileNetV2, illustrated in *Figure 8*, demonstrates these adjustments.

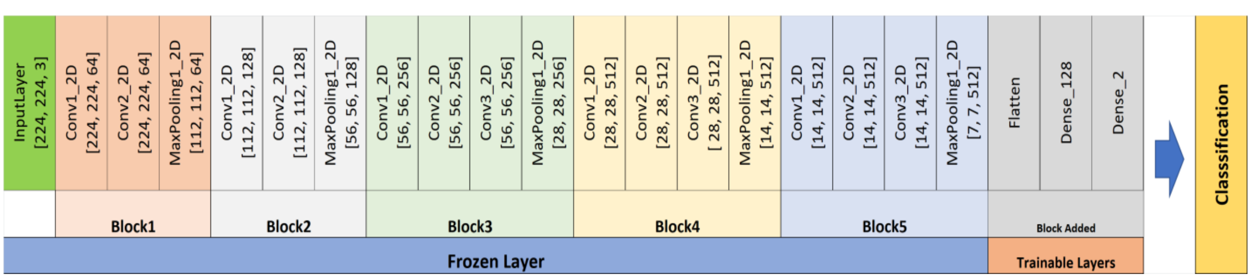


Figure 7 Proposed VGG-16 fine tune architecture

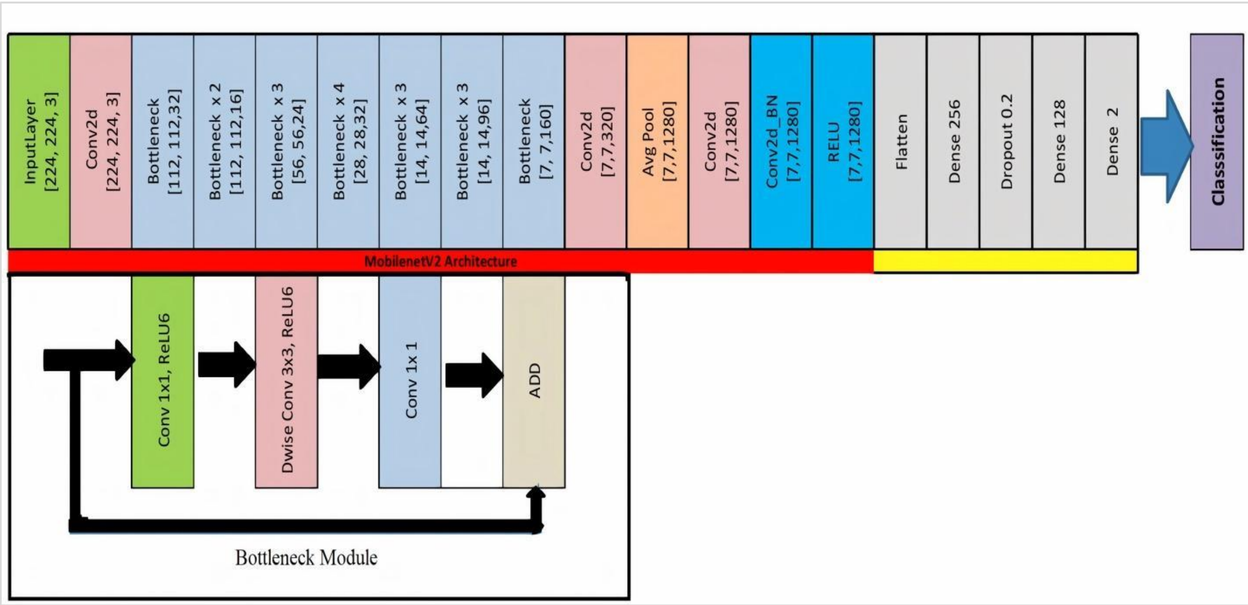


Figure 8 Proposed MobileNetV2 fine tune architecture

3.4 Evaluation of performance metrics

The predicted model's performance is analysed using standard performance measures including accuracy, precision, recall (sensitivity), and F1 score. The accuracy-loss curves of training and validation, confusion matrices, and AUC curves are obtained. The training and validation accuracy curves identify if the model leads to overfitting or underfitting. Overfitting occurs when the model excessively fits the training data, while underfitting happens when the model fails to capture the patterns from the training set. Both situations tend to decrease the data generalization ability of a model.

Number of true positive (TP), true negative (TN), FP, and FN predicts the classifier's performance in terms of confusion matrix. Some of the performance measures are formulated as shown below.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

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

$$Recall = \frac{TP + TN}{TP + FN}$$

$$F1\ score = 2 \times \frac{Prpecision \times Recall}{Prepcision + Recall}$$

3.5 Graphical user interface

The proposed model incorporates a user interface developed using Python Gradio. This interface allows users to interact with the ML model through a web browser, the deployment of the model and facilitating its sharing with others. This enables users to remotely access the model to perform X-ray image binary classification.

4. Results and discussion

For these experiments, the proposed model was developed using the TensorFlow/Keras Python framework on an Intel Xeon Gold 5118 processor with a clock speed of 2.3 GHz. Several pre-trained CNN networks, namely ResNet50, VGG-16, VGG-19, and MobileNetV2, were employed, alongside a custom CNN network with modified layers. The evaluation of model involved the use of standard performance metrics, including a normalized confusion matrix that visually represents the predicted and true labels for each class.

To examine the impact of number of layers on the efficiency of the CNN architecture, a series of experiments were carried out. CNN architectures were specifically created with 4 layers, 5 layers, and 6 layers from scratch. These architectures were then trained using identical hyperparameters, which consisted of 25 epochs, batch size 6, and RMSprop optimizer with a 0.1 learning rate.

According to experimental results presented in *Table 1*, it is evident that the 5-layer CNN architecture outperformed both the 4-layer and 6-layer architectures in terms of accuracy on the dataset. This finding suggests that increasing the depth of the CNN architecture beyond a certain point does not necessarily result in higher accuracy. Additionally, other factors such as the total images in the dataset, epochs, batch size, and optimizer greatly influence the accuracy of the model. To validate our findings, the model was tested on both balanced and unbalanced datasets. The balanced dataset with 2500 COVID-19 and 2500 normal images and unbalanced datasets comprising 3500 COVID-19 and 1500 normal images are supplied to the network. The experiments were conducted using a 5-layer CNN architecture built from scratch with the same set of hyperparameters. *Table 2* shows the results of these tests.

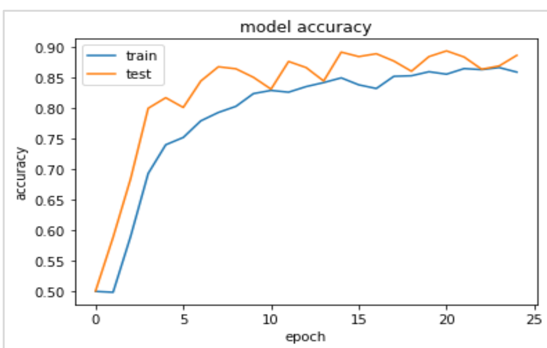
Table 1 Comparison performance measures of number of layers of CNN

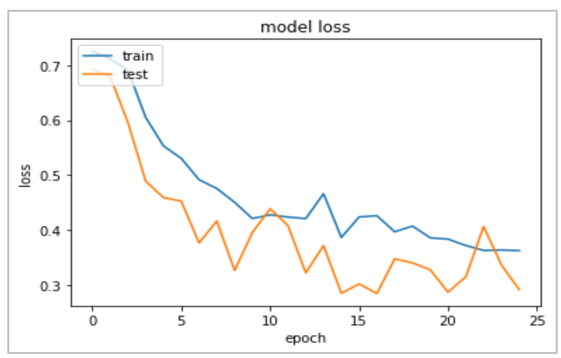
CNN scratch	4 Layers	5 Layers	6 Layers
Training Accuracy	0.73	0.77	0.79
Validation Accuracy	0.70	0.78	0.82
Precision	0.83	0.89	0.86
Recall	0.76	0.79	0.77
F1 Score	0.78	0.82	0.80
AUC	0.74	0.82	0.78

Table 2 Performance measurement for balanced and unbalanced dataset

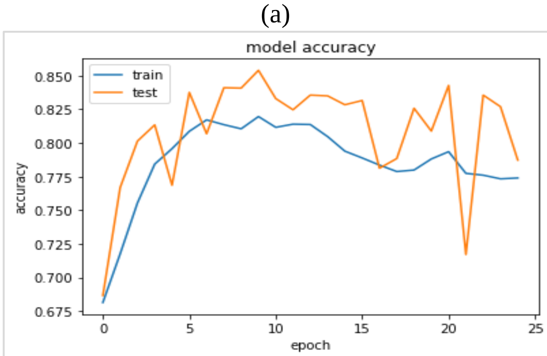
Scratch CNN	Unbalanced dataset	Balanced dataset
Training Accuracy	0.73	0.77
Validation Accuracy	0.70	0.78
Precision	0.83	0.89
Recall	0.76	0.79
F1 Score	0.78	0.82
AUC	0.74	0.82

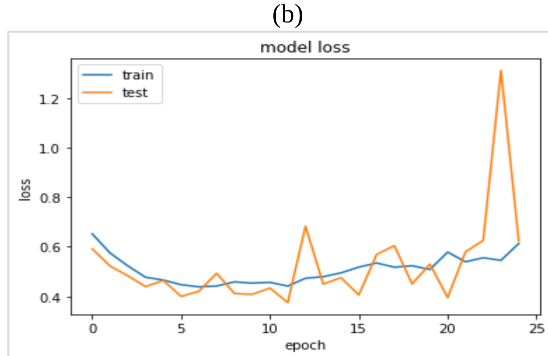
Table 2 highlights performance of the model concerning achieved accuracy during the training-validation stages, as well as higher score of parameters (precision, recall, F1 score, and AUC) with balanced dataset. *Figure 9* displays the accuracy-loss curves obtained with 5-layer CNN architecture trained using RMSprop optimizer (Epochs=25) balanced and unbalanced datasets.





(a) (b)





(c) (d)

Figure 9 Accuracy-loss curve for CNN-5 layers Model (a, b) balanced dataset (c, d) unbalanced dataset

The experimental findings from our study suggest that the balanced dataset demonstrates a less pronounced disparity between training and testing accuracy/loss. This observation serves as an indication that the model exhibits improved generalization on the validation set. The equal distribution of both classes in the balanced dataset facilitates equal learning opportunities for the model across all categories, leading to enhanced overall performance. Consequently, we underscore the importance of dataset balance in the pursuit of achieving optimal accuracy for our classification problem. As part of this study, the COVID-19 radiography database was divided into two distinct sets: a training and a test or validation set. Random assignment was performed for each image, resulting in the following proportions: 80% for training and 20% for testing, 70% for training and 30% for validation, and 60% for training and 40% for validation. Analysis of our experimental findings that the most favourable outcomes were achieved when utilizing a 70% training and a 30% validation set.

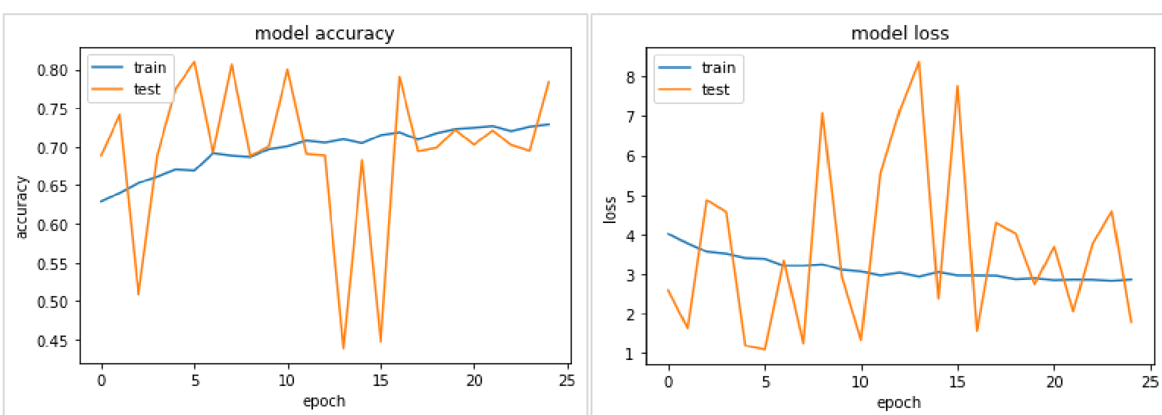
In this study, a TL approach was employed, specifically retraining only the classification layer while keeping the remaining layers fixed. The primary aim was to compare the performance of various pre-trained networks, VGG-16, VGG-19, ResNet50, InceptionV3, and MobileNetV2, in the binary classification of COVID-19. The hyperparameters utilized for all pre-trained architectures are found in *Table 3*.

Table 3 Hyper-parameters of all pretrained architectures

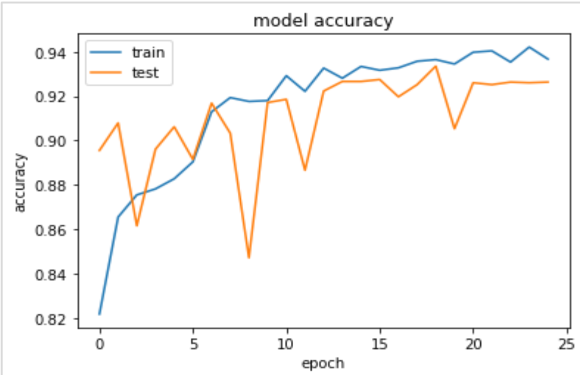
Hyper parameters	Values
Pretrained weights	Imagenet
Bath size	6
No of epochs	25
Optimizer	Adam
Learning rate	0.001
Train –Test spilt	70 -30
Weight decay	None

The accuracy-loss curves depicted in *Figure 10* provide valuable insights into the training performance of each architecture. Based on our experimental findings, ResNet-50 and VGG-19 exhibited slower convergence rates, suggesting that a longer training duration may be necessary to achieve optimal performance. InceptionV3, on the other hand, displayed a noticeable gap in its accuracy-loss curve, indicating potential issues with overfitting or underfitting. In contrast, VGG-16 and MobileNetV2 exhibited smaller gaps in their training and validation accuracy/loss curves, indicating more stable and consistent performance. Notably, MobileNetV2 demonstrated stable results after 15 epochs, suggesting its potential as an efficient architecture for the specific task at hand. These observations underscore the importance of selecting appropriate pre-trained architectures with favourable convergence properties to ensure effective model training and performance.

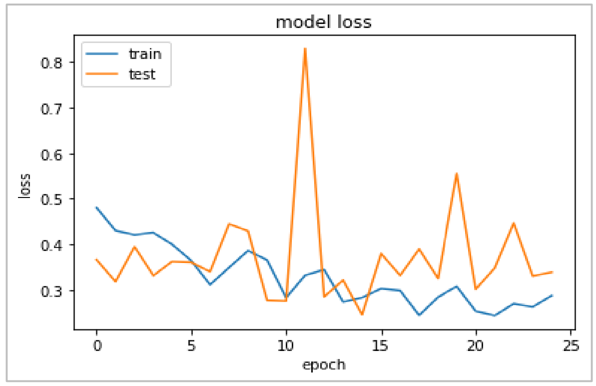
Table 4 presents the approaches employed in the study along with the achieved performance measures. Based on these metrics, VGG-16 and MobileNetV2 exhibited higher accuracy and overall performance compared to the other approaches for COVID-19 classification. It is crucial to highlight, however, that there is still potential for improvement in terms of closing the gap between the training and validation curves. Further experimentation or fine-tuning of the model could be explored to enhance its generalization capability and address this aspect.



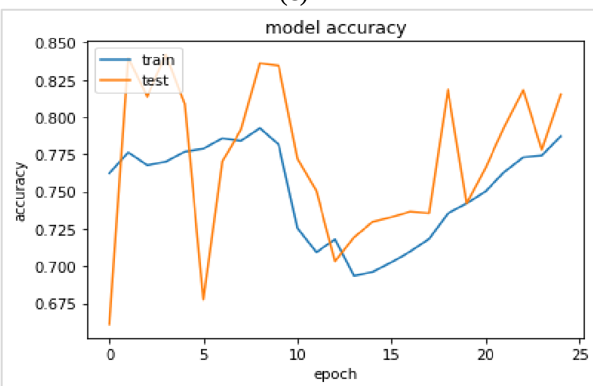
(a) (b)



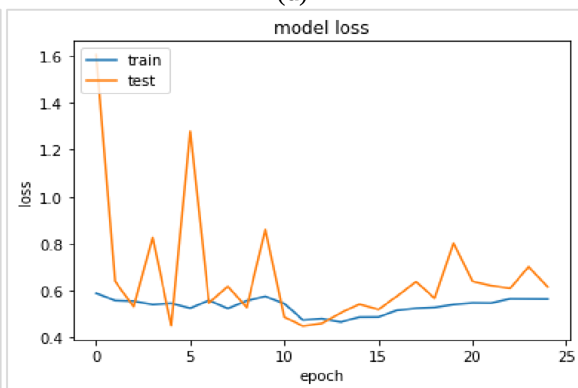
(c)



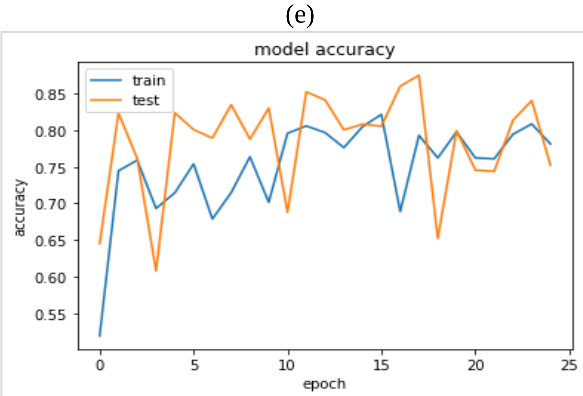
(d)



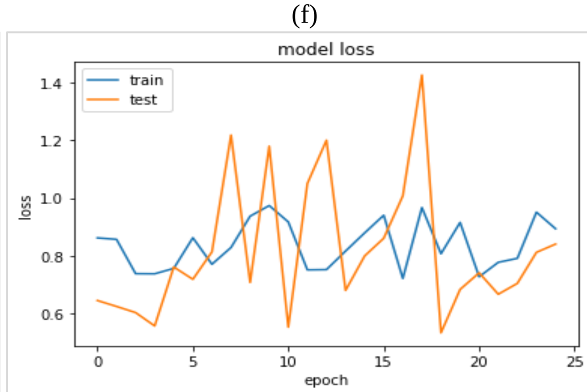
(e)



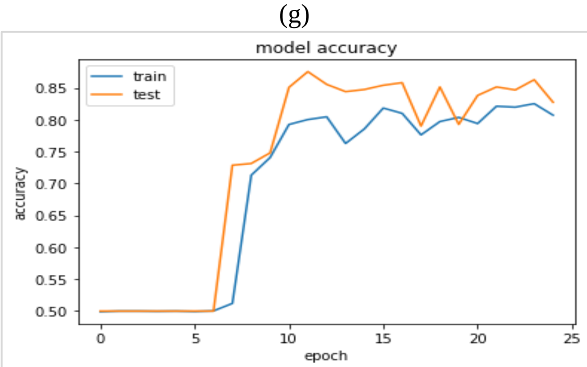
(f)



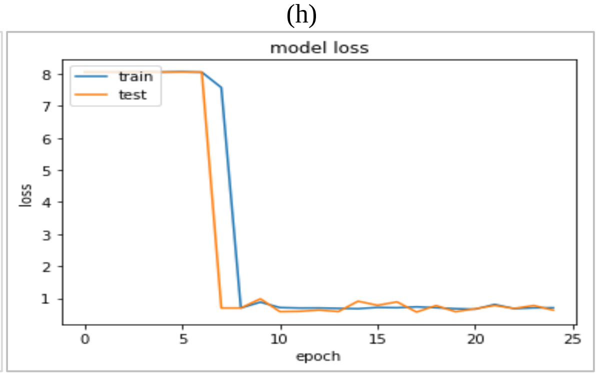
(g)



(h)



(i)



(j)

Figure 10 Accuracy-loss curve of various pretrained architectures (a, b) ResNet-50 (c, d) VGG-16 (e, f) VGG-19 (g, h) InceptionV3 (i, j) MobileNetV2

To optimize the VGG-16 network, the original classification layer was replaced with a customized classification layer designed for the specific classification objective. The first three layers, responsible for detecting low-level features, were fixed, while layers 4 and 5 were made trainable to capture high-level features relevant to the task. However, the accuracy and loss curves obtained from fine-tuning VGG-16, as shown in *Figure 11*, revealed overfitting to the training data and a subsequent decline in accuracy. In order to mitigate this issue, various techniques such as dropout regularization and early stopping were applied, but they did not yield improvements in accuracy. It was hypothesized that the complex architecture of VGG-16 and more trainable parameters, coupled with the limited size of the dataset, contributed to the observed overfitting problem.

Table 4 Performance metrics of various pretrained architectures

Approaches Used	Training Accuracy	Validation Accuracy	Precision	Recall	F1 Score	AUC
Resnet-50	0.72	0.78	0.92	0.78	0.82	0.85
InceptionV3	0.78	0.76	0.89	0.86	0.87	0.83
VGG-19	0.78	0.81	0.88	0.77	0.80	0.78
VGG-16	0.93	0.91	0.92	0.91	092	0.89
MobileNetV2	0.80	0.82	0.87	0.83	0.83	0.85

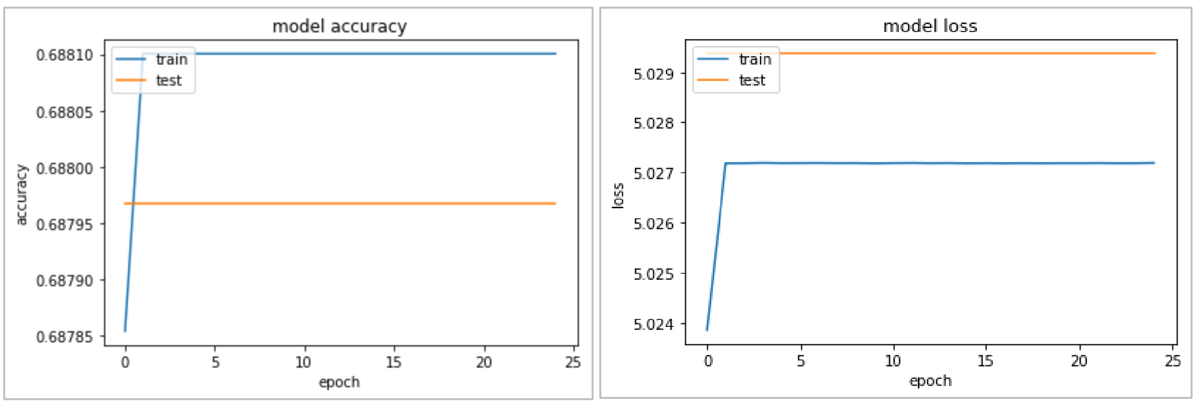


Figure 11 Accuracy-loss curve of VGG-16 tuning model

To enhance the accuracy of the VGG-16 architecture, the experiment involved making modifications to the pretrained network. Specifically, the classification layer was customized, two additional layers were added, and the remaining layers were kept frozen. The proposed VGG-16 architecture is presented in *Figure 7*. These modifications aimed to enable the network to learn more specific and relevant features for the classification task, while avoiding issues of overfitting and underfitting. The results of the modified VGG-16 model showed enhanced accuracy compared to the pre-trained VGG-16 model. The training accuracy achieved a level of 94%, indicating effective learning from the training data. Similarly, the validation accuracy reached 93%, demonstrating that the model's performance generalized well to testing data. *Figure 12* visually illustrates the accuracy achieved during the training process, with a consistent and steady increase in both training and validation accuracy. This signifies that the modified VGG-16 model has effectively learned from the dataset, leading to improved performance.

To enhance the accuracy of the MobileNetV2 model, we conducted fine-tuning by modifying its classification layer and introducing additional trainable layers. MobileNetV2 was configured to have initial layers non trainable and last 24 trainable layers. The MobileNetV2 architecture proposed in this study is depicted in *Figure 8*. A loss versus learning rate curve was plotted for the proposed MobileNetV2 architecture to determine the optimal optimizer and learning rate for the fine-tuning process. *Figure 13* illustrates the model's potential for the loss and different learning rates to guide in choosing the selection of most suitable optimizer and learning rate. With ' RMSprop ' optimizer and learning rate value between 0.0001 to 0.001 yielded the best performance. By fine-tuning the MobileNetV2 model in this manner, the model was adapted specifically to the classification task, resulting in improved results.

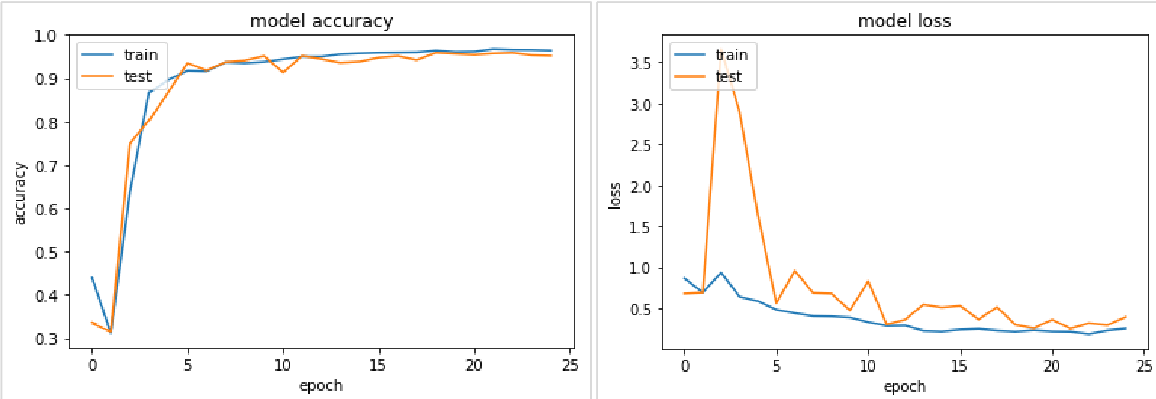
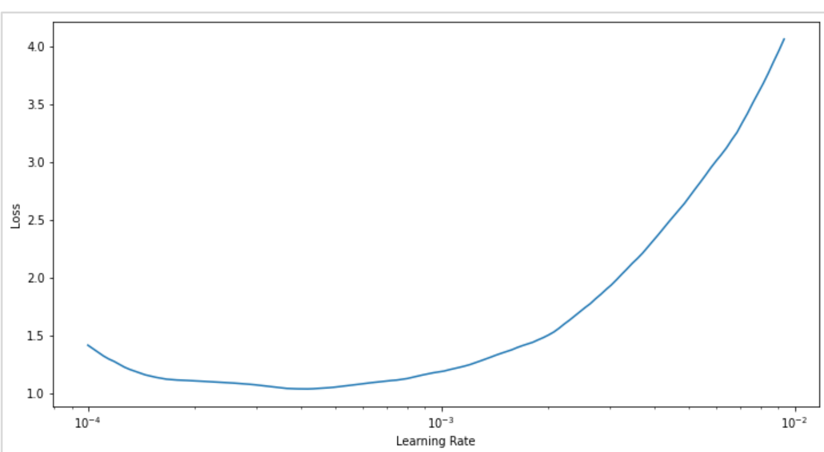


Figure 12 Accuracy-loss curves Modified VGG-16

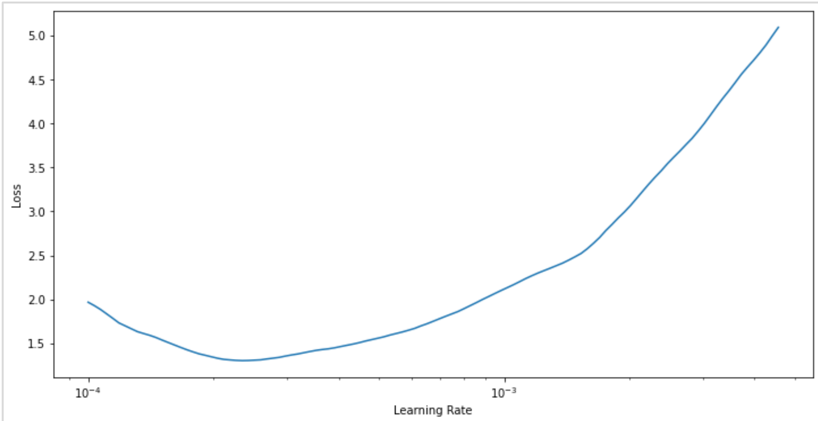
Figure 14 shows the perfectly matched accuracy-loss curves obtained with tuned MobileNetV2 achieving training-validation accuracy of 96%. *Table 5* summarizes the performance metrics for the fine-tuned VGG-16 and MobileNetV2 architectures. Both models demonstrated excellent performance in the binary classification of COVID-19. The VGG-16 model attained 94% and 93% validation accuracy, respectively, while the MobileNetV2 achieved even higher accuracy, with 96% training and validation accuracy. These results underscore the efficacy of fine-tuning in adapting pre-trained models to the specific classification task, resulting in improved accuracy and overall performance. *Figure 15* presents the confusion matrix for all pre-trained architectures, depicting the model's ability to accurately classify COVID-19 and healthy images with only a few misclassifications. The remarkable accuracy and minimal misclassification rate of the model demonstrate its potential for assisting in the diagnosis of COVID-19 from chest radiography. This study demonstrates the effectiveness of TL with a fine-tuning approach in developing accurate models for COVID-19 detection from chest X-rays. The tuned MobileNetV2 achieved a validation accuracy of 96%, along with high precision, recall, F1 score, and AUC.

Figure 16 compares the performance of VGG-16 and MobileNetV2 models in terms of accuracy, training time, and memory usage. These results suggested that the MobileNetV2 model is a more efficient choice for COVID-19 X-ray image classification tasks, as it achieved higher accuracy with less training time and memory usage.

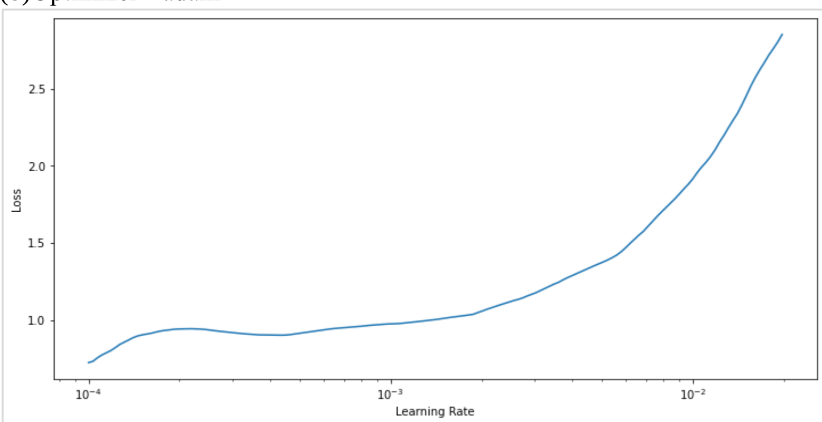
A user interface was developed using the Python Gradio framework to facilitate interaction with the trained model. This interface enables users to upload an X-ray image for prediction by accessing a web link. The results of the prediction are then displayed, as shown in *Figure 17*.



(a)Optimizer='sgd'



(b)Optimizer='adam'



(c) optimizer=' RMSprop'

Figure 13 Learning rate versus loss curve of different optimizer for proposed MobileNetV2 architecture

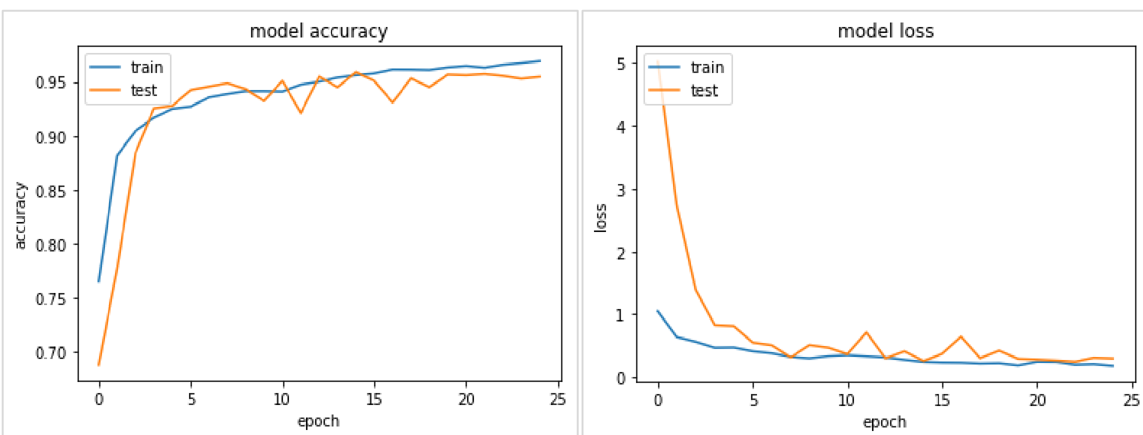


Figure 14 Accuracy-loss curves of proposed tuned MobileNetV2

Table 5 Performance metrics of modified VGG-16 and MobileNetV2 fine-tuned pretrained architecture

Approaches Used	Training accuracy	Validation accuracy	Precision	Recall	F1 Score	AUC
VGG-16 (Fine Tuning)	0.94	0.93	0.93	0.92	0.91	0.91
MobilentV2 (Fine Tuning)	0.96	0.96	0.96	0.96	0.96	0.96

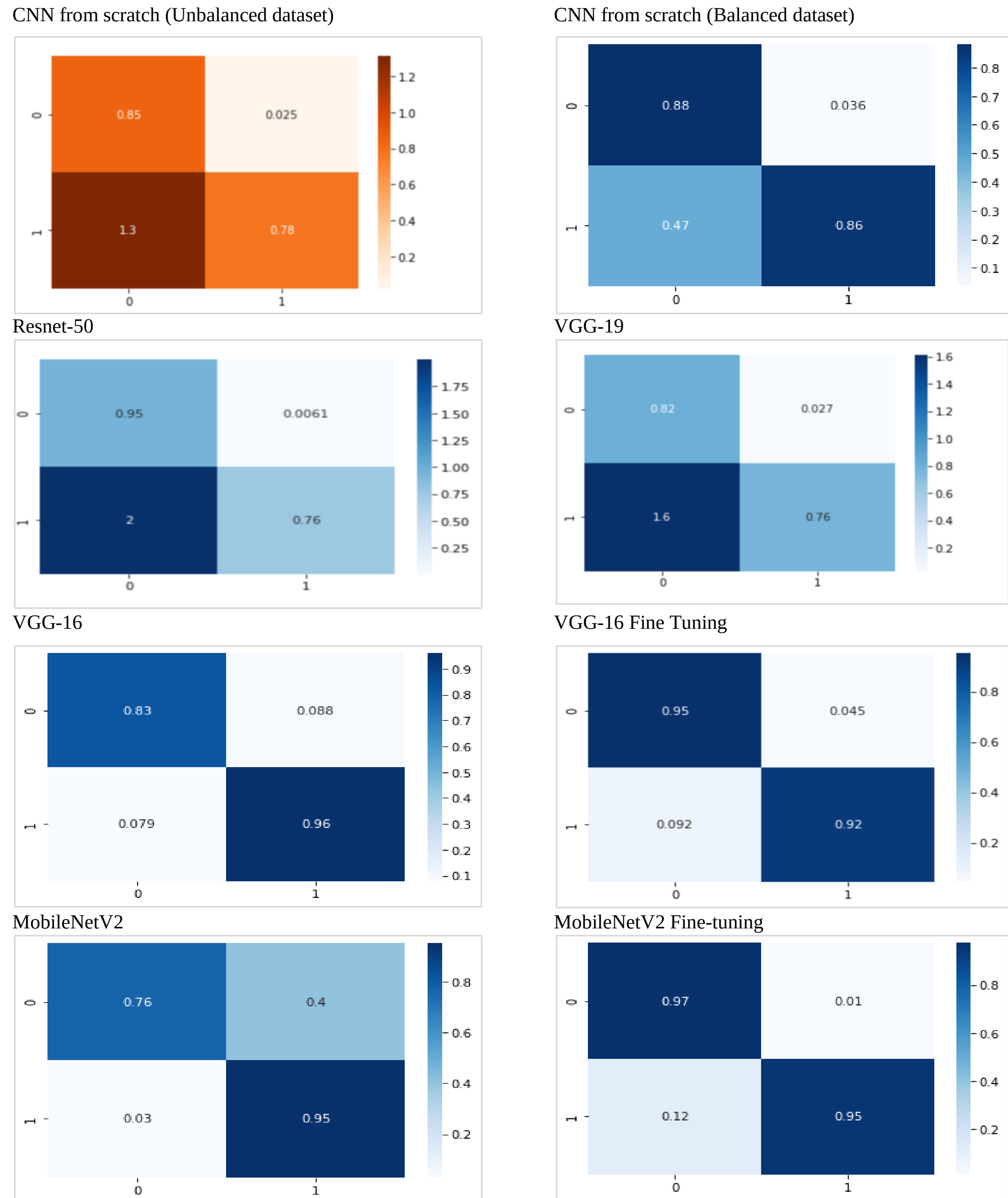


Figure 15 Confusion matrix

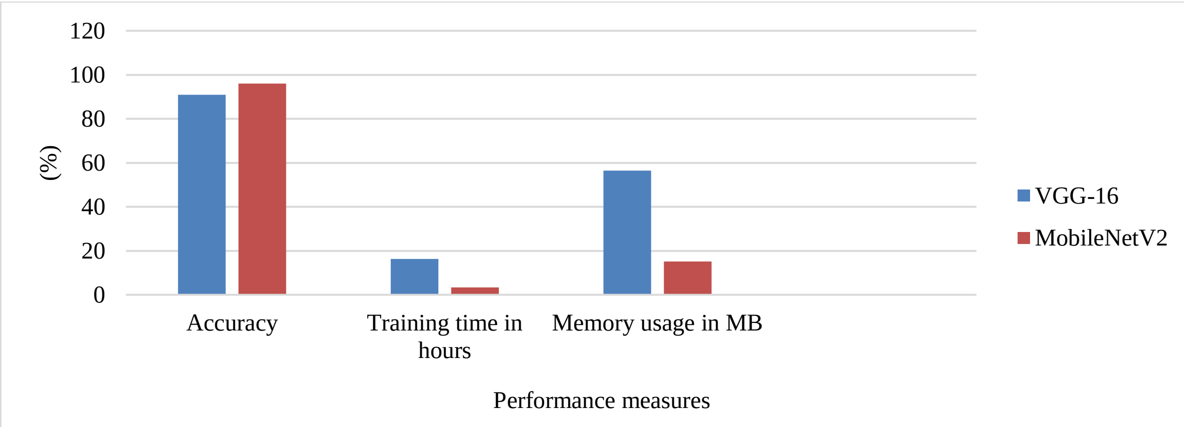


Figure 16 Comparisons of VGG-16 and MobilenetV2

Figure 17 User interface for COVID classification

4.1Result comparison with similar work

Recently, numerous research studies have been conducted to classify COVID-19. In *Table 6,* we compare our approach with previous work that used different DL models. A complete list of abbreviations is shown in *Appendix I.*

Table 6 Comparison of proposed model with similar work

References	DL model	Dataset used	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
[49]	EfficientnetB0 tuned	COVID_19 radiography	95	90	97	
[50]	COVIDX-Net	IEEE Covid chest X-ray dataset	90	91	90	90
[51]	Deep CNN from scratch	X-ray and metadata collected from Wuhan	93	--	--	--
[52]	COVIDGAN for CNN	Combined three datasets COVID-19 radiography, IEEE Covid chest X-ray, COVID-19 chest X-ray initiative	95	95	95	95
[53]	CovidSORT-Ensemble model built on the majority voting approach	Publicly available chest X-rays	96	--	--	--
[54]	ResNet101	COVID-QU dataset -three class	96	96	96	96
Proposed	MobileNetV2 Tuned	COVID-19 radiography	96	96	96	96

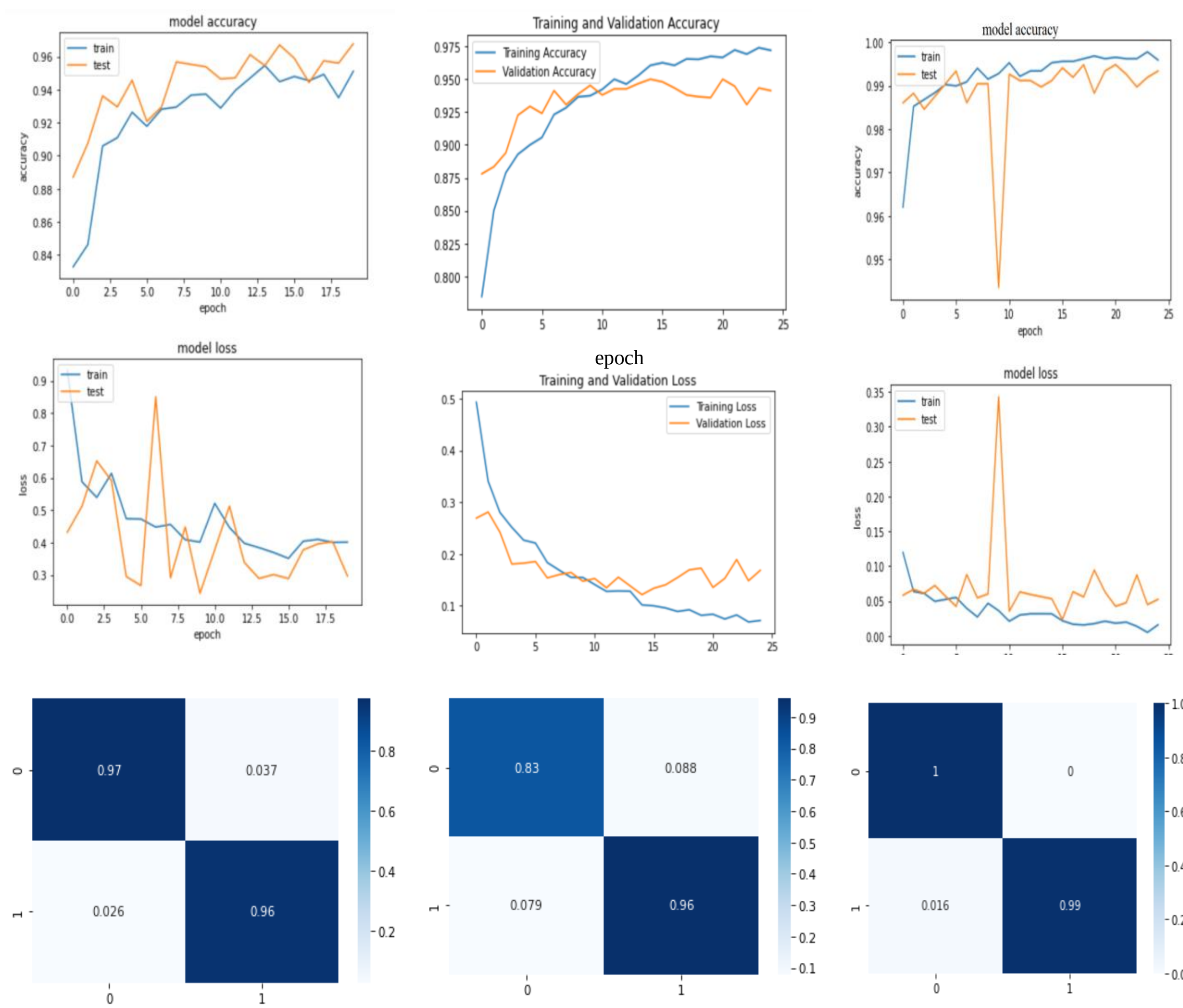
This suggests that the proposed MobileNetV2 model beats other models in terms of performance measures. The evaluation is based on a specific dataset, and further validation on larger datasets is necessary to confirm the findings. To validate the credibility of the proposed work, tests were conducted on various radiography datasets as presented in *Table 7*. *Figure 18* displays the training and validation accuracy-loss curves as well as the confusion matrix for these datasets. Based on this evaluation, it can be concluded that the model performed satisfactorily for all datasets.

4.2Limitation of the work

The obtained model performance relies various factors like size of dataset, complexity, and quality. The evaluation is conducted on a specific dataset, and additional validation on larger datasets is required to validate the results. The study solely relies on chest radiographs in detecting COVID-19. The inclusion of primary clinical investigations and travel history can enhance the accuracy of the achieved findings further. Therefore, any AI system developed using this dataset should be used alongside other diagnostic tools and clinical assessments performed by healthcare professionals.

Table 7 Performance measures on various dataset

Reference and dataset	Number of COVID-19 images	Number of normal images	Validation accuracy	Precision	Recall	F1 Score	AUC
[55] Curated	1281	3270	0.96	0.96	0.96	0.96	0.96
[56] COVID-19 Xray-CT images	4044	5500	0.92	0.92	0.92	0.91	0.90
[57] IEEE Covid chest X-ray	299	349	0.99	0.99	0.99	0.99	0.99



[52] Curated dataset [53] X-ray-CT images dataset [54] Covid-19 dataset

Figure 18 Accuracy-loss curve and confusion matrix of various datasets

5. Conclusion and future scope

The present study aimed to conduct a comprehensive comparison between CNN-based transfer learning (TL) and fine-tuning approaches for the detection of COVID-19 using X-ray images. Among the evaluated models, MobileNetV2 exhibited outstanding performance, achieving an impressive accuracy of 96%. Notably, this approach demonstrated efficient resource utilization, with a training time of 3.37 hours and a memory utilization of only 15.2 MB. When compared to other pre-trained models, MobileNetV2 with fine-tuning outperformed them in terms of accuracy, memory utilization, training time, and AUC. Its lightweight and accurate architecture rendered it highly suitable for deployment on web and remote service platforms. This study underscored the importance of dataset balance and the careful selection of hyperparameters to achieve optimal model performance. The findings of this research contributed to the development of rapid and cost-effective techniques for the automatic detection of COVID-19 using chest radiographs. The resultant model, along with its web-based system, holds great potential for providing valuable support to the medical community in the early detection of COVID-19.

To further enhance the model's performance and contribute to a more accurate and reliable diagnosis of lung disorders, future research directions should encompass exploring noise removal techniques, incorporating metadata, feature selection, and ensemble learning. These approaches have the potential to improve the model's efficiency, as well as enhance the accuracy and reliability of the diagnosis. Furthermore, the proposed model shows promise for detecting various lung-related disorders in the future.

Acknowledgment

We are thankful to the facilitators at Department of Information Technology, Vishwakarma Government Engineering College, Chandkheda, Ahmadabad, to carry out the experiment work under their supervision.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contribution statements

Mamta Patel: Conceptualization, data acquisition, simulations, interpretation of results, original draft writing, review and editing. **Mehul Shah:** Conceptualization, supervision, review and editing, analysis and review of the result.

References

[1] WHO. Coronavirus overview, prevention and symptoms. https://www.who.int/emergencies/diseases/novel-coronavirus-2019. Accessed 20 November 2020.

[2] Okolo GI, Katsigiannis S, Althobaiti T, Ramzan N. On the use of deep learning for imaging-based COVID-19 detection using chest X-rays. Sensors. 2021; 21(17):5702.

[3] Sharma A, Tiwari S, Deb MK, Marty JL. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): a global pandemic and treatment strategies. International Journal of Antimicrobial Agents. 2020; 56(2):1-13.

[4] Abbasi-oshaghi E, Mirzaei F, Farahani F, Khodadadi I, Tayebinia H. Diagnosis and treatment of coronavirus disease 2019 (COVID-19): laboratory, PCR, and chest CT imaging findings. International Journal of Surgery. 2020; 79:143-53.

[5] Wu YC, Chen CS, Chan YJ. The outbreak of COVID-19: an overview. Journal of the Chinese Medical Association. 2020; 83(3):217-20.

[6] Himoto Y, Sakata A, Kirita M, Hiroi T, Kobayashi KI, Kubo K, et al. Diagnostic performance of chest CT to differentiate COVID-19 pneumonia in non-high-epidemic area in Japan. Japanese Journal of Radiology. 2020; 38:400-6.

[7] Woloshin S, Patel N, Kesselheim AS. False negative tests for SARS-CoV-2 infection—challenges and implications. New England Journal of Medicine. 2020; 383(6):1-3.

[8] Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, et al. The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner society. Radiology. 2020; 296(1):172-80.

[9] Moghadam SO. A review on an ongoing pandemic caused by the severe acute respiratory syndrome coronavirus 2: the pathogenesis, epidemiology, immunological features, and currently available diagnostic tests. Reviews in Medical Microbiology. 2022; 33(1):e212-23.

[10] Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, et al. Can AI help in screening viral and COVID-19 pneumonia? IEEE Access. 2020; 8:132665-76.

[11] Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review. 2020; 53:5455-516.

[12] Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine. 2020; 43:635-40.

[13] Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In artificial intelligence in healthcare 2020 (pp. 25-60). Academic Press.

[14] Alzubaidi L, Zhang J, Humaidi AJ, Al-dujaili A, Duan Y, Al-shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data. 2021; 8:1-74.

[15] Alafif T, Tehame AM, Bajaba S, Barnawi A, Zia S. Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions. International Journal of Environmental Research and Public Health. 2021; 18(3):1117.

[16] Nayak SR, Nayak DR, Sinha U, Arora V, Pachori RB. Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: a comprehensive study. Biomedical Signal Processing and Control. 2021; 64:102365.

[17] Gatto A, Accarino G, Aloisi V, Immorlano F, Donato F, Aloisio G. Limits of compartmental models and new opportunities for machine learning: a case study to forecast the second wave of COVID-19 hospitalizations in Lombardy, Italy. Informatics. 2021; 8(3): 57.

[18] Paul SG, Saha A, Biswas AA, Zulfiker MS, Arefin MS, Rahman MM, et al. Combating Covid-19 using machine learning and deep learning: Applications, challenges, and future perspectives. Array. 2022:100271.

[19] Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. Annals of Translational Medicine. 2020; 8(11):1-15

[20] Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. NPJ Digital Medicine. 2021; 4(1):1-9.

[21] Kim M, Yun J, Cho Y, Shin K, Jang R, Bae HJ, et al. Deep learning in medical imaging. Neurospine. 2019; 16(4):657-68.

[22] Arias-londoño JD, Gomez-garcia JA, Moro-velazquez L, Godino-llorente JI. Artificial intelligence applied to chest X-ray images for the automatic detection of COVID-19. A thoughtful evaluation approach. IEEE Access. 2020; 8:226811-27.

[23] Yang D, Martinez C, Visuña L, Khandhar H, Bhatt C, Carretero J. Detection and analysis of COVID-19 in medical images using deep learning techniques. Scientific Reports. 2021; 11(1):1-13.

[24] Toraman S, Alakus TB, Turkoglu I. Convolutional capsnet: a novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. Chaos, Solitons & Fractals. 2020; 140:1-11.

[25] He X, Yang X, Zhang S, Zhao J, Zhang Y, Xing E, et al. Sample-efficient deep learning for COVID-19 diagnosis based on CT scans. IEEE Transactions on Medical Imaging. 2020:1-10.

[26] Kassania SH, Kassanib PH, Wesolowskic MJ, Schneidera KA, Detersa R. Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: a machine learning based approach. Biocybernetics and Biomedical Engineering. 2021; 41(3):867-79.

[27] Basha SS, Dubey SR, Pulabaigari V, Mukherjee S. Impact of fully connected layers on performance of convolutional neural networks for image classification. Neurocomputing. 2020; 378:112-9.

[28] Haque KF, Haque FF, Gandy L, Abdelgawad A. Automatic detection of COVID-19 from chest X-ray images with convolutional neural networks. In international conference on computing, electronics & communications engineering 2020 (pp. 125-30). IEEE.

[29] Wang L, Lin ZQ, Wong A. Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Scientific Reports. 2020; 10(1):1-2.

[30] Diaz-escobar J, Ordóñez-guillén NE, Villarreal-reyes S, Galaviz-mosqueda A, Kober V, Rivera-rodriguez R, et al. Deep-learning based detection of COVID-19 using lung ultrasound imagery. Plos one. 2021; 16(8): e0255886.

[31] Wang N, Liu H, Xu C. Deep learning for the detection of COVID-19 using transfer learning and model integration. In 10th international conference on electronics information and emergency communication 2020 (pp. 281-4). IEEE.

[32] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In proceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 2097-106).

[33] Kaur P, Harnal S, Tiwari R, Alharithi FS, Almulihi AH, Noya ID, et al. A hybrid convolutional neural network model for diagnosis of COVID-19 using chest X-ray images. International Journal of Environmental Research and Public Health. 2021; 18(22):1-17.

[34] Bahuguna A, Yadav D, Senapati A, Nath Saha B. k NN-SVM with deep features for COVID-19 pneumonia detection from chest X-ray. In international conference on mathematics and computing 2022 (pp. 103-15). Singapore: Springer Nature Singapore.

[35] Biswas S, Chatterjee S, Majee A, Sen S, Schwenker F, Sarkar R. Prediction of covid-19 from chest CT images using an ensemble of deep learning models. Applied Sciences. 2021; 11(15):1-16.

[36] Hossain MB, Iqbal SH, Islam MM, Akhtar MN, Sarker IH. Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images. Informatics in Medicine Unlocked. 2022; 30:1-10.

[37] Mercaldo F, Belfiore MP, Reginelli A, Brunese L, Santone A. Coronavirus covid-19 detection by means of explainable deep learning. Scientific Reports. 2023; 13(1):462.

[38] Duong LT, Nguyen PT, Iovino L, Flammini M. Automatic detection of Covid-19 from chest X-ray and lung computed tomography images using deep neural networks and transfer learning. Applied Soft Computing. 2023; 132:109851.

[39] https://www.sciencedirect.com/topics/computer-science/imaging-modality. Accessed 10 May 2023.

[40] https://www.kaggle.com/tawsifurrahman/covid19-radiography-database. Accessed 10 May 2023.
[41] Yamashita R, Nishio M, Do RK, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights into Imaging. 2018; 9:611-29.
[42] Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Computer Science. 2021; 2(6):1-20.
[43] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-8).
[44] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
[45] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In proceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 1-9).
[46] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Communications of the ACM. 2017; 60(6):84-90.
[47] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. 2017.
[48] Choudhary K, Decost B, Chen C, Jain A, Tavazza F, Cohn R, et al. Recent advances and applications of deep learning methods in materials science. NPJ Computational Materials. 2022; 8(1):59.
[49] Nikolaou V, Massaro S, Fakhimi M, Stergioulas L, Garn W. COVID-19 diagnosis from chest x-rays: developing a simple, fast, and accurate neural network. Health Information Science and Systems. 2021; 9:1-11.
[50] Hemdan EE, Shouman MA, Karar ME. Covidx-net: a framework of deep learning classifiers to diagnose covid-19 in x-ray images. Computer Vision and Pattern Recognition. 2020:1-14.
[51] Medhi K, Jamil M, Hussain MI. Automatic detection of COVID-19 infection from chest X-ray using deep learning. Medrxiv. 2020: 2020-05.
[52] Waheed A, Goyal M, Gupta D, Khanna A, Al-turjman F, Pinheiro PR. Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection. IEEE Access. 2020; 8:91916-23.
[53] Tammina S. Covidsort: detection of novel covid-19 in chest x-ray images by leveraging deep transfer learning models. In ICDSMLA: proceedings of the 2nd international conference on data science, machine learning and applications 2022 (pp. 431-47). Springer Singapore.
[54] Constantinou M, Exarchos T, Vrahatis AG, Vlamos P. COVID-19 classification on chest X-ray images using deep learning methods. International Journal of Environmental Research and Public Health. 2023; 20(3):2035.
[55] Sait U, Lal KG, Prajapati S, Bhaumik R, Kumar T, Sanjana S, et al. Curated dataset for COVID-19 posterior-anterior chest radiography images (X-Rays). Mendeley Data. 2020;1.
[56] https://data.mendeley.com/datasets/8h65ywd2jr. Accessed 10 May 2023.
[57] Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988. 2020.

Mamta Patel is a Ph.D. research scholar in Instrumentation and Control Engineering at Gujarat Technological University. She holds a Master's degree in Applied Instrumentation from L.D. College of Engineering in Ahmedabad. Currently, Mamta serves as an Assistant Professor at Vishwakarma Government Engineering College in Ahmedabad, India.
Email: mamtavp@vgecg.ac.in

Dr. Mehul Shah is an Associate Professor in the Department of Instrumentation and Control Engineering at Vishwakarma Government Engineering College in Chandkheda, Ahmedabad, India. With over 25 years of teaching experience, he has gained extensive knowledge in the field. He earned his Ph.D. in Instrumentation and Control Engineering from Gujarat Technological University.
Email: mehul.shah@vgecg.ac.in

Appendix I

S. No.	Abbreviation	Description
1	AUC	Area under the curve
2	CAD	Computer-Aided Diagnosis
3	CT	Computed Tomography
4	CLAHE	Contrast Limited Adaptive Histogram Equalization
5	CapsNet	Capsule Network
6	CNN	Convolutional Neural Network
7	DCNN	Deep Convolutional Neural Network
8	DL	Deep Learning
9	DNN	Deep Neural Network
10	DST	Department of Science and Technology
11	FN	False Negative
12	FP	False Positive
13	HE	Histogram Equalization
14	KNN	K-Nearest Neighbor
15	ML	Machine Learning
16	ResNet	Residual Network
17	RF	Random Forest
18	TN	True Negative
19	TP	True Positive
20	TL	Transfer Learning
21	VGG	Visual Geometry Group
22	WHO	World Health Organization