

On the Detection of COVID-19 from Chest X-Ray Images Using CNN-Based Transfer Learning

Mohammad Shorfuzzaman^{1,*} and Mehedi Masud¹

Abstract: Coronavirus disease (COVID-19) is an extremely infectious disease and possibly causes acute respiratory distress or in severe cases may lead to death. There has already been some research in dealing with coronavirus using machine learning algorithms, but few have presented a truly comprehensive view. In this research, we show how convolutional neural network (CNN) can be useful to detect COVID-19 using chest X-ray images. We leverage the CNN-based pre-trained models as feature extractors to substantiate transfer learning and add our own classifier in detecting COVID-19. In this regard, we evaluate performance of five different pre-trained models with fine-tuning the weights from some of the top layers. We also develop an ensemble model where the predictions from all chosen pre-trained models are combined to generate a single output. The models are evaluated through 5-fold cross validation using two publicly available data repositories containing healthy and infected (both COVID-19 and other pneumonia) chest X-ray images. We also leverage two different visualization techniques to observe how efficiently the models extract important features related to the detection of COVID-19 patients. The models show high degree of accuracy, precision, and sensitivity. We believe that the models will aid medical professionals with improved and faster patient screening and pave a way to further COVID-19 research.

Keywords: Deep learning, artificial neural network, COVID-19, transfer learning.

1 Introduction

The novel coronavirus disease 2019 (COVID-19) started in the Wuhan city, China, in December 2019 [Zhu, Zhang, Wang et al. (2020); Li, Guan, Wu et al. (2020)]. The common symptoms observed at that time were fever, cough, myalgia, or fatigue after clinical diagnosis of 41 patients and all the patients had pneumonia [Huang, Wang, Li et al. (2020)]. The complications were acute heart injury, severe respiratory distress syndrome, and other infections. Thirteen COVID-19 patients were hospitalized for intensive care unit (ICU) and six died. A research team Chan et al. [Chan, Yuan, Kok et al. (2020)] at the University of Hong Kong first time discovered the evidence of human-to-human transmission of COVID-19.

¹ College of Computers and Information Technology, Taif University, Taif, 21974, Saudi Arabia.

* Corresponding Author: Mohammad Shorfuzzaman. Email: m.shorf@tu.edu.sa.

Received: 01 May 2020; Accepted: 28 May 2020.

Now, COVID-19 has turned into a pandemic and has affected almost 209 countries and territories around the world by April 08, 2020 [Worldometer (2020)]. Fig. 1 shows the spread of COVID-19 around the world as of April 20, 2020. On 30th January 2020, World Health Organization (WHO) has declared it as a Public Health Emergency of International Concern (PHEIC) [ICTV (2020)]. Unknown treatment, shortage of facilities, and stringent conditions of laboratory environment were seriously delaying precise diagnosis of suspected patients. This has created major challenges to stop the spread of the infection. Correct and fast COVID-19 diagnosis of suspected patients at the premature stage may play an important role in timely quarantine and treatment. Fast detection of COVID-19 is very important for patients' prognosis, epidemic control, and the security of public health.

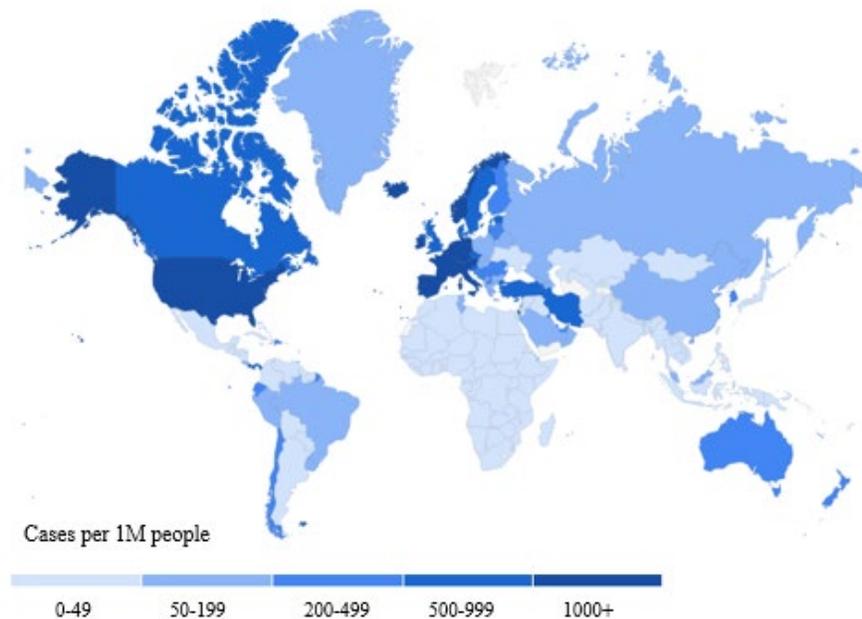


Figure 1: Coronavirus disease (COVID-19) map [Google N. (2020)] as of April 20, 2020

Detection of nucleic acid is generally considered to identify the COVID-19 infection. Now, faster detection kits for coronavirus are available. Nevertheless, chest X-rays and computed tomography (CT) scans remain largely as an efficient modality to detect and evaluate the pneumonia severity of patients [Chen, Zhou, Dong et al. (2020)]. In the initial stage of the COVID-19 infection, thinner layer scanning is usually required as an alternative of traditional CT scan for diagnosis. This process is very time-consuming, and the radiologists become burdened that lingers early isolation and diagnosis of patients, delays patients' treatment and prognosis, and eventually, causes problem to control COVID-19 epidemic.

COVID-19 epidemic is still uncontrolled in most of the countries and number of infected cases and deaths are rising every day. Researchers are also focusing on the learning-based mechanisms for detecting COVID-19 infection. Chest X-rays and CT scans are the adopted method of imaging to diagnose COVID-19 pneumonia. This approach can be cost-effective

and will possibly take less time to perform the test. Machine learning and computer vision technology are now largely considered in the field of medical diagnosis. Machine learning models can identify early signs of a disease by examining medical images that doctors would overlook. However, the models generally require a huge data to learn from.

Deep learning is a significant development in the domain of artificial intelligence during the last decade. It has great potential to extract minute features in image analysis [Liu, Liu and Zhang (2019)]. Hence, deep learning models are mostly considered for object recognition, image segmentation and classification, image change detection and in various areas of computer vision and image processing [Wang, Li, Zou et al. (2020); Wang, Jiang, Luo et al. (2019)]. Deep learning architecture such as CNN has extensive applications in medical image processing. A substantial amount of research has been conducted using deep learning for pulmonary nodules diagnosis, benign and malignant tumors classification, and disease prediction.

Modern deep learning techniques based on CNN are also being applied for automatic image analysis in medical fields to facilitate clinical treatment [Celik, Talo and Yildirim (2020); Saba, Mohamed, El-Affendi et al. (2020); Yildirim, Talo, Ay et al. (2019)]. Large amount of data is often required to train these deep learning models with longer training period. However, one of the major hindrances faced by the researchers in medical data analysis is the limited availability of datasets. Transfer learning has been proven to be very useful in such scenario [Raghu, Zhang, Kleinberg et al. (2019); Ravishankar, Sudhakar, Venkataramani et al. (2016); Yosinski, Clune, Bengio et al. (2014)]. With transfer learning, knowledge learned from a deep learning model trained on a very large dataset is utilized in the target model to solve a similar problem. In other words, choose a model trained on a large dataset and apply its knowledge to a model with smaller dataset in a new domain.

In the context of computer vision and deep learning, off-the-shelf pre-trained models constitute the core of transfer learning. These pre-trained models are based on large CNNs with significant depth and can learn hierarchical feature representations automatically. A typical CNN consists of a convolutional base created from a stack of convolutional and pooling layers and a classifier comprising of fully connected layers. The idea is that lower level convolutional layers compute mostly general high-level features that are similar across different images and can be reused. On the other hand, some of the upper level layers compute features that are specific to the chosen dataset. In this paper, we used a suit of pre-trained models built to solve computer vision problems which are trained on data from the ImageNet challenge [Canziani, Paszke and Culurciello (2016); Deng, Dong, Socher et al. (2009)]. There are two common strategies (as shown in Fig. 2) to use these pre-trained models for image classification problems: i) using the pre-trained models as feature extractors without fine-tuning, and ii) using the pre-trained models after fine-tuning. We have adopted the second approach where we decided to fine-tune the models by re-training one third of the layers towards the upper level rather than re-train the models entirely. Consequently, we obtain fine-tuned pre-trained models with improved performance and reduced training time. We have used five different fine-tuned pre-trained models and analyzed their performance. Thereafter, we employed an ensemble model where the predictions from all chosen pre-trained models are combined to generate a single output.

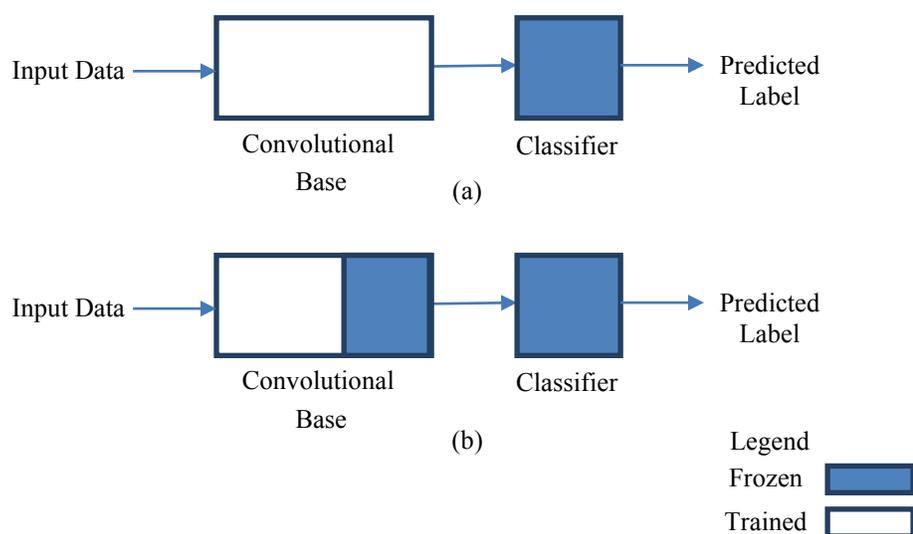


Figure 2: Strategies to use pre-trained models: (a) Freeze the convolutional base, (b) Train some upper-level layers and leave others frozen

To evaluate the performance of the models used in this study, we have prepared an image dataset by collecting chest X-ray images of COVID-19 positive cases in addition to healthy and non-COVID pneumonia cases from two publicly available data sources [Cohen (2020); Kermany, Goldbaum, Valentim et al. (2018)]. Thus, the studied models are expected to effectively identify COVID-19 patients from other non-COVID pneumonia patients and healthy population. Moreover, we investigated by using two different visualization techniques that our models can extract important features (also called imaging biomarkers) related to coronavirus infection that will help health practitioners gain profound understanding associated to the detection of COVID-19 patients.

The rest of the paper is organized as follows. ‘Related work’ section reviews the state-of-the-art techniques used in COVID-19 detection. ‘Materials and methods’ provide detailed description of our model, its configuration, dataset used, and performance evaluation metrics. ‘Results and discussion’ section presents the performance results obtained by the proposed models, qualitative analysis through visualization techniques, and state-of-the-art comparison. Finally, the paper concludes with ‘Conclusions’ and outlines some potential future work.

2 Related work

After the outbreak of COVID-19, a handful of related research has been conducted. We can categorize the areas of COVID-19 research as: (i) recommend antiviral drugs; (ii) forecast rates of infected patients and patient prognosis for facilitating healthcare service providers to prepare for better healthcare actions and find resources; (iii) develop COVID-19 diagnostic tools to predict coronavirus from chest X-rays or CT images; (iv) mining social media data to better understand the coronavirus spread pattern in the areas and countries. In the following, we discuss some of the research progress. Deep learning techniques are being applied by researchers to model molecules or molecular interactions. In particular,

they are trying to identify the features that might possibly enable them to discover the vaccines or an operative antiviral. Zhang et al. [Zhang, Saravanan and Yang (2020)] made an effort to come up with an antiviral from other current antiviral using a deep learning model with a hope to benefit patients with coronavirus infection. They have used DenseNet transfer learning model to anticipate contacts of protein-ligand. They used coronavirus's RNA sequences in addition to chemical compounds to recommend the best possible drug. Bo et al. [Bo, Bonggun, Yoonjung et al. (2020)] proposed a model that works based on calculating affinity values between COVID-19 target proteins and commercially available antiviral drugs to identify the drug that works the best.

There has been some research to develop methods to model the coronavirus spread. Majority of the techniques work primarily based on shallow approaches for predicting spread of disease and mortality of patients. Hence, machine learning and deep learning-based models can play a significant role to obtain better results. Yan et al. [Yan, Zhang, Xiao et al. (2020)] presented a machine model which works based XG-Boost technique to predict the probability of survival for a COVID-19 infected patient considering some vital factors such as patient's age. This aids health provides to decide whether the patient needs intensive care or not. In a subsequent effort, Fong et al. [Fong, Li, Deyet et al. (2020)] developed a machine learning model to predict early outbreak of COVID-19 leveraging various different techniques such as data augmentation, model ensemble, and parameter fine-tuning to attain the maximum possible accuracy. Anastassopoulou et al. [Anastassopoulou, Lucia, Athanasios et al. (2020)] presented a model to forecast coronavirus spread in Hubei, China, by estimating the key epidemiological parameters using the publicly available data for a period of one month. Ran [Ran (2020)] proposed a model that can predict the spread of the virus with high accuracy for a short period such as one day. The model used Kalman filter algorithm that can demonstrate a future trend in the spread of the virus.

Chakraborty [Chakraborty (2020)] proposed a deep learning approach to detect COVID-19 using chest X-ray images. The reason is that chest X-ray images are more accessible than obtaining CT scans particularly in remote and rural areas. The model performance was significant even with the small dataset. However, the model is neither generalized nor fine-tuned. We believe that better results can be obtained with fine tuning the model and generalizing the model with more data. In an earlier open source research effort, COVID-Net [Wang and Wong (2020)] was proposed which is based on a CNN-based deep learning model for COVID-19 detection using chest X-ray images. However, the model uses a relatively fewer number (less than 100) of COVID-19 images for training and testing the model's performance in addition to about 16,000 images of normal and pneumonia categories. Furthermore, Apostolopoulos et al. [Apostolopoulos and Mpesiana (2020)] presented a subsequent study on COVID-19 detection using transfer learning with chest X-ray images from open source data repositories.

CT scan-based diagnosis of pneumonia cases related to coronavirus can potentially reduce diagnosis time and facilitate improved treatment. Radiologists become burdened quickly with large numbers of patients with ICUs. Deep learning techniques using CT scans can easily reduce the burden. There has been some research conducted on CT scan-based images.

Chen et al. [Chen, Wu, Zhang et al. (2020)] considered a UNet++ for extracting relevant

features from CT scans for classification. The model is trained on 40 k scans from 106 patients. The model performed well with respect to precision and recall. Specifically, the model shows 100% sensitivity, 93.55% specificity, 95.24% accuracy, 84.62% PPV, and 100% NPV on a patient level analysis. Using the model, radiologists' reading time greatly reduced by 65%. However, the model needs to be tested with more cases.

Wang et al. [Wang, Kang, Ma et al. (2020)] developed a deep learning technique to produce COVID-19's graphical features from COVID-19 radiographical changes in CT images for providing a medical diagnosis before the pathogenic test, hence saving crucial time for controlling disease. Authors applied their model on 1065 CT images in which there are 325 images pathogen-confirmed COVID-19 cases with other 740 images of viral pneumonia, previously diagnosed.

Xu et al. [Xu, Jiang, Ma et al. (2020)] proposed a similar approach as in Wang et al. [Wang, Kang, Ma et al. (2020)]. However, they used a relatively larger number of patients. The samples were collected from 509 patients that include 175 healthy ones from three hospitals in China. The precision and recall values obtained were considerably lower than in Xu et al. [Xu, Jiang, Ma et al. (2020)].

Jiang et al. [Jiang, Coffee, Bari et al. (2020)] proposed an algorithmic solution to identify COVID-19 clinical characteristics and developed an AI capability tool to predict patients' severe outcomes at the initial stage including the possibility of developing acute respiratory distress syndrome.

At present, the WWW can be employed to monitor diseases using search engines and social networking sites. Tracking the trends of different diseases can be possible between seven to ten days faster than the government agencies [Alessa and Faezipour (2018)]. Social media contains huge information about symptoms and spread of the disease. This can help administration to take actions based on geographical locations such as areas and cities. Despite this, little research has been reported so far that works based on social media information.

3 Materials and methods

3.1 Pre-trained CNN models

We have used the following CNN-based pre-trained predictive models to detect COVID-19 cases from chest X-ray images: (a) VGG16 [Simonyan and Zisserman (2014)]; (b) ResNet50V2 [He, Zhang, Ren et al. (2016)]; (c) Xception [Chollet (2017)]; (d) MobileNet [Howard, Zhu, Chen et al. (2017)]; (e) DenseNet121 [Huang, Liu and Weinberger (2016)]. VGG16 architecture was introduced by researchers from Oxford University which uses 3×3 convolutional layers in increasing depth and max pooling layers to reduce volume size. It has two fully connected layers consisting each layer of 4096 nodes and achieves 92.7% test classification accuracy on the ImageNet dataset posed by ILSVRC-2012 classification challenge which contains more than 14 million images from over 20,000 categories. It contains 16 weight layers achieving performance improvement over its pioneer AlexNet [Krizhevsky, Sutskever and Hinton (2017)] by using multiple smaller sized filters. The ResNet50V2 model is based on the concept of deep residual learning which effectively addresses the vanishing gradient problem in deeper networks which dictates that with increasing depth (in traditional sequential network such as VGG), accuracy gets saturated

and drops abruptly. In residual learning, the network learns an identity function represented by a skip connection which allows the network to pass the input through the residual block devoid of passing it through some other weight layers. Thus, ResNet architecture solves the problem of degrading accuracy and allows the training of extremely deep network with standard SGD optimizer and skipping through less relevant layers by means of residual modules. The ResNet50V2 has 50 weight layers and shows significant reduction in model size and FLOPs compared to its plain counterparts (such as VGG16 and VGG19). The Xception model is an extreme version of Inception which extends the concept of Inception module by separately dealing with each output channel through the mapping of spatial correlations. This is followed by another step of capturing inter-channel correlation by performing 1×1 convolutions. MobileNet provides a simplified version of Xception focusing on a compressed structure while preserving the model accuracy. Finally, DenseNet121 represents the densely connected convolutional networks which need lesser parameters than a comparable customary CNN. In contrast to ResNets, layers in DenseNets are incredibly constricted and apply merely 12 filters with a small number of new feature-maps. DenseNet121 improves training time by directly obtaining gradient values from the loss function and gaining direct access to input image. This substantially lowers computation cost and makes this a better alternative.

In this study, we used the above pre-trained models with weights from their convolutional layers as feature extractors to substantiate transfer learning in detecting COVID-19 cases from chest X-ray images. We removed the predicting layers (i.e., classifier) of the pre-trained CNNs and added our own classifier consisting of a global average pooling layer (GAP) and two dense layers containing 256 and 3 neurons, respectively. Global average pooling (GAP) layer is considered right after the last block of convolution as a better replacement of flattening to reduce overfitting by minimizing the size of model parameters. The GAP layer reduces spatial dimensions of a 3-dimensional tensor having size $h \times w \times d$ to $1 \times 1 \times d$ tensor by simply taking the average of all hw pixel values of each $h \times w$ feature map to single number [Lin, Chen and Yan (2013)]. Fig. 3 shows the block diagram of the modified pre-trained models with our new classifier.

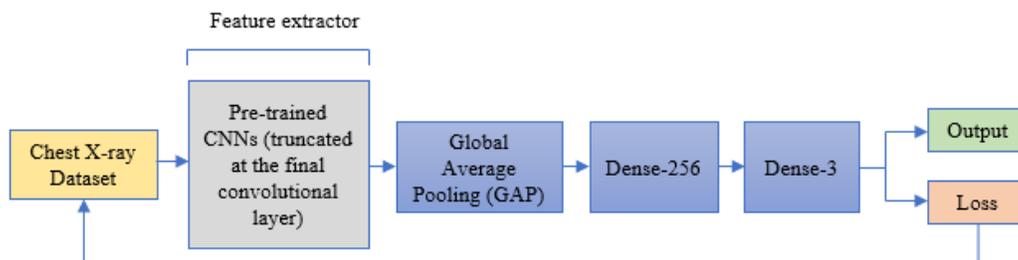


Figure 3: Modified architecture of pre-trained models

We used the above pre-trained models by fine-tuning the models in part to minimize the categorical cross-entropic loss without abruptly changing the pre-trained weights with an appropriate selection (details will follow in the section of model configurations) of learning rate and optimizer. Since lower level convolutional layers in the networks learn common features, we decided to freeze their weights in the pre-trained models during the

training. On the other hand, high level layers compute features that are specific to the task at hand. Thus, we decided to fine-tune the models by re-training one third of the layers towards the upper level rather than re-train the models entirely. Consequently, we obtain fine-tuned pre-trained models with improved performance and reduced training time.

3.1.1 Building the model ensemble

Ensemble of models represents the process of combining predictions from various machine learning and deep learning models to produce a final predictive output. This guides to the opening of diversified prospect of representational capability of the models. Model ensembles are incredibly popular improvement to customary machine learning models such as forming random forests from decision trees. Unlike machine learning models, deep learning models require longer training time and hence constructing an ensemble of models by training deep learning models from scratch is impractical. We build the ensemble of models (as shown in Fig. 4) by adopting a simple technique where the predictions from various pre-trained CNN models are combined to generate a single prediction vector and a final output is obtained by majority voting. This works by summing up the output probabilities from *softmax* layer of each pre-trained model in the ensemble and the final output corresponds to the maximum value from the resultant vector. The best performing model ensemble is then selected from the set of all model combinations.

3.2 Dataset generation and computation resources

In this study, we have used chest X-ray images to detect COVID-19 cases since previous studies have shown great success in diagnosing infectious and other diseases such as pneumonia, malaria, lung cancer, breast cancer and so on using X-ray images. Moreover, it is justified to use X-ray images to detect COVID-19 cases by analyzing patients' lungs since the novel coronavirus infects the epithelial cells in the lining of the respiratory tract. Currently, due to the unavailability of an appropriate COVID-19 radiography image datasets, we have created an image dataset to be used in this study by collecting chest X-ray images of COVID-19 positive and non-COVID patients from multiple open access data sources. COVID-19 X-ray images were collected from a publicly available GitHub repository created by Cohen [Cohen (2020)]. This data also contains X-ray and CT images of patients suffering from other diseases such as ARDS (acute respiratory distress syndrome), MERS (Middle East respiratory syndrome), SARS (severe acute respiratory syndrome) and pneumonia. To facilitate our study, we also collected non-COVID chest X-ray images from the Guangzhou Women and Children's Medical Center [Kermany, Goldbaum, Valentim et al. (2018)] which contains images of normal, pneumonia bacterial and pneumonia viral categories.

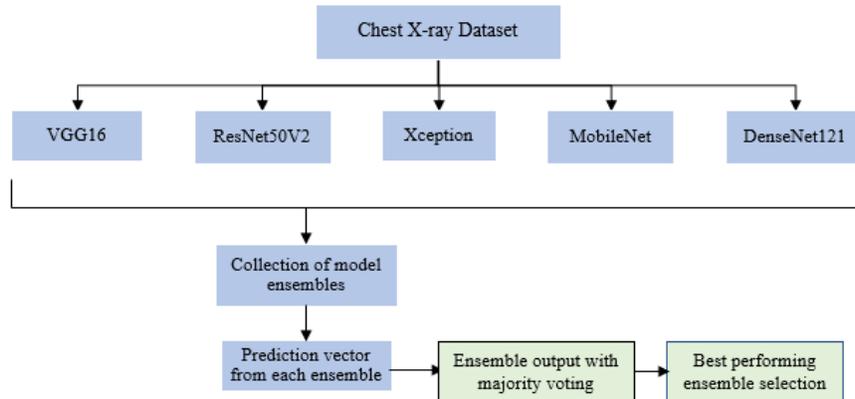


Figure 4: Optimal ensemble selection from pre-trained models using majority voting

It is important to note that a very limited amount of public X-ray data is available for COVID-19 cases which emphasizes the need for improving the existing repository of data as more new cases are identified to expand the dataset. As such, we have collected a total of 226 images with confirmed COVID-19 cases from the first repository mentioned above. To make a balanced dataset, we have collected non-COVID chest X-ray images from the second repository consisting of 452 images divided into 226 images with normal condition and 226 images with both bacterial and viral pneumonia cases. Hence, our curated dataset contains a total of 678 images from all three categories. As more images with COVID-19 cases become available over time, we will increase the size of our curated dataset by adding the new COVID-19 images with the same number of images of each category from the second data repository.

The collected COVID-19 and non-COVID images in the prepared dataset are not of equal sizes. We plan to resize the images to 224×224 and 299×299 which are the standard input image sizes of the selected pre-trained CNN models for faster model convergence. Fig. 5 shows some sample images from both normal and infected categories. The normal chest X-ray image shows clear lungs and does not have any irregular “opacification” area. Bacterial pneumonia image usually shows a focal non-segmental pattern in the image indicated with white arrows in the upper right lobe whereas viral pneumonia image seems to display thin “interstitial” pattern in both lungs. The proposed deep transfer learning models will be used to identify these patterns in X-ray images to effectively differentiate COVID patients.

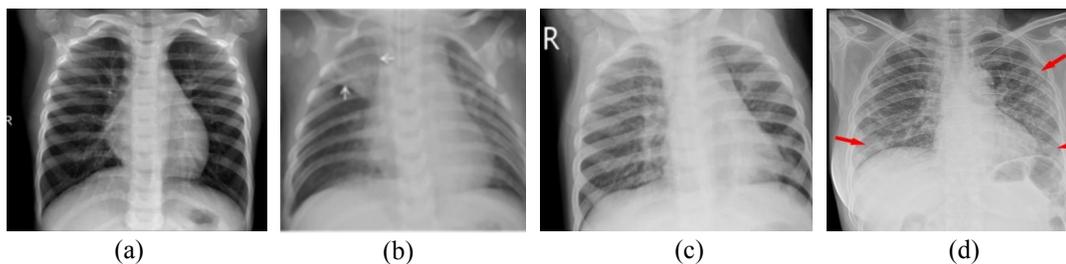


Figure 5: Sample X-ray images from our curated dataset (a) Normal; (b) Bacterial Pneumonia; (c) Viral Pneumonia; (d) COVID-19

3.2.1 Pre-processing and image augmentation

Data scaling is an important pre-processing task for training and evaluating neural network models. Unscaled input image data often results in a slow or unstable learning process. We have scaled the input data using a scaling technique called normalization. Normalization refers to the process of rescaling input data from original range to the range between 0 and 1. Since our dataset contains 8-bit RGB color images, the range of pixel values is between 0 and 255. Hence, we rescale our data using following formulation:

$$Z = \frac{x - \min(x)}{\max(x) - \min(x)} = \frac{x}{255} \quad (1)$$

Furthermore, we attempted to improve the quality and size of the dataset by using various data augmentation techniques which have been used in improving performances in image classification problems in medical fields [Araújo, Guilherme, Eduardo et al. (2017); Frid-Adar, Eyal, Michal et al. (2018)]. Image data augmentation will artificially increase the size of our training dataset to address the limited size of our dataset. To this end, certain data augmentation techniques are applied on training images which also assist in addressing overfitting problems and in increasing model's generalizability during training. Tab. 1 summarizes the augmentation techniques applied on our training dataset.

The zoom range is used to zoom in (magnify) or zoom out (reduce) the image based on a randomly picked value in the range of 1 ± 0.1 (in this case). If the value is less than 1.0 it zooms in the image and if it is greater than 1.0 then it zooms out the image. The shear range slants the shape of the image in counterclockwise direction with a value of 0.1 radian. The rotation range refers to the rotation angle in degrees (i.e., 15 degrees) which is then used to produce images randomly in the range -15 to +15. We used width shifting of 0.1 which specifies the upper bound on the fraction of total width to shift the mages randomly in horizontal manner (i.e., x-direction). Similarly, height shifting of 0.1 specifies the upper bound on the fraction of total height to shift the mages randomly in vertical manner (i.e., y-direction). We also used a nearest type of fill mode in which the empty values are filled by choosing and repeating the closest pixel values. Lastly, horizontal flip causes the images to be flipped horizontally.

Table 1: Image augmentation configurations

Augmentation type	Values
zoom range	0.1
shear range	0.1
rotation range	15
width shift	0.1
height shift	0.1
horizontal flip	true
fill mode	nearest

3.2.2 Cross-validation studies

Generally, dataset is split into training and test sets to train the model with training set and to test the model with test set. This process, however, is not very consistent since the performance results can be very diverse across multiple test sets. K -fold cross validation is utilized to address this problem where the entire data samples are divided into K equal parts and each part in turn is used as validation data to test the model. Hence, our pre-trained models used in this study were evaluated through 5-fold cross validation to lower generalization error. The results obtained from cross-validation are then averaged to yield a single estimate. We split the data (as shown in Fig. 6) into training, validation, test sets randomly with a ratio of 60:20:20 for the purpose of cross-validation studies and final evaluation of the model. The number of X-ray images for each set of data in different categories are given in Tab. 2.



Figure 6: Dataset partitioning

3.2.3 Computational resources

For training and performance evaluation of our proposed pre-trained CNN models, we used Google Colab [Colab (2020)] which is a free Jupyter notebook environment entirely running on the cloud. Colab offers a fully configured run time for deep learning and access to a robust graphical processing unit (GPU) at no cost. Currently, it provides a single CUDA enabled 16 GB NVIDIA Tesla P100 GPU and comes with pre-installed Python 3 with Keras 2.2.5 API and TensorFlow 1.15.0 at the backend.

Table 2: Training, validation, and test data sets

Set	Count			Percent
	Normal	Pneumonia	COVID-19	
Training	135	135	135	60%
Validation	46	46	45	20%
Test	45	45	46	20%
Total	226	226	226	100%

3.3 Model configurations and evaluation metrics

We have used dropout regularization (with a dropout ratio of 0.15) in our custom classifier to reduce overfitting and to improve network generalization error by randomly dropping out nodes during model training [Srivastava, Hinton, Krizhevsky et al. (2014)]. The output from the GAP layer followed by a dropout is passed to the first (fully connected) dense layer having 256 neurons. First dense layer output is then fed to a dropout and then passed to the second dense layer with three neurons and a Softmax classifier.

We have used Adam optimizer [Kingma and Ba (2015)] for training and optimizing the model in order to minimize categorical cross-entropic loss. The optimizer is configured with an initial learning rate of 0.001 and a decay of 0.00002 which is calculated by

dividing the initial learning rate by the number of epochs used for model training. Learning rate is thought to be one of the most dominating hyperparameters in a neural network configuration. We used an automatic optimal learning rate finder technique first introduced by Smith [Smith (2017)] to find an optimal learning rate that the models can start training with.

3.3.1 Automatic learning rate finder

Smith [Smith (2017)] proposes an algorithm to find optimal learning rates automatically for model training. Fig. 7 shows different steps of determining minimum and maximum learning rates for our neural network architecture. The algorithm starts by setting a very small ($1e^{-10}$) and a very large ($1e^{+1}$) value for lower and upper bound of learning rates to train the model. As training continues, learning rate is increased exponentially after every batch update and loss is recorded as well. Typically, training runs for 3 to 5 epochs before the learning rate hits the upper bound. At this point, we plot a smoothed curve for loss and learning rate over the training epochs and identify two values for learning rates. First, the learning rate which causes the loss start decreasing and the second value of learning rate that causes the loss start increasing. Fig. 8 demonstrates the loss for various learning rates for our model using automatic learning rates finder.

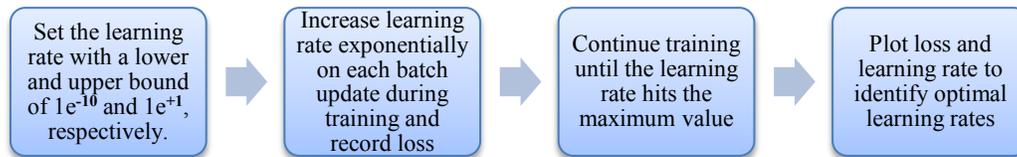


Figure 7: The process of finding optimal learning rates automatically

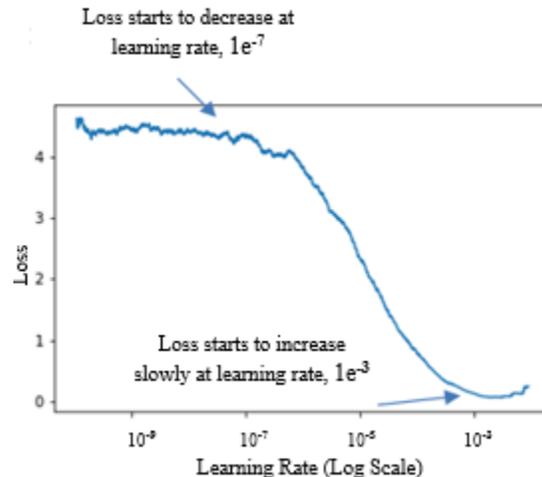


Figure 8: Model loss for various learning rates to identify optimal lower and upper bounds on learning rate

As seen from the plot, loss remains constant until the learning rate drops to approximately $1e^{-8}$. This implies that the model is not learning due to very low initial learning rates. When the learning rate reaches approximately $1e^{-7}$ the loss starts to decrease implying that the model starts to learn due to a large enough learning rate. The loss continues to decrease sharply indicating that the model is learning rapidly until the learning rate decreases to approximately $1e^{-3}$ where loss starts to increase again. Based on this observation we choose $1e^{-3}$ as the initial learning rate in our optimizer for model training. Finally, Tab. 3 shows the summary of our model configurations including hyperparameters.

3.3.2 Evaluation metrics

As stated earlier, we have chest X-ray data samples in our created dataset from three different categories namely normal, pneumonia, and COVID-19. Based on this, we consider our COVID-19 detection problem as three-class classification problem which classifies all three types (normal, pneumonia, and COVID-19) of images.

Table 3: Model configuration summary including hyperparameters

Parameter	Value/Type
Epochs	50
Batch Size	8
Optimizer	Adam
Learning Rates	Initial rate ($1e^{-3}$) with a decay of (initial_rate/epochs)
Loss Function	Categorical cross entropy
Input Shape	224×224 , 299×299
Pooling	GlobalAverage (Flatten layer)
Activation	Softmax (Final dense layer)
Dropout rate	0.15
CNN Weights	ImageNet

The performance evaluation of our models was done in terms of accuracy, precision, recall, F1-score, specificity, and Area Under Curve (AUC). Accuracy is considered as our primary metric since this is an important metric if there are nearly balanced target classes, which is true for our classification problem. It is the number of correct predictions over all predictions made by the model. We used confusion matrix which contains False Positives (FP), True Positives (TP), False Negatives (FN), and True Negatives (TN) to determine precision, recall, F1-score, specificity, and also classification report to obtain values of precision and recall for each target class. In this case, TP specifies the correctly identified COVID-19 cases, while FP specifies normal or pneumonia cases that were incorrectly identified as COVID-19. In addition, TN represents normal or pneumonia cases that were identified as non COVID-19 cases, while FN specifies COVID-19 cases that were incorrectly classified as normal or common pneumonia cases.

Precision measures the proportion of patients that are identified as COVID-19 infected are really infected by the virus. Recall or sensitivity measures the proportion of patients that are infected are diagnosed by the model as COVID-19 patients. Specificity is the opposite of recall which measures the proportion of patients that are not infected are diagnosed by the model as not carrying the virus. F1-score provides a single metric out of precision and recall by calculating their harmonic mean.

4 Results and discussions

We have adopted the following approach to assess the performance of the proposed pre-trained CNN models for identifying COVID-19 patients using chest X-ray images. First, we evaluated the performance of fine-tuned models in which one third of the layers towards the upper level are retrained to attain improved performance and reduced training time. Second, proposed ensemble models are evaluated in an attempt to investigate further performance improvement.

Table 4: Performance metrics of pre-trained models with fine-tuning for the dataset with normal, pneumonia, and COVID-19 images

Model	Accuracy	AUC	Precision	Sensitivity	Specificity	F1-Score
VGG16	94.11	95.47	95.83	95.65	97.67	95.73
ResNet50V2	98.15	98.34	97.87	98.26	98.89	98.06
MobileNet	97.94	98.45	97.83	97.83	98.86	97.83
Xception	97.35	98.01	97.83	97.83	97.67	97.83
DenseNet121	97.79	98.33	97.38	100	97.78	98.67

Tab. 4 shows the performance metrics of our pre-trained models with fine tuning on the holdout test dataset. It is observed that ResNet50V2 and MobileNet outperformed the other pre-trained models almost in all performance metrics including accuracy, sensitivity and specificity. In this case, we can see that these two models achieve very impressive classification accuracy of about 98.15% and 97.94%, respectively on the holdout test dataset towards classifying COVID-19, pneumonia and healthy patients. Both the models show high values of sensitivity (98.26% and 97.83% respectively) and specificity (98.89% and 98.86% respectively) which are two very important performance measures for medical applications. Sensitivity of 98.26% implies that out of 100 positive COVID-19 cases, the models would only miss 1.74 cases in really identifying them as positive. In addition, specificity results imply that of the cases that are COVID-19 negative, the models would accurately detect them as COVID-19 negative 98.89% and 98.86% of the times respectively. It is worthwhile to mention that DenseNet121 showed the best value for sensitivity but achieved similar or lower values for accuracy and specificity as the other two models.

Generic high-level features learned from ImageNet dataset by these models greatly contribute to the success of this classification task. A steady decrease in training and validation losses (as shown in Fig. 9) indicates a modest learning process by the models during the training period. The learning curves also indicate that the models are not overfitting to the training dataset albeit the size of the data is limited. This is largely due to the dropout regularization technique applied to the custom classifier parts of the pre-trained models and the use of image augmentation to combat the scarcity of available COVID-19 data samples.

Tabs. 5 and 6 show the confusion matrix for the studied pre-trained models with fine-tuning using the datasets containing COVID-19, normal and pneumonia samples. The values of TP, FP, TN and FN in Tab. 5 are calculated with respect to COVID-19 cases. First, we can see that False Negatives (FN) counts for both the top performing models (ResNet50V2 and MobileNet) are very few (i.e., 0 and 1) which contribute to higher values of sensitivity. FN indicates that the models identify a COVID-19 patient to be

healthy whereas the patient is infected. This is very detrimental to patient treatment and increases the risk of disease transmission. Second, the models also show a very few False Positive (FP) cases (i.e., 1) that are misidentified as COVID-19 infected which ultimately contributes to higher values of specificity as well as precision. It is very important to limit

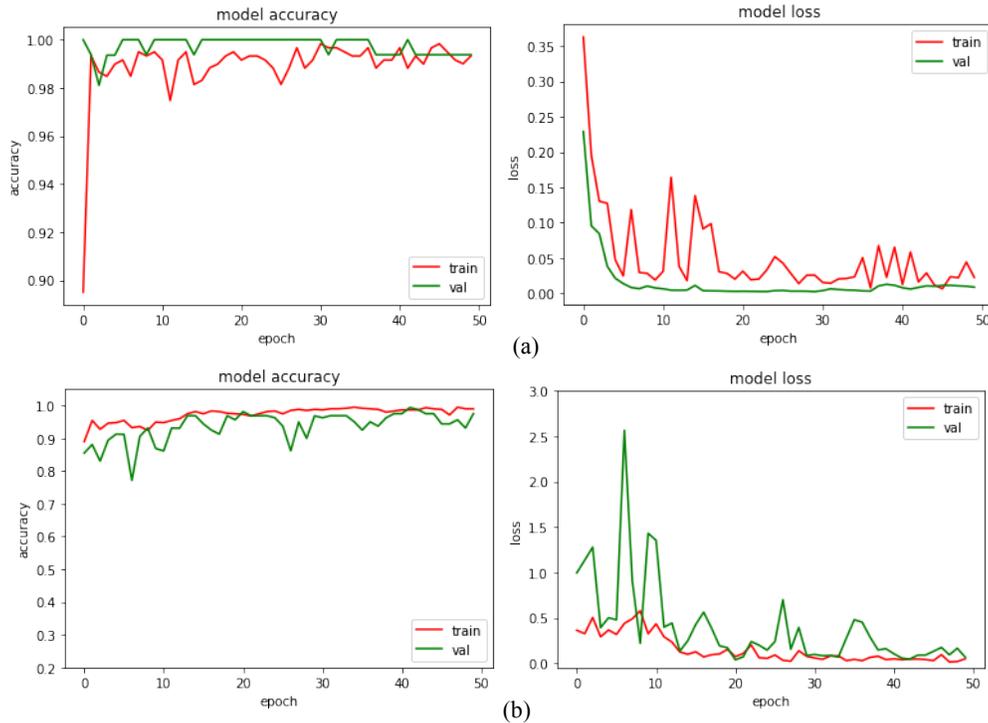


Figure 9: Training and validation loss and accuracy of two best performing models (a) ResNet50V2 (b) MobileNet, with fine-tuning using normal, pneumonia and COVID-19 images

FP counts which otherwise unnecessarily put financial burden on health providers. Based on these obtained results, we consider ResNet50V2 as our best performing model. Tab. 7 shows the values for precision and sensitivity for all three classes in the dataset using ResNet50V2. It shows highest sensitivity for COVID-19 class while demonstrating highest precision for pneumonia class.

Table 5: Confusion matrix for fine-tuned pre-trained models with respect to COVID-19 class using dataset containing normal, pneumonia and COVID-19 samples

Model	TP	FP	TN	FN
VGG16	43	2	85	3
ResNet50V2	46	1	87	0
MobileNet	45	1	87	1
Xception	44	0	88	2
DenseNet121	46	2	86	0

Table 6: Confusion matrix for top two *fine-tuned* pre-trained models with dataset containing normal, pneumonia, and COVID-19 samples

Model	Actual Values	Predicted Values		
		COVID-19	Normal	Pneumonia
ResNet50V2	COVID-19	46	0	0
	Normal	1	44	0
	Pneumonia	0	2	43
MobileNet	COVID-19	45	0	1
	Normal	1	44	0
	Pneumonia	0	2	43

Table 7: Precision and sensitivity for each class using fine-tuned ResNet50V2 model

Class	Precision	Sensitivity	F1-Score
COVID-19	98.0	100	99.0
Normal	96.0	98.0	97.0
Pneumonia	100	96.0	98.0

To further improve the performance of the studied pre-trained CNN models, we evaluated the ensemble model (as shown in Tab. 8) by combining the prediction results from top four performing CNN models (such as ResNet50V2, MobileNet, Xception, and DenseNet121) to generate a single prediction vector and a final output is obtained by majority voting. This ensemble model used the same inputs as the individual fine-tuned CNN models and outperformed them in all performance metrics. A substantial diversity in the base CNN learners in this model ensemble resulted in less correlation in the predictions and thus yielded enhanced performance and generality.

Table 8: Performance metrics of ensemble model for the dataset with normal, pneumonia, and COVID-19 images

Model	Accuracy	AUC	Precision	Sensitivity	Specificity	F1-Score
Ensemble	99.26	99.44	97.87	100	98.89	98.89

4.1 Heatmap visualization

We also investigated how the studied models come to their conclusions in distinguishing COVID-19 cases from normal and other non COVID infections such as pneumonia. In other words, it is important to understand what the models are learning from the supplied data during training and validation. To accomplish this, we have used the concept of gradient-weighted class activation mapping (Grad-CAM) [Selvaraju, Abhishek, Ramakrishna et al. (2019)] that uses important areas of an image to predict the target classes. More specifically, Grad-CAM can be used to find out if our models are activating the correct locations in the images to come to the prediction decision. This gradient-based class activation technique works by locating the last convolutional layer of the model and then investigating the gradient data that is coming into that layer. The output represents a

heatmap visualization with respect to a particular class label that is used to verify in which portion of the image the model is looking at. Fig. 10 illustrates heatmap visualizations from ResNet50V2 of some example COVID-19 positive images indicating the critical (highlighted) areas within the lungs of the infected patients. This could be validated with clinical notes that would take experienced medical professionals to carry out rigorous testing and verification of these results. This is important to ensure that the proposed COVID-19 detector is relying on appropriate information to make decisions. To verify the consistency of prediction explanation obtained through Grad-CAM technique, we have also leveraged another feature importance visualization technique called LIME (Local Interpretable Model-Agnostic Explanations) [Marco, Sameer and Carlos (2016)] that provides local model interpretability. It is observed that both the methods are referring to nearly the same area of images as critical regions in identifying COVID-19 symptoms.

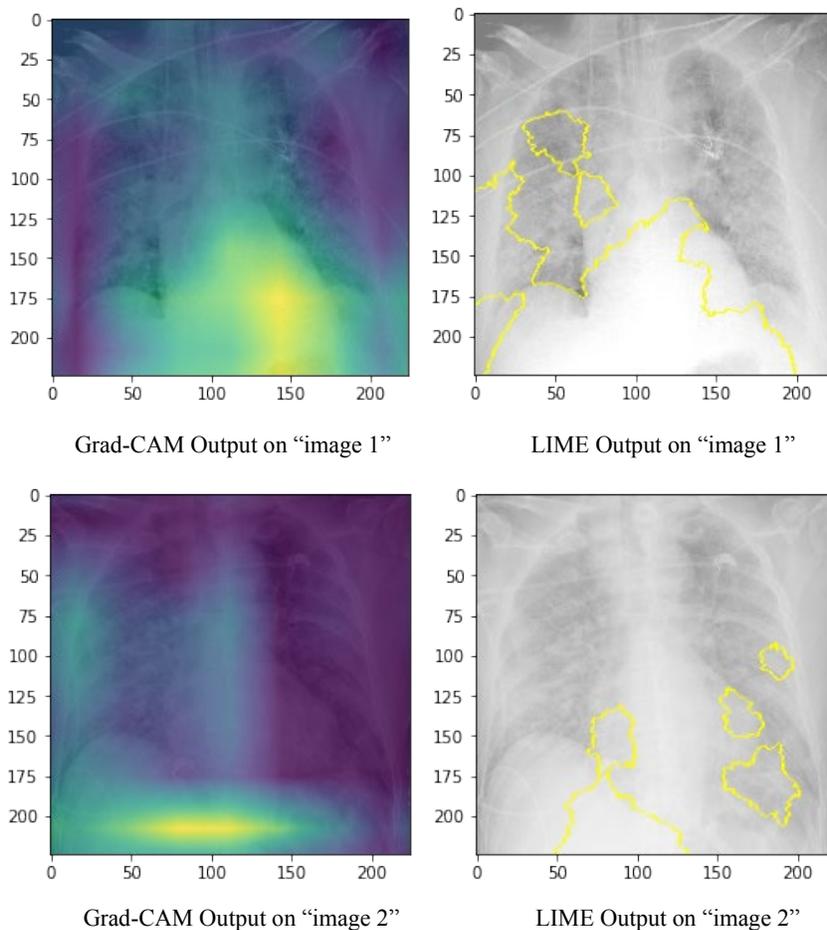


Figure 10: Visualizing activation maps for COVID-19 positive images with Grad-CAM and LIME techniques (ResNet50V2 model)

4.2 Discussion

We compared the performance results obtained by our best performing model and the model ensemble (as shown in Tab. 9) with the results obtained by two recent research on COVID-19 detection using similar datasets. First, COVID-Net [Wang and Wong (2020)] based on a CNN-based deep learning model represents a relatively earlier effort on COVID-19 detection using chest X-ray images. COVID-Net uses a relatively fewer number (less than 100) of COVID-19 images for training as compared to our dataset. However, the number of training images used in each of normal and pneumonia categories is much higher (close to 10000) which makes their dataset rather imbalance. It reported accuracy, sensitivity, and precision (positive predictive value) of 92.4%, 91.33% and 88.67% respectively. COVID-Net leverages 111.6 million parameters to achieve this performance accuracy whereas our fine-tuned ResNet50V2 model attains higher accuracy (98.15%), sensitivity (98.26%) and precision (97.87%) with only about 24.1 million parameters. This results in a significant savings in terms of computation. Second, the authors in Apostolopoulos et al. [Apostolopoulos and Mpesiana (2020)] present a subsequent study on COVID-19 detection using transfer learning with chest X-ray images from a dataset containing the same number of COVID-19 positive images as our dataset. Their dataset is also largely imbalanced where they have used 504 healthy and 714 pneumonia cases. As reported in their paper, they have achieved best performance using pre-trained MobileNetV2 network with sensitivity of 98.66% which is better than our fine-tuned ResNet50V2 model (98.26%). In terms of specificity, our ResNet50V2 model performs better than its counterpart (MobileNetV2). Overall, the proposed ensemble model outperforms the state-of-the-art approaches to COVID-19 detection in terms of accuracy and sensitivity while showing similar results for precision and specificity.

Table 9: Performance comparison of the proposed and state-of-the-art models

Model	Accuracy	Precision	Sensitivity	Specificity	AUC
COVID-Net [Wang and Wong (2020)]	92.4	91.33	88.67	-	-
[Apostolopoulos and Mpesiana (2020)]	94.72	-	98.66	96.46	-
ResNet50V2 in this paper	98.15	97.87	98.26	98.89	98.34
Ensemble model in this paper	99.26	97.87	100	98.89	99.44

Based on the performance and comparative results, it is worthwhile to deduce that CNN based deep learning models have significant impact on automatic detection of coronavirus disease patients by effectively extracting important features from chest X-ray images. At the same time, it is imperative to highlight some of the limitations of the current study which can potentially be addressed in future research. The major limitation is the inadequate availability of chest X-ray images from COVID-19 patients and thus the models were trained and tested using a small dataset containing few hundreds of images. An in-depth analysis of the models can be done as more new data becomes available.

Since accurate diagnostics necessitate a deeper understanding of radiological features (bio-markers) evident in chest X-ray images, the model interpretability results obtained by different tools need to be clinically verified by trained medical professionals. Nevertheless, our present study offers an automatic, faster and computationally cost-effective solution to the diagnosis of COVID-19 patients. Furthermore, even though an appropriate line of treatment cannot be determined by merely screening an X-ray image, it can aid in adopting some early measures such as quarantining the positive cases until a complete checkup and treatment are prescribed.

5 Conclusions and future work

In this study, we proposed an end-to-end deep learning model based on pre-trained CNNs to offer an early and automatic detection of COVID-19 patients to prevent the spread of the disease. The study leverages open source chest X-ray images of healthy, pneumonia, and COVID-19 cases. The curated dataset consists of 678 images containing 226 samples from each category. Our best performing fine-tuned model, ResNet50V2, achieves an accuracy of 98.15% in classifying healthy, coronavirus and pneumonia infected patients with a high degree of precision (97.87%) and sensitivity (98.26%). We also constituted a model ensemble consisting of four fine-tuned pre-trained models which further improves the performance metrics and outperforms two state-of-the-art approaches. Despite having the limitations discussed early, we believe that the results obtained from this work will benefit the radiologists and health practitioners to earn profound insights into important factors related to coronavirus infection. We also have demonstrated a good control on the model development by defining the COVID-19 detection problem as a three-class (normal, pneumonia, and COVID-19) classification problem to check if the model is really detecting coronavirus infection or merely detecting pulmonary edema (a condition causing excessive fluid in the lungs) which is usually looked for in the X-ray images to detect symptoms caused by many other common diseases. In future, to make our model more robust, we will increase the size of our curated dataset by adding the new COVID-19 images as they become available. We also plan to better address COVID-19 detection problem as a multi-modal problem where heterogeneous patient data will be collected from various sources such as patient vitals, area, density of population and so on.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- Alessa, A.; Faezipour, M.** (2018): A review of influenza detection and prediction through social networking sites. *Theoretical Biology and Medical Modeling*, vol. 15, no. 1, pp. 1-27.
- Anastassopoulou, C.; Lucia, R.; Athanasios, T.; Constantinos, S.** (2020): Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS One*, vol. 15, no. 3, pp. 1-21.

- Apostolopoulos, I. D.; Mpesiana, T. A.** (2020): COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, <https://doi.org/10.1007/s13246-020-00865-4>.
- Araújo, T.; Guilherme, A.; Eduardo, C.; José, R.; Paulo, A. et al.** (2017): Classification of breast cancer histology images using convolutional neural networks. *PLoS One*, vol. 12, no. 6, pp. 6035-6047.
- Bo, R. B.; Bonggun, S.; Yoonjung, C.; Sungsoo, P.; Keunsoo, K.** (2020): Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Computational and Structural Biotechnology Journal*, vol. 18, pp. 784-790.
- Canziani, A.; Paszke, A.; Culurciello, E.** (2016): An analysis of deep neural network models for practical applications. *5th International Conference on Learning Representations*, pp. 1-7.
- Celik, Y.; Talo, M.; Yildirim, O.; Karabatak, M.; Acharya, U. R.** (2020): Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. *Pattern Recognition Letters*, vol. 133, pp. 232-239.
- Chakraborty, S.** (2020): An attempt-detection of COVID-19 presence from chest X-ray scans using CNN and class activation maps. <https://towardsdatascience.com/detection-of-covid-19-presence-from-chest-x-ray-scans-using-cnn-class-activation-maps-c1ab0d7c294b>.
- Chan, J. F.; Yuan, S.; Kok, K. H.; To, K. K.; Chu, H. et al.** (2020): A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, vol. 395, no. 10223, pp. 514-523.
- Chen, J.; Wu, L.; Zhang, J.; Zhang, L.; Gong, D. et al.** (2020): Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. *medRxiv*. <https://doi.org/10.1101/2020.02.25.20021568>.
- Chen, N.; Zhou, M.; Dong, X.; Qu, J.; Gong, F. et al.** (2020): Epidemiological and clinical characteristics of 99 cases of 371 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*, vol. 395, no. 10223, pp. 507-513.
- Chollet, F.** (2017): Xception: deep learning with depth wise separable convolutions. *International Conference on Computer Vision and Pattern Recognition*, pp. 1800-1807.
- Cohen, J. P.** (2020): COVID-19 chest X-ray dataset. <https://github.com/ieee8023/covid-chestxray-dataset>.
- Colab** (2020): Google Colab. <https://colab.research.google.com/>.
- Deng, J.; Dong, W.; Socher, R.; Li, L. J.; Li, K. et al.** (2009): ImageNet: a large-scale hierarchical image database. *International Conference on Computer Vision and Pattern Recognition*, pp. 248-255.
- Fong, S. J.; Li, G.; Dey, N.; Crespo, R. G.; Herrera-Viedma, E.** (2020): Finding an accurate early forecasting model from small dataset: a case of 2019-ncov novel coronavirus outbreak. *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, pp. 51-61.

- Frid-Adar, M.; Eyal, K.; Michal, A.; Jacob, G.; Hayit, G.** (2018): Synthetic data augmentation using GAN for improved liver lesion classification. *15th IEEE International Symposium on Biomedical Imaging*, pp. 289-293.
- Google, N.** (2020): Coronavirus (COVID-19). <https://google.com/covid19-map/?hl=en>.
- He, K.; Zhang, X.; Ren, S.; Sun, J.** (2016): Deep residual learning for image recognition. *International Conference on Computer Vision and Pattern Recognition*, pp. 770-778.
- Howard, A. G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W. et al.** (2017): MobileNets: efficient convolutional neural networks for mobile vision applications. *Google Inc.* <https://towardsdatascience.com/using-kalman-filter-to-predict-corona-virus-spread-72d91b74cc8>. <https://www.worldometers.info/coronavirus/>.
- Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J. et al.** (2020): Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, vol. 395, no. 10223, pp. 497-507.
- Huang, G.; Liu, Z.; Laurens, V. D. M.; Weinberger, L.** (2016): Densely connected convolutional networks. *International Conference on Computer Vision and Pattern Recognition*, pp. 2261-2269.
- ICTV** (2020): International committee on taxonomy of viruses (ICTV). <https://talk.ictvonline.org/>.
- Jiang, X.; Coffee, M.; Bari, A.; Wang, J.; Jiang, X. et al.** (2020). Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Computers, Materials & Continua*, vol. 63, no. 1, pp. 537-551.
- Kermany, D. S.; Goldbaum, M.; Cai, W.; Valentim, C. C. S.; Liang, H. et al.** (2018): Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, vol. 172, no. 5, pp. 1122-1131.
- Kingma, D. P.; Ba, J. L.** (2015): Adam: a method for stochastic optimization. *3rd International Conference for Learning Representations*, pp. 1-15.
- Krizhevsky, A.; Sutskever, I.; Hinton, G. E.** (2017): ImageNet classification with deep convolutional neural networks. *Communication of the ACM*, vol. 60, pp. 84-90.
- Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L. et al.** (2020): Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *The New England Journal of Medicine*, vol. 382, no. 13, pp. 1199-1207.
- Lin, M.; Chen, Q.; Yan, S.** (2013): Network in network, *International Conference on Learning Representations*, pp. 1-10.
- Liu, Y.; Liu, G.; Zhang, Q.** (2019): Deep learning and medical diagnosis. *The Lancet*, vol. 394, no. 10210, pp. 1710-1711.
- Marco, T. R.; Sameer, S.; Carlos, G.** (2016): Why should I trust you? Explaining the predictions of any classifier. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135-1144.

Raghu, M.; Zhang, C.; Kleinberg, J. M.; Bengio, S. (2019): Transfusion: understanding transfer learning with applications to medical imaging. *33rd Conference on Neural Information Processing Systems*, pp. 1-22.

Ran, K. (2020): Using Kalman filters to predict spread of coronavirus.

Ravishankar, H.; Sudhakar, P.; Venkataramani, R.; Thiruvankadam, S.; Annangi, P. et al. (2016): Understanding the mechanisms of deep transfer learning for medical images. *Deep Learning in Medical Image Analysis*.

Saba, T.; Mohamed, A. S.; El-Affendi, M.; Amin, J.; Sharif, M. (2020): Brain tumor detection using fusion of hand crafted and deep learning features. *Cognitive Systems Research*, vol. 59, pp. 221-230.

Selvaraju, R. R.; Abhishek, D.; Ramakrishna, V.; Michael, C.; Devi, P. et al. (2019): Grad-CAM: visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336-359.

Simonyan, K.; Zisserman, A. (2014): Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556

Smith, L. N. (2017): Cyclical learning rates for training neural networks. *IEEE Winter Conference on Applications of Computer Vision*, pp. 464-472.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. (2014): Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, vol. 15, pp. 1929-1958.

Wang, L.; Wong, A. (2020): COVID-Net: a tailored deep convolutional neural network. *Design for Detection of COVID-19 Cases from Chest Radiography Images*. arXiv preprint.

Wang, S.; Kang, B.; Ma, J. L.; Zeng, X. J.; Xiao, M. M. et al. (2020): A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). medRxiv.

Wang, W.; Jiang, Y. B.; Luo, Y. H.; Li, J.; Wang, X. et al. (2019): An advanced deep residual dense network (DRDN) approach for image super-resolution. *International Journal of Computational Intelligence Systems*, vol. 12, no. 2, pp. 1592-1601.

Wang, W.; Li, Y. T.; Zou, T.; Wang, X.; You, J. Y. et al. (2020): A novel image classification approach via Dense-MobileNet models. *Mobile Information Systems*, <https://doi.org/10.1155/2020/7602384>.

Worldometer (2020): COVID-19 coronavirus pandemic.

Xu, X.; Jiang, X.; Ma, C.; Du, P.; Li, X. et al. (2019): Automated detection of diabetic subject using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals. *Computers in Biology and Medicine*, vol. 113, pp. 103387.

Yan, L.; Zhang, H. T.; Xiao, Y.; Wang, M.; Sun, C. et al. (2020): Prediction of criticality in patients with severe COVID-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan, medRxiv.

Yildirim, O.; Talo, M.; Ay, B.; Baloglu, U. B.; Aydin, G. et al. (2019): Automated detection of diabetic subject using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals. *Computers in Biology and Medicine*, vol. 113.

Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H.; Ghahramani, Z. et al. (2014): How transferable are features in deep neural networks? *The 27th International Conference on Neural Information Processing Systems (NeurIPS)*, vol. 2, pp. 3320-3328.

Zhang, H.; Saravanan, K. M.; Yang, Y.; Hossain, M. T.; Li, J. et al. (2020): Deep learning-based drug screening for novel coronavirus 2019-nCov. *Preprints*.

Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B. et al. (2020): A novel coronavirus from patients with pneumonia in China, 2019. *The New England Journal of Medicine*, vol. 382, pp. 727-733.