# Detection of COVID-19 using Chest X-Ray Scans

Anuraag Shankar
*Department of Computer Engineering*
*Pune Institute of Computer Technology*
Pune, India
anuraag.k.shankar@gmail.com

Yash Sonar
*Department of Computer Engineering*
*Pune Institute of Computer Technology*
Pune, India
yashsonar213@gmail.com

Dr. Kavita A. Sultanpure
*Department of Information Technology*
*Pune Institute of Computer Technology*
Pune, India
kasultanpure@pict.edu

***Abstract*—COVID-19 has proven to be the unseen and unforeseen pandemic nobody was prepared to face. Healthcare professionals and radiologists have been under a lot of pressure ever since the outbreak to treat and develop faster ways to detect the disease. The proposed research involves the classification of chest X-ray scans to identify whether a patient has been infected with COVID-19 or not using the concept of Transfer Learning. Two methods of classifying the images have been presented in the research. The first approach classifies a given image into two categories being COVID-19 and Non COVID-19. The second approach classifies a given image into three categories namely COVID-19, Healthy and Pneumonia. These two models have obtained unprecedented evaluation metrics and could prove to be extremely useful when it comes to fast and accurate detection of the disease.**

***Index Terms*—Classification, Convolutional Neural Networks (CNNs), COVID-19, Pneumonia, Transfer Learning**

## I. Introduction

Retrospective investigations by Chinese authorities started identifying first cases of Novel Coronavirus or COVID-19 in early December 2019. These cases were found to be originating from Wuhan, the capital of the Hubei province. Early attempts to control the virus were aimed by considering it as pneumonia. After a few attempts, when the patients showed no signs of recovery and their condition deteriorated, it was found that this is a new virus. The spreading of the virus has been extremely rapid resulting in a global pandemic. The virus is prevalent in at least 188 countries and territories with over 4.9 million infections and increasing daily [1].
Severe Pneumonia requires extended hospital stay. In some cases, the patients need to be placed in an ICU with mechanical ventilators. It is necessary that these patients be treated in regular hospitals rather than those treating COVID-19 patients. Attempts to identify the virus were carried out through CT scans and chest X-rays. This method required experienced doctors to classify the images manually based on past experiences. This method proved to be helpful when the number of cases were inconsequential. But due to the exponential growth of the virus, there is an urgent need for faster alternatives that can classify the scans with accurate results.



An earlier study shows the DarkCOVIDNet [4] being used for detection of the disease. This model used a modified DarkNet architecture as its base to classify the images. Another research [5] shows the AlexNet model being compared with a user-defined model for classifying CT scans and X-rays into two categories namely COVID-19 and Normal. The two studies have been explained in detail in Section II.

This paper proposes two methods that have been used to classify the chest X-ray scans of patients. The first method classifies a given X-ray into two categories as follows -

- COVID-19
- Non COVID-19

The class "Non COVID-19" refers to patients who are either healthy or infected by Pneumonia (Bacterial or Viral).
The second method of classification categorizes X-rays into one of the following three categories -

- COVID-19
- Healthy
- Pneumonia

The classes "Healthy" and "Pneumonia" provide a clearer perspective of the patient's health. This would not be possible using the two-way classifier as both healthy as well as Pneumonia infected results would come under a common category.

This research is focused on the most important issues regarding this domain:

1) Collection of X-Ray scans [2] of patients infected with COVID-19. Publicly available images of CT scans are very few in number, leading to classification being performed only on X-Rays. For all Non COVID-19 cases, Healthy and Pneumonia infected images have been obtained [3]. The data distribution has been given in Section III-A.
2) To maximize the output with the limited amount of data, the most efficient approach to solve the given problem would be using Transfer Learning. The state-of-the-art models being used for Transfer Learning have been mentioned in Section III-B.
3) Several metrics have been calculated for the models. The metrics used for the two-way and three-way classifiers

have been presented in Section III-C.

Section II, Literature Review, describes the state-of-the-art models existing in the same domain and discusses their approach. Section III discusses the several approaches implemented in order to get the best results. Section IV shows the results obtained by performing cross-validation and data augemntations on the dataset. The research concludes in Section V followed by Acknowledgements and References.

## II. Literature Review

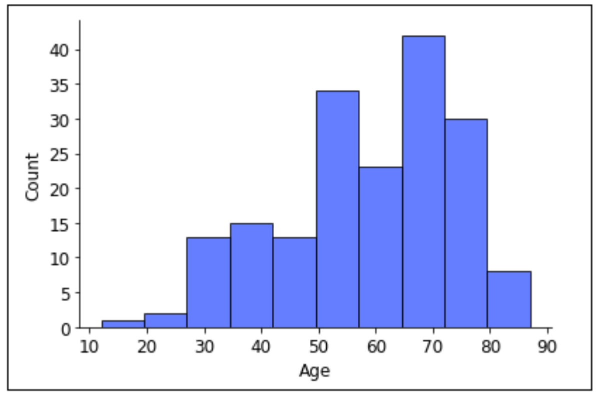
Evaluating and Classifying images based on X-ray scans and CT scans is widely practiced by researchers all over the world. The authors in [4] were able to gather a dataset containing 127 COVID-19 positive X-ray images, 500 Pneumonia images and 500 no-finding images. The model used as a starting point in this case was the Darknet-19 model, which forms the basis of a real time object detection system named YOLO [4]. They designed the DarkCOVIDNet architecture which was inspired from the DarkNet-19 architecture [4].

It is observed that the DarkNet architecture has been modified to reduce the number of layers and filters as compared to the original architecture. The number of filters have been gradually increased such as 8, 16, 32 [4]. “The proposed model has 17 convolution layers” [4]. “Similar to the Darknet-19 model, the Maxpool method is used in all the pooling operations” [4]. When used for binary classification task, the classifier classifies images into COVID-19 or no-findings whereas for multi-class classification, it categorizes images into 3 classes namely COVID-19, Pneumonia or no-findings [4]. The model has been trained for 100 epochs and the performance of the model has been evaluated using 5-fold cross-validation procedure for both binary and triple classification problems [4]. It shows an average binary classification accuracy of 98.08% and a triple class classification accuracy of 87.02% [4].

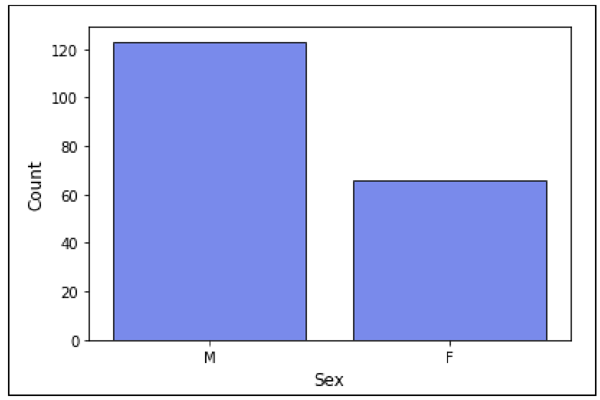
In another attempt, the authors in [5] have worked on the same issue in a different way. The model here is trained using 120 X-ray images (60 COVID-19 and 60 normal) and 339 CT scan images (192 COVID-19 and 147 normal) [5]. The dataset is divided into two categories: 50% to train the CNN and remaining 50% to validate the model 3 times in each epoch [5]. The testing of the model was done on a total of 67 images comprising both X-ray and CT scan images.

The proposed model here consists of a CNN with one convolution layer which is followed by a BatchNorm layer followed by ReLU activation [5]. The fully connected layer is followed by a Softmax layer which outputs ‘0’ or ‘1’. The concepts of transfer learning have also been implemented here and the pretrained model AlexNet has been used which has been trained on over a few million images on ImageNet and in the range of 1000 classes [5]. The last layer of the AlexNet has been replaced to obtain binary results [5].

The comparisons in the results show that the proposed CNN performs better than the AlexNet in case of CT scans with an accuracy of 94.1% as compared to 82% for the AlexNet [5]. The AlexNet overshadows the proposed CNN in case of X-rays with an accuracy of 98% and compared to 94% of the proposed CNN [5].




(a) By Age




(b) By Sex

Fig. 1: Distribution of COVID-19 positive patients.

TABLE I: Data distribution for two-way classification

| Class | Train | Validation | Test | Total |
|---|---|---|---|---|
| COVID-19 | 125 | 47 | 47 | 219 |
| Non COVID-19 | 250 | 57 | 60 | 367 |
| Combined | 375 | 104 | 107 | 586 |

TABLE II: Data distribution for three-way classification

| Class | Train | Validation | Test | Total |
|---|---|---|---|---|
| COVID-19 | 170 | 24 | 25 | 219 |
| Healthy | 170 | 25 | 25 | 220 |
| Pneumonia | 170 | 25 | 25 | 220 |
| Combined | 510 | 74 | 75 | 659 |

## III. Methodology

### A. Data Collection

For both the methods of classification mentioned in Section I, common sources of data have been used but with varying distribution.

- Two-way classification:

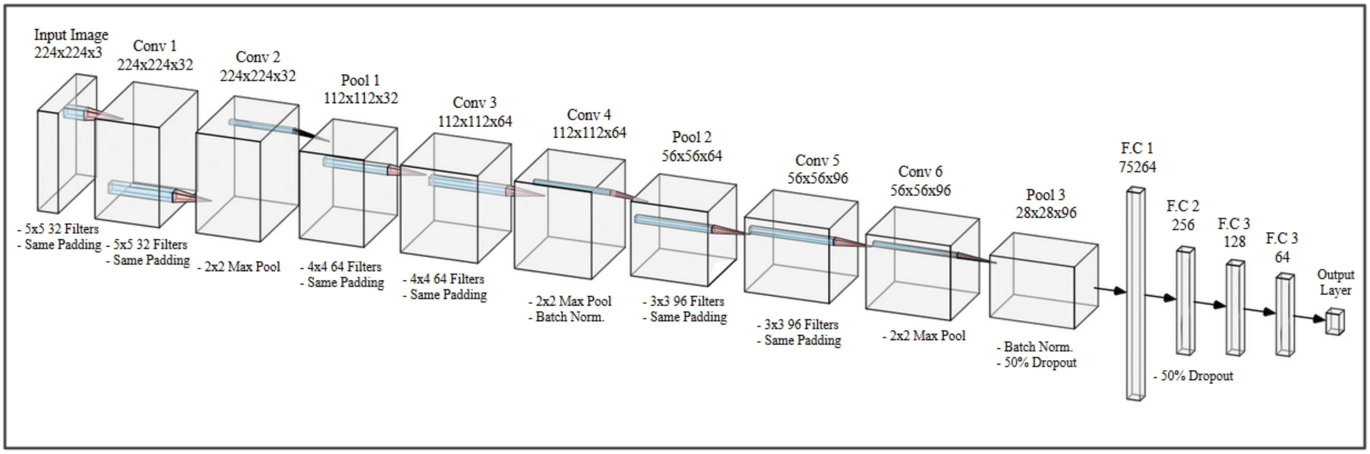



Fig. 2: User-defined model architecture.

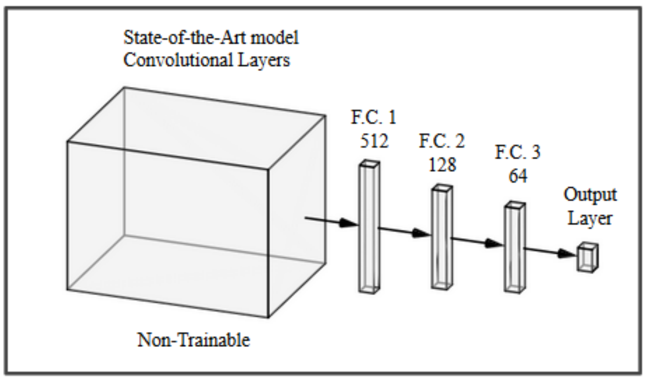



Fig. 3: Transfer learning model architecture.

- A total of 219 images of front-view chest X-rays were obtained for patients infected with COVID-19 [2].
- A total of 367 images were chosen from a shuffled combined dataset of patients infected with Viral Pneumonia, Bacterial Pneumonia and Healthy patients [3].

- Three-way classification:
  - A total of 219 images of front-view chest X-rays were obtained for patients infected with COVID-19.
  - A total of 220 images were selected from a shuffled dataset of healthy patients.
  - A total of 220 images were selected from a combined shuffled dataset of patients infected with Viral and Bacterial Pneumonia.

Tables I and II show the distribution of the dataset for two-way and three-way classification.

On further exploration of the dataset, it is observed that the Age and Sex could also act as influential factors to decide whether the person has been infected or not as shown in Fig. 1. From Fig. 1a, it can be seen that patients with age greater than 50 have a higher chance of being infected by the virus. The distribution in Fig. 1b shows that the number of male infected patients are almost twice as the number of female infected patients.

### B. Models Used

The models were made using Keras, and Tensorflow for the backend. The dataset was trained and tested on a variety of models, containing several state-of-the-art models. Out of these models, the following three were chosen:

- VGG-19 [6]
- ResNet50 [7]
- MobileNet [8]

These models were further analyzed and compared with each other and with a user-defined model. The architecture of the user-defined model is shown in Fig. 2.

Each layer in the model is followed by a ReLu activation function except the Output Layer, where a Softmax activation has been applied. The dimensionality of the Output Layer is conditioned upon the mode of classification.
For the transfer learning networks, their convolutional layers were frozen and their fully-connected layers were replaced as shown in Fig. 3.
The input-shape of images for the models is as follows:

- VGG-19: 224x224x3
- ResNet50: 299x299x3
- MobileNet: 224x224x3

### C. Metrics

The models were tested on the following metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall(Sensitivity) = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

TABLE III: Confusion matrix for two-way classification

| Model Used | | Predicted COVID-19 | Predicted Non COVID-19 |
|---|---|---|---|
| **VGG-19** | **Actual COVID-19** | 47 | 0 |
| | **Actual Non COVID-19** | 0 | 60 |
| **User-defined Model** | **Actual COVID-19** | 46 | 1 |
| | **Actual Non COVID-19** | 4 | 56 |
| **ResNet50** | **Actual COVID-19** | 42 | 5 |
| | **Actual Non COVID-19** | 0 | 60 |
| **MobileNet** | **Actual COVID-19** | 39 | 8 |
| | **Actual Non COVID-19** | 0 | 60 |

TABLE IV: Confusion matrix for three-way classification

| Model Used | | Predicted COVID-19 | Predicted Healthy | Predicted Pneumonia |
|---|---|---|---|---|
| **VGG-19** | **Actual COVID-19** | 25 | 0 | 0 |
| | **Actual Healthy** | 0 | 24 | 1 |
| | **Actual Pneumonia** | 0 | 0 | 25 |
| **User-defined Model** | **Actual COVID-19** | 25 | 0 | 0 |
| | **Actual Healthy** | 0 | 21 | 4 |
| | **Actual Pneumonia** | 0 | 0 | 25 |
| **ResNet50** | **Actual COVID-19** | 25 | 0 | 0 |
| | **Actual Healthy** | 0 | 21 | 4 |
| | **Actual Pneumonia** | 1 | 0 | 24 |
| **MobileNet** | **Actual COVID-19** | 24 | 0 | 1 |
| | **Actual Healthy** | 0 | 16 | 9 |
| | **Actual Pneumonia** | 0 | 1 | 24 |

TABLE V: Metrics calculated for two-way classification (%)

| Model Used | Overall Accuracy | COVID-19 | | Non COVID-19 | |
|---|---|---|---|---|---|
| | | Recall | Precision | Recall | Precision |
| **VGG-19** | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| **User-defined Model** | 95.33 | 97.87 | 92.00 | 93.33 | 98.25 |
| **ResNet50** | 95.33 | 89.36 | 100.00 | 100.00 | 92.31 |
| **MobileNet** | 92.52 | 82.97 | 100.00 | 100.00 | 88.24 |

TABLE VI: Metrics calculated for three-way classification (%)

| Model Used | Overall Accuracy | COVID-19 | | Healthy | | Pneumonia | |
|---|---|---|---|---|---|---|---|
| | | Recall | Precision | Recall | Precision | Recall | Precision |
| **VGG-19** | 98.67 | 100.00 | 100.00 | 96.00 | 100.00 | 100.00 | 96.15 |
| **ResNet50** | 94.67 | 100.00 | 100.00 | 84.00 | 100.00 | 100.00 | 86.21 |
| **User-defined Model** | 93.33 | 100.00 | 96.15 | 84.00 | 100.00 | 96.00 | 85.71 |
| **MobileNet** | 85.33 | 96.00 | 100.00 | 64.00 | 94.12 | 96.00 | 70.59 |

where TP, TN, FP, FN are True Positives, True Negatives, False Positives and False Negatives respectively.
Tables III and IV describe the confusion matrix of the various models for the two classifiers. Tables V and VI show the various metrics calculated for each model.

## IV. Results and Discussion

Cross-validation was performed for both the classifiers for every model. The method of cross-validation chosen for this purpose was the K-Fold Cross Validation. Selecting the value of K as five, metrics for each model were calculated.

Tables VII and VIII show the results after performing 5-fold cross-validation for both the classification tasks.
Here, Accuracy, Recall and Precision have been calculated from (1), (2), (3). The additional metrics have been calculated as:

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{TP + FN} \tag{5}$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP + TN} \tag{6}$$

$$\begin{aligned} &Receiver\ Operating\ Characteristic(ROC)\ Curve \\ &= Curve\ describing\ TPR\ vs\ FPR\ relationship \end{aligned} \tag{7}$$

$$ROC\_AUC\ score = Area\ under\ ROC\ Curve \tag{8}$$

A variety of Data Augmentation methods were applied on the respective test sets which can be labelled as follows:

$A = ZCA\ Whitening\ (ZCA\ Epsilon = 10^{-6})$
$B = Horizontal\ Flip$
$C = Vertical\ Flip$
$D(x) = Height\ and\ Width\ Shifted\ by\ a\ factor\ of\ “x”$

An augmentation of B, D(0.2) would mean flipping the image horzontally and shifting the image vertically as well as horizontally by 20%.

TABLE VII: Average metrics calculated after 5-fold cross-validation over 586 images in two-way classification (%)

| Model Used | Overall Accuracy | COVID-19 | | | Non COVID-19 | | | ROC_AUC Score |
|---|---|---|---|---|---|---|---|---|
| | | Recall | Precision | F1 Score | Recall | Precision | F1 Score | |
| **VGG-19** | 99.32 | 99.56 | 98.56 | 99.08 | 99.16 | 99.74 | 99.44 | 99.13 |
| **ResNet50** | 96.25 | 90.43 | 99.49 | 94.74 | 99.73 | 94.26 | 96.92 | 95.07 |
| **User-defined Model** | 93.39 | 92.64 | 90.89 | 91.75 | 93.78 | 95.84 | .94.80 | 93.21 |
| **MobileNet** | 89.74 | 79.25 | 99.26 | 88.13 | 99.26 | 86.43 | 92.39 | 86.34 |

TABLE VIII: Average metrics calculated after 5-fold cross-validation over 659 images in three-way classification (%)

| Model Used | Overall Accuracy | COVID-19 | | | Healthy | | | Pneumonia | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Recall | Precision | F1 Score | Recall | Precision | F1 Score | Recall | Precision | F1 Score |
| **VGG-19** | 97.80 | 99.07 | 97.76 | 98.41 | 98.01 | 97.06 | 97.53 | 9667 | 98.7 | 97.67 |
| **User-defined Model** | 93.63 | 97.25 | 96.40 | 96.82 | 91.36 | 94.22 | 92.76 | 92.27 | 91.19 | 91.72 |
| **ResNet50** | 89.31 | 77.33 | 100.00 | 87.21 | 94.32 | 85.87 | 89.89 | 96.02 | 86.38 | 90.94 |
| **MobileNet** | 81.15 | 68.13 | 99.52 | 80.88 | 78.64 | 93.23 | 85.31 | 97.27 | 69.90 | 81.34 |

TABLE IX: VGG-19 results on test set after performing data augmentation for two-way classification (%)

| Augmentations Performed | Overall Accuracy | COVID-19 | | | Non COVID-19 | | | ROC_AUC Score |
|---|---|---|---|---|---|---|---|---|
| | | Recall | Precision | F1 Score | Recall | Precision | F1 Score | |
| **A, B** | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| **A, B, C** | 98.13 | 95.74 | 100.00 | 97.82 | 100.00 | 96.77 | 98.36 | 97.87 |
| **A, B, D(0.2)** | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| **A, B, D(0.3)** | 93.46 | 100.00 | 87.04 | 93.07 | 88.33 | 100.00 | 93.80 | 94.17 |
| **A, B, D(0.4)** | 92.52 | 100.00 | 85.45 | 92.15 | 86.67 | 100.00 | 92.86 | 93.33 |

TABLE X: VGG-19 results on test set after performing data augmentation for three-way classification (%)

| Augmentations Performed | Overall Accuracy | COVID-19 | | | Healthy | | | Pneumonia | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Recall | Precision | F1 Score | Recall | Precision | F1 Score | Recall | Precision | F1 Score |
| **A, B** | 96.00 | 100.00 | 96.15 | 98.04 | 92.00 | 95.83 | 93.88 | 96.00 | 96.00 | 96.00 |
| **A, B, C** | 85.33 | 100.00 | 100.00 | 100.00 | 56.00 | 100.00 | 71.79 | 100.00 | 69.44 | 81.96 |
| **A, B, D(0.2)** | 81.33 | 100.00 | 92.59 | 96.15 | 44.00 | 100.00 | 61.11 | 100.00 | 67.57 | 80.65 |
| **A, B, D(0.3)** | 74.67 | 96.00 | 85.71 | 90.56 | 36.00 | 100.00 | 52.94 | 100.00 | 60.53 | 75.41 |
| **A, B, D(0.4)** | 69.33 | 96.00 | 77.42 | 85.71 | 28.00 | 100.00 | 43.75 | 84.00 | 56.76 | 67.74 |

Tables IX and X show the results obtained after performing five different combinations of augmentations for both the classifiers on the trained VGG-19 model.

It is clearly visible from Tables VII and VIII that the VGG-19 model outperforms every other model and shows results comparable its trained equivalent (Tables V and VI) when 5-fold cross-validation is performed. The ResNet50 model shows better results than the User-defined model for two-way classification but fails to get a high recall on the COVID-19 class in three-way classification resulting in a lower overall performance. The User-defined model shows almost identical accuracies for both the classifiers but performs better on the COVID-19 class in three-way classification. The MobileNet model shows decent results for the two-way classification task and below-par performance for the three-way classification task. It is seen that the MobileNet model fails to classify quite a few images of the COVID-19 class in both cases and a few of the Healthy class images in the three-way classification task. The reason for lower performance of MobileNet model in comparison to the other models could be the shallower architecture of MobileNet model [8].

From Table IX, it is seen that as the augmentations get more complex, it becomes difficult for the model to classify images from the Non COVID-19 category, but it still shows a high recall for the same. The overall performance of the model remains high even after performing many augmentations and it classifies the COVID-19 class perfectly almost every time.

A similar result is visible for the three-way classifier as shown in Table X, where it is seen that the performance of the model on the Healthy class deteriorates drastically, but decreases at a slower pace for the other two classes. The poor performance over the Healthy class in turn affects the overall performance of the model which shows a noticeable decay as the augmentations become more complex.

Among the various methods to train the classifier, it is observed that the metrics are not the best for dense models when trained from scratch with the limited amount of data available. Thus transfer learning was proven to be the perfect choice for developing state-of-the-art models.

## V. Conclusion

In this paper, we analyzed two methods of classifying the chest X-rays of a patient to detect COVID-19 using the concept of transfer learning. By comparing the two methods of classification, it can be concluded that the two-way classification model performs better and shows metrics which are unmatched by any other model. Of all the models tested, the VGG-19 yielded the best results. This model can act as a crucial and fast method to detect COVID-19 in patients thus resulting in quicker and more appropriate treatment.

## ACKNOWLEDGMENT

The authors would like to thank Mr. Kapil Mirchandani and Mr. Kushal Chordiya for their help and support in making this work possible.

## REFERENCES

[1] Coronavirus: Which countries have confirmed cases?: https://www.aljazeera.com/news/2020/01/countries-confirmed-cases-coronavirus-200125070959786.html.

[2] COVID-19 image data collection: https://github.com/ieee8023/covid-chestxray-dataset.

[3] Chest X-Ray Images (Pneumonia): https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia.

[4] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, U. Rajendra Acharya, "Automated Detection of COVID-19 cases using Deep Neural Network with X-ray images," Computers in Biology and Medicine, Volume 121, June 2020, 103792.

[5] Halgurd S. Maghdid, Aras T. Asaad, Kayhan Zrar Ghafoor, Ali Safaa Sadiq, Muhammad Khurram Khan, "Diagnosing COVID-19 Pneumonia from X-ray and CT images using deep learning and transfer learning algorithms, " arXiv:2004.00038.

[6] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition," arXiv:1512.03385.

[8] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv:1704.04861.