Ahsanullah University of Science and Technology

Department of Computer Science and Engineering

CSE4238: Soft Computing Lab

Spring 2021

Final Project Report

A Deep Learning Approach for COVID-19 and Viral Pneumonia Screening with X-ray Images

---

Submitted To

H M Zabir Haque Assistant Professor

Nibir Chandra Mandal Lecturer

Department of Computer Science and Engineering

Submitted by

Takia Maliha	170204037
Tahiya Ahmed Chowdhury	170204048
Shreya Chakraborty	170204050

Lab Section: A2

Group: 5

14 March, 2022

Contents

1 Introduction

Covid-19 virus, which has emerged in the Republic of China in an undetermined cause, has affected the whole world quickly. It is important to detect positive cases early to prevent further spread of the outbreak. In the diagnostic phase, radiological images of the chest are determinative as well as the RT-PCR (Reverse Transcription-Polymerase Chain Reaction) test. Chest X-Ray image analysis can be done and infected individuals can be identified quickly by convolutional neural network. The experimental results are encouraging in terms of the use of computer-aided in the field of pathology. It can also be used in situations when RT-PCR tests are insufficient. Employing a deep learning approach to this problem can help increase efficiency in detecting positive cases of the disease. While most of the people infected with the COVID-19 experienced mild to moderate respiratory illness, some developed a deadly pneumonia. There are assumptions that elderly people with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, renal or hepatic diseases and cancer are more likely to develop serious illness.

Many studies have been conducted on the automating of the detection of the disease from chest x-ray images using artificial intelligence. In this study, chest X-ray images created using more than one data set were studied. The proposed deep learning model produced an average classification accuracy of 90% and F1-Score of 88.6% after performing 5-fold cross-validation on a multi-class dataset consisting of COVID-19, Viral Pneumonia, and normal X-ray images.

2 Motivation

Approximately 268 Million people are affected and 5.29 Million have died infected by COVID-19. To optimize resource usage and clinical workflow, development of quick and specialized strategies is necessary. We aim to develop a model that will assist clinicians in detecting COVID-19 and COVID-19 Pneumonia from a patient's x-ray, which is less time consuming and quicker. As we have known from various medical research papers visual signs of COVID-19 and Viral Pneumonia can be difficult to spot even to experts. For this, wrong assumptions regarding medication can be very dangerous. This is what motivated us to take this project to make the process easier for an experienced specialist to identify a sub-type of Pneumonia in affected patients. We believe this can put a huge impact in medical field.

3 Methodology

3.1 Description of the Model

The proposed model contains five convolutional layers. Each convolutional layer is followed by batch normalization and 2x2 max pooling filters, as well as dropout of 0.25. A fully connected layer with 512 neurons processes the final convolutional layer, which is followed by a final layer with three neurons representing each X-ray category. ReLU was used as the activation function for each layer, and softmax was used for the final dense layer. Figure 1. depicts the model architecture, whereas Figure 2. lists the specifics for each layer. Each convolutional layer's number of filters was gradually increased from 32 to 64 to 128, then returned to 64 and 32, with 3x3 strides for each layer. To standardize the inputs, batch normalization was utilized, and max pooling was used to generate spacial variation, which was used to account for different appearances.

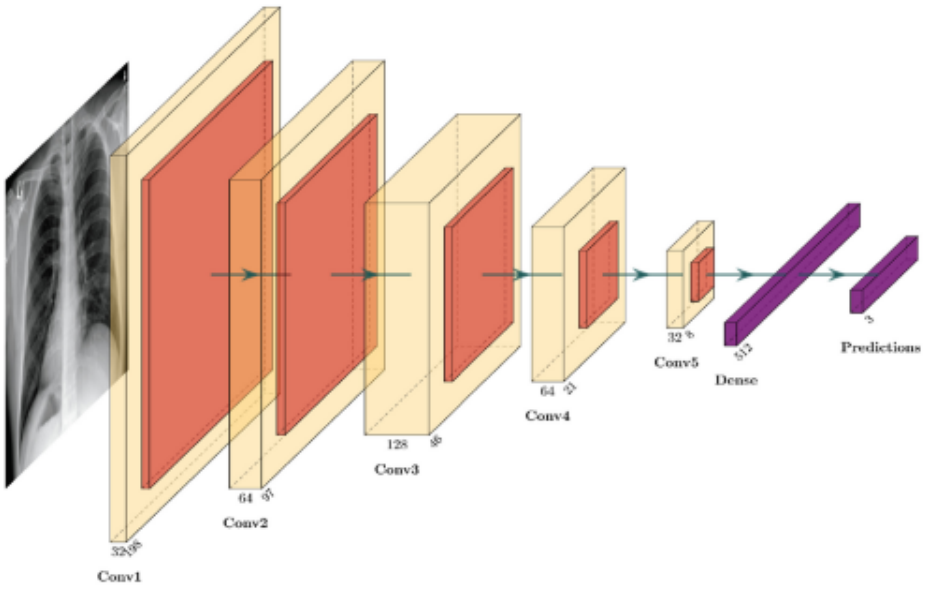


Figure 1: Model Architecture

Layer (type)	Output Shape	Param #
conv2d_7 (Conv2D)	(None, 198, 198, 32)	896
batch_normalization_8 (Batch	(None, 198, 198, 32)	128
max_pooling2d_7 (MaxPooling2	(None, 99, 99, 32)	0
dropout_8 (Dropout)	(None, 99, 99, 32)	0
conv2d_8 (Conv2D)	(None, 97, 97, 64)	18496
batch_normalization_9 (Batch	(None, 97, 97, 64)	256
max_pooling2d_8 (MaxPooling2	(None, 48, 48, 64)	0
dropout_9 (Dropout)	(None, 48, 48, 64)	0
conv2d_9 (Conv2D)	(None, 46, 46, 128)	73856
batch_normalization_10 (Batc	(None, 46, 46, 128)	512
max_pooling2d_9 (MaxPooling2	(None, 23, 23, 128)	0
dropout_10 (Dropout)	(None, 23, 23, 128)	0
conv2d_10 (Conv2D)	(None, 21, 21, 64)	73792
batch_normalization_11 (Batc	(None, 21, 21, 64)	256
max_pooling2d_10 (MaxPooling	(None, 10, 10, 64)	0
dropout_11 (Dropout)	(None, 10, 10, 64)	0
conv2d_11 (Conv2D)	(None, 8, 8, 32)	18464
batch_normalization_12 (Batc	(None, 8, 8, 32)	128
max_pooling2d_11 (MaxPooling	(None, 4, 4, 32)	0
dropout_12 (Dropout)	(None, 4, 4, 32)	0
flatten_2 (Flatten)	(None, 512)	0
dense_2 (Dense)	(None, 512)	262656
batch_normalization_13 (Batc	(None, 512)	2048
dropout_13 (Dropout)	(None, 512)	0
predictions (Dense)	(None, 3)	1539

Figure 2: Model Architecture

The developed model has 453,027 parameters, with a loss function of categorical cross entropy, an optimizer of rmsprop, and a batch size of 5. The learning rate was initially set to 0.0001, but if the validation accuracy does not increase for two steps, we reduced the learning rate by a factor of 0.5 to prevent the accuracy from plateauing during training.

3.2 Abstraction of the Model

In this approach, We tried to classify a patient's medical condition such as COVID-19 or Viral Pneumonia by using radiography images specifically Chest X-Rays. At the first step, we collected the Chest X-Ray images and applied augmentation before passing it to the model. Annotating images is required to make the model understand about the difference between the X-Ray images. In the training process, the model learns how to differentiate between the Covid-19 X-rays, Viral Pneumonia X-rays and Normal X-rays by using the annotated images. After the training process, the model can predict and classify different Chest X-Ray images provided which may or may not be seen anytime by the model. Figure 3. depicts the abstract workflow of the methodology.

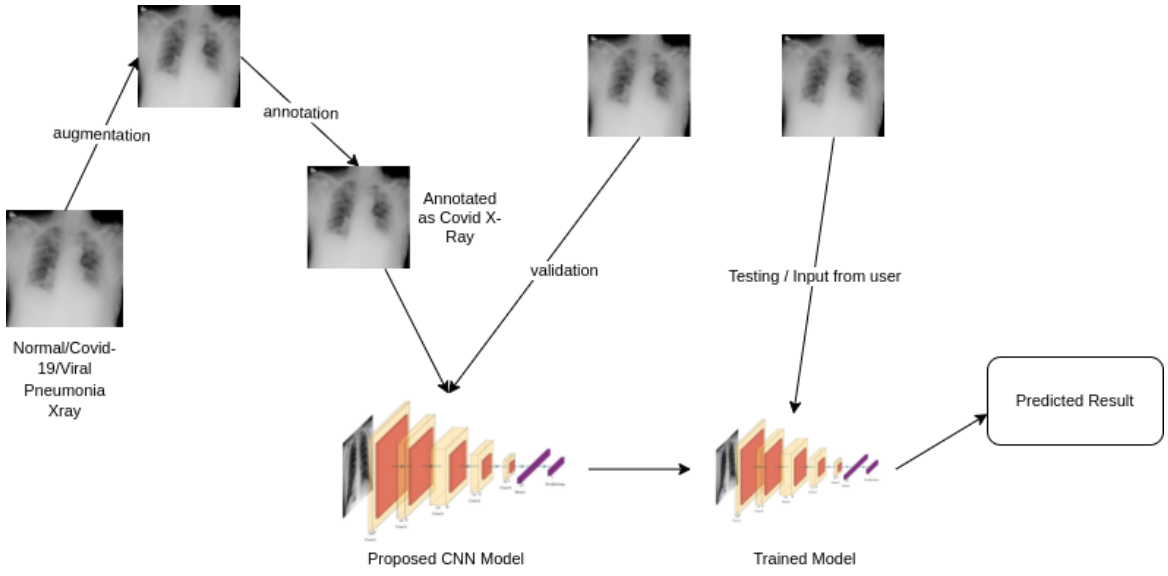


Figure 3: Abstraction of the Model

4 Experiments

4.1 Dataset

The X-ray pictures used in this research obtained from the Kaggle dataset, which was produced in partnership with medical doctors to create an open-source dataset to aid researchers in better understanding COVID-19. The X-ray pictures were

extracted from a variety of public sources, including publications and indirect collection from hospitals and clinicians. There are 15,153 images in the dataset used in this research, including 3,616 COVID-19, 1,345 Viral Pneumonia, and 10,192 Normal X-rays. All the images are in Portable Network Graphics (PNG) file format and resolution are 299x299 pixels. Figure 4. shows the categorized count of the dataset.

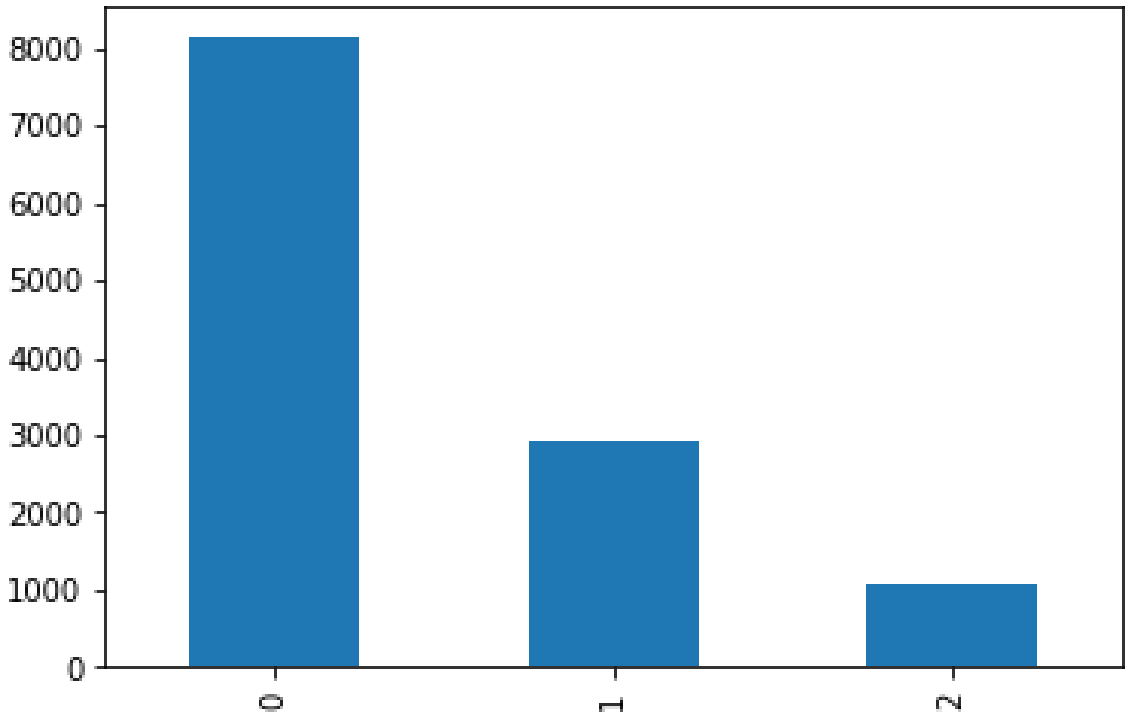


Figure 4: Categorized Count of the dataset. (Normal: 0, Covid-19: 1, Viral Pneumonia: 2)

4.2 Experimental Design

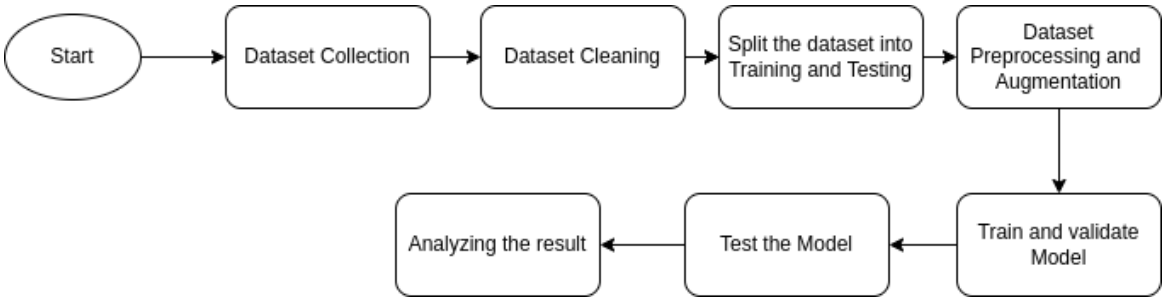


Figure 5: Workflow of the study

Figure 5 depicts the workflow of the experiment. The model was trained and tested using a dataset that included 10192 Normal X-rays, 3616 COVID-19 X-rays, and 1345 Viral Pneumonia X-rays, each shrunk to 200 by 200 pixels in height and width. Sixty percent of the data was used for training, while 20% was used

for validation and remaining 20% was used for testing. To further evaluate the model, we performed 5-fold cross-validation on the entire dataset of 15,153 X-ray images, where 80% of the dataset was used for training and 20% percent was used for testing. Of the training set, 20% was used as the validation set, with 9698 total images used for training, 2424 images used for validation, and 3031 images used for testing for each fold of cross-validation. A visual representation of this procedure is shown in Figure 6.

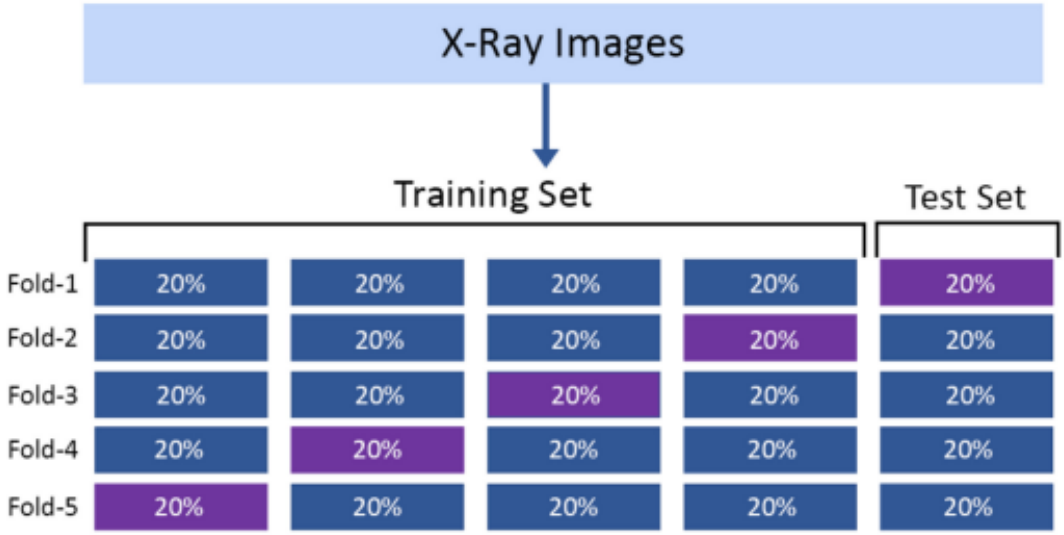


Figure 6: Schematic representation of the 5-fold cross-validation procedure.

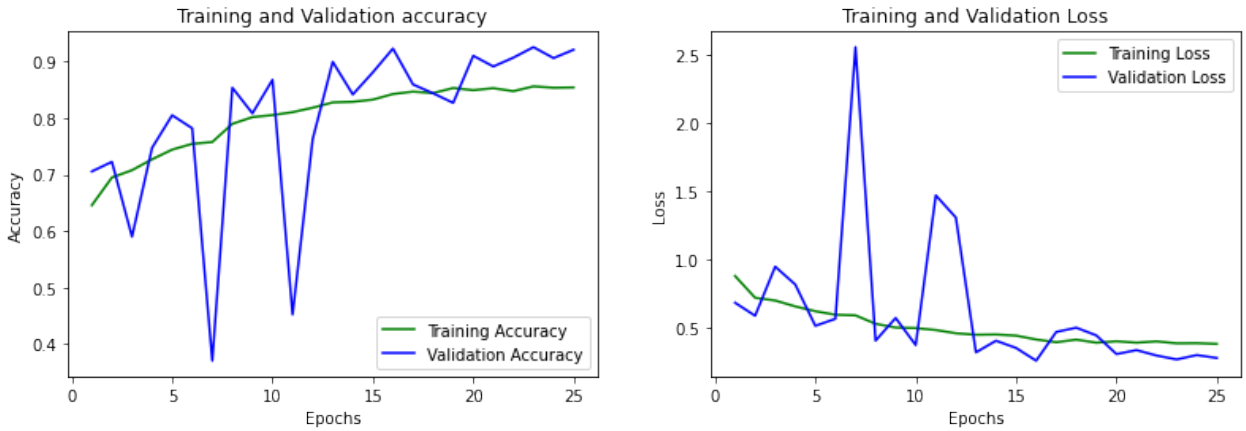


Figure 7: Training, validation accuracy, and loss of the model after 25 epochs.

4.3 Results and Performance Metrics

The proposed model attained a 92 percent accuracy on the testing set, with an F1-score of 90 percent, for the initial method. Figure 8 shows the confusion matrix of the model's performance, which shows that the model can successfully identify between the incidence of diseases with high precision and recall for each

X-ray category. With an average accuracy of 90 percent and an F-1 Score of 88.6 percent, the model performed similarly for each fold of cross-validation, as shown in Table 1 and 9. Figure 7 depicts the training-validation accuracy and loss.

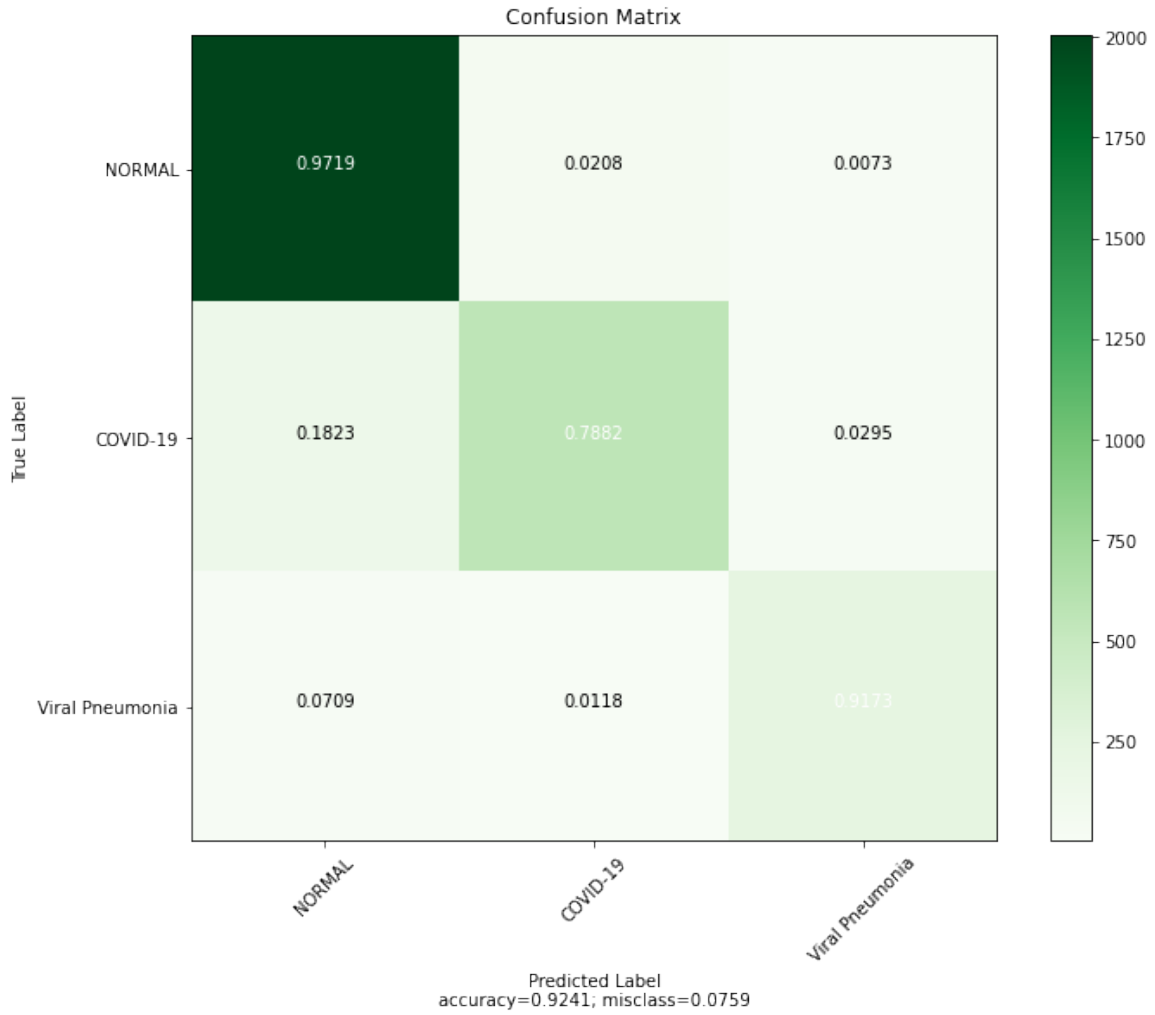


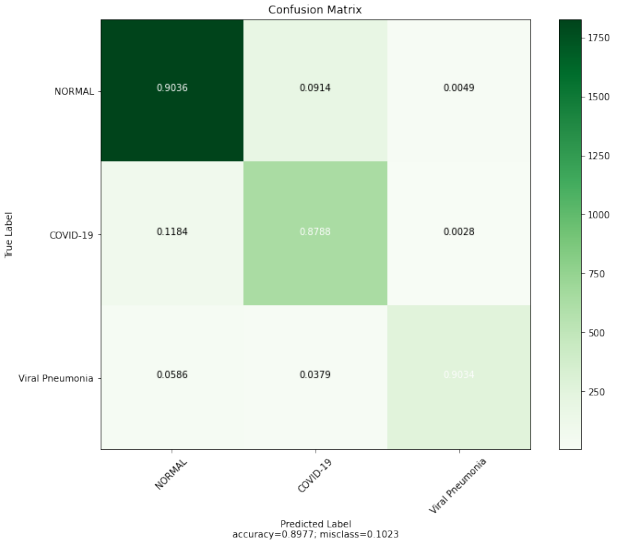
Figure 8: Confusion matrix for model performance on testing set.

Fold	Accuracy	Precision	Recall	F1-Score
Fold-1	90	89	90	89
Fold-2	82	90	74	77
Fold-3	92	91	90	93
Fold-4	93	91	91	93
Fold-5	93	91	91	91
Average	90	90.4	87.2	88.6

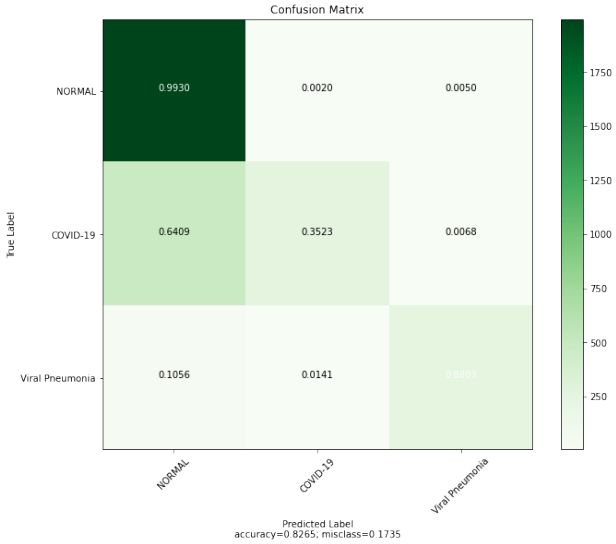
Table 1: Model Performance (%) for Each Fold of Cross-validation

4.4 Comparison

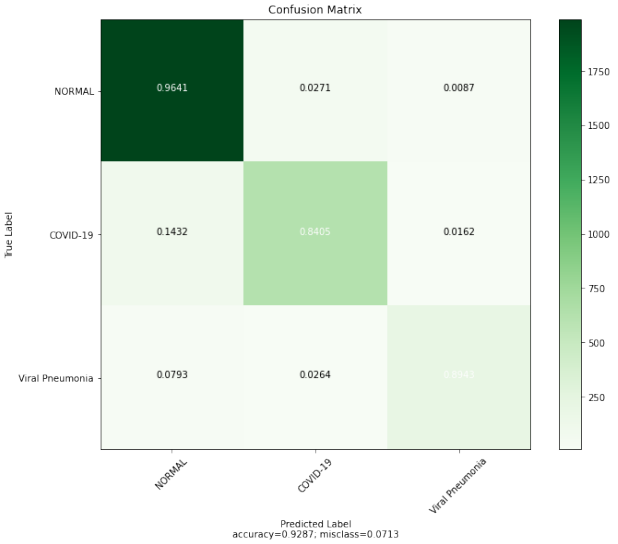
Table 2. depicts the comparison between our result and the original paper result. In case of K-Fold Cross Validation, we used the average value of the performance



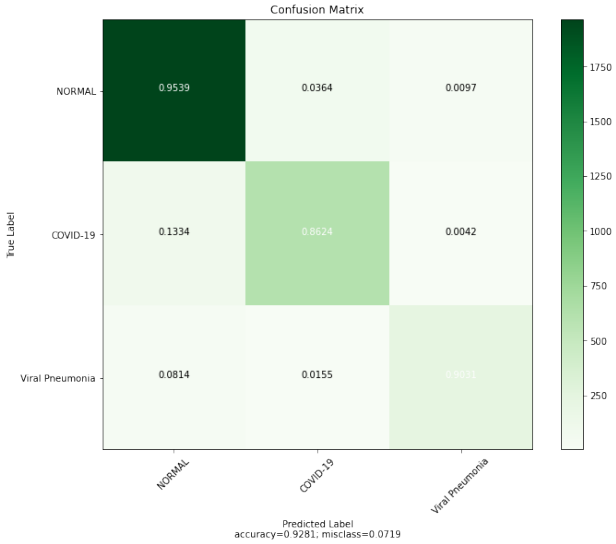
(a) Fold-1



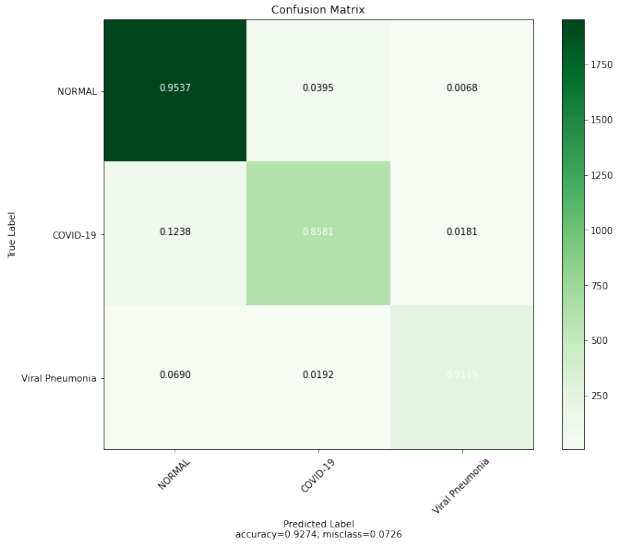
(b) Fold-2



(c) Fold-3



(d) Fold-4



(e) Fold-5

Figure 9: Confusion matrices for each fold of the cross-validation procedure.

metrics to compare the results. In the original paper, they trained the model on a small and balance dataset whereas we trained the model with a large and imbalance dataset because of the huge addition of the Chest X-Ray images in the dataset.

Performance Metrics	Our Result	Original Paper Result
Accuracy	90	90.64
Recall	87.2	89
Precision	90.4	92
F1-Score	88.6	89.8

Table 2: Comparison between our result and original paper result

4.5 Discussion

In the original paper, they performed the experiment with a dataset of 550 Viral Pneumonia Chest X-Ray images, 550 Normal Chest X-Ray images and 289 Covid-19 Chest X-Ray images which yielded a better result than we did because the dataset was balanced. To categorize different lung infections such as COVID-19 and Viral Pneumonia, we used a dataset that included 10,192 Normal Chest X-Ray images, 3,616 COVID-19 Chest X-Ray images, and 1,345 Viral Pneumonia Chest X-Ray images. We used the whole imbalanced dataset to train our model, resulting in a minor reduction in performance metrics for our model. When we dissect our model's prediction by different category classifications, we can see that it performs poorly for one category, which is expected because an imbalance dataset never contains the same proportion of images from each category, resulting in false predictions when testing for that category. Table 3. depicts the performance metrics of each category. For example in fold-2, we can see high precision and low recall value for Covid-19 samples which means it's returning the correct predictions for Covid-19 samples but the total amount of returned value is low.

Performance Metrics	Normal	Covid-19	Viral Pneumonia
Accuracy	90	87	90
Precision	95	76	96
Recall	90	88	90
F1-Score	92	82	93

(a) Fold-1

Performance Metrics	Normal	Covid-19	Viral Pneumonia
Accuracy	99	35	88
Precision	80	97	94
Recall	99	35	88
F1-Score	89	52	91

(b) Fold-2

Performance Metrics	Normal	Covid-19	Viral Pneumonia
Accuracy	96	84	89
Precision	94	91	89
Recall	96	84	87
F1-Score	95	87	88

(c) Fold-3

Performance Metrics	Normal	Covid-19	Viral Pneumonia
Accuracy	95	86	90
Precision	94	89	91
Recall	95	86	90
F1-Score	95	87	91

(d) Fold-4

Performance Metrics	Normal	Covid-19	Viral Pneumonia
Accuracy	95	86	91
Precision	95	88	90
Recall	95	86	91
F1-Score	95	87	90

(e) Fold-5

Table 3: Categorized Performance Metrics for Each Fold