Fast and Accurate Diagnosis of COVID-19 from X-Ray Scans using Deep Learning

Ritvik
2018csb1115
Rohit Tuli
2018csb1116
Praful Gupta
2018csb1112

Indian Institute of Technology, Ropar
Rupnagar, Punjab
India

Abstract

Coronavirus Disease (COVID-19) is a novel infectious disease caused by a newly mutated strain of Coronavirus and has infected 1 Million+ people from 196 Countries. The lack of a fast and accurate diagnosis has been one of the major hurdles in containing the spread of this disease. It has been proven by recent research that the manifestations of computed tomography(CT)imaging of COVID-19 had their own characteristics, which are different from other types of viral pneumonia, such as Influenza-A viral pneumonia. These differences are subtle and difficult to detect by naked eye but can be detected using Computer Vision. Here we have developed a deep learning-based lung X-Ray Scan diagnosis system to detect the patients with COVID-19, which can automatically extract radiographic features of the novel pneumonia and correctly label whether the given patient is infected or not.

1 Rationale of Study

Yvonne Doyle, the medical director and the director of health protection for Public Health England, tells The Scientist in an email that once a sample is received by a laboratory, it takes 24–48 hours to get a result via the current RT-PCR kits. This takes up a lot of medical resources and the precious early stage time is lost. Moreover, due to the high infectability of the virus, the number of infected patients is increasing at a high rate, with a total of over 1,400,000 cases so far. Hospitals have neither the resources nor the capacity to attend so many patients at once. Hence, there is an immediate need for faster and a more accurate method for correctly diagnosing the patients who show the symptoms of the virus, such as pneumonia. Chest X Ray Scans are very detailed, and inexpensive compared to physical testing kits and is not prone to shortages. A Chest X-Ray takes lesser time to generate, and our proposed model also takes a few seconds for analysis. On the other hand, traditional PCR tests make take a few days to generate results.

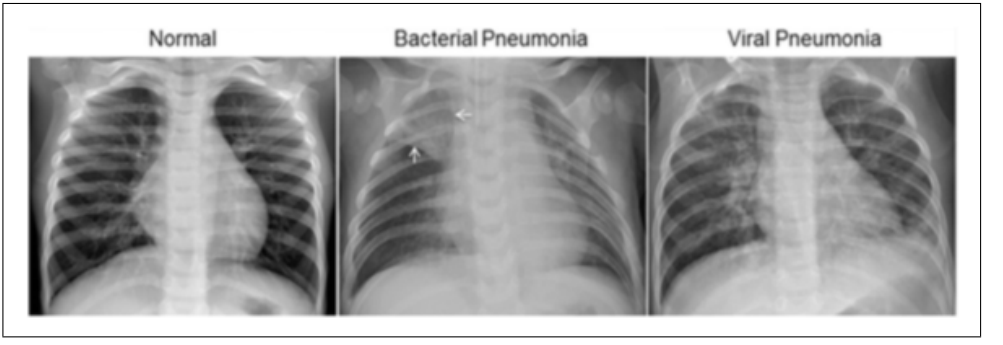

Figure 1: Difference in X-Ray Scans of healthy lungs, Bacterial Pneumonia infected lungs, Viral Pneumonia infected lungs

2 Objectives and Methodology

Most symtoms of coronavirus are also common in Common Cold, Flu, Pneumonia etc, due to which the conventional testing time is higher.

To effectively classify patients into Common Cold / Pneumonia category or Coronavirus category we are proposing a Convolutional Neural Network Based Deep Learning Model, which can differenciate between the two classes in a fast and accurate manner.

To achieve this we have created two Deep Learning models. The first model takes Chest X-Ray of the patient as input and classifies it into Pneumonia or Healthy. If the first model predicts Pneumonia, then the second model predicts whether the pneumonia is due to COVID-19 or from some other bacteria / virus.

3 Dataset Collection and Info

We used two different datasets for both models. For the Pneumonia classification model, we have used Chest X Ray Images dataset available at Kaggle. The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal).

The dataset contains chest X-ray images (anterior-posterior) selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care. All the images present in the dataset have been pre-validated by physicians.

For the Covid-19 Model we are using a Chest X-Ray/ CT Scan Dataset which is made freely available on Github by Joseph Paul Cohen. It contains scans which are classified into Pneumonia only and Covid 19 categories. It has 92 Covid Positive and 92 Covid Negative scans. As this dataset is small, we will be trying to find more data to train our model better. Presently we included some random scans from the previous pneumonia dataset to train the model better.

Model	Accuracy
Pneumonia	89.49%
Covid-19	91.38%
Overall	81.77%

Table 1: Accuracy for the models

4 Current Status

Presently we have created both the models in Python using various libraries such as Keras, Pandas, Numpy, Matplotlib etc. They are both sequential models consisting of 3 2D-Convolution Layers, 2 Max-Pooling Layers and 2 Dense Layers. ReLu and Sigmoid have been used as Activation Functions, Binary Cross-entropy Loss Function and Adam as optimiser. We have used 60:30:10 Ratio to split dataset for Training, Testing and Validation. So far using this configuration we have achieved accuracy, with loss at 0.1048, validation loss at 0.2570 and validation accuracy at 91.38 for the second model.

5 Deliverables

1. Google colab file containing our source code: here
2. Datasets used:
   (a) Covid-19 detector training dateset
   (b) Pneumonia detector training dataset

6 References

1. Deep learning Enables Accurate Diagnosis of Novel Coronavirus , Song Ying 2020.02.14.20023028
2. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography , Jun Chen 2020.02.25.20021568
3. A deep learning algorithm using CT images to screen for CoronaVirus Disease , Shuai Wang 2020.02.23.20026930

```
Epoch 1/10
163/163 [==============================] - 90s 551ms/step - loss: 0.3855 - acc: 0.8196 - val_loss: 2.5884 - val_acc: 0.3783
Epoch 2/10
163/163 [==============================] - 82s 506ms/step - loss: 0.2928 - acc: 0.8735 - val_loss: 1.4988 - val_acc: 0.6284
Epoch 3/10
163/163 [==============================] - 81s 498ms/step - loss: 0.2581 - acc: 0.8963 - val_loss: 1.1351 - val_acc: 0.3970

Epoch 00003: ReduceLROnPlateau reducing learning rate to 0.0003000000142492354.
Epoch 4/10
163/163 [==============================] - 81s 495ms/step - loss: 0.2027 - acc: 0.9197 - val_loss: 0.3323 - val_acc: 0.8463
Epoch 5/10
163/163 [==============================] - 81s 500ms/step - loss: 0.1909 - acc: 0.9294 - val_loss: 0.2530 - val_acc: 0.9139

Epoch 00005: ReduceLROnPlateau reducing learning rate to 9.000000427477062e-05.
Epoch 6/10
163/163 [==============================] - 81s 495ms/step - loss: 0.1639 - acc: 0.9423 - val_loss: 0.3316 - val_acc: 0.8834
Epoch 7/10
163/163 [==============================] - 80s 492ms/step - loss: 0.1625 - acc: 0.9387 - val_loss: 0.2403 - val_acc: 0.8919

Epoch 00007: ReduceLROnPlateau reducing learning rate to 2.700000040931627e-05.
Epoch 8/10
163/163 [==============================] - 80s 490ms/step - loss: 0.1587 - acc: 0.9423 - val_loss: 0.2732 - val_acc: 0.9122
Epoch 9/10
163/163 [==============================] - 81s 496ms/step - loss: 0.1575 - acc: 0.9419 - val_loss: 0.2605 - val_acc: 0.9054

Epoch 00009: ReduceLROnPlateau reducing learning rate to 8.100000013655517e-06.
Epoch 10/10
163/163 [==============================] - 80s 490ms/step - loss: 0.1633 - acc: 0.9423 - val_loss: 0.2589 - val_acc: 0.9155
```

Figure 2: Test and Validation accuracy per epoch - Pneumonia detection model

```
CONFUSION MATRIX ------------------
[[191  43]
 [ 13 377]]

TEST METRICS ----------------------
Accuracy: 91.02564102564102%
Precision: 89.76190476190476%
Recall: 96.66666666666667%
F1-score: 93.08641975308642

TRAIN METRIC ----------------------
Train acc: 94.23
```

Figure 3: Performance Metrics - Pneumonia detection model

```
Epoch 1/10
5/5 [==============================] - 8s 2s/step - loss: 0.1042 - acc: 0.9545 - val_loss: 0.1944 - val_acc: 0.9138
Epoch 2/10
5/5 [==============================] - 8s 2s/step - loss: 0.1278 - acc: 0.9545 - val_loss: 0.2377 - val_acc: 0.9483
Epoch 3/10
5/5 [==============================] - 8s 2s/step - loss: 0.0656 - acc: 0.9740 - val_loss: 0.2866 - val_acc: 0.9310
Epoch 4/10
5/5 [==============================] - 8s 2s/step - loss: 0.1257 - acc: 0.9220 - val_loss: 0.7126 - val_acc: 0.7414
Epoch 5/10
5/5 [==============================] - 8s 2s/step - loss: 0.1745 - acc: 0.9415 - val_loss: 0.2355 - val_acc: 0.9310
Epoch 6/10
5/5 [==============================] - 7s 1s/step - loss: 0.0919 - acc: 0.9545 - val_loss: 0.2286 - val_acc: 0.8793
Epoch 7/10
5/5 [==============================] - 8s 2s/step - loss: 0.1308 - acc: 0.9415 - val_loss: 0.2364 - val_acc: 0.9138
Epoch 8/10
5/5 [==============================] - 7s 1s/step - loss: 0.0638 - acc: 0.9740 - val_loss: 0.3684 - val_acc: 0.9310
Epoch 9/10
5/5 [==============================] - 8s 2s/step - loss: 0.3001 - acc: 0.8960 - val_loss: 0.3037 - val_acc: 0.8966
Epoch 10/10
5/5 [==============================] - 7s 1s/step - loss: 0.1048 - acc: 0.9480 - val_loss: 0.2570 - val_acc: 0.9138
```

Figure 4: Test and Validation accuracy per epoch - Covid-19 detection model