

# Data-set preparation

- Importing the libraries
- Loading the data and moving the data to google colab storage (for faster executions)
- creating train and test data frame.
- Labeling the binary data as 0's and 1's (COVID = 0 and non-COVID = 1)
- Splitting the training data into training and validation data.

Total training images = 1400

Number of images after the split of 0.15 of training data

Training images = 1190

Validation images = 210

- I am using the same image size 100\*100
- Data generation: generating batches for training using the flow from the data frame function

## Architecture used

- creating the model using 4 convolution layers of sizes (16, (3,3)), (32, (3,3)), (64, (3,3)), (64, (3,3)) respectively.
- I have used Maxpooling (2,2) for all the convolution layers.
- 'Relu' Activation function is used in the hidden layers, and the 'sigmoid' activation function is used as an outer layer.
- As I faced overfitting (high accuracy for the validating data and low accuracy for the test data) to reduce this, I have introduced dropout (0.4) in the first three convolution layers.
- Training the data, and Prediction on test data
- confusion matrix

```
array([[297,    5],
       [273,   11]])
```

- printing the report

| precision    | recall | f1-score | support |     |
|--------------|--------|----------|---------|-----|
| 0            | 0.52   | 0.98     | 0.68    | 302 |
| 1            | 0.69   | 0.04     | 0.07    | 284 |
| accuracy     |        |          | 0.53    | 586 |
| macro avg    | 0.60   | 0.51     | 0.38    | 586 |
| weighted avg | 0.60   | 0.53     | 0.39    | 586 |