Covid19 Detection

**Introduction:**

Chest X-Ray is one of the methods used to identify whether a subject is COVID-19 positive or negative. The project aims to classify X-Ray images, whether it is positive or negative based on training a model using CNN model (Deep Learning) on a pre-defined data.

**Accessing the data:**

The project and the results are based 29000 Images, which can be accessed through this link form Kaggle. <https://www.kaggle.com/datasets/andyczhao/covidx-cxr2>

Notes to be taken care of, while running the code:

**Steps of the project:**

**Step 1: Data collection:**

In the downloaded folder, the followings were found:

Train Folder: it contains all the images of the train dataset, where the images named by patient id only (Not mentioned positive or negative).

Test Folder: it contains all the images of the train dataset, where the images named by File name only (Not mentioned positive or negative).

Train and test txt file contain (Patient ID, File name, Class, Data source).

**Step 2: data preprocessing:**

First, we need to connect the file name, which is associated with the name of the images, with the txt file, so that we know the class of each image whether it is positive or negative.

At this stage, we have imbalance between positive and negative X-Ray images. Hence, Random Under Sampler function was used to balance the 2 classes in the dataset.

Next, the dataset is divided into train, validation and test sets.

Then, ImageDataGenerator function from Keras were used to normalize and resizing the data and to connect the class of each file name of the txt file with the image, so that we get to know the class of each image.

**Step 3: Building and training the model:**

A CNN Deep Learning model is built to classify X-Ray images and it was trained using Adam Optimizer.

**Step 4: Model prediction:**

At this stage, the test set is used to make predictions based on the trained model.

**Step 5: Model Evaluation:**

Several Model Evaluation methods are involved to get a clear vision of the performance of the Model such as Accuracy, Precision, Recall, F1-Score, Confusion Matrix.

**Step 6: Model Tuning:**

Error Type I (False Positives) means that the model predicts the Negative images as Positive. Where Error Type II (False Negatives) means that the model predicts the Positive images as Negative. Both Errors in the Confusion Matrix are important to be considered. But our case, False negative (Error Type II) is more crucial to be considered and has to be decreased.

Several Model Tuning Methods are involved to improve the performance of the model such as Dropouts, Adding Hidden Layers, increase and decrease of the Number of Neurons, Data Augmentation, Regularization etc. After trying different methods, the following results are achieved based on:

* 2 Convolutional Layers (32 and 64 Neurons).
* 1 Hidden Layer (500 Neurons).
* Learning Rate of 0.001
* No Dropout.
* No Data Augmentation.
* 10 Epochs.

