CT Scan Image Classification

Project Description:

In this project, we aim to develop an AI solution to classify CT scan images as COVID-19 positive (COVID) or negative (non-COVID).

The COVID-19 pandemic has highlighted the need for efficient diagnosis, and CT scans have shown promise in aiding detection. Leveraging deep learning techniques, our project focuses on training a model to analyze CT scan images and provide accurate predictions for COVID-19 infection.

The developed model can assist healthcare providers by offering an additional diagnostic resource, potentially contributing to faster and more accurate patient management during a critical time

Dataset Description:

1)This dataset contains 1252 CT scans that are positive for SARS-CoV-2 infection (COVID-19) and 1230 CT scans for patients non-infected by SARS-CoV-2, 2482 CT scans in total.

2)These data have been collected from real patients in hospitals from Sao Paulo, Brazil.

3)The aim of this dataset is to encourage the research and development of artificial intelligent methods which are able to identify if a person is infected by SARS-CoV-2 through the analysis of his/her CT scans.

Approach to the Problem

Data Preprocessing:

1)Downloaded the data.zip file and extracted the COVID and Non-COVID directories.

2)Loaded the images into list.

3)Resized the image into fixed sized of 224 X 224 and then converting images to RGB mode.

4)Normalized the pixel values of images

Data Augmentation:

1)Applied data augmentation techniques on images using Keras preprocessing “ImageDataGenerator” function.

2)Included rotation, shifting, shear, zoom, and flipping to increase data diversity.

Data Generator & Splitting the data:

Combining COVID and Non-COVID data with corresponding labels. 2)Splitting the data into Training and validation sets.

Model Building:

1)We used ResNet50 Model architecture which offers better performance and capacity to capture complex features.

2)Adding custom layers like Average pooling 2D layer, Dense Layers

3)Frozing the layers of the ResNet base model to avoid overwriting learned features. 4)Compiled the model using binary cross-entropy, Adam optimizer and performance metrics as accuracy.

Model Training:

1)Defined early stopping and model checkpoint callbacks.

2)Trained the model using fit method with data generators for training and validation sets.

3)Saved the best model.

Model Evaluation & Prediction:

1)Loaded the best saved model.

2)Evaluated model using metrics like accuracy, precision, recall, F1-Score, and confusion matrix on validation set.

3)Loaded a test image, preprocessed it and used model to predict the class whether

4)COVID or Non-COVID of the test image.

Fine Tuning the best model:

1)Firstly, we fine-tuned the model by unfreezing the last 20 layers of pretrained ResNet50 model & running for the 500 epochs.

2)Then again, we adjusted to 500 epochs and by unfreezing last 40 layer's

3)Both times the model resulted 87.26%

4)Accuracy i.e., almost 25% increase in accuracy of ResNet50 Model.

Last Approach of the Project:

Tech Stack Used:

1. Python, Google Collab, Jupyter Notebooks & VS code
2. TensorFlow & Keras.
3. Visualization libraries like
4. Matplotlib.Pyplot, and Seaborn.
5. NumPy, Pandas, PIL (Pillow module), and Sklearn