DETECTING LUNG DISEASES USING CNN

EXECUTIVE SUMMARY

I have presented CNN designed to classify lung disease. A series of convolutional layers, pooling layers, along with other layers like dropout layers, batch normalization layers, etc. are used in extracting features related to the input images which are used to classify the images into the 4 different categories of lung which are affected by COVID-19, Lungs Opacity, Viral Pneumonia and Normal Lungs.

This project uses the deep learning to predict whether a person is affected by lungs diseases like lung opacity, Covid-19, Pneumonia by using X-ray images. The aim of the project is to predict the lung disease or a healthy lung from X-ray images using, TensorFlow, Keras, and Scikit learn. The dataset consists of 21,165 images of different diseases. I converted the dataset into array of numbers to model them. I have created a function that preprocess the images to arrays by using keras image to array method that converts image into the numerical arrays. Later I converted categorical target label to number using LabelBinarizer for the dataset. Minmax scaling is performed on the dataset to convert them into the range of 0 to 1. To perform them, the array is divided by 255, which is the maximum value any array of an image can take. Testing the performance of the model using the hidden dataset becomes necessary. So, I have split the dataset into train and test set. I have set aside 20% of the data as test set.

To use TensorFlow, my data needs to be in the format of tensors. So, I converted all the data into tensors. I have used the Image Data Generator function from the TensorFlow to do the same. I set all the parameters as default to get the original images as tensors. After various regressive modifications and analyzing the performance of each model by changing various hyper parameters, I was able to train a better performing model. I applied relu activation to each convolution layers so that all the negative values are not passed to the next layer.

I got total trainable params in model summary was 58,021,636. Same Adam optimizer was used on this model too, to train the model. After training template model and CNN models. I was able to train a better performing model with good accuracy. The model can be optimizing after only 30 epochs with a good accuracy and very less model loss. The accuracy of the model was 94.7% on train set and 84.03% on the test set. I can train model with CT scan images dataset, can provide various options in future to detect other problems of body using various medical reports.