Covid19 identification by analysis lungs' CT scan images

Problem: To classify/identify if the person is infected to SARS-CoV-2 (COVID-19) by analyzing the CT scan image of its lungs.

Dataset: The 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. These data have been collected from real patients in hospitals from Sao Paulo, Brazil.

Below steps are followed for solving the problem-

1. **Load dataset**
   Data is available in zip file. Unzip the file and mount on the python notebook directory.

2. **Resize images**
   All images has to be same size for deep learning/Resnet. Let's take the size be (224x224).

3. **Splitting data for Train, Validation & test**
   We have around 1200 images for each covid & non-covid. Split the dataset in three parts – train, validation, test in ratio (0.7,0.2,0.1) using split-folder module.

4. **Data Augmentation**
   Process the image data (data augmentation) by using variables- rescale, rotation_range, width_shift_range, height_shift_range, shear_range and zoom_range.

5. **Generate dataset**
   Generate dataset from the image data using imagedatagenerator.

6. **Model defining**
   Resnet50 model created using sequential, and layers & Dropout added.

7. **Adding callbacks**
   Add modelcheckpoints and early stopping.

8. **Compiling model**
   Compile model for binary_crossentropy loss, .0001 LR, and evaluation metrics = Accuracy, precision, recall, and AUC.

9. **Model fitting**
   Fit the model on train data for 100 epochs and step = total data/batch_size.

10. **Model evaluation**
    Model's history of metrics & loss is saved and plotted for comparison.

11. Final model **saved** for future reference.