Victor Chen

MAIS Deliverable 2 - COVID 19 CAT Scan Predictor

1. Problem Statement:
   An application that will detect COVID 19 from a CAT scan image.
2. Data Preprocessing:
   a. Dataset:
      Continued with current dataset, sourced from Kaggle at https://www.kaggle.com/bachrr/covid-chest-xray
   b. Changes:
      No major deviations from my original plan.
      Due to the complexity of the modality, I searched for the view/modality with the most images and only used that data. This ended up dropping about 59 images/values from the dataset.
   c. Methods:
      Currently I have done exploratory data analysis, analyzing which modality would provide the most data. I then proceeded to drop the features that were irrelevant to our target, such as the ID of the patient, the location of the image, the filename, the DOI, url, licenses, and credit.
      NaN values were removed, as I am not a (qualified) medical professional and cannot manually identify the images.
      Then I worked with OS to rename all the images for easier visibility. The difficult part about dropping images from the dataset is that not only do I have to drop from the dataset, but the images themselves have to be dropped as well.
3. Machine Learning Model:
   I've decided to swap over to using convolutional neural networks to analyze my images using Keras.layers, keras.models, and keras.preprocessing. After doing some more research, I believe this will be more efficient for me to learn.
   I have used Pandas, NumPy, Os, and the various Keras modules so far.
   As of right now, there are no validation methods.
   The challenges I faced were mentioned above. Data preprocessing took a while because I had to learn Os and edit and label images in tandem with the dataframe, but the rest should be smooth from now on.
4. Preliminary Results:
   The results of my work is a cleaned dataset consisting of a Pandas dataframe and labelled images, both ready to test.
5. Next Steps:
   My next steps include creating a running a model on my data, and then deploying the model on a web application. This should be feasible, as I've done a lot of research on Keras, and have already thought about the design of my web application as well as how to host it.