COMP 5413
Topics in Smart Health Informatics
Assignment 3

Marks distribution:

This assignment has a problem with five parts. Each part carries equal mark of 3.
The total assignment is worth 5 X 3 = 15 marks.

Release date: 30th July 2020

Submission deadline (D2L): 06th August 2020

Problem:

The supplementary file "Dataset.zip" contains a dataset of 1500 CT scans of various parts of the body such as abdomen, chest and head. You are required to train machine learning models in order to classify the provided images. You are allowed to use all the necessary libraries to execute the tasks.

Write a python program to perform the following tasks:

1. Load the dataset for all three classes and resize each image to (32 x 32). Apply the required pre-processing steps to employ the data into Machine Learning / Deep Learning algorithms.
2. Construct a Convolutional Neural Network (CNN) architecture from scratch to extract features from the images. (HINT: Extract features for train and test set separately. Extract the features constructed by the convolutional layers from an intermediate dense layer. Please refrain from using any pre-trained model for implementing this step)
3. Apply the K-Nearest Neighbor (KNN) algorithm to the extracted features from CNN and find the optimal value of K. The value of K can be considered as [3, 5, 7, 9]. Determine the performance of the model using an appropriate performance metric. Draw a graph of K values and their corresponding performance in order to represent your results.
4. Apply Random Forest (RF) algorithm to the extracted features from CNN. Tune at least two hyperparameters using random search. Determine the model's optimal performance, the confusion matrix, and the value of hyperparameters producing the optimal performance.
5. Report the performance of each model and explain your results. (eg. overfitting, underfitting, etc.)

For steps 3 and 4, save the models into pickle files after training and load the saved models in order to predict the class of the test data. Make sure the models are validated.

You must submit the python programs (Student-ID.py) or (Student-ID.ipynb), a report, and the saved models (Student-ID-KNN.pkl, Student-ID-RF.pkl) and screenshots of your results (Student-ID-KNN.jpg, Student-ID-RF.jpg). Your program should have sufficient comments so that the steps in the program are understandable. In the report, you must include the pre-processing steps that were applied, the graphs, the hyperparameter values, the optimal performance of each model and their explanation.