Instructions:

- The homework should be submitted via 'Microsoft Teams'. No other modes of submission will be accepted. The due date for the submission is 29th Nov 2022 (Tuesday), 3:30 PM.
- A single ZIP file ('name-ds261-hw2.zip') should be uploaded. The file should contain all codes and a latex generated report. Please use 'Python' to complete the assigned tasks. You can use inbuilt functions.
- The dataset to be used for this homework has been posted in 'Microsoft Teams'. The dataset consists of high dimensional embeddings of COVID-19 CT Scans used in homework 1. Remember that these CT scans were labelled as Normal, Mild and, Severe in Q1 of homework 1. Use the same labels for corresponding high dimensional embeddings shared in this homework.

***

Q1 **2 points**

Perform principal component analysis (PCA) and project the high dimensional CT embeddings to 100 and 50 dimensions. Also, report the reconstruction error as a plot of number of principal components.

Q2 **3 points**

Split the PCA projected embeddings into 70% training, 10% validation and 20% testing. Apply SVM with Linear kerenl and RBF kernel to classify these embeddings into Normal, Mild and Severe categories. Report the Accuracy and F1-score on training and testing set for all the classes.

***