Friday March 29- with Dr.Diggans (mentor)

Last week: Due to the original approach to explore a correlation between the entropy value of subsamples and the test accuracy of CNN models trained on those subsamples is not working well, we switched to class-by-class approach(test CNN models' performance on data **from each class**). This approach shows some promising results.

Dr.Diggans suggested some refined approaches based on the results we attained so far:

- Try NOT truncating the singular value vector (because of potential feature space information loss) and also NOT including the limits [0, float('nan')]. If still we want to truncate, then keep the truncation to around 10+ values so that we are not too restrictive.

- Plot the relative error in testing accuracy on the y axis, similar to how we altered the x-axis to be relative BSIE error. To do this, we will have to go back to the fully trained model and evaluate the accuracy on each class of the test data. Then we will have a vector of length 10 of accuracy values per class using the whole training set to train.

Professor Chen (later in the meeting):

- The current approach only provides a vague correlation (between the entropy and test accuracy) which might not be convincing. Also, we might consider other approaches because the entropy approach itself is taking too many intermediate steps which might cause information loss during the process.

Next week: practice the refined approach Dr.Diggans suggested. Consider better ways to find representative subsets to train CNN models.