Imbalanced dataset, as we don’t have the capabilities to handle imbalanced datasets as of right now, we decided to use different metrics to represent the results better.

Before outlier removal, the categories of the outliers should be checked to ensure that categories with low sample number don’t get heavily affected by the removal of a few datasets.

We should check the outliers, if they aren’t part of the minor categories, remove them

Sensitivity score

Confusion matrix English for each class