ROC Curve on the test set with 5 different train/validation splits