Protein: Q07869 - metric: balanced_accuracy

Total (train): 1504

ROC curve

- AUC_train = 0.996
- AUC_test = 0.935
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 1806)
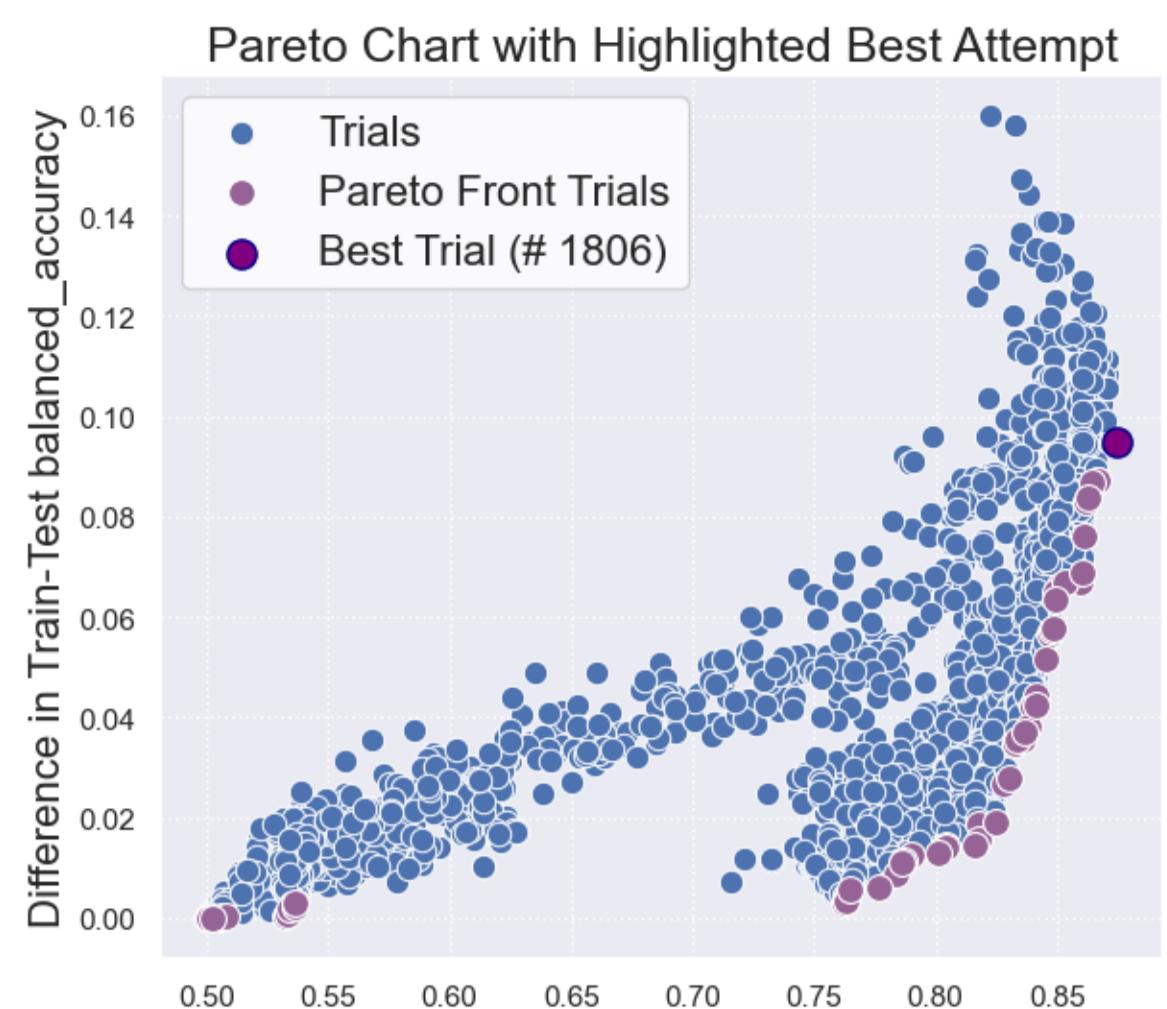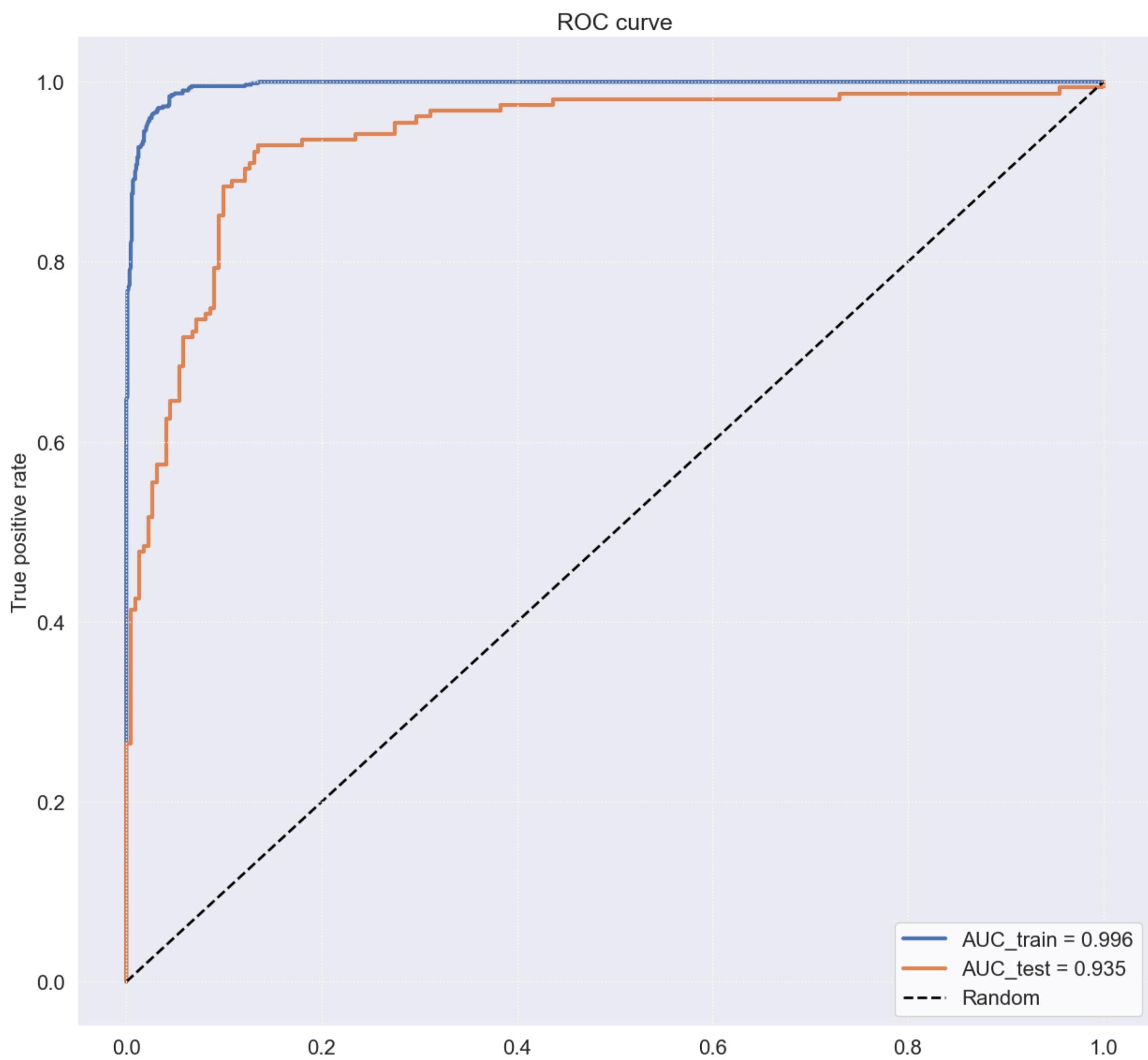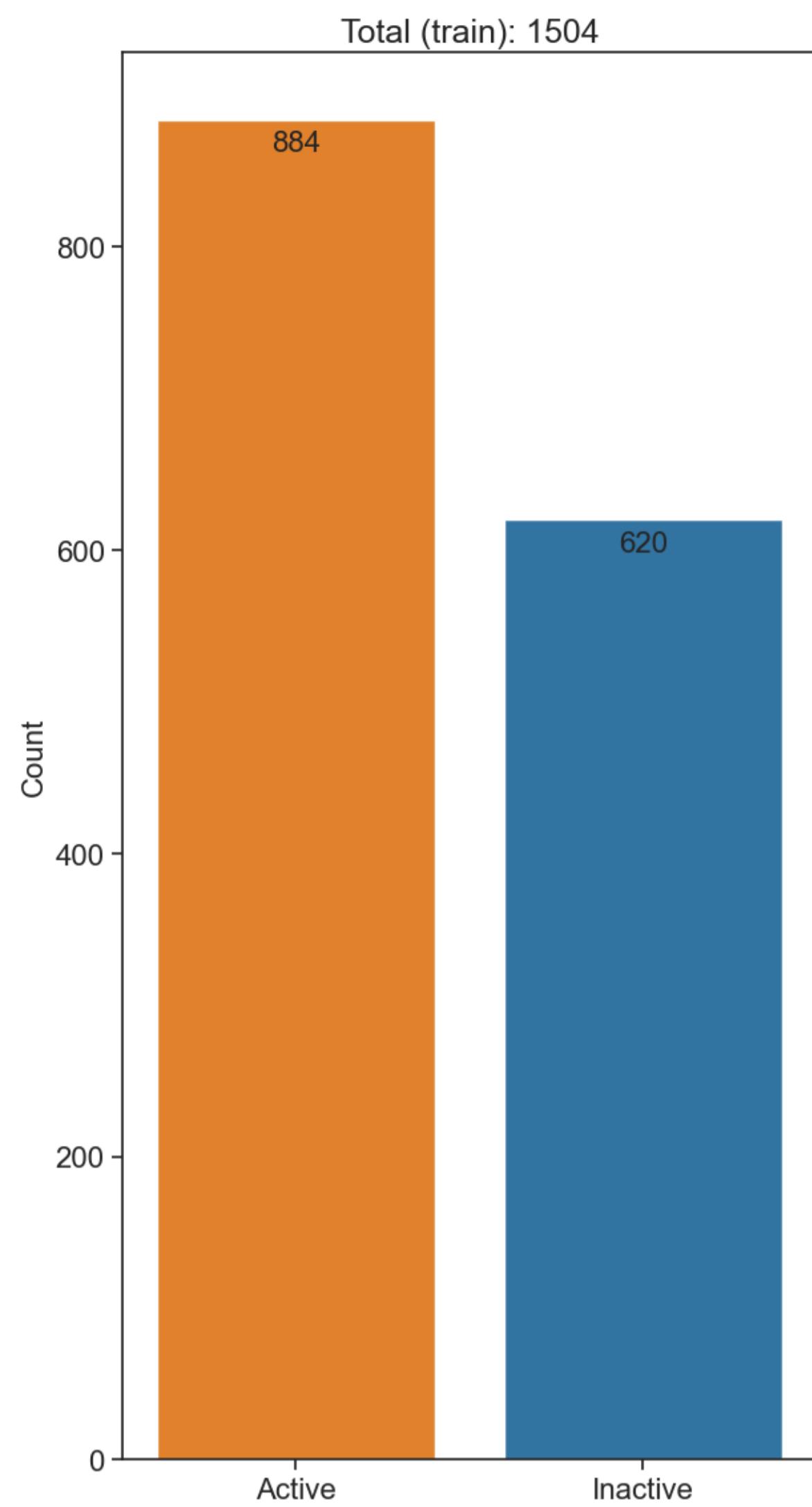
XGBoost Parameters	Value
alpha	3.8932
colsample_bytree	0.8901
gamma	0.0346
lambda	1.7306
learning_rate	0.0416
max_depth	10.0
min_child_weight	3.0
n_estimators	635.0
reg_alpha	0.7349
reg_lambda	0.0227
scale_pos_weight	1.6153
subsample	0.5057

Confusion Matrix

195	27
15	140

Confusion Matrix (nomr.)

0.88	0.12
0.097	0.9

Calibration curve

- train: 0.0304
- test: 0.0921
- perfectly_calibrated

Active/Inactive probability

Inactive zone   Active zone

type
- active (155)
- inactive (222)
- decoy (377)
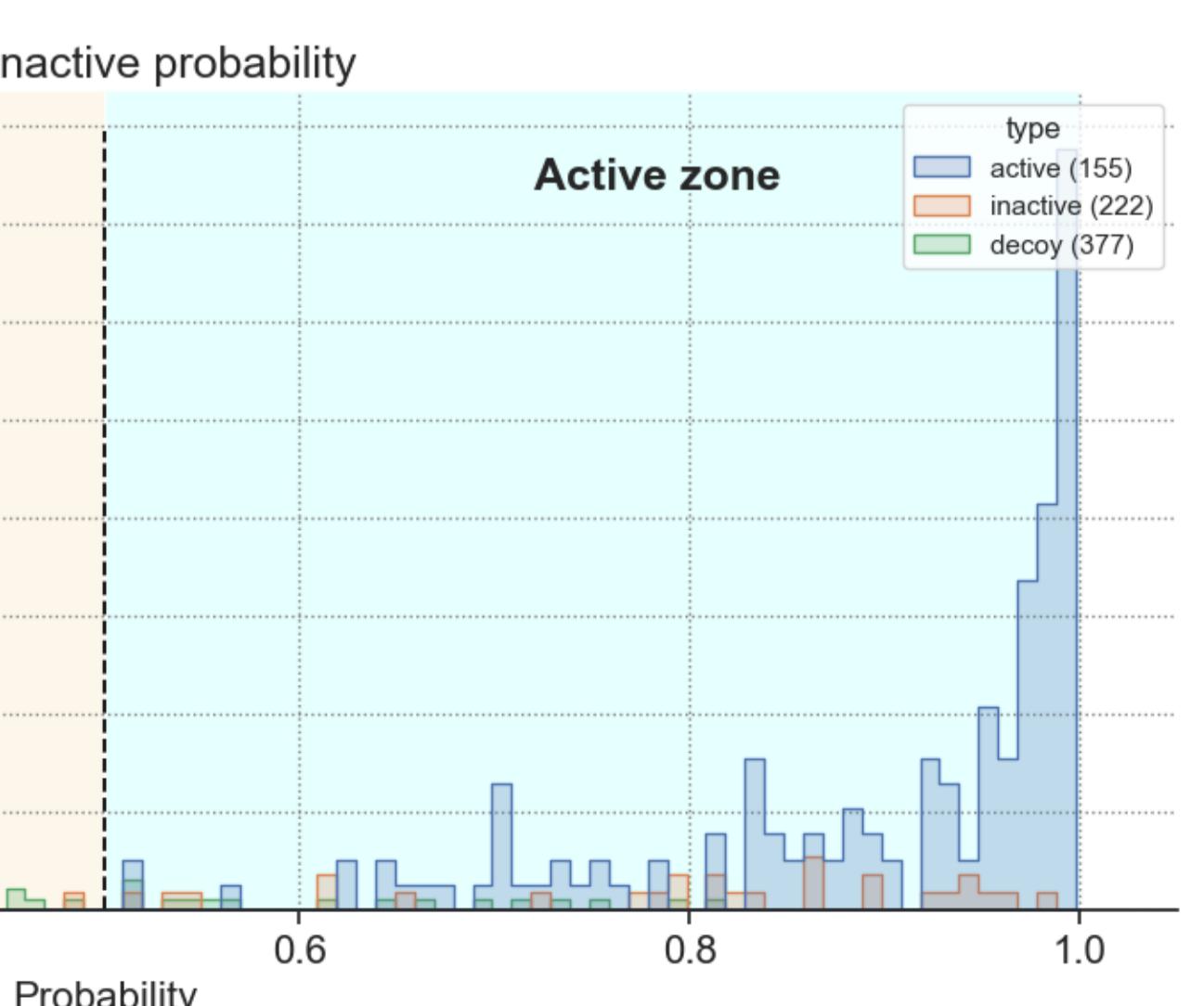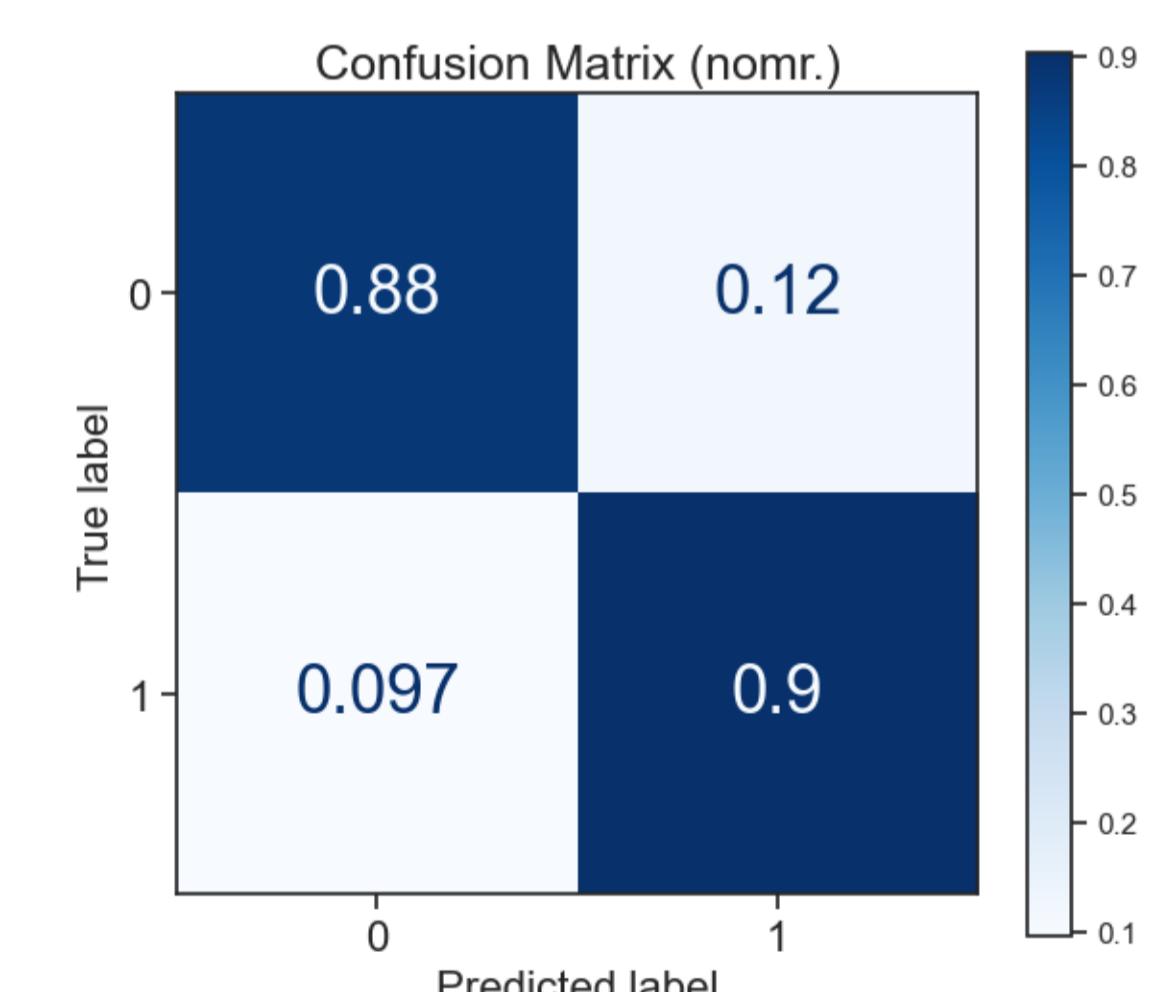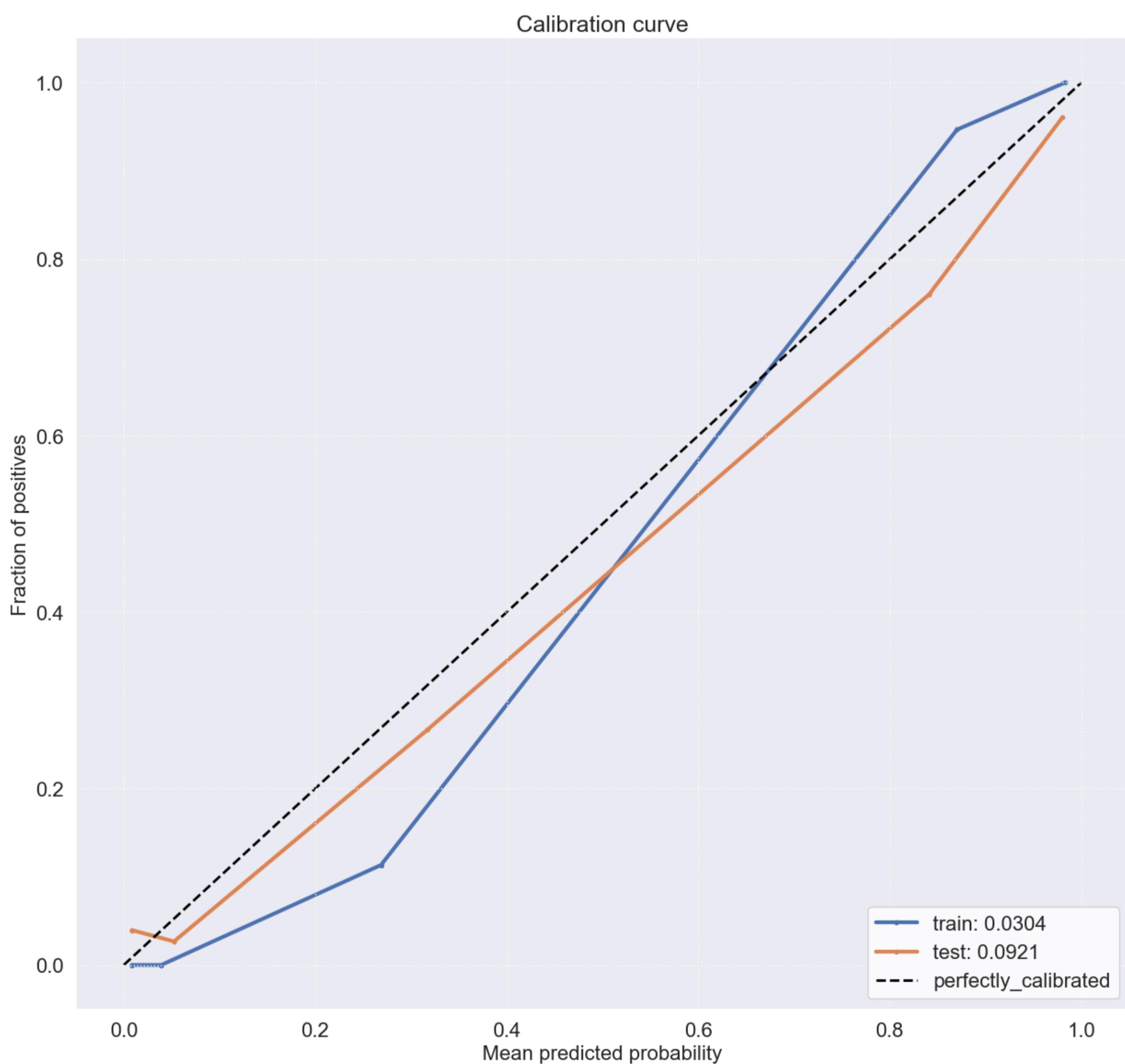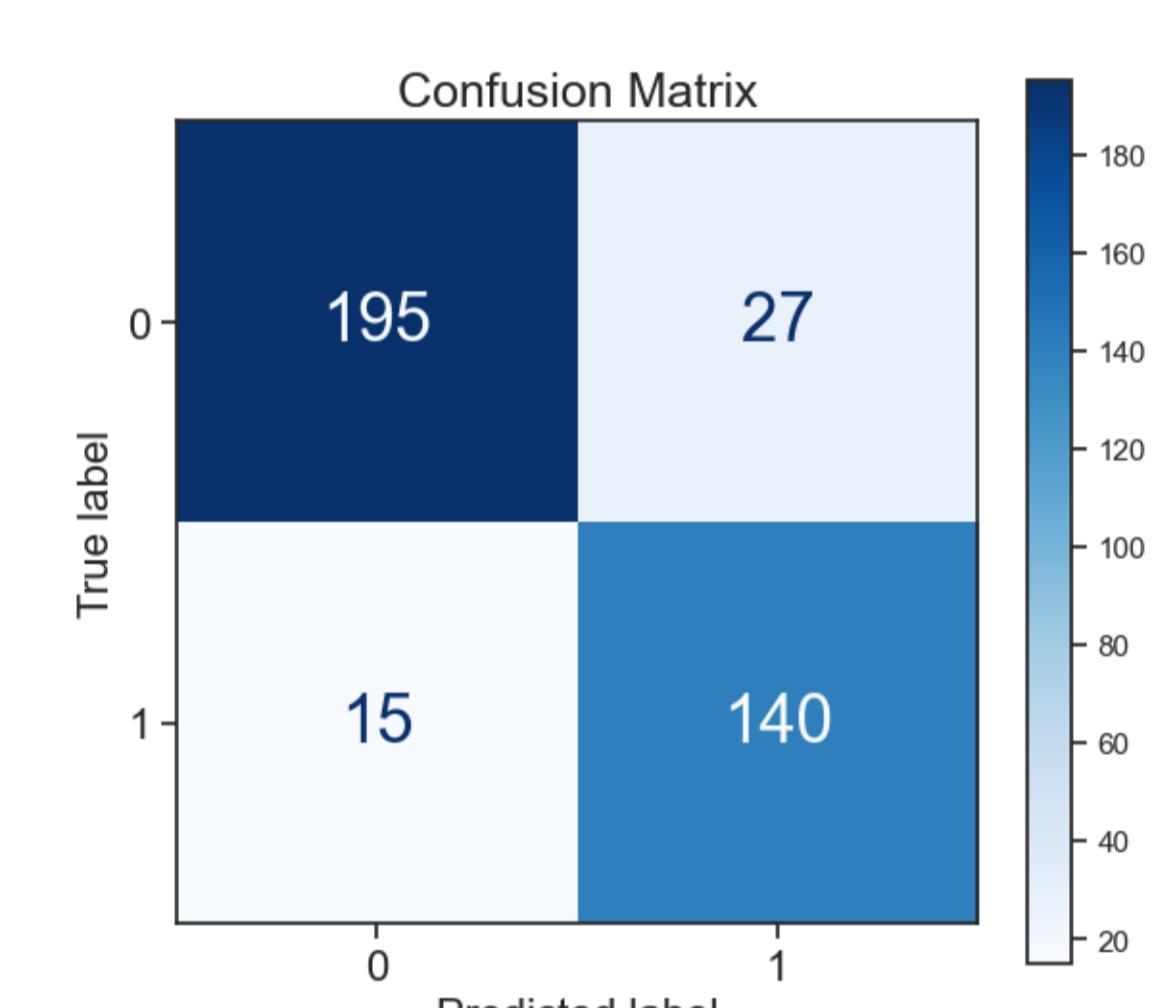
Metric	Train	Test
AUC	0.9961	0.9355
accuracy	0.9661	0.8886
balanced accuracy	0.9683	0.8908
recall_(sens)	0.9806	0.9032
specificity	0.9559	0.8784
precision	0.9397	0.8383
f1_score	0.9662	0.8891

Protein: Q07869 - metric: f1

Total (train): 1504

Active: 884
Inactive: 620

ROC curve

AUC_train = 0.998
AUC_test = 0.938
Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 782)

X-axis: Score f1 (Test)
Y-axis: Difference in Train-Test f1
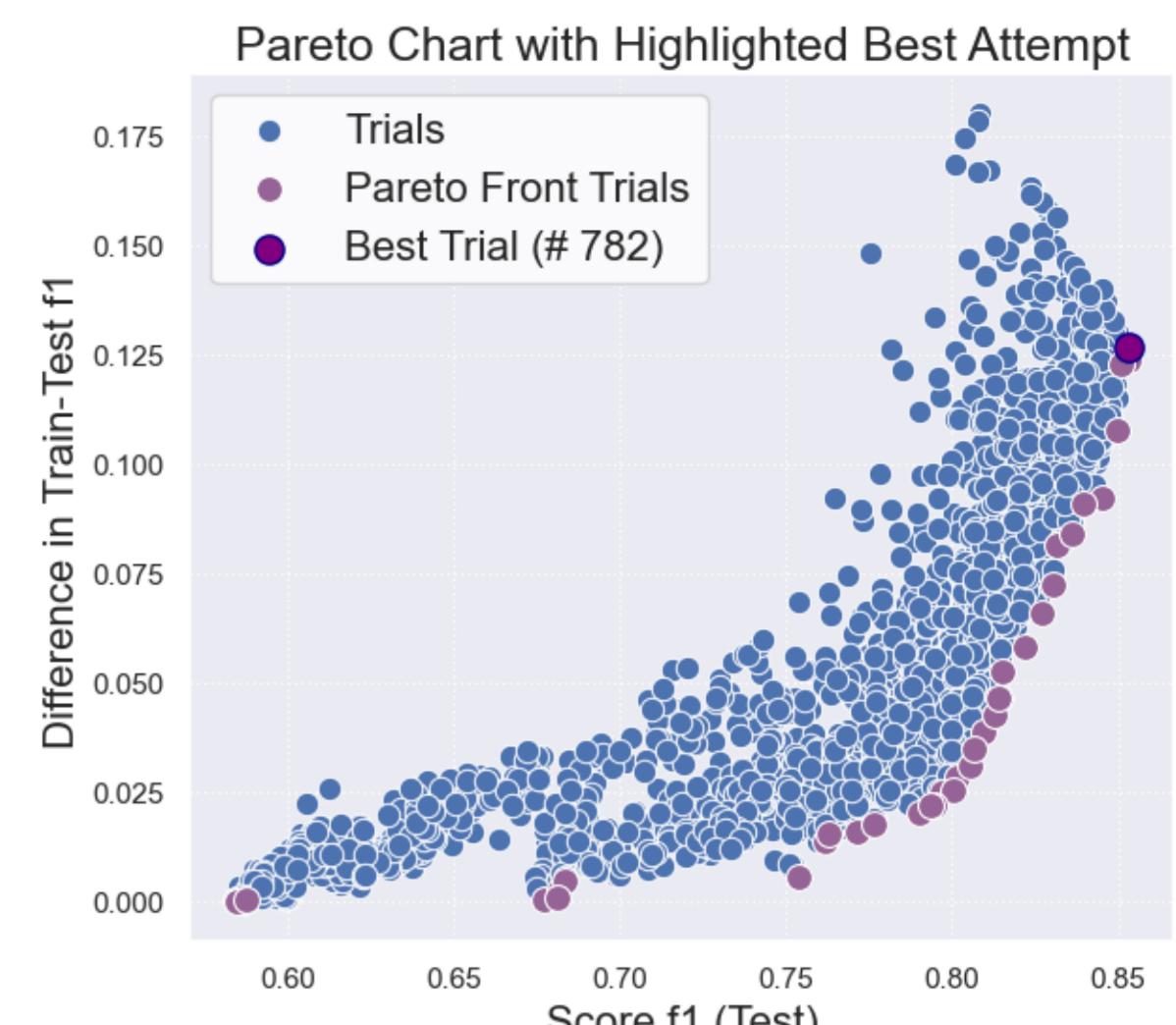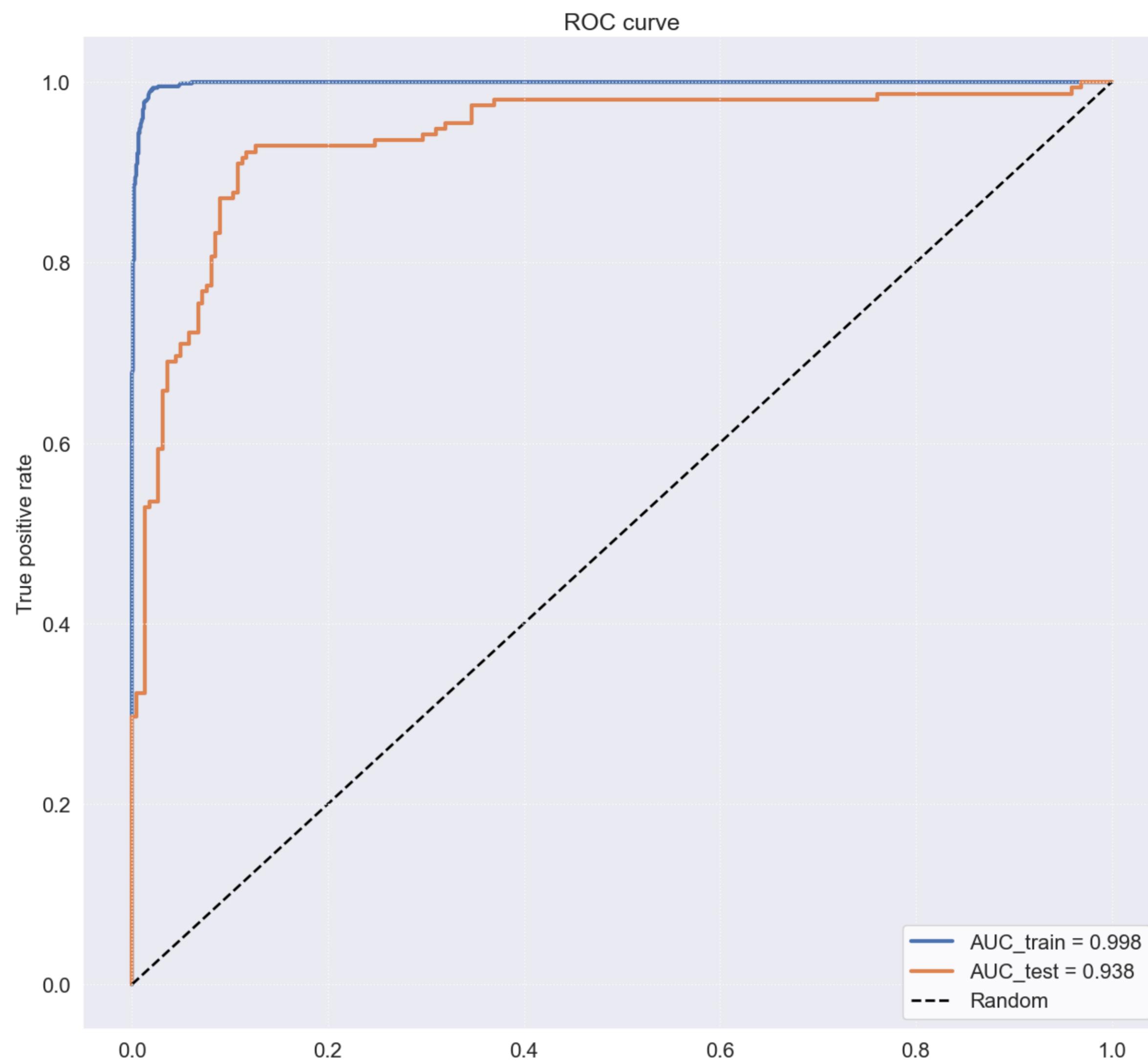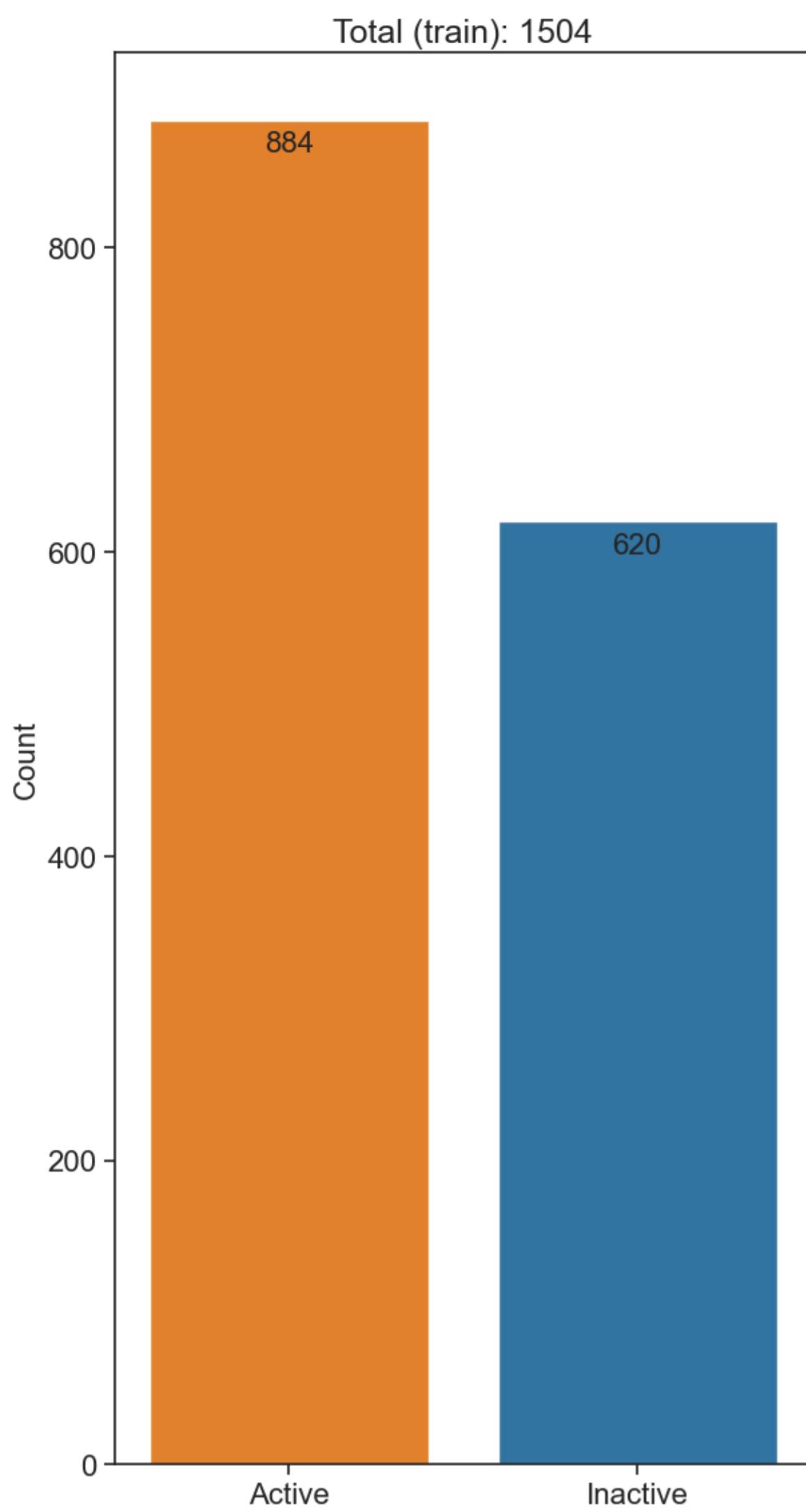
XGBoost Parameters	Value
alpha	2.219
colsample_bytree	0.5377
gamma	0.0037
lambda	3.0693
learning_rate	0.0381
max_depth	9.0
min_child_weight	1.0
n_estimators	604.0
reg_alpha	0.6293
reg_lambda	0.4899
scale_pos_weight	1.9593
subsample	0.649

Confusion Matrix

	0	1
0	196	26
1	13	142

Confusion Matrix (nomr.)

	0	1
0	0.88	0.12
1	0.084	0.92

Calibration curve

train: 0.02
test: 0.0891
perfectly_calibrated

Active/Inactive probability

Inactive zone | Active zone

type:
- active (155)
- inactive (222)
- decoy (377)
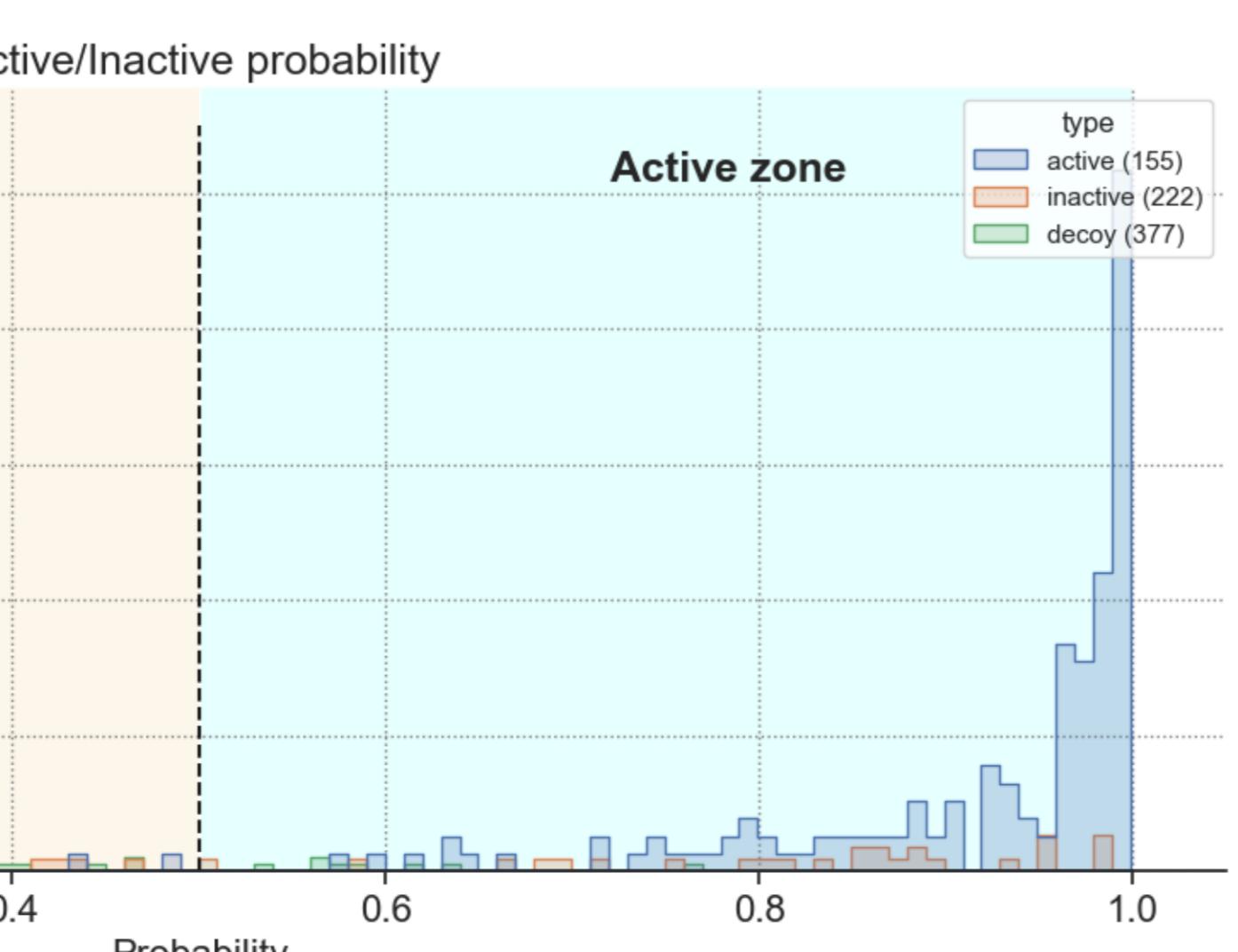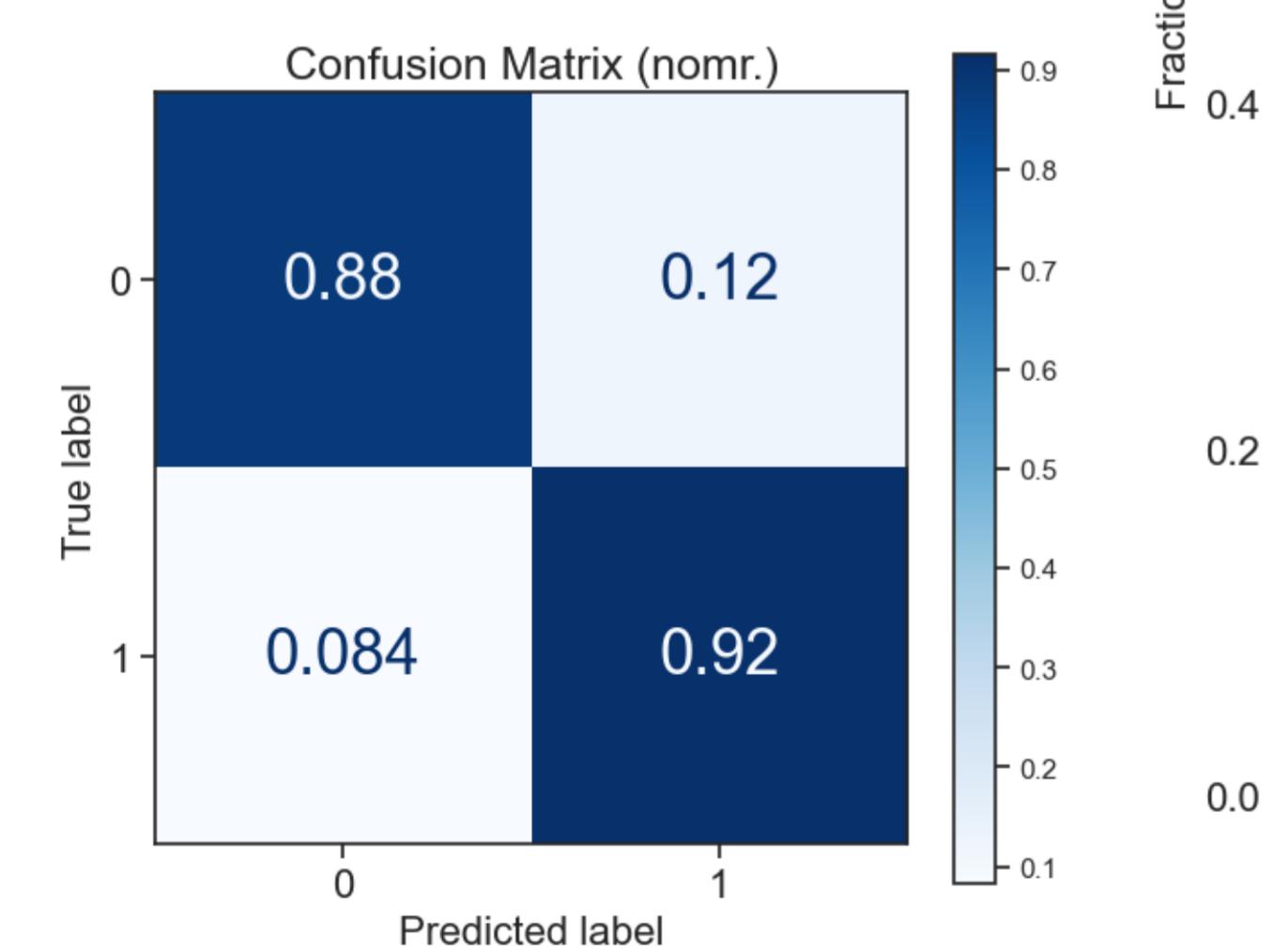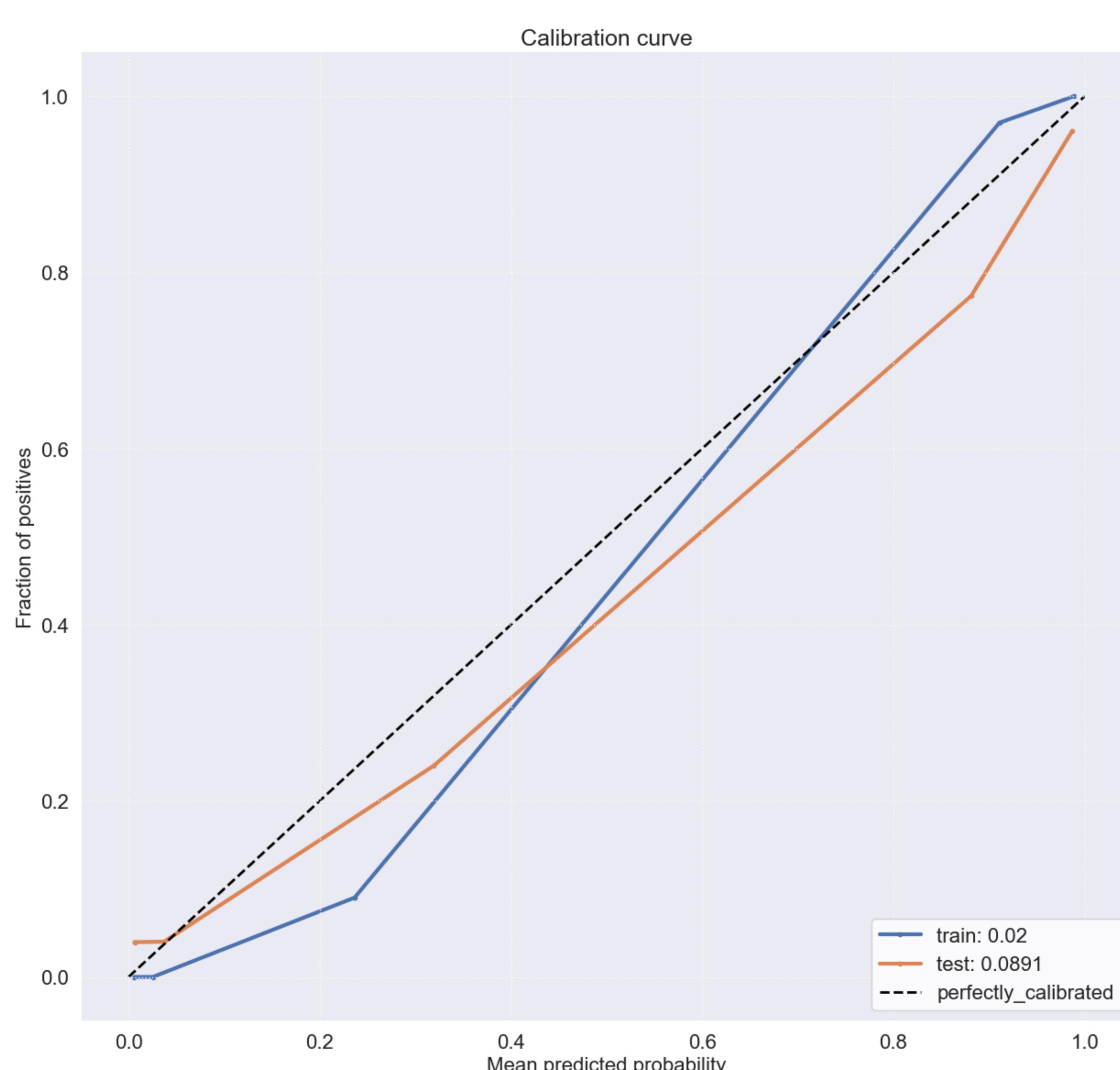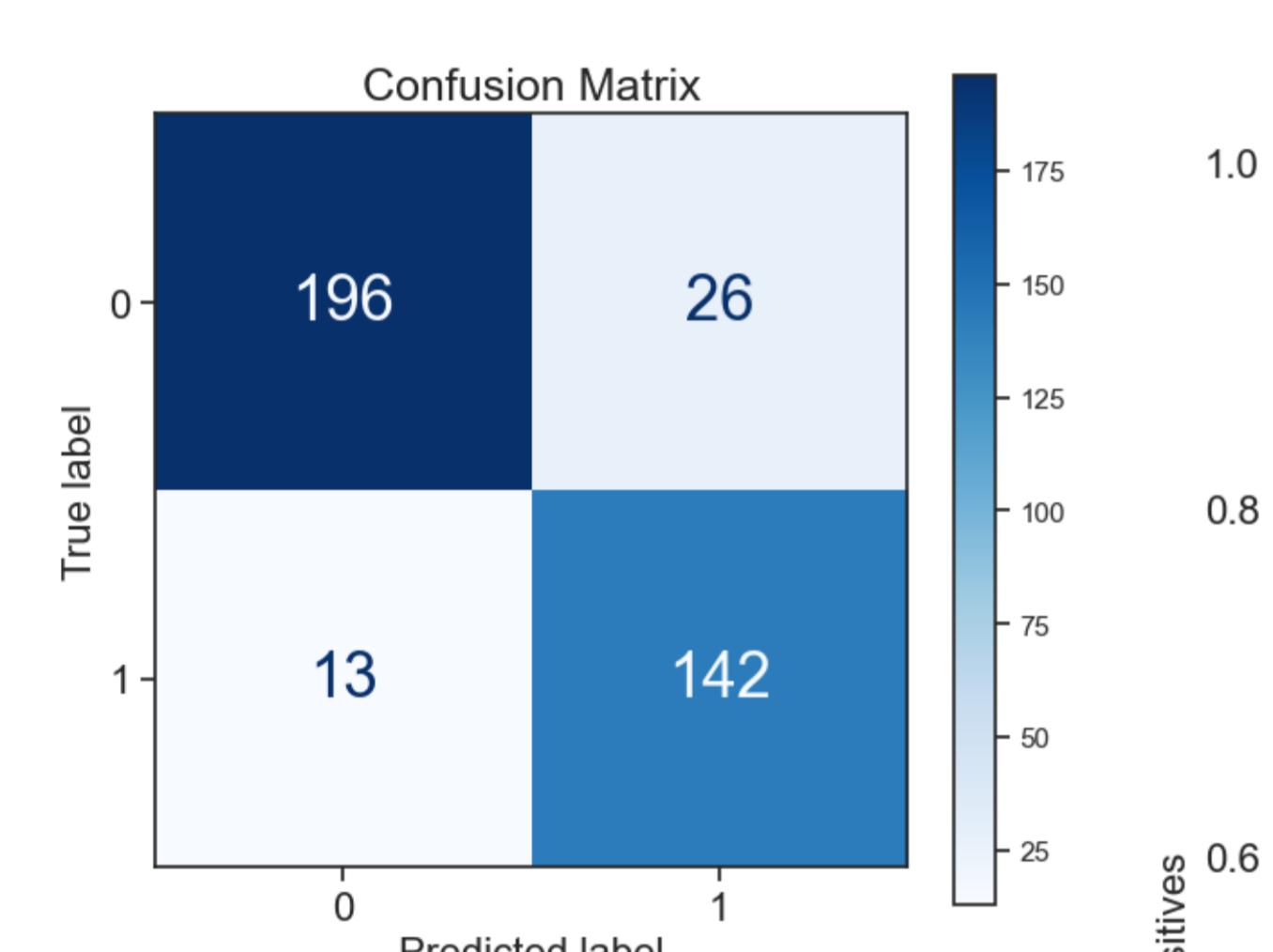
Metric	Train	Test
AUC	0.9984	0.9381
accuracy	0.9807	0.8966
balanced accuracy	0.9829	0.8995
recall_(sens)	0.9952	0.9161
specificity	0.9706	0.8829
precision	0.9596	0.8452
f1_score	0.9808	0.8971

Protein: Q07869 - metric: precision

Total (train): 1504

ROC curve

AUC_train = 0.998
AUC_test = 0.934
Random

Pareto Chart with Highlighted Best Attempt

Trials
Pareto Front Trials
Best Trial (# 2227)
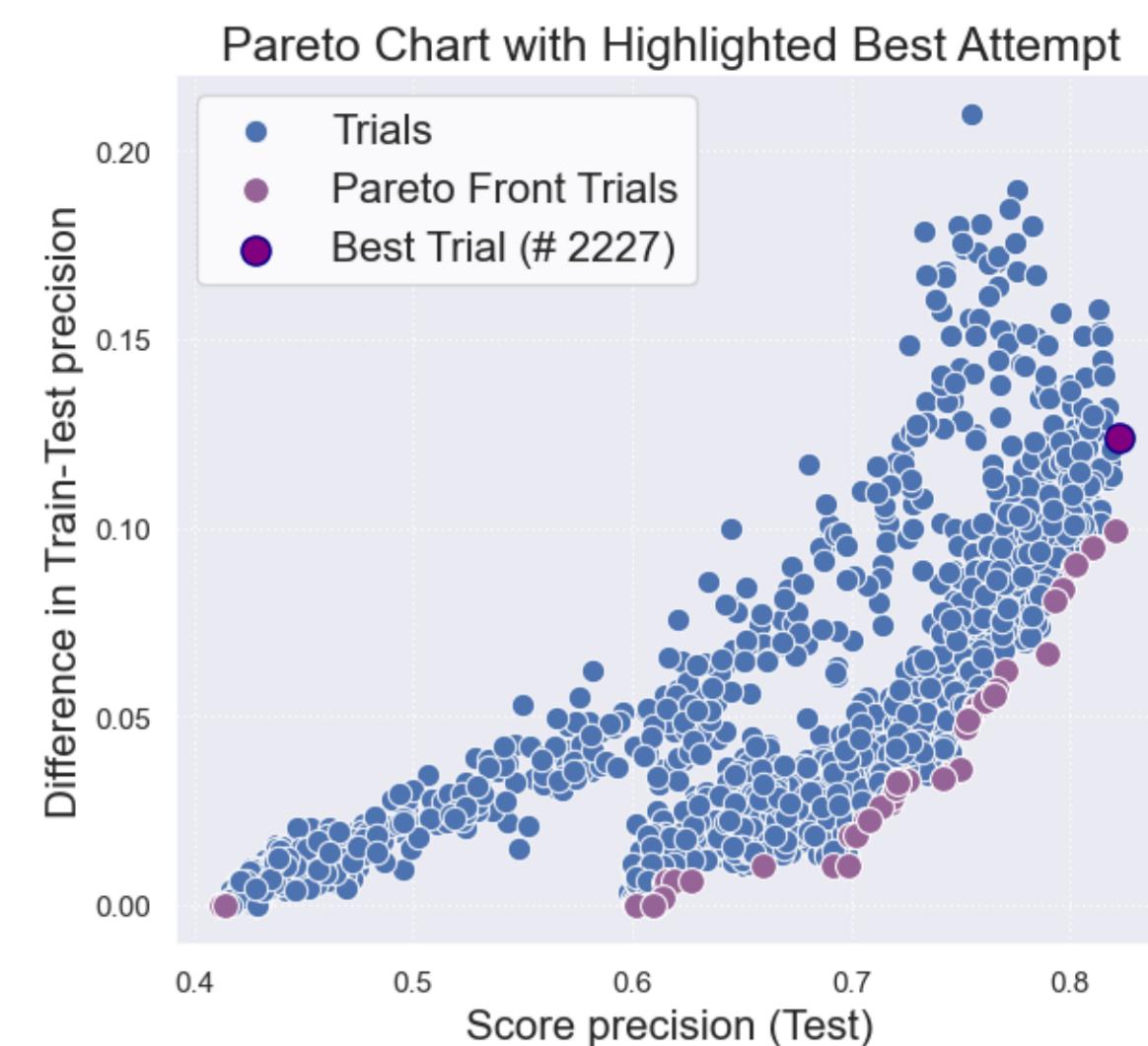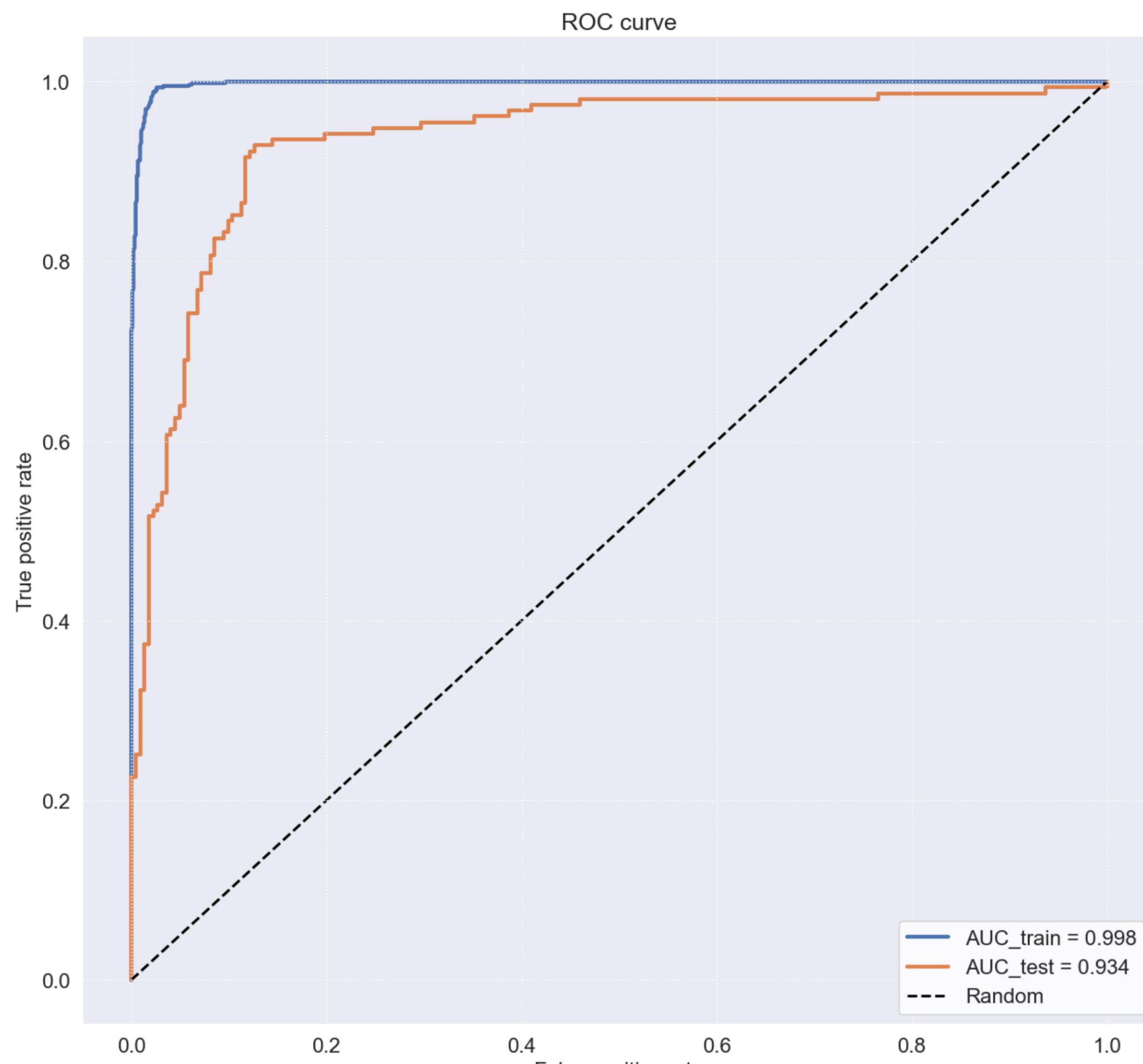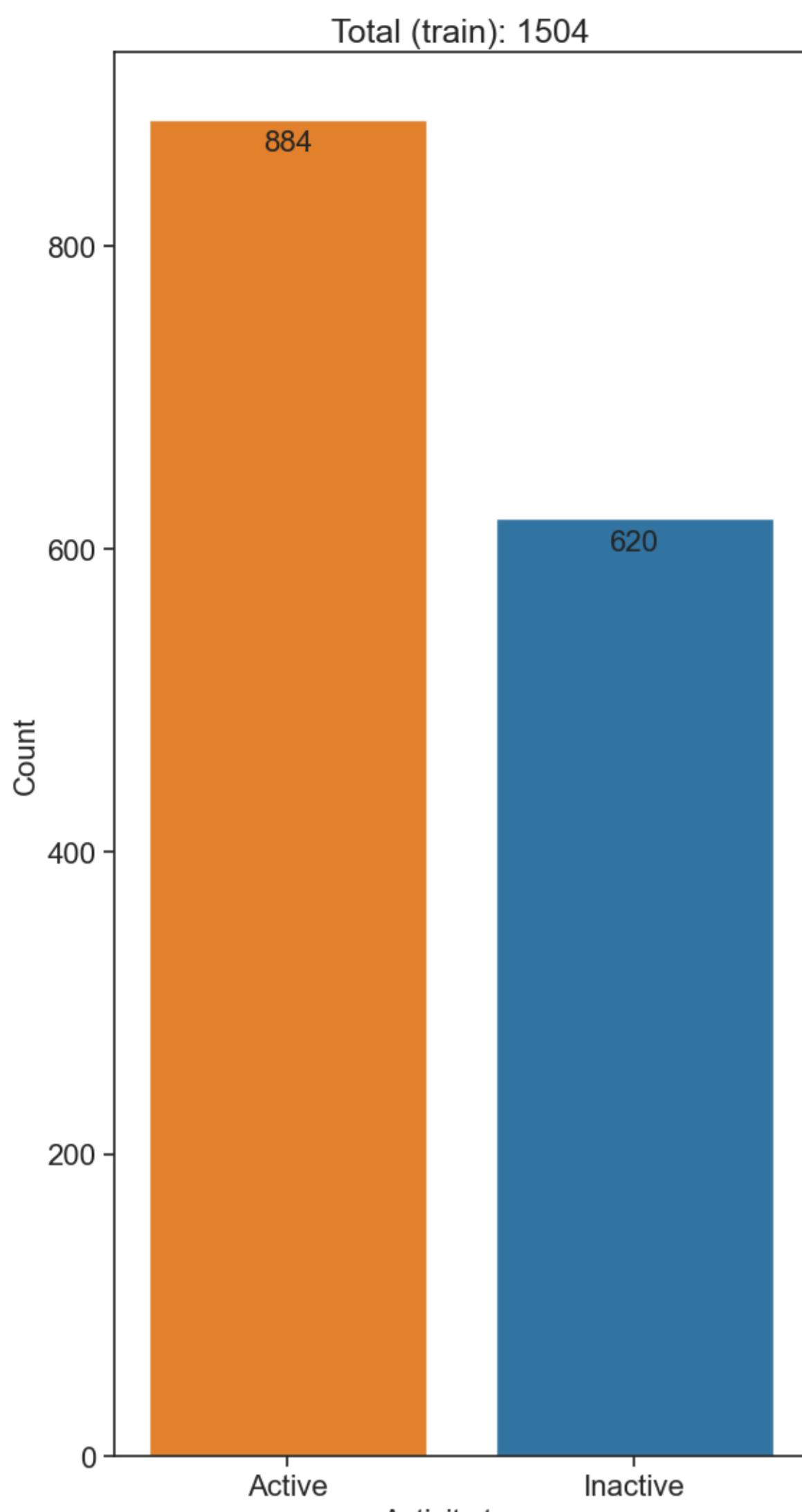
XGBoost Parameters	Value
alpha	4.6742
colsample_bytree	0.8802
gamma	0.6787
lambda	4.4581
learning_rate	0.1114
max_depth	8.0
min_child_weight	4.0
n_estimators	493.0
reg_alpha	0.1061
reg_lambda	0.1717
scale_pos_weight	2.2971
subsample	0.7011

Confusion Matrix

| | 196 | 26 |
| | 13 | 142 |

Confusion Matrix (nomr.)

| | 0.88 | 0.12 |
| | 0.084 | 0.92 |

Calibration curve

train: 0.0241
test: 0.0915
perfectly_calibrated

Active/Inactive probability

Inactive zone   Active zone

type
active (155)
inactive (222)
decoy (377)
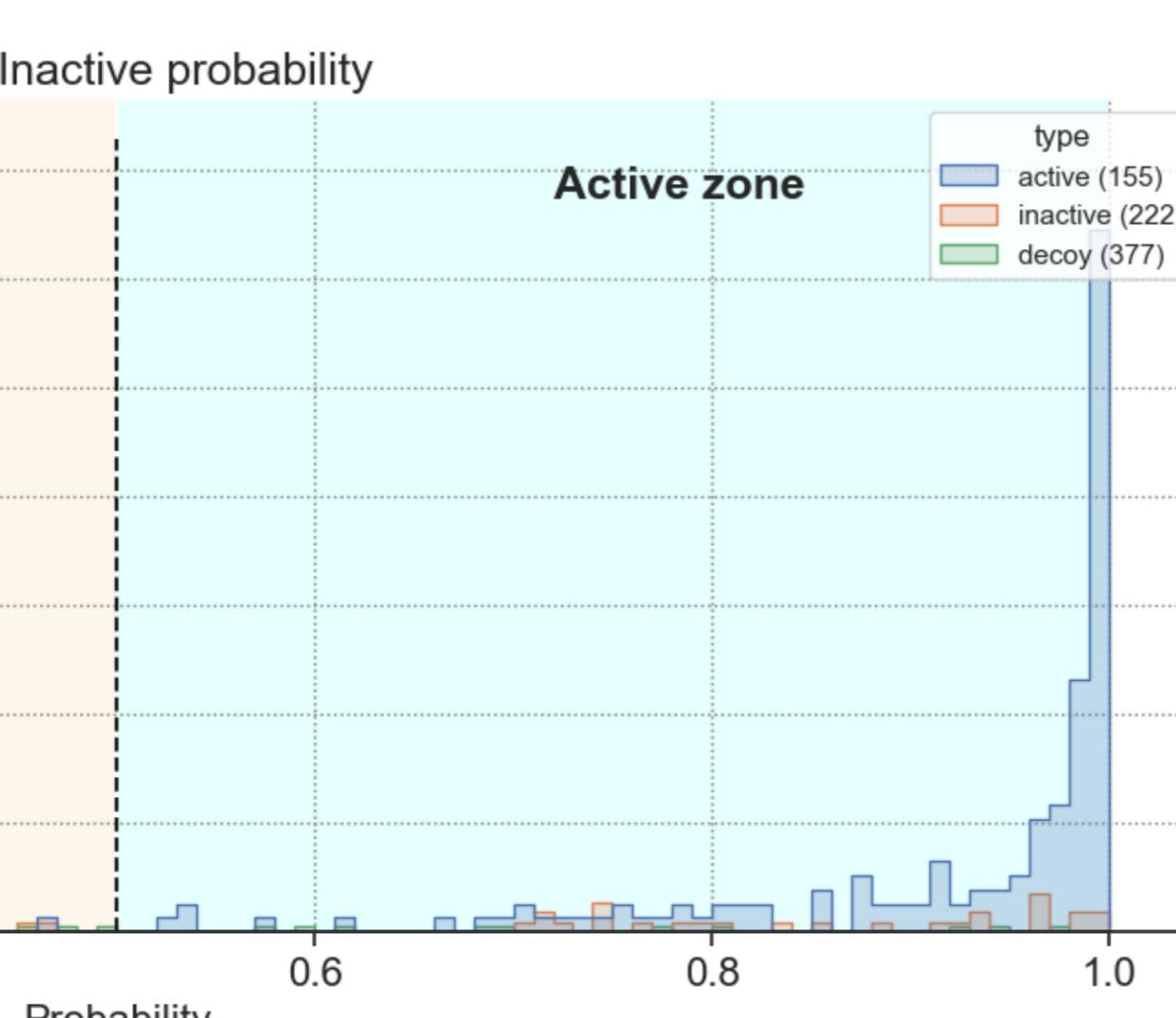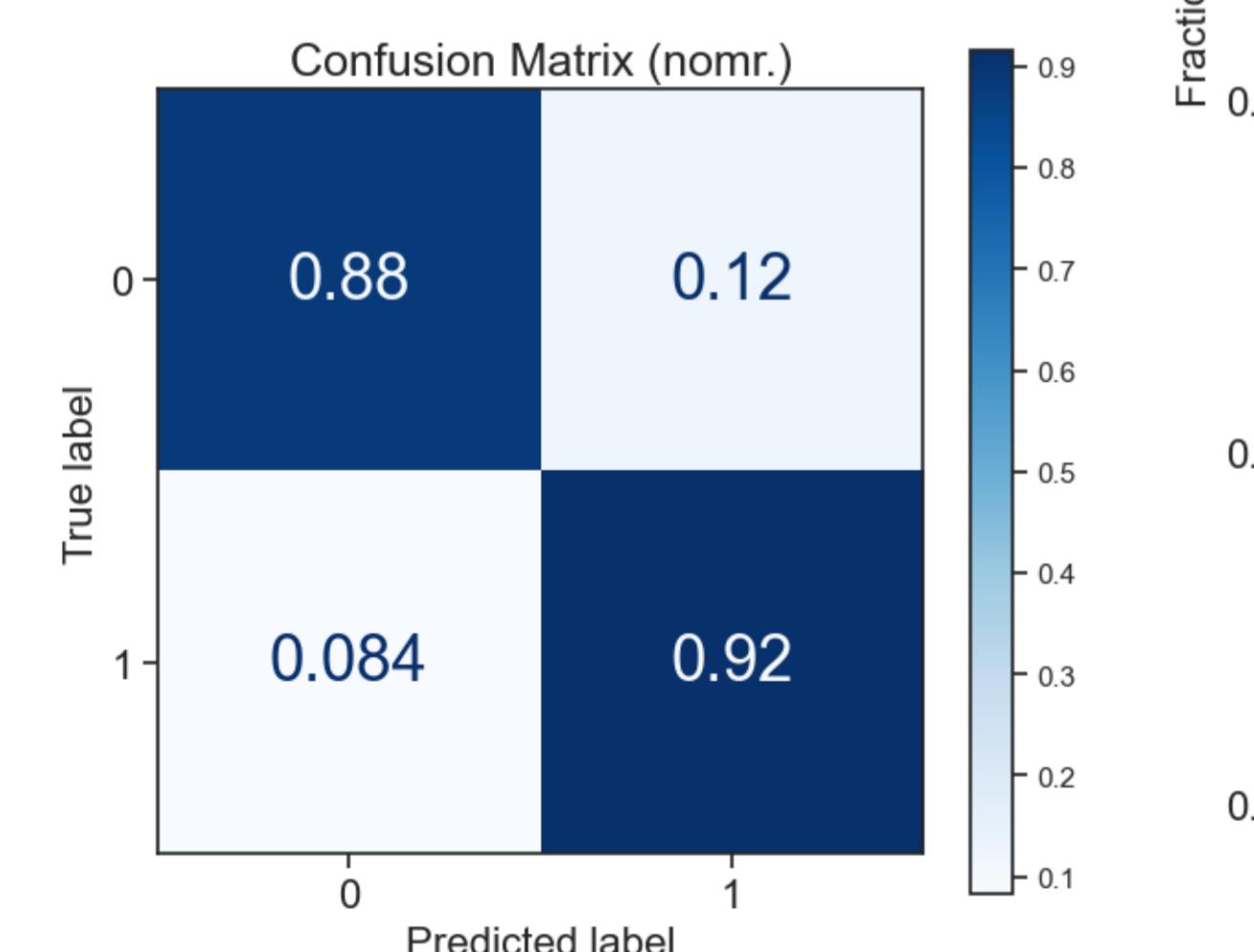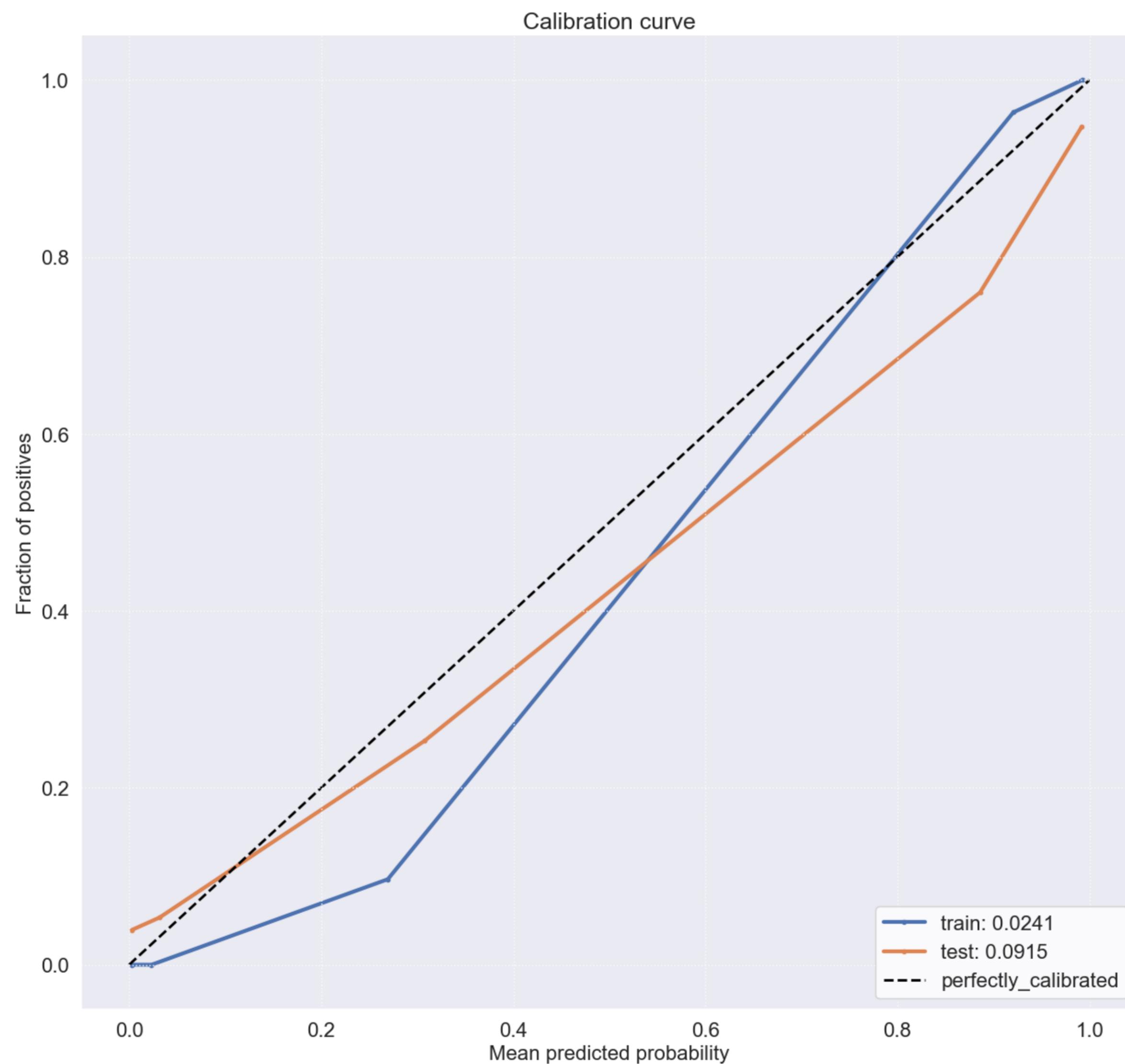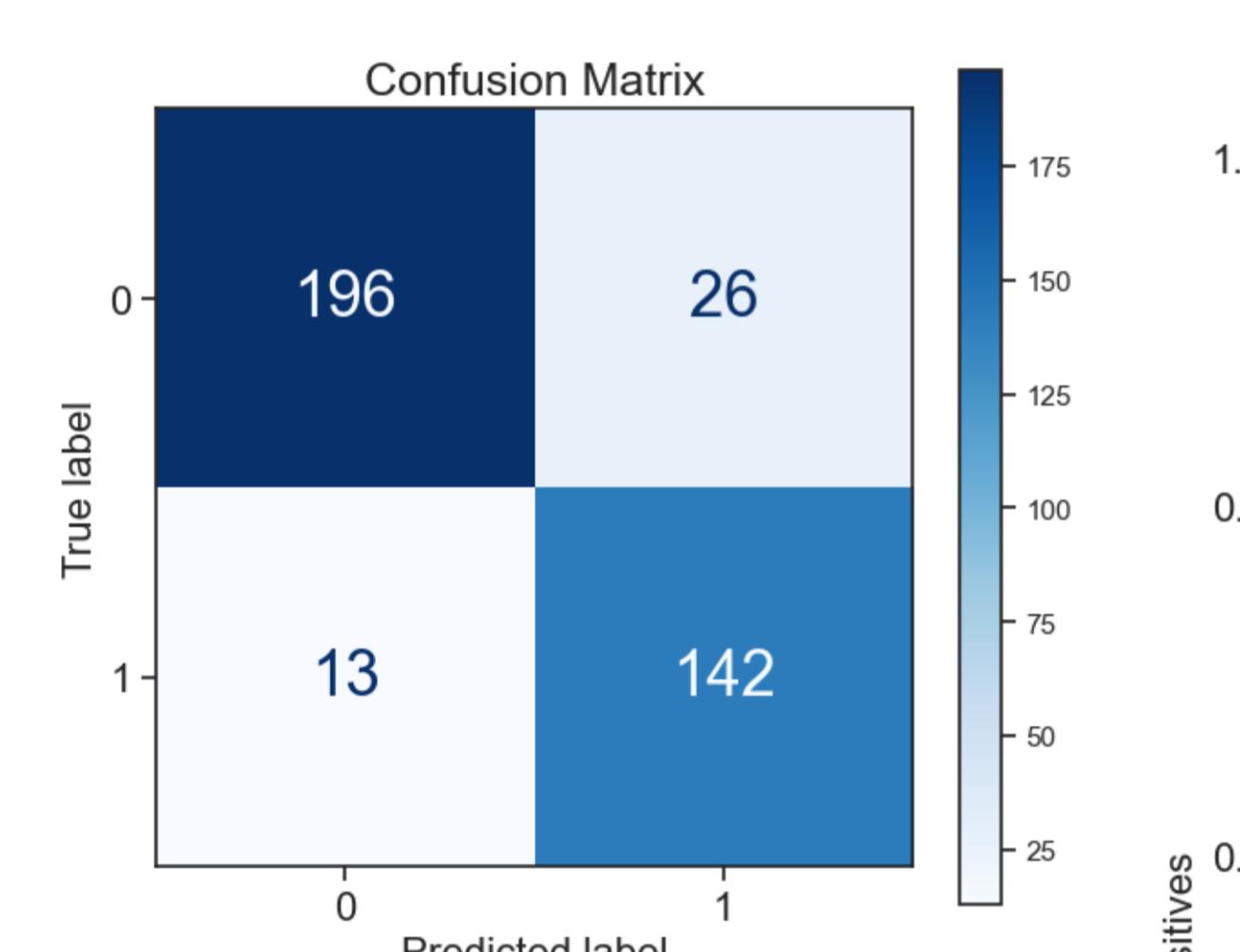
Metric	Train	Test
AUC	0.9978	0.9341
accuracy	0.9734	0.8966
balanced accuracy	0.9767	0.8995
recall_(sens)	0.9952	0.9161
specificity	0.9581	0.8829
precision	0.9434	0.8452
f1_score	0.9735	0.8971

Protein: Q07869 - metric: SoftVotingXGBoostEnsemble
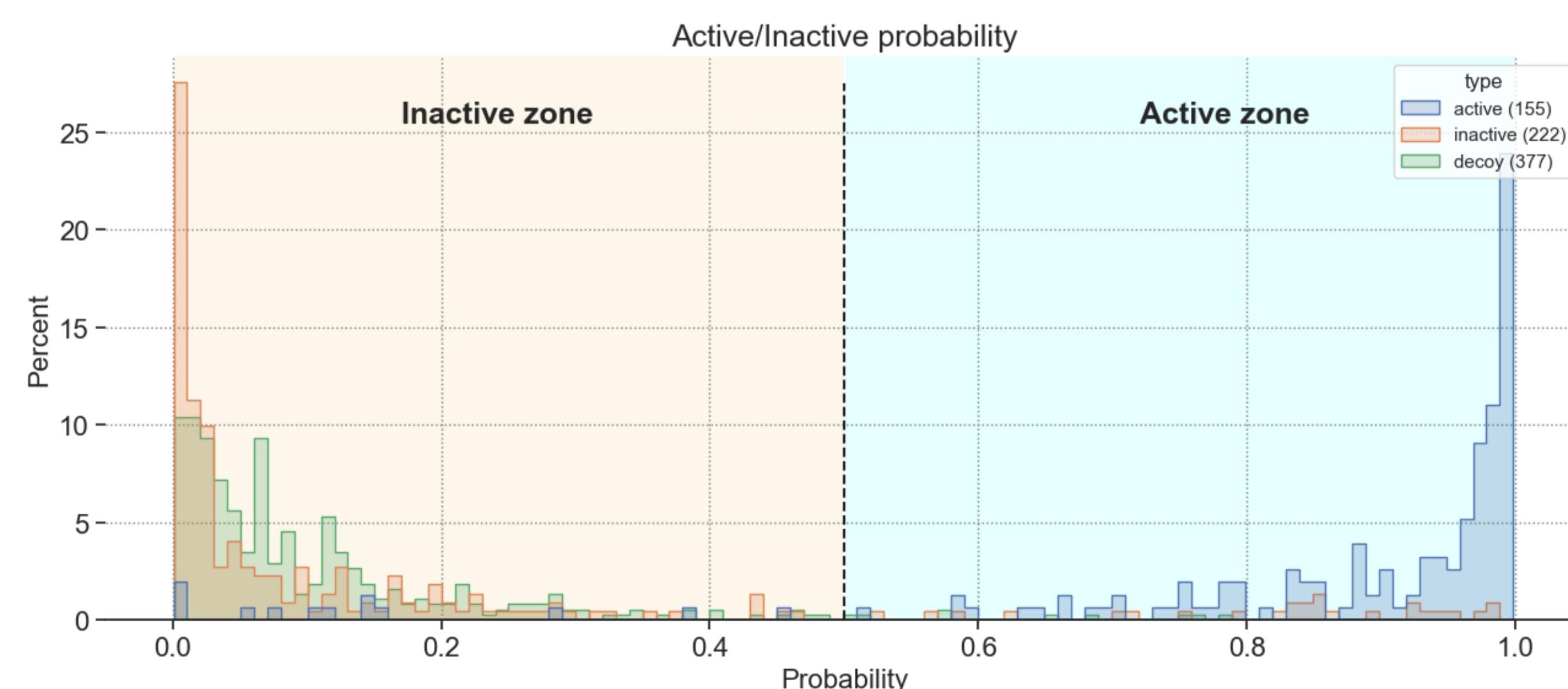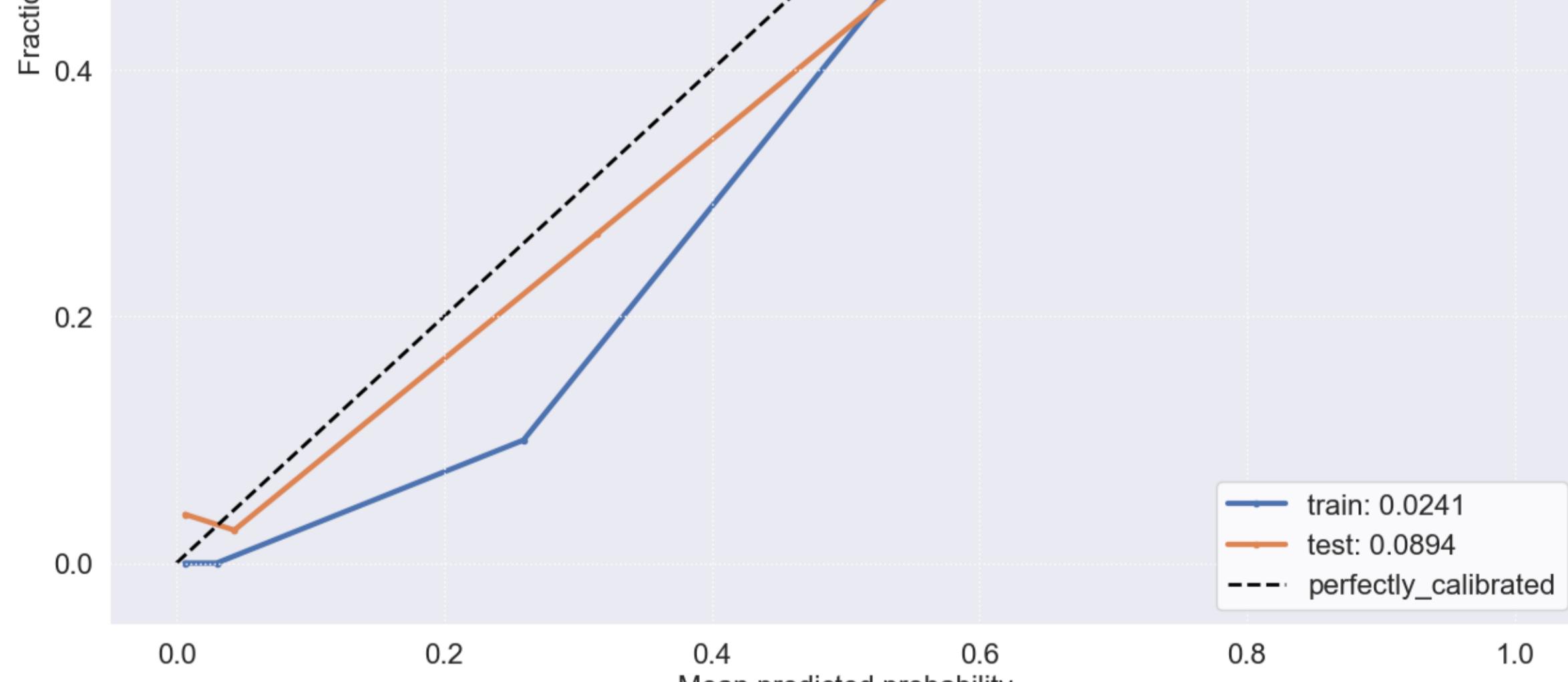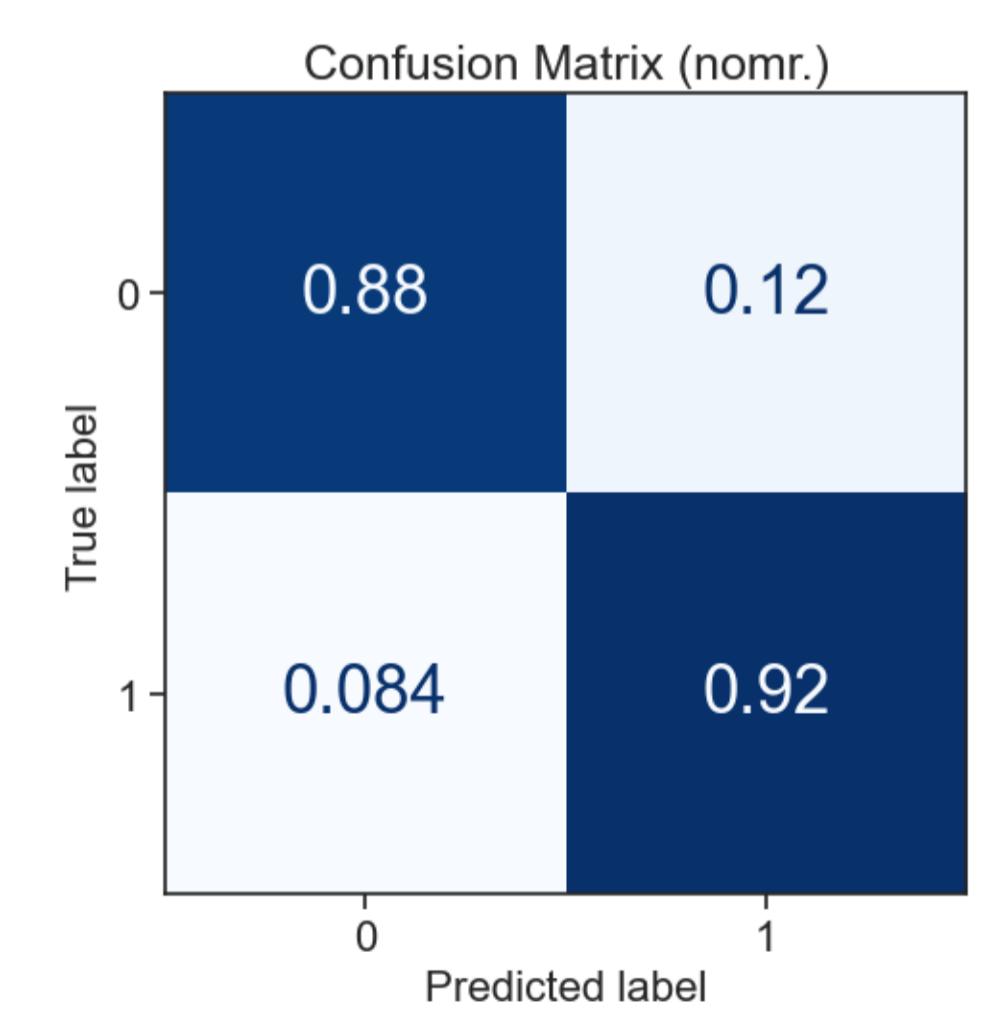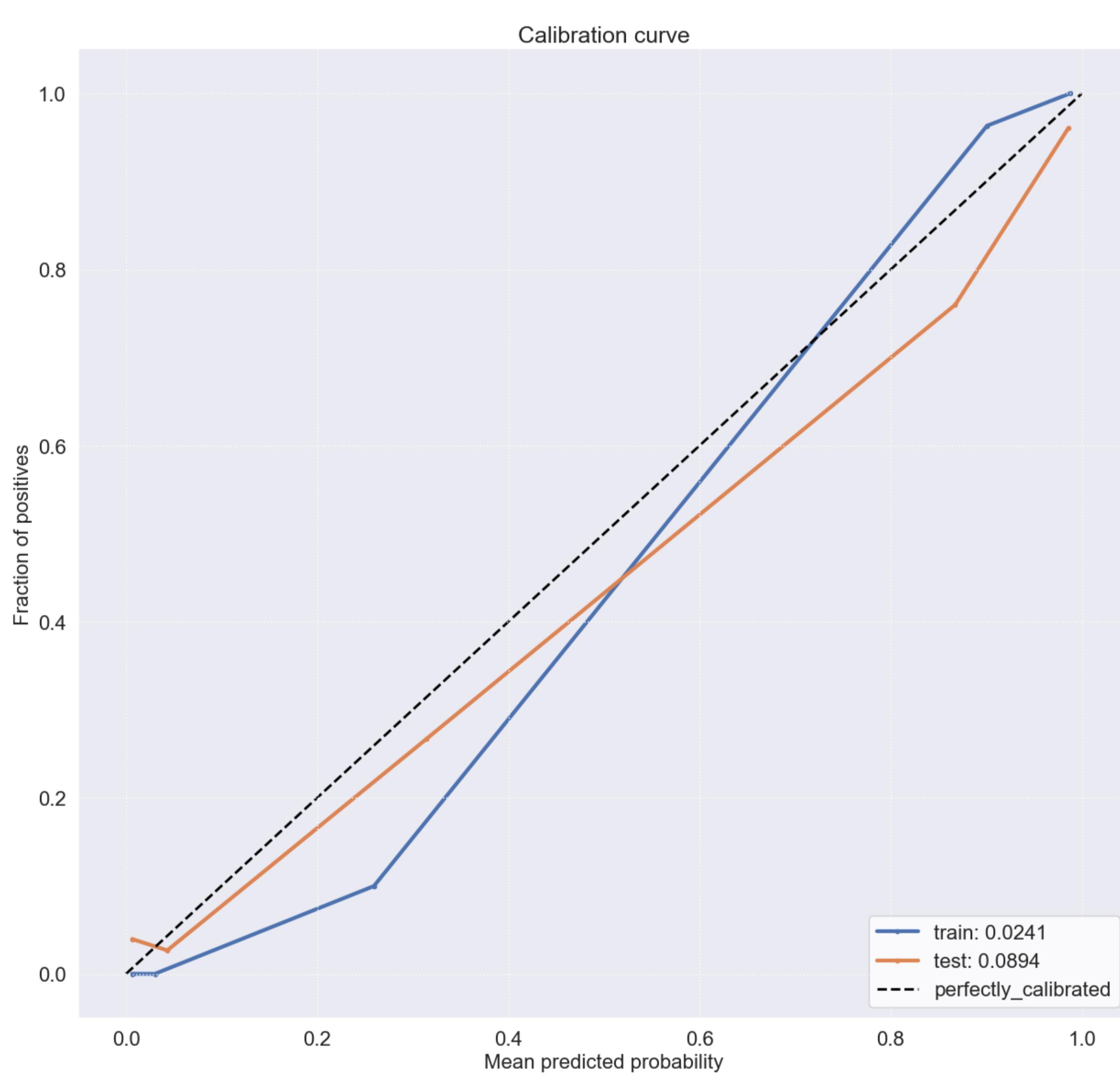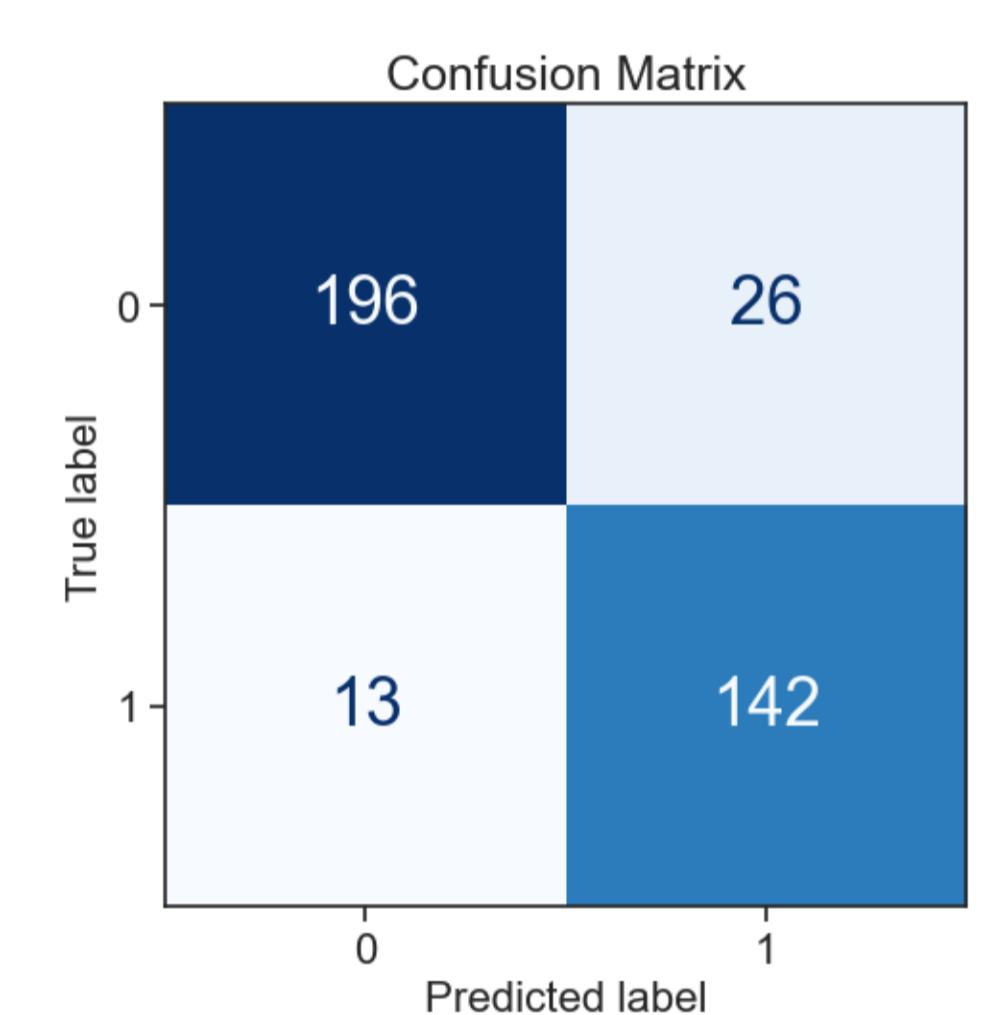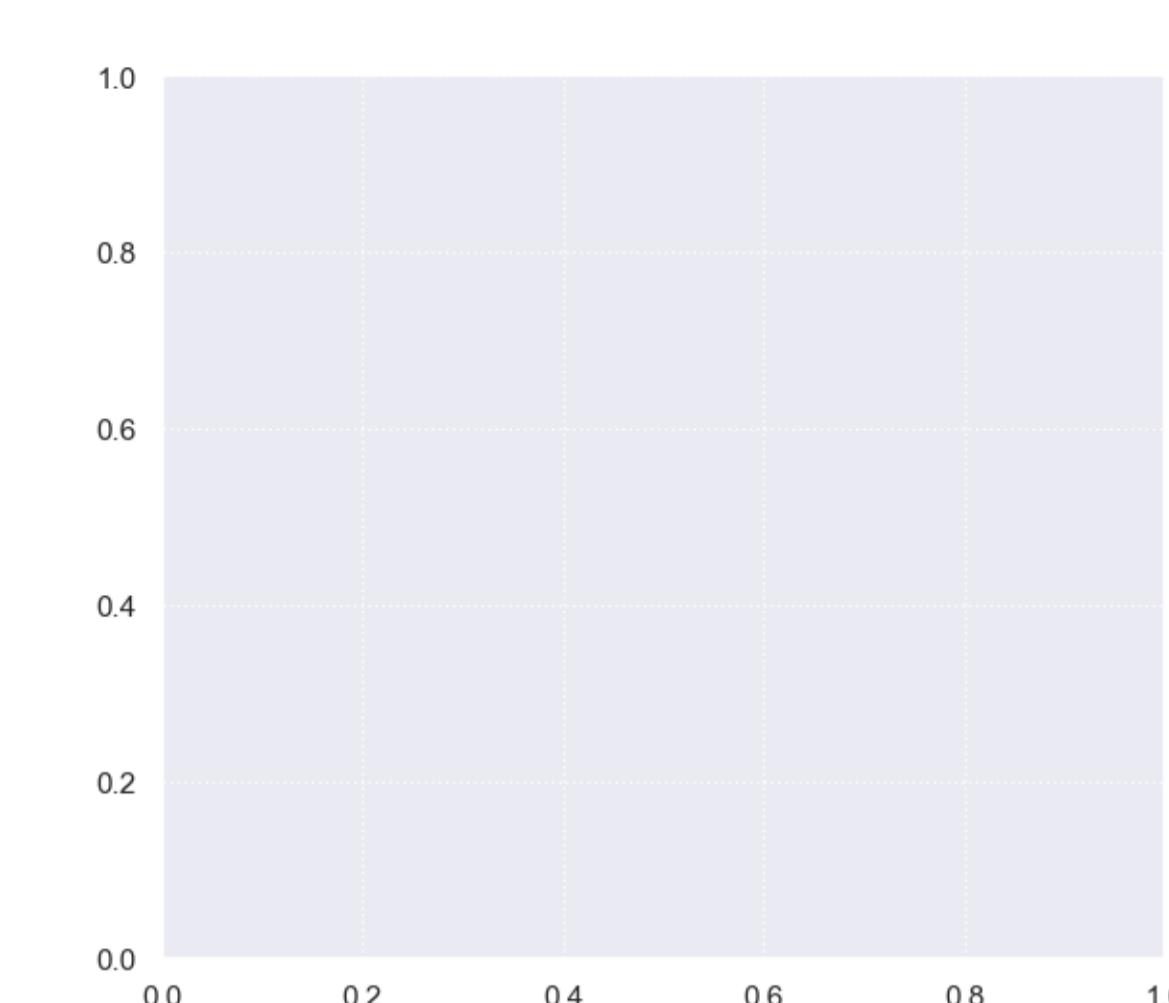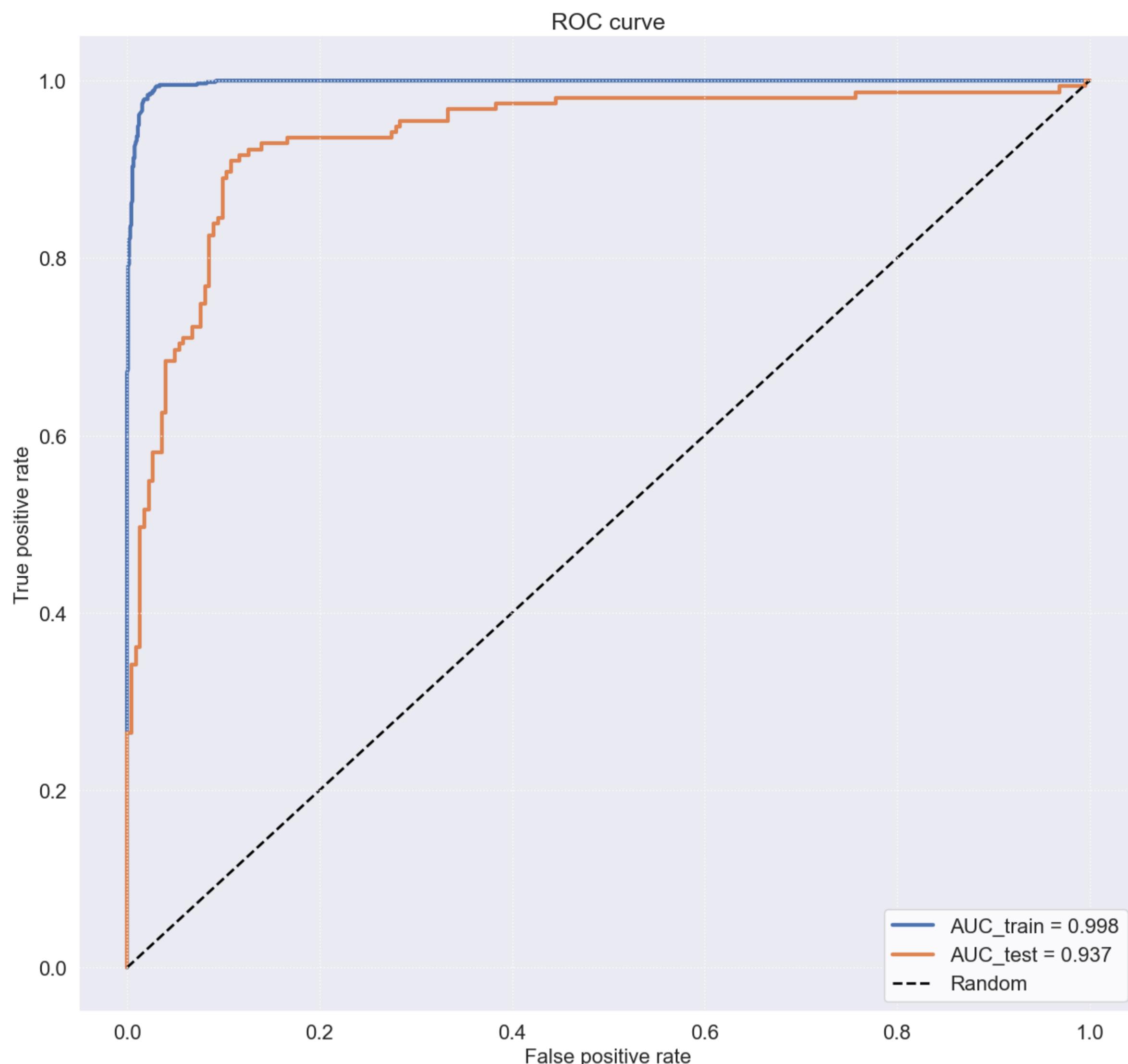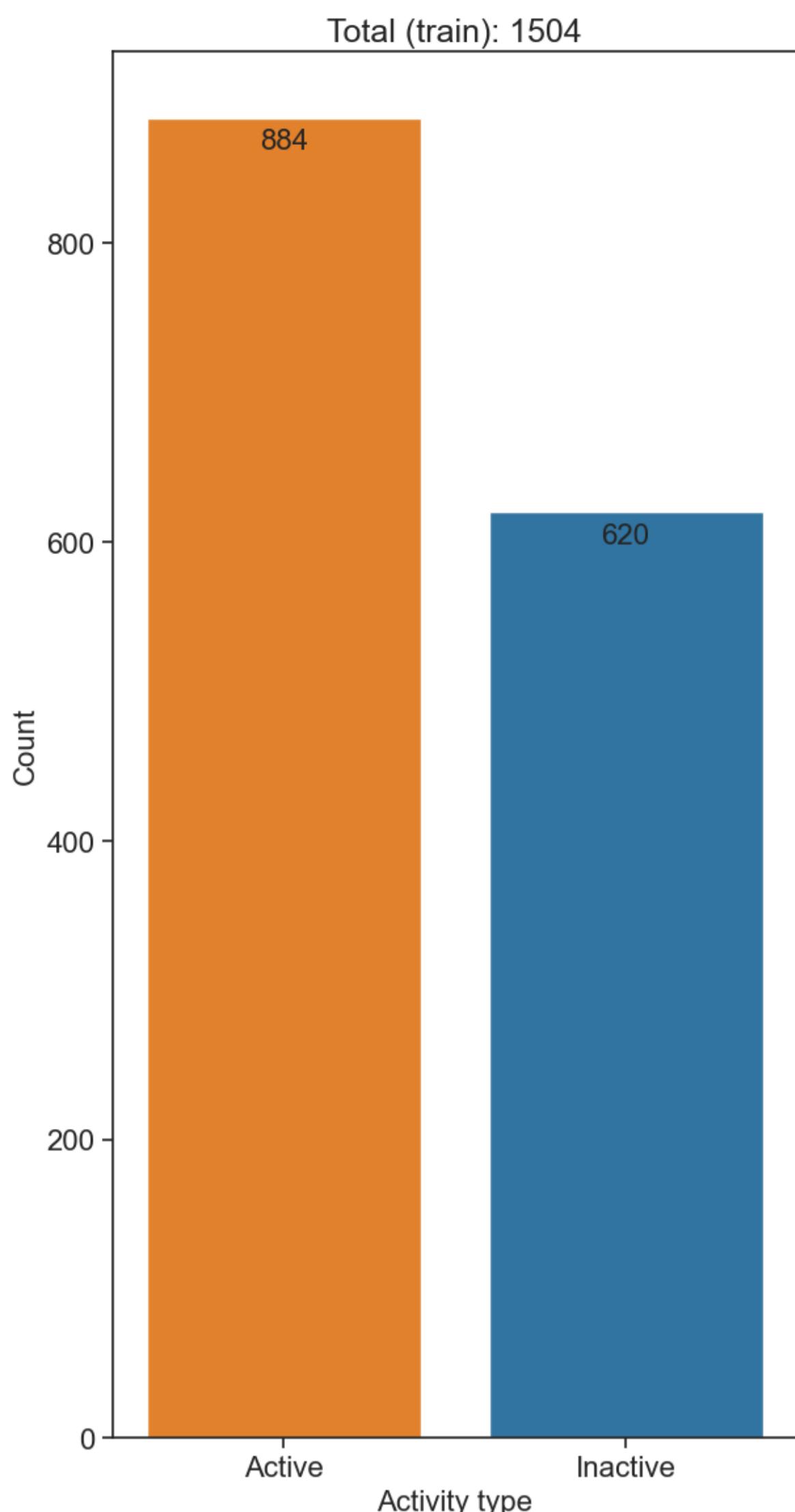
Metric	Train	Test
AUC	0.9977	0.9365
accuracy	0.9774	0.8966
balanced accuracy	0.98	0.8995
recall_(sens)	0.9952	0.9161
specificity	0.9649	0.8829
precision	0.9522	0.8452
f1_score	0.9775	0.8971