Protein: P06239 - metric: balanced_accuracy

Total (train): 1209

Active: 746
Inactive: 463

ROC curve

- AUC_train = 0.998
- AUC_test = 0.920
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
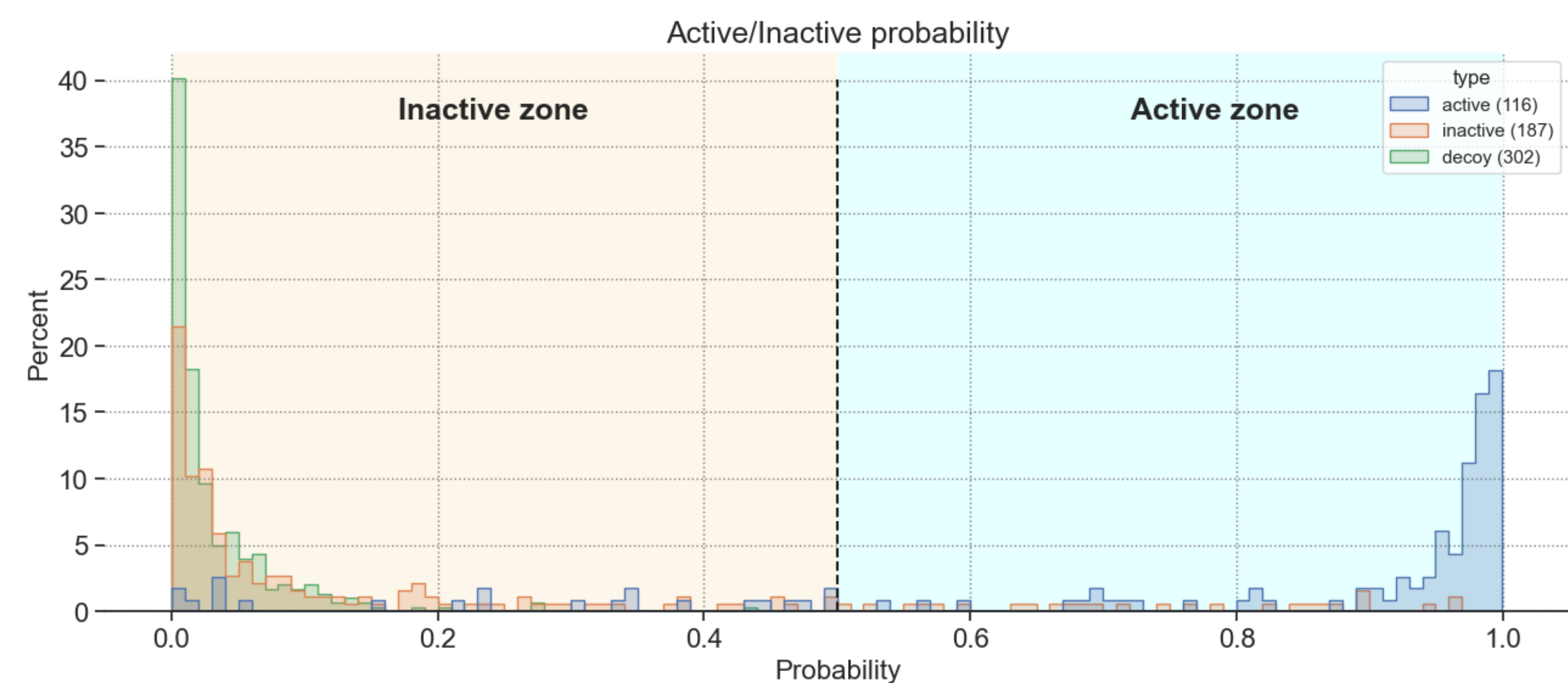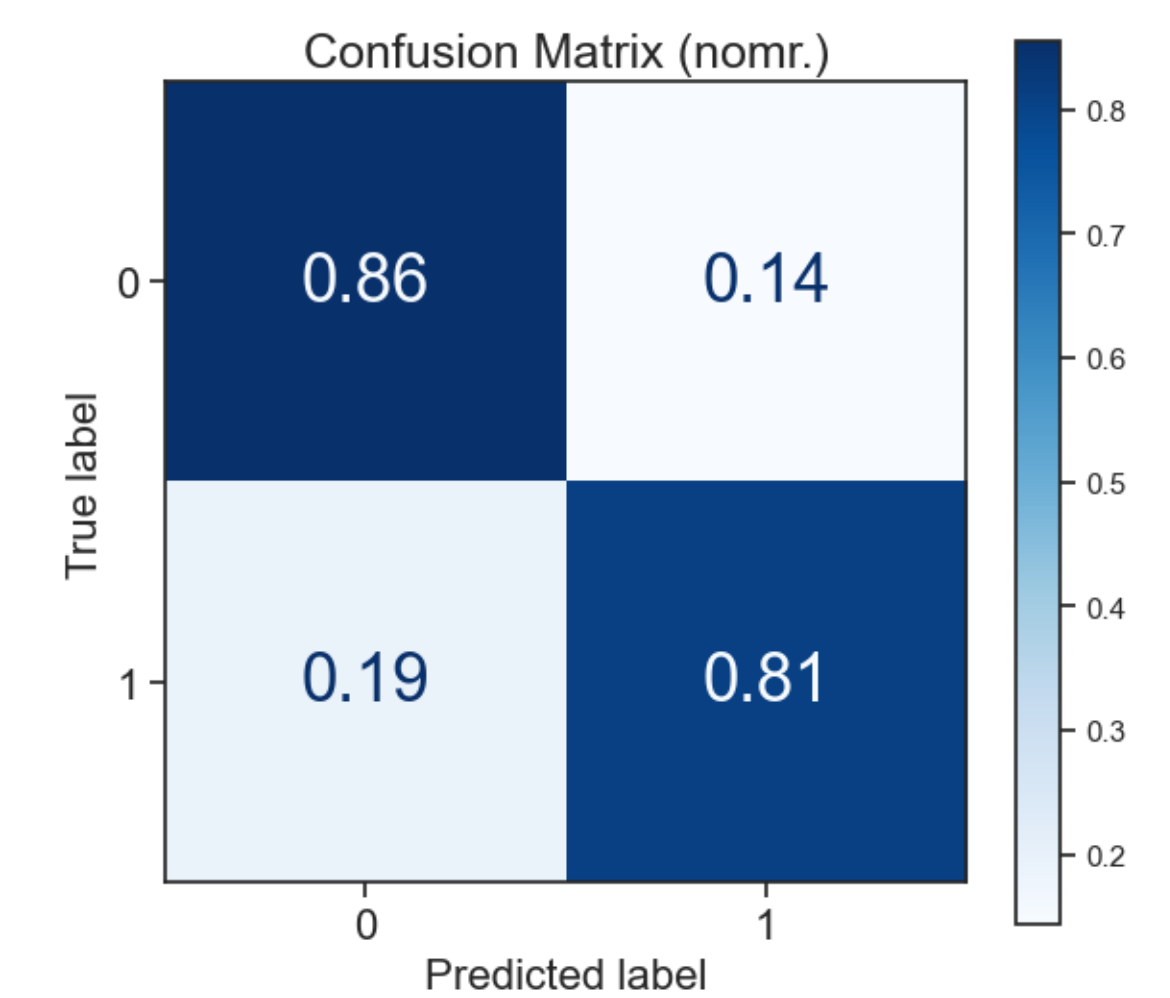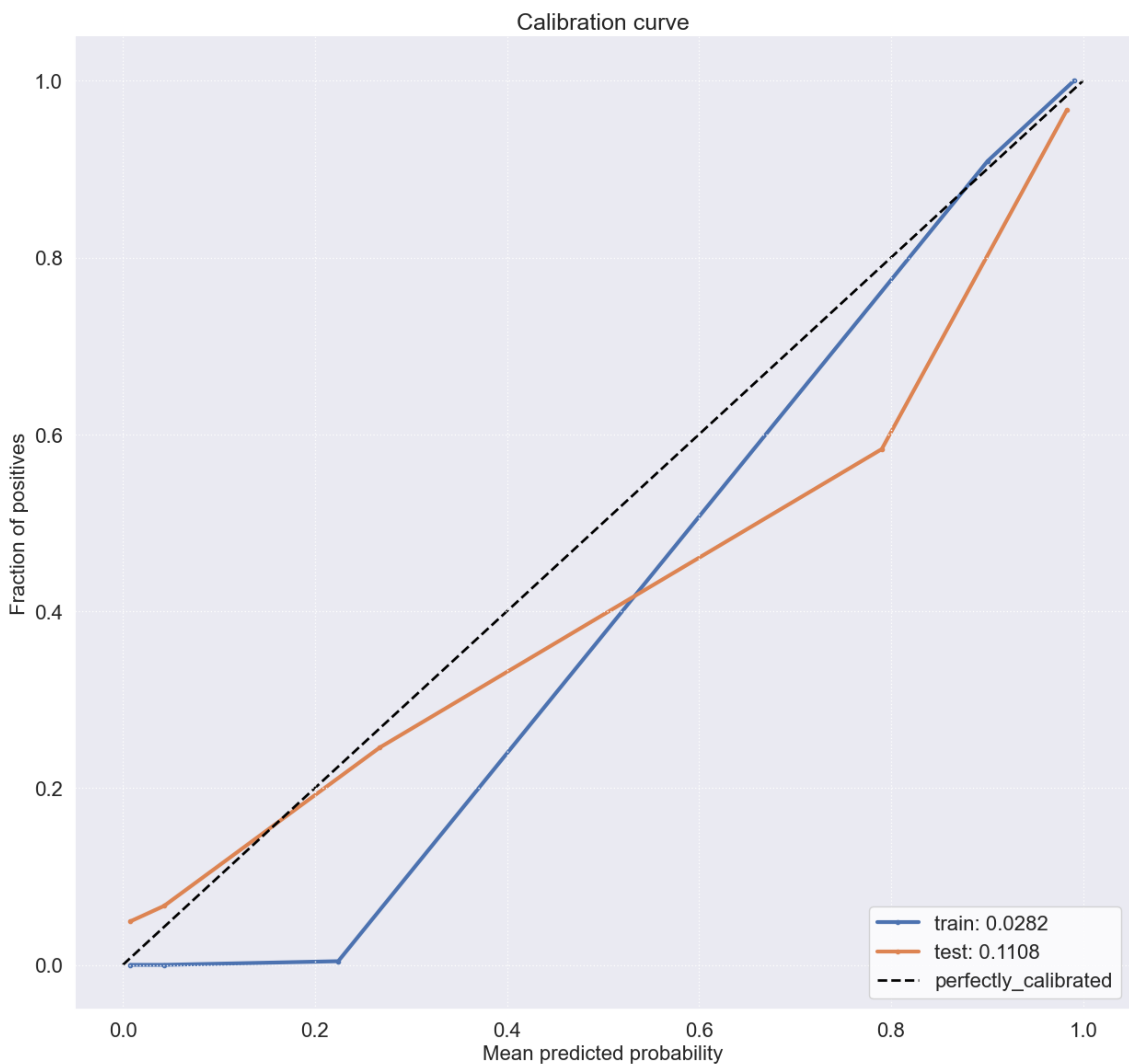- Pareto Front Trials
- Best Trial (# 2879)
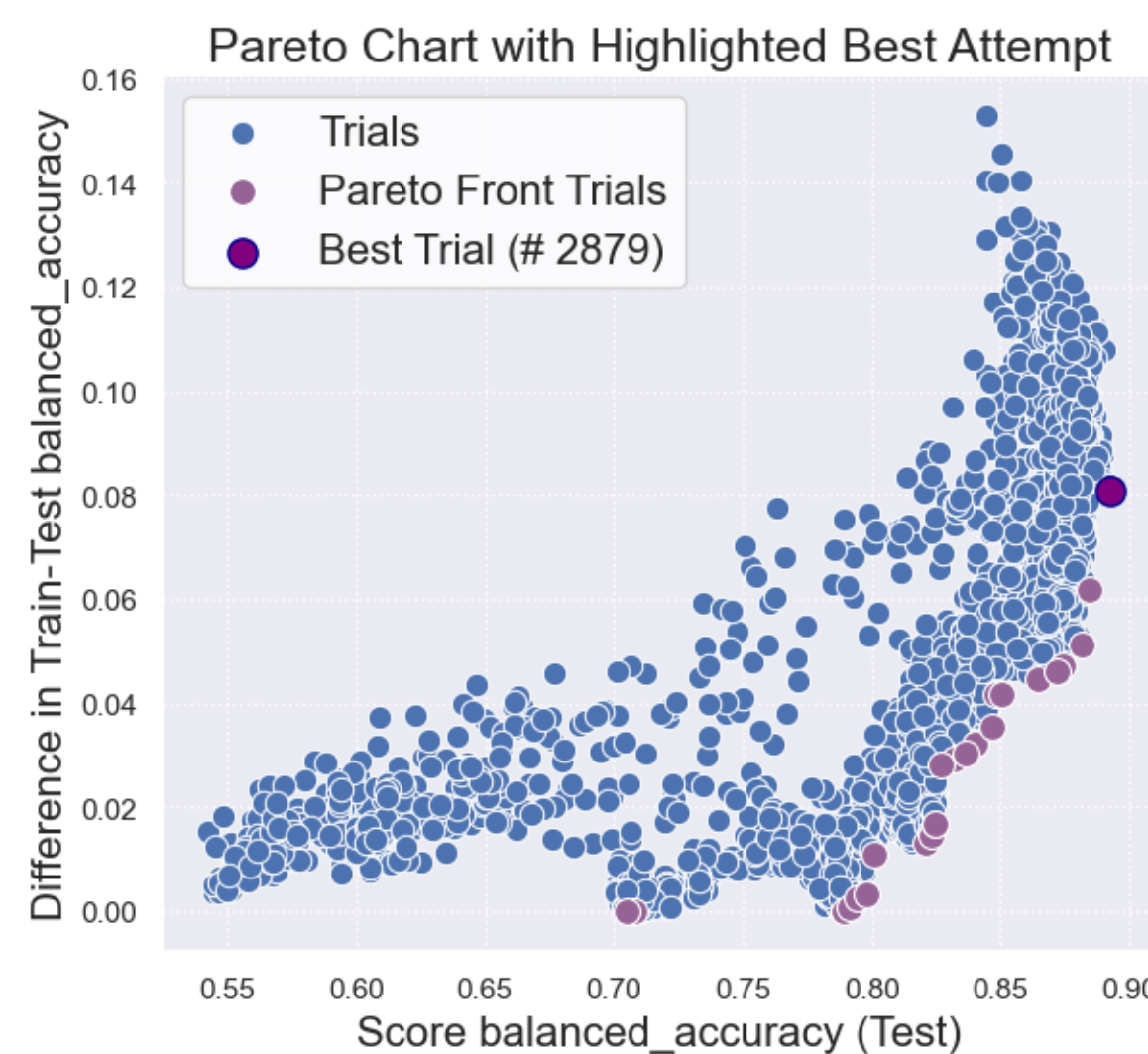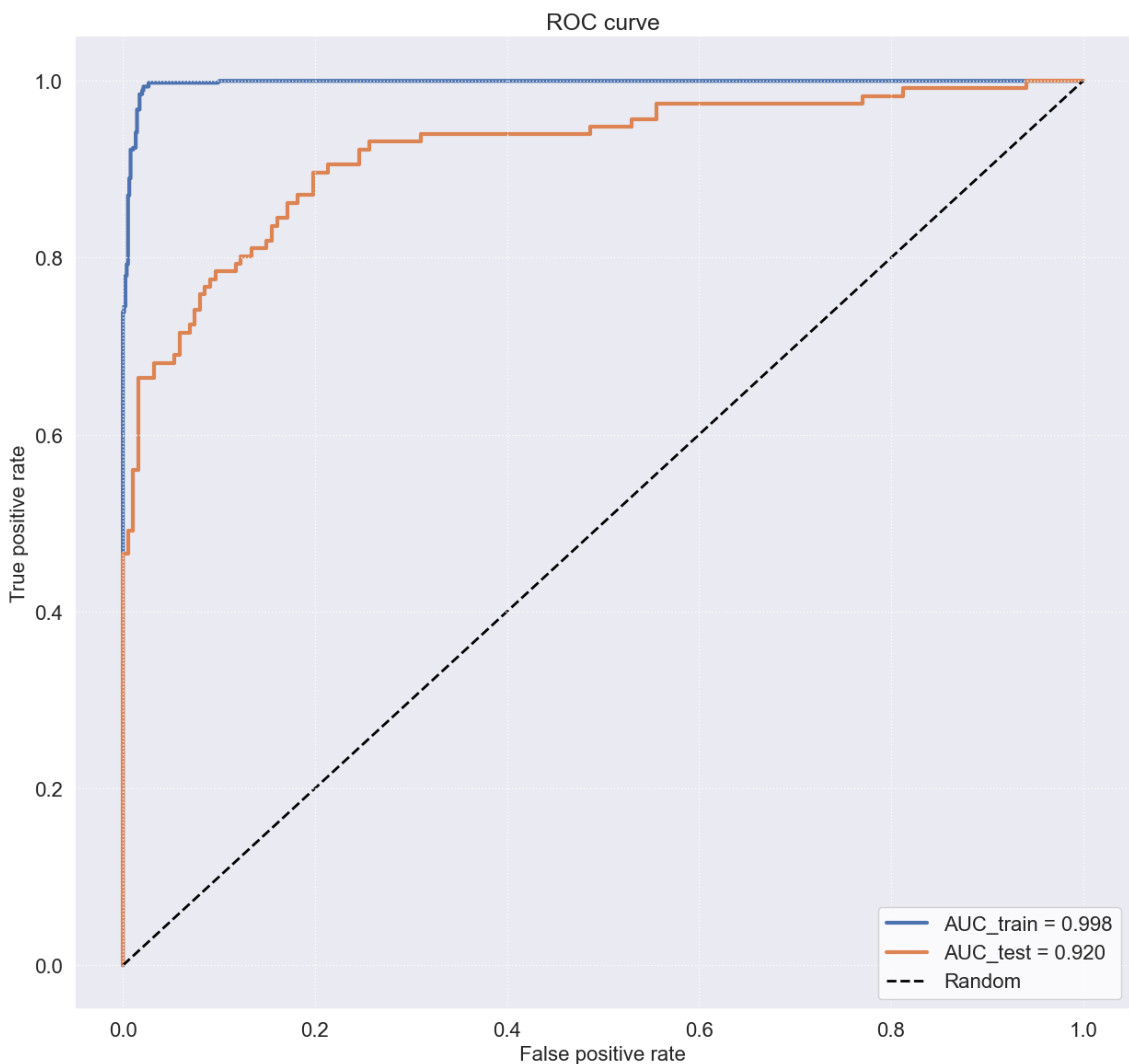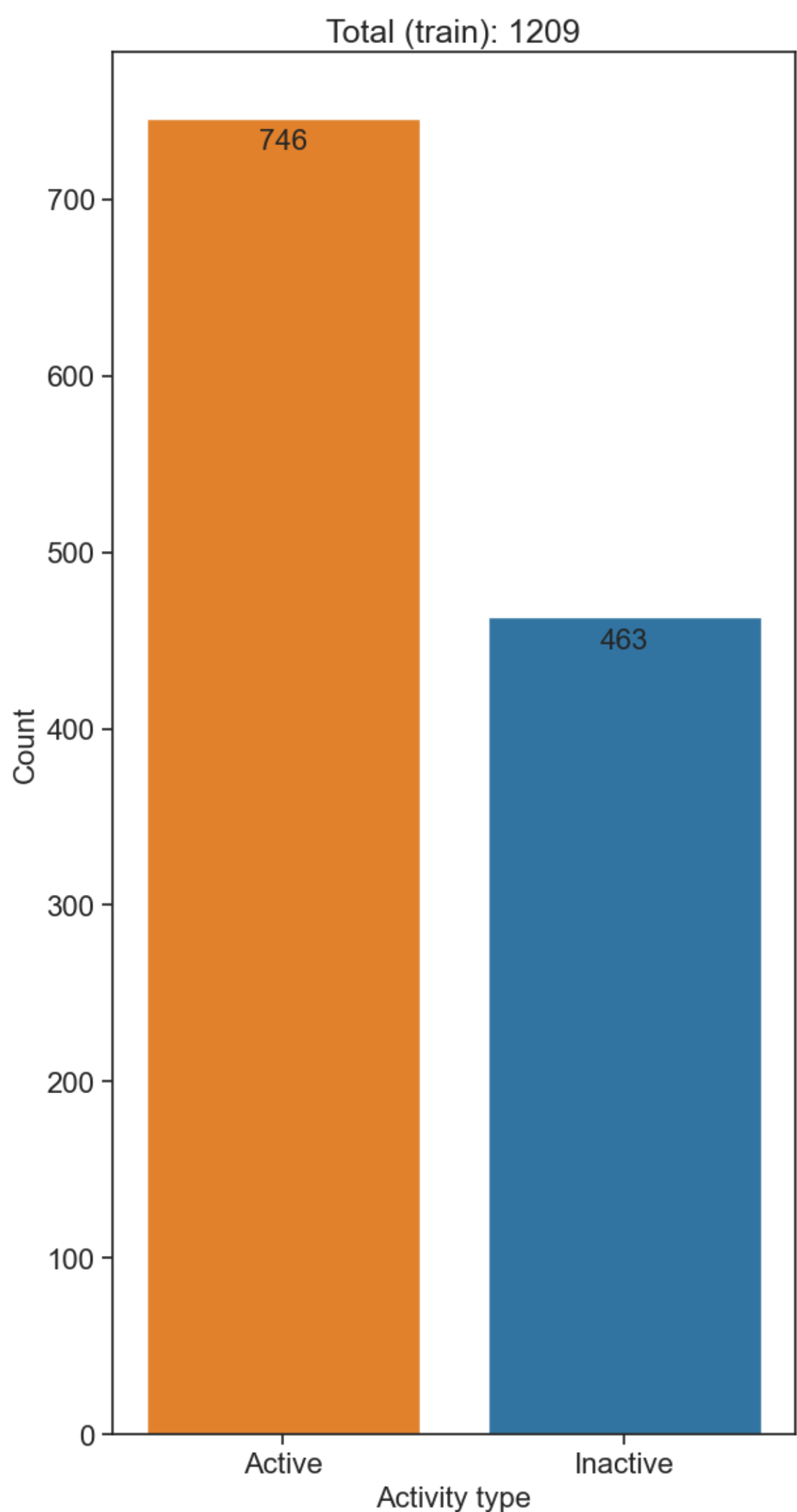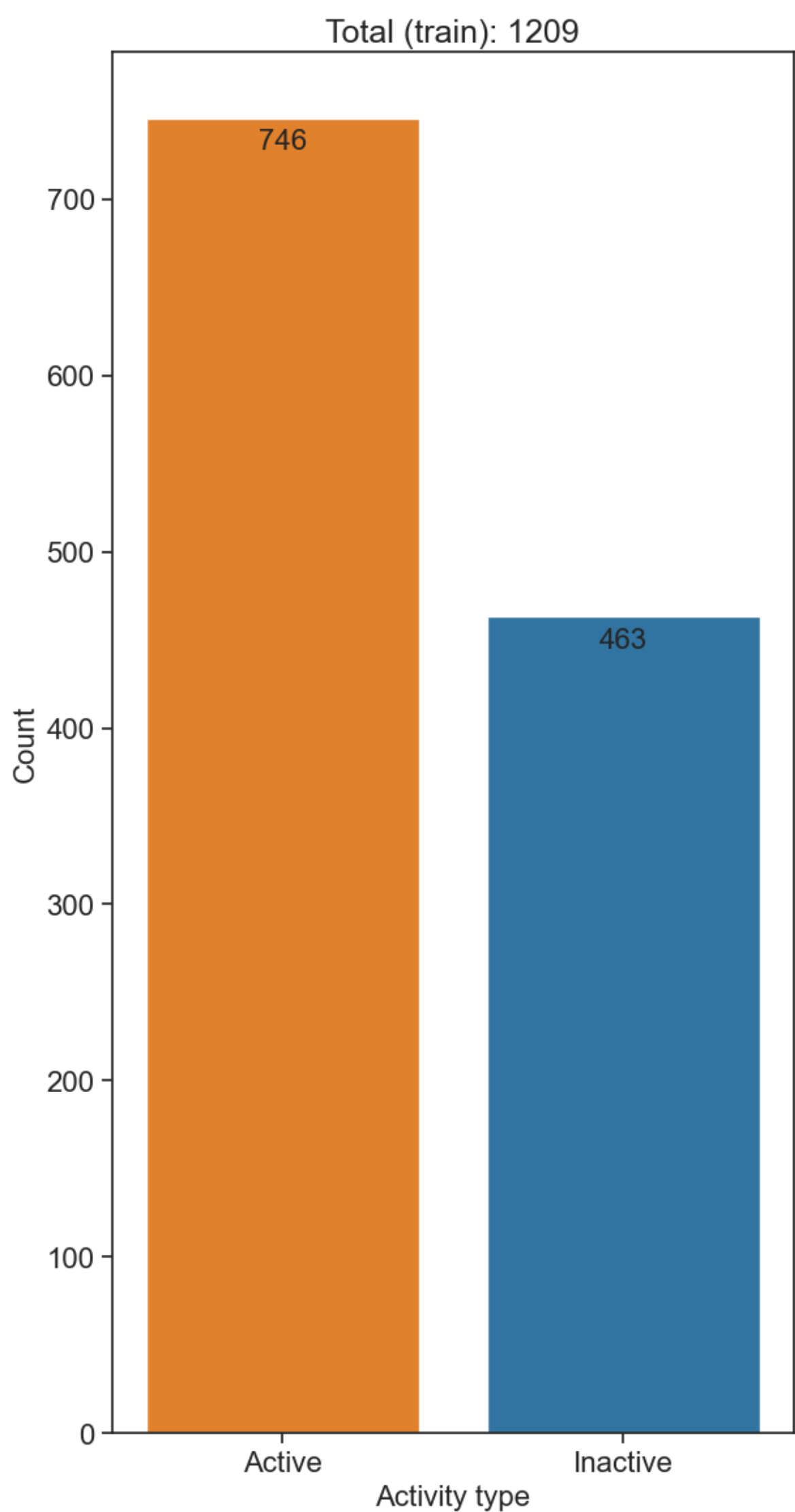
XGBoost Parameters	Value
alpha	4.2928
colsample_bytree	0.6343
gamma	0.2277
lambda	1.4789
learning_rate	0.0391
max_depth	9.0
min_child_weight	6.0
n_estimators	693.0
reg_alpha	0.9624
reg_lambda	0.7318
scale_pos_weight	3.8231
subsample	0.6615

Confusion Matrix

	0	1
0	160	27
1	22	94

Confusion Matrix (nomr.)

	0	1
0	0.86	0.14
1	0.19	0.81

Calibration curve

- train: 0.0282
- test: 0.1108
- perfectly_calibrated

Active/Inactive probability

Inactive zone | Active zone

- active (116)
- inactive (187)
- decoy (302)
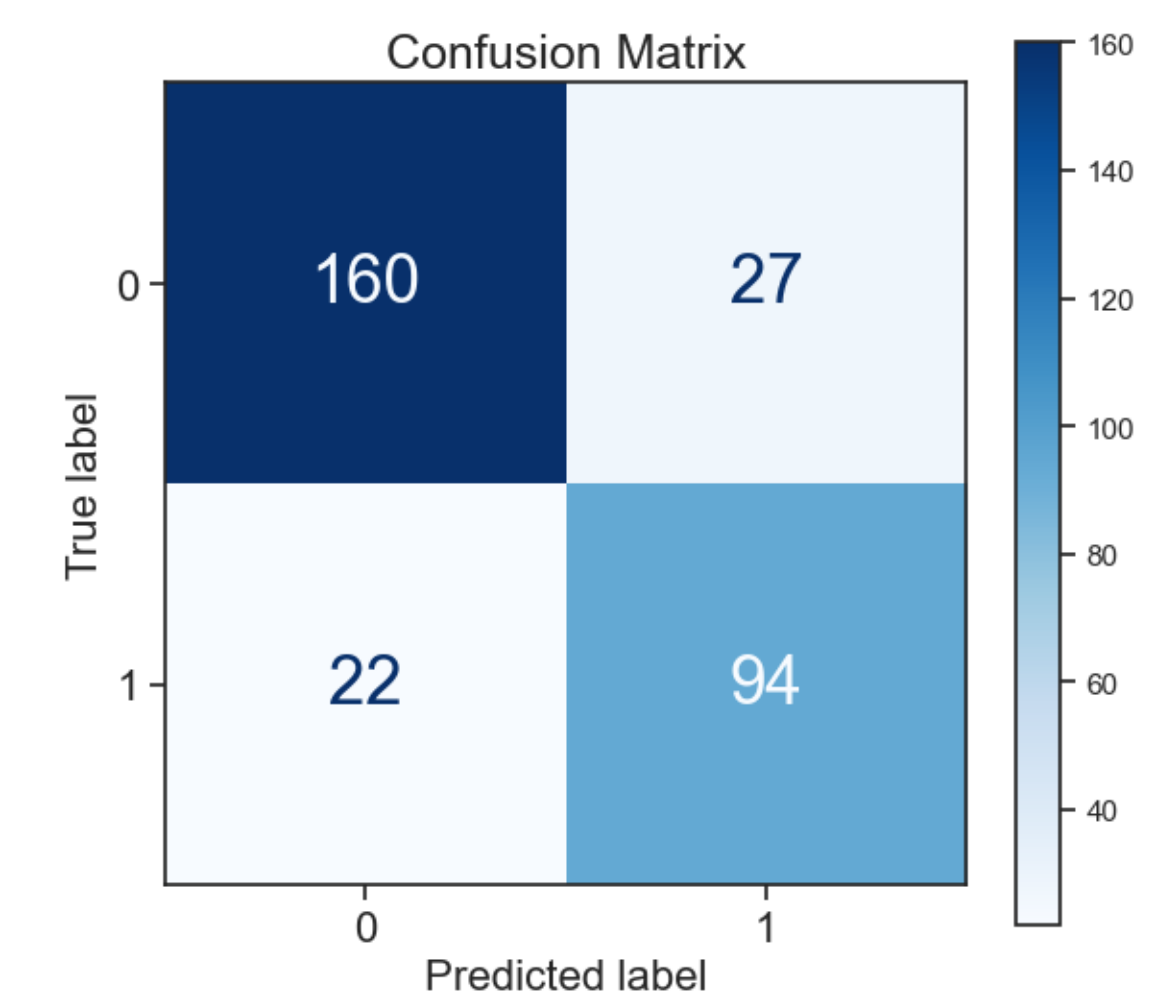
Metric	Train	Test
AUC	0.9976	0.92
accuracy	0.9702	0.8383
balanced accuracy	0.9755	0.833
recall_(sens)	0.9978	0.8103
specificity	0.9531	0.8556
precision	0.9296	0.7769
f1_score	0.9704	0.8389

Protein: P06239 - metric: f1

Total (train): 1209

ROC curve

- AUC_train = 0.999
- AUC_test = 0.917
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 1632)
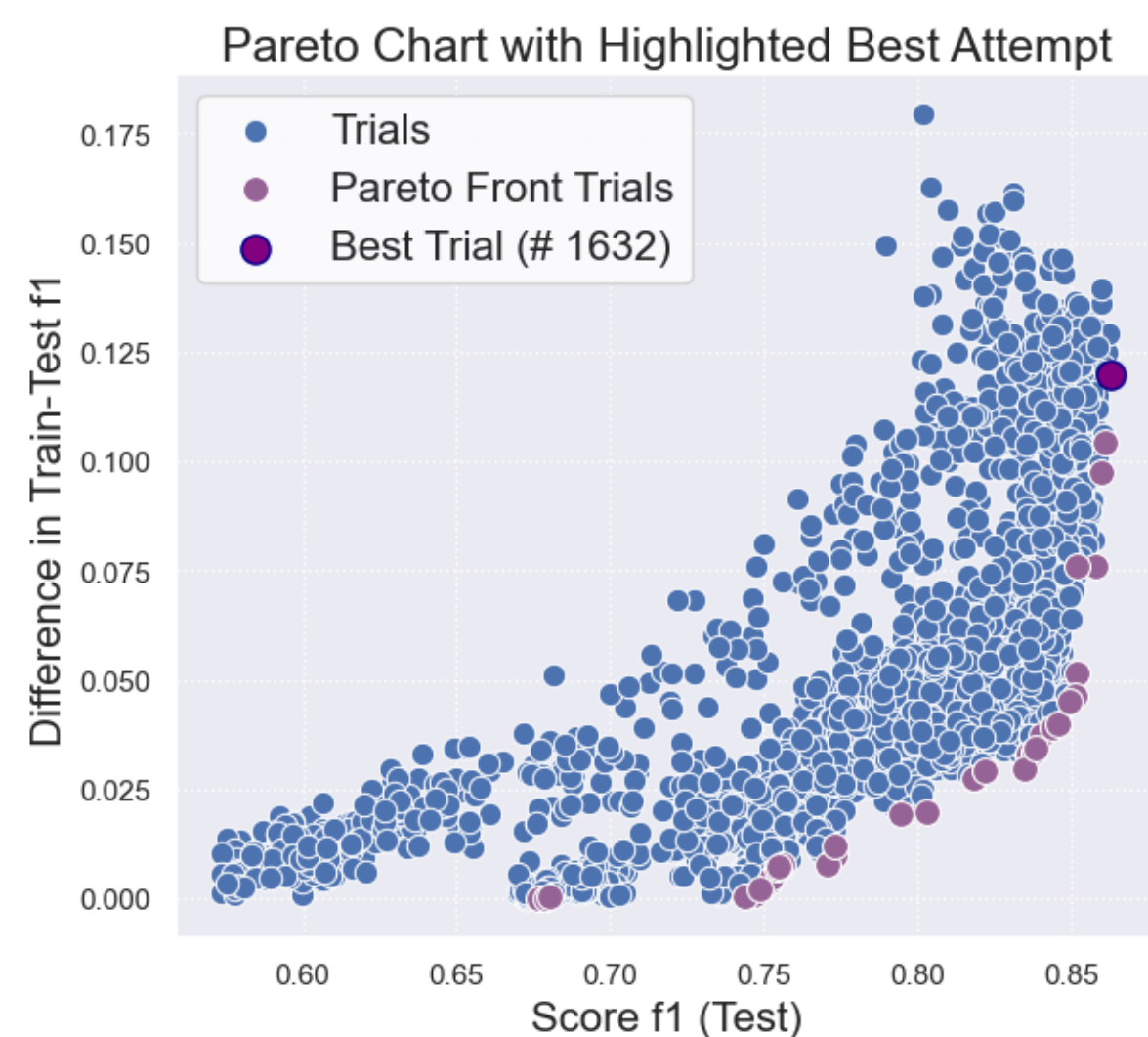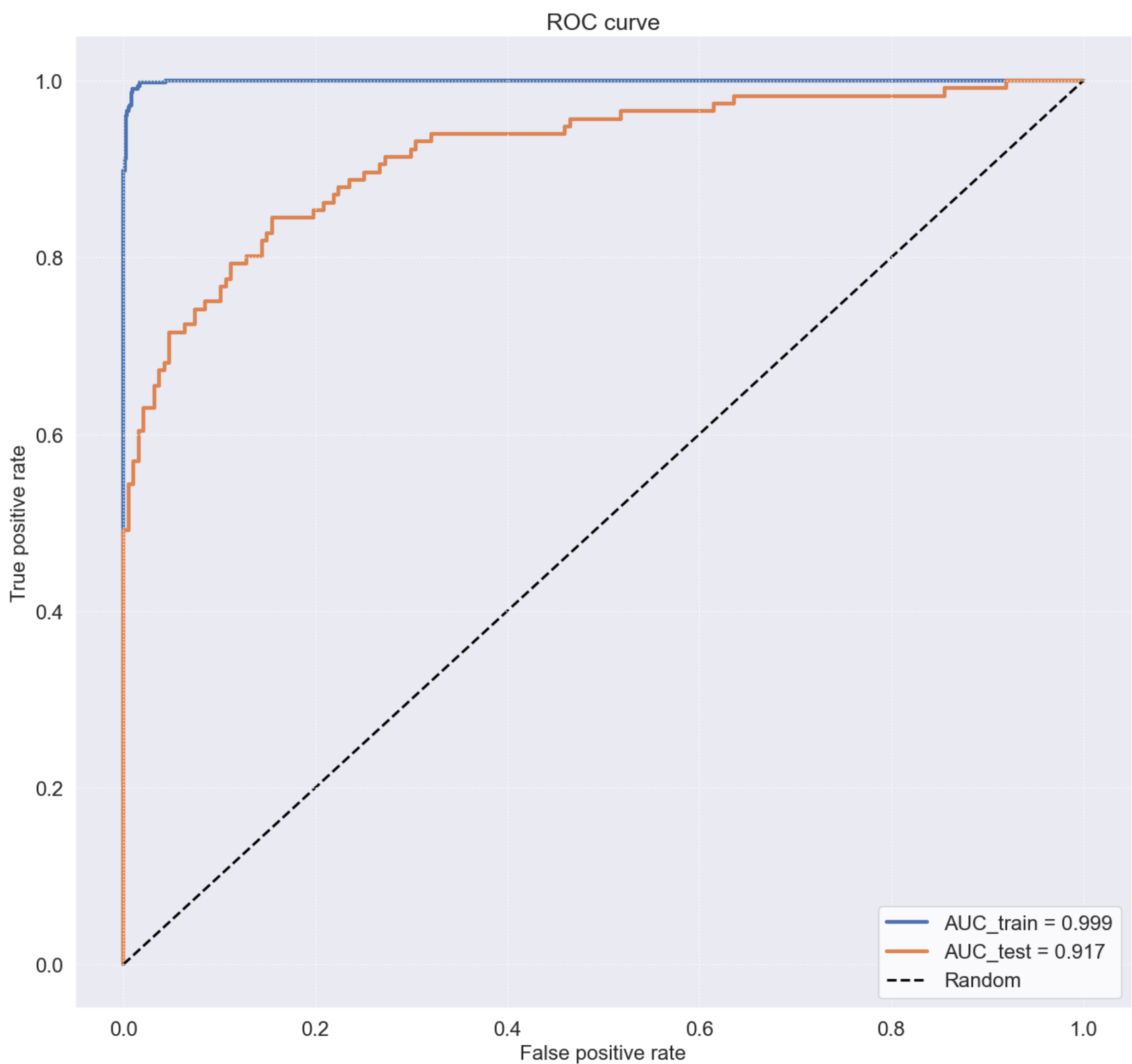
XGBoost Parameters	Value
alpha	3.9246
colsample_bytree	0.7854
gamma	0.0509
lambda	0.4696
learning_rate	0.0625
max_depth	10.0
min_child_weight	4.0
n_estimators	964.0
reg_alpha	0.2798
reg_lambda	0.9517
scale_pos_weight	1.8483
subsample	0.5424

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

- train: 0.0132
- test: 0.1115
- perfectly_calibrated

Active/Inactive probability

- active (116)
- inactive (187)
- decoy (302)
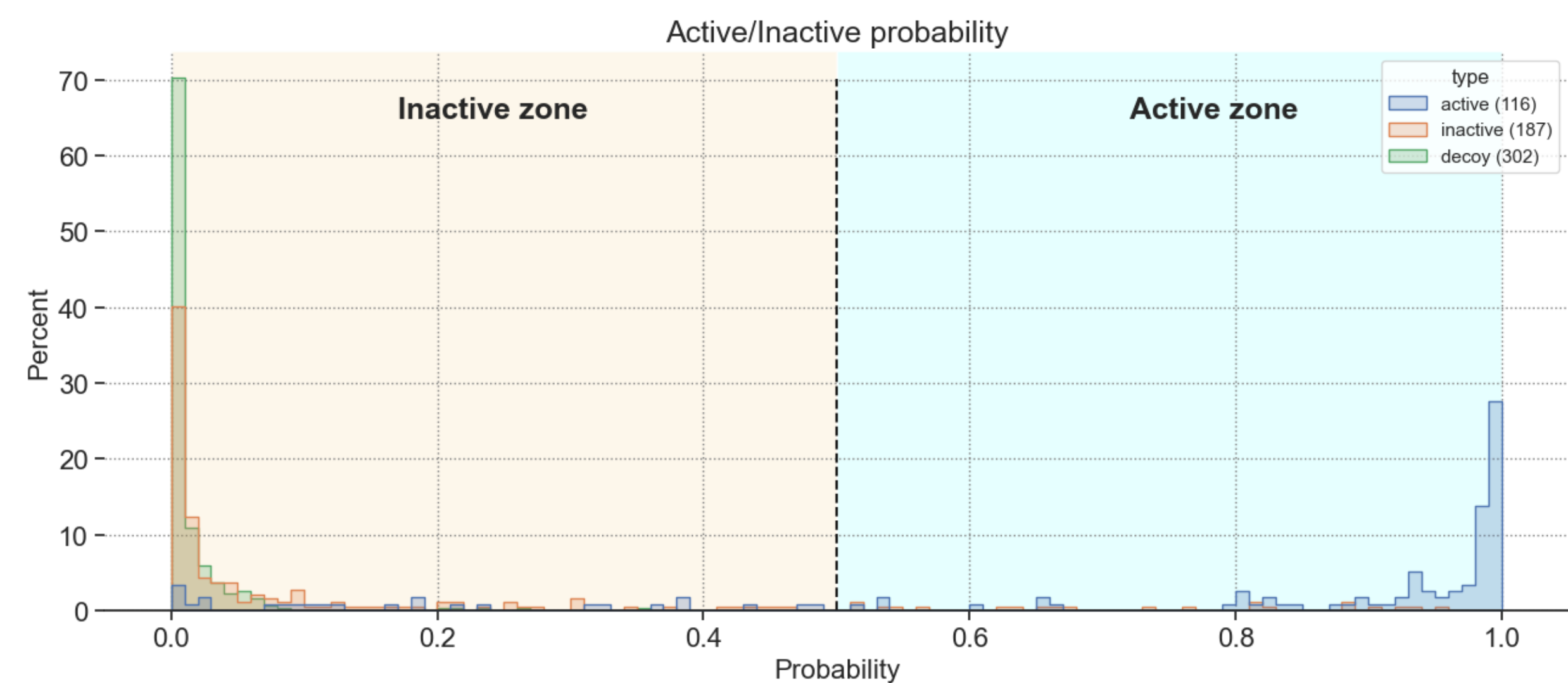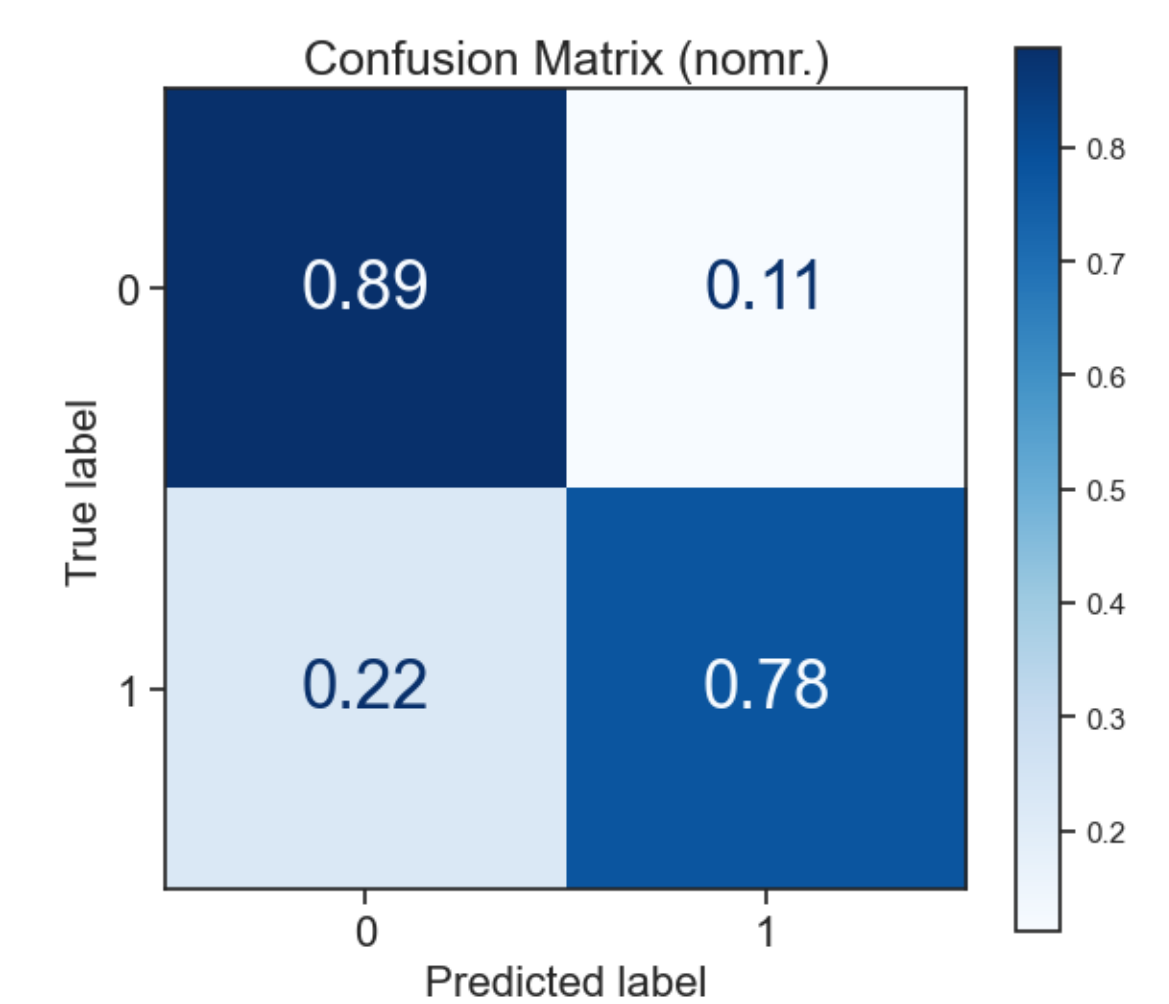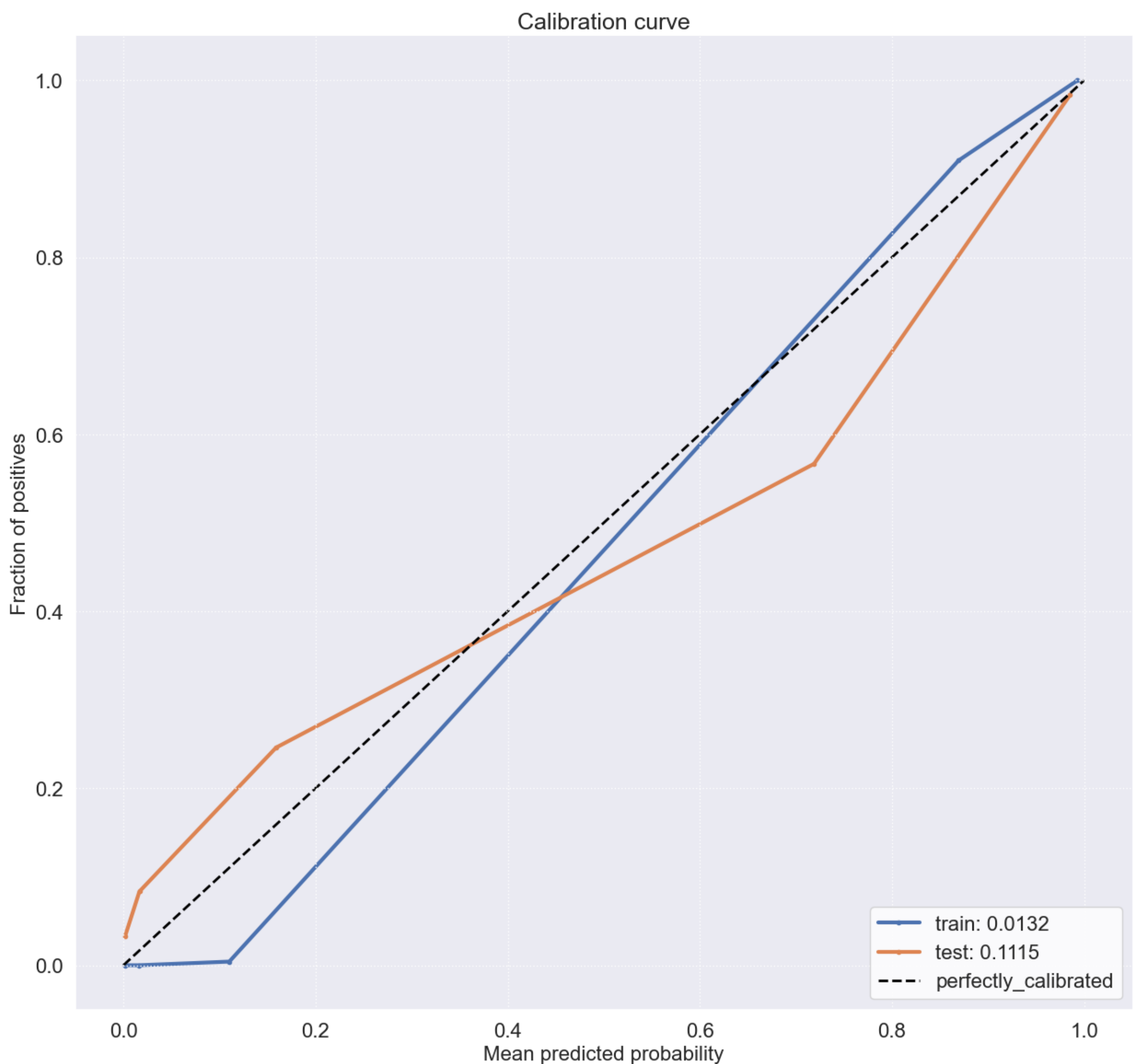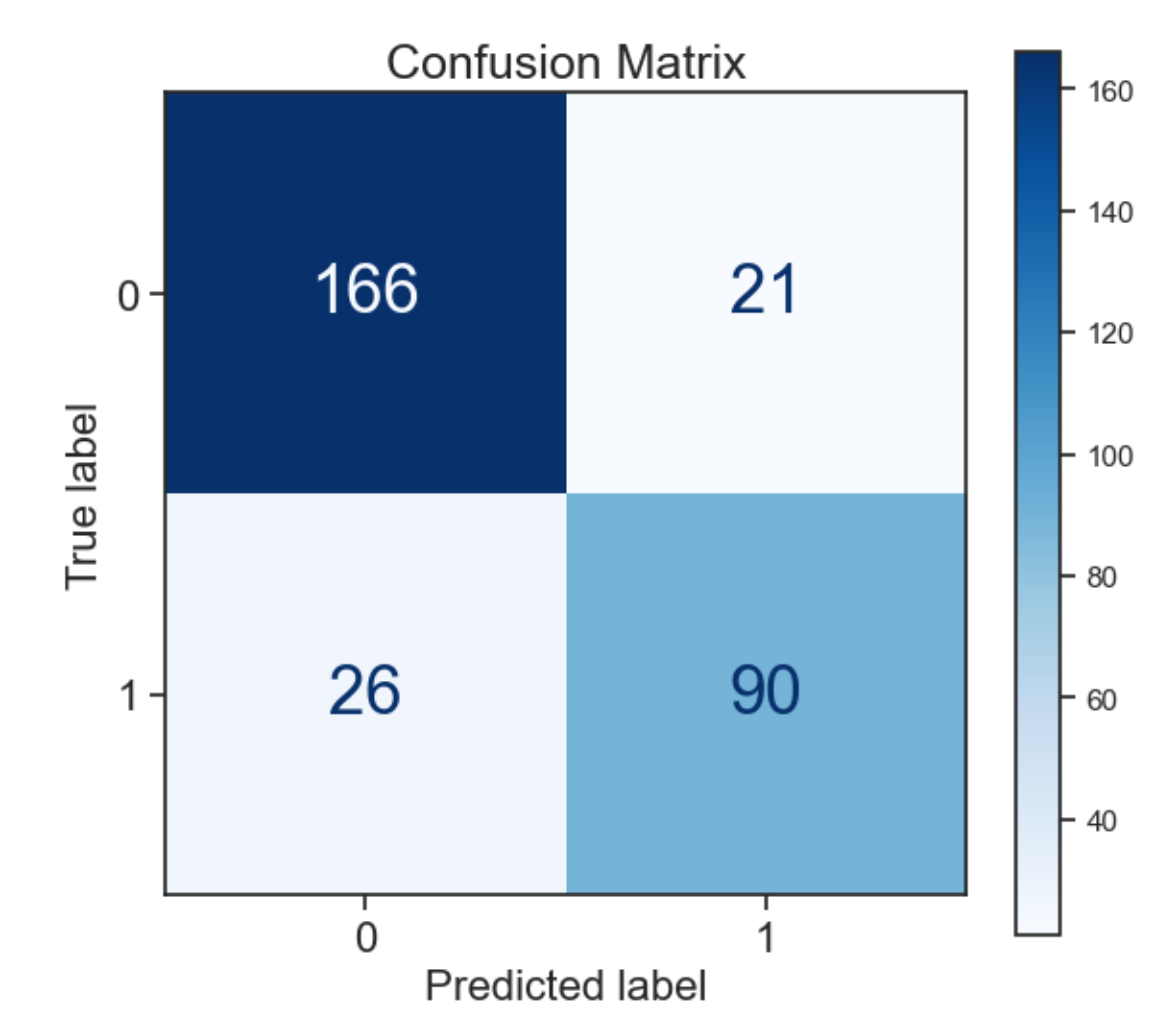
Metric	Train	Test
AUC	0.9994	0.9174
accuracy	0.9884	0.8449
balanced accuracy	0.9902	0.8318
recall_(sens)	0.9978	0.7759
specificity	0.9826	0.8877
precision	0.9726	0.8108
f1_score	0.9884	0.8442

Protein: P06239 - metric: precision

Total (train): 1209

ROC curve

AUC_train = 1.000
AUC_test = 0.916
Random

Pareto Chart with Highlighted Best Attempt

Trials
Pareto Front Trials
Best Trial (# 2923)

Confusion Matrix

	0	1
0	165	22
1	27	89

Confusion Matrix (nomr.)

	0	1
0	0.88	0.12
1	0.23	0.77

Calibration curve

train: 0.0071
test: 0.1141
perfectly_calibrated
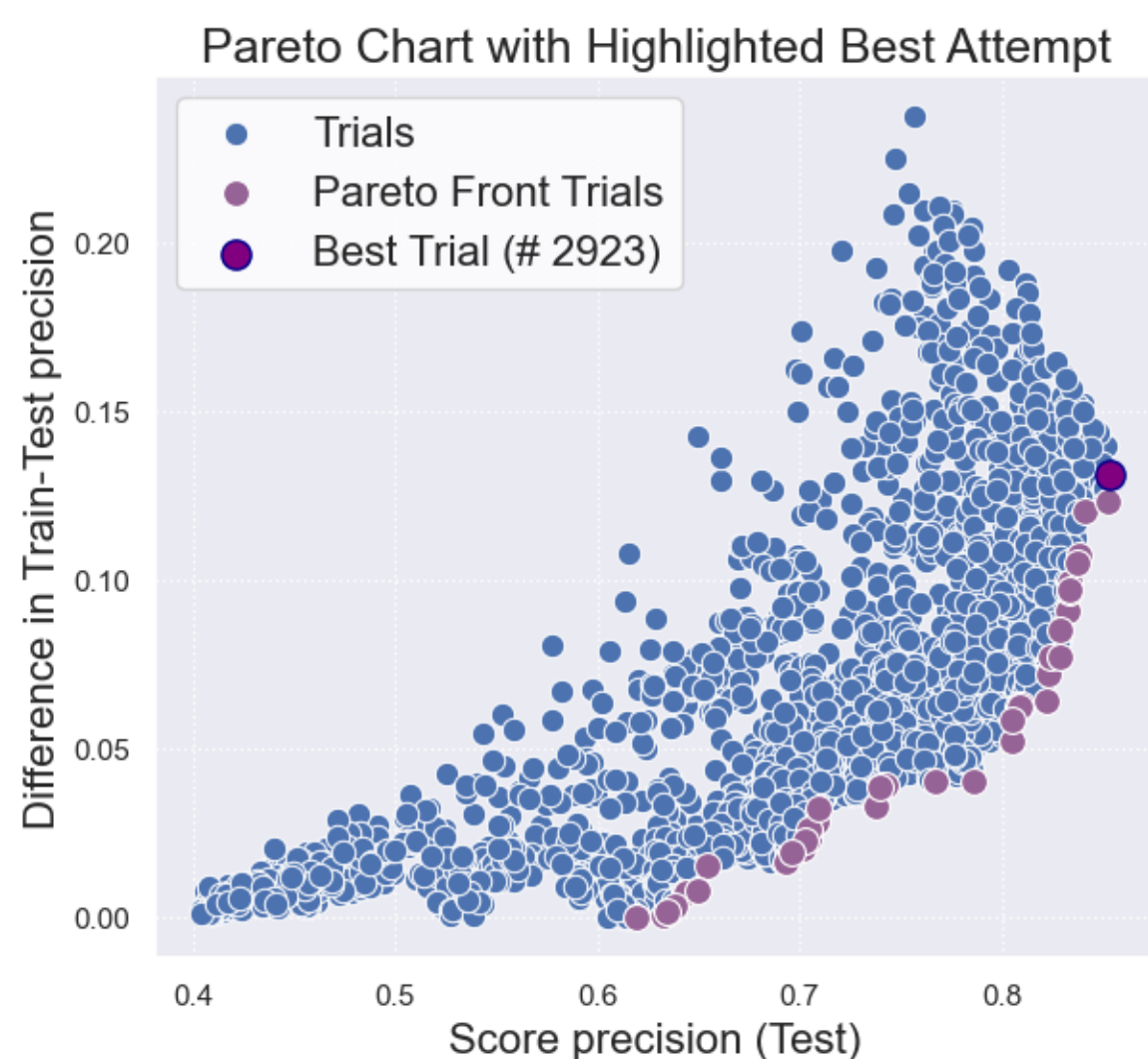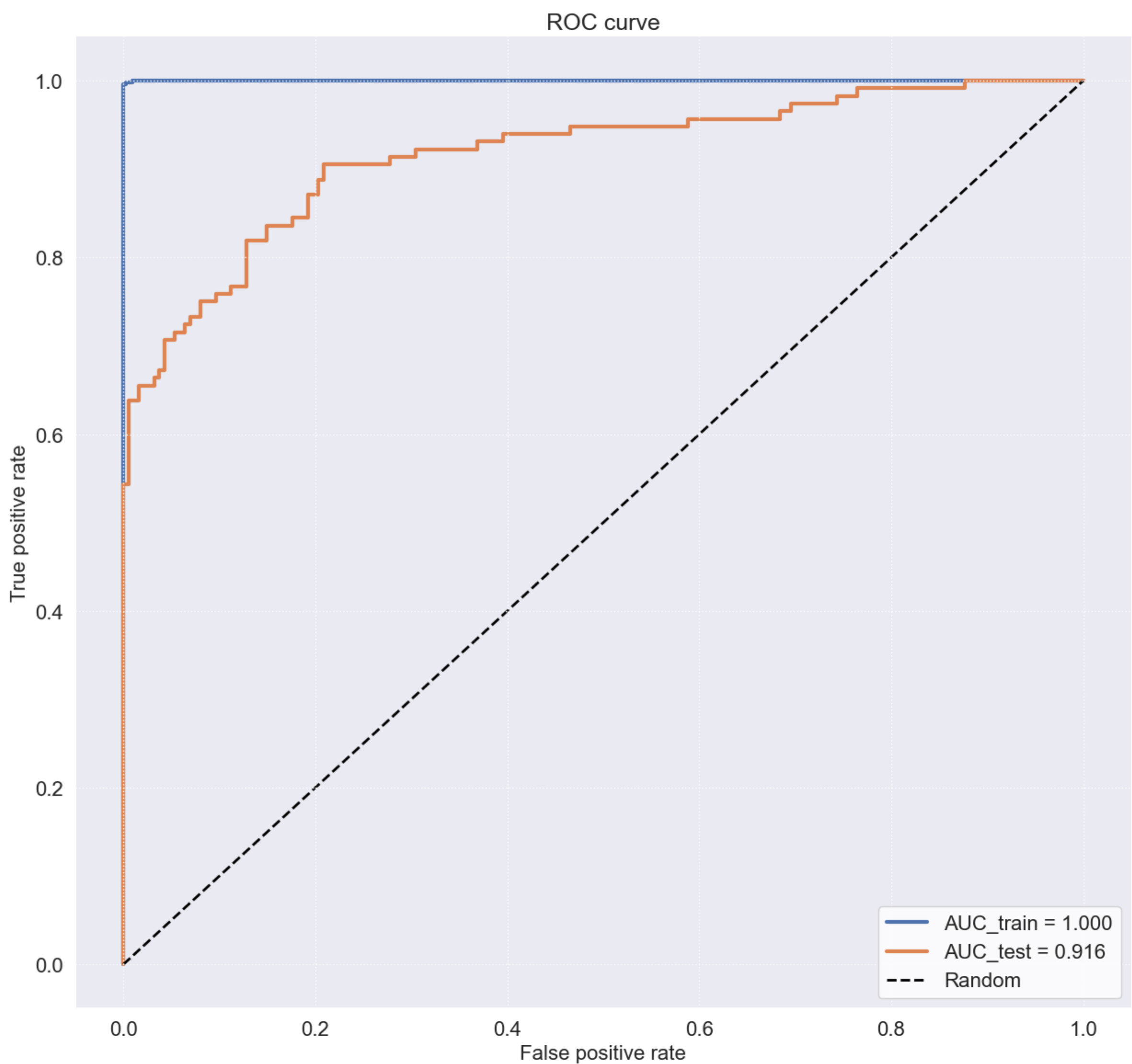
XGBoost Parameters	Value
alpha	4.8434
colsample_bytree	0.5829
gamma	0.1139
lambda	4.4875
learning_rate	0.0759
max_depth	10.0
min_child_weight	3.0
n_estimators	510.0
reg_alpha	0.2808
reg_lambda	0.441
scale_pos_weight	2.5458
subsample	0.7969

Active/Inactive probability

type
active (116)
inactive (187)
decoy (302)

Inactive zone   Active zone
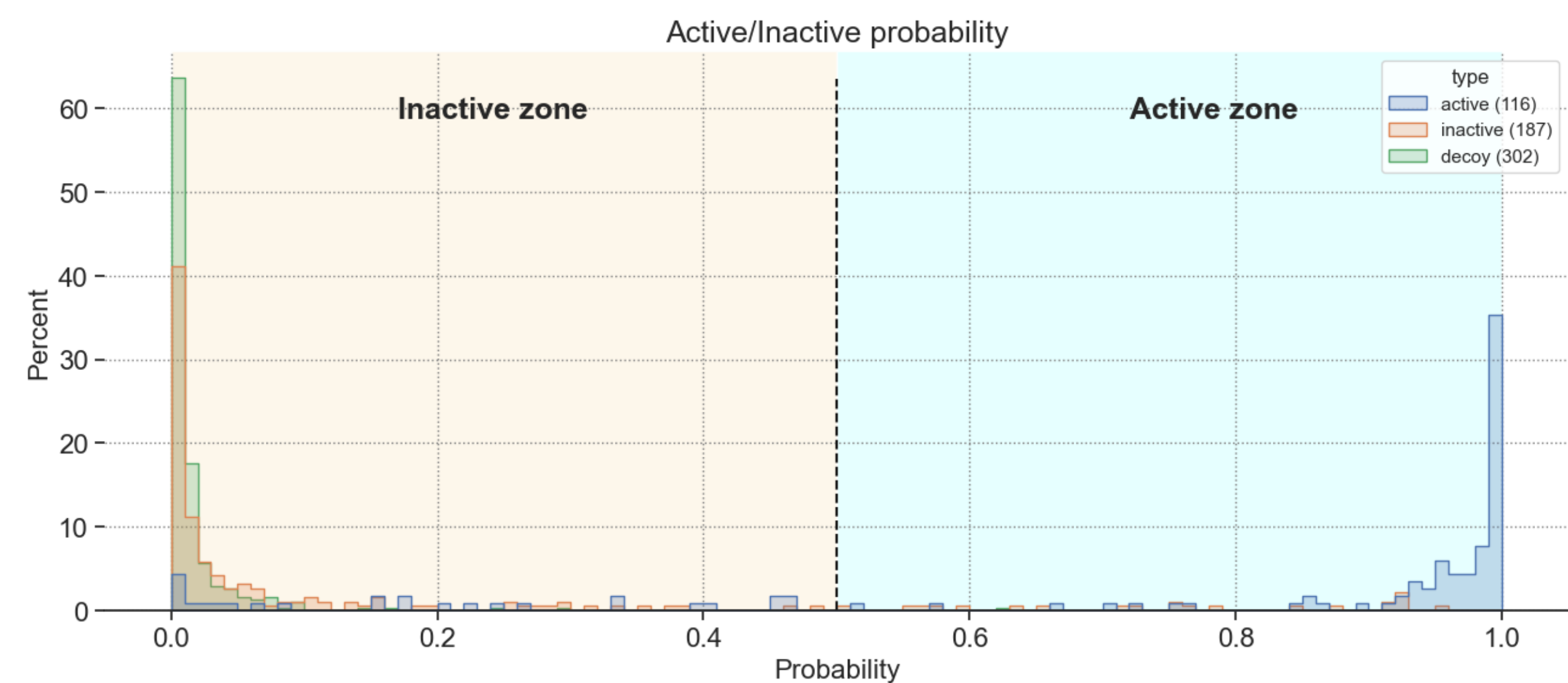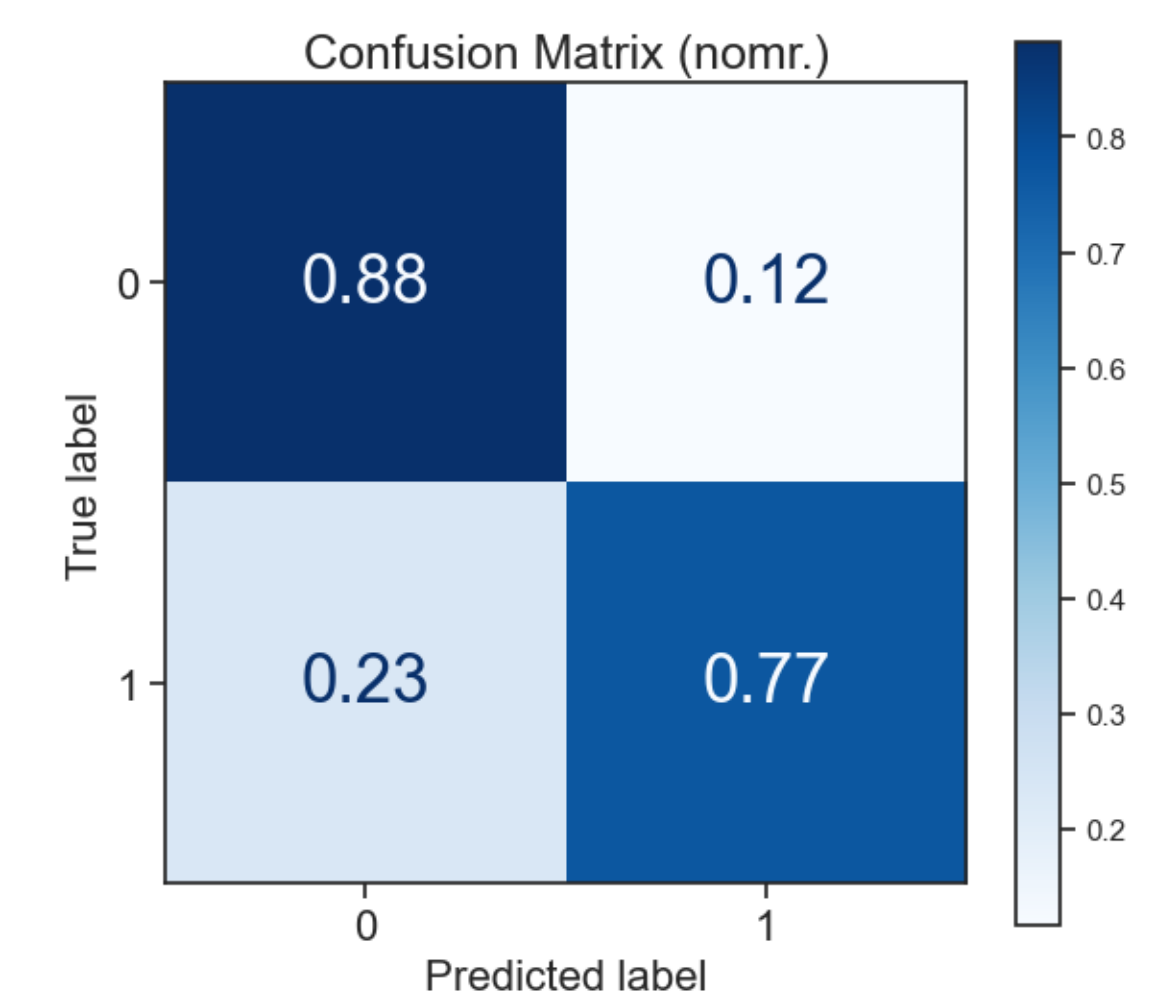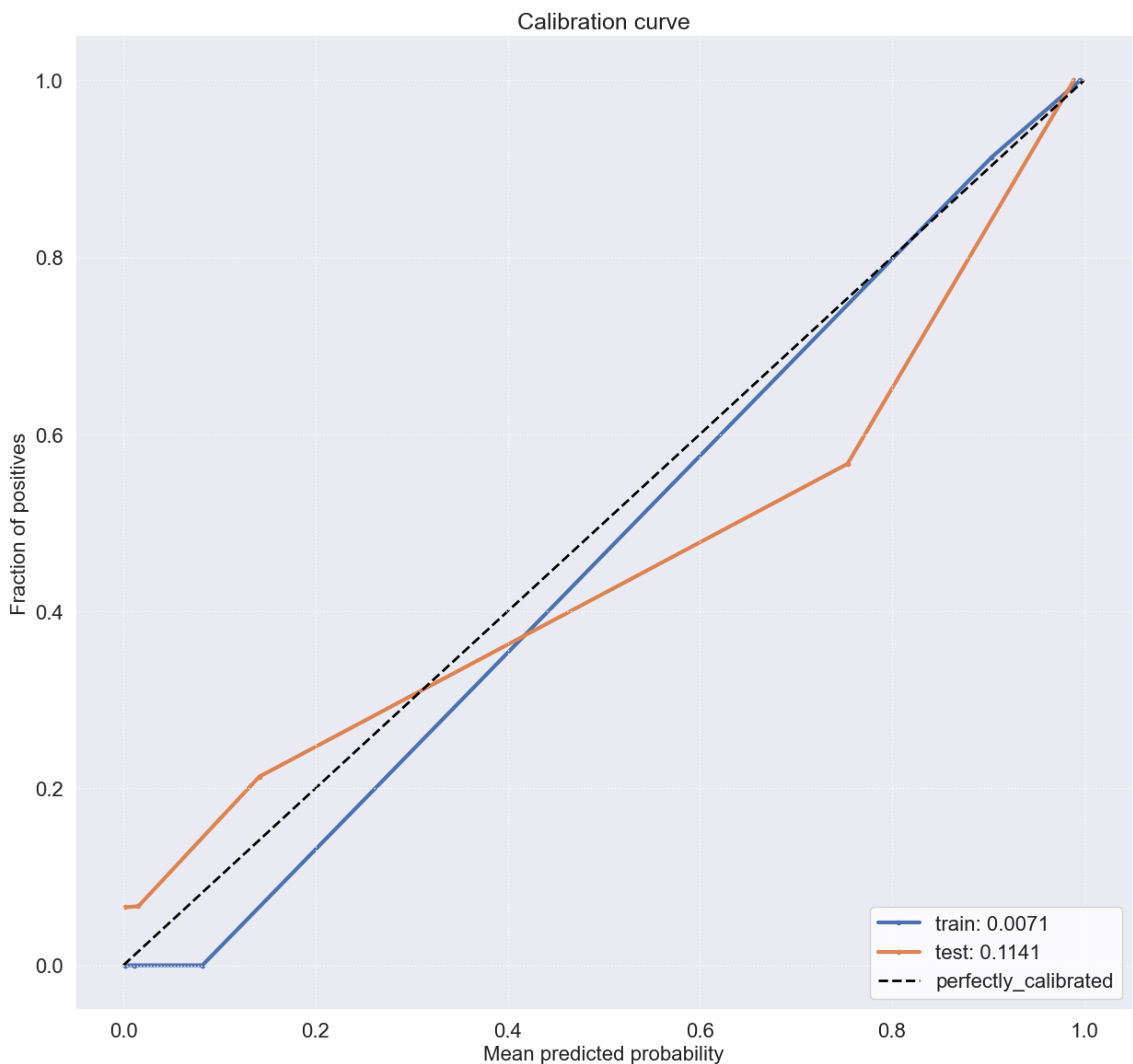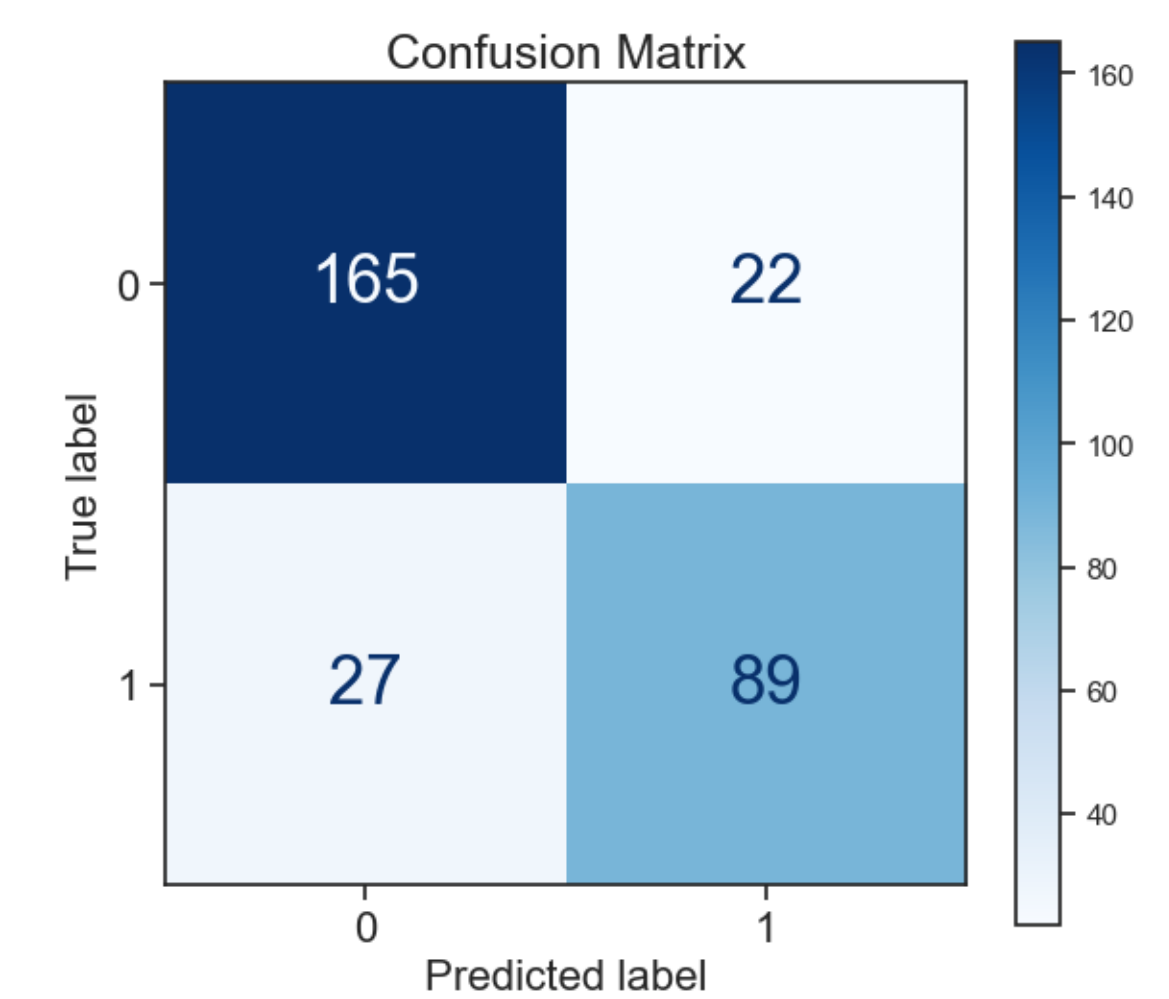
Metric	Train	Test
AUC	1.0	0.9164
accuracy	0.9942	0.8383
balanced accuracy	0.9953	0.8248
recall_(sens)	1.0	0.7672
specificity	0.9906	0.8824
precision	0.9851	0.8018
f1_score	0.9942	0.8376

Protein: P06239 - metric: SoftVotingXGBoostEnsemble

Total (train): 1209

Active: 746
Inactive: 463

ROC curve

AUC_train = 1.000
AUC_test = 0.919
Random

Confusion Matrix

	0	1
0	166	21
1	23	93

Confusion Matrix (nomr.)

	0	1
0	0.89	0.11
1	0.2	0.8

Calibration curve

train: 0.0143
test: 0.1094
perfectly_calibrated

Active/Inactive probability

type
active (116)
inactive (187)
decoy (302)

Inactive zone | Active zone
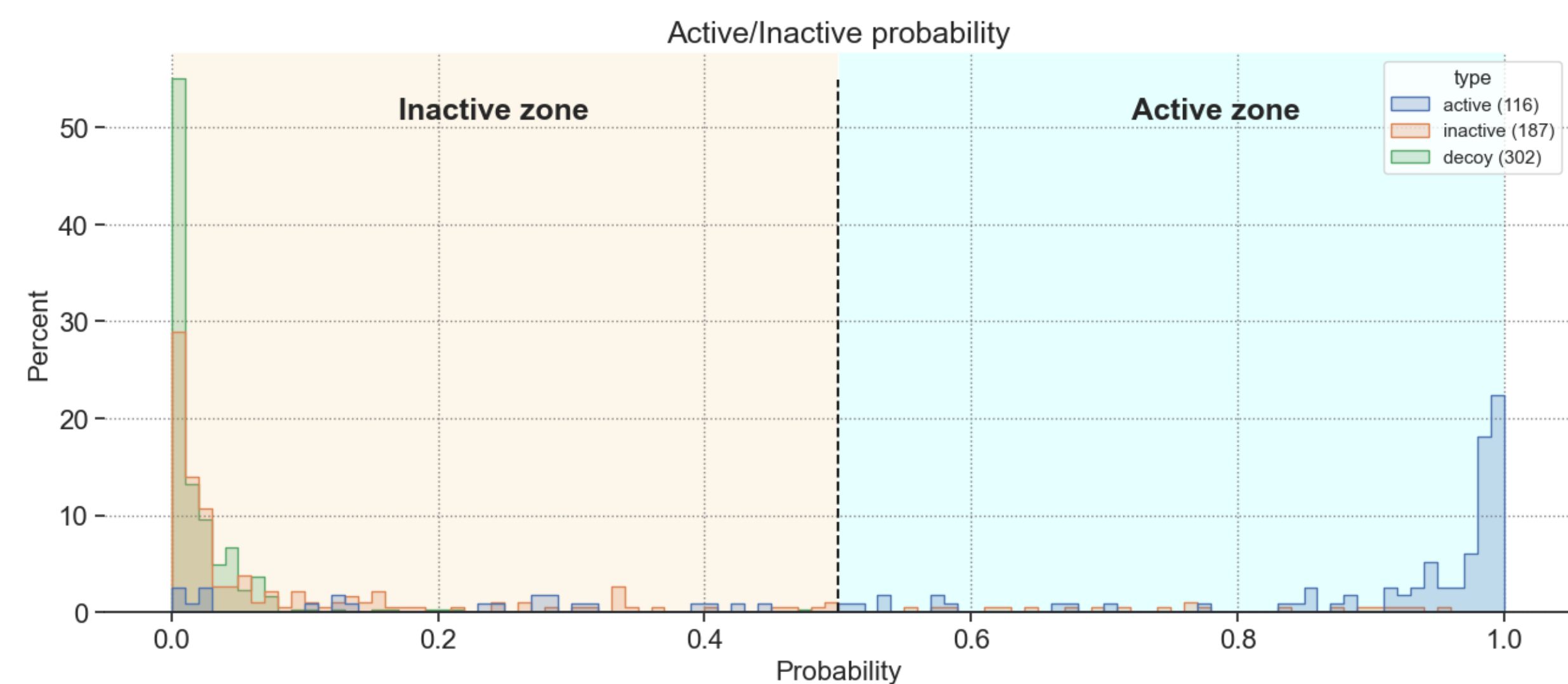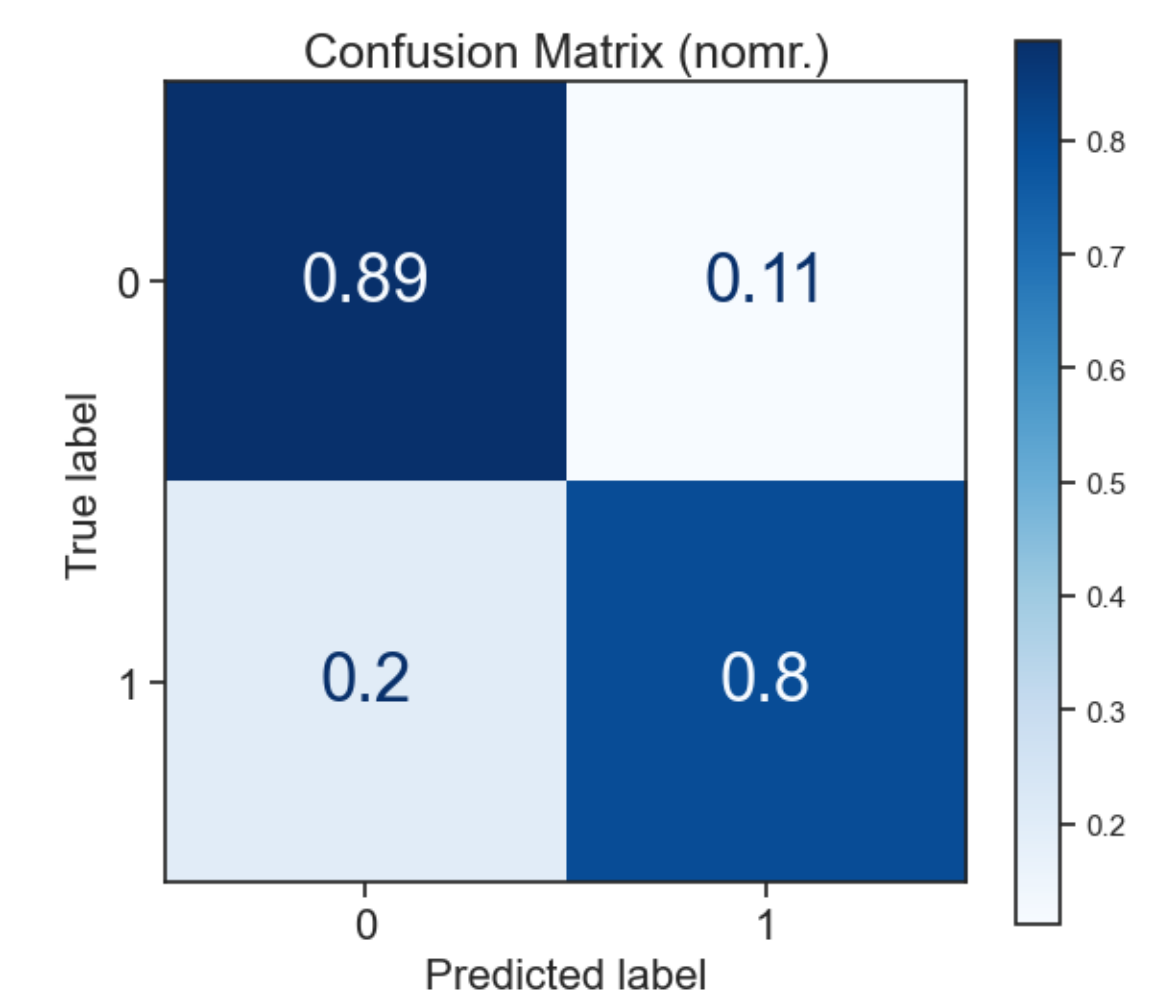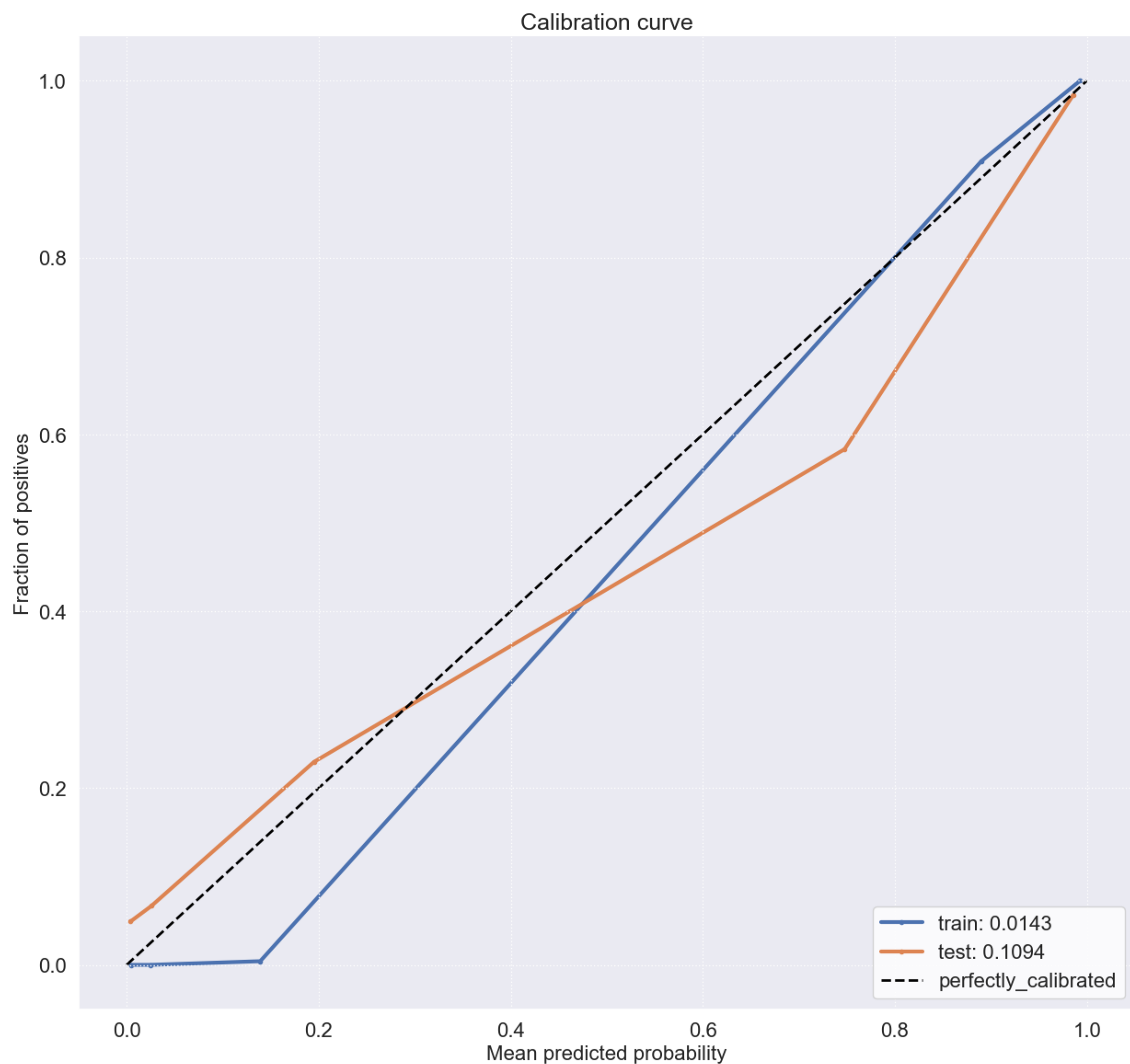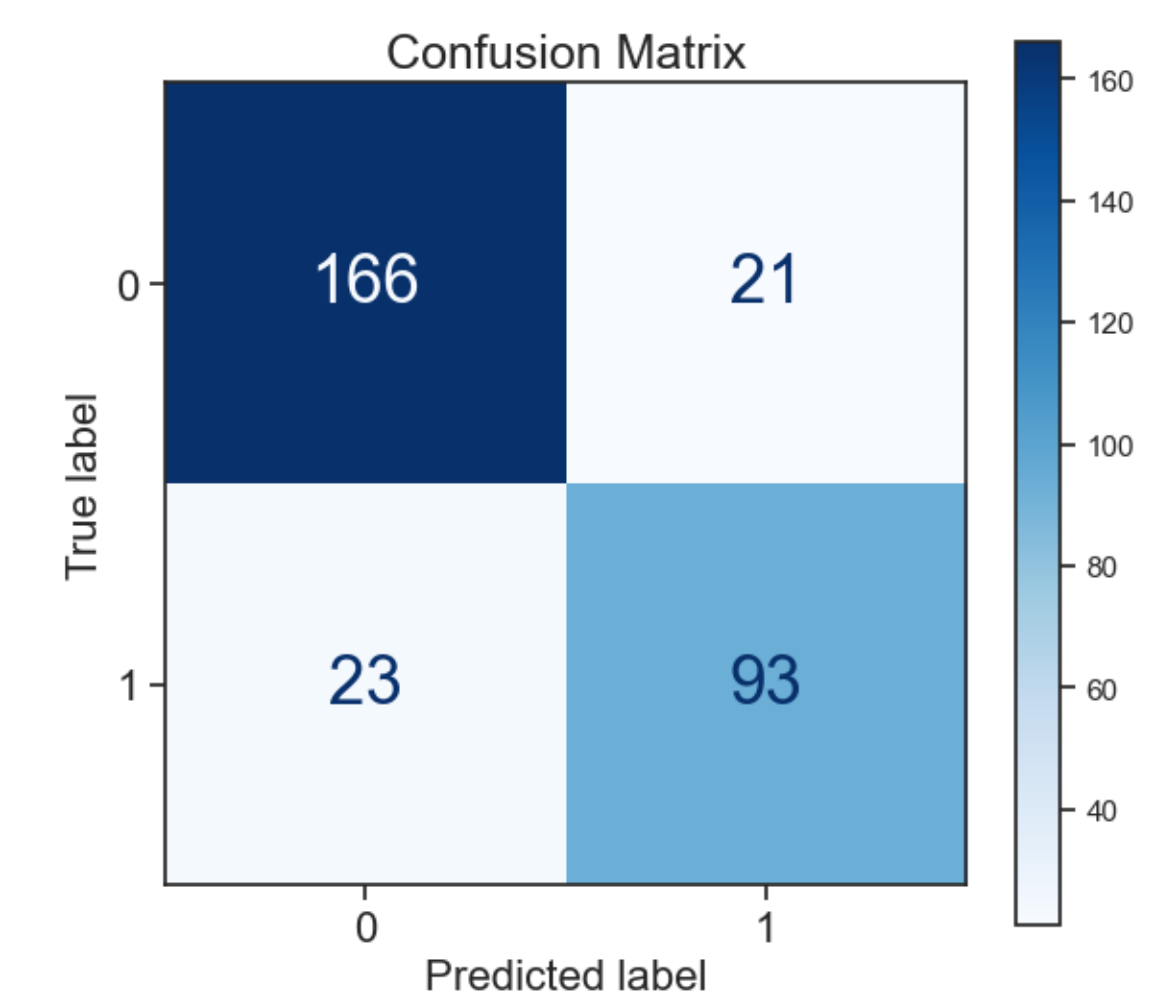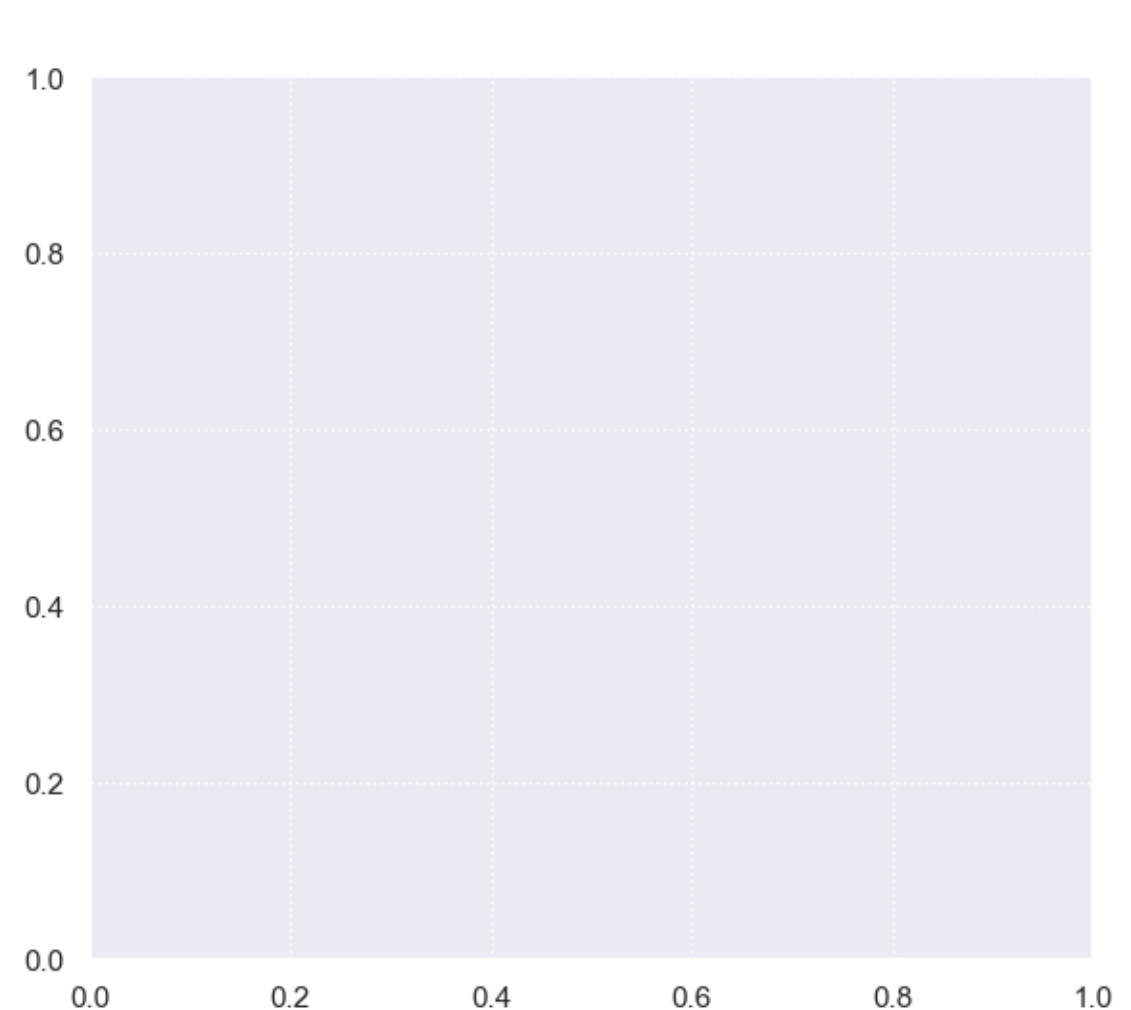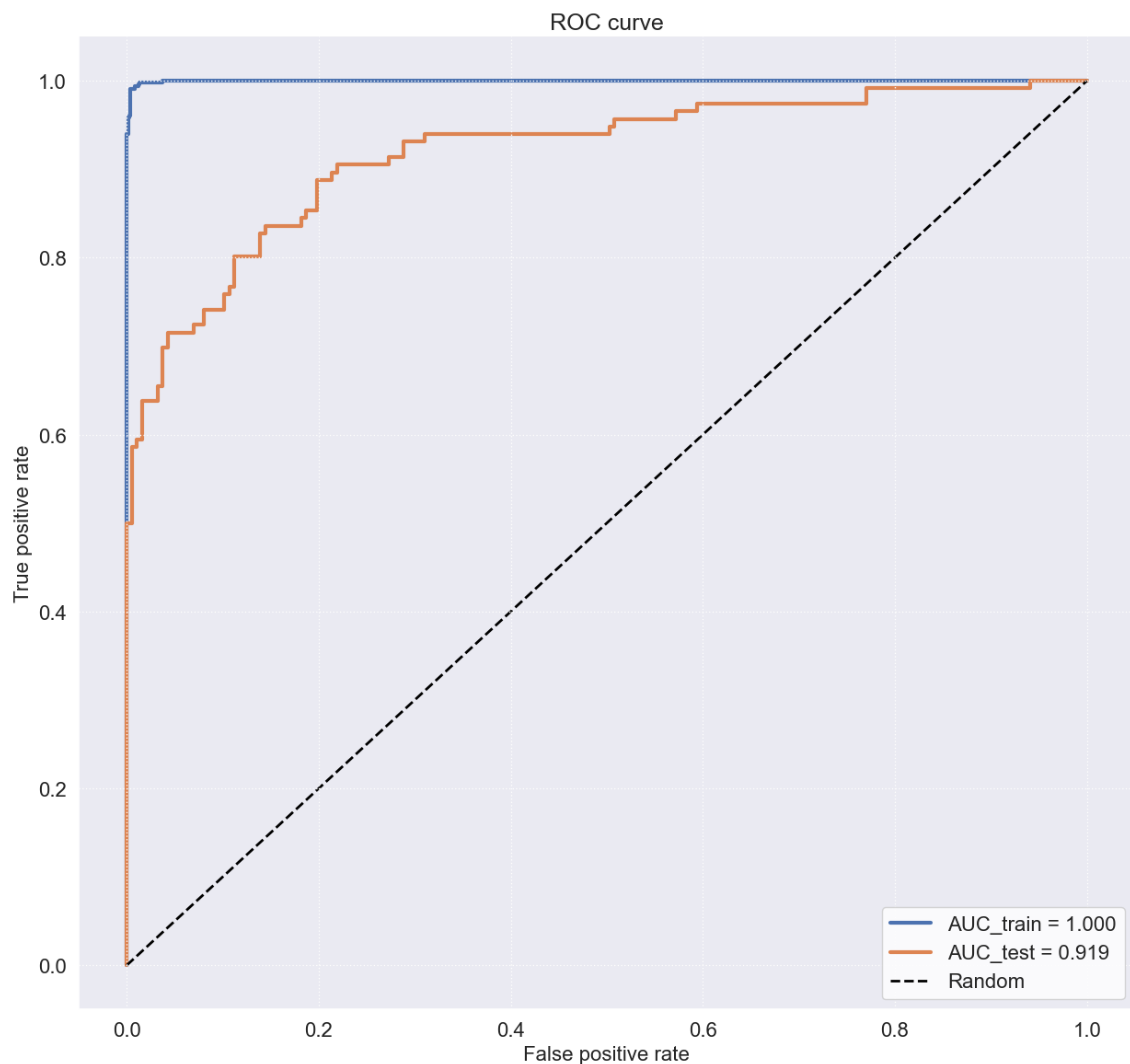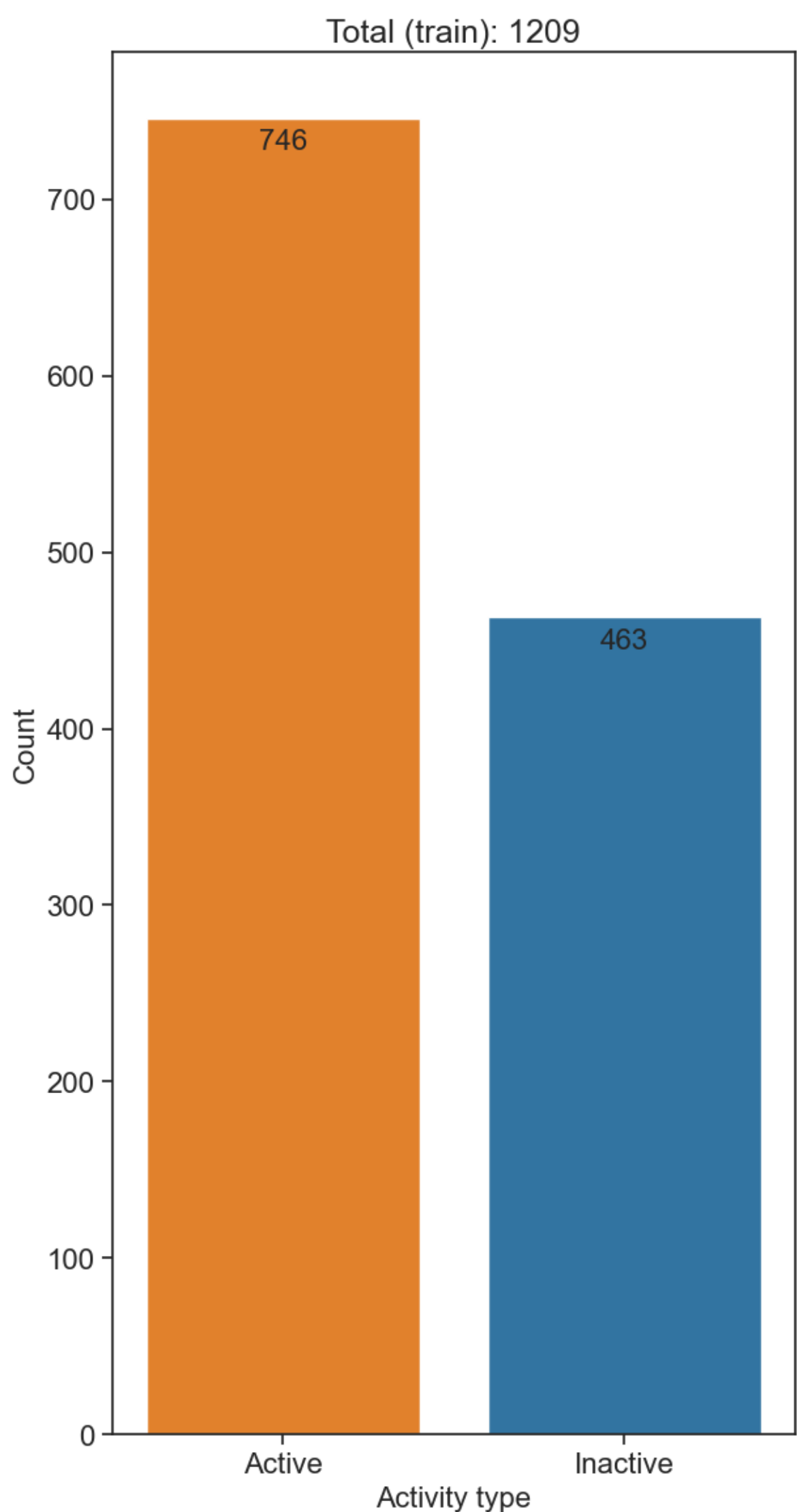
Metric	Train	Test
AUC	0.9997	0.9186
accuracy	0.9876	0.8548
balanced accuracy	0.9895	0.8447
recall_(sens)	0.9978	0.8017
specificity	0.9812	0.8877
precision	0.9706	0.8158
f1_score	0.9876	0.8545