Protein: P07550 - metric: balanced_accuracy

Total (train): 1279

ROC curve

- AUC_train = 0.998
- AUC_test = 0.962
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
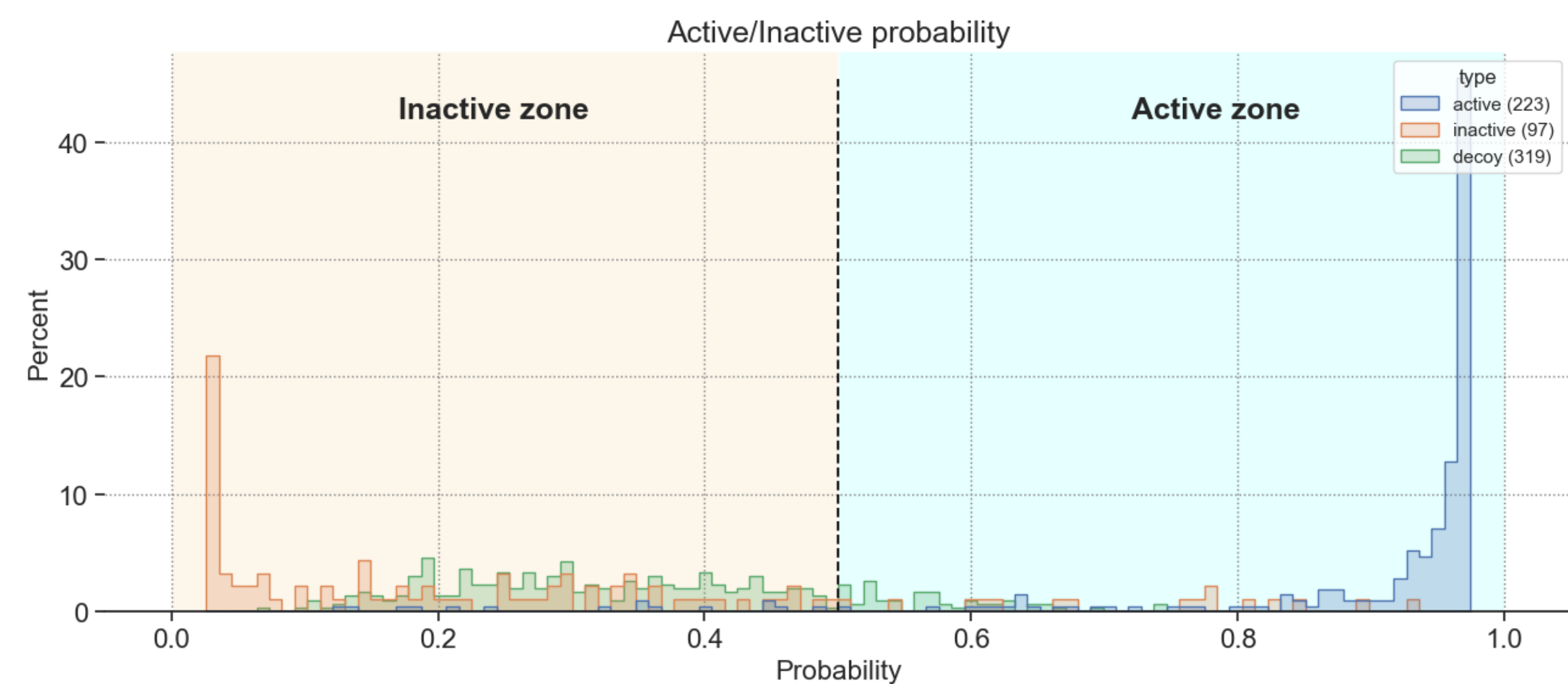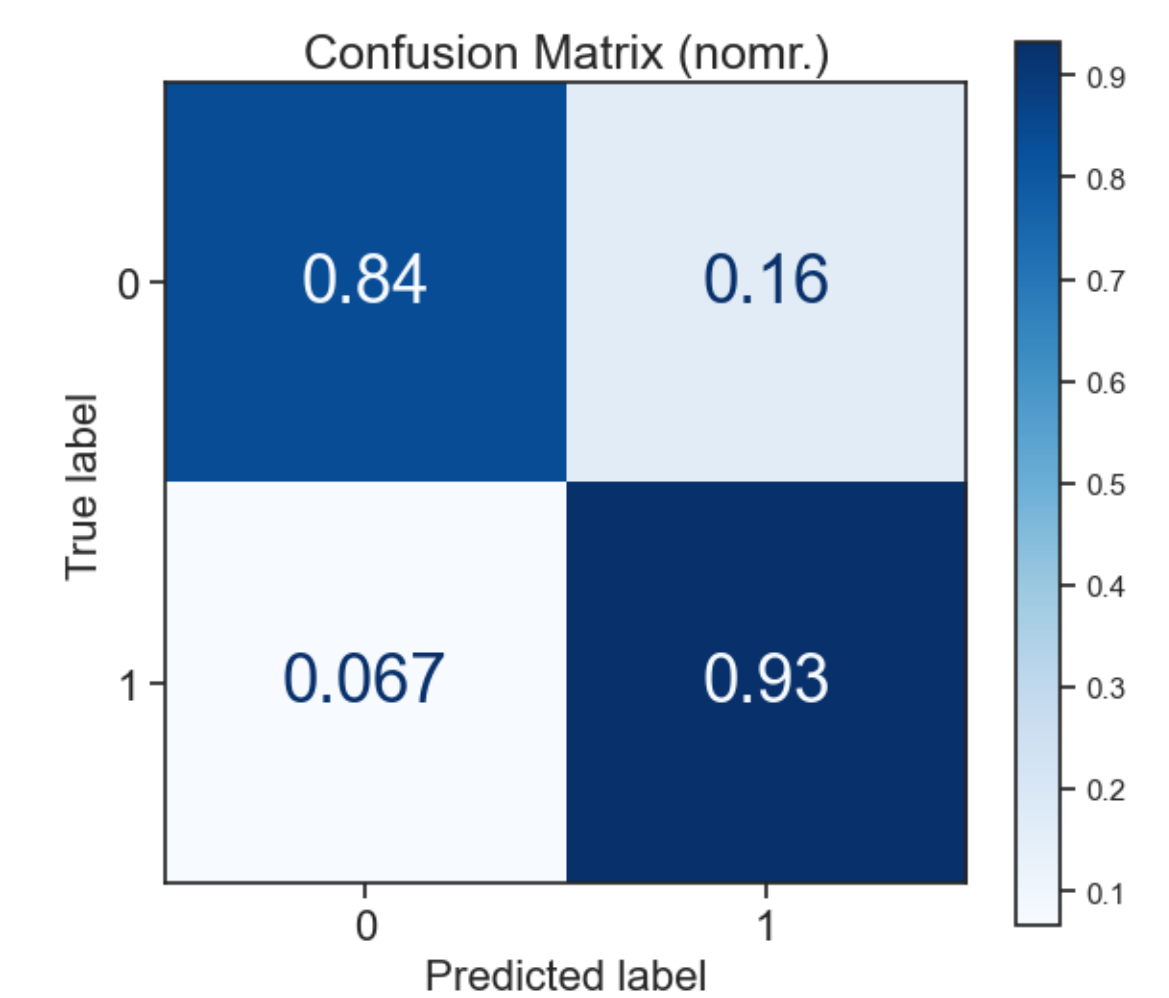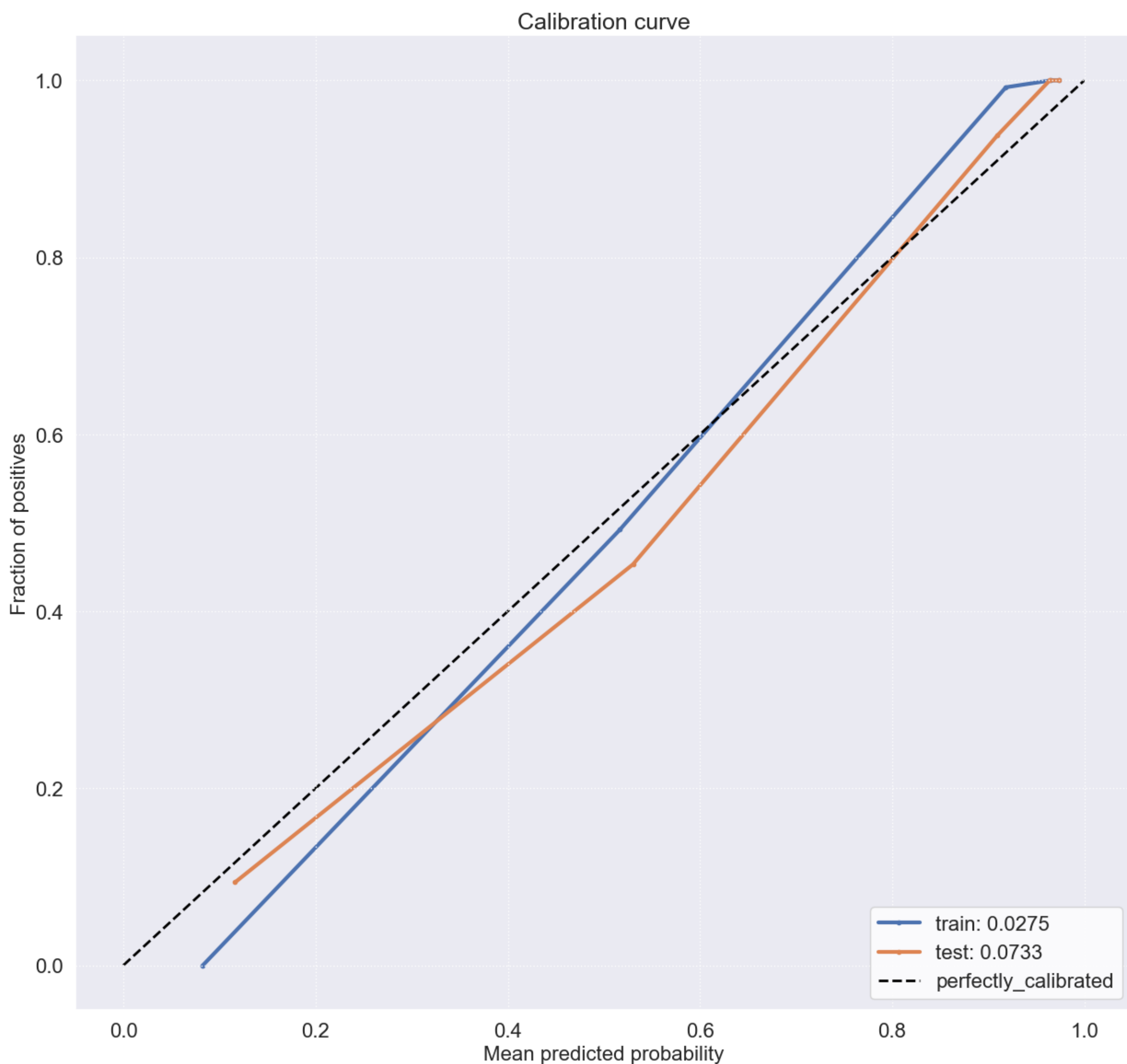- Pareto Front Trials
- Best Trial (# 2363)
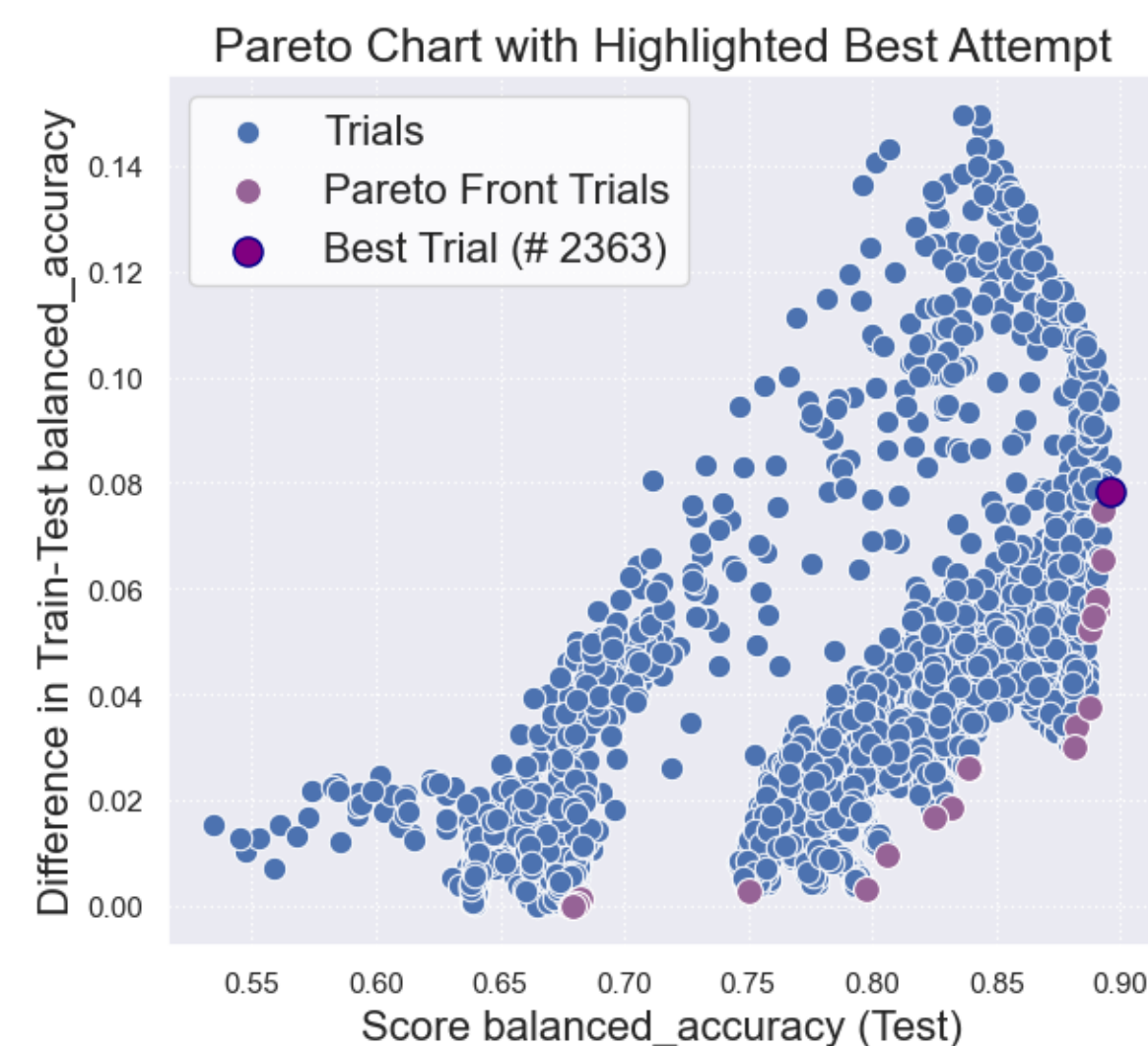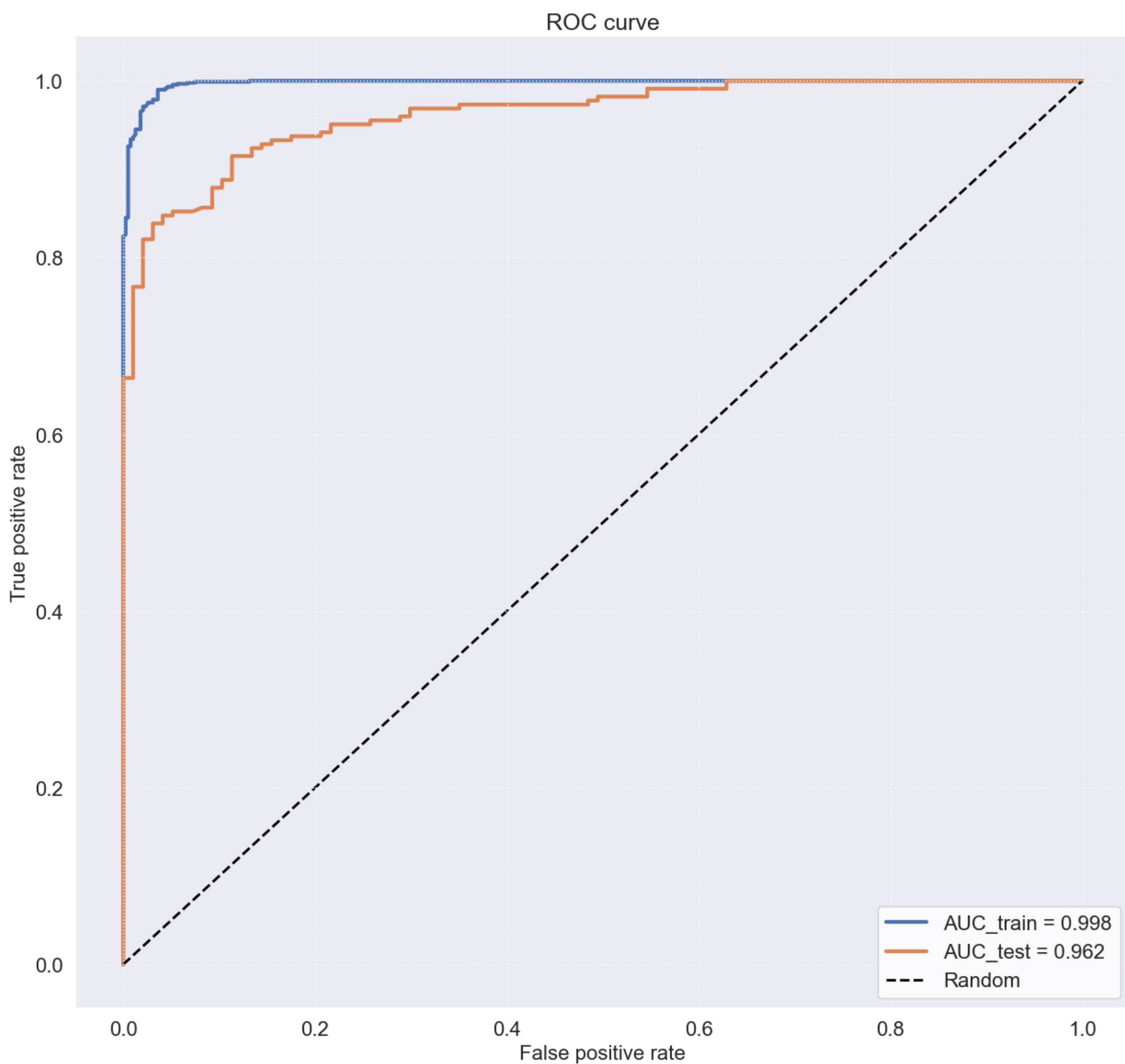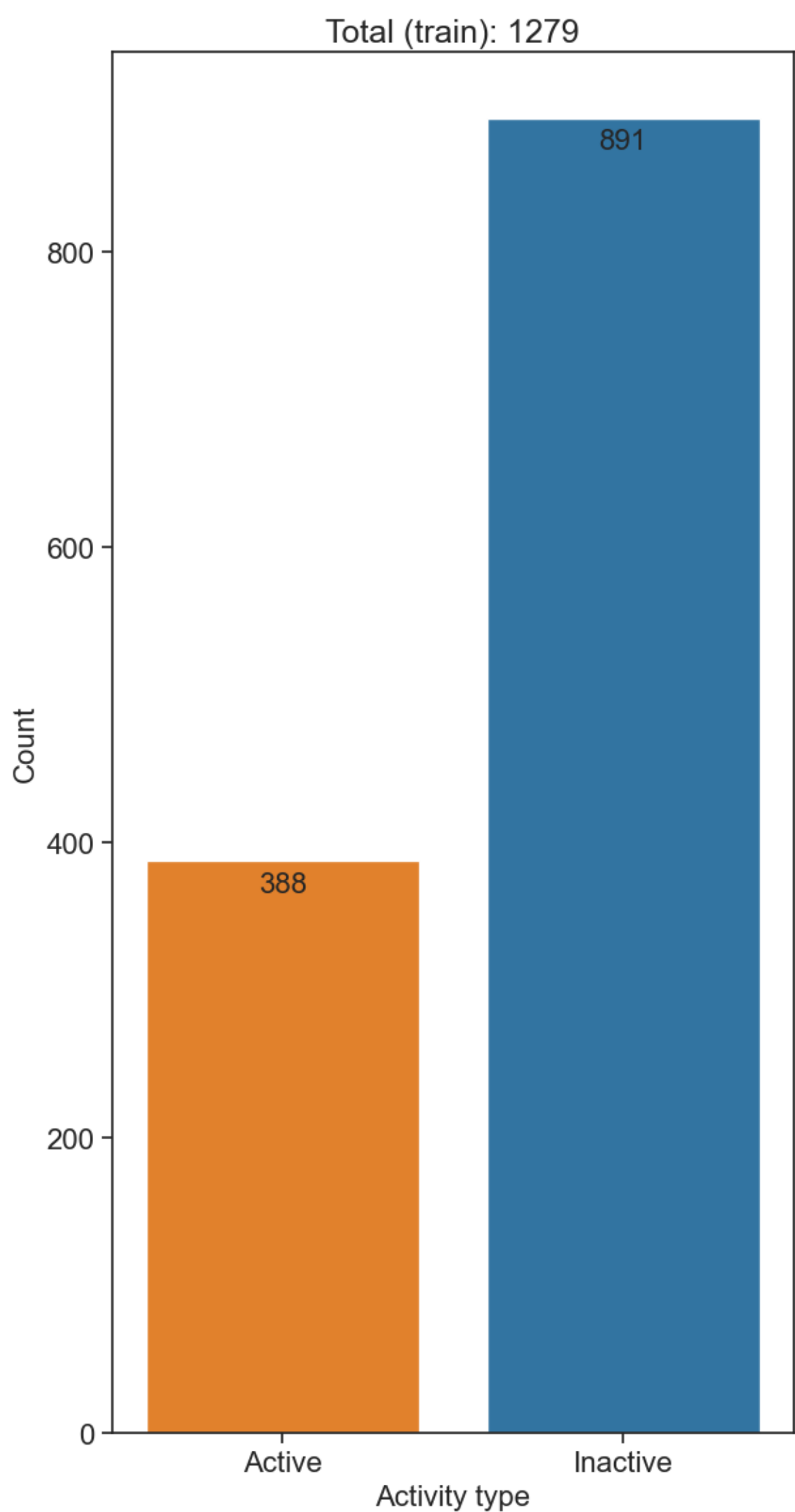
XGBoost Parameters	Value
alpha	3.2357
colsample_bytree	0.6117
gamma	0.2456
lambda	1.4022
learning_rate	0.0078
max_depth	10.0
min_child_weight	1.0
n_estimators	377.0
reg_alpha	0.0429
reg_lambda	0.0327
scale_pos_weight	1.1287
subsample	0.9513

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

- train: 0.0275
- test: 0.0733
- perfectly_calibrated

Active/Inactive probability

Inactive zone / Active zone

- active (223)
- inactive (97)
- decoy (319)
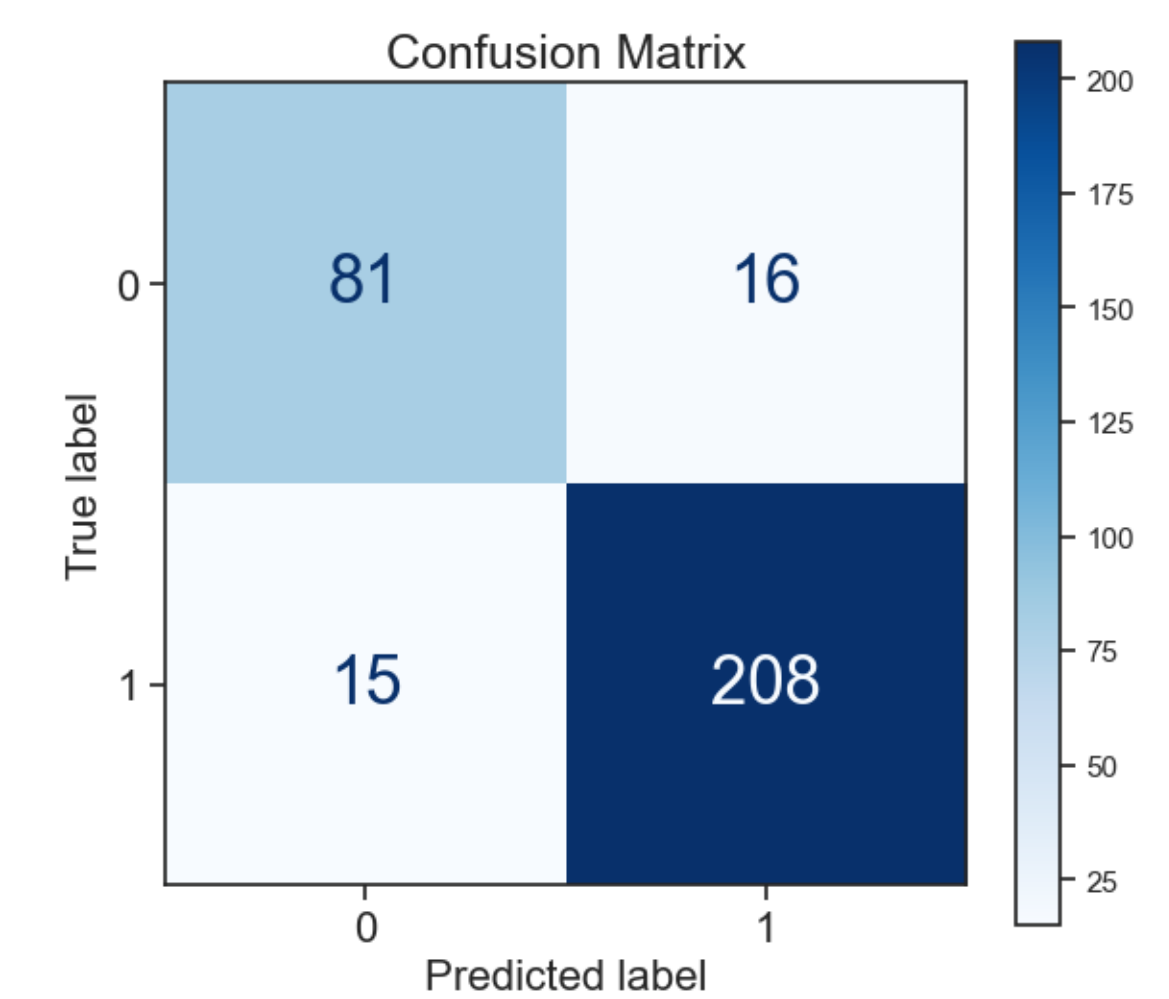
Metric	Train	Test
AUC	0.9976	0.9624
accuracy	0.9805	0.9031
balanced accuracy	0.9707	0.8839
recall_(sens)	0.9955	0.9327
specificity	0.9459	0.8351
precision	0.9769	0.9286
f1_score	0.9803	0.903

Protein: P07550 - metric: f1

Total (train): 1279

Active: 388
Inactive: 891

ROC curve

AUC_train = 0.998
AUC_test = 0.961
Random

Pareto Chart with Highlighted Best Attempt

Trials
Pareto Front Trials
Best Trial (# 1841)
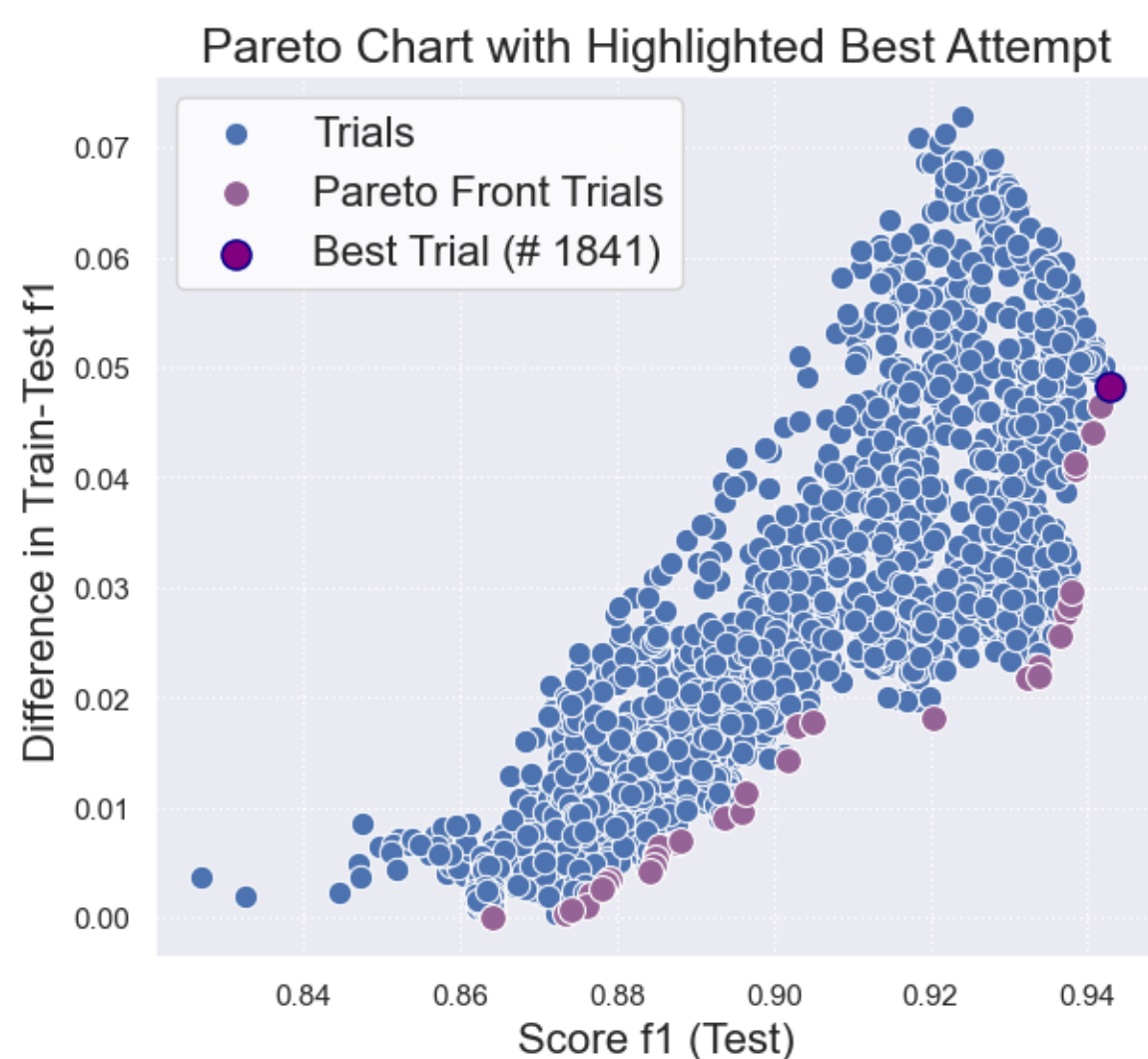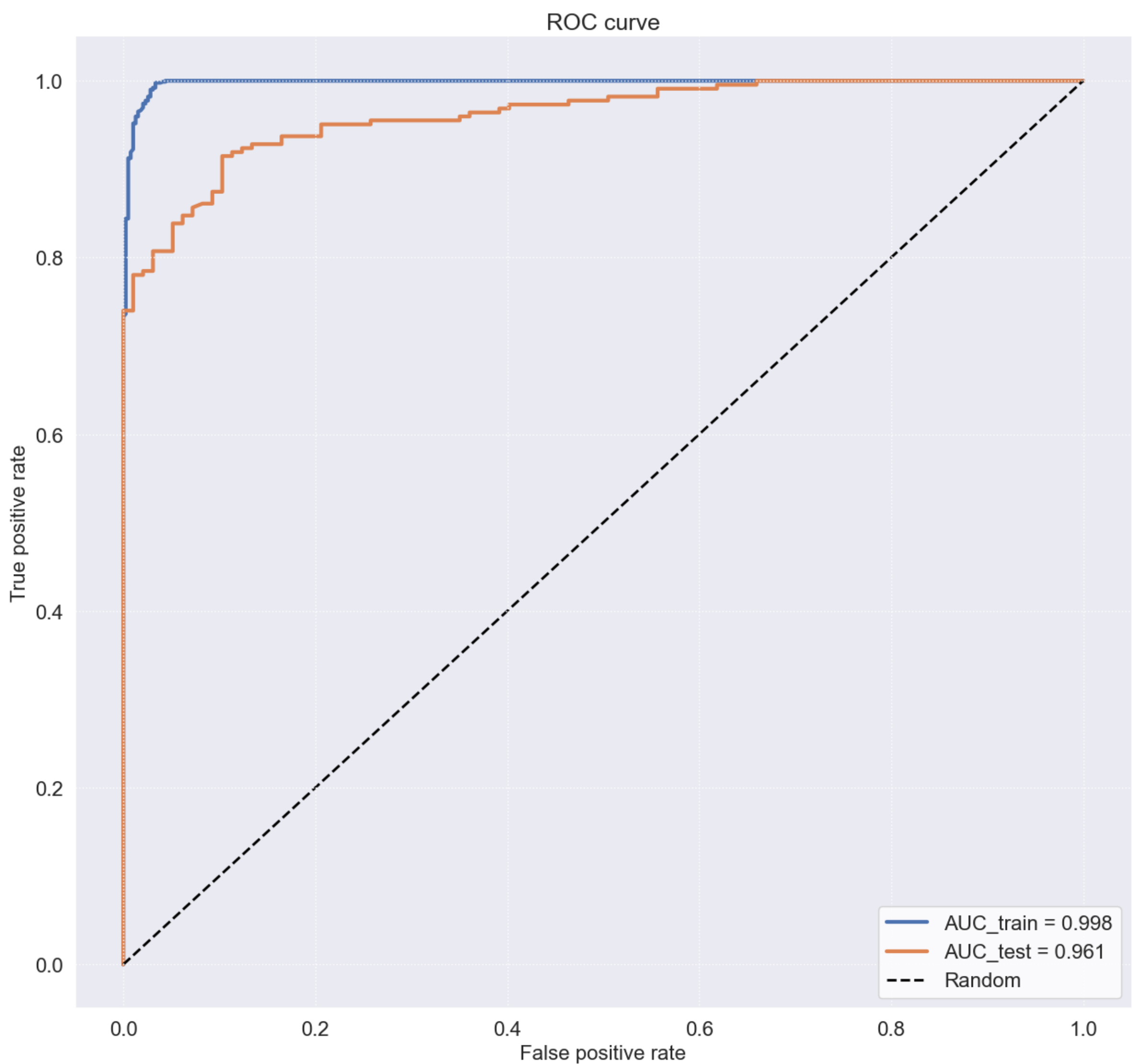
XGBoost Parameters	Value
alpha	1.6698
colsample_bytree	0.7879
gamma	0.8943
lambda	2.9037
learning_rate	0.0313
max_depth	10.0
min_child_weight	1.0
n_estimators	479.0
reg_alpha	0.0196
reg_lambda	0.1294
scale_pos_weight	1.3308
subsample	0.9342

Confusion Matrix

	Predicted 0	Predicted 1
True 0	78	19
True 1	14	209

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.8	0.2
True 1	0.063	0.94

Calibration curve

train: 0.0171
test: 0.0754
perfectly_calibrated

Active/Inactive probability

Inactive zone
Active zone

type
active (223)
inactive (97)
decoy (319)
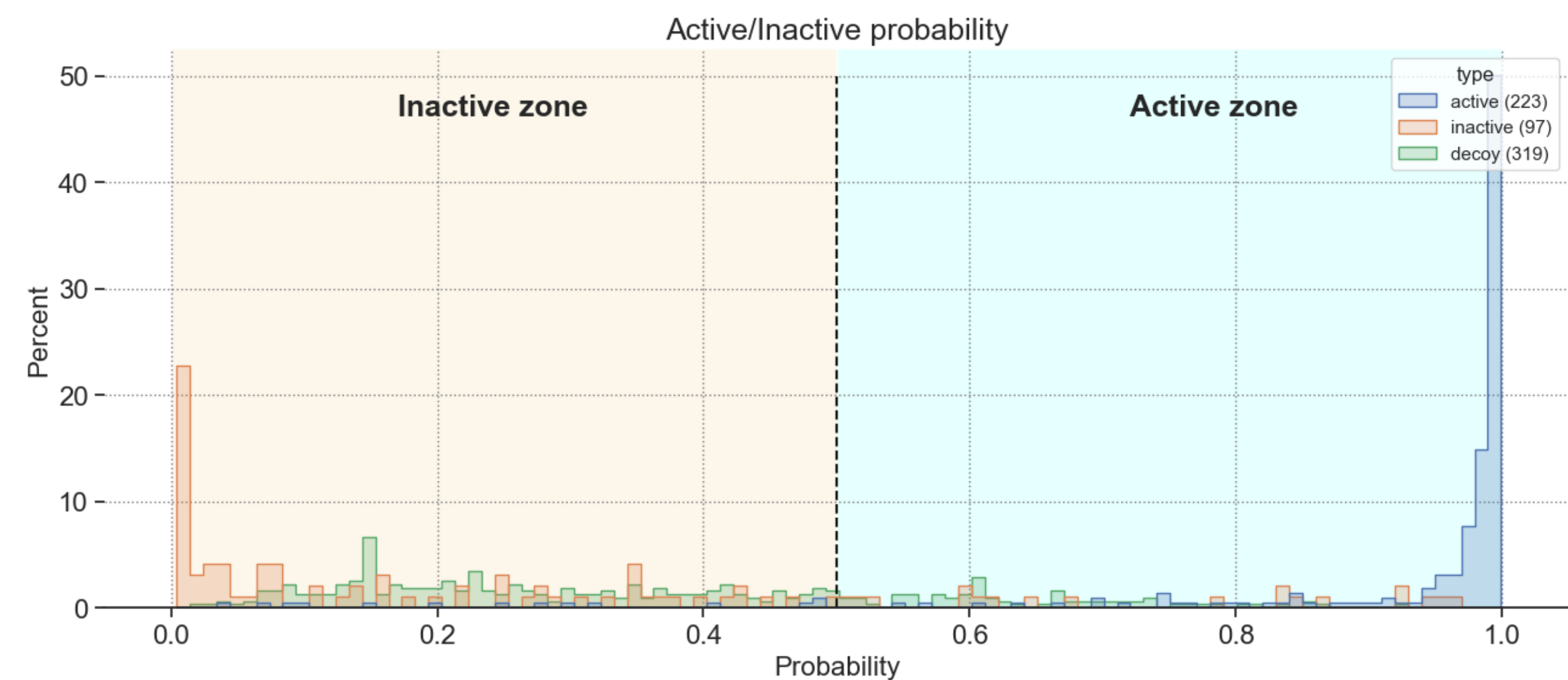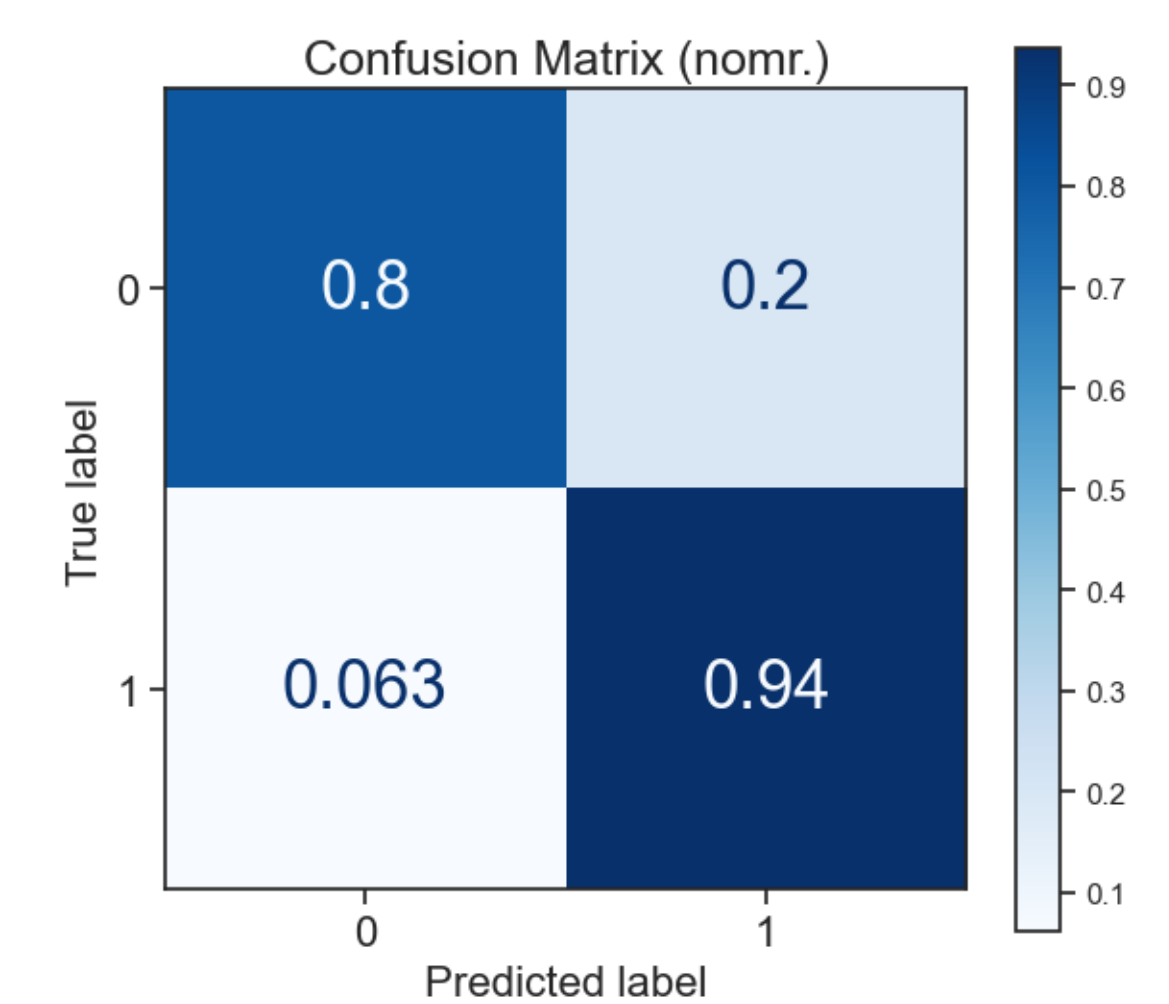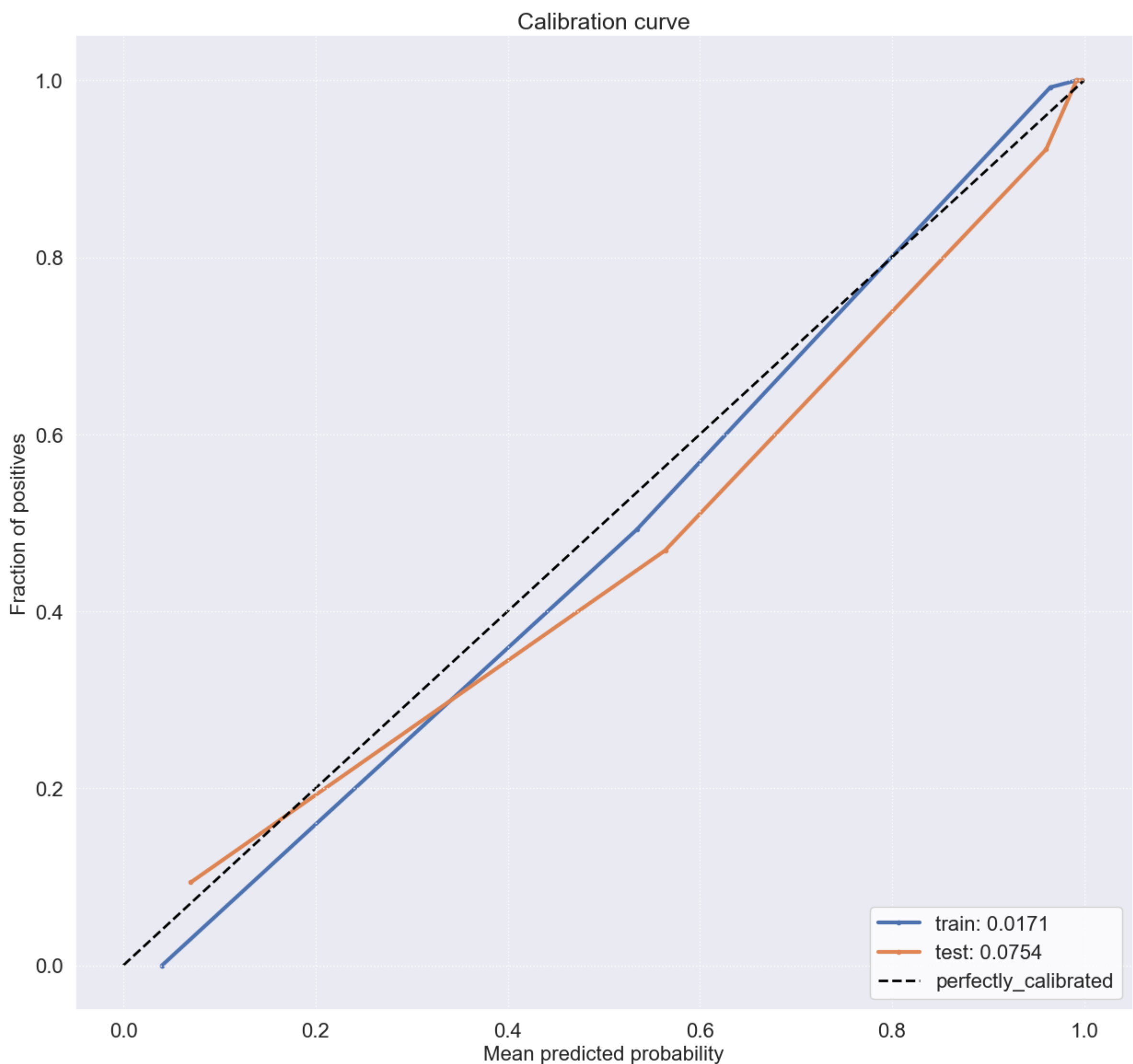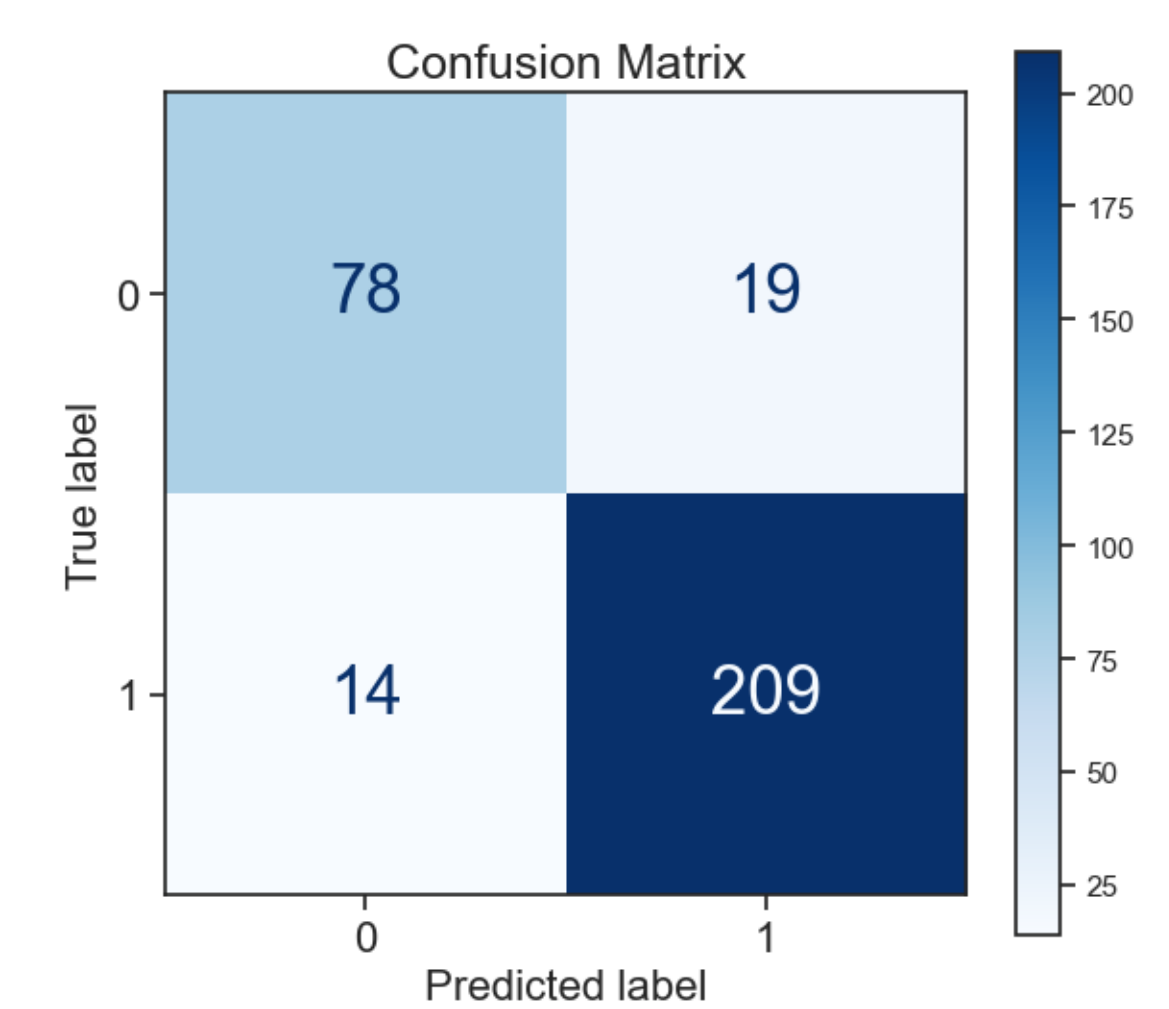
Metric	Train	Test
AUC	0.9978	0.9613
accuracy	0.9875	0.8969
balanced accuracy	0.9808	0.8707
recall_(sens)	0.9978	0.9372
specificity	0.9639	0.8041
precision	0.9845	0.9167
f1_score	0.9874	0.8961

Protein: P07550 - metric: precision

Total (train): 1279

ROC curve
- AUC_train = 0.999
- AUC_test = 0.964
- Random

Pareto Chart with Highlighted Best Attempt
- Trials
- Pareto Front Trials
- Best Trial (# 1413)
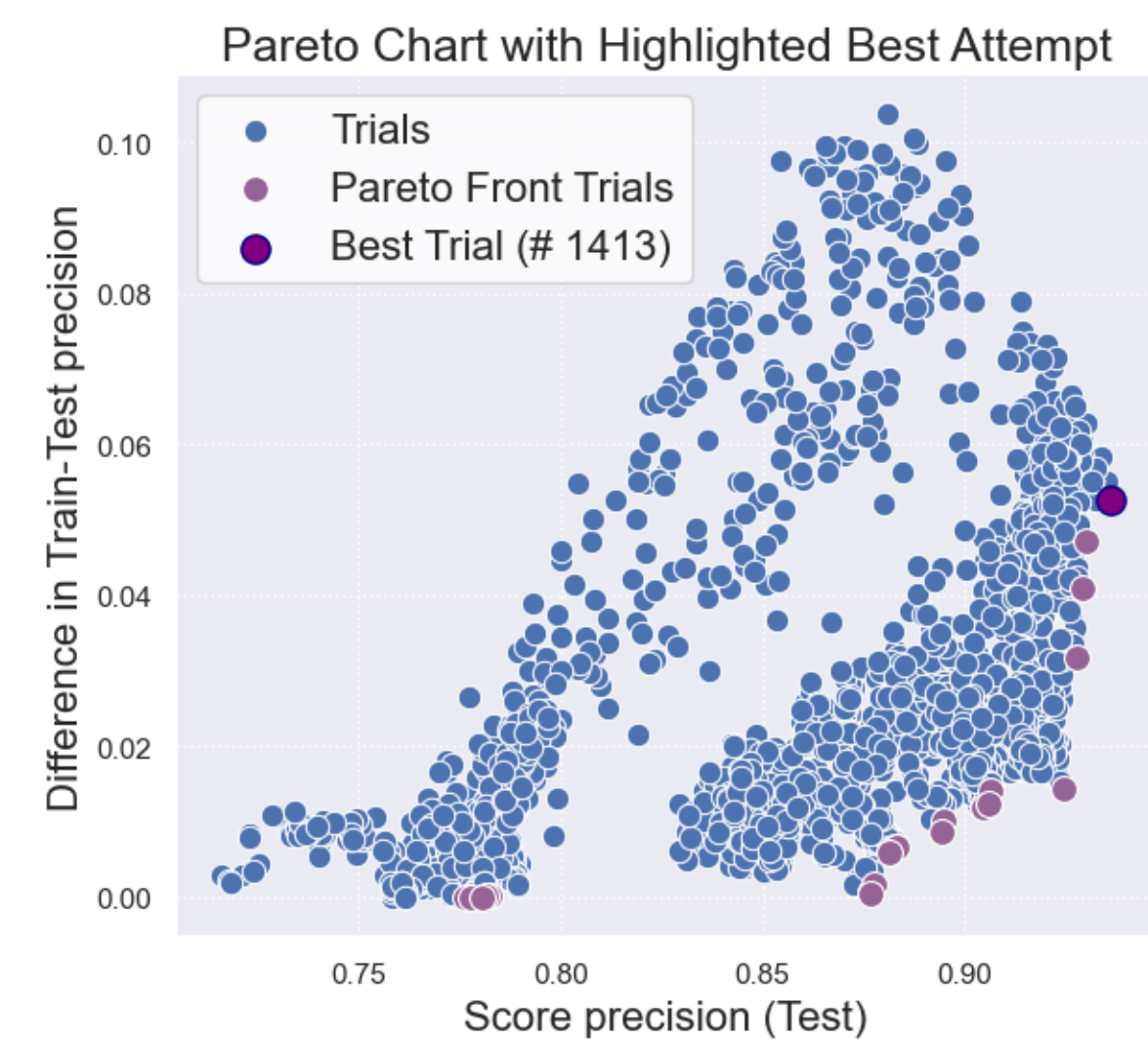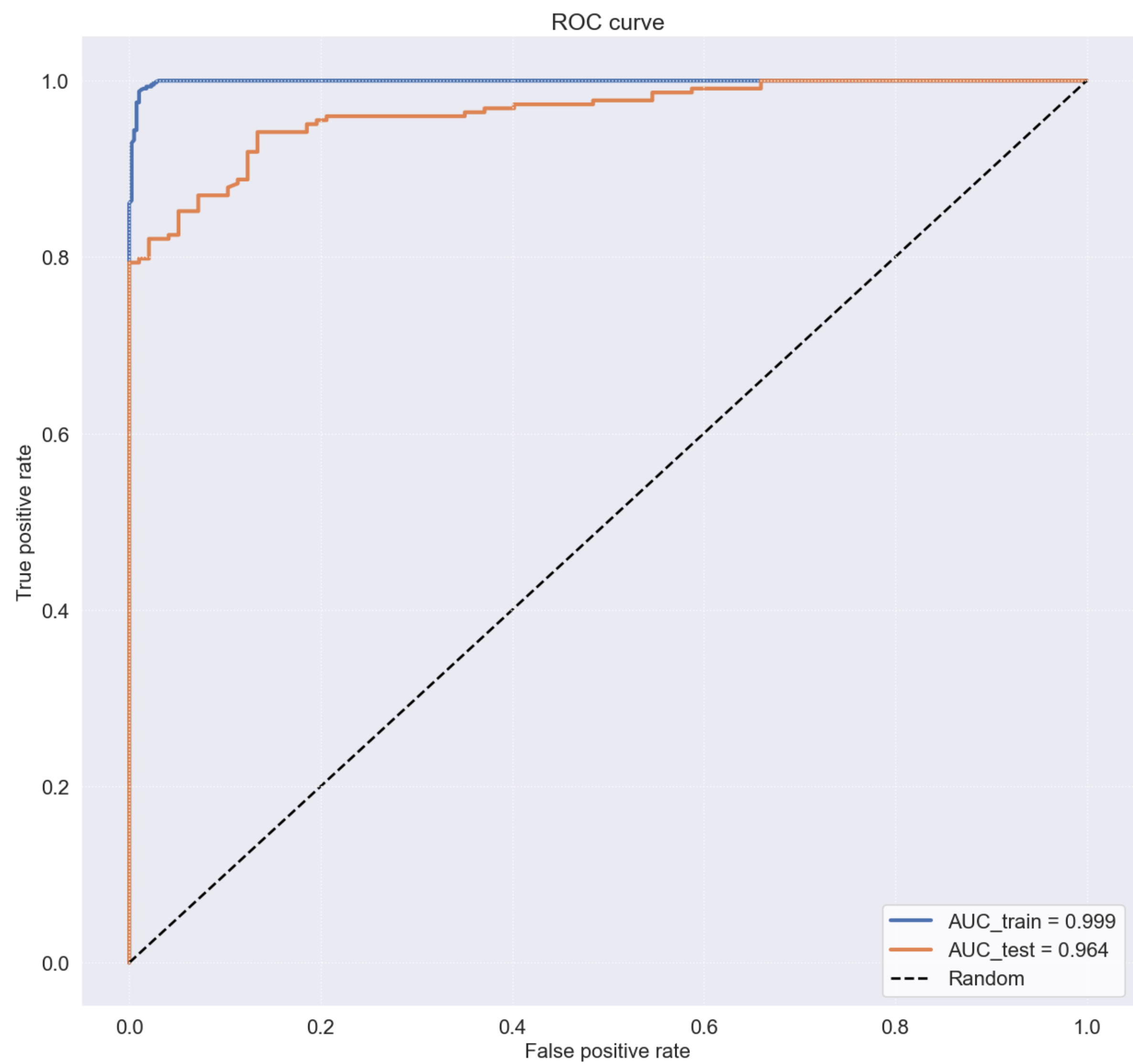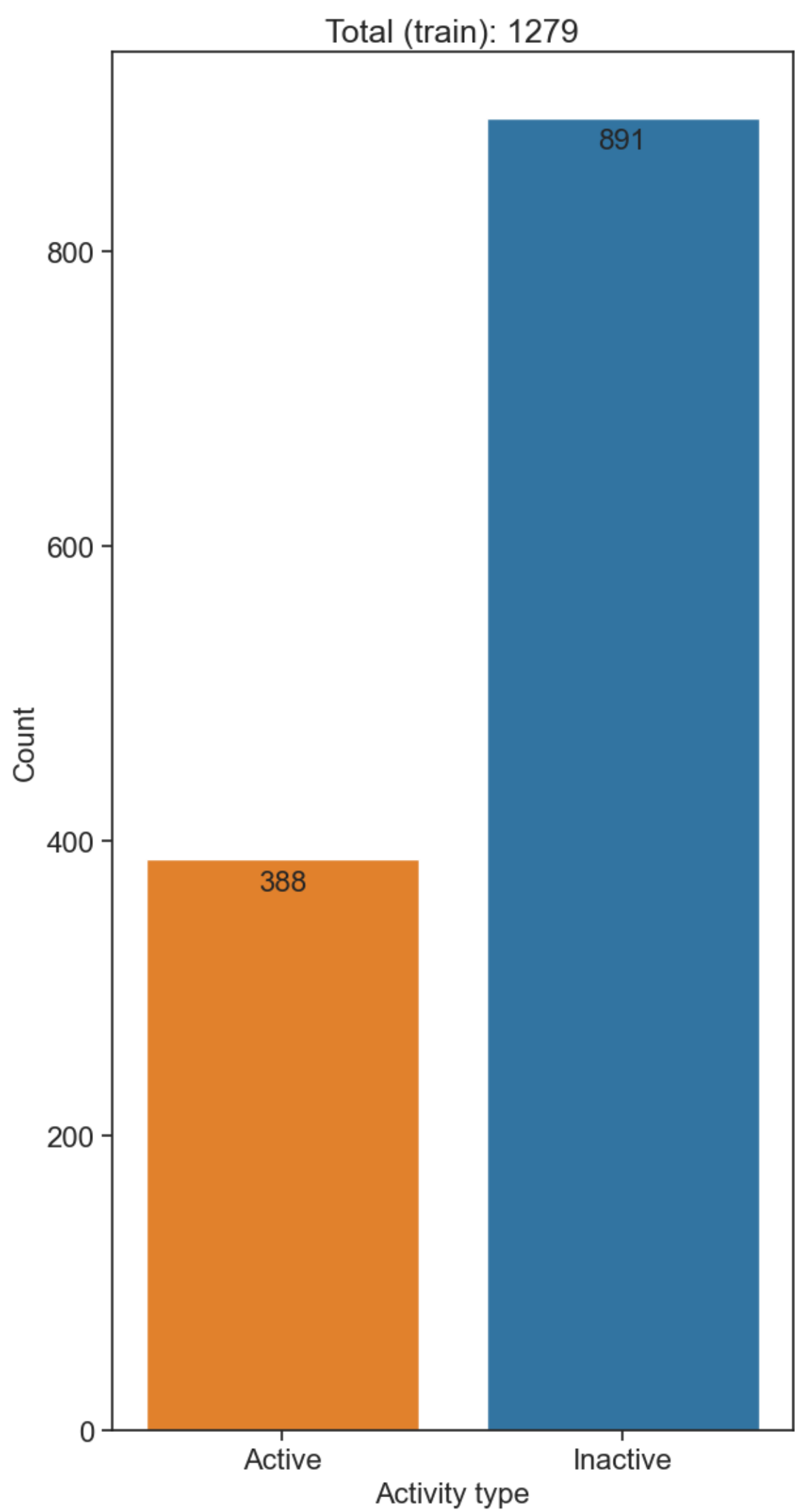
XGBoost Parameters	Value
alpha	0.4467
colsample_bytree	0.6565
gamma	0.4405
lambda	1.823
learning_rate	0.0616
max_depth	8.0
min_child_weight	1.0
n_estimators	624.0
reg_alpha	0.0689
reg_lambda	0.1964
scale_pos_weight	1.1275
subsample	0.9087

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve
- train: 0.0112
- test: 0.0722
- perfectly_calibrated

Active/Inactive probability
- type: active (223), inactive (97), decoy (319)
- Inactive zone / Active zone
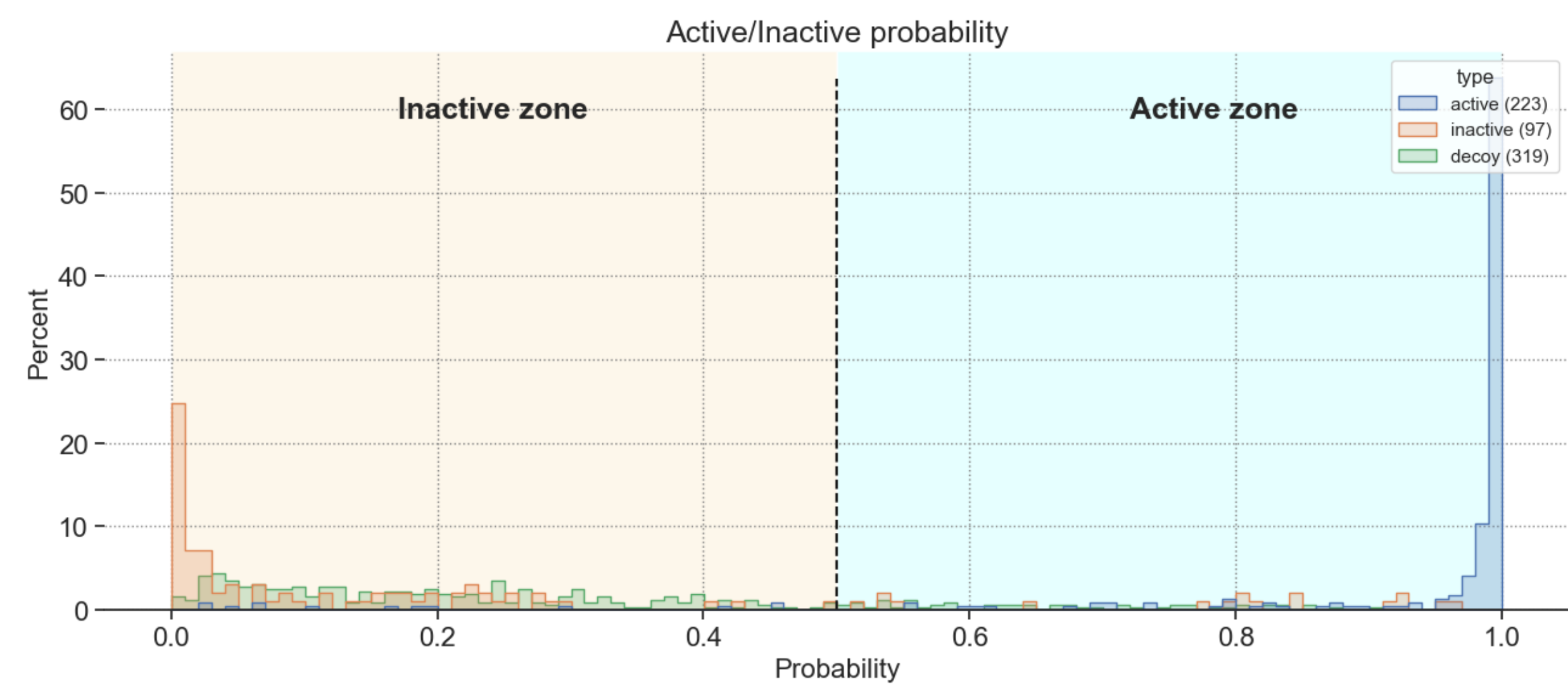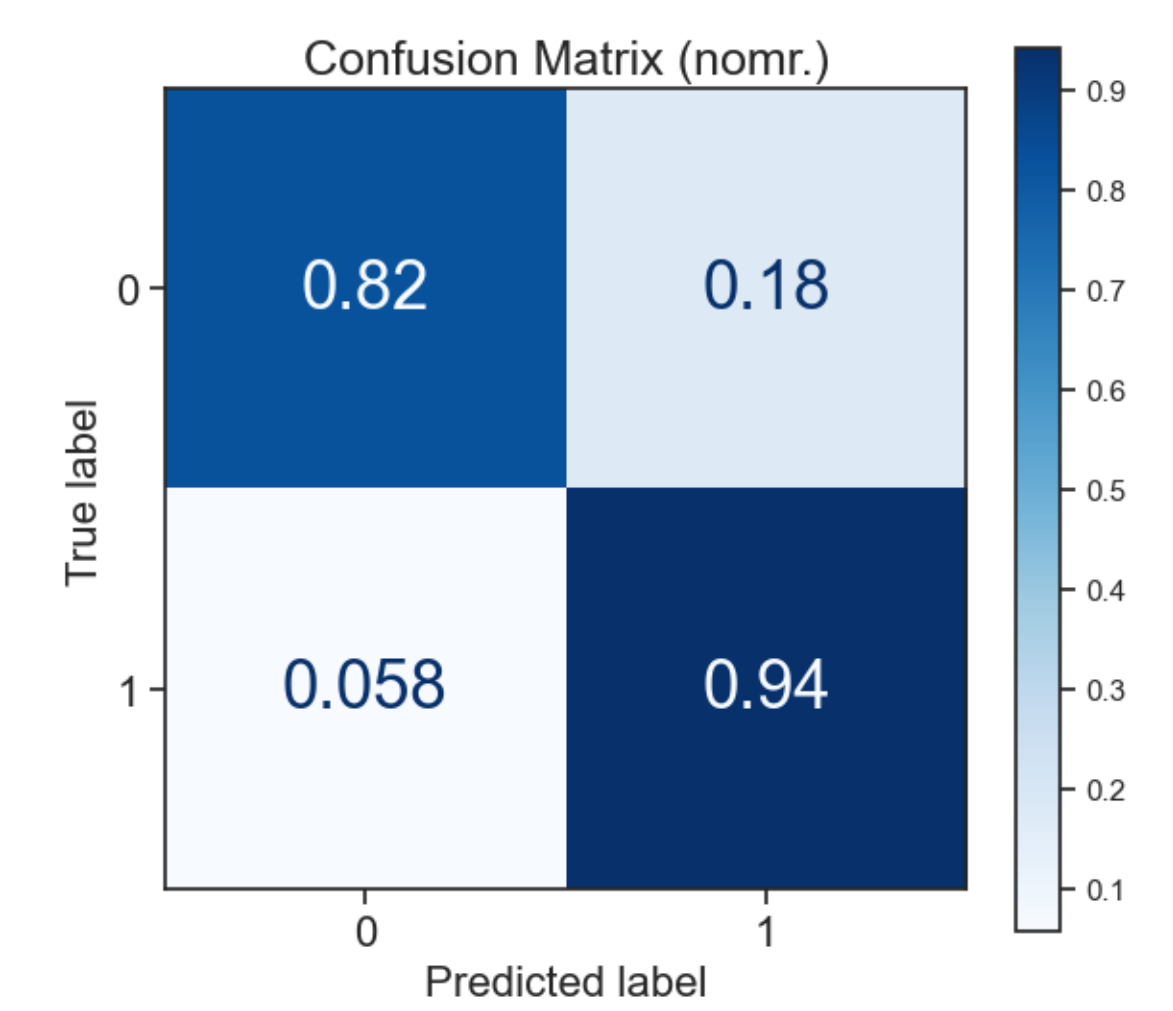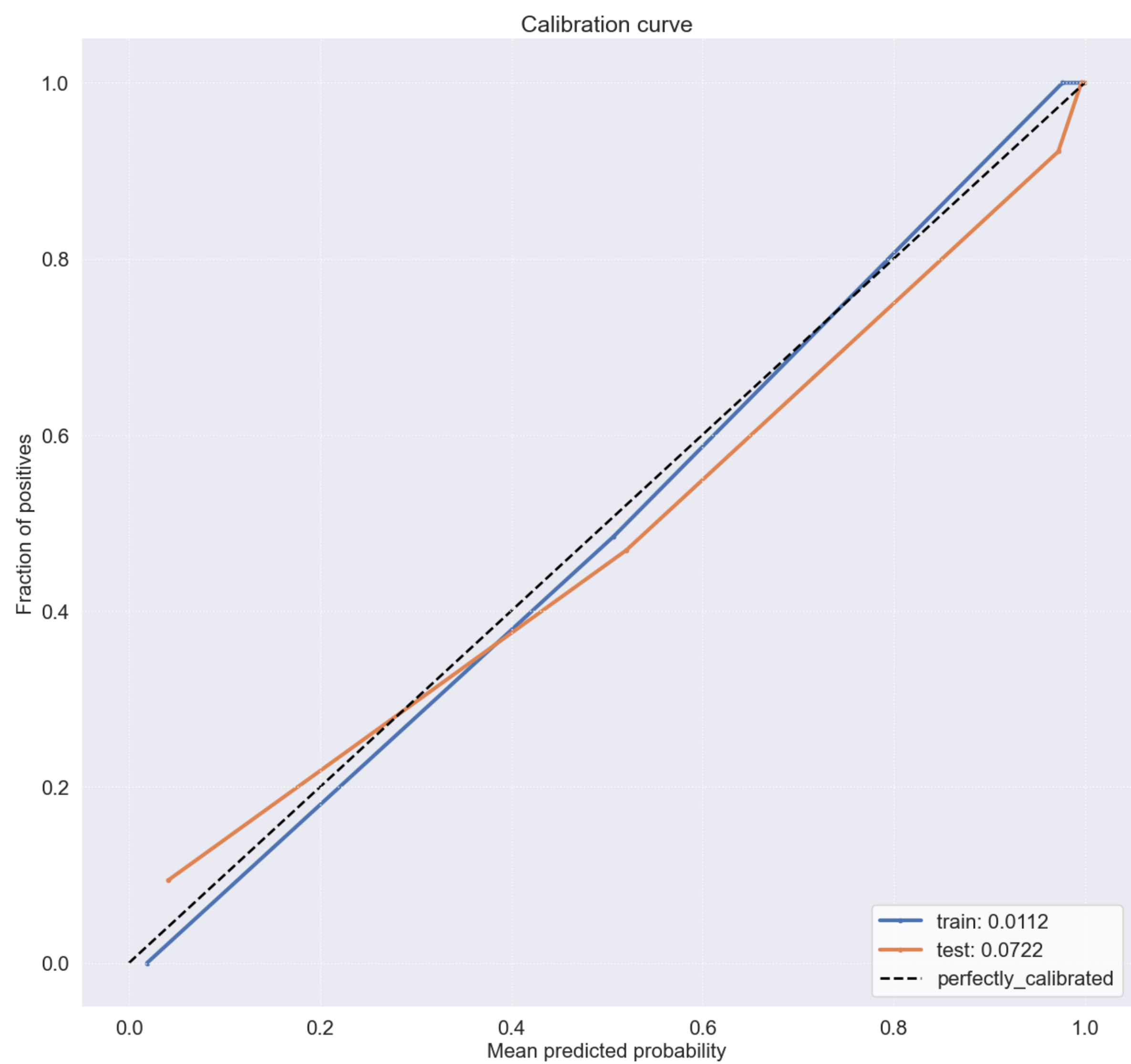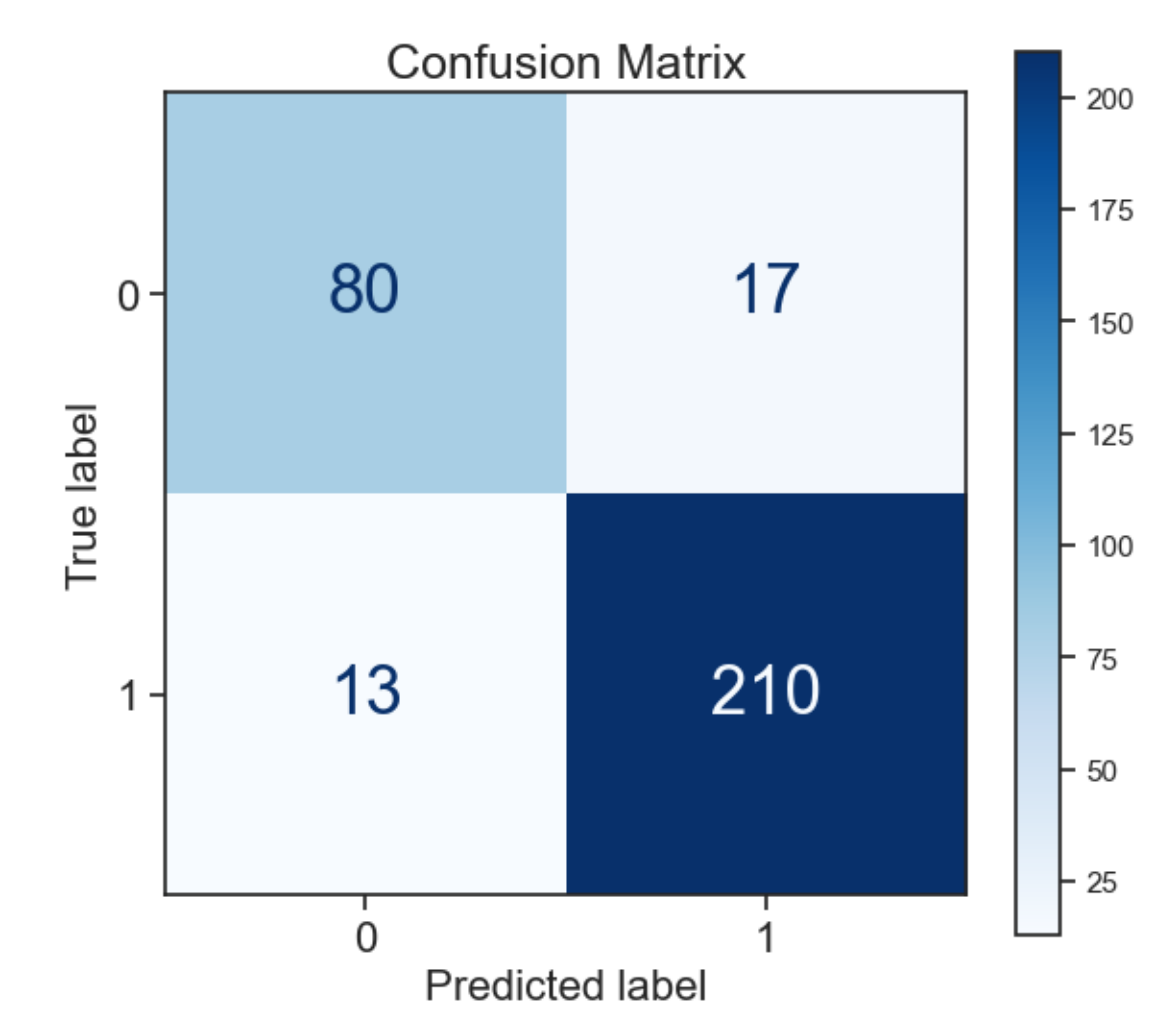
Metric	Train	Test
AUC	0.9991	0.9637
accuracy	0.9898	0.9062
balanced accuracy	0.9847	0.8832
recall_(sens)	0.9978	0.9417
specificity	0.9716	0.8247
precision	0.9878	0.9251
f1_score	0.9898	0.9057

Protein: P07550 - metric: SoftVotingXGBoostEnsemble
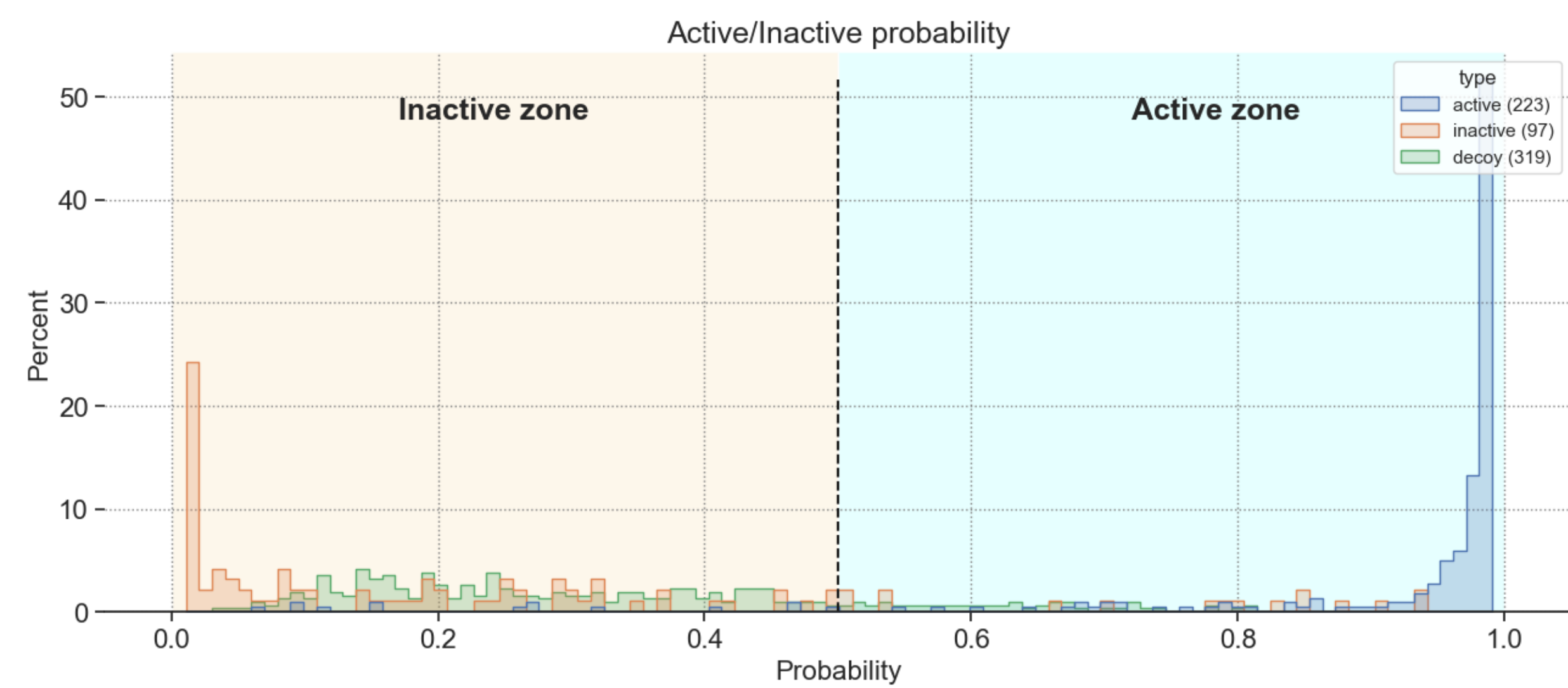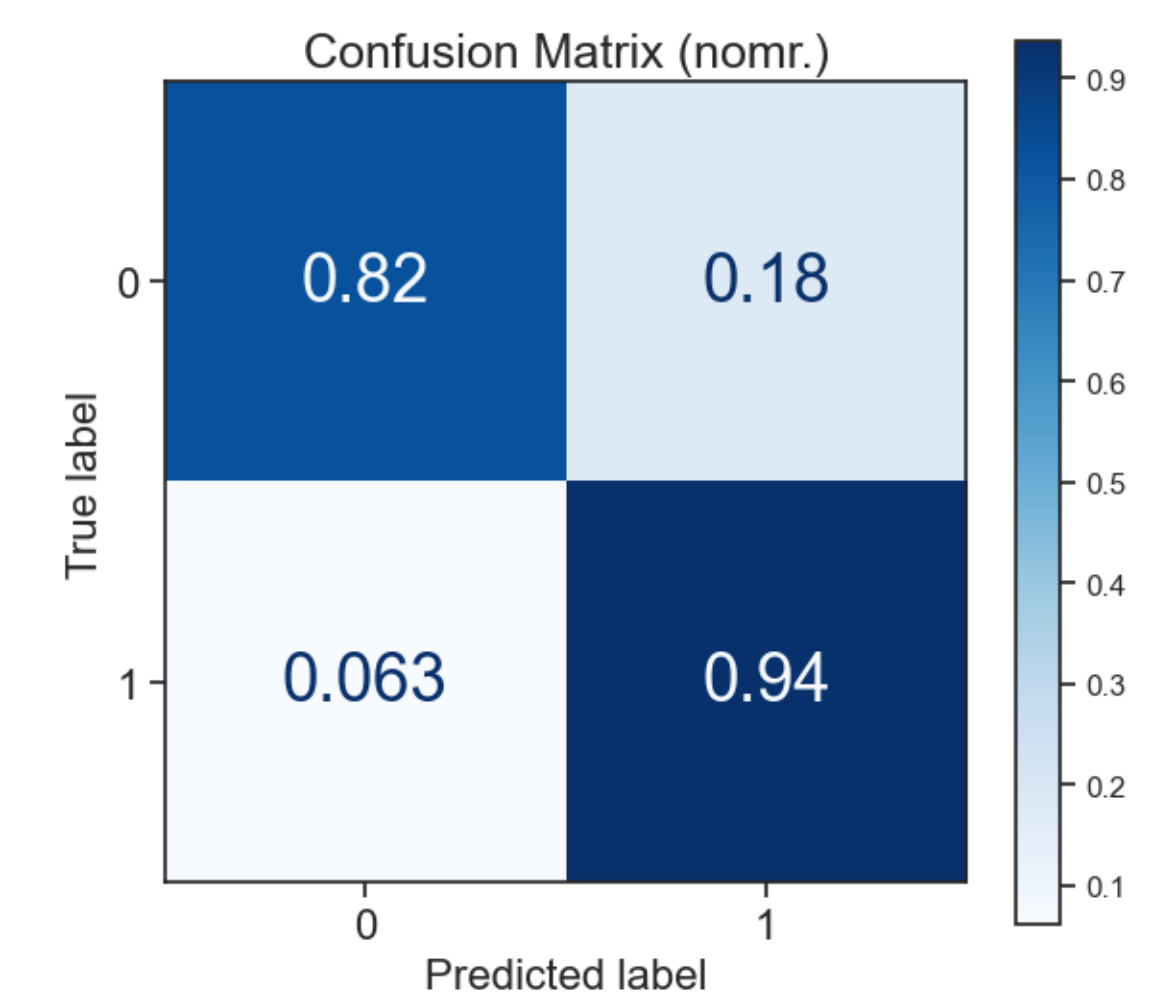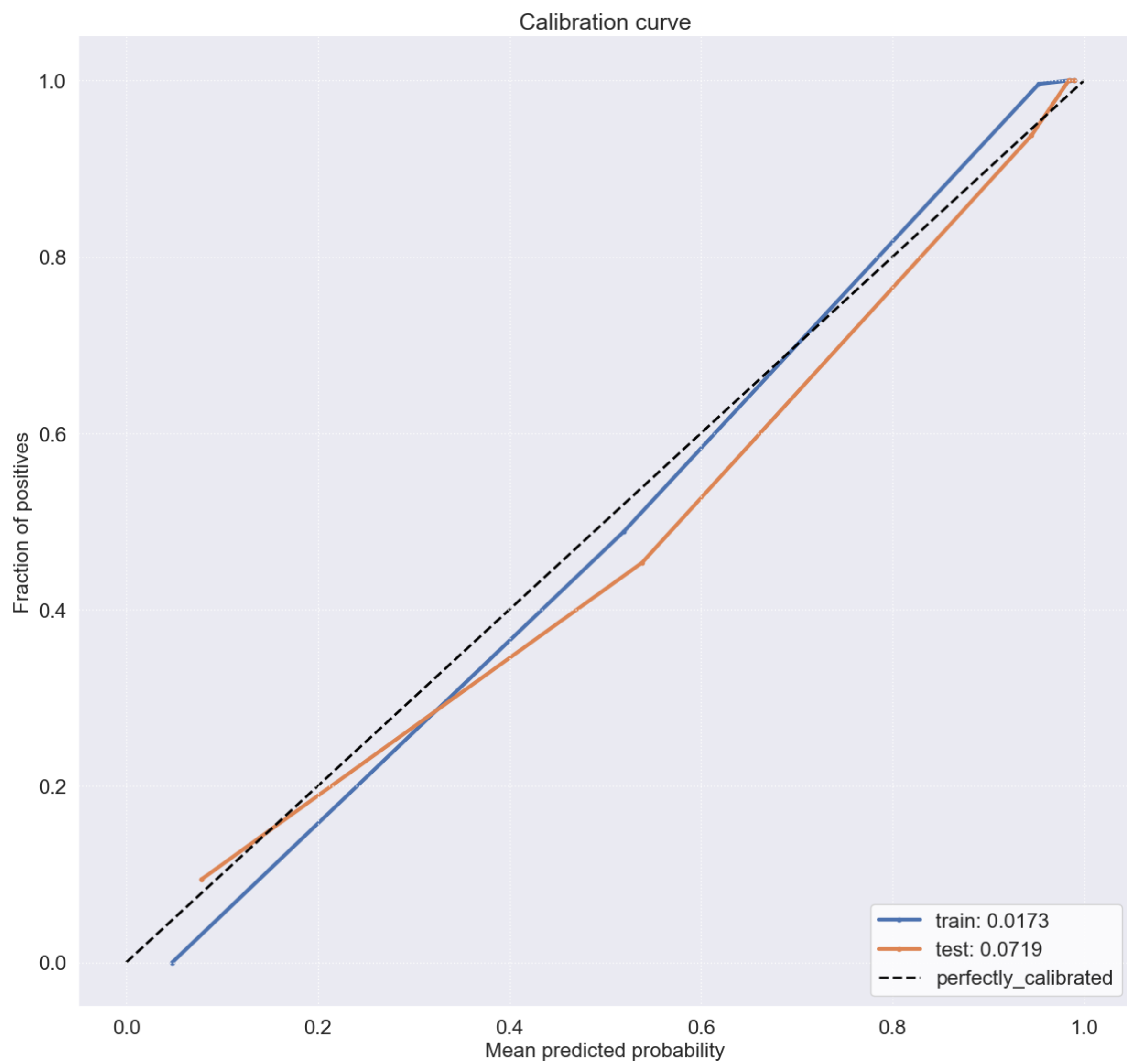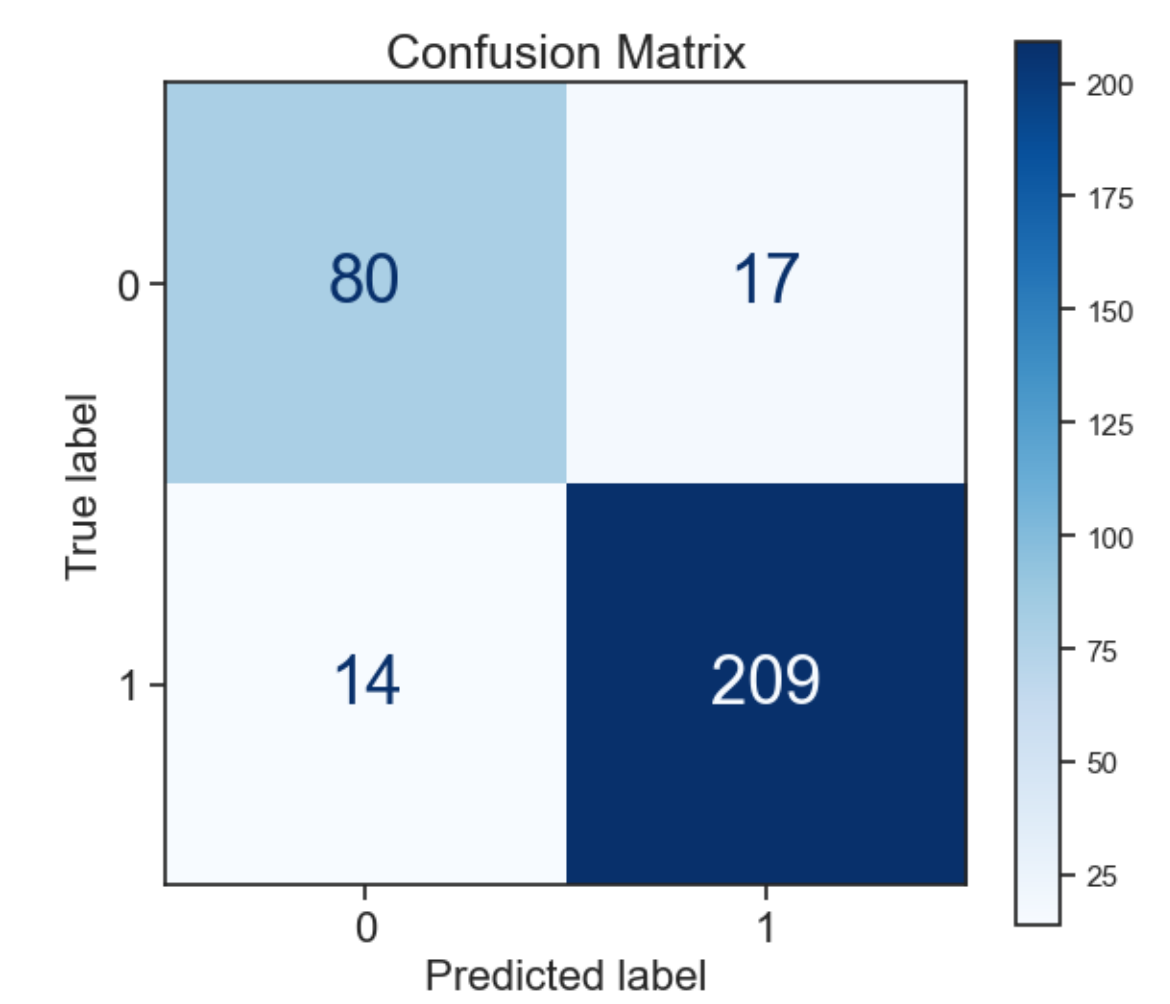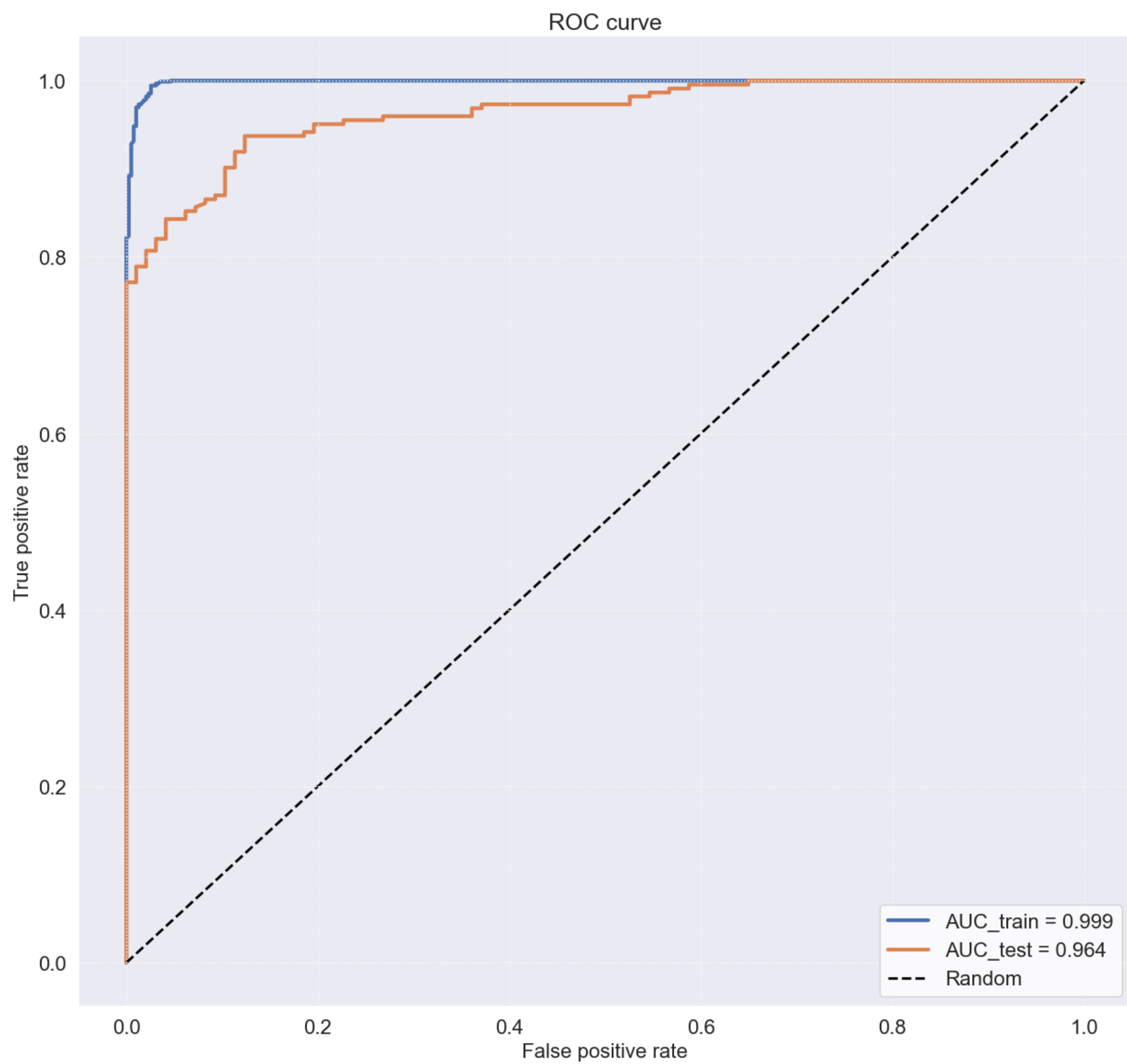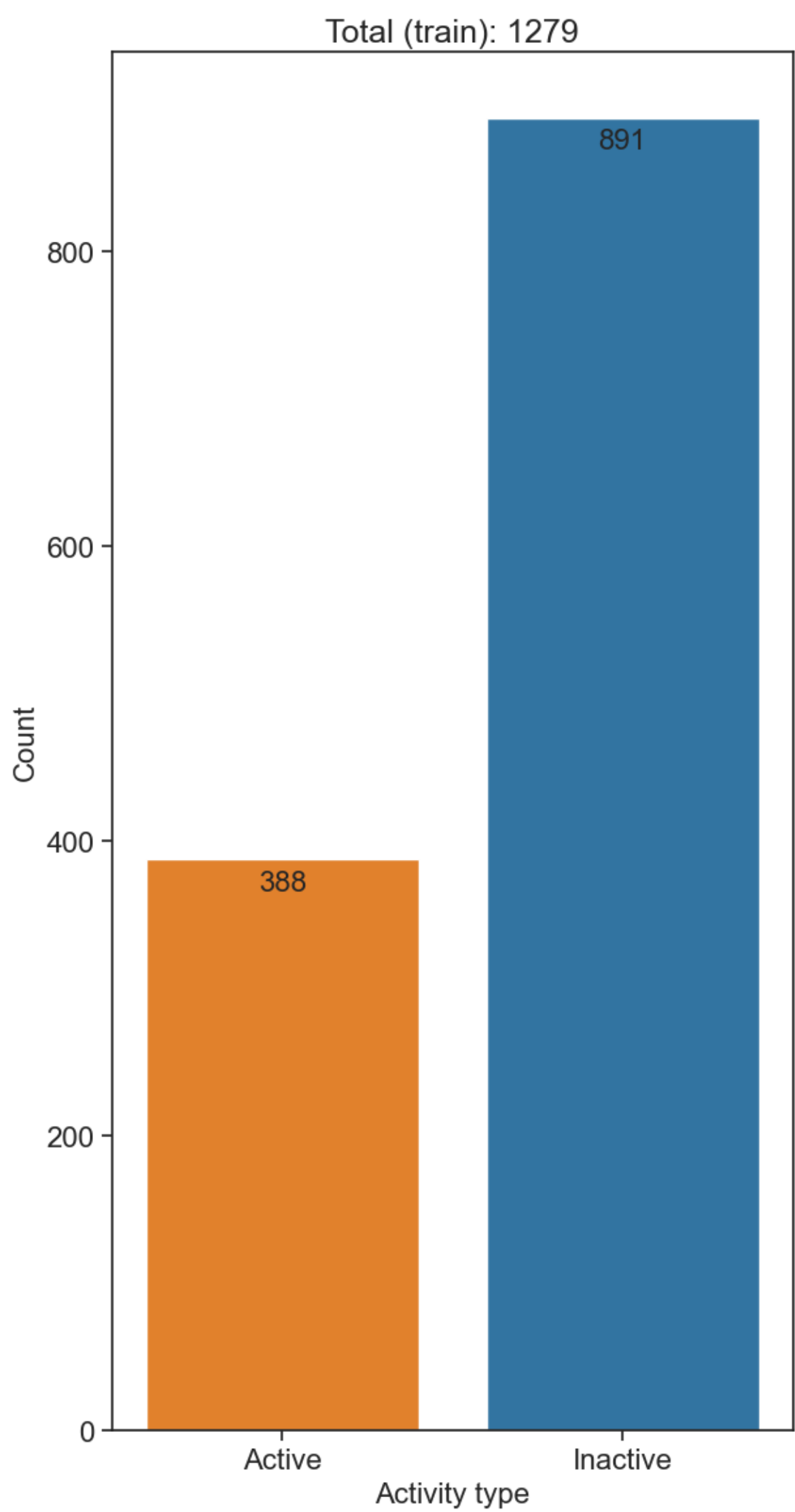
Metric	Train	Test
AUC	0.9985	0.9636
accuracy	0.9883	0.9031
balanced accuracy	0.9821	0.881
recall_(sens)	0.9978	0.9372
specificity	0.9665	0.8247
precision	0.9856	0.9248
f1_score	0.9882	0.9027