Protein: O76074 - metric: balanced_accuracy

Total (train): 1276

Active: 422, Inactive: 854

ROC curve

- AUC_train = 0.997
- AUC_test = 0.965
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2755)

Confusion Matrix

	Predicted 0	Predicted 1
True 0	90	16
True 1	13	201

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.85	0.15
True 1	0.061	0.94

Calibration curve

- train: 0.0206
- test: 0.0679
- perfectly_calibrated

Active/Inactive probability

Inactive zone / Active zone

type:
- active (214)
- inactive (106)
- decoy (320)

XGBoost Parameters	Value
alpha	2.0617
colsample_bytree	0.5892
gamma	0.9741
lambda	0.1609
learning_rate	0.0729
max_depth	5.0
min_child_weight	3.0
n_estimators	523.0
reg_alpha	0.1237
reg_lambda	0.3237
scale_pos_weight	1.4201
subsample	0.5071
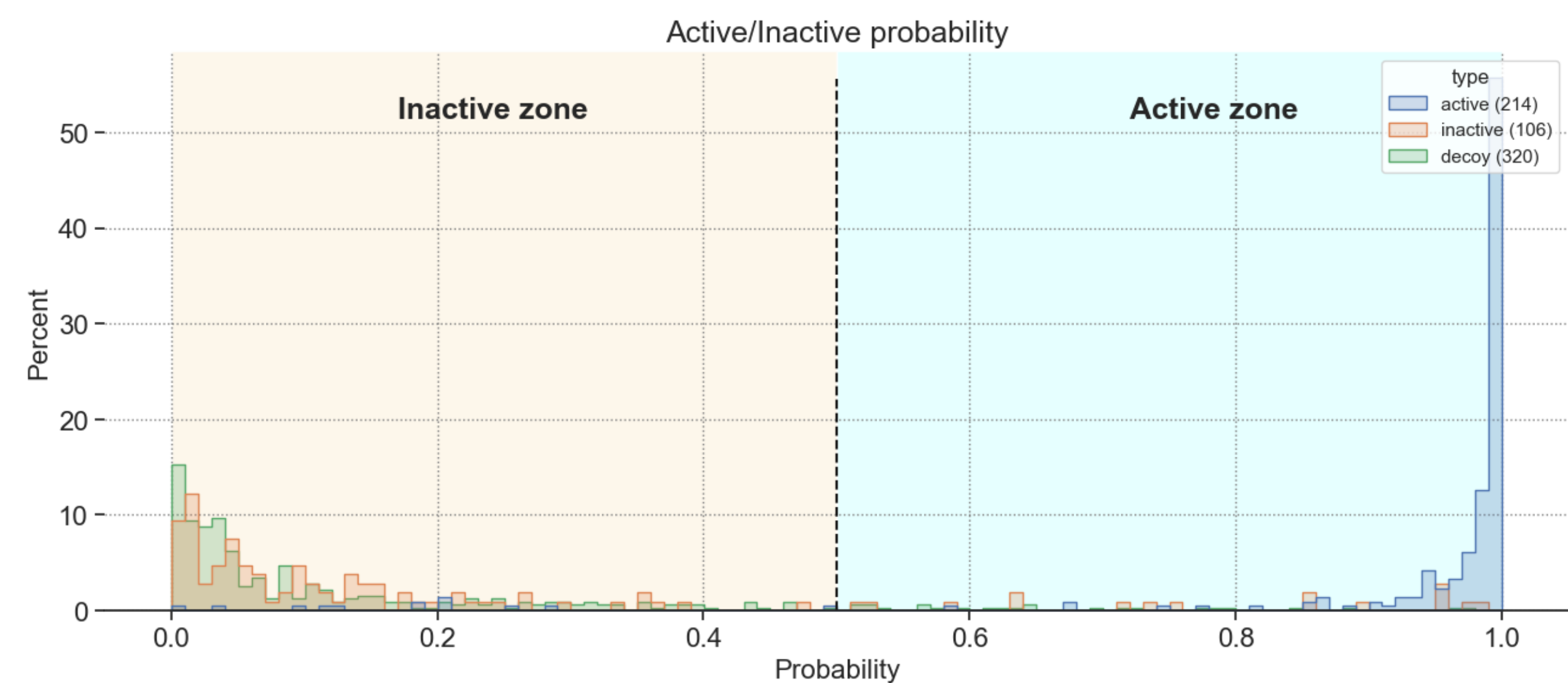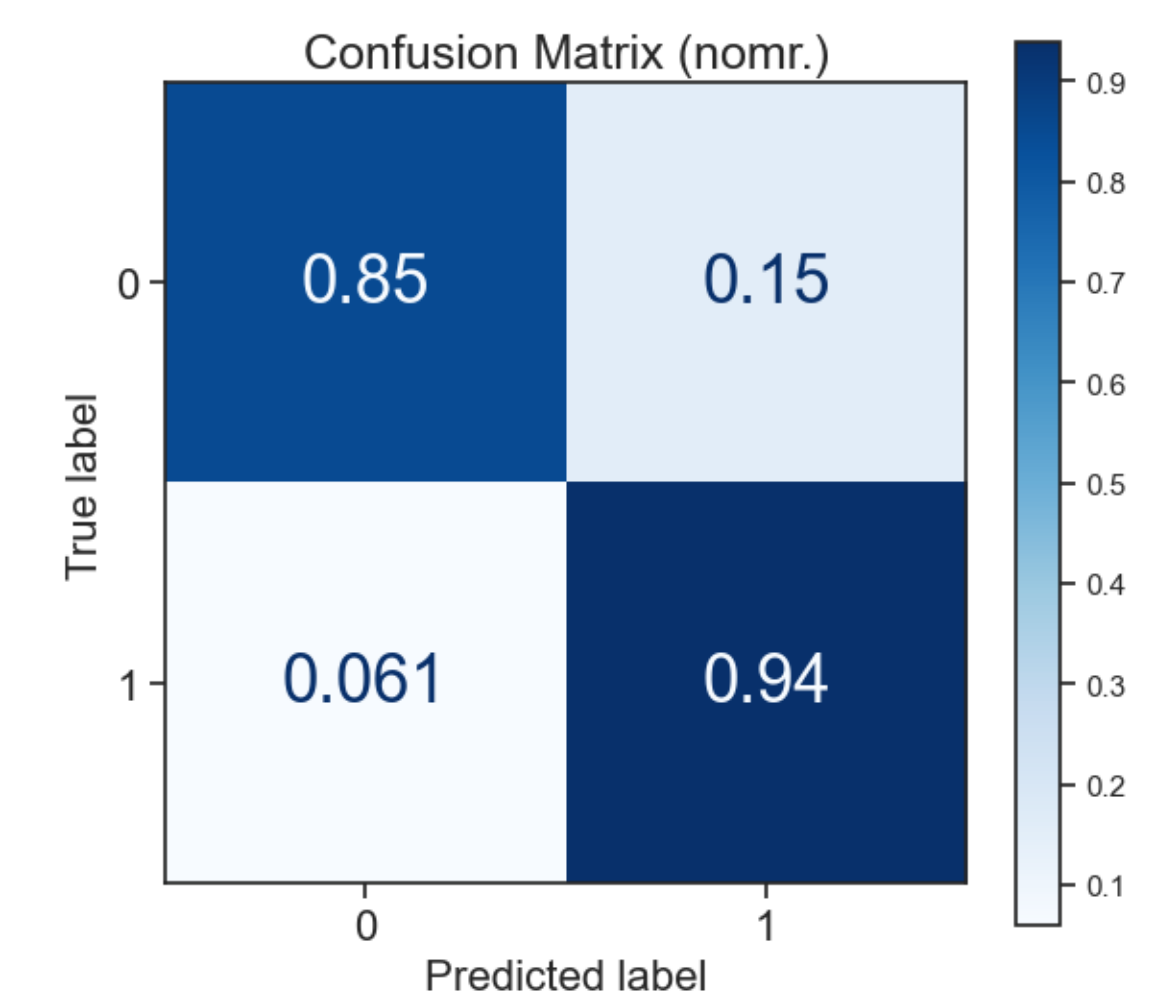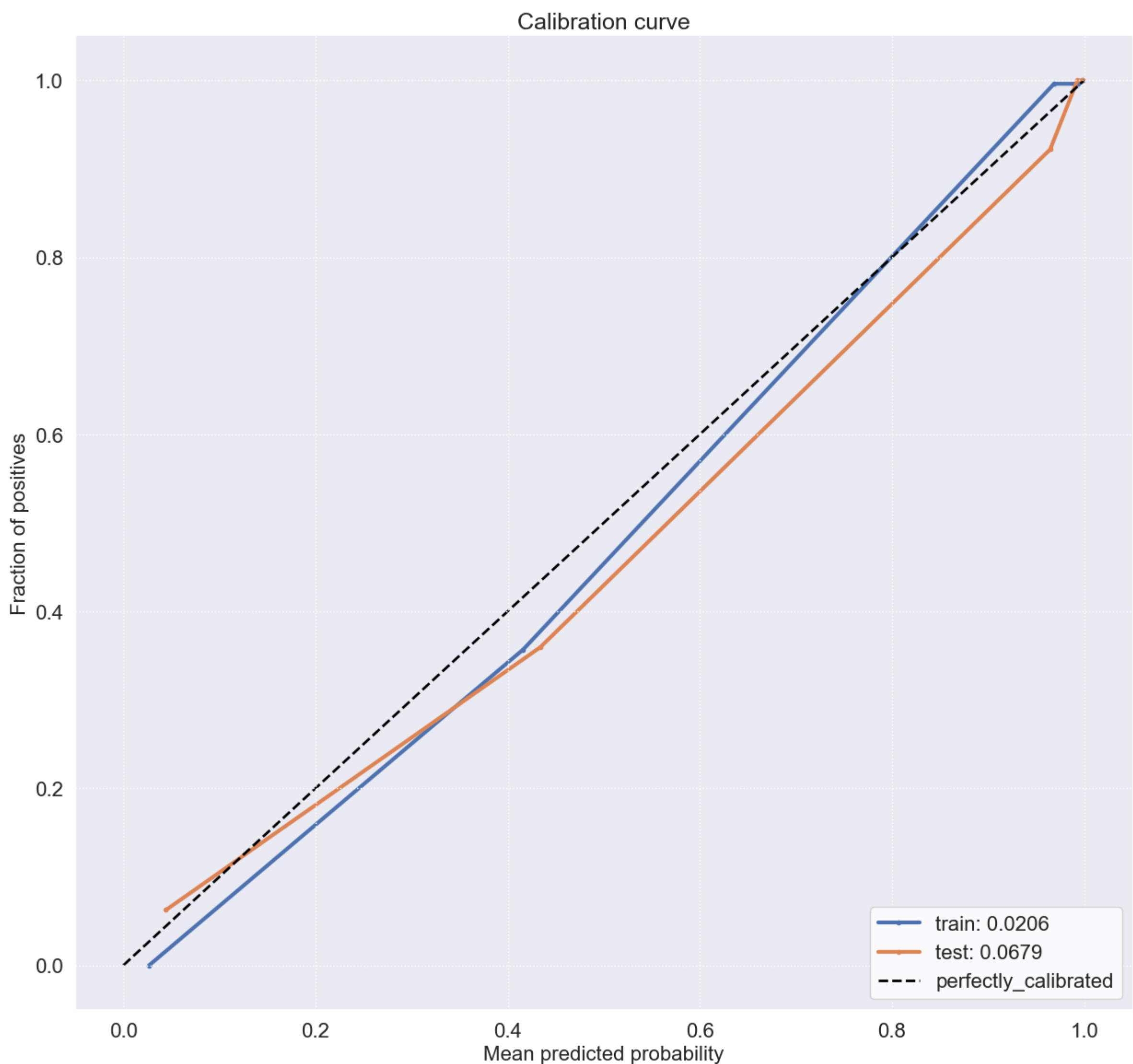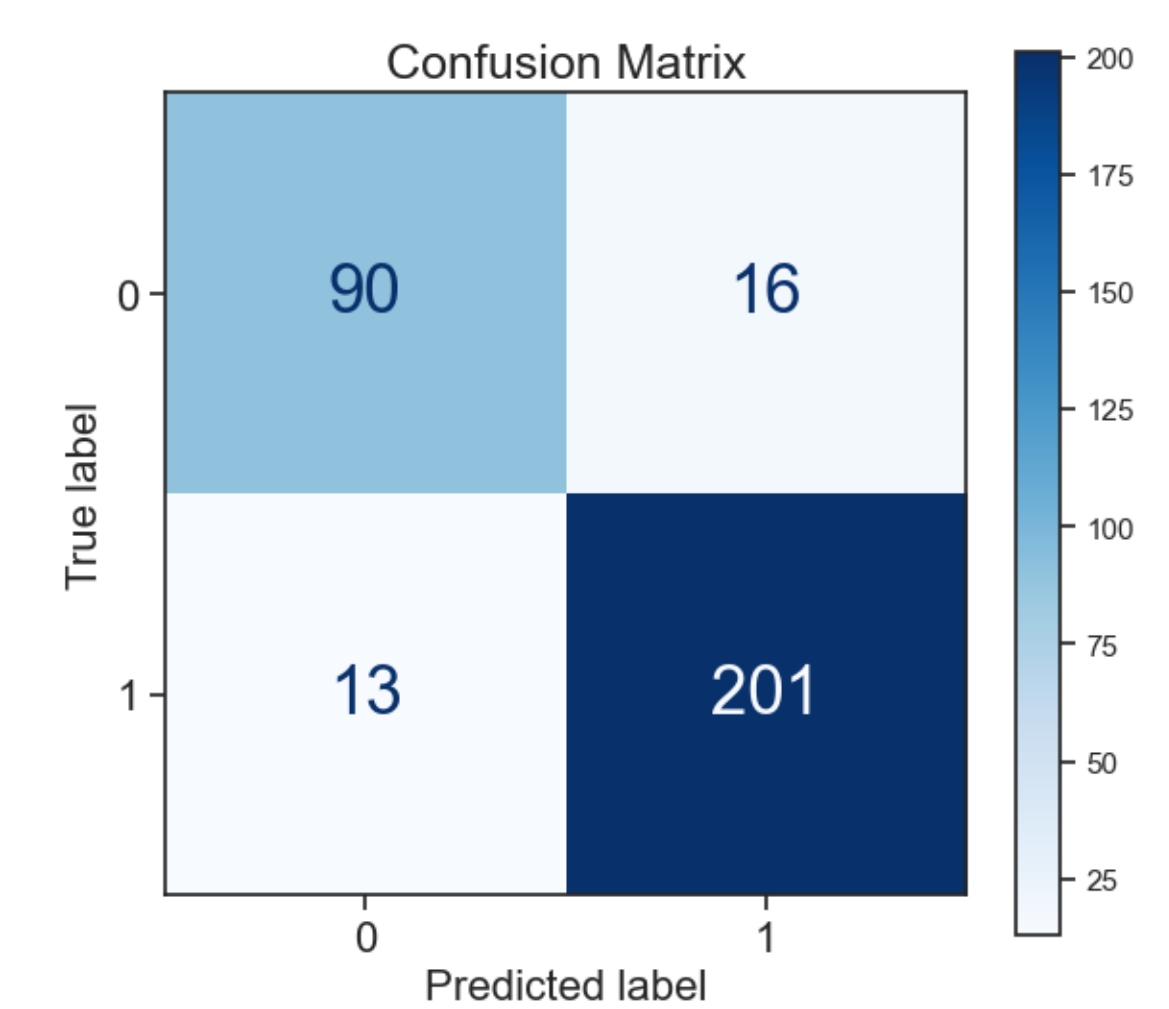
Metric	Train	Test
AUC	0.9967	0.9654
accuracy	0.9765	0.9094
balanced accuracy	0.9716	0.8942
recall_(sens)	0.9859	0.9393
specificity	0.9573	0.8491
precision	0.9791	0.9263
f1_score	0.9764	0.909

Protein: O76074 - metric: f1

Total (train): 1276

ROC curve

AUC_train = 1.000
AUC_test = 0.975
Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2317)
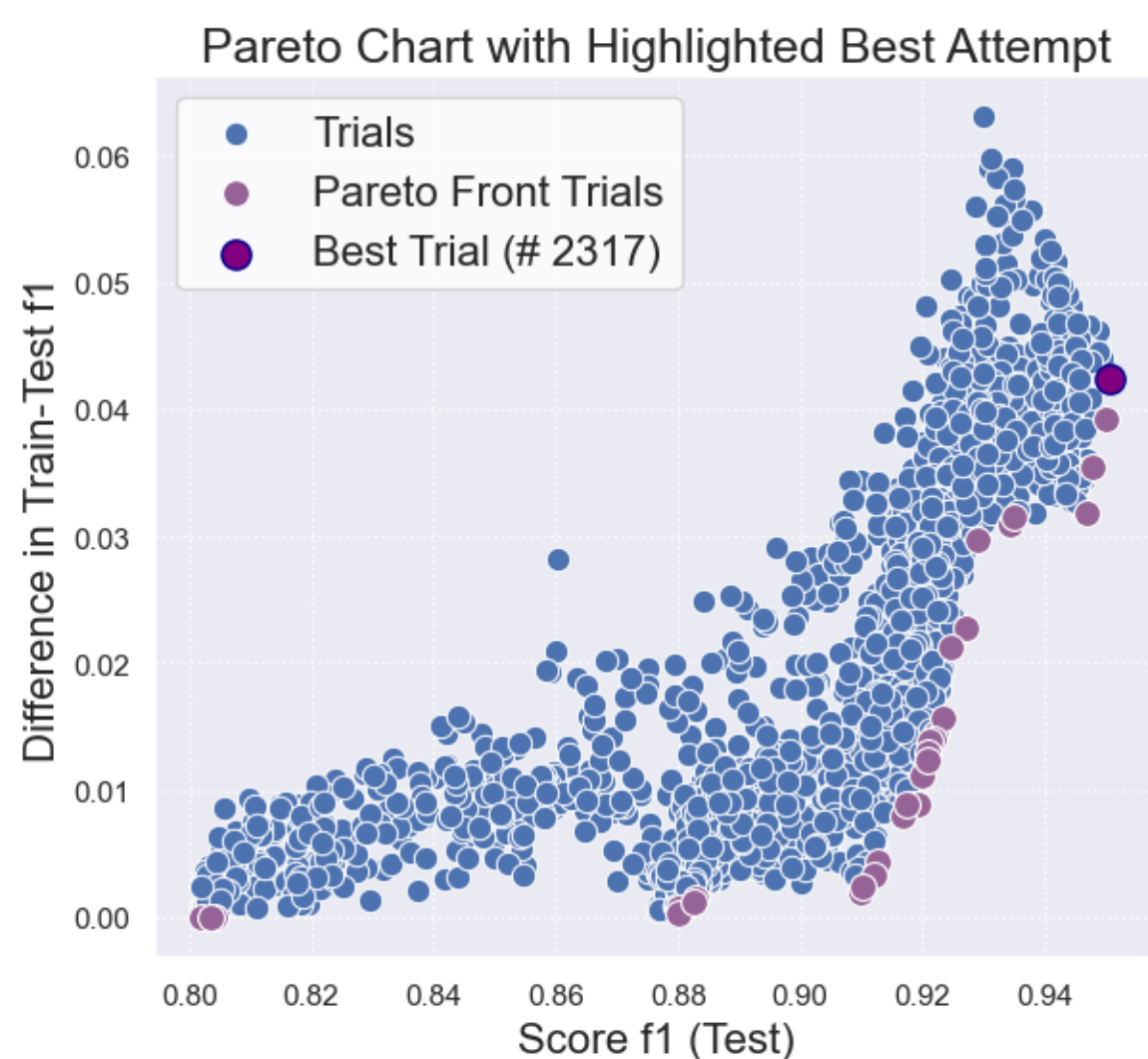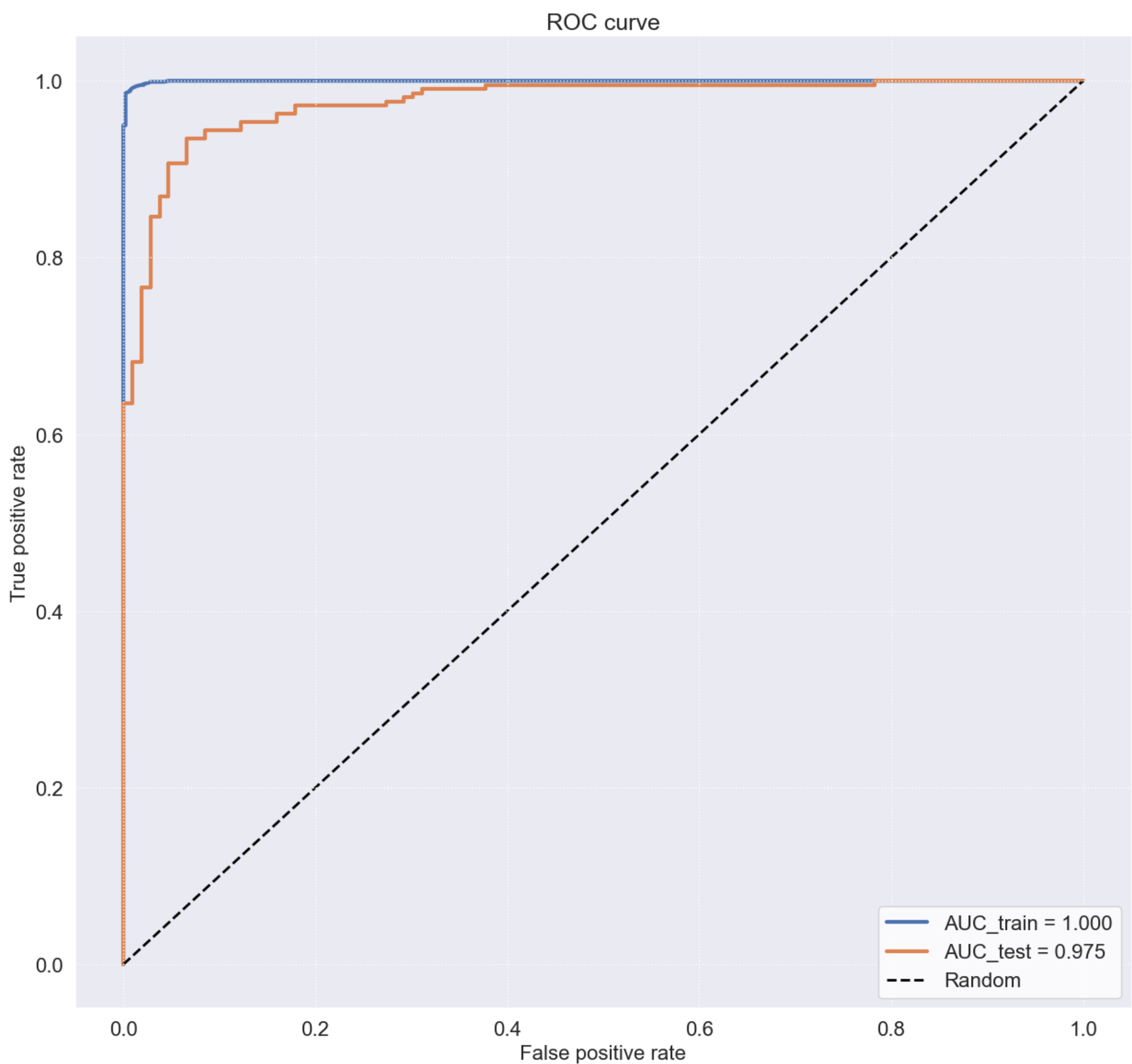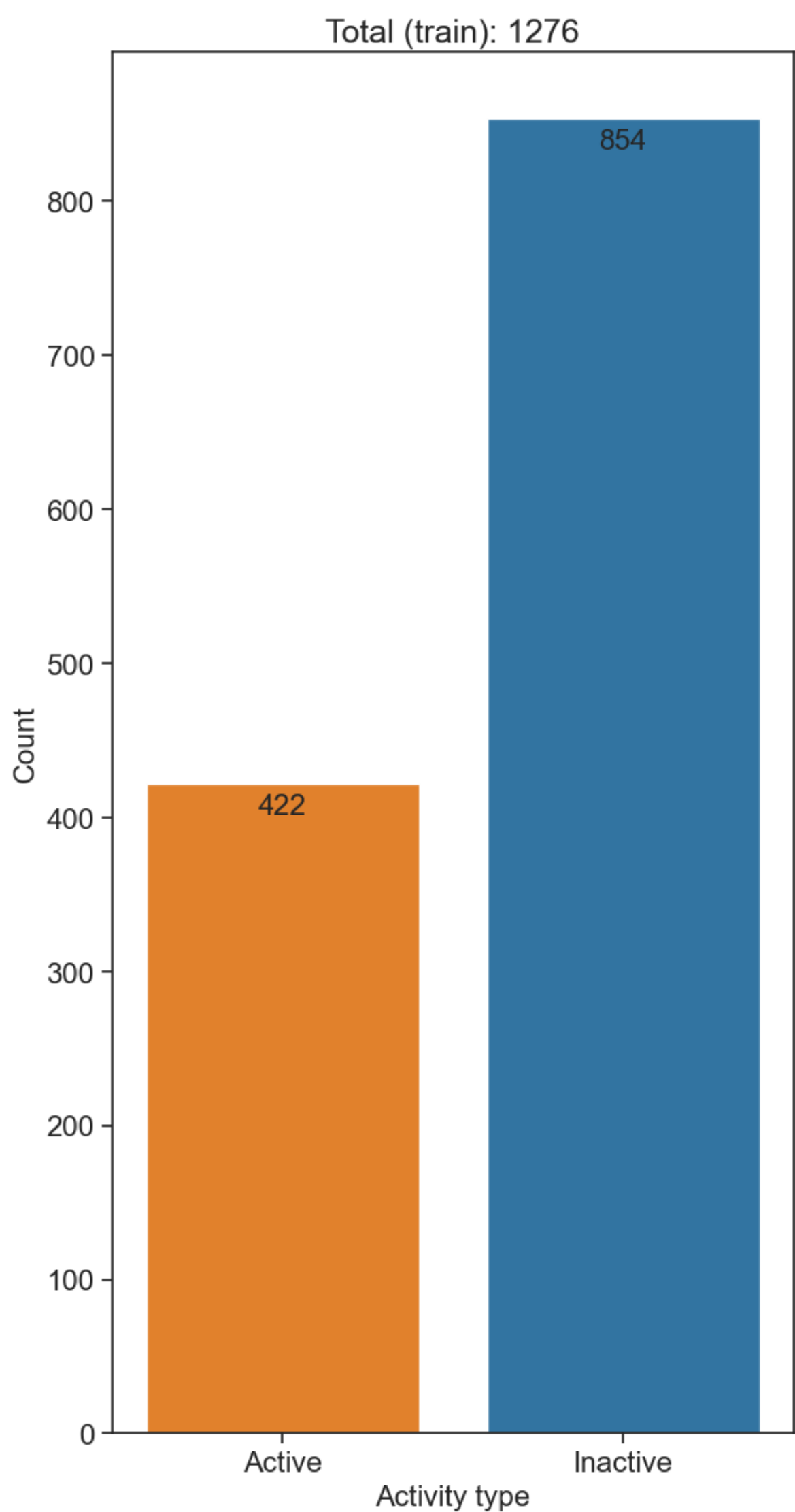
XGBoost Parameters	Value
alpha	4.3534
colsample_bytree	0.7359
gamma	0.0244
lambda	3.4388
learning_rate	0.0878
max_depth	10.0
min_child_weight	2.0
n_estimators	451.0
reg_alpha	0.1454
reg_lambda	0.1742
scale_pos_weight	1.3989
subsample	0.587

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

train: 0.0081
test: 0.0567
perfectly_calibrated

Active/Inactive probability

Inactive zone | Active zone

type
active (214)
inactive (106)
decoy (320)
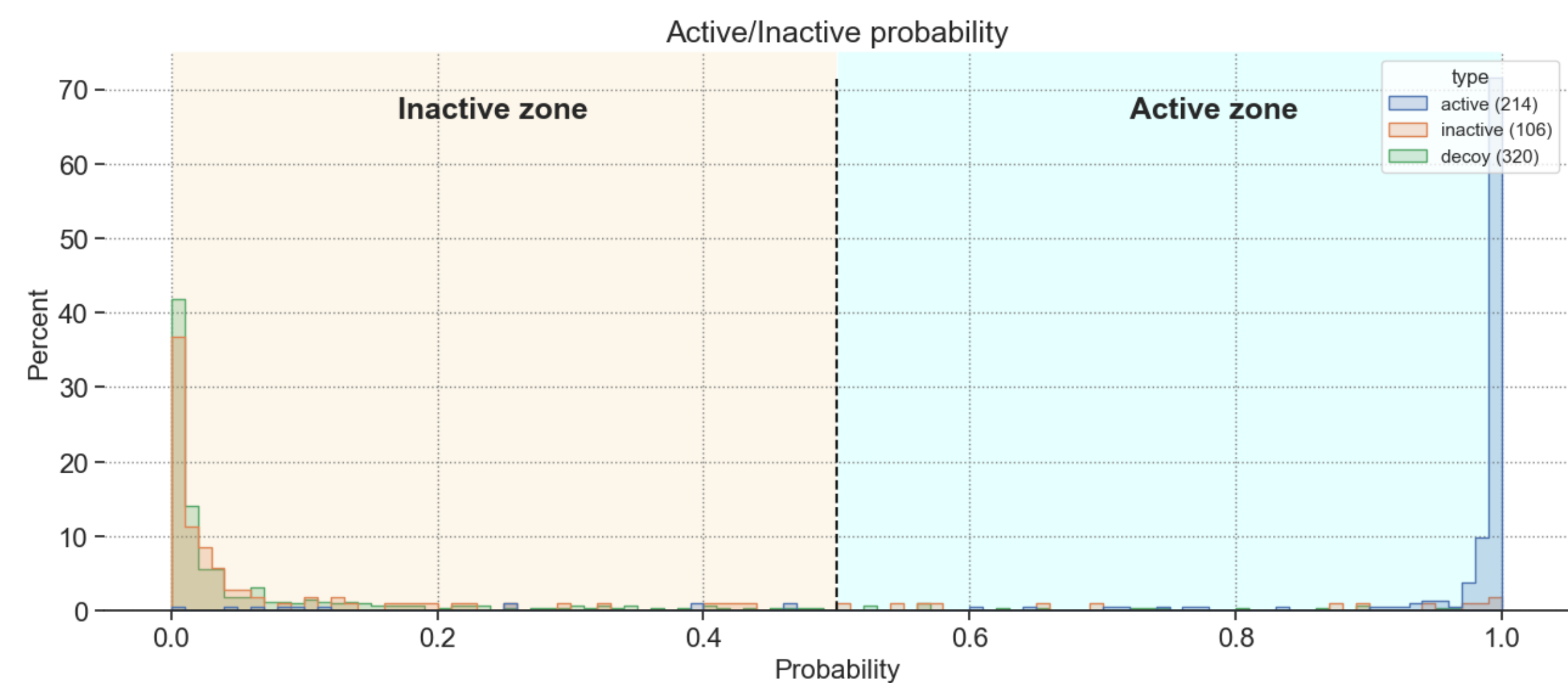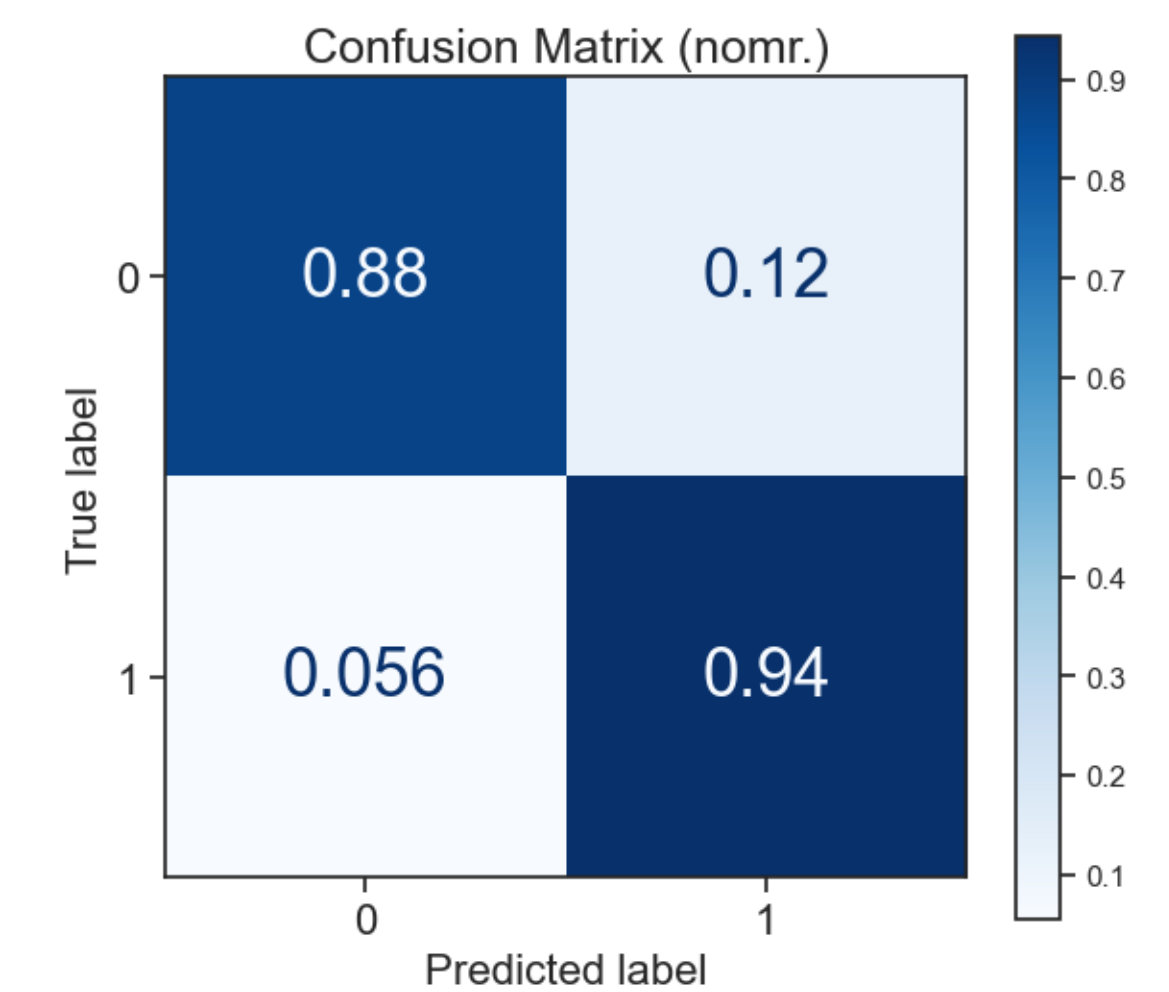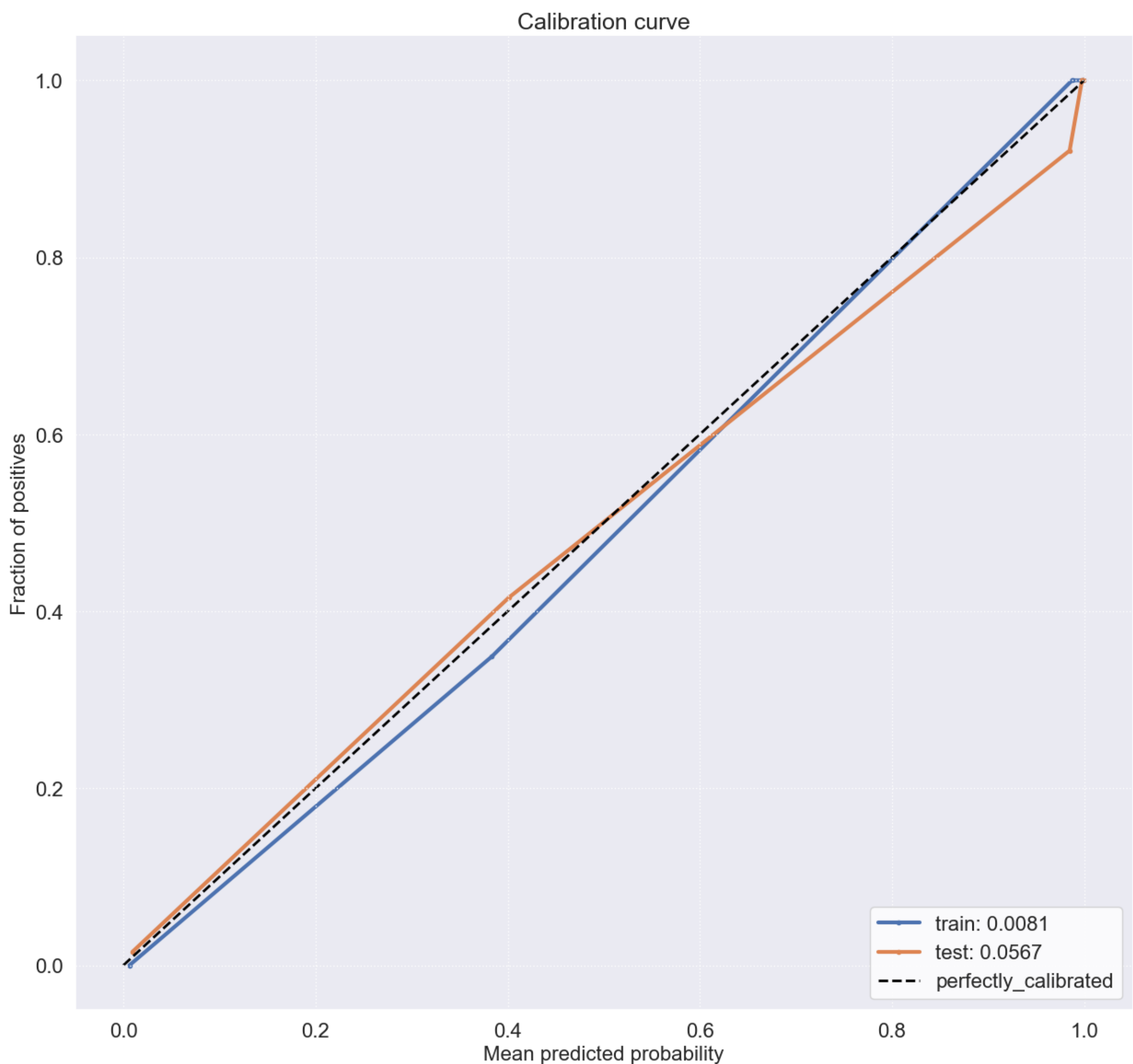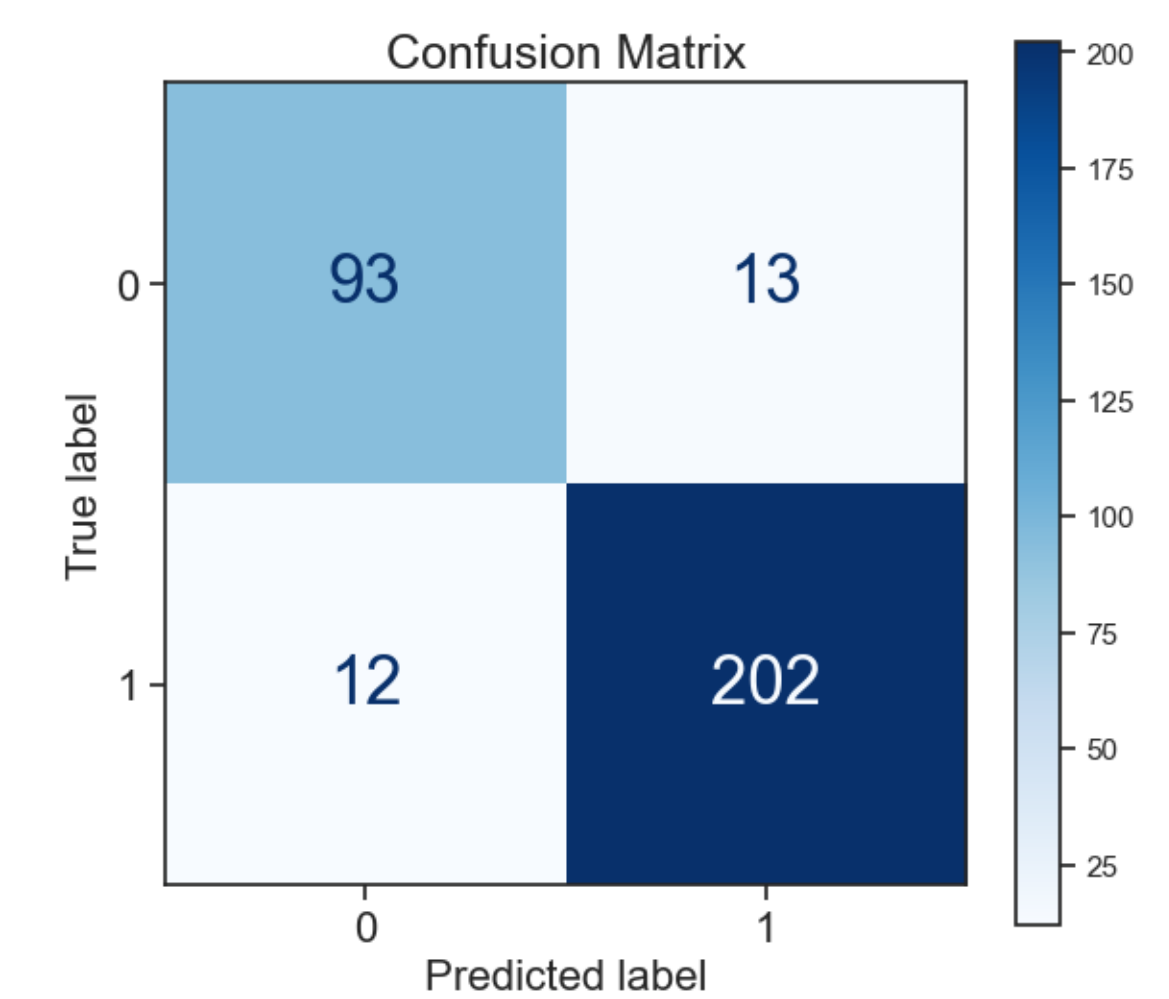
Metric	Train	Test
AUC	0.9997	0.9752
accuracy	0.9898	0.9219
balanced accuracy	0.9858	0.9106
recall_(sens)	0.9977	0.9439
specificity	0.9739	0.8774
precision	0.9873	0.9395
f1_score	0.9898	0.9218

Protein: O76074 - metric: precision

Total (train): 1276

ROC curve

- AUC_train = 0.999
- AUC_test = 0.976
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 766)
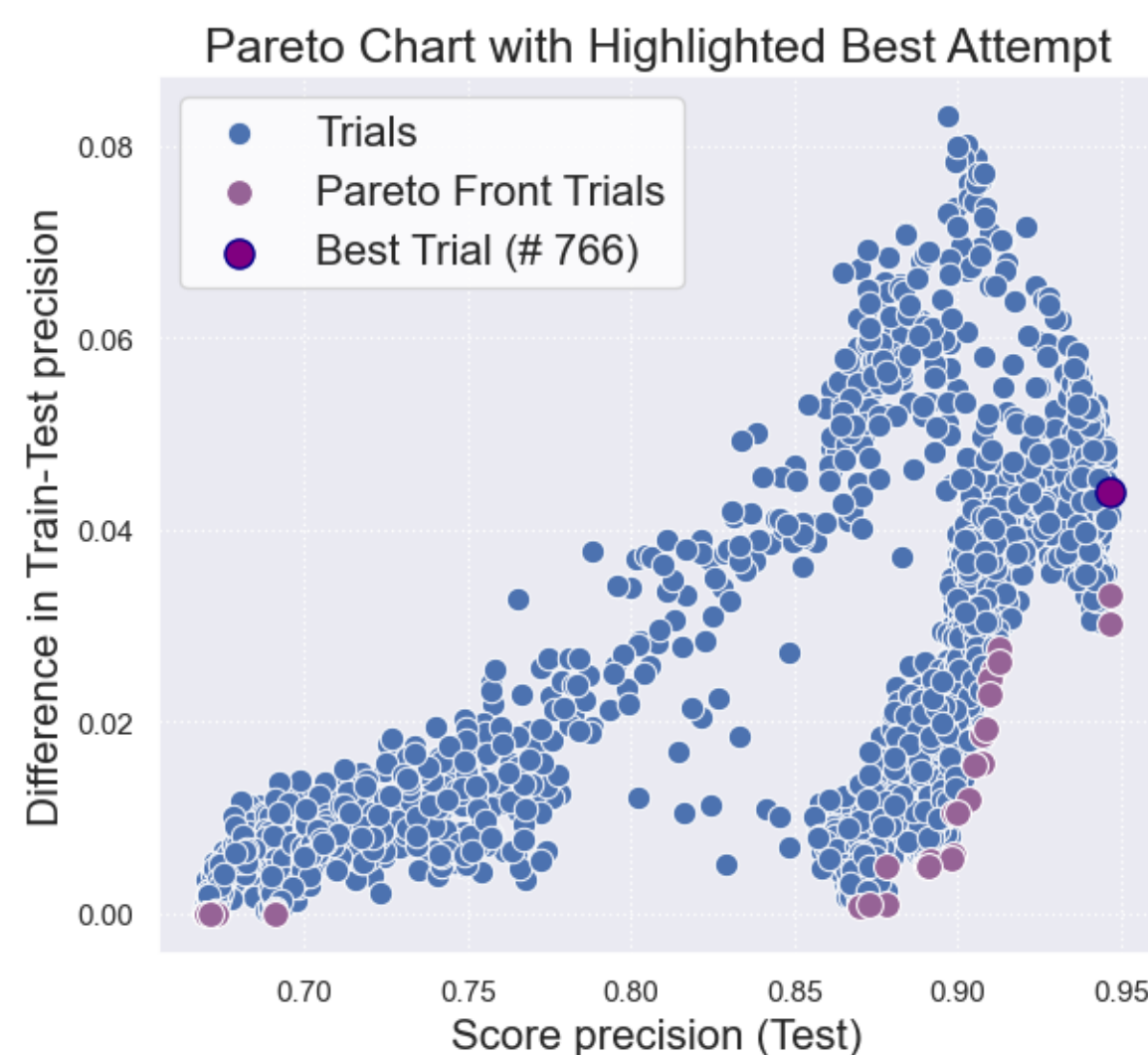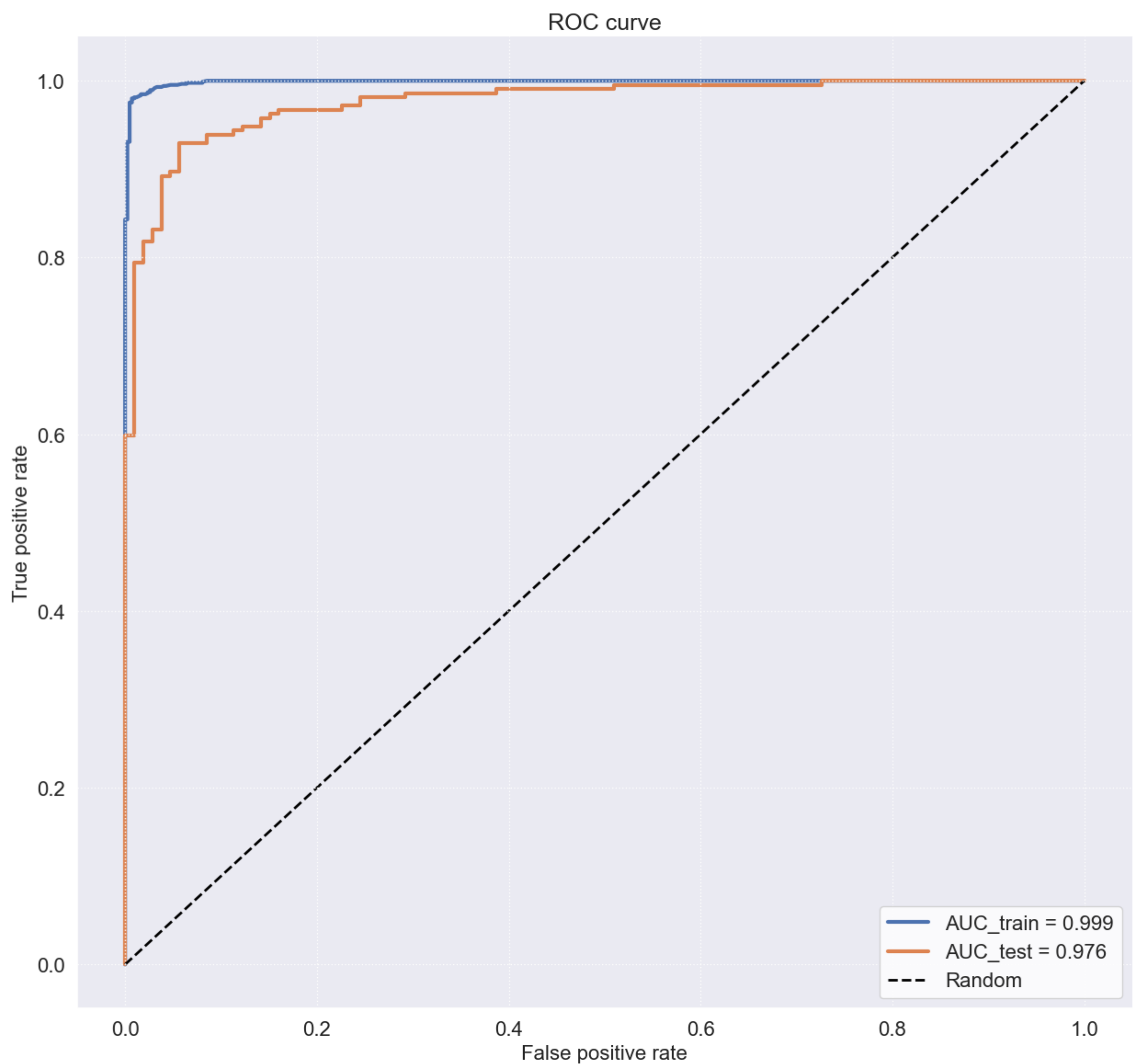
XGBoost Parameters	Value
alpha	0.8118
colsample_bytree	0.5977
gamma	0.1306
lambda	4.0969
learning_rate	0.1234
max_depth	10.0
min_child_weight	2.0
n_estimators	166.0
reg_alpha	0.5474
reg_lambda	0.5239
scale_pos_weight	1.0925
subsample	0.6609

Confusion Matrix

Calibration curve

- train: 0.0144
- test: 0.0543
- perfectly_calibrated

Confusion Matrix (nomr.)

Active/Inactive probability

type
- active (214)
- inactive (106)
- decoy (320)
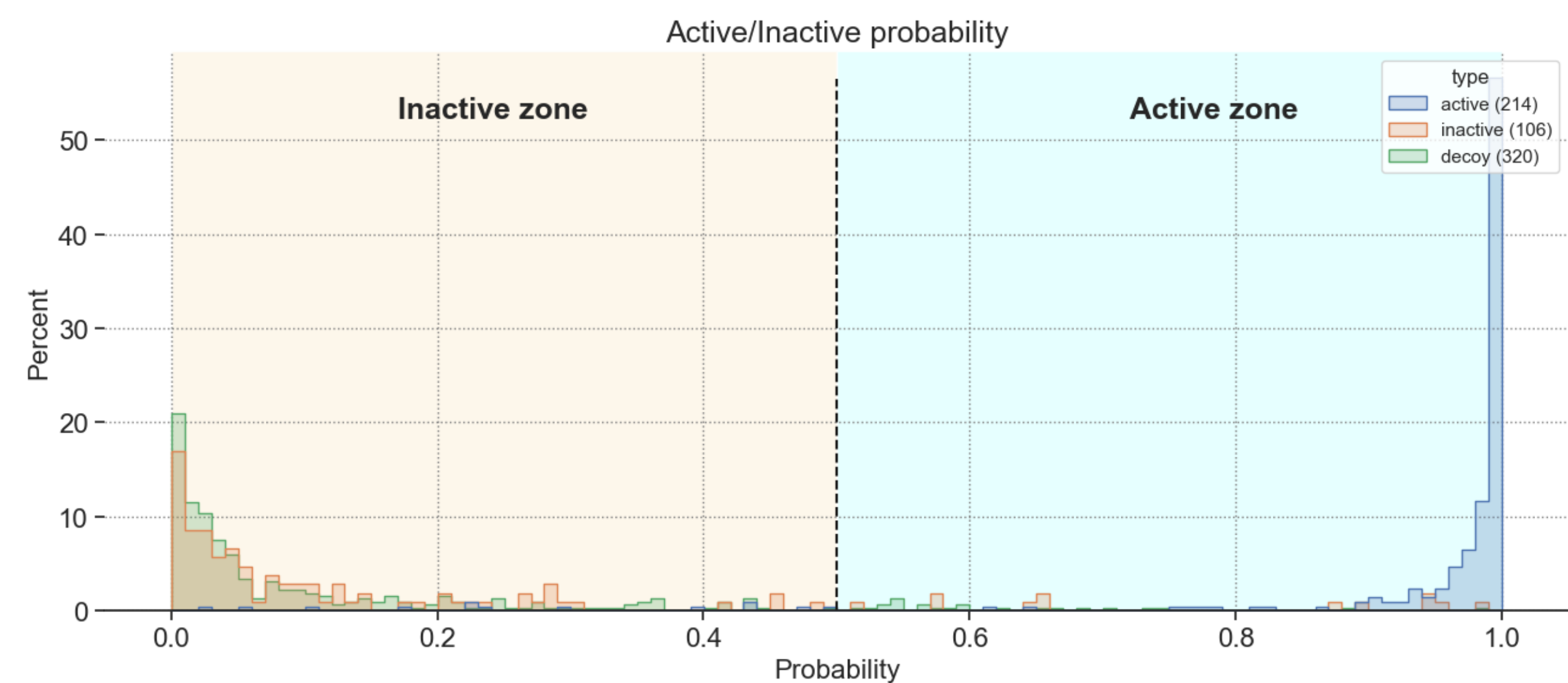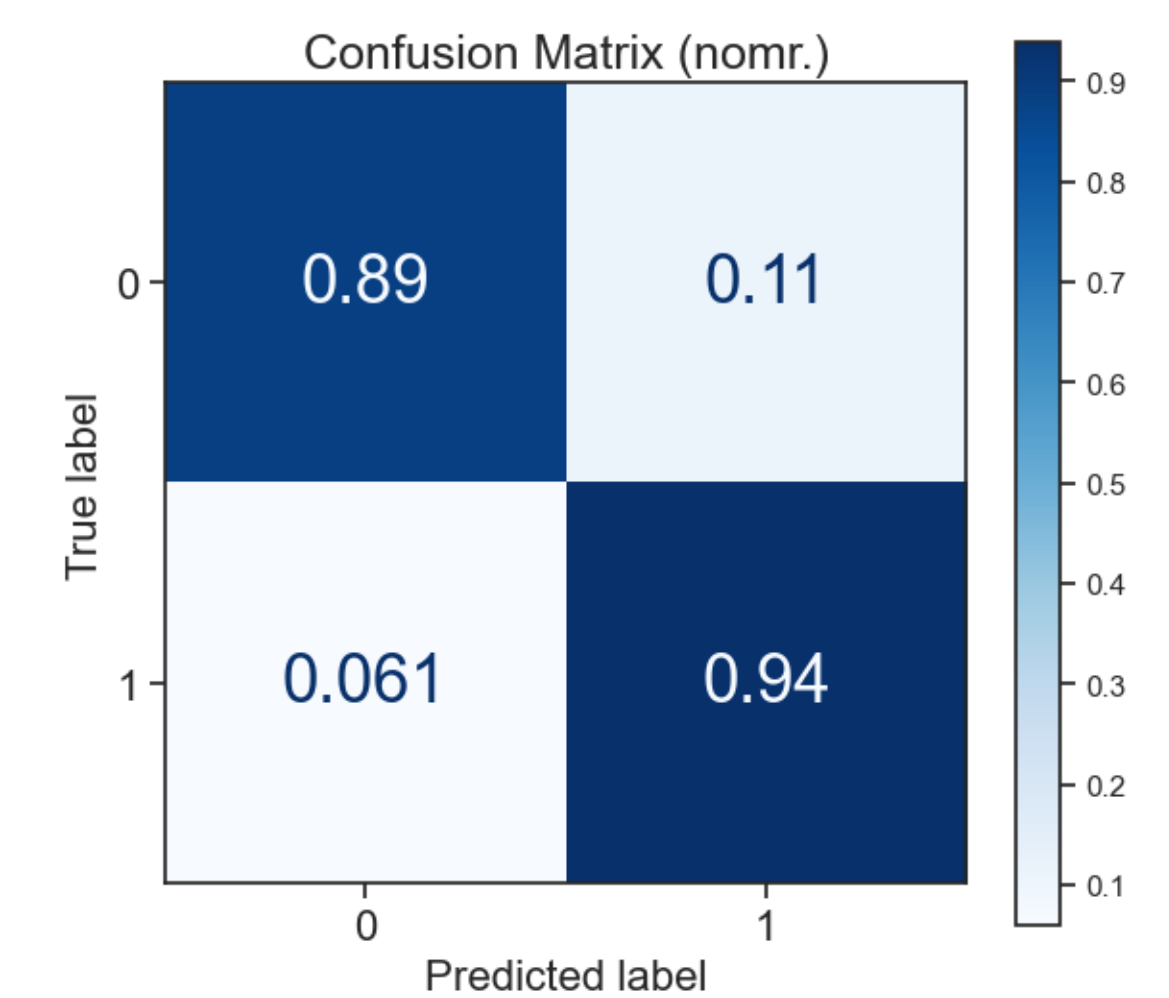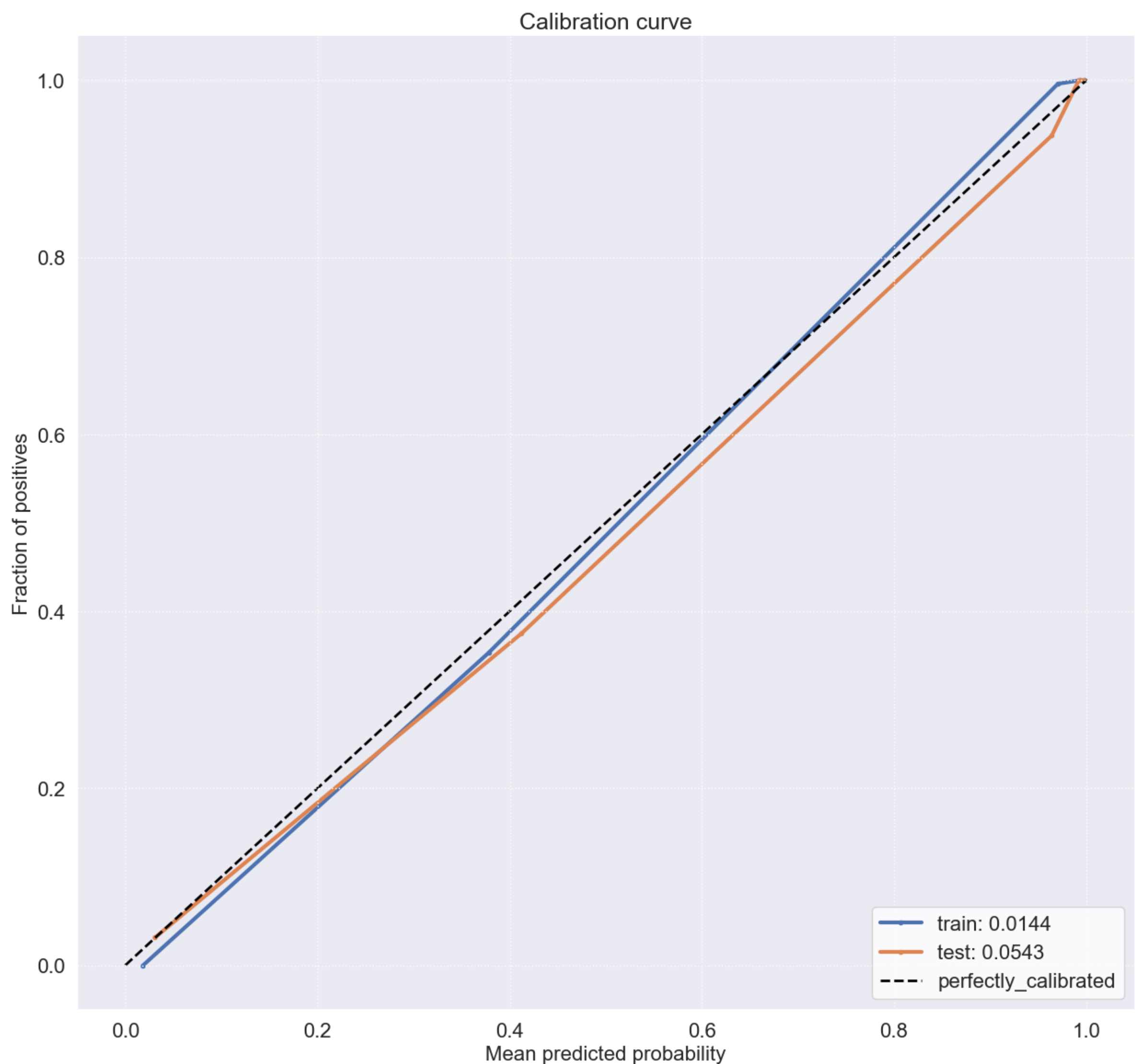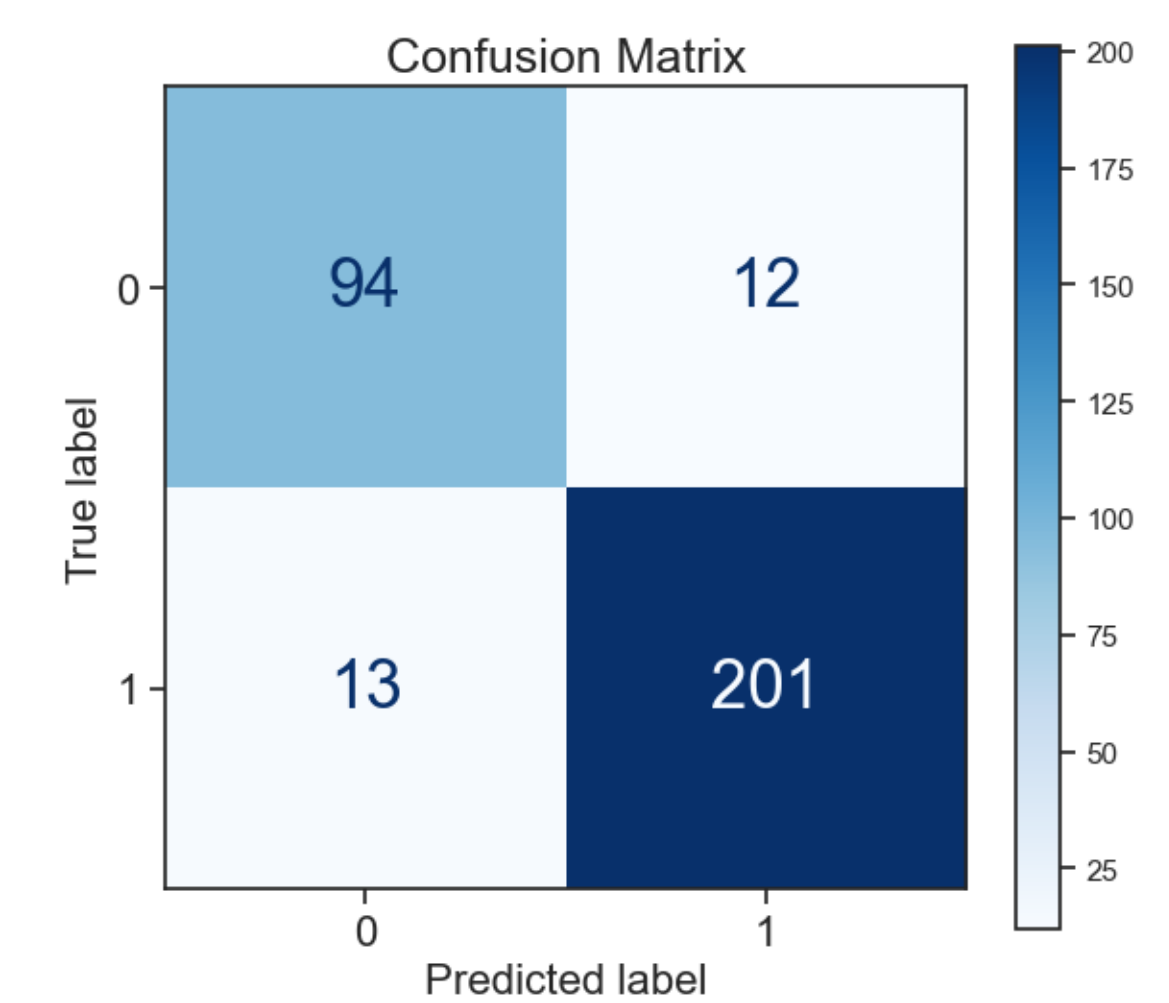
Metric	Train	Test
AUC	0.9988	0.976
accuracy	0.9835	0.9219
balanced accuracy	0.9829	0.913
recall_(sens)	0.9848	0.9393
specificity	0.981	0.8868
precision	0.9906	0.9437
f1_score	0.9836	0.922

Protein: O76074 - metric: SoftVotingXGBoostEnsemble
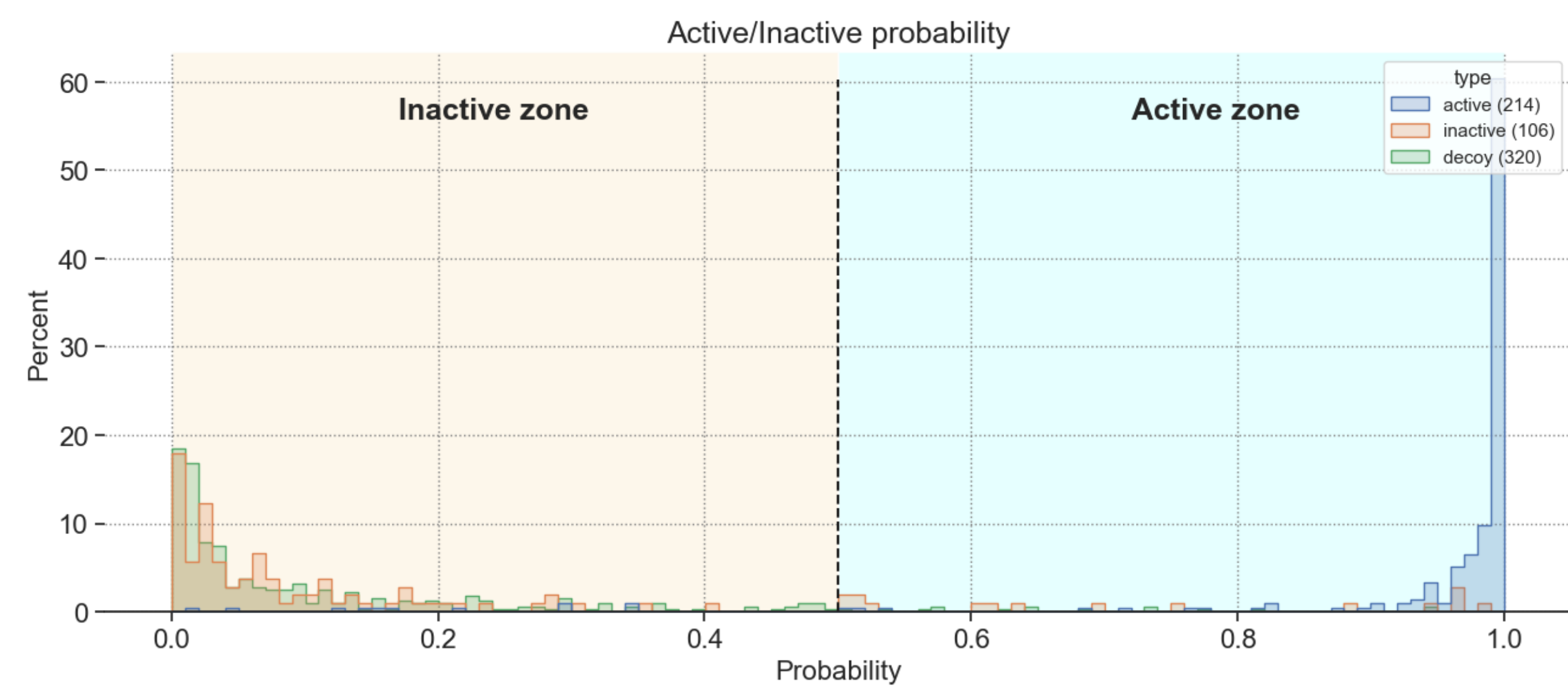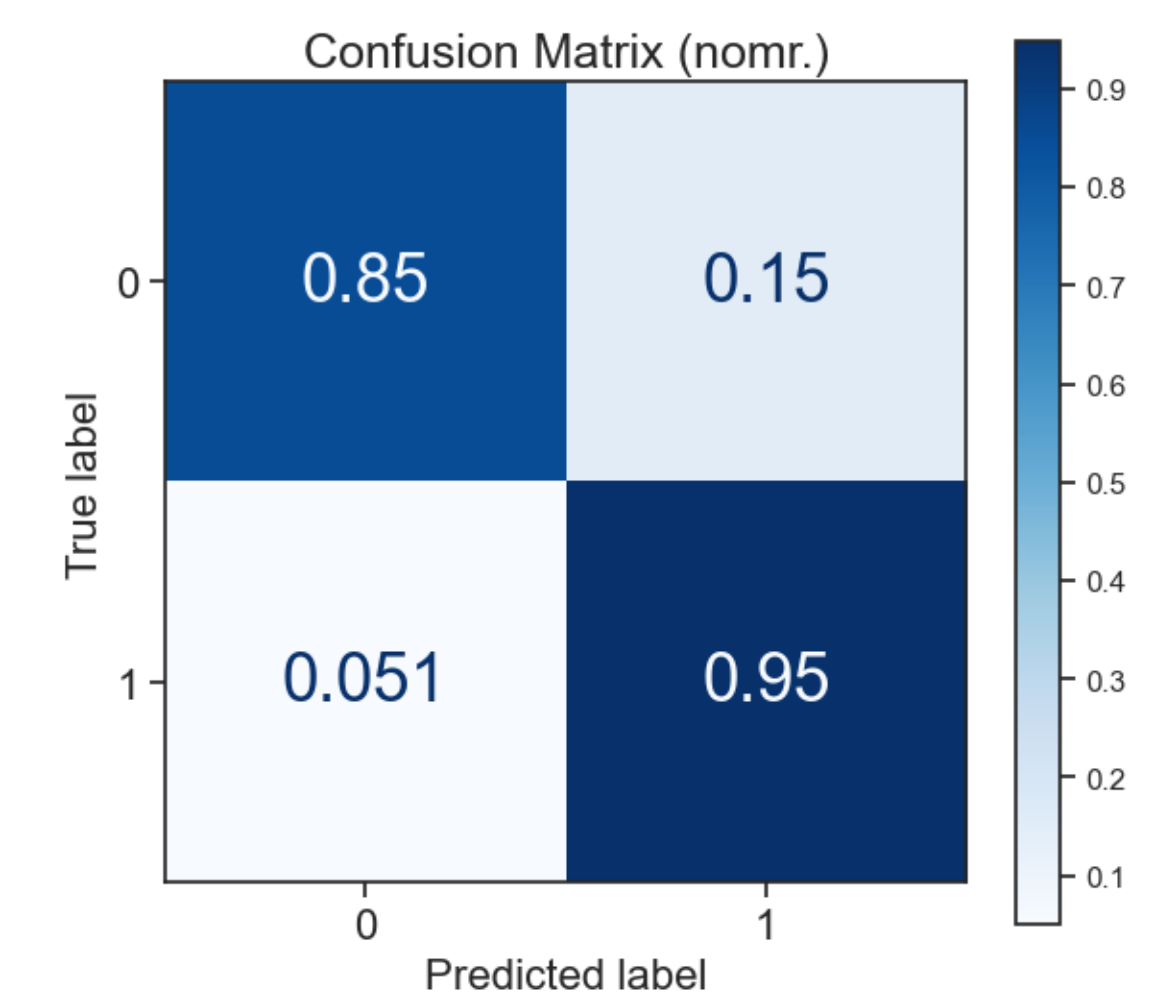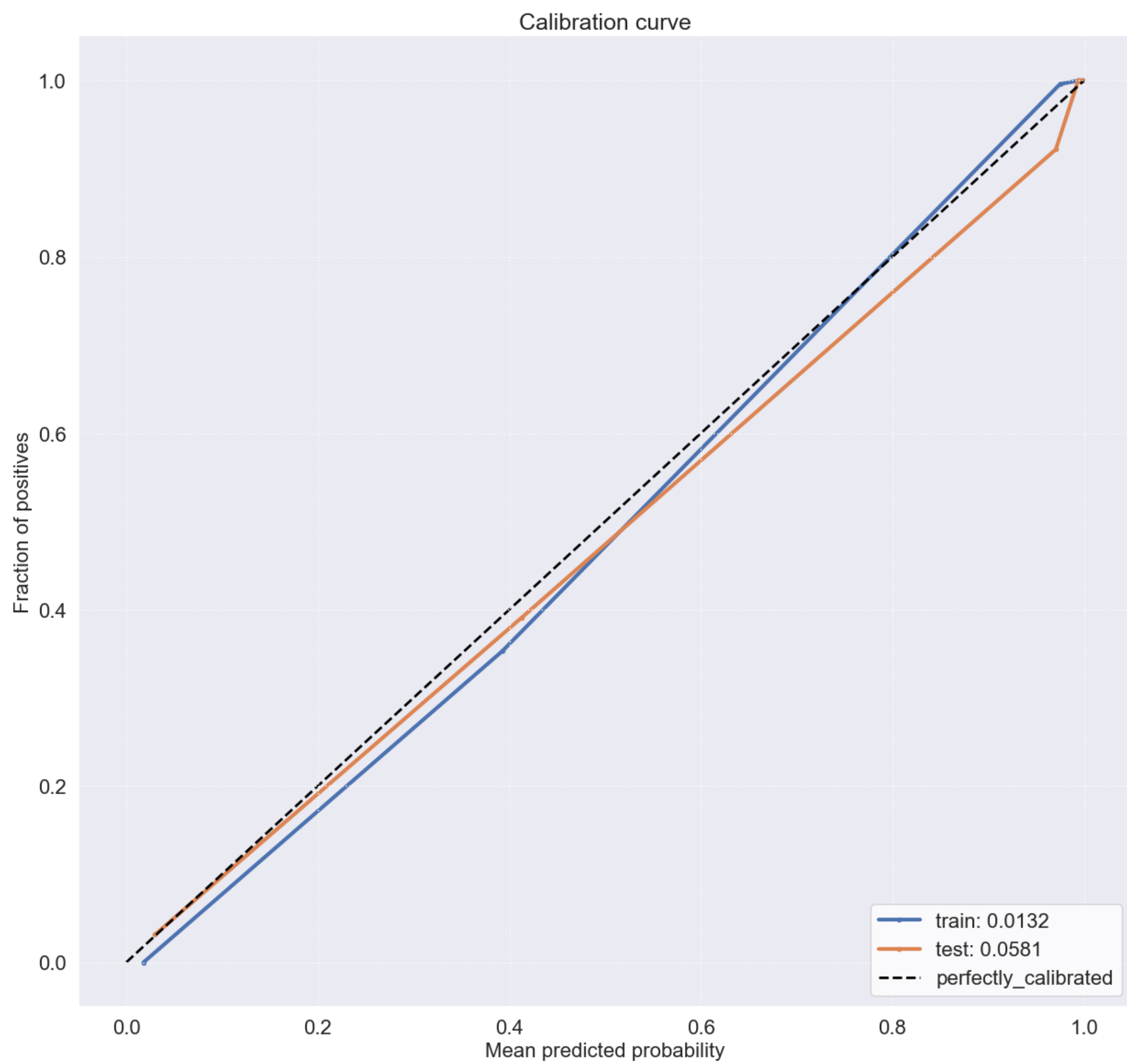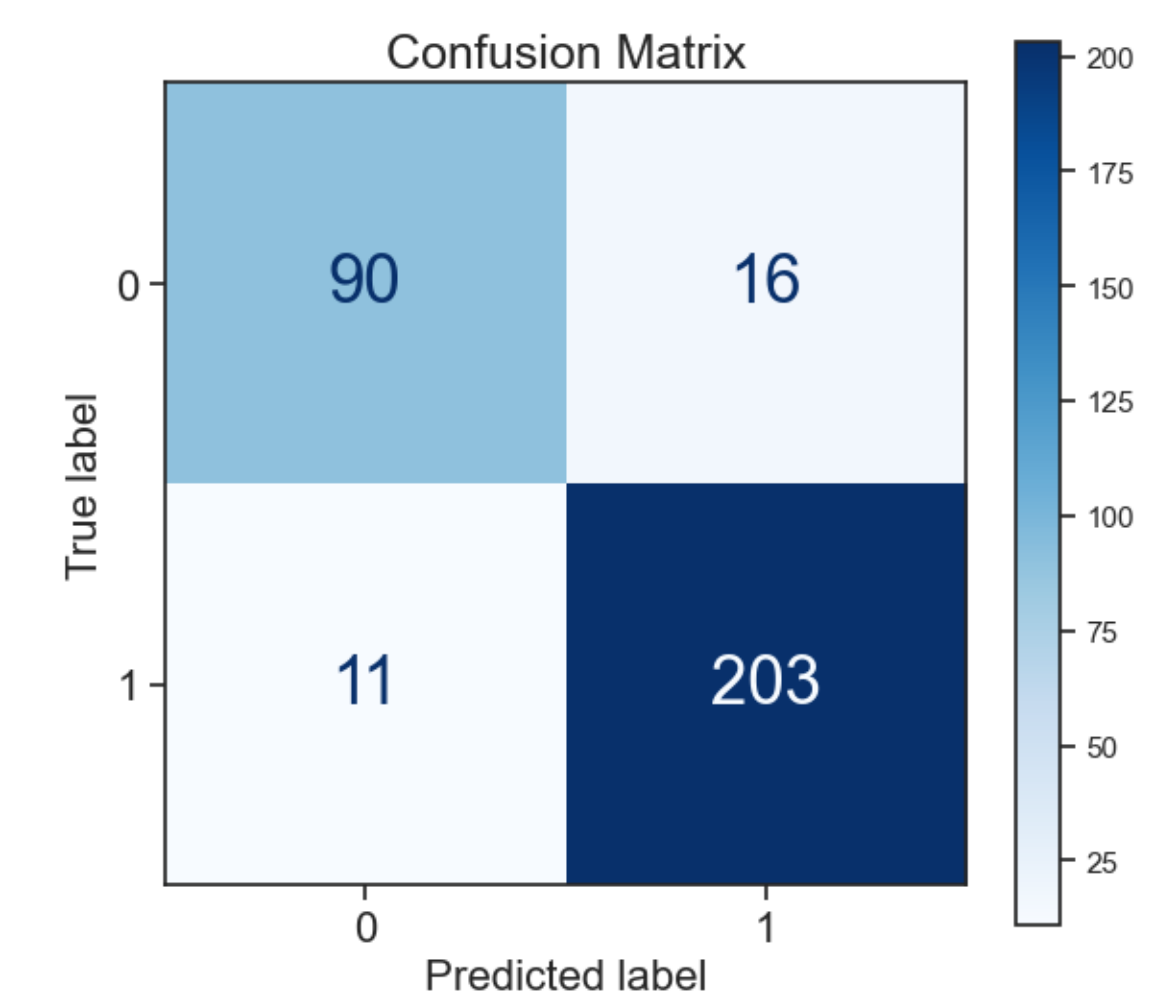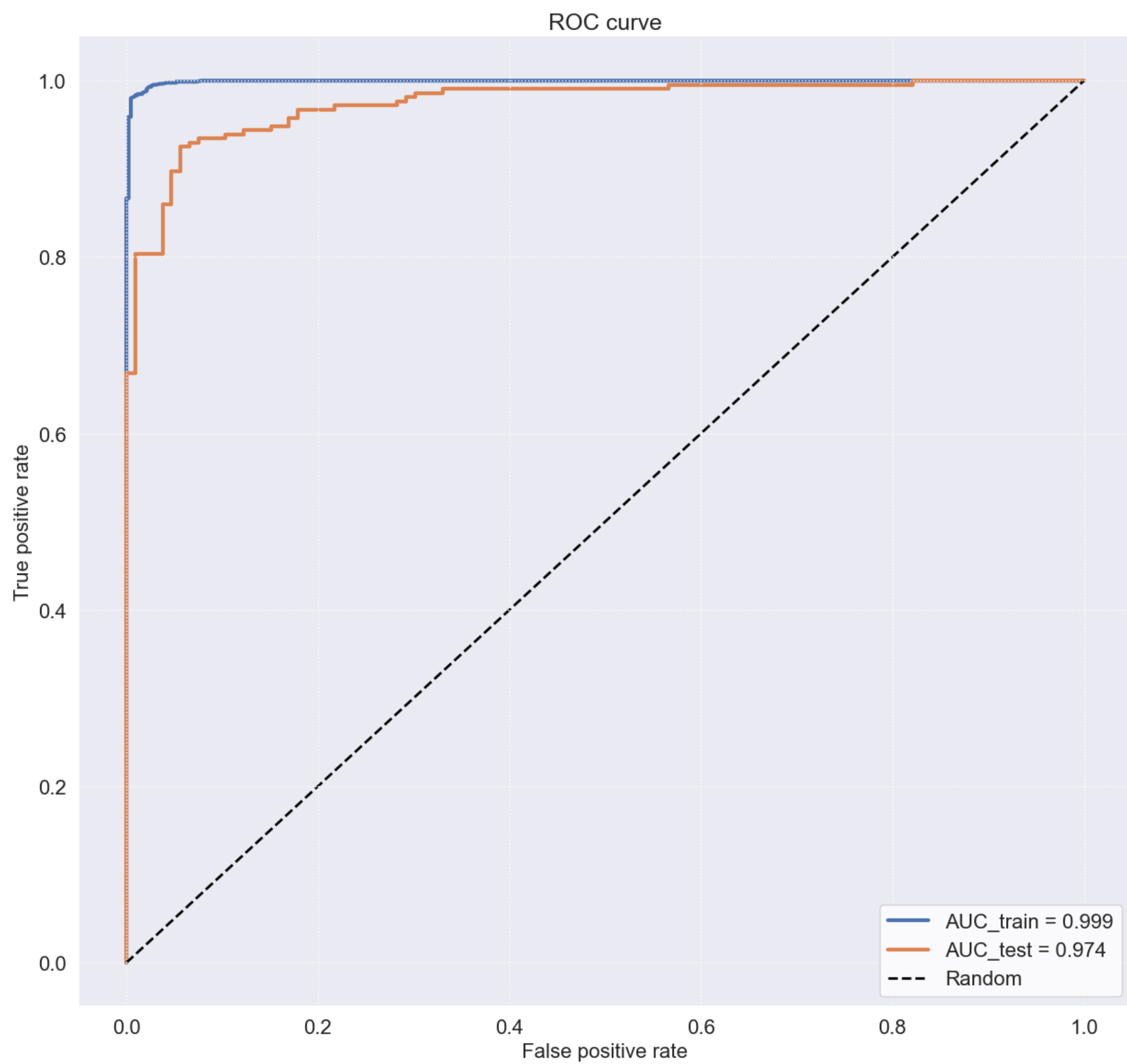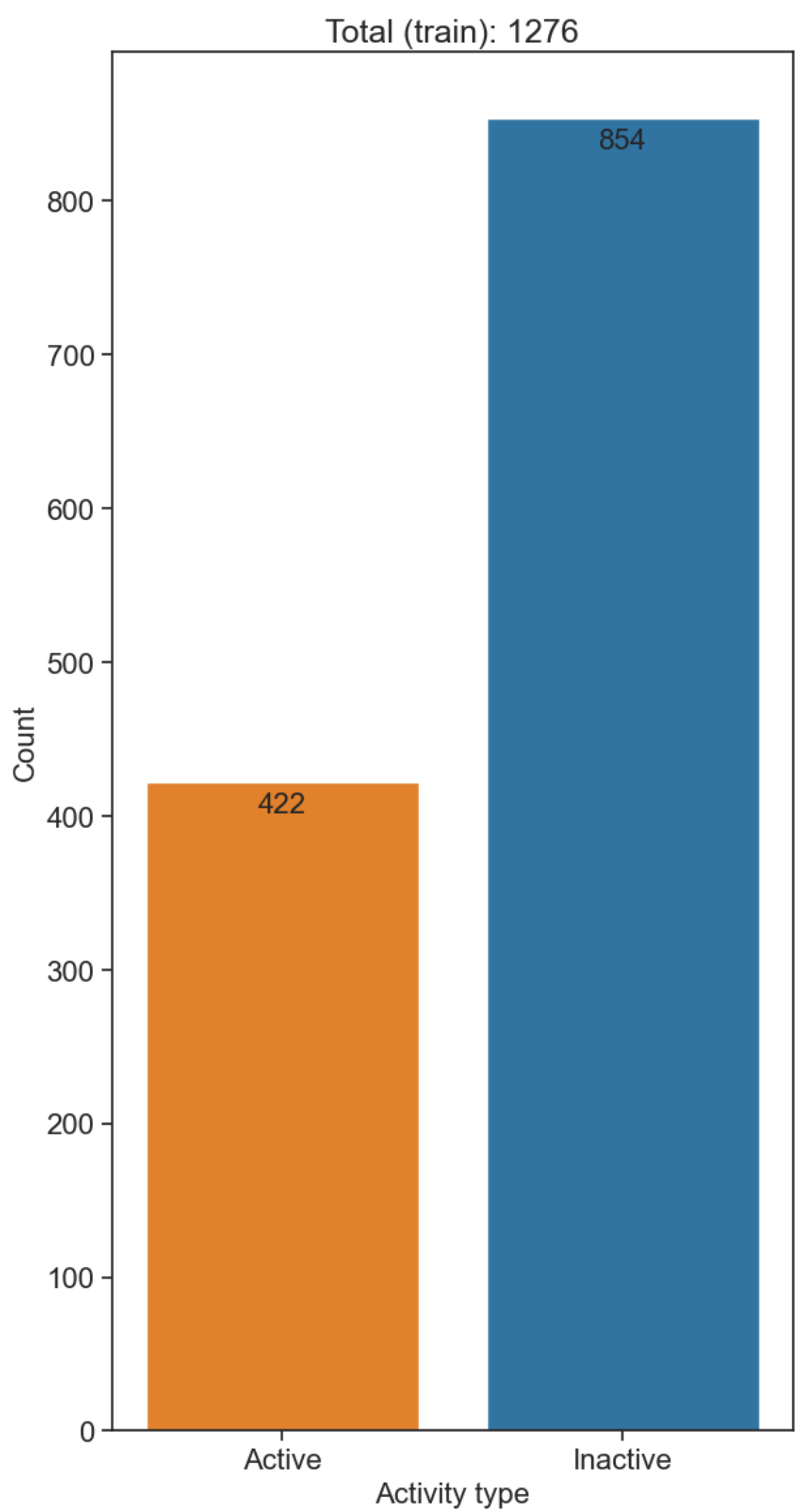
Metric	Train	Test
AUC	0.9992	0.9744
accuracy	0.9867	0.9156
balanced accuracy	0.9847	0.8988
recall_(sens)	0.9906	0.9486
specificity	0.9787	0.8491
precision	0.9895	0.9269
f1_score	0.9867	0.9151