Protein: P48039 - metric: balanced_accuracy

Total (train): 624

ROC curve

AUC_train = 0.996
AUC_test = 0.935
Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2807)

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

train: 0.0335
test: 0.1003
perfectly_calibrated

Active/Inactive probability

type
- active (86)
- inactive (70)
- decoy (156)

Inactive zone   Active zone
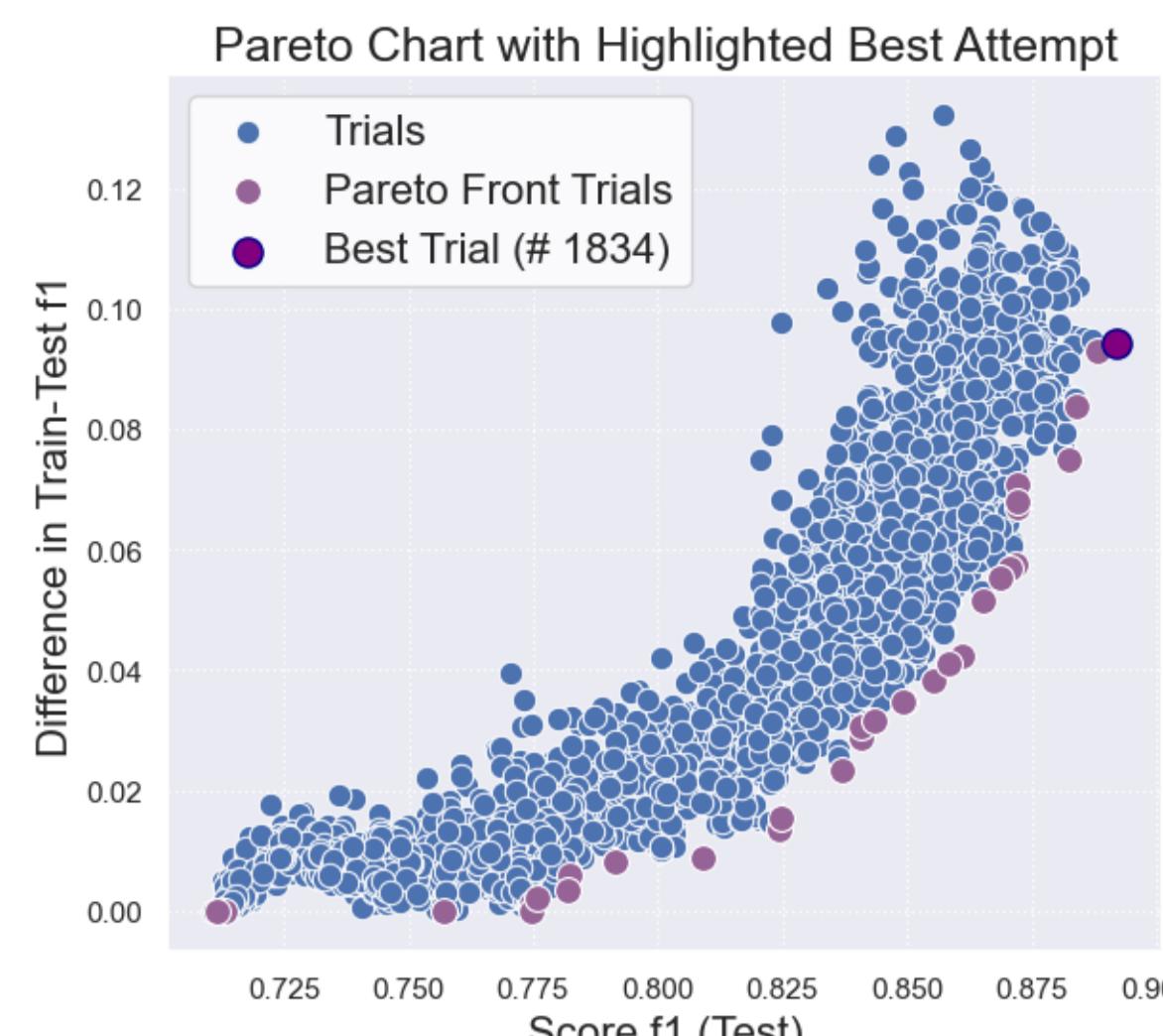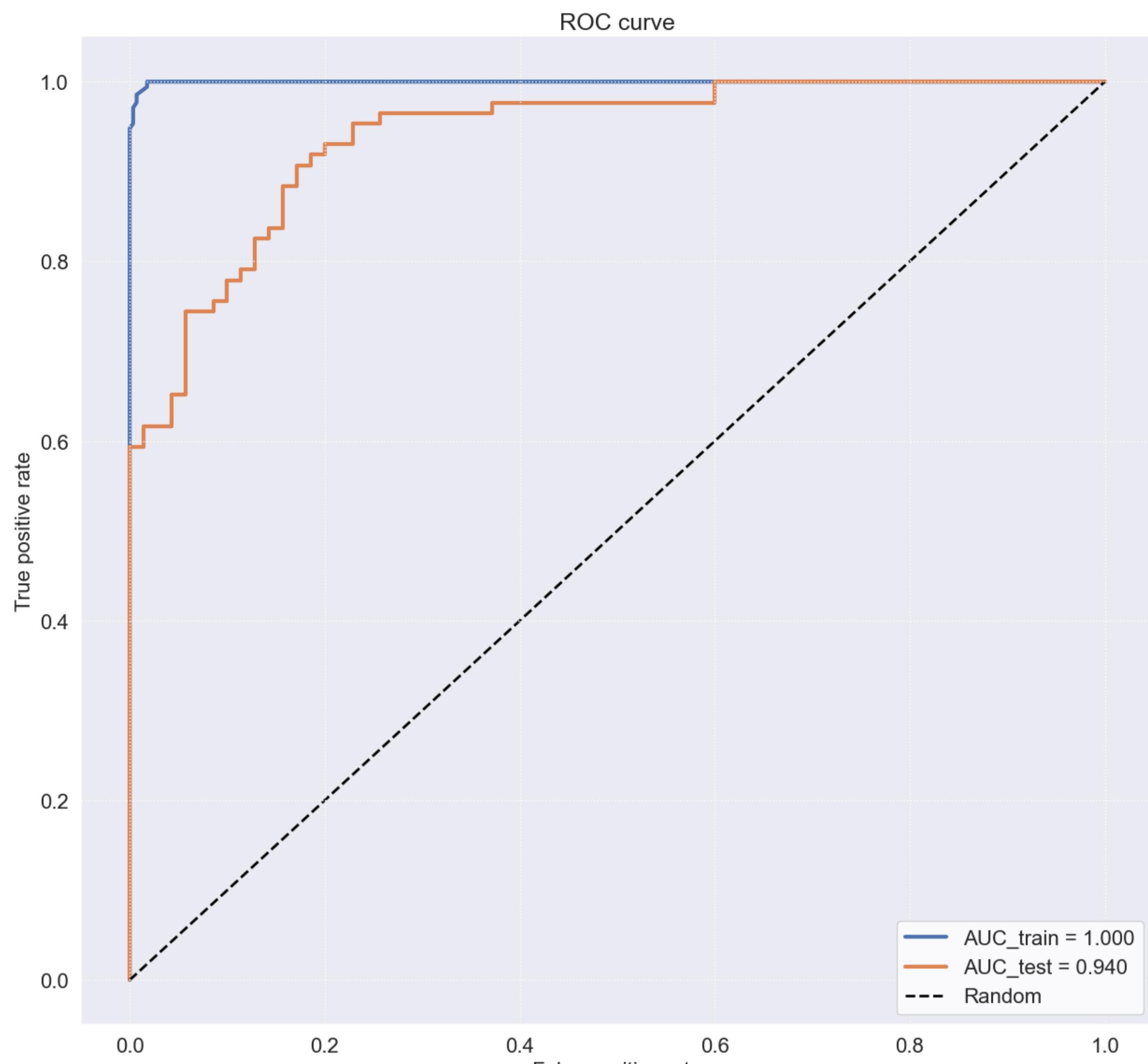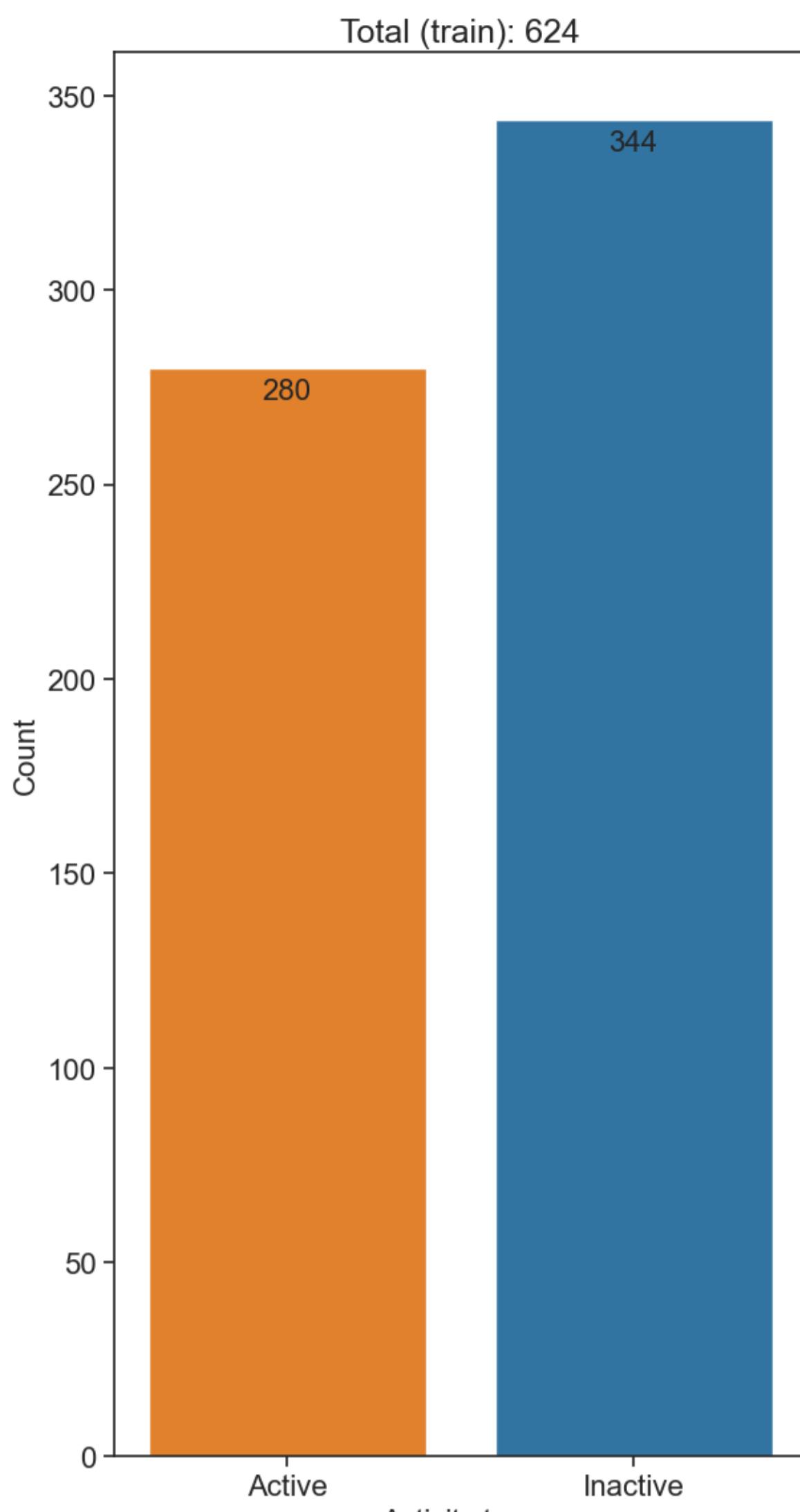
XGBoost Parameters	Value
alpha	0.862
colsample_bytree	0.5079
gamma	0.9621
lambda	2.5963
learning_rate	0.0384
max_depth	9.0
min_child_weight	1.0
n_estimators	968.0
reg_alpha	0.0176
reg_lambda	0.8057
scale_pos_weight	1.7173
subsample	0.8527
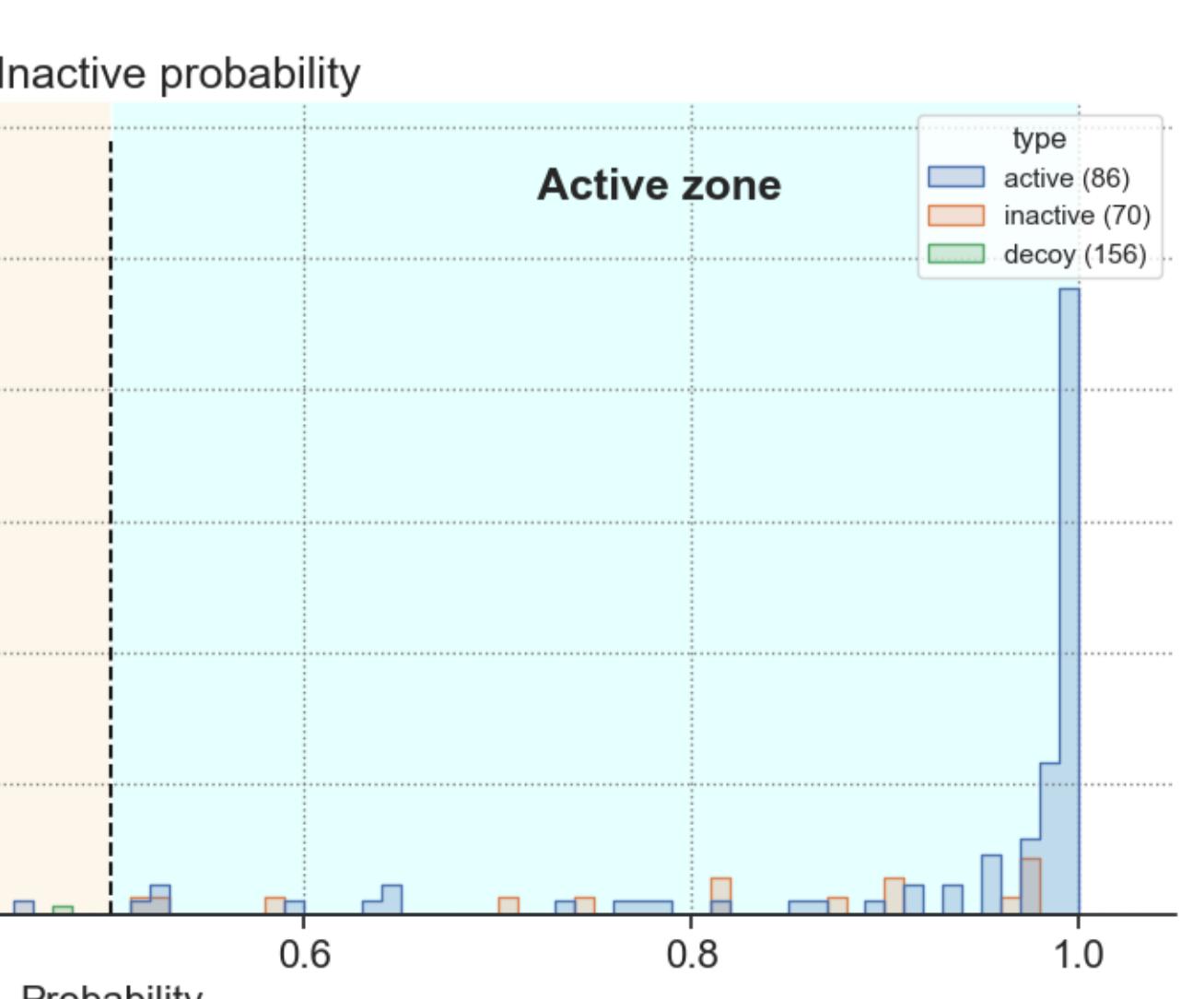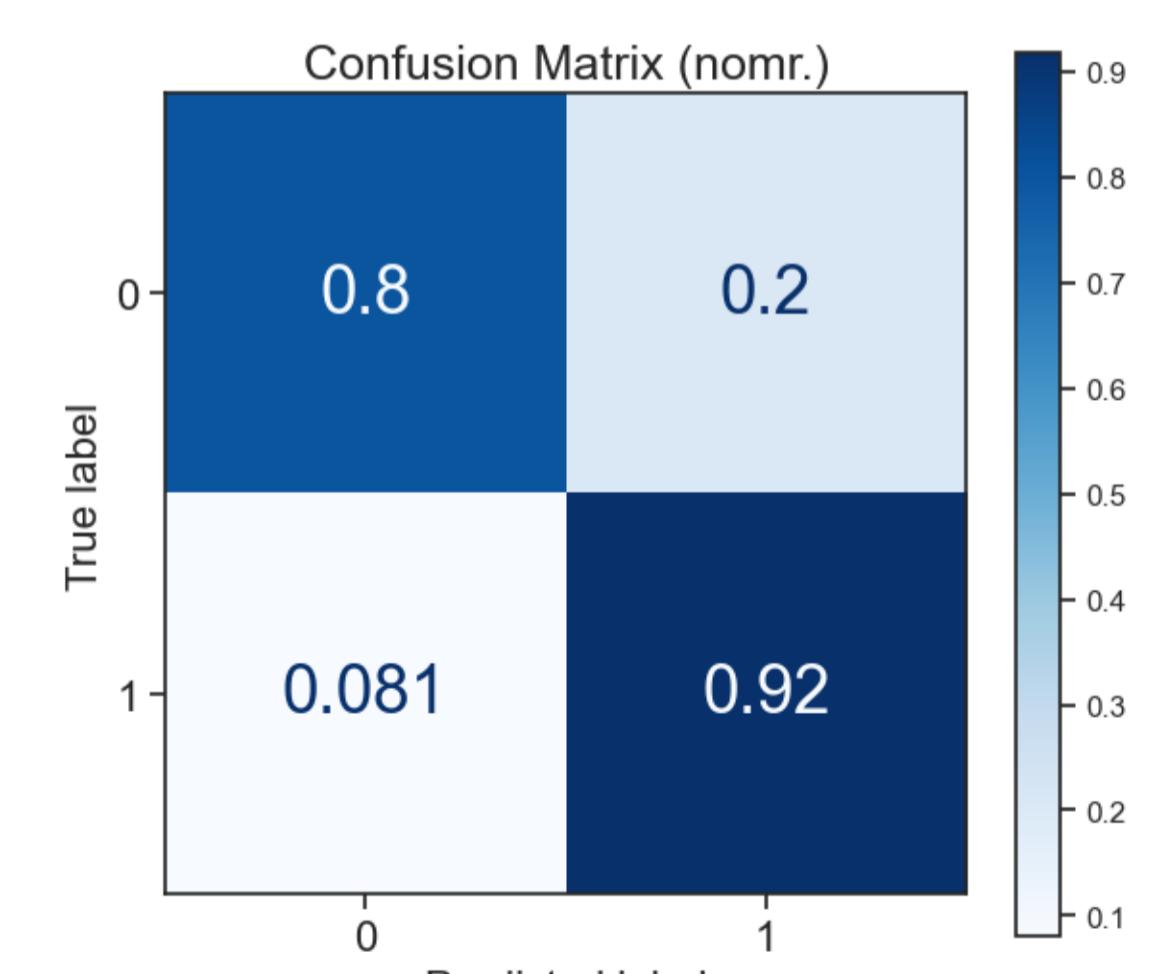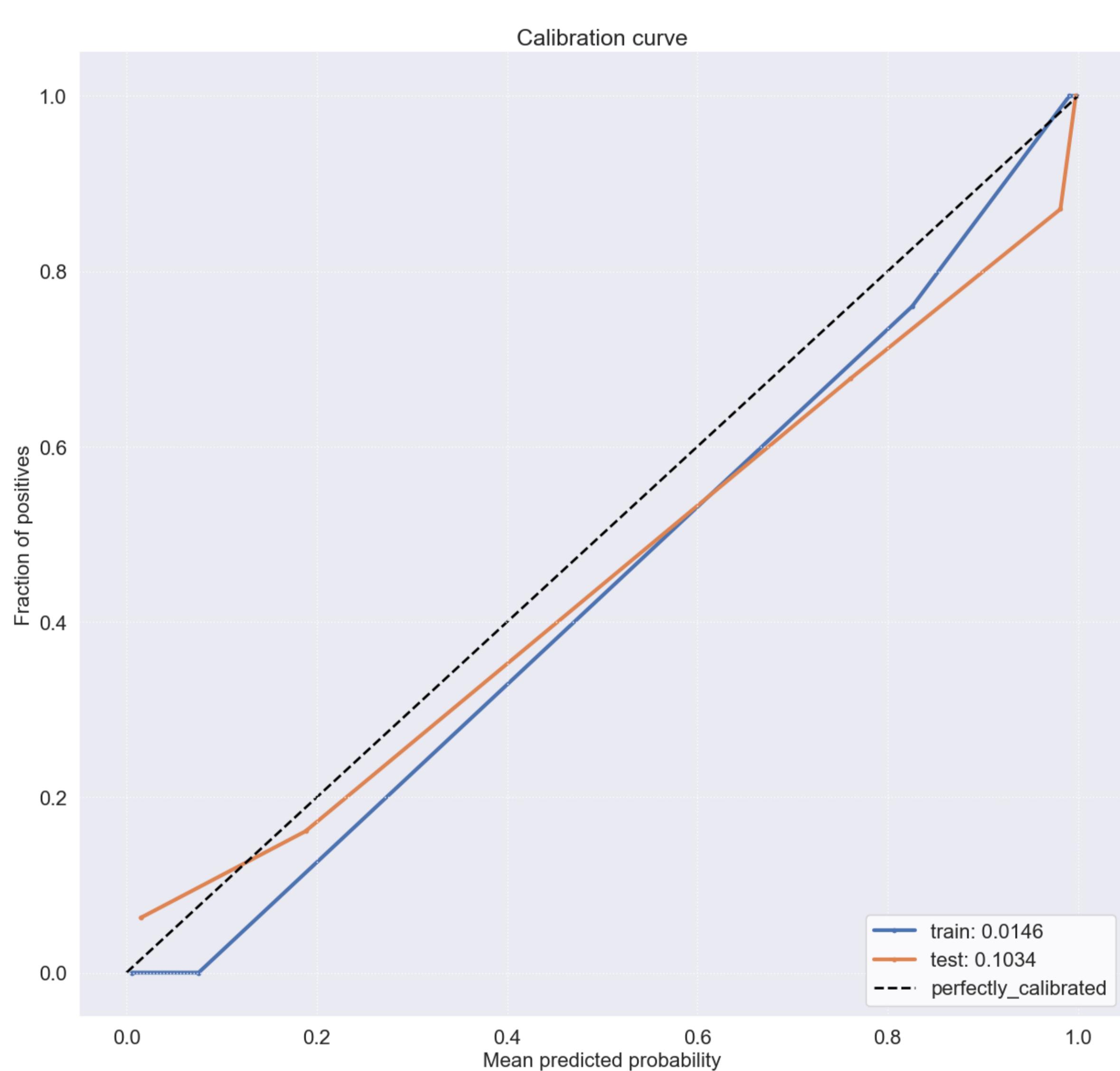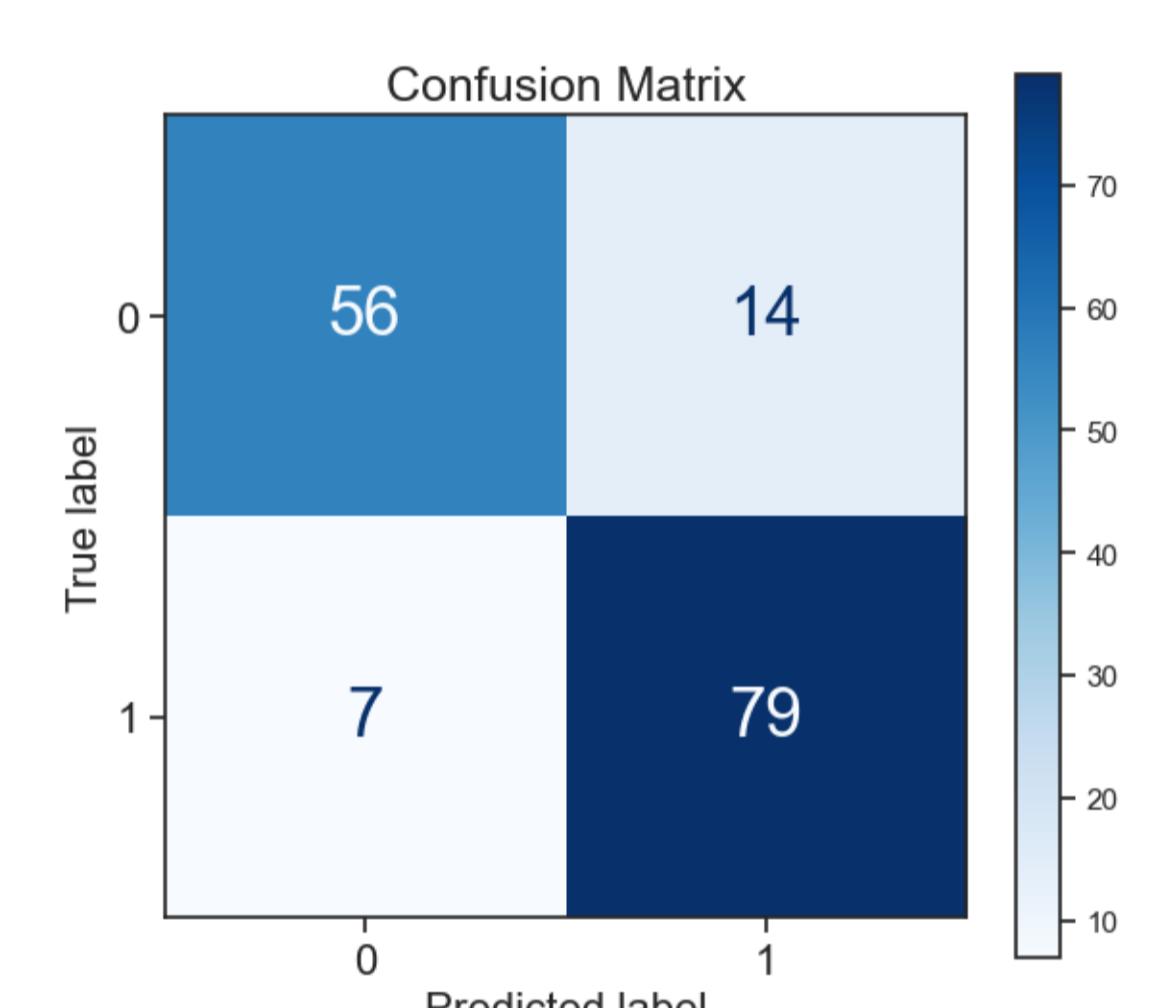
Metric	Train	Test
AUC	0.9961	0.9352
accuracy	0.9663	0.8718
balanced accuracy	0.9625	0.8691
recall_(sens)	1.0	0.8953
specificity	0.925	0.8429
precision	0.9425	0.875
f1_score	0.9662	0.8716

Protein: P48039 - metric: f1

Total (train): 624

ROC curve

- AUC_train = 1.000
- AUC_test = 0.940
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 1834)
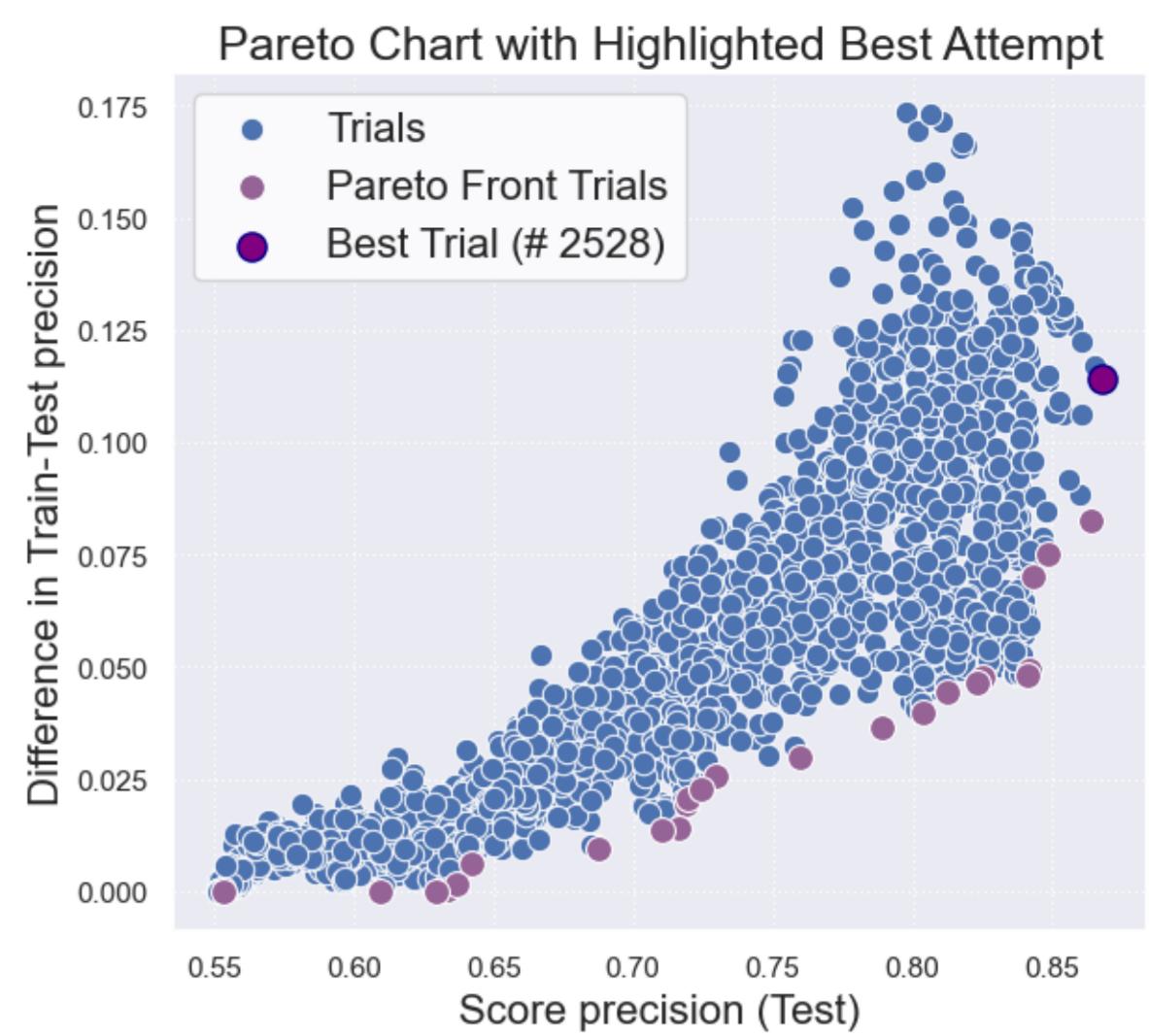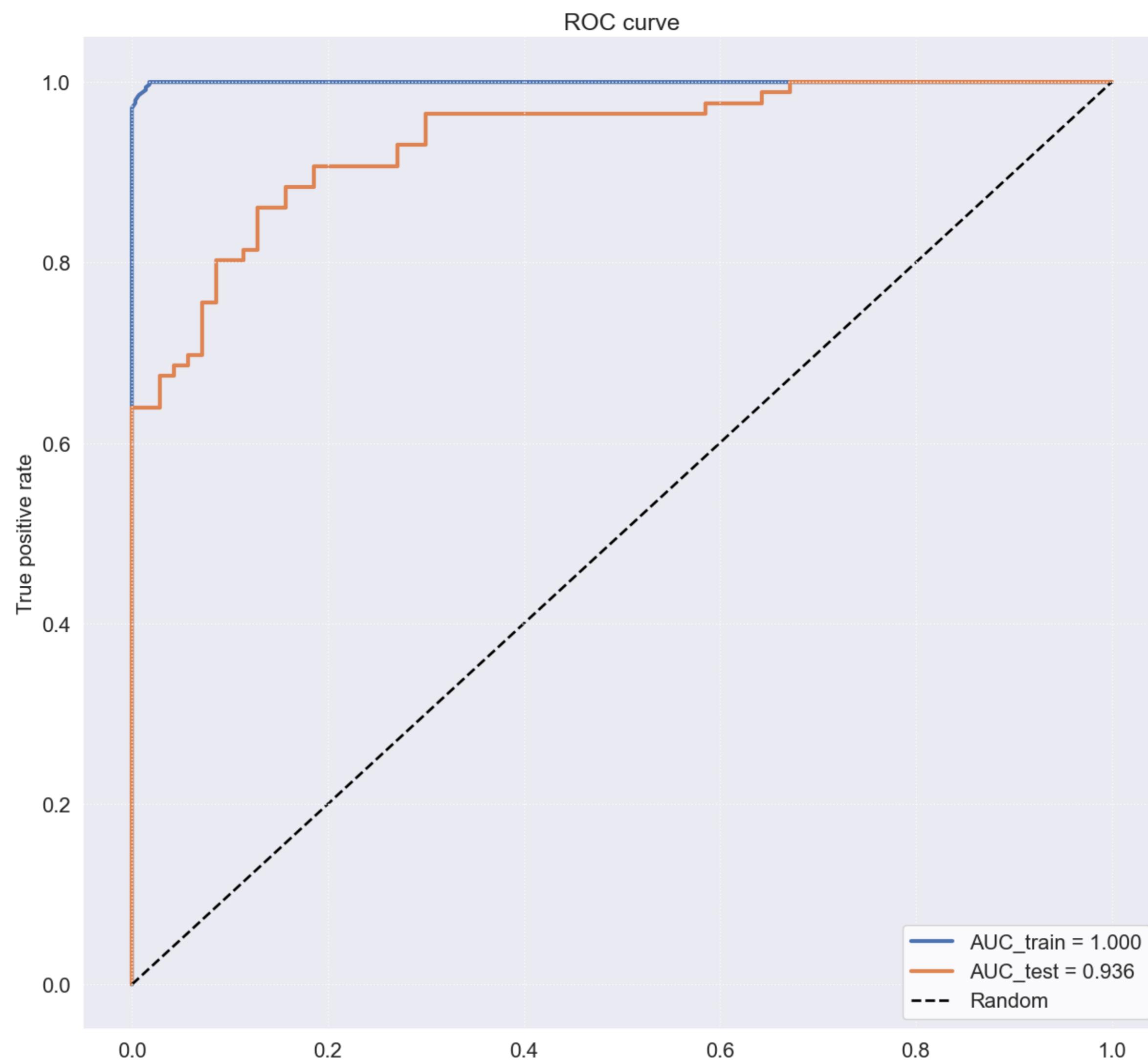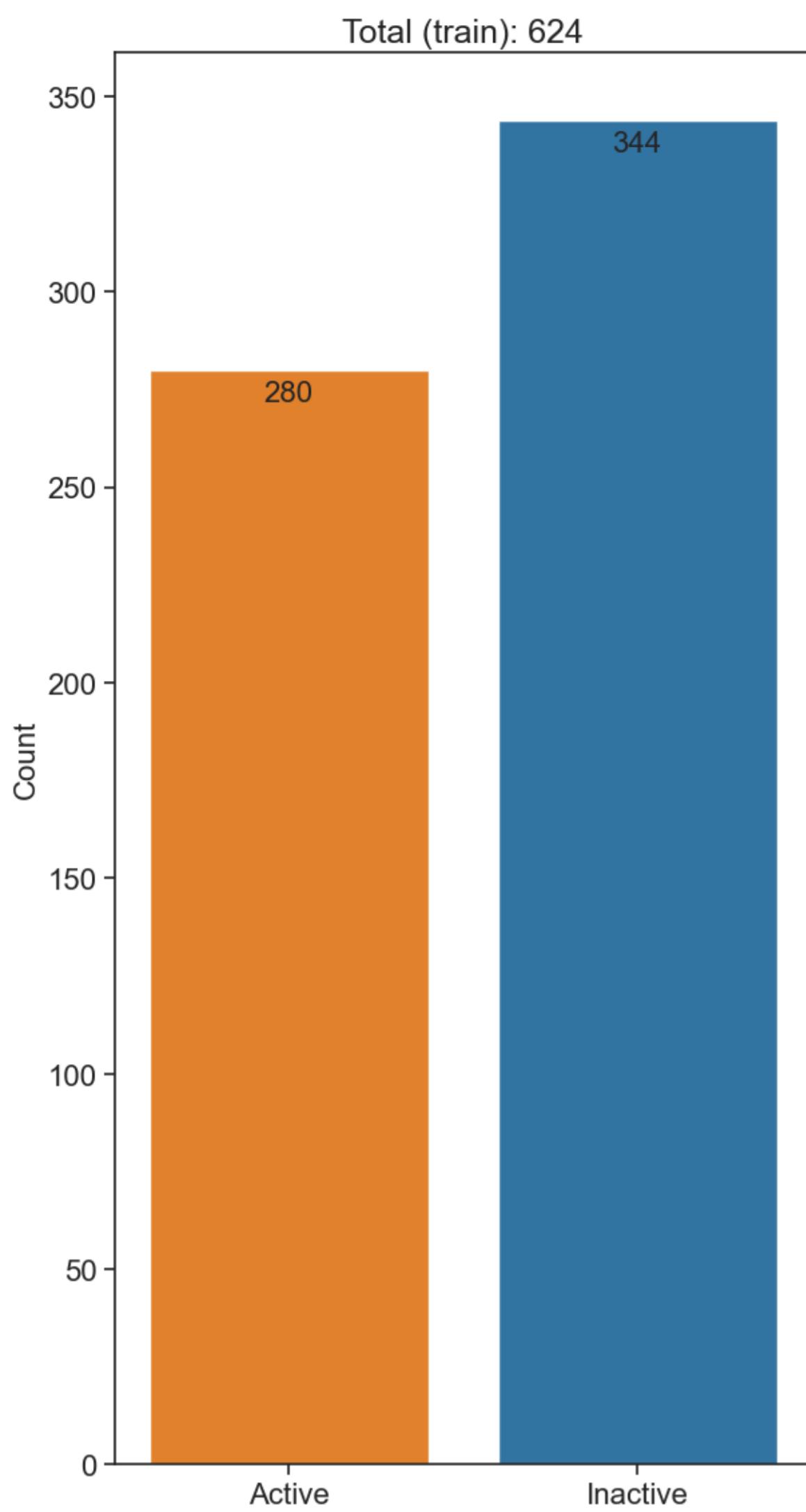
XGBoost Parameters	Value
alpha	0.29
colsample_bytree	0.9376
gamma	0.1099
lambda	3.1744
learning_rate	0.1354
max_depth	9.0
min_child_weight	1.0
n_estimators	414.0
reg_alpha	0.8461
reg_lambda	0.0783
scale_pos_weight	3.6378
subsample	0.8629

Confusion Matrix

	0	1
0	56	14
1	7	79

Confusion Matrix (nomr.)

	0	1
0	0.8	0.2
1	0.081	0.92

Calibration curve

- train: 0.0146
- test: 0.1034
- perfectly_calibrated

Active/Inactive probability

- Inactive zone
- Active zone
- type
  - active (86)
  - inactive (70)
  - decoy (156)
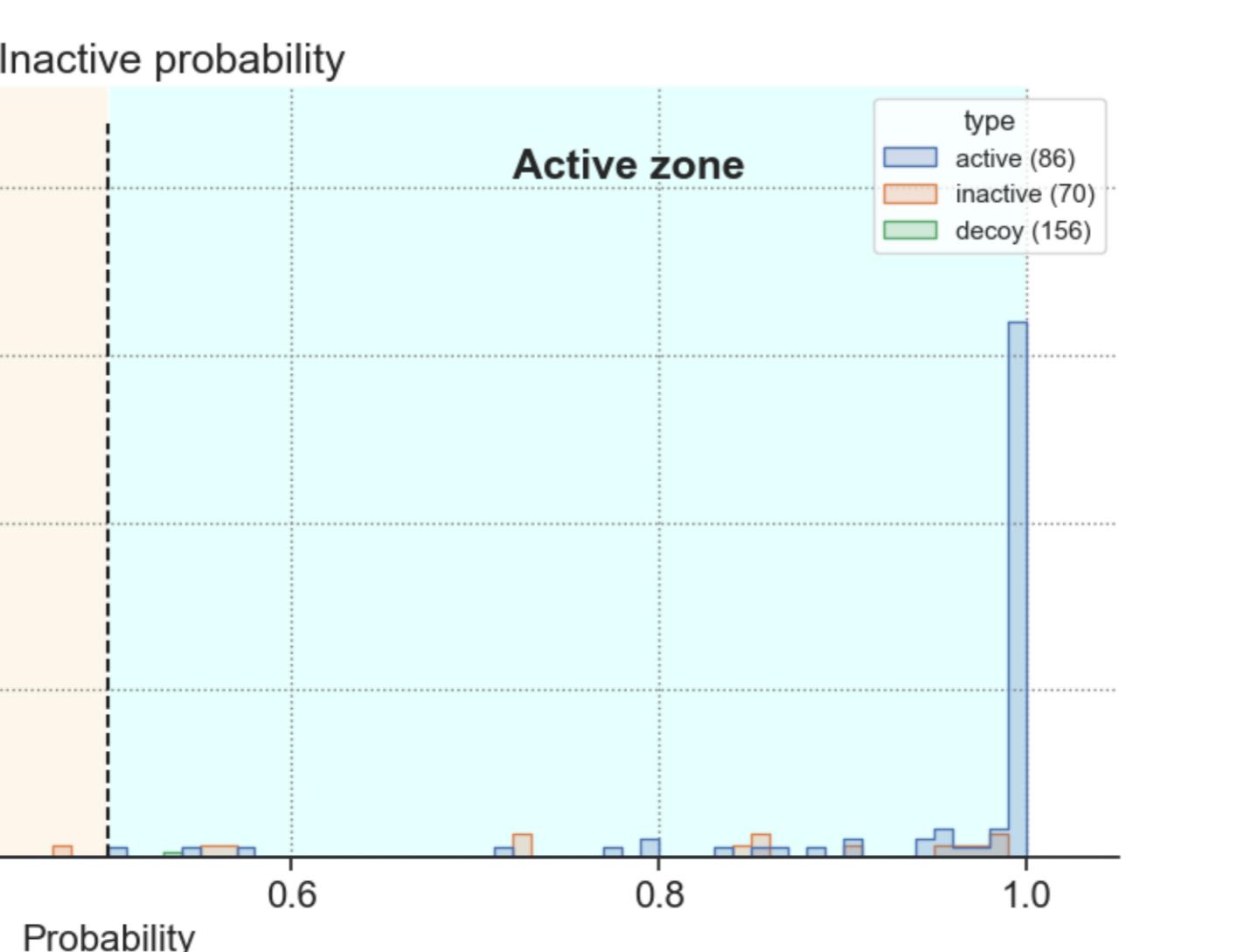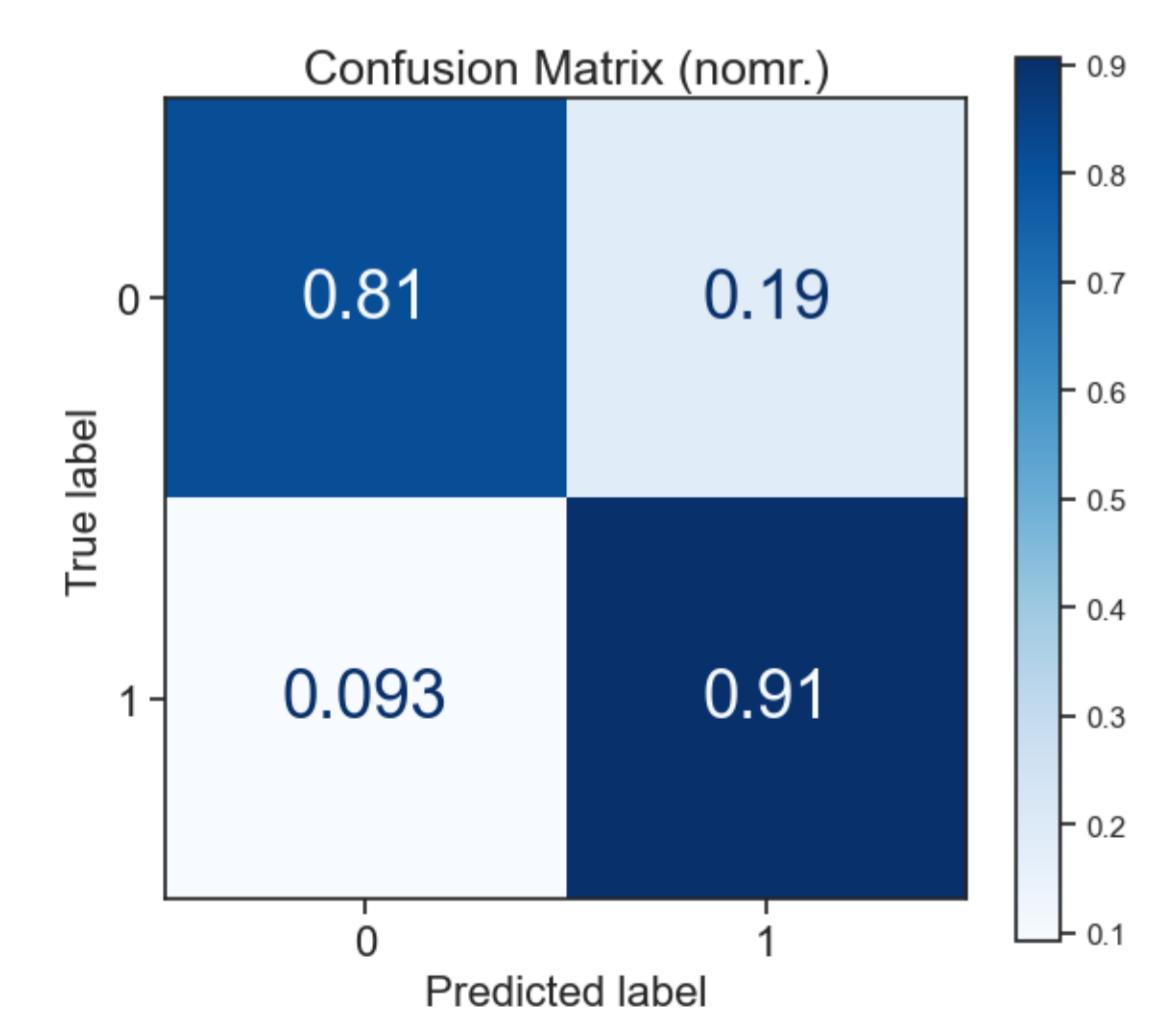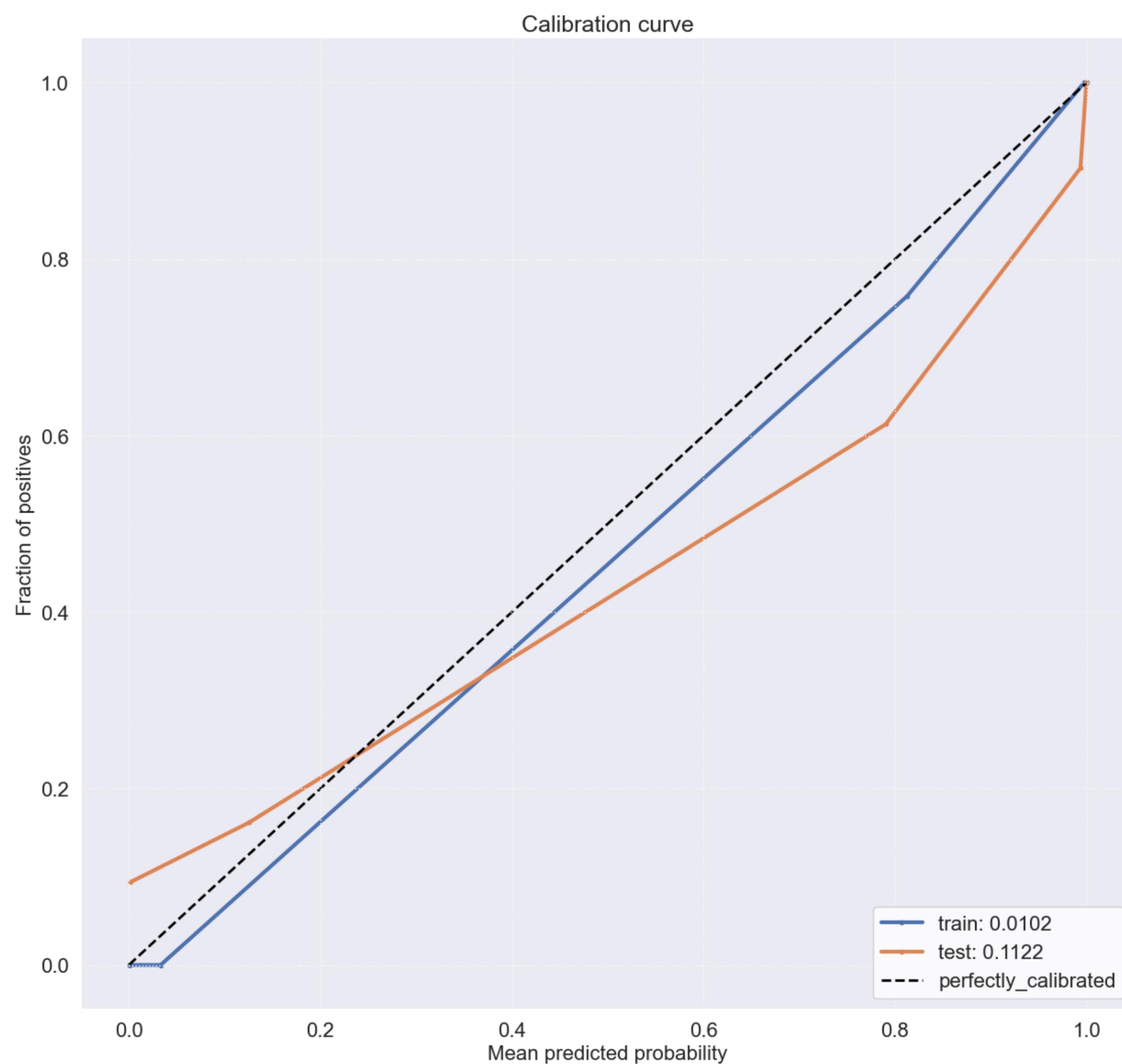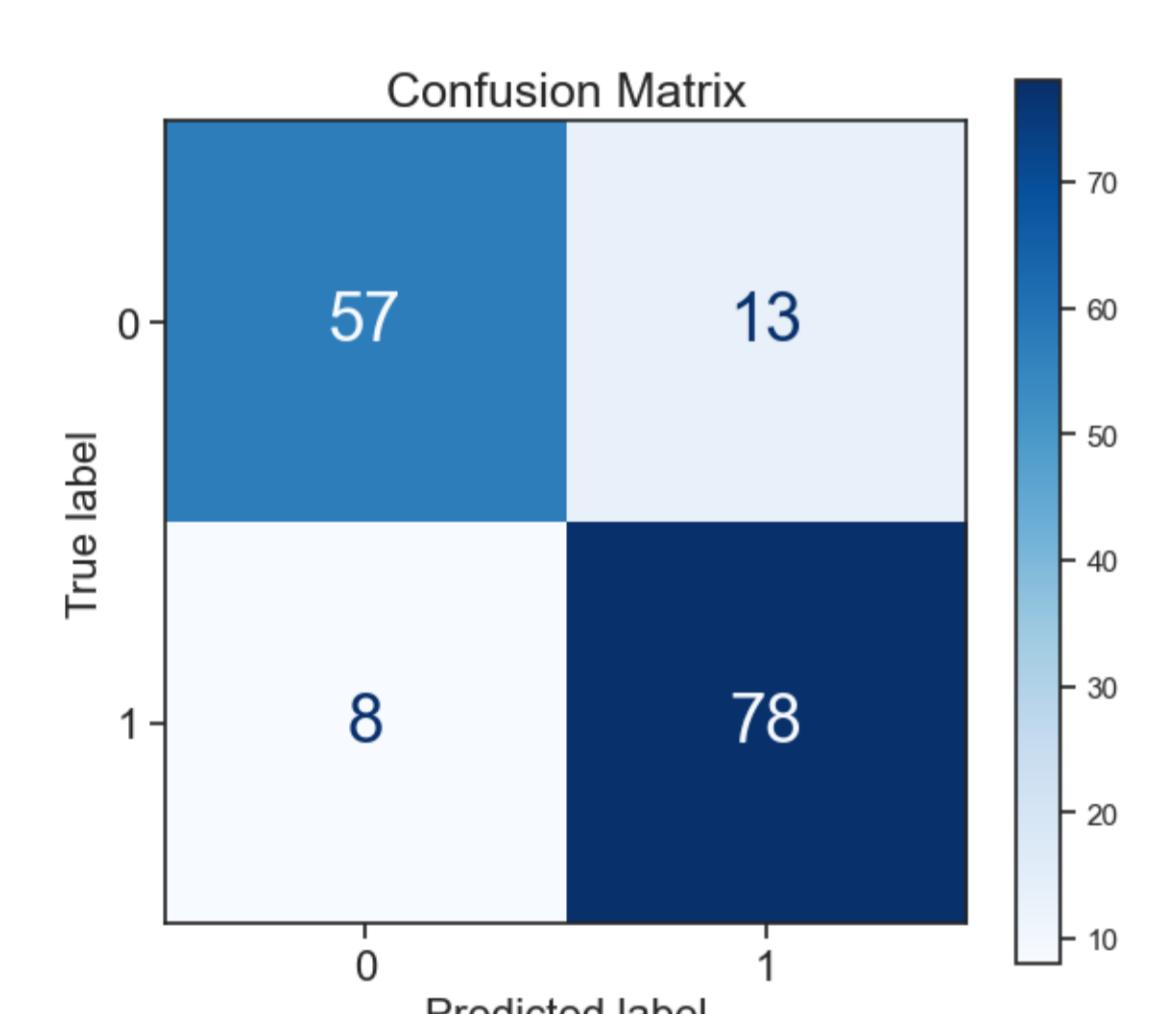
Metric	Train	Test
AUC	0.9996	0.9397
accuracy	0.9856	0.8654
balanced accuracy	0.9839	0.8593
recall_(sens)	1.0	0.9186
specificity	0.9679	0.8
precision	0.9745	0.8495
f1_score	0.9856	0.8645

Protein: P48039 - metric: precision

Total (train): 624

ROC curve

AUC_train = 1.000
AUC_test = 0.936
Random

Pareto Chart with Highlighted Best Attempt

Trials
Pareto Front Trials
Best Trial (# 2528)
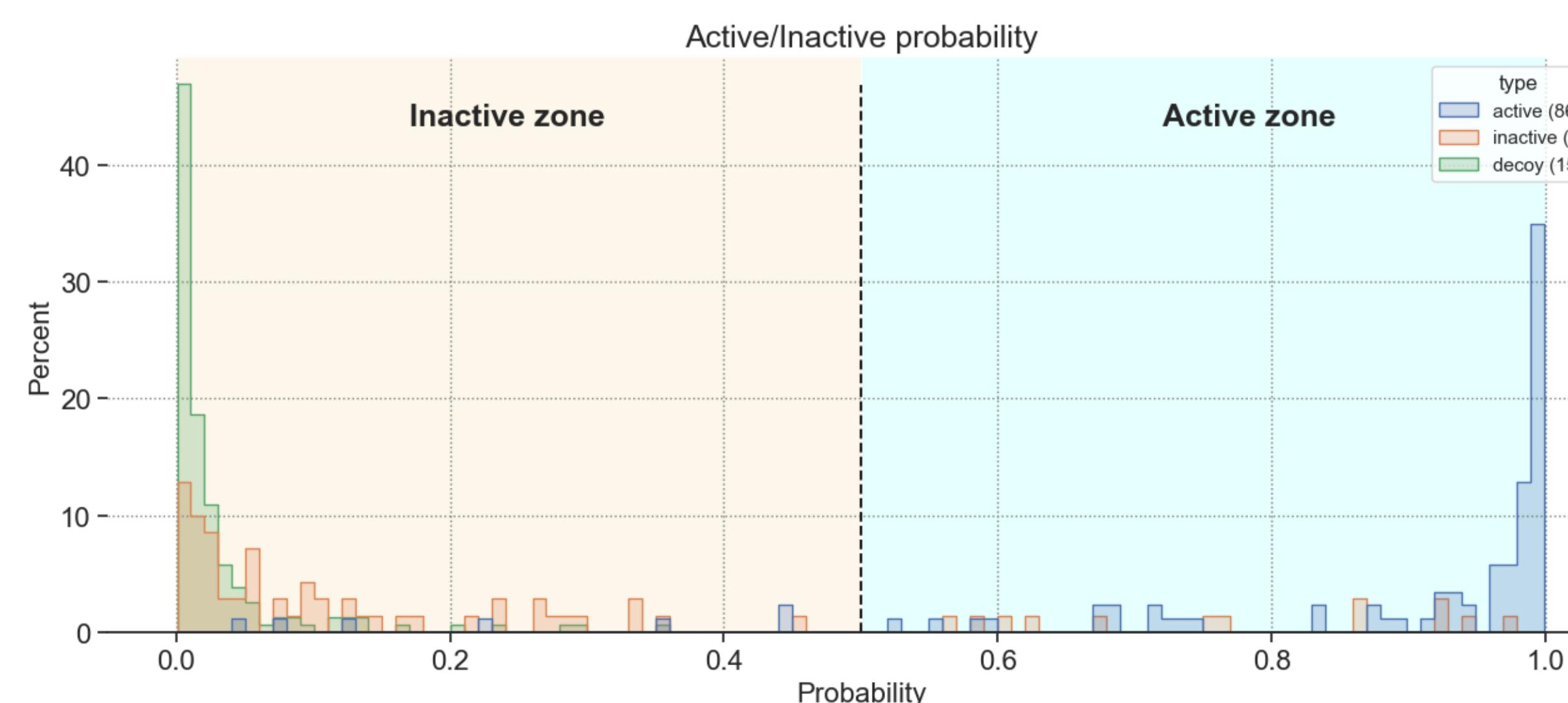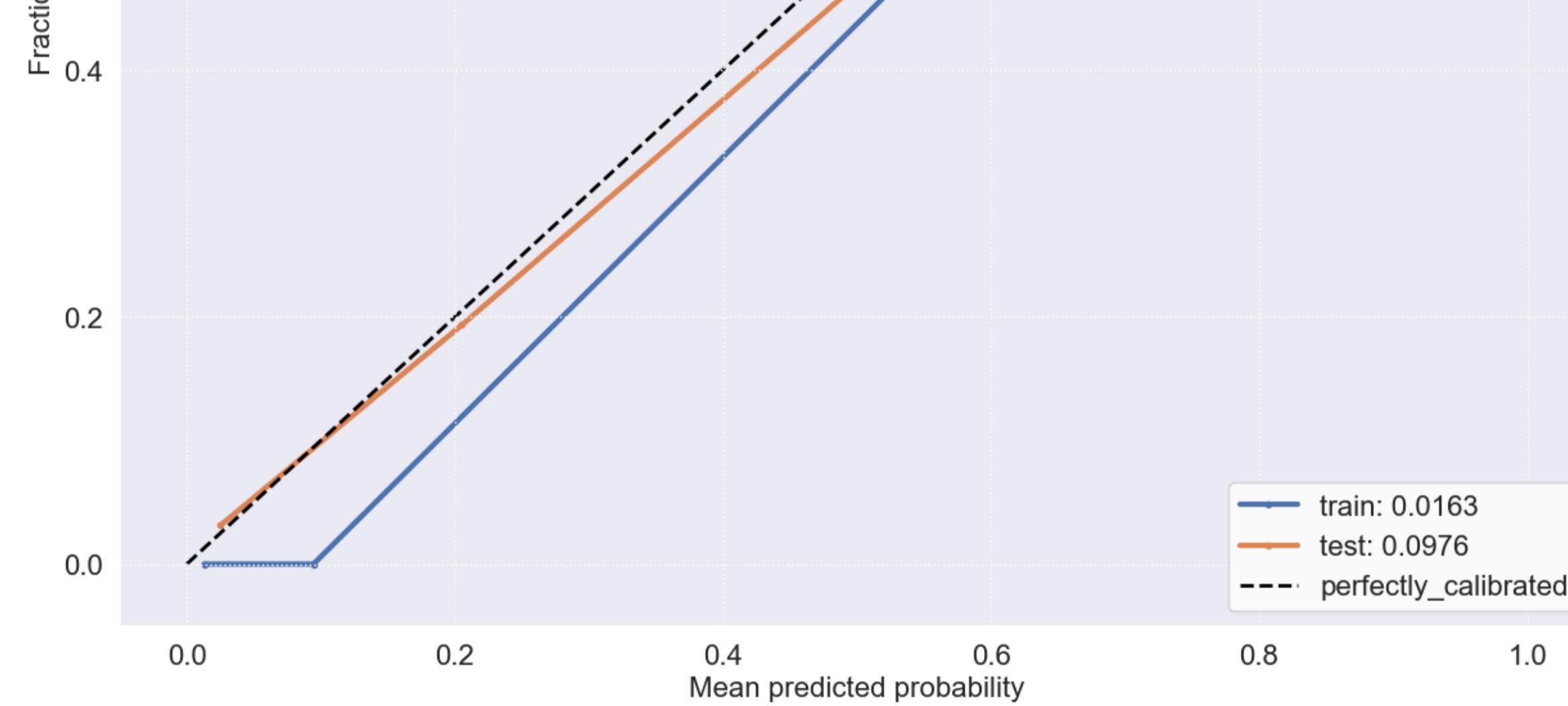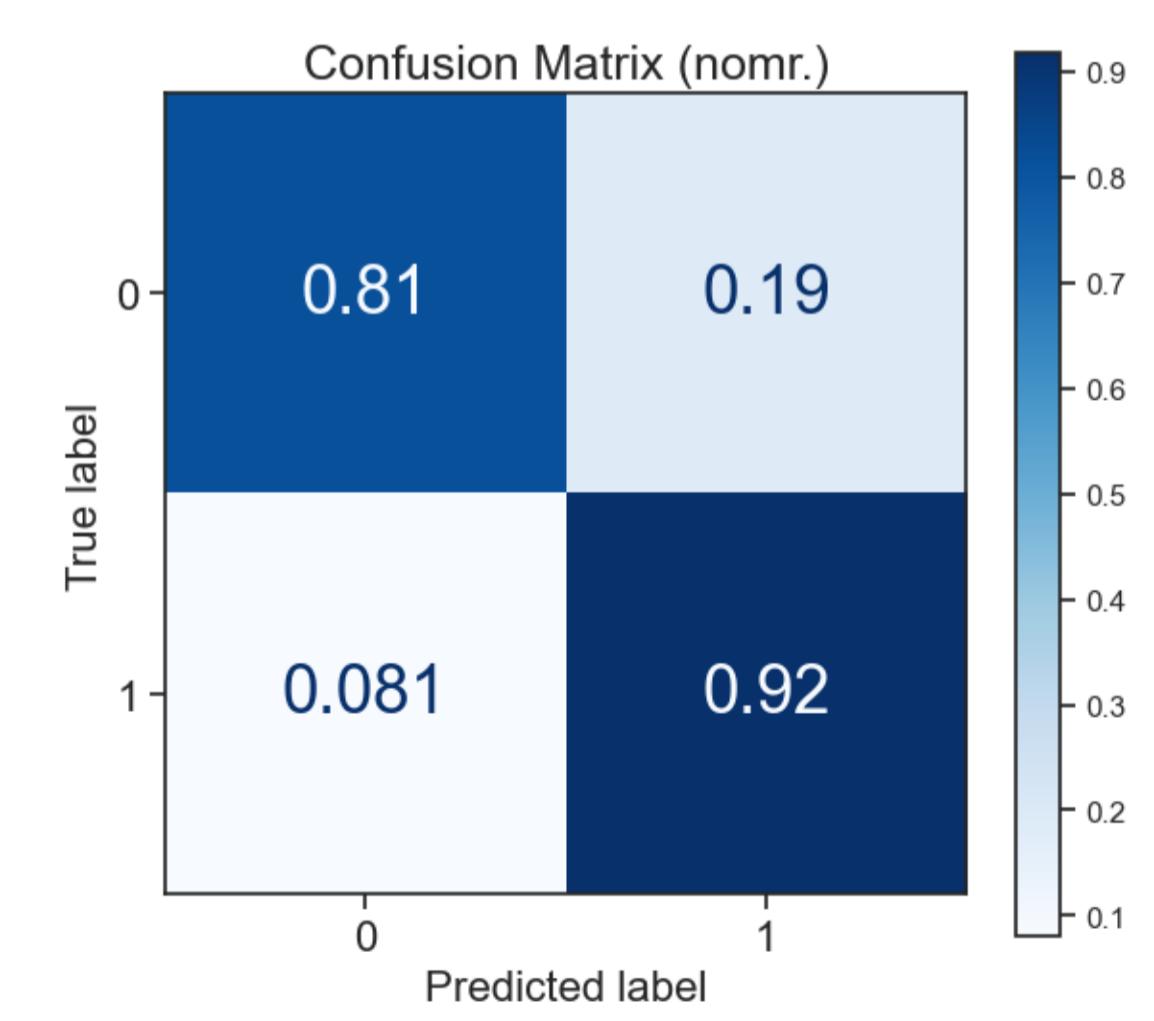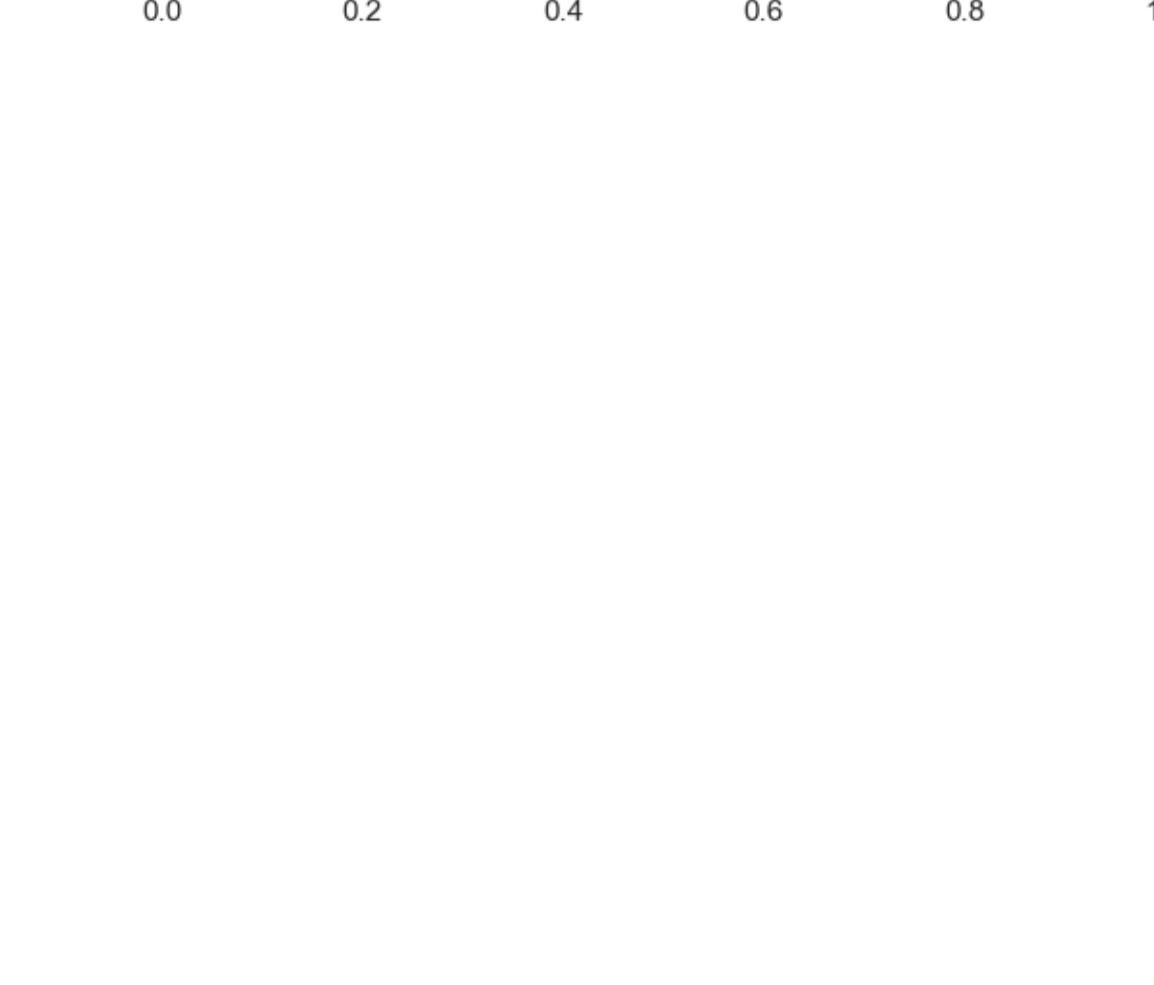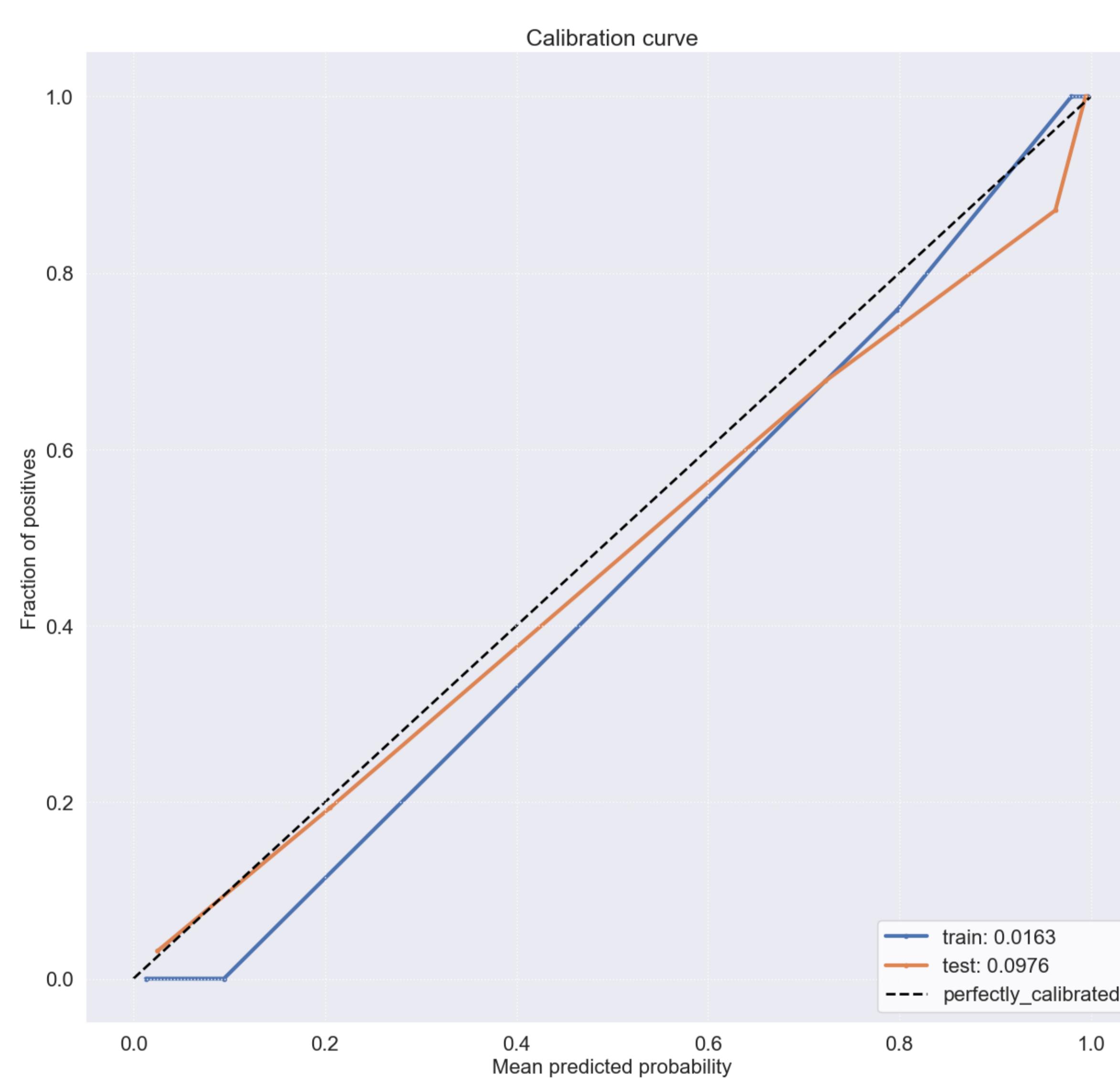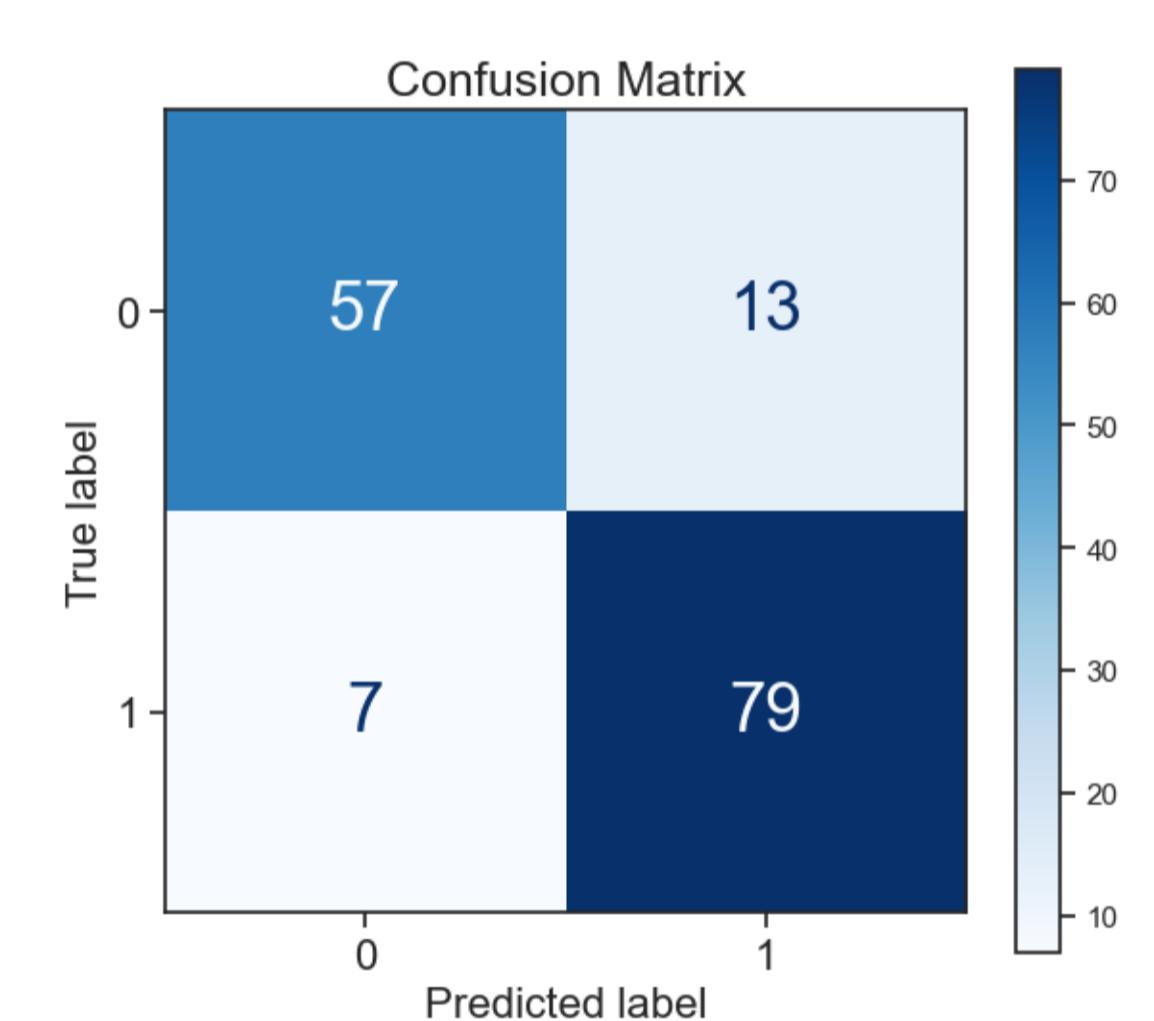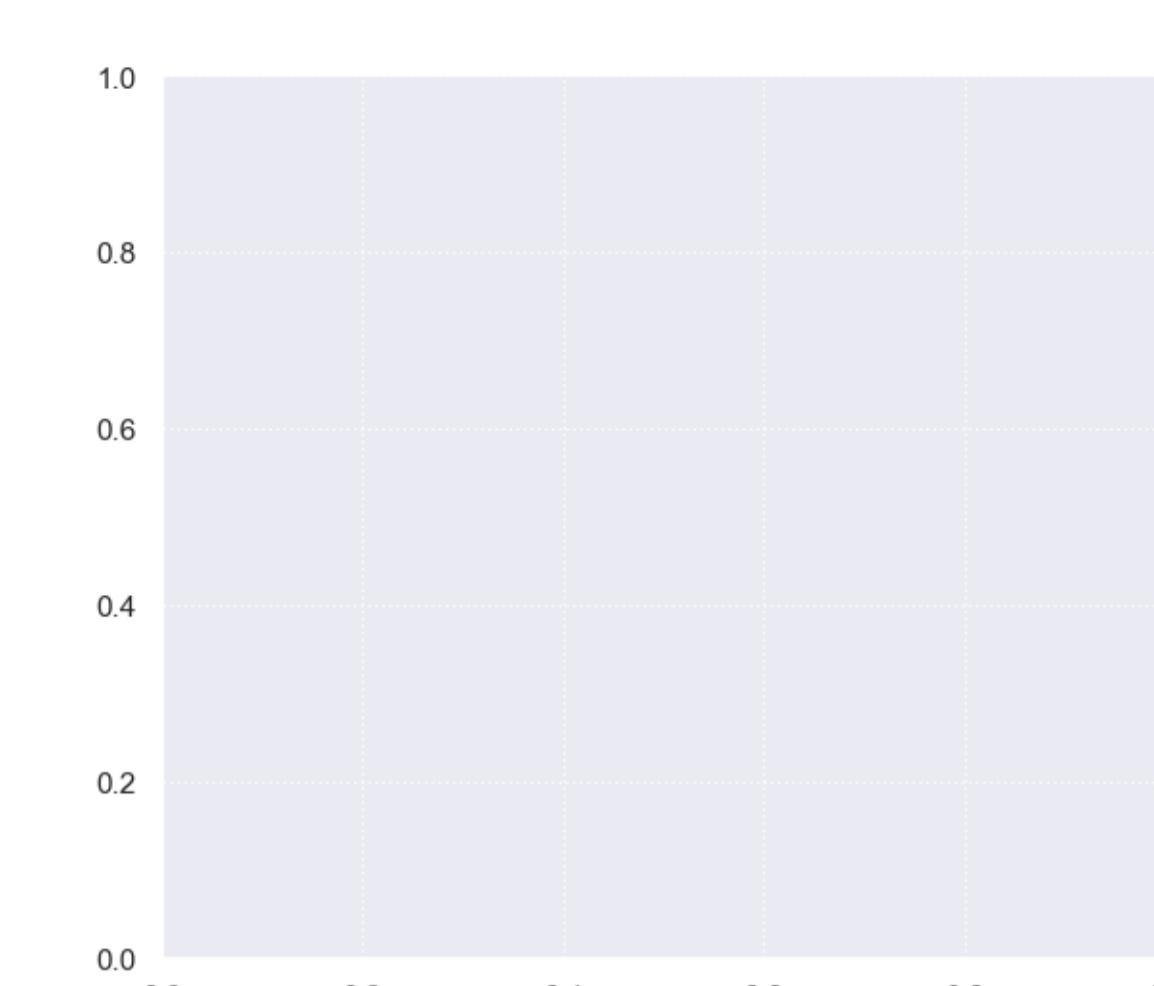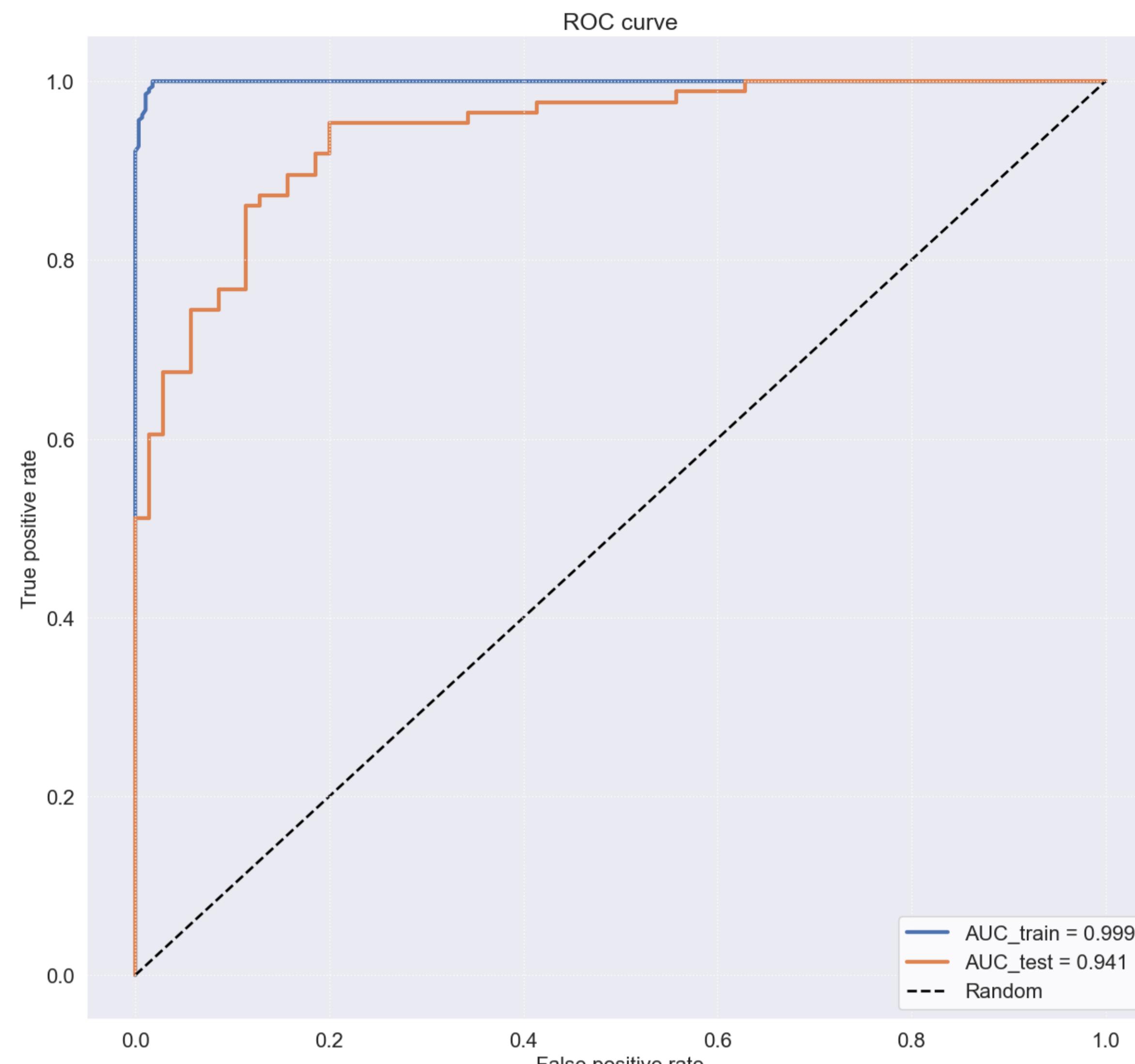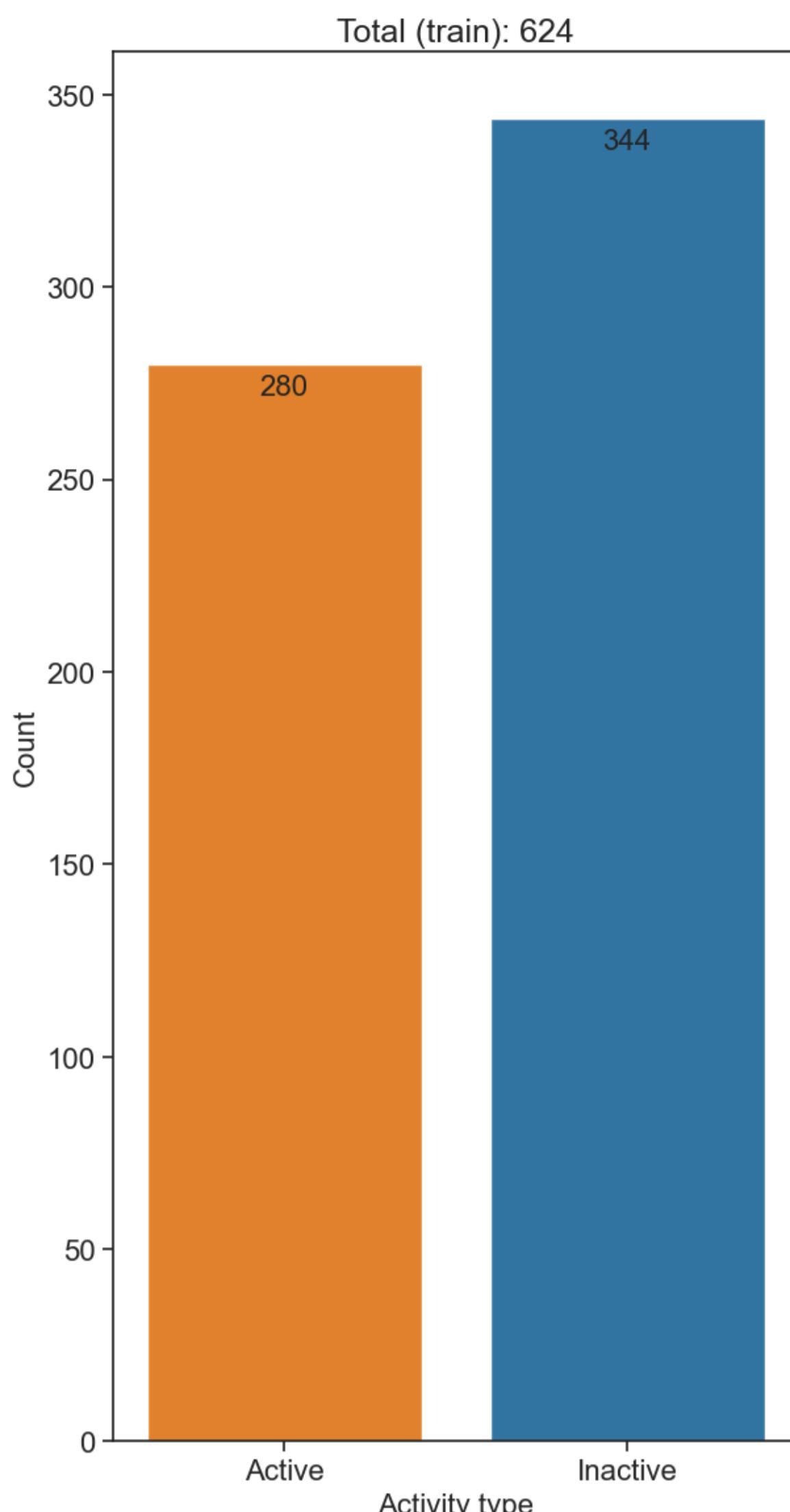
XGBoost Parameters	Value
alpha	3.8401
colsample_bytree	0.8441
gamma	0.438
lambda	1.9722
learning_rate	0.3497
max_depth	7.0
min_child_weight	1.0
n_estimators	488.0
reg_alpha	0.1234
reg_lambda	0.4879
scale_pos_weight	3.9827
subsample	0.5224

Confusion Matrix

57	13	
8	78	

Confusion Matrix (nomr.)

0.81	0.19	
0.093	0.91	

Calibration curve

train: 0.0102
test: 0.1122
perfectly_calibrated

Active/Inactive probability

Inactive zone   Active zone

type
active (86)
inactive (70)
decoy (156)

Metric	Train	Test
AUC	0.9997	0.9355
accuracy	0.9888	0.8654
balanced accuracy	0.9875	0.8606
recall_(sens)	1.0	0.907
specificity	0.975	0.8143
precision	0.9801	0.8571
f1_score	0.9888	0.8648

Protein: P48039 - metric: SoftVotingXGBoostEnsemble

Metric	Train	Test
AUC	0.9994	0.941
accuracy	0.984	0.8718
balanced accuracy	0.9821	0.8664
recall_(sens)	1.0	0.9186
specificity	0.9643	0.8143
precision	0.9718	0.8587
f1_score	0.9839	0.8711