Protein: P28335 - metric: balanced_accuracy

Total (train): 1788

Active: 553
Inactive: 1235

ROC curve

AUC_train = 1.000
AUC_test = 0.924
Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
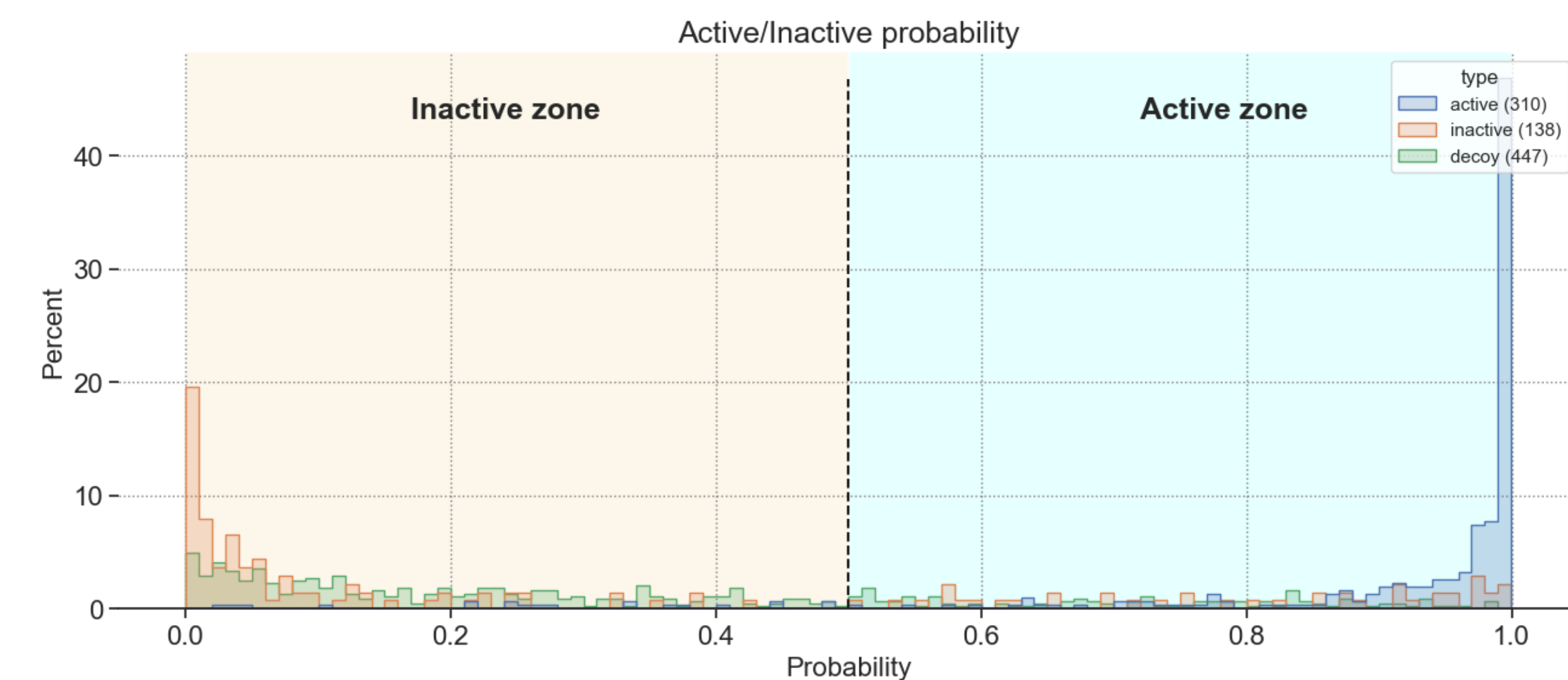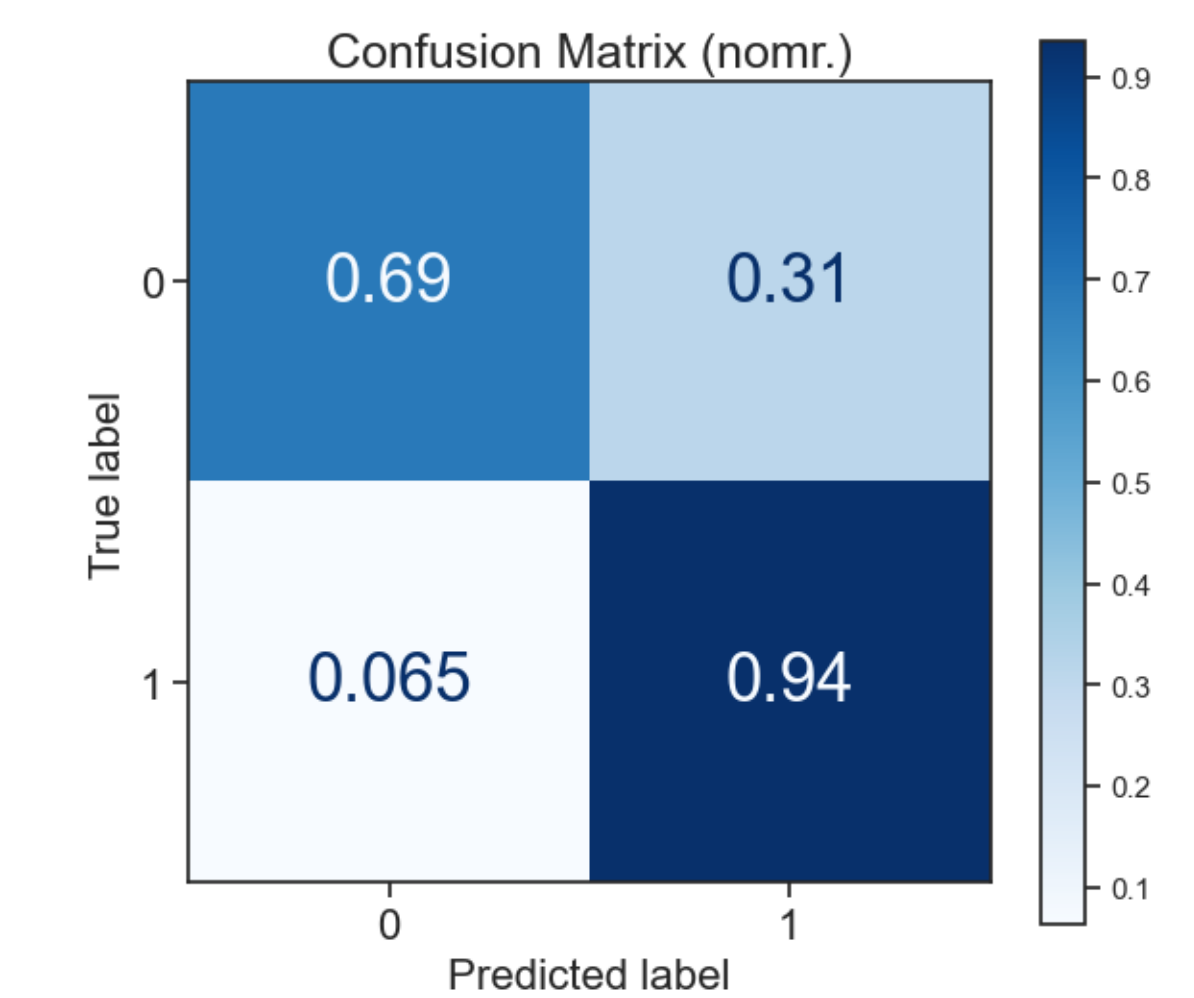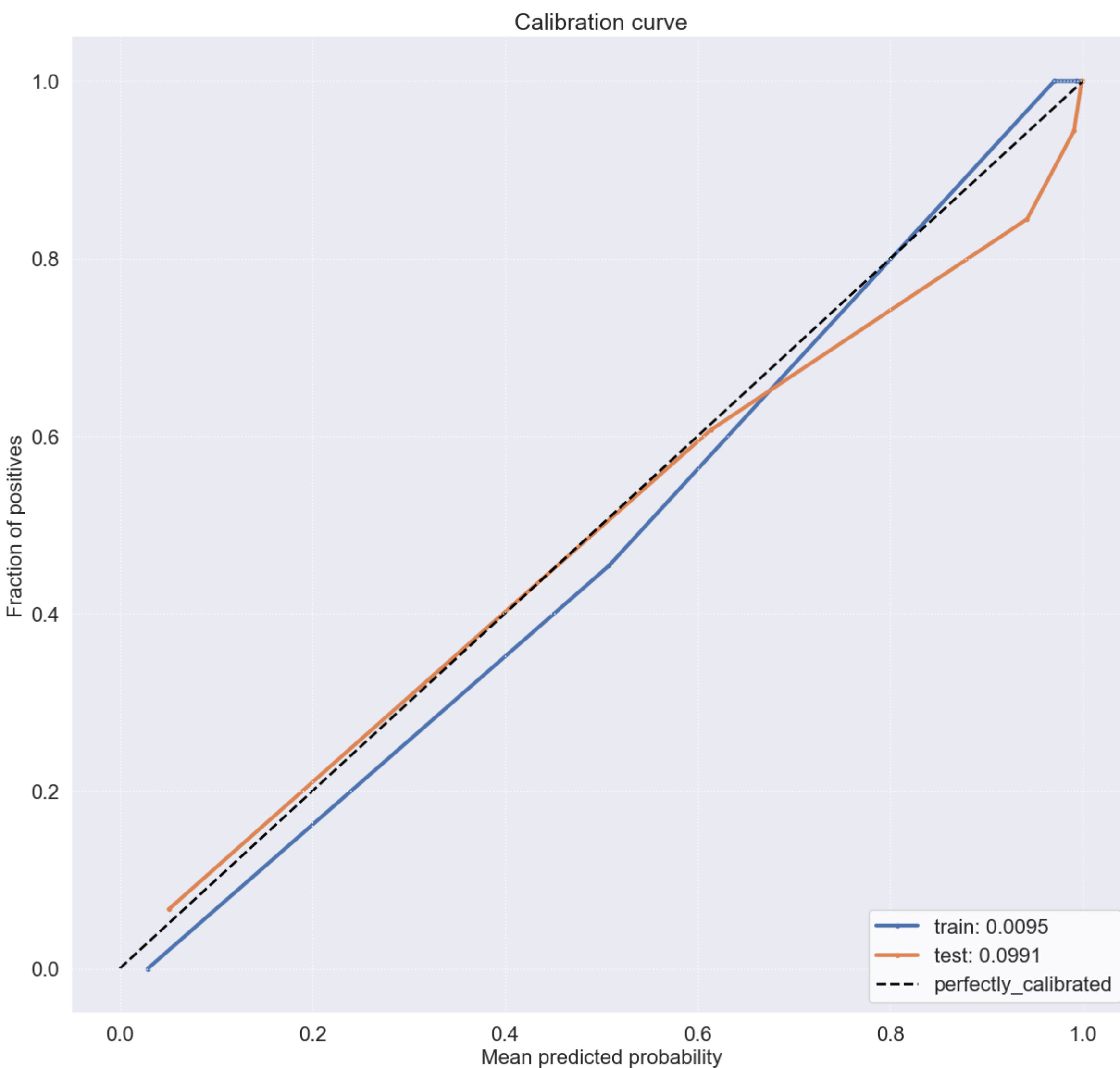- Best Trial (# 578)

X-axis: Score balanced_accuracy (Test)
Y-axis: Difference in Train-Test balanced_accuracy
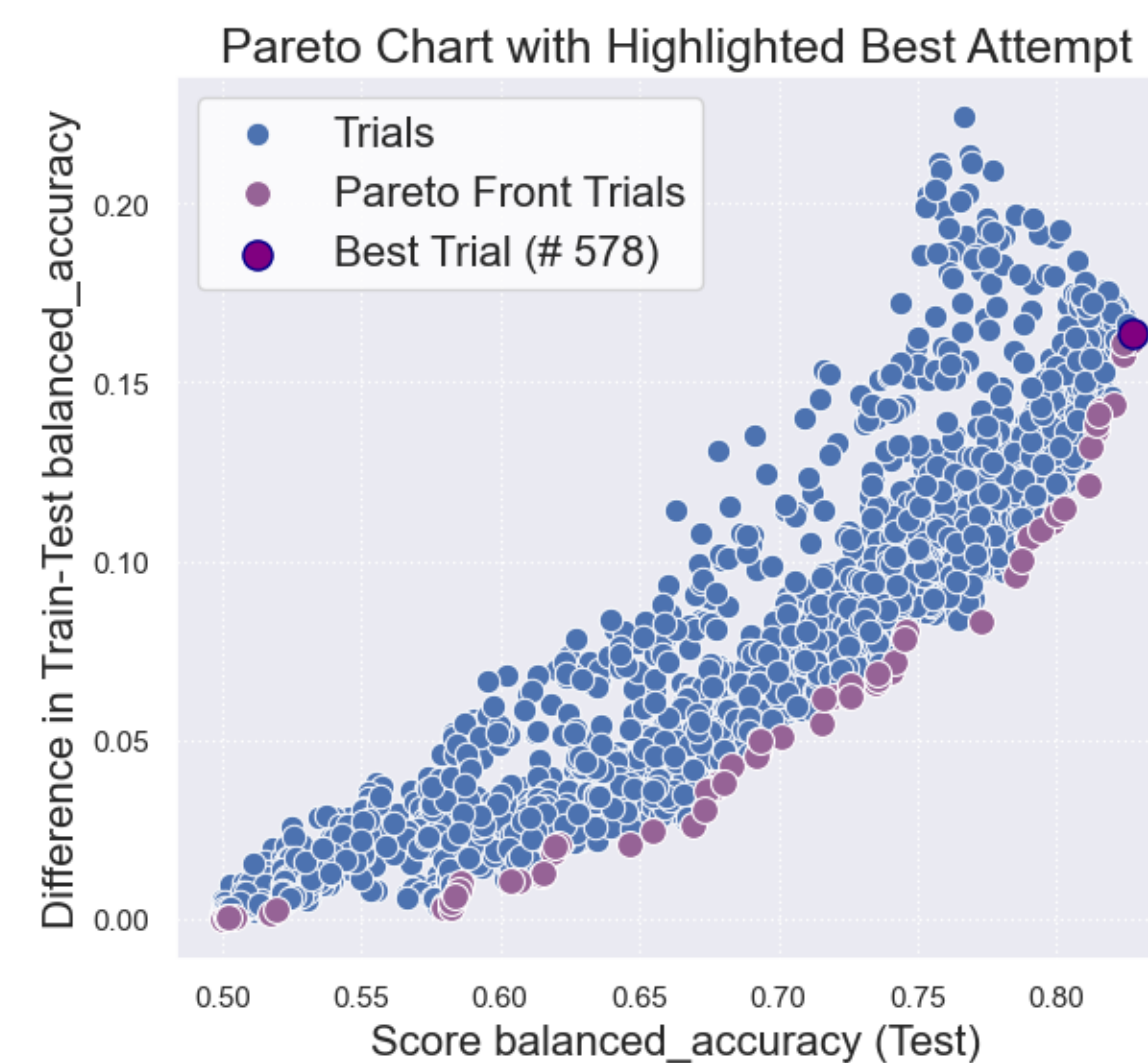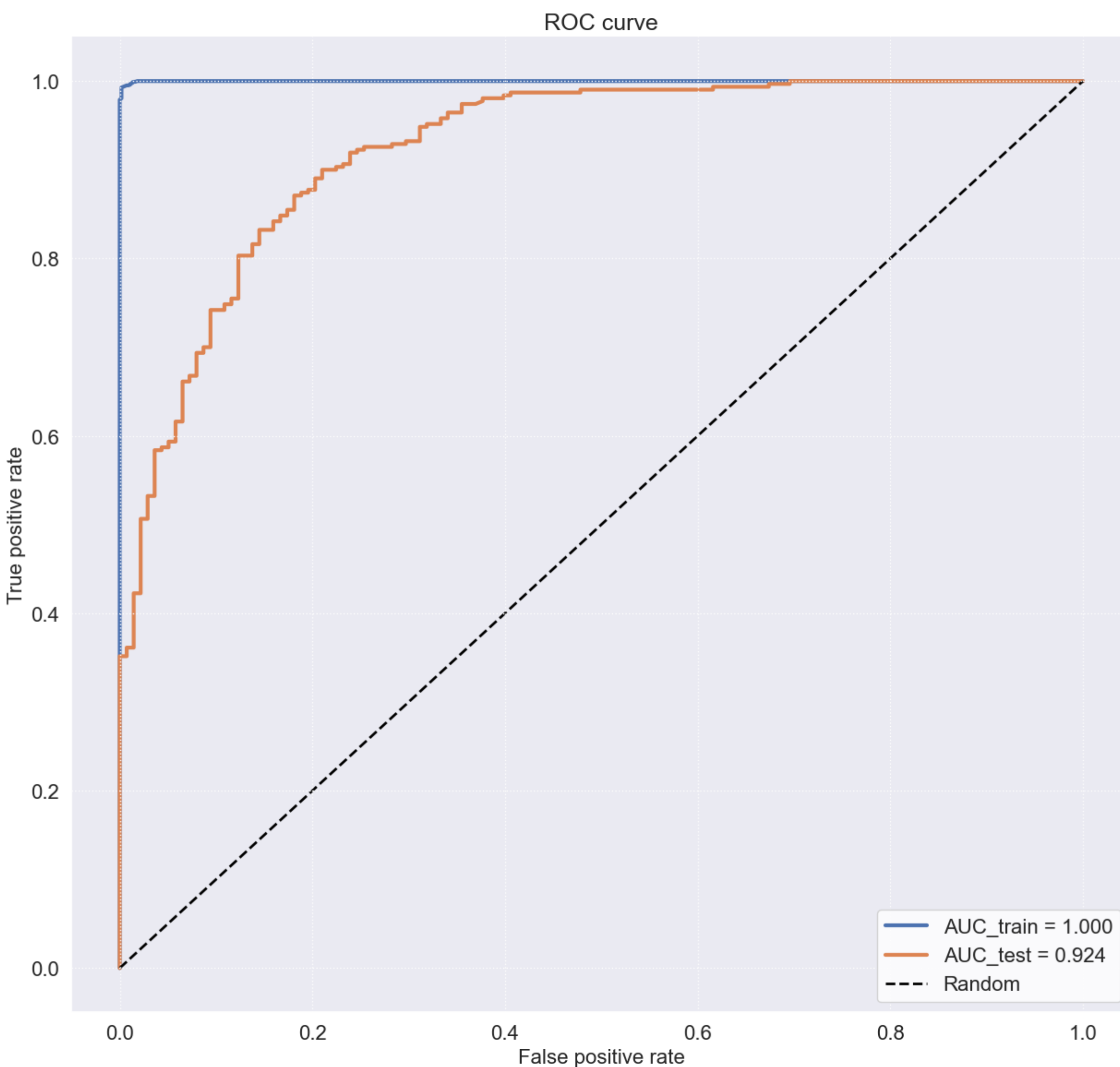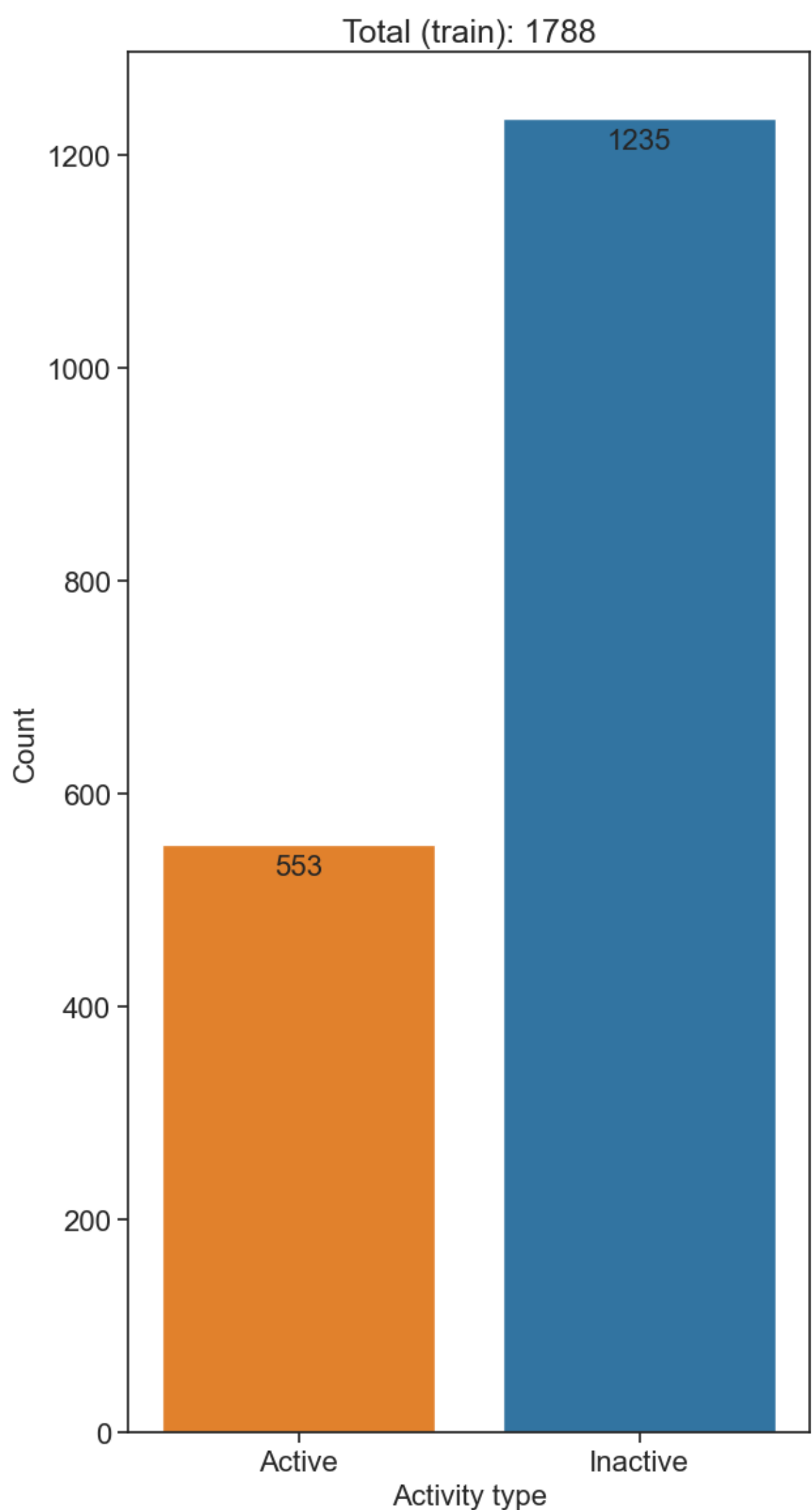
XGBoost Parameters	Value
alpha	4.0852
colsample_bytree	0.6915
gamma	0.3595
lambda	2.0192
learning_rate	0.0923
max_depth	9.0
min_child_weight	1.0
n_estimators	942.0
reg_alpha	0.4933
reg_lambda	0.3234
scale_pos_weight	1.4468
subsample	0.5909

Confusion Matrix

	0	1
0	95	43
1	20	290

Confusion Matrix (nomr.)

	0	1
0	0.69	0.31
1	0.065	0.94

Calibration curve

train: 0.0095
test: 0.0991
perfectly_calibrated

Active/Inactive probability

Inactive zone
Active zone

type:
- active (310)
- inactive (138)
- decoy (447)
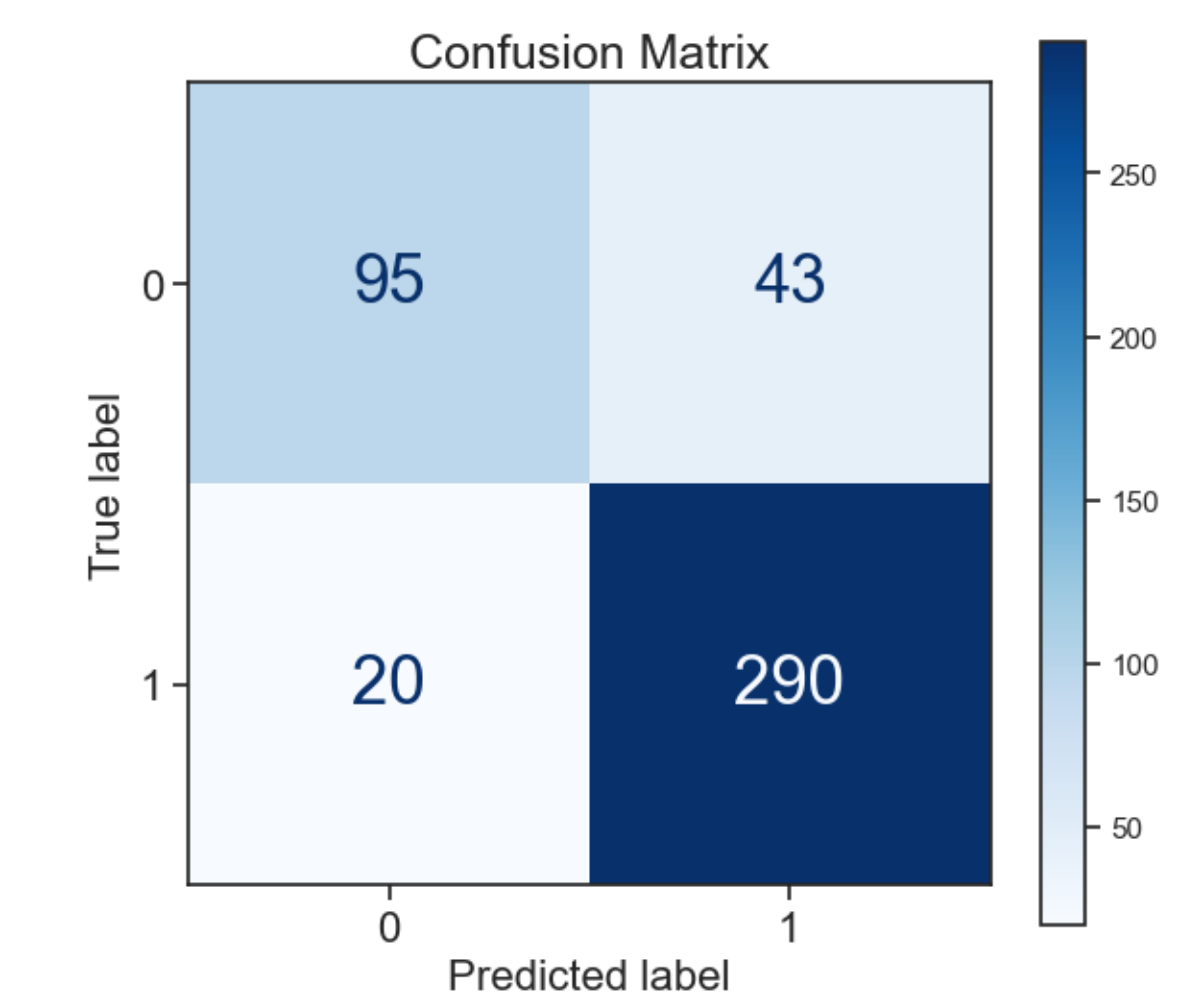
Metric	Train	Test
AUC	0.9999	0.9239
accuracy	0.9938	0.8594
balanced accuracy	0.9901	0.8119
recall_(sens)	1.0	0.9355
specificity	0.9801	0.6884
precision	0.9912	0.8709
f1_score	0.9938	0.8555

Protein: P28335 - metric: f1

Total (train): 1788

Activity type	Count
Active	553
Inactive	1235

ROC curve

AUC_train = 1.000
AUC_test = 0.919
Random

Pareto Chart with Highlighted Best Attempt

Trials
Pareto Front Trials
Best Trial (# 2252)

Confusion Matrix

	0	1
0	94	44
1	13	297

Confusion Matrix (nomr.)

	0	1
0	0.68	0.32
1	0.042	0.96

Calibration curve

train: 0.0116
test: 0.1031
perfectly_calibrated

Active/Inactive probability

Inactive zone — Active zone

type
active (310)
inactive (138)
decoy (447)

XGBoost Parameters	Value
alpha	2.6167
colsample_bytree	0.6624
gamma	0.4658
lambda	1.934
learning_rate	0.0529
max_depth	8.0
min_child_weight	1.0
n_estimators	960.0
reg_alpha	0.2427
reg_lambda	0.4197
scale_pos_weight	2.9851
subsample	0.526
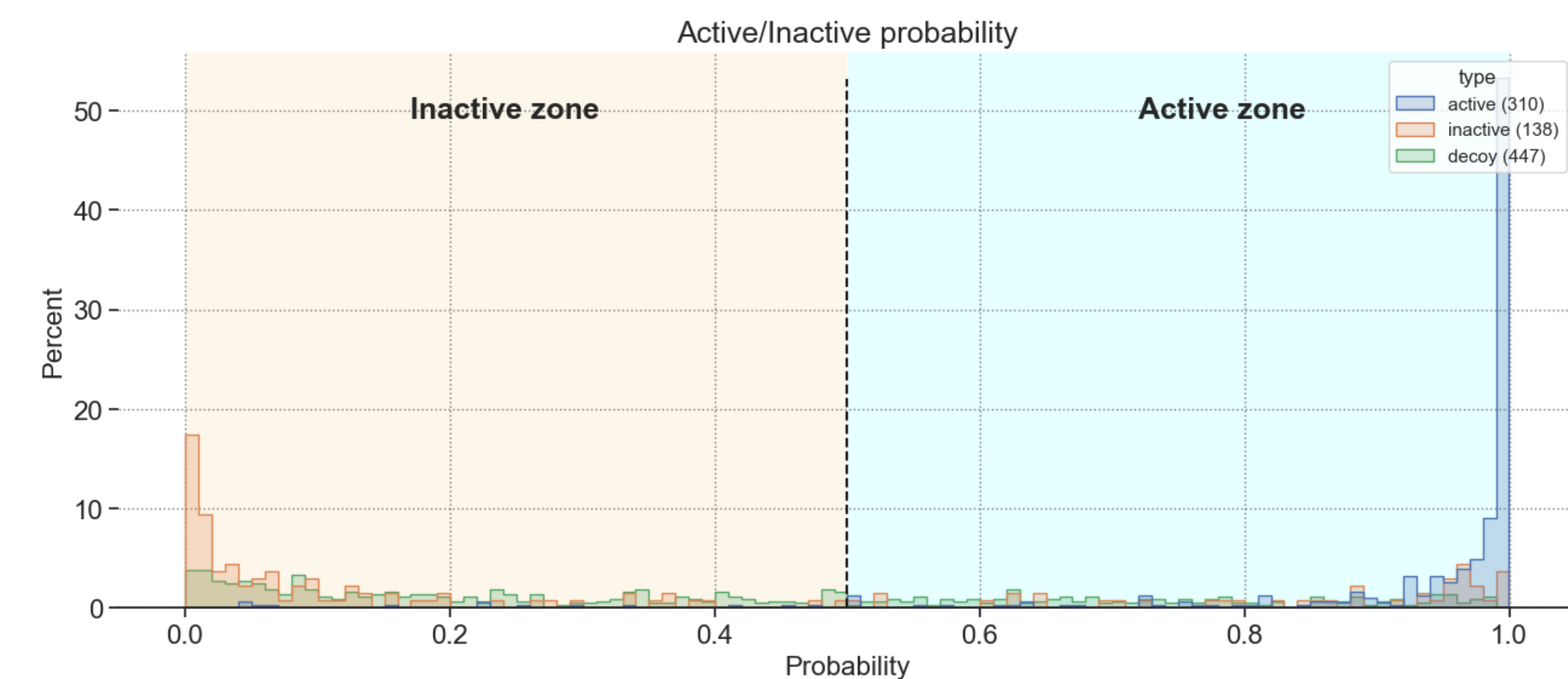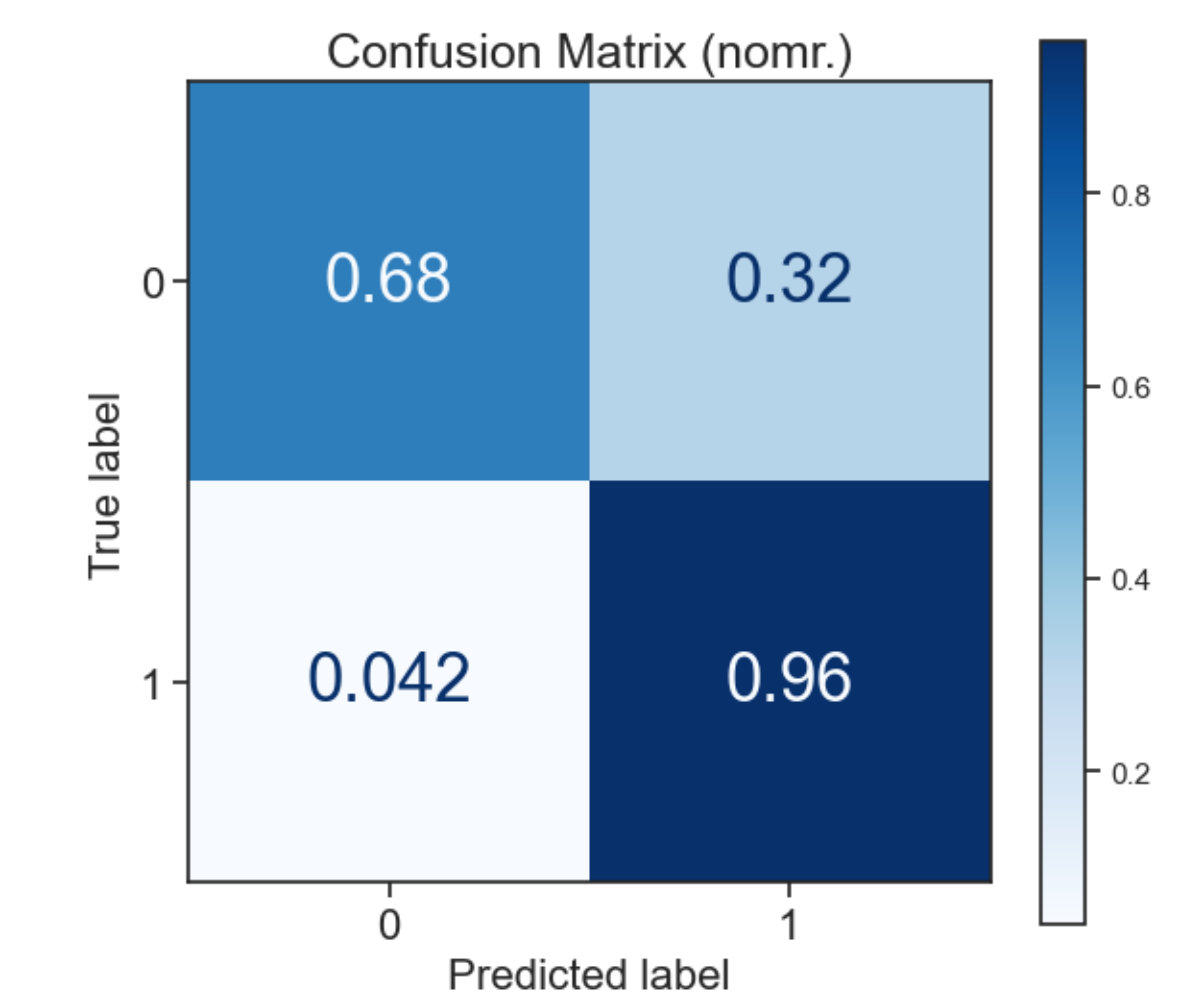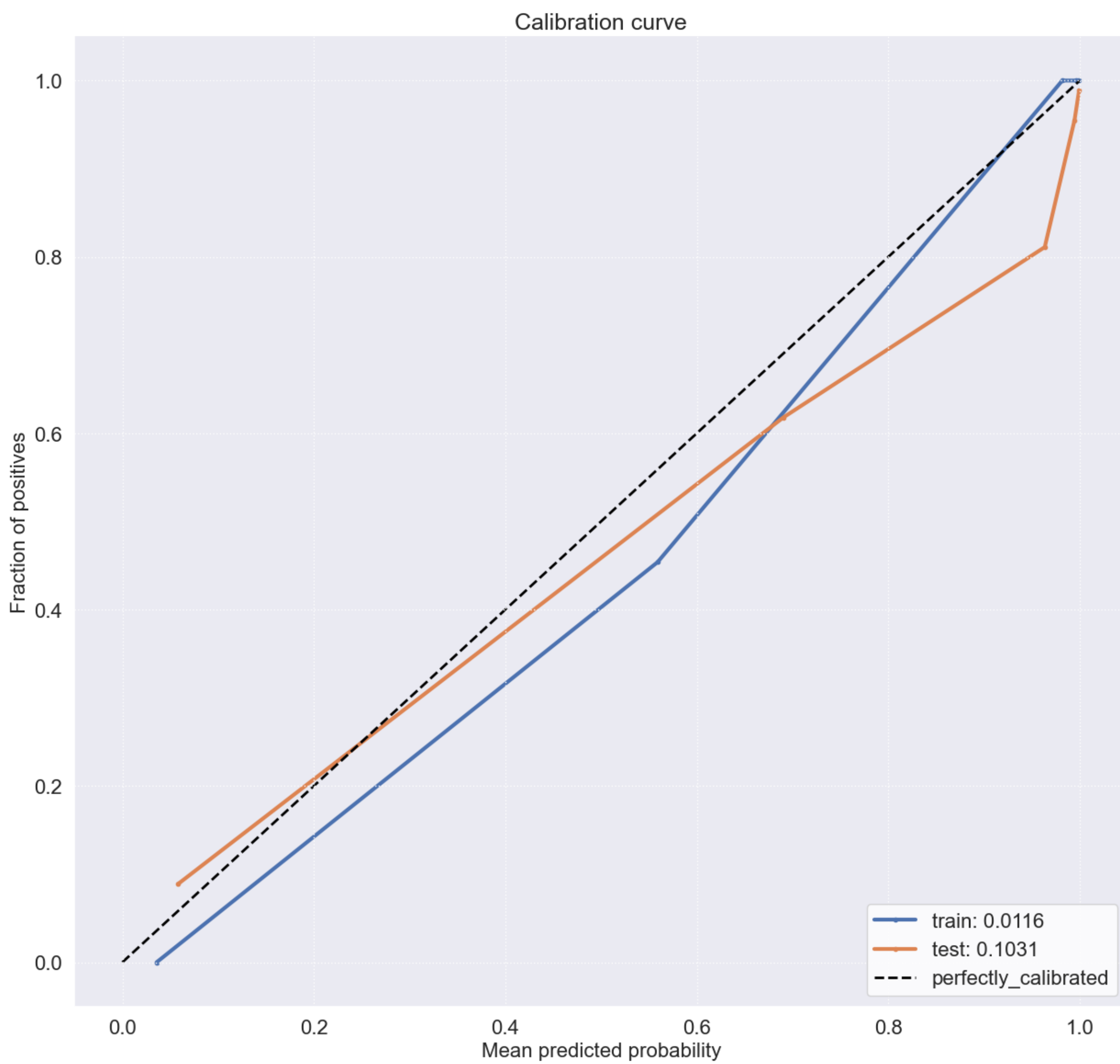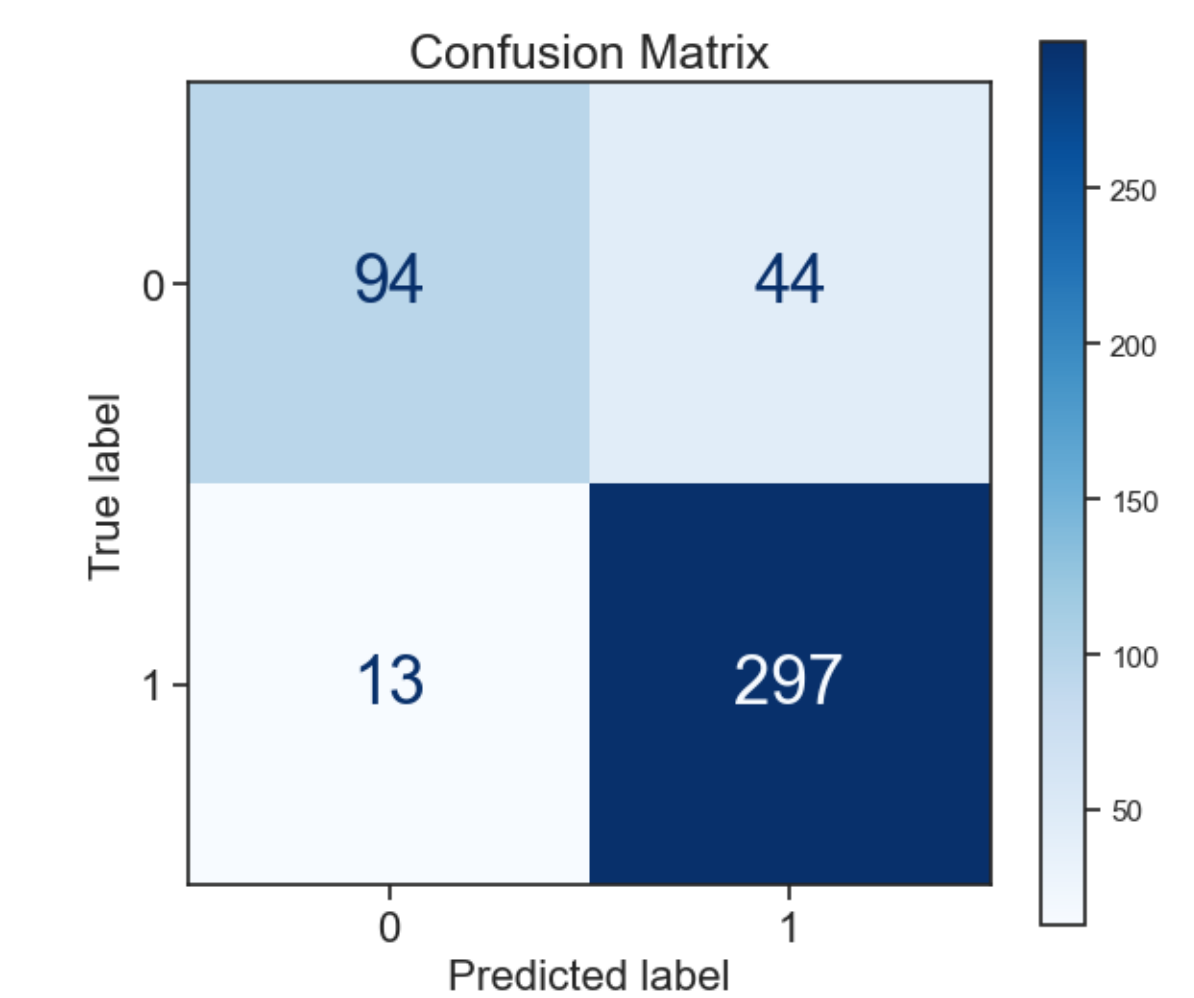
Metric	Train	Test
AUC	0.9999	0.9191
accuracy	0.9899	0.8728
balanced accuracy	0.9837	0.8196
recall_(sens)	1.0	0.9581
specificity	0.9675	0.6812
precision	0.9856	0.871
f1_score	0.9899	0.8677

Protein: P28335 - metric: precision

Total (train): 1788

Activity type	Count
Active	553
Inactive	1235

ROC curve

- AUC_train = 0.999
- AUC_test = 0.918
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2391)

Confusion Matrix

	0	1
0	103	35
1	28	282

Confusion Matrix (nomr.)

	0	1
0	0.75	0.25
1	0.09	0.91

Calibration curve

- train: 0.02
- test: 0.108
- perfectly_calibrated

Active/Inactive probability

- Inactive zone
- Active zone
- type
  - active (310)
  - inactive (138)
  - decoy (447)

XGBoost Parameters	Value
alpha	2.9929
colsample_bytree	0.9595
gamma	0.1978
lambda	2.4528
learning_rate	0.2287
max_depth	3.0
min_child_weight	3.0
n_estimators	754.0
reg_alpha	0.5149
reg_lambda	0.0868
scale_pos_weight	1.0302
subsample	0.7295
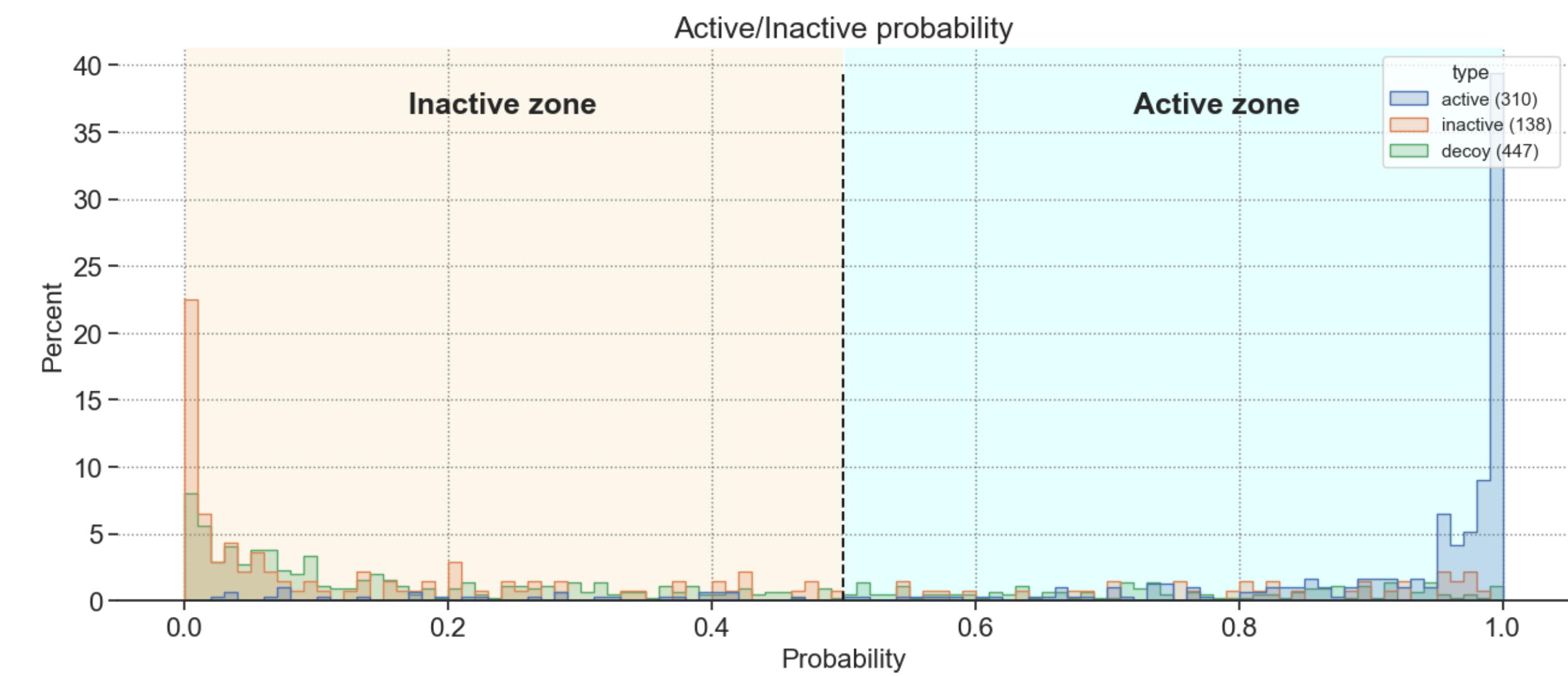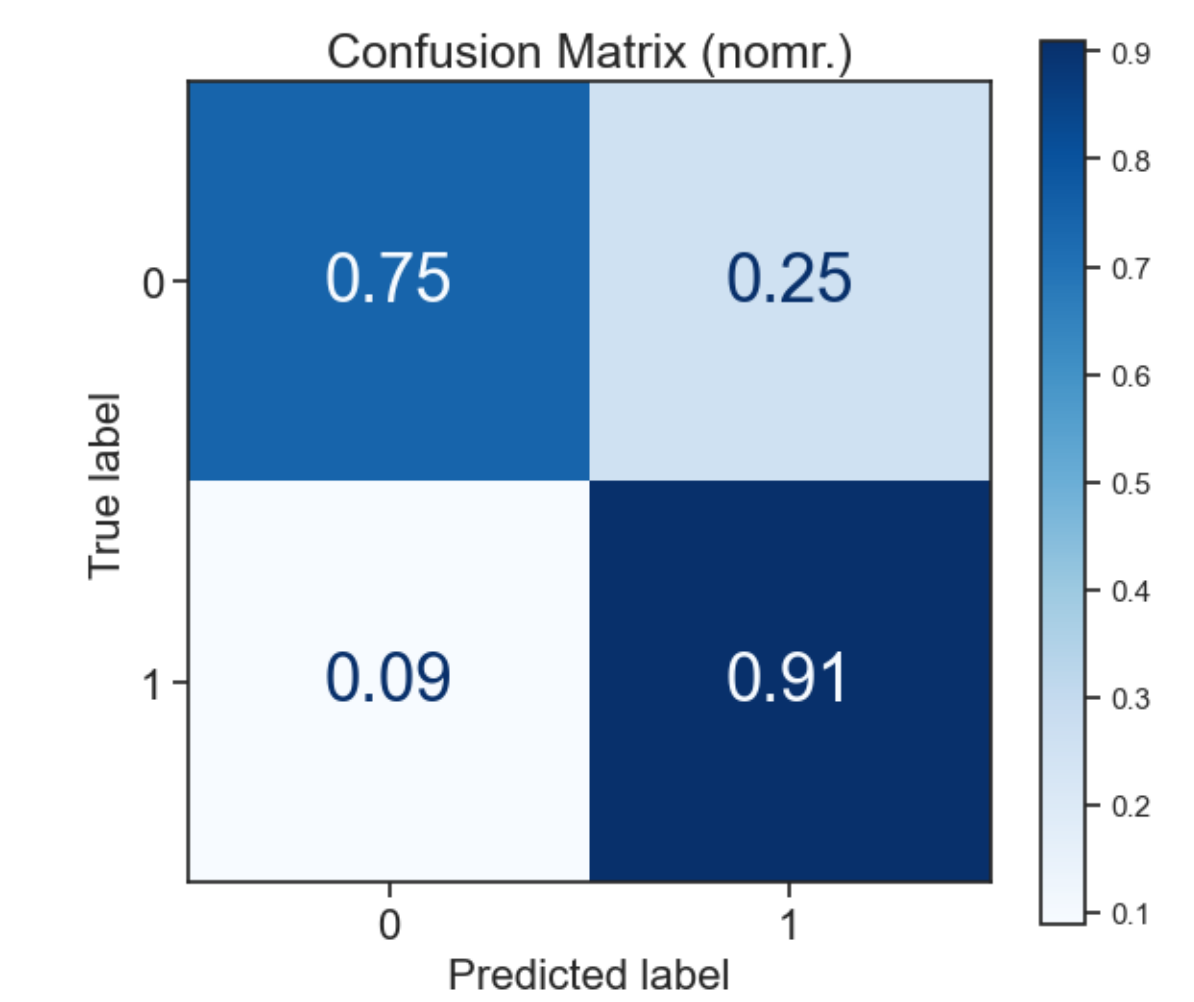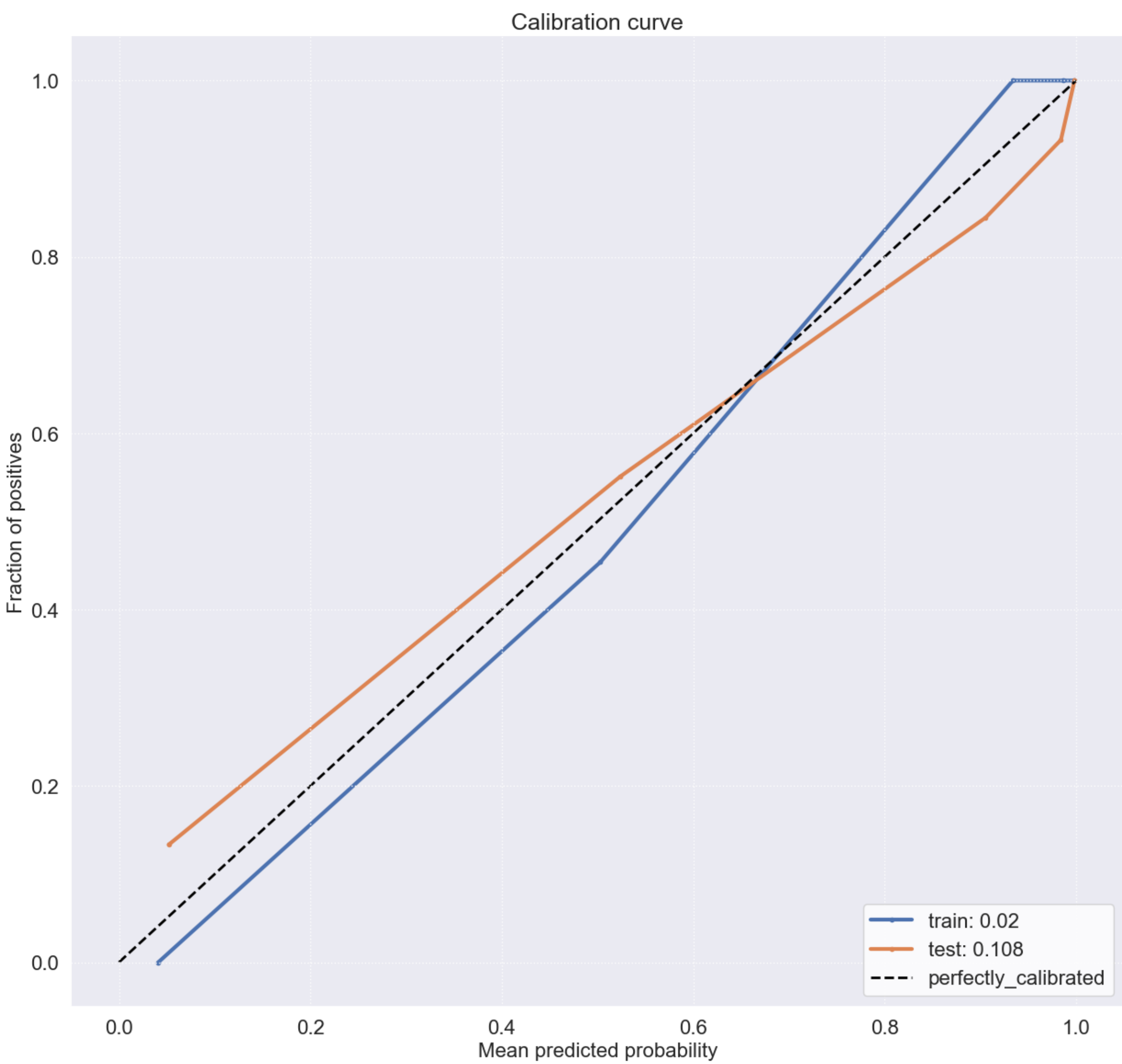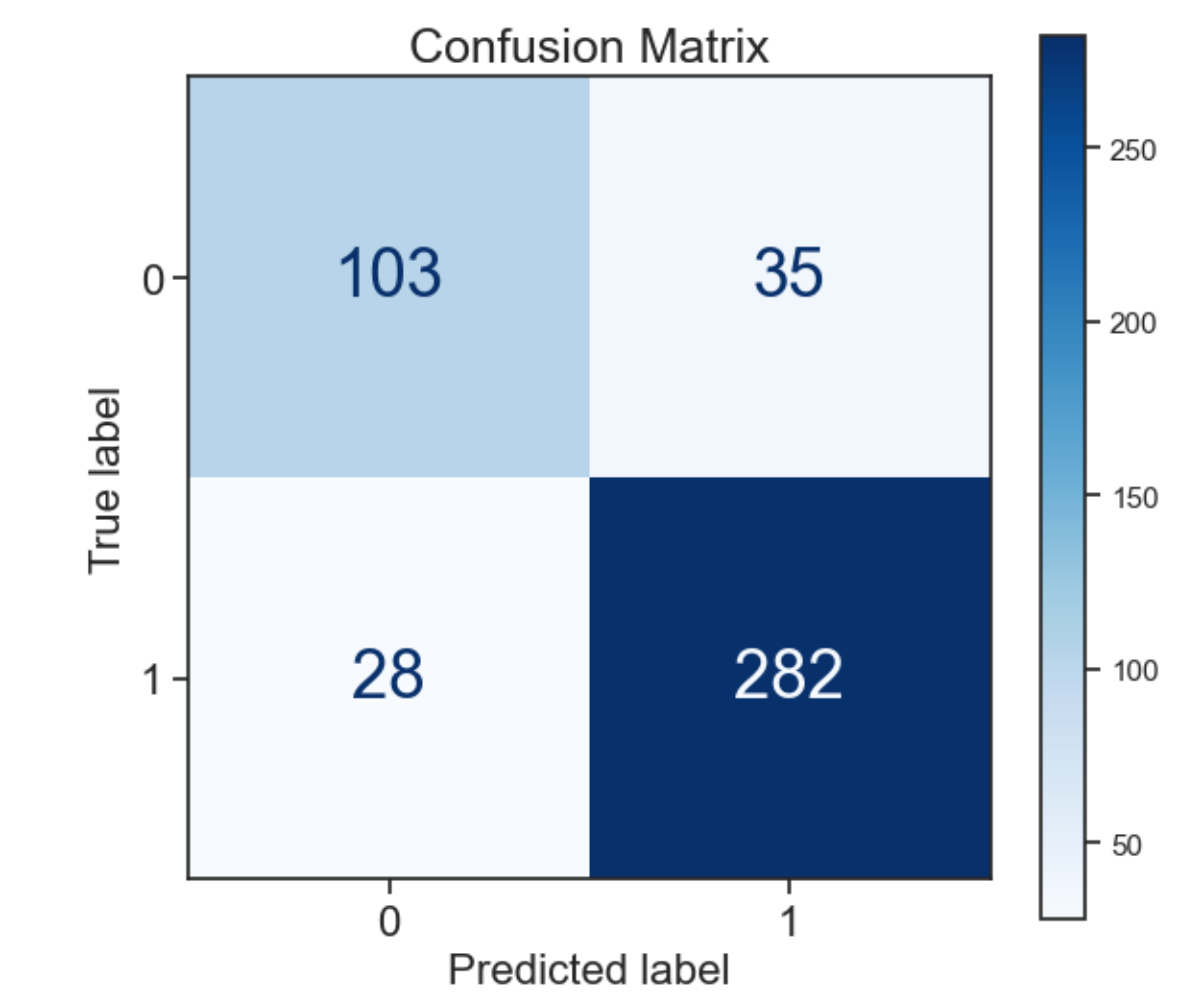
Metric	Train	Test
AUC	0.9991	0.918
accuracy	0.9827	0.8594
balanced accuracy	0.9745	0.828
recall_(sens)	0.996	0.9097
specificity	0.953	0.7464
precision	0.9793	0.8896
f1_score	0.9826	0.8583

Protein: P28335 - metric: SoftVotingXGBoostEnsemble
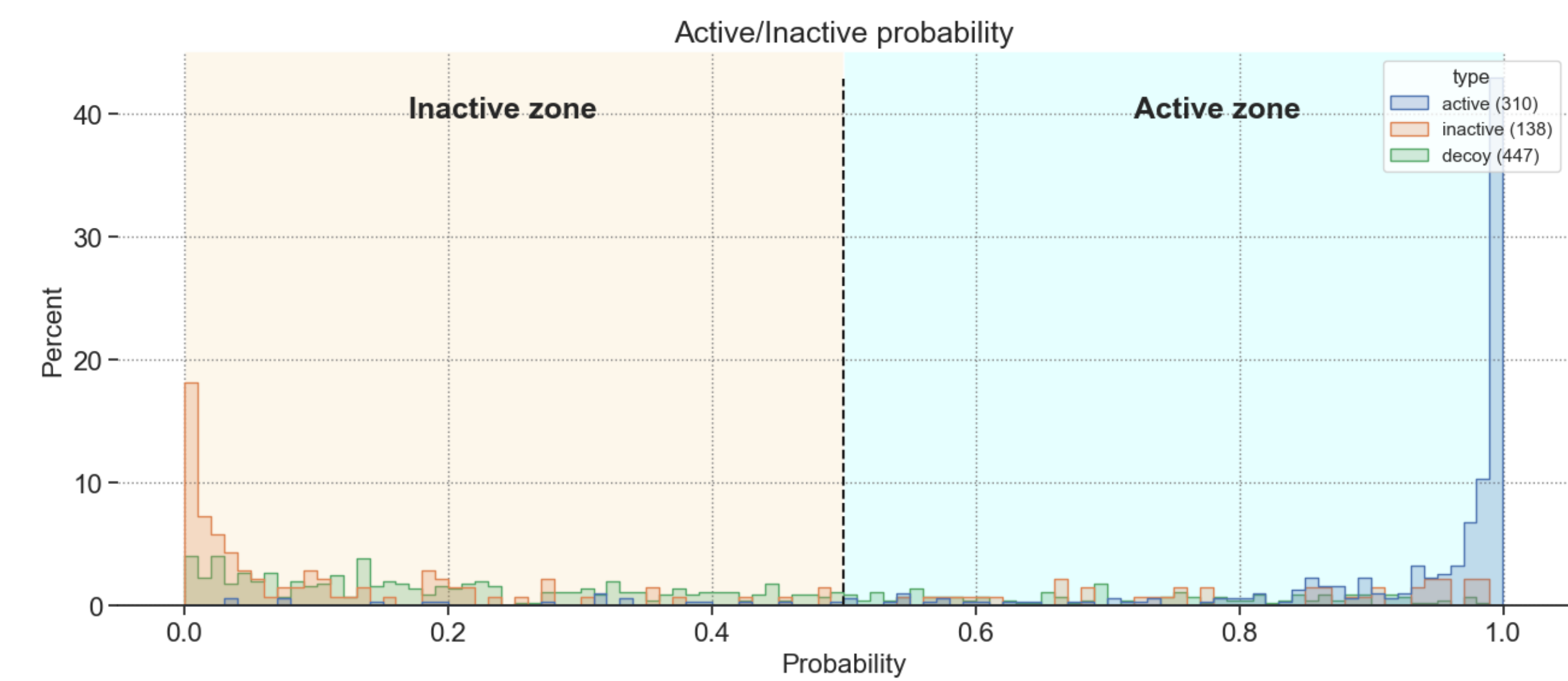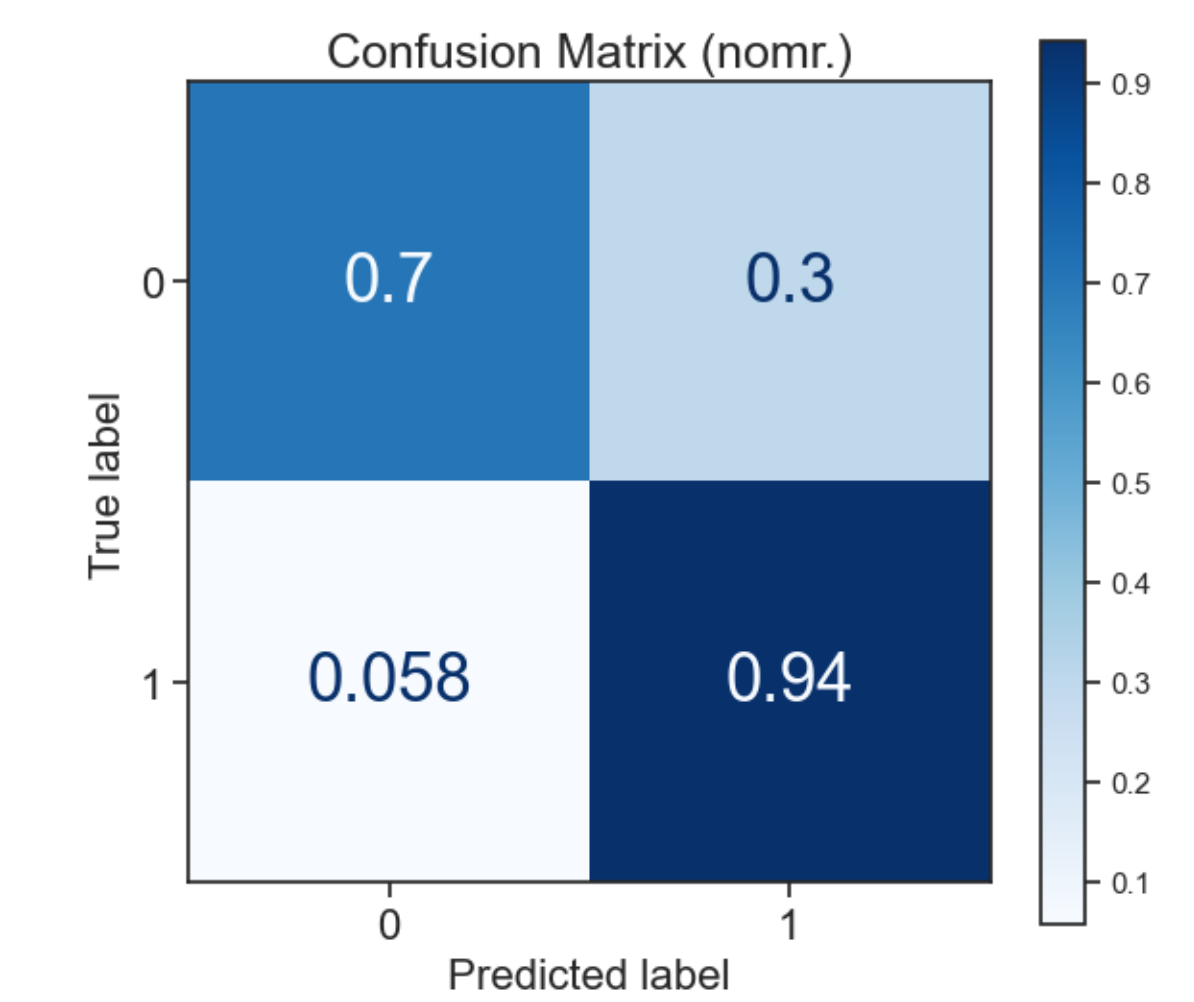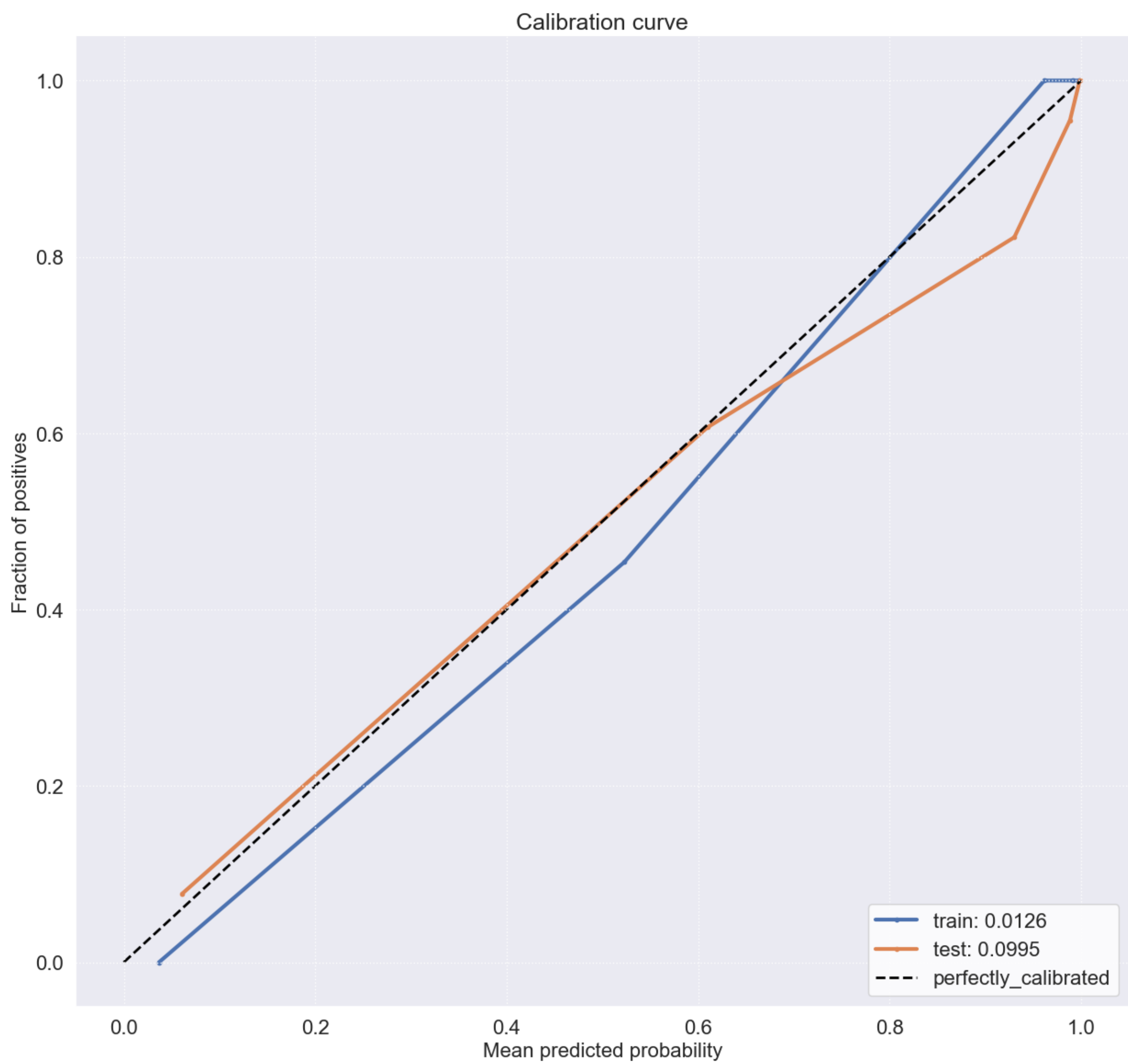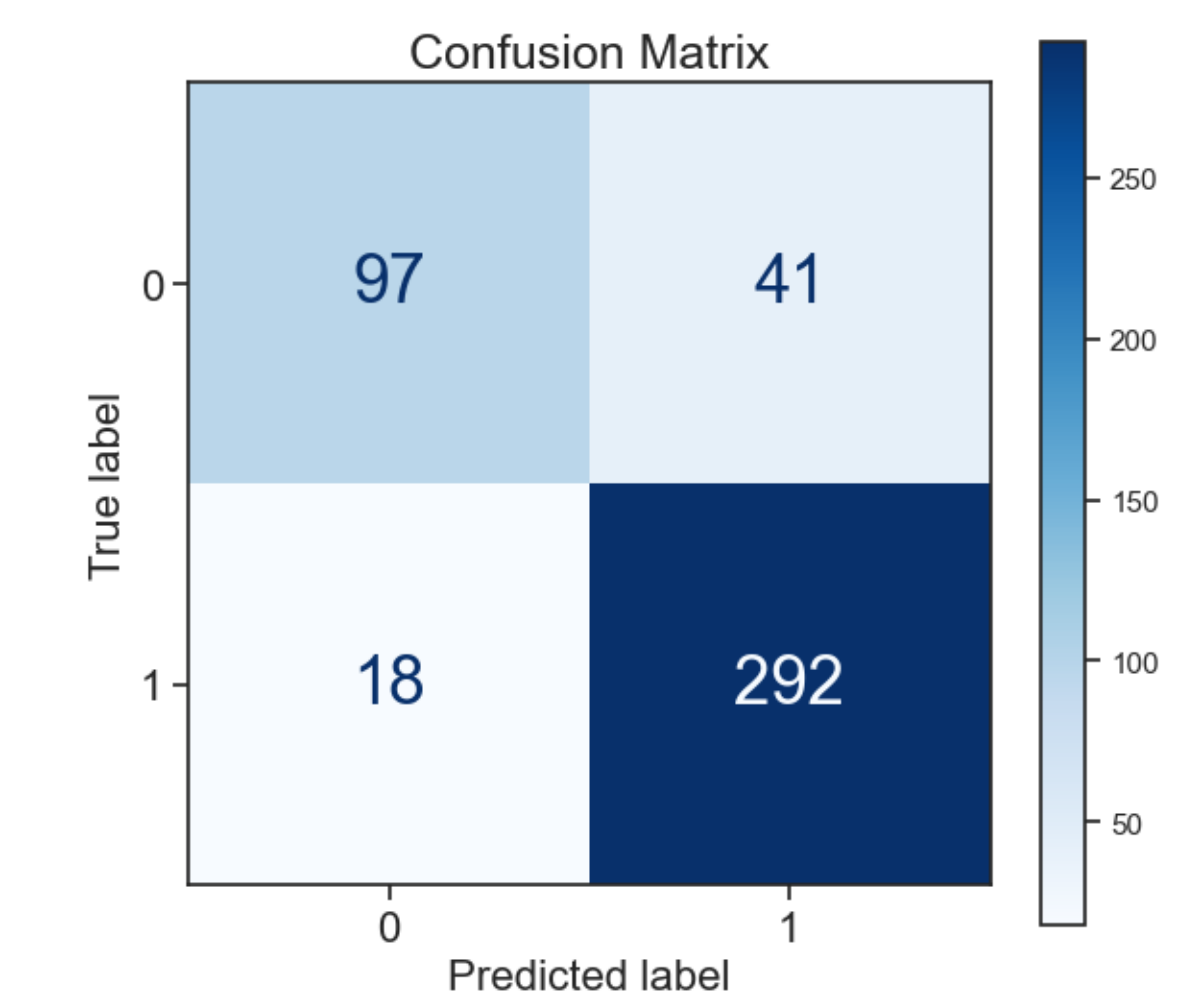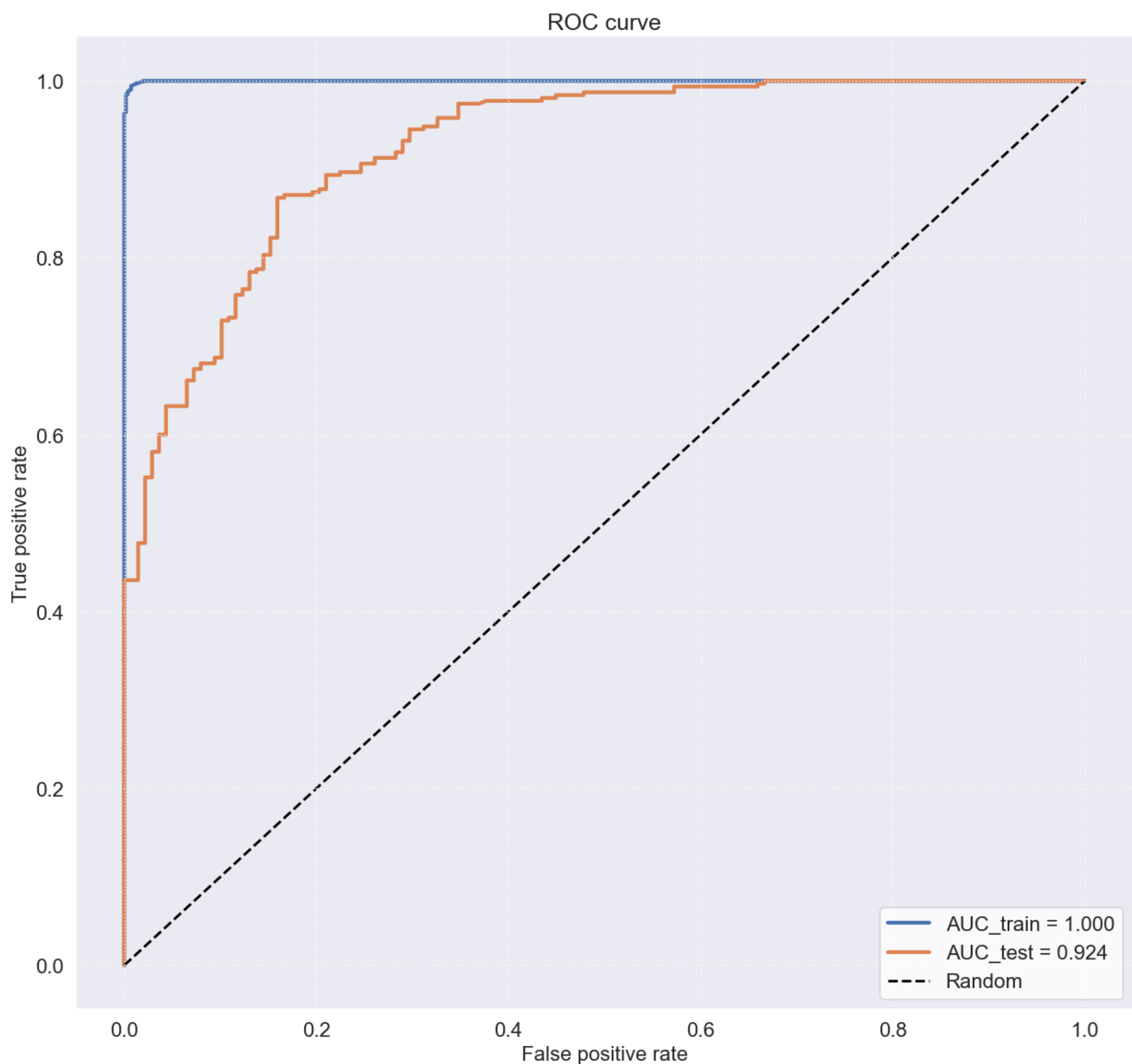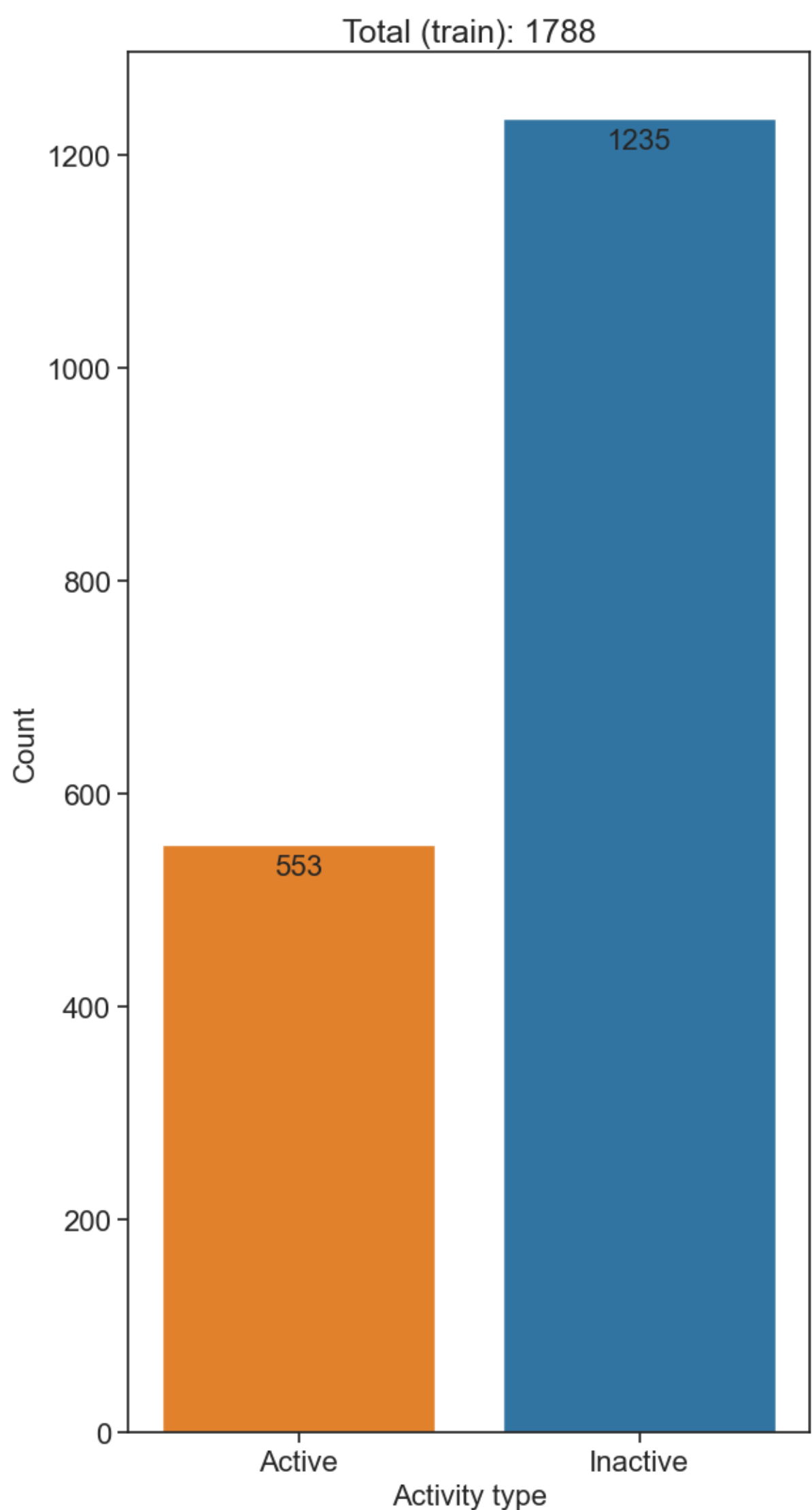
Metric	Train	Test
AUC	0.9998	0.9243
accuracy	0.9916	0.8683
balanced accuracy	0.9864	0.8224
recall_(sens)	1.0	0.9419
specificity	0.9729	0.7029
precision	0.988	0.8769
f1_score	0.9916	0.8647