Protein: P03372 - metric: balanced_accuracy

Total (train): 2150
- Inactive: 792
- Active: 1358

ROC curve
- AUC_train = 0.998
- AUC_test = 0.965
- Random

Pareto Chart with Highlighted Best Attempt
- Trials
- Pareto Front Trials
- Best Trial (# 2620)

X-axis: Score balanced_accuracy (Test)
Y-axis: Difference in Train-Test balanced_accuracy

Confusion Matrix

	Predicted 0	Predicted 1
True 0	162	36
True 1	14	326

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.82	0.18
True 1	0.041	0.96

Calibration curve
- train: 0.0234
- test: 0.0682
- perfectly_calibrated

X-axis: Mean predicted probability
Y-axis: Fraction of positives

Active/Inactive probability

- Inactive zone
- Active zone

type:
- active (340)
- inactive (198)
- decoy (538)

X-axis: Probability
Y-axis: Percent

XGBoost Parameters

XGBoost Parameters	Value
alpha	1.475
colsample_bytree	0.8453
gamma	0.834
lambda	2.0559
learning_rate	0.078
max_depth	3.0
min_child_weight	1.0
n_estimators	979.0
reg_alpha	0.081
reg_lambda	0.8909
scale_pos_weight	1.3844
subsample	0.5104
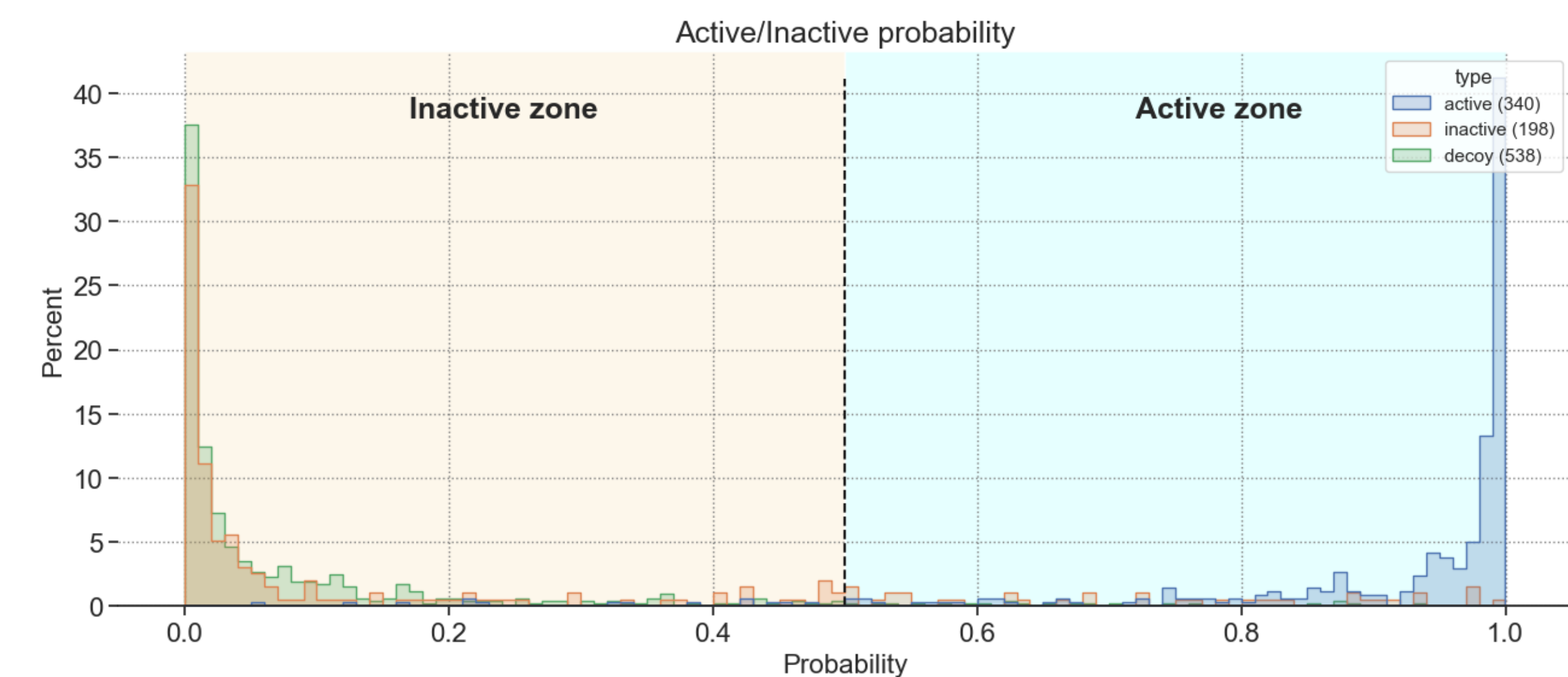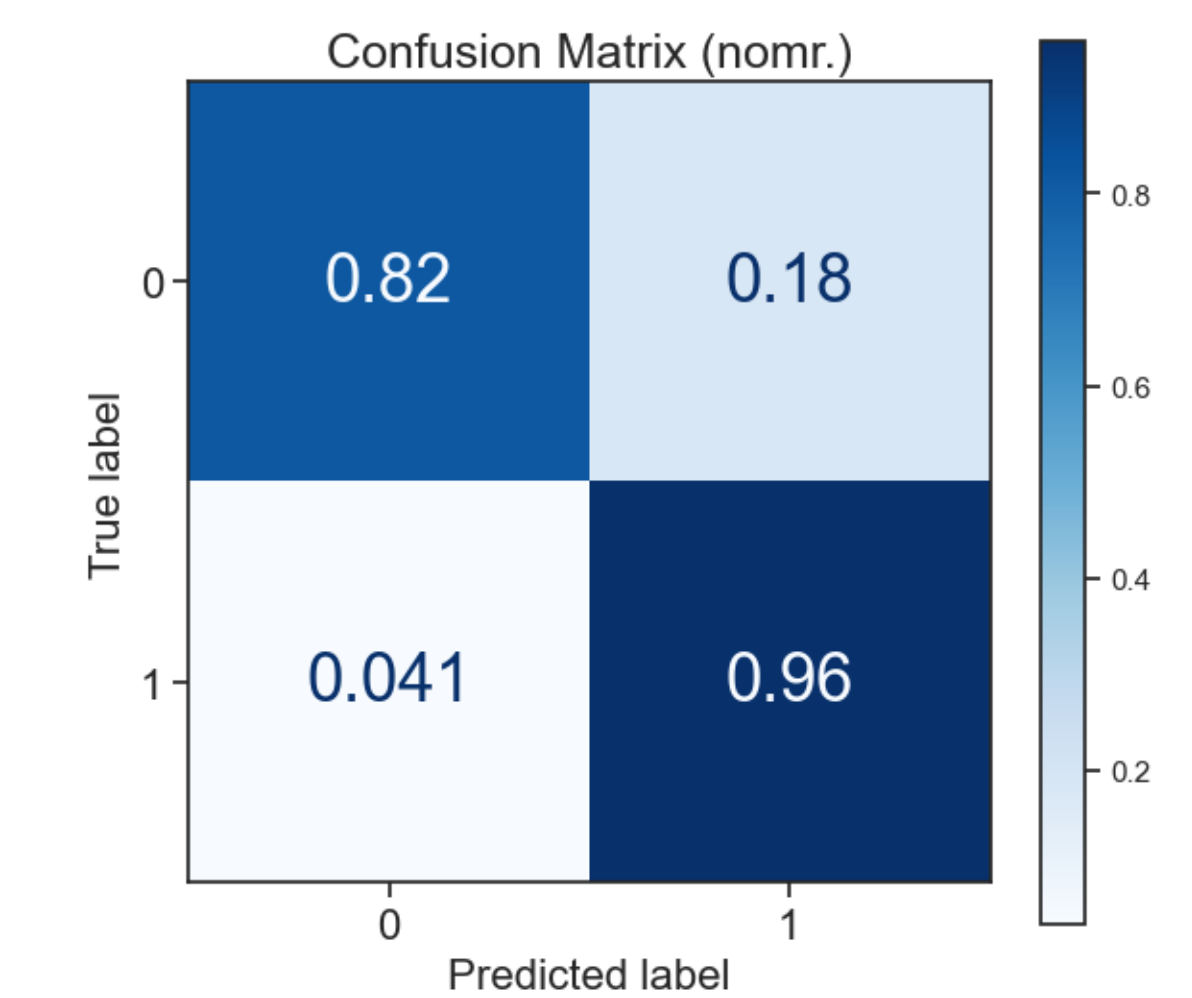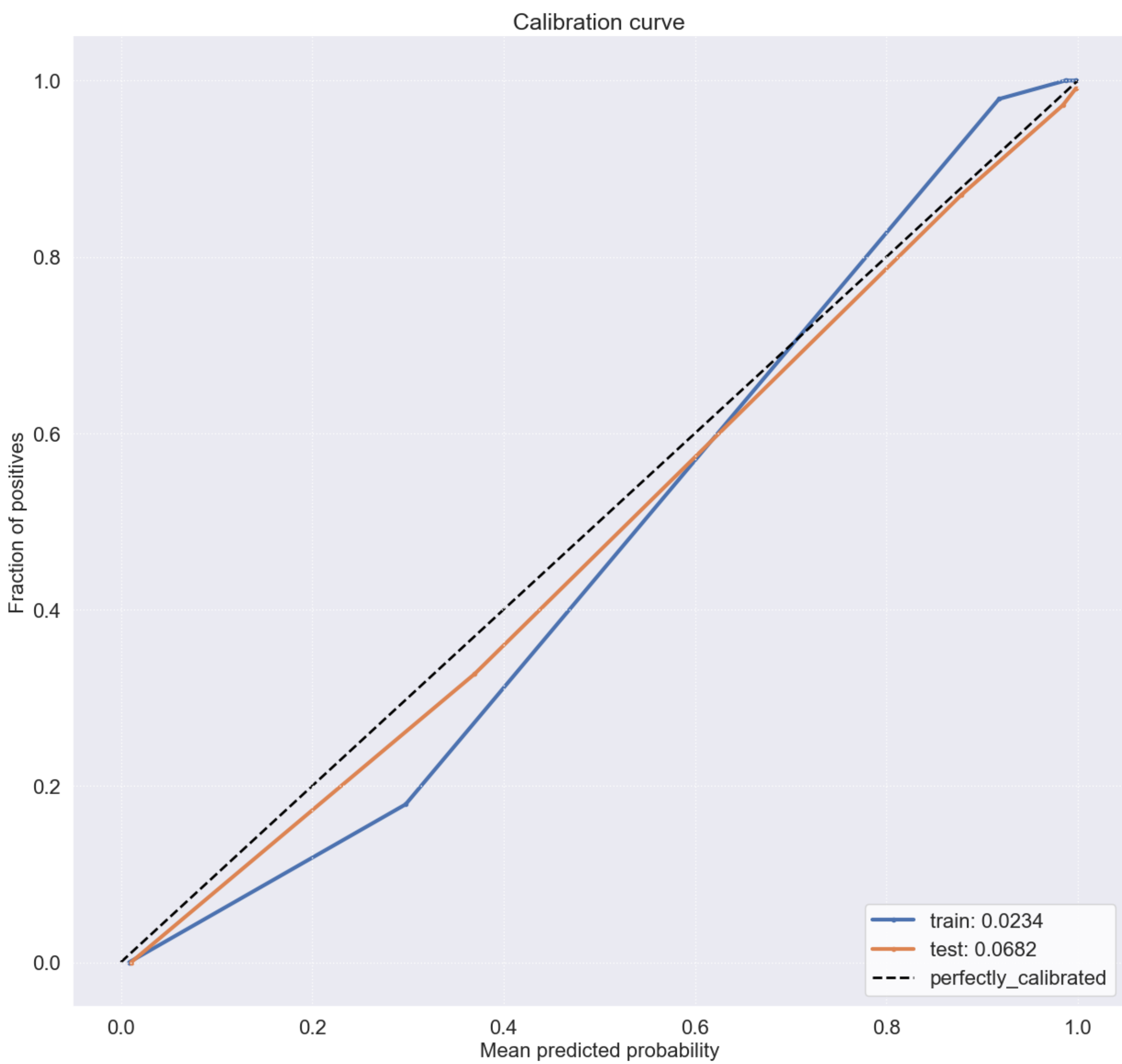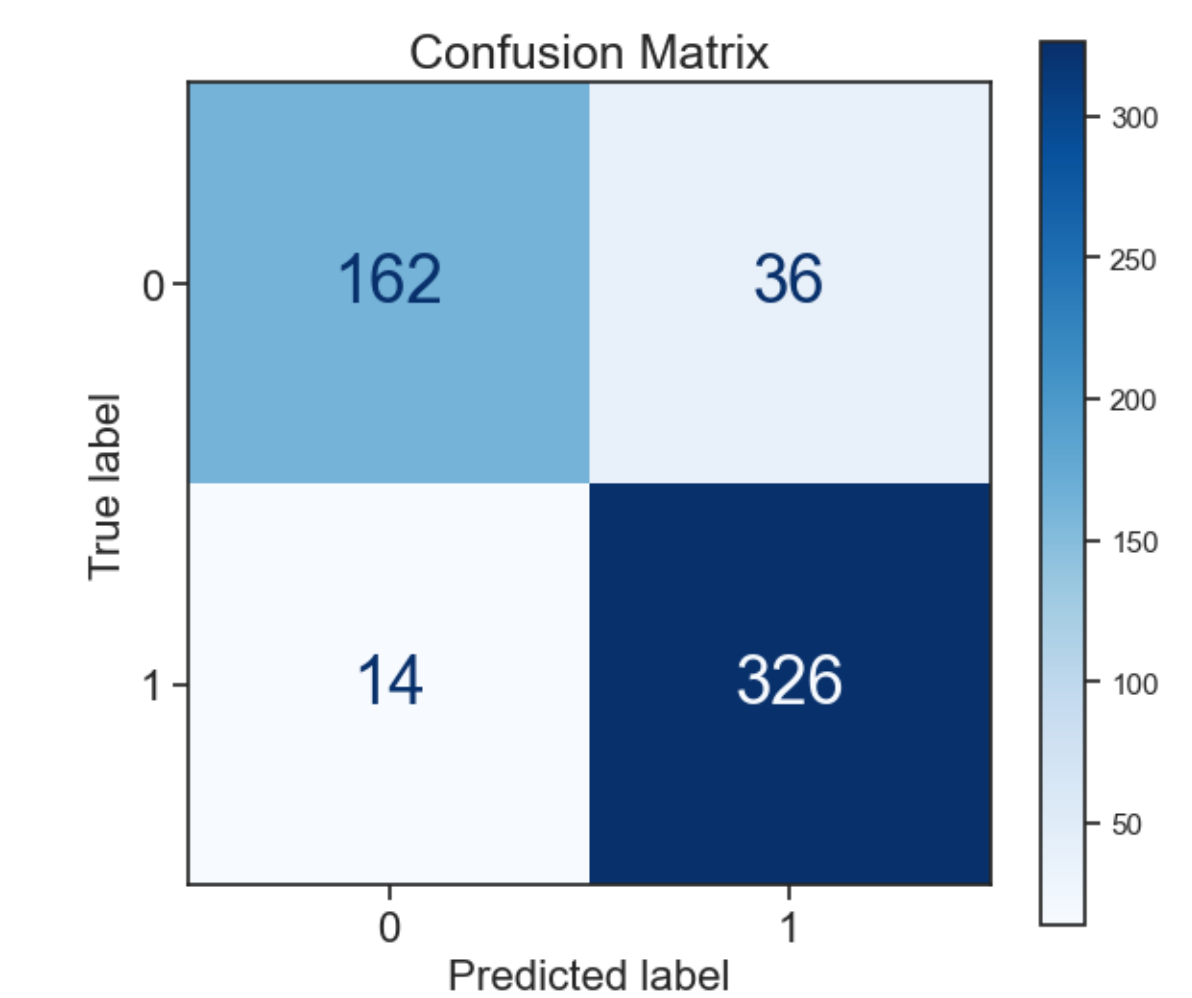
Metrics

Metric	Train	Test
AUC	0.9975	0.9654
accuracy	0.9758	0.9071
balanced accuracy	0.9688	0.8885
recall_(sens)	0.9956	0.9588
specificity	0.9419	0.8182
precision	0.9671	0.9006
f1_score	0.9757	0.9058

Protein: P03372 - metric: f1

Total (train): 2150

ROC curve

AUC_train = 1.000
AUC_test = 0.967
Random

Pareto Chart with Highlighted Best Attempt

Trials
Pareto Front Trials
Best Trial (# 2641)
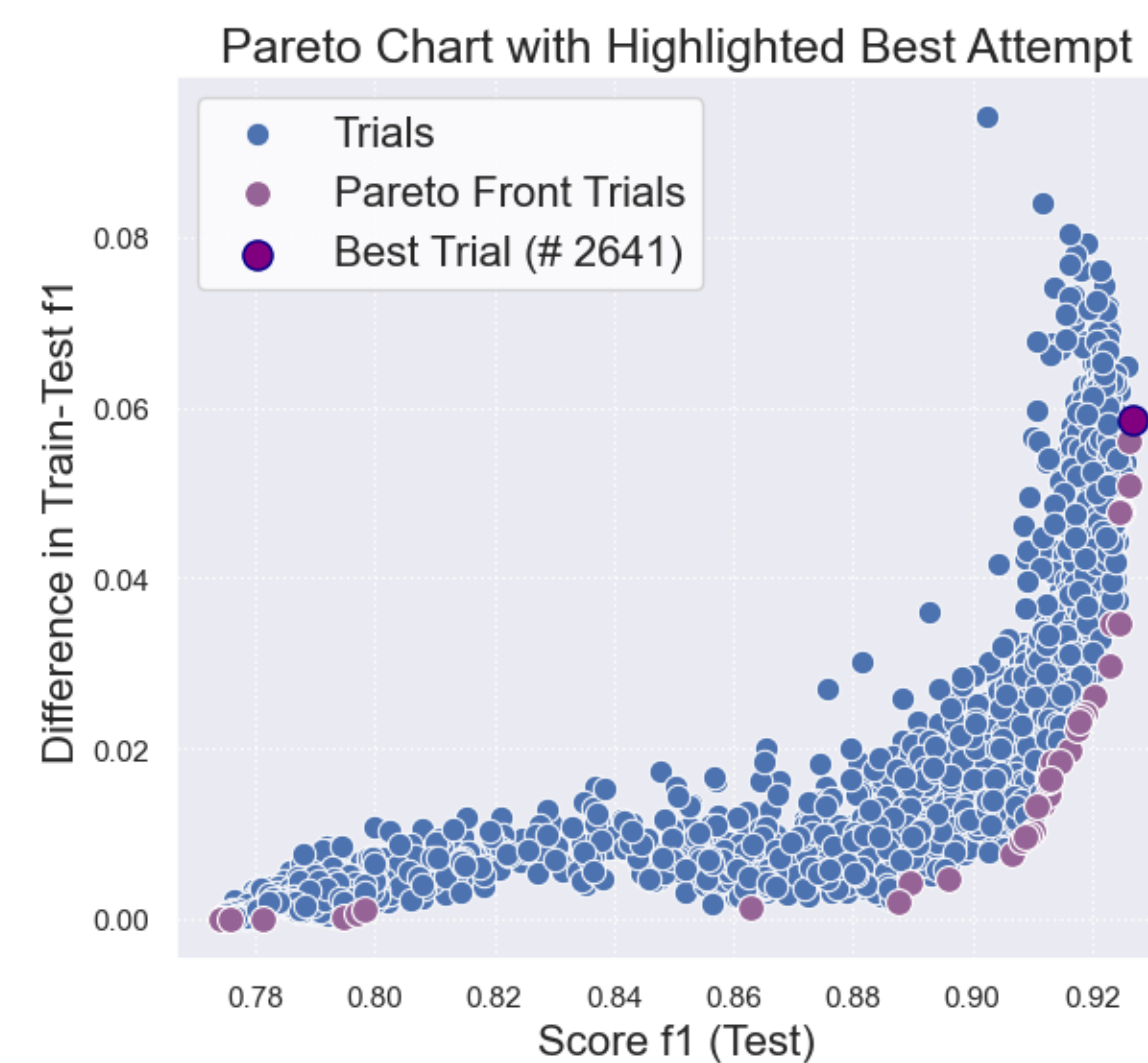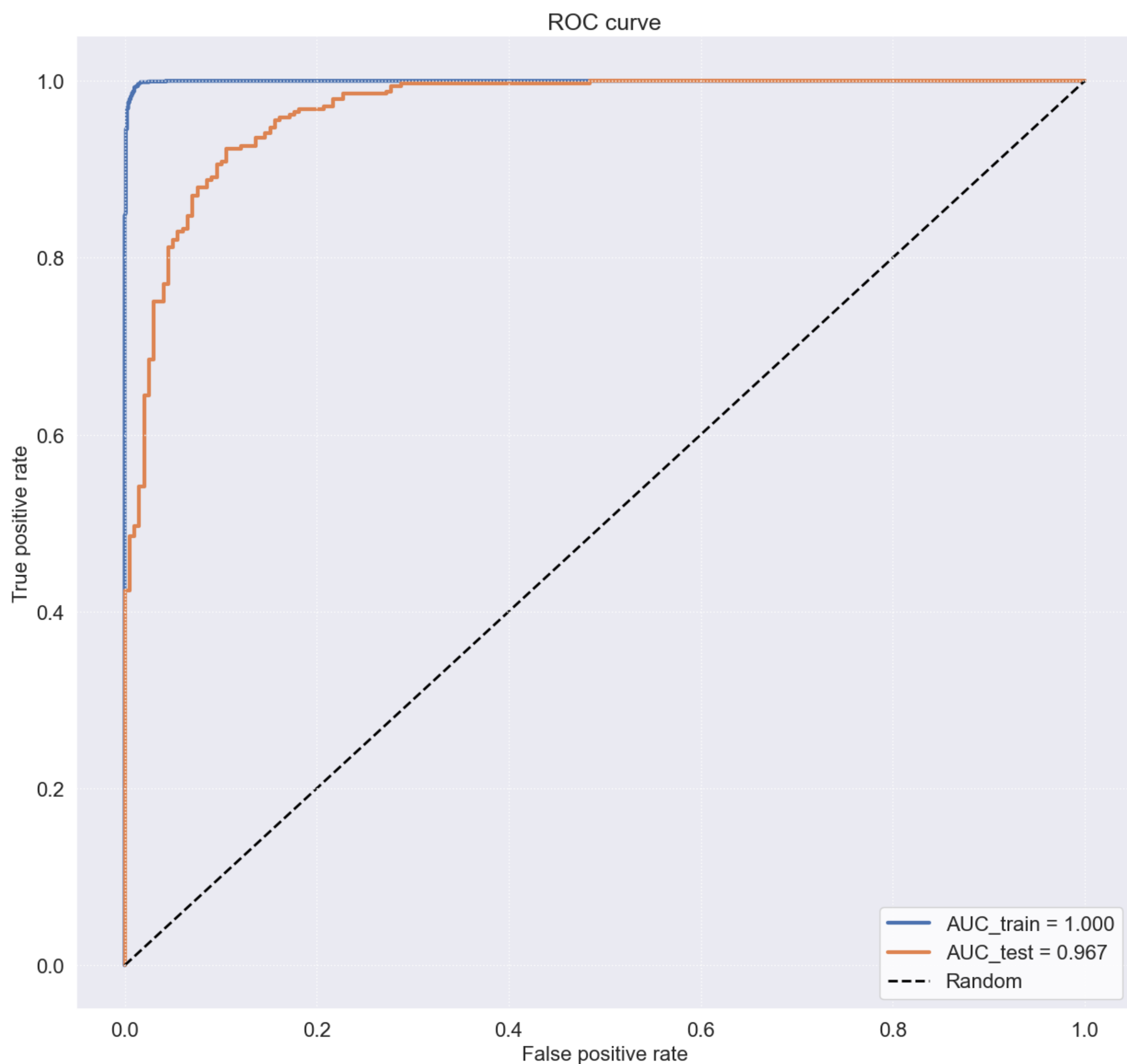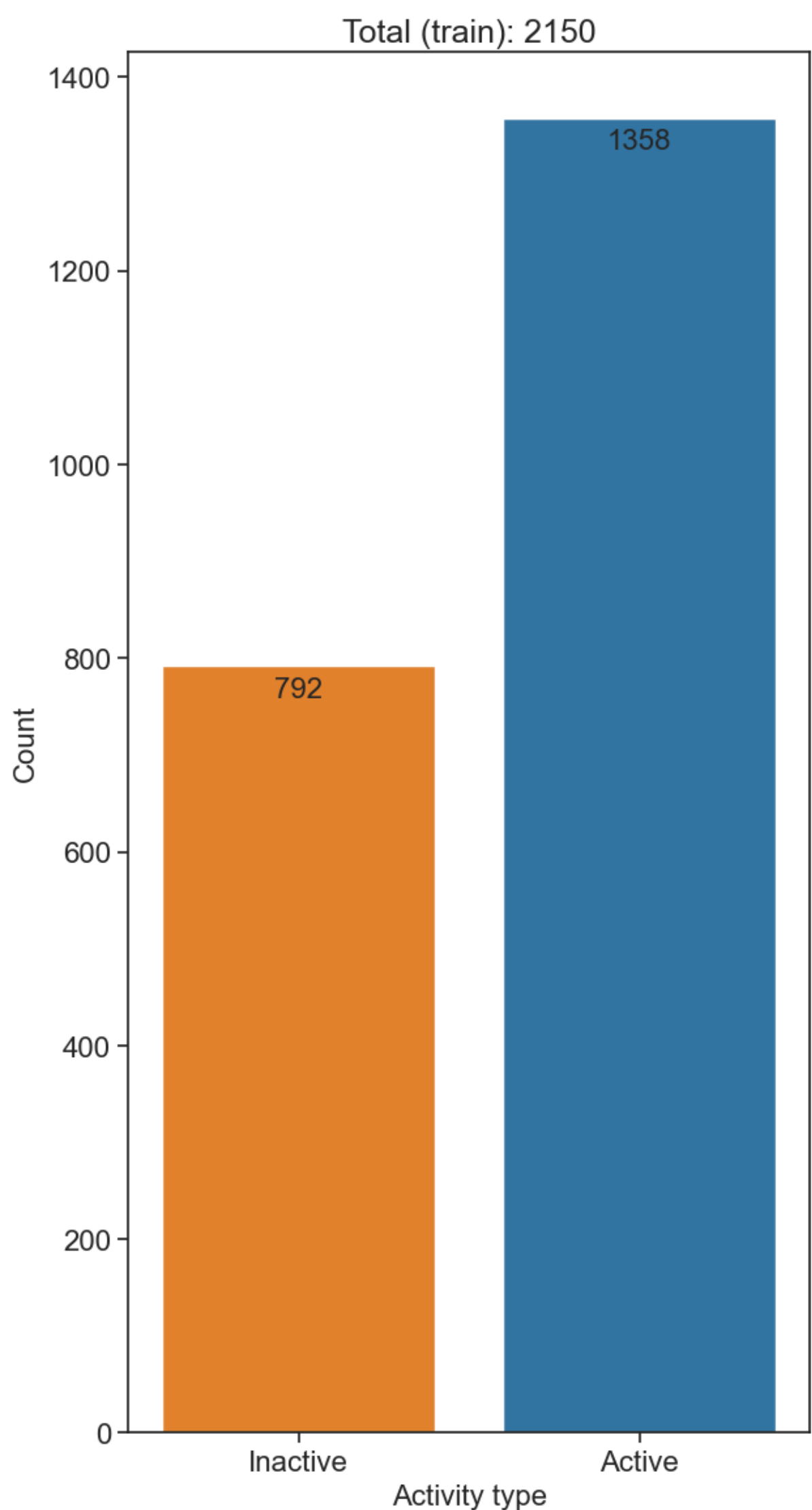
XGBoost Parameters	Value
alpha	4.9598
colsample_bytree	0.5145
gamma	0.99
lambda	0.3936
learning_rate	0.1089
max_depth	7.0
min_child_weight	1.0
n_estimators	789.0
reg_alpha	0.2189
reg_lambda	0.6804
scale_pos_weight	2.5442
subsample	0.5926

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

train: 0.0155
test: 0.0674
perfectly_calibrated

Active/Inactive probability

type
active (340)
inactive (198)
decoy (538)
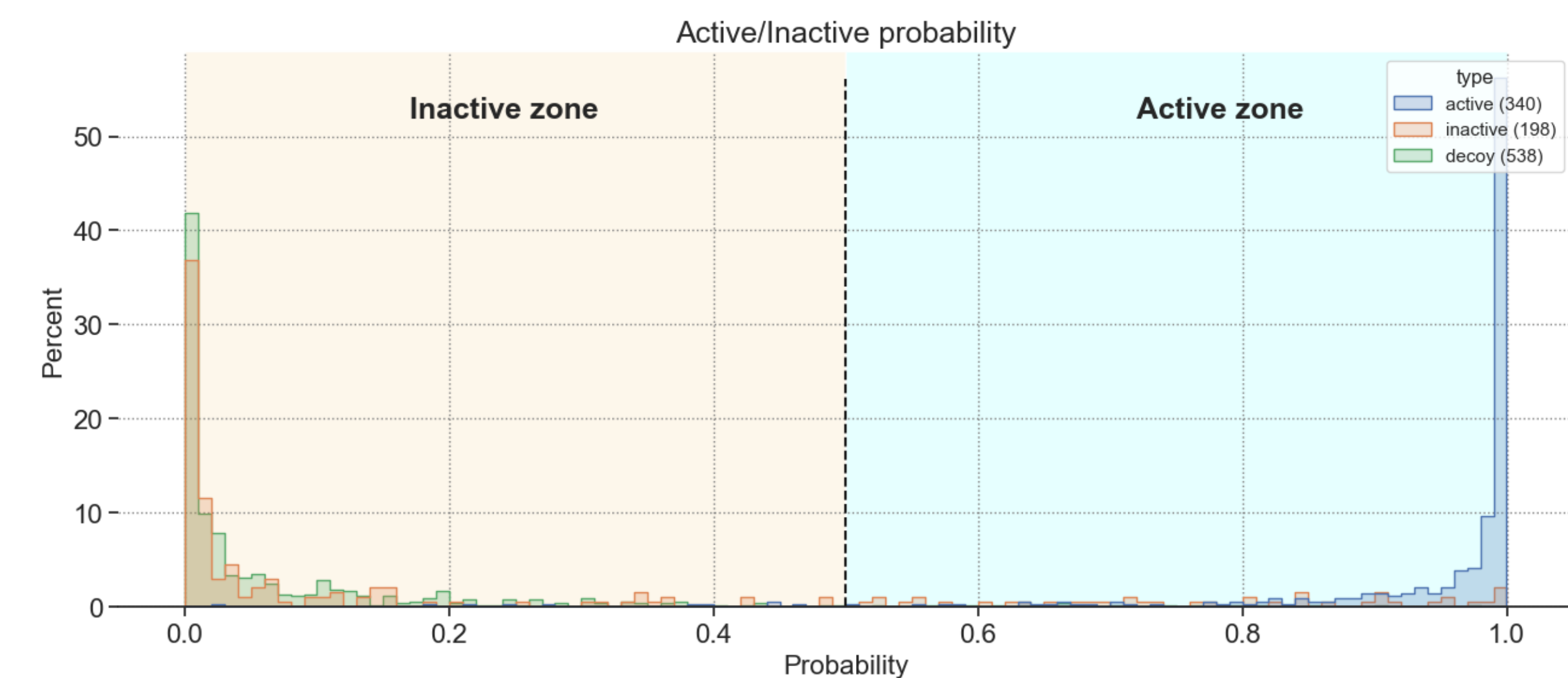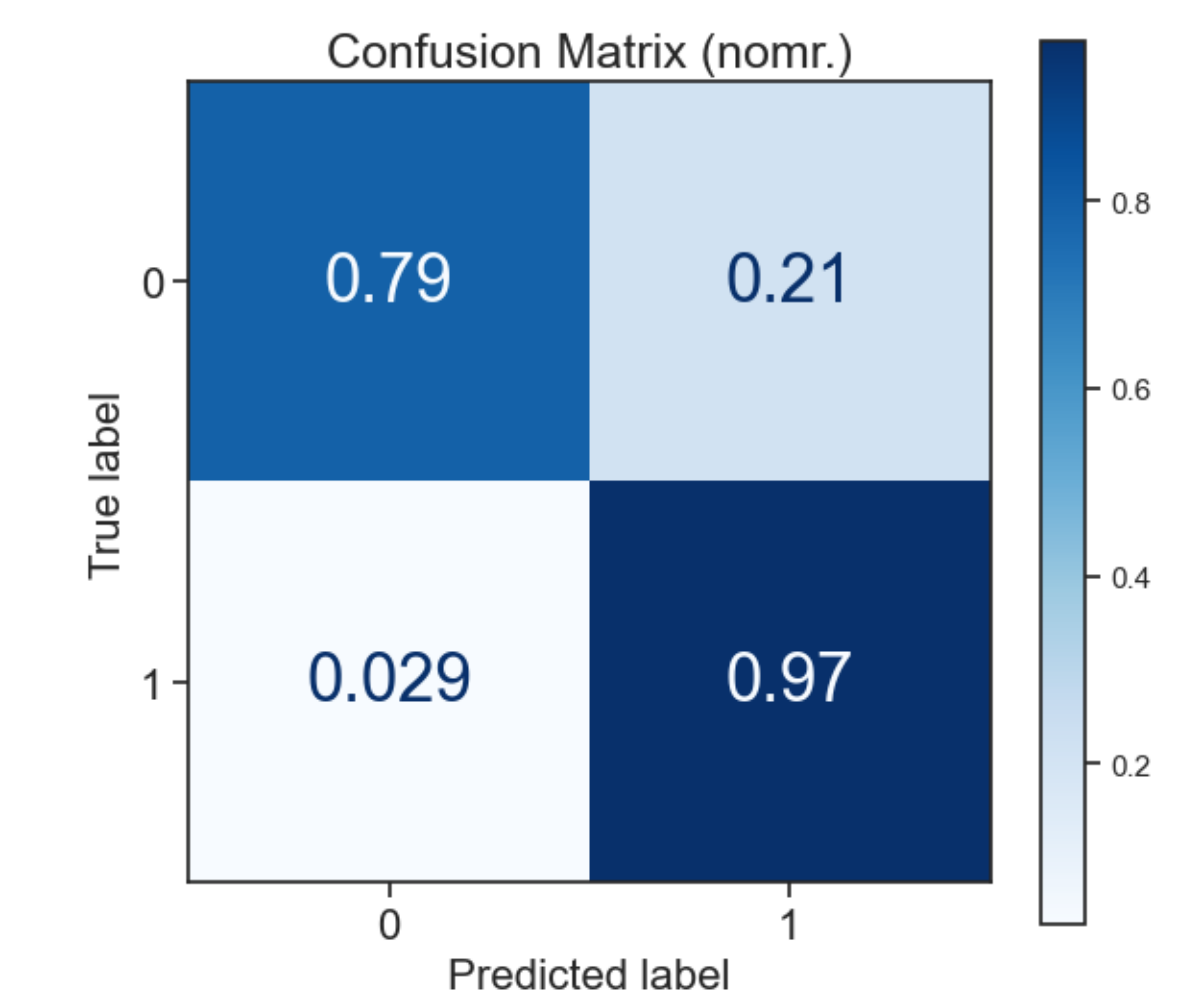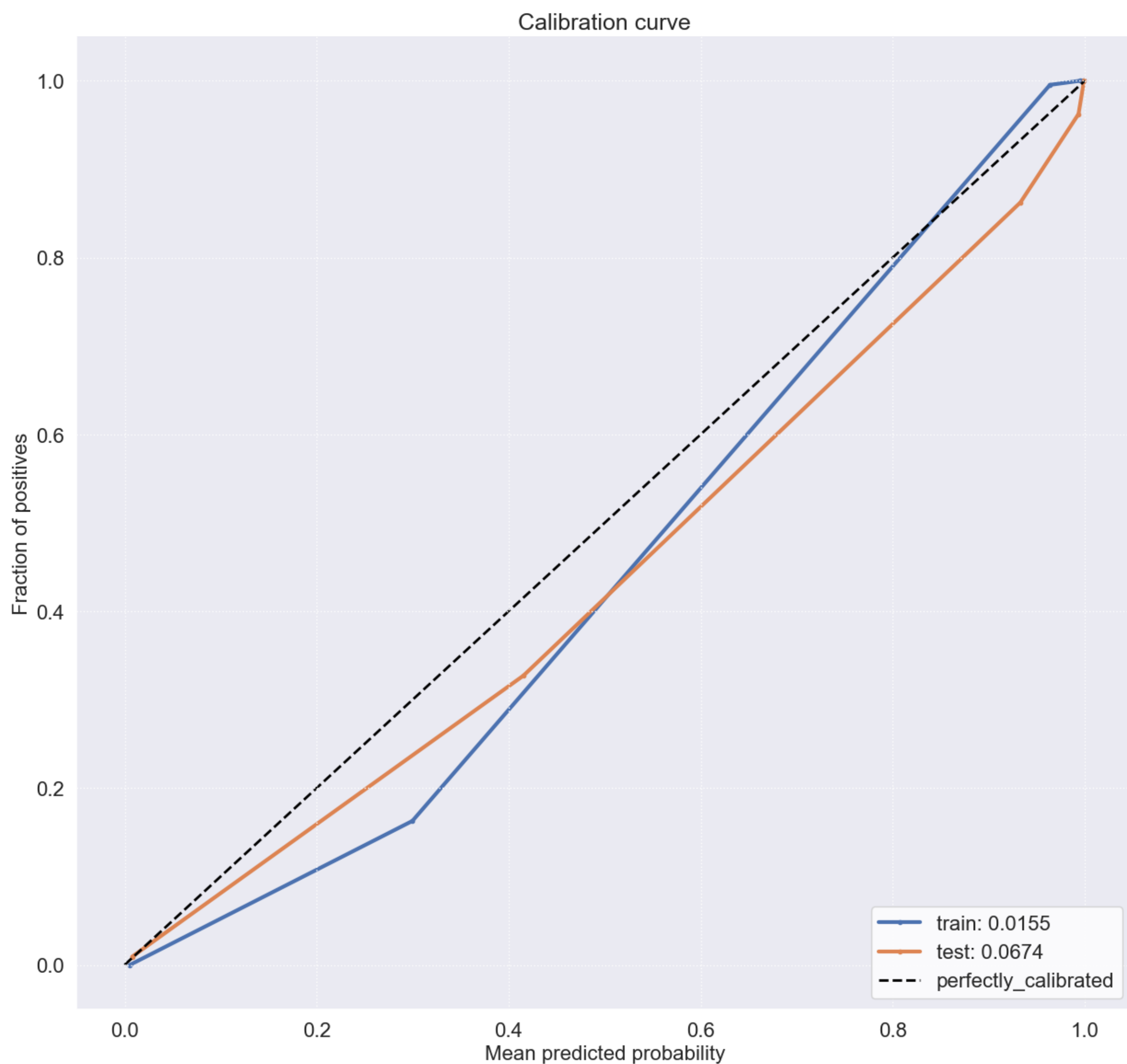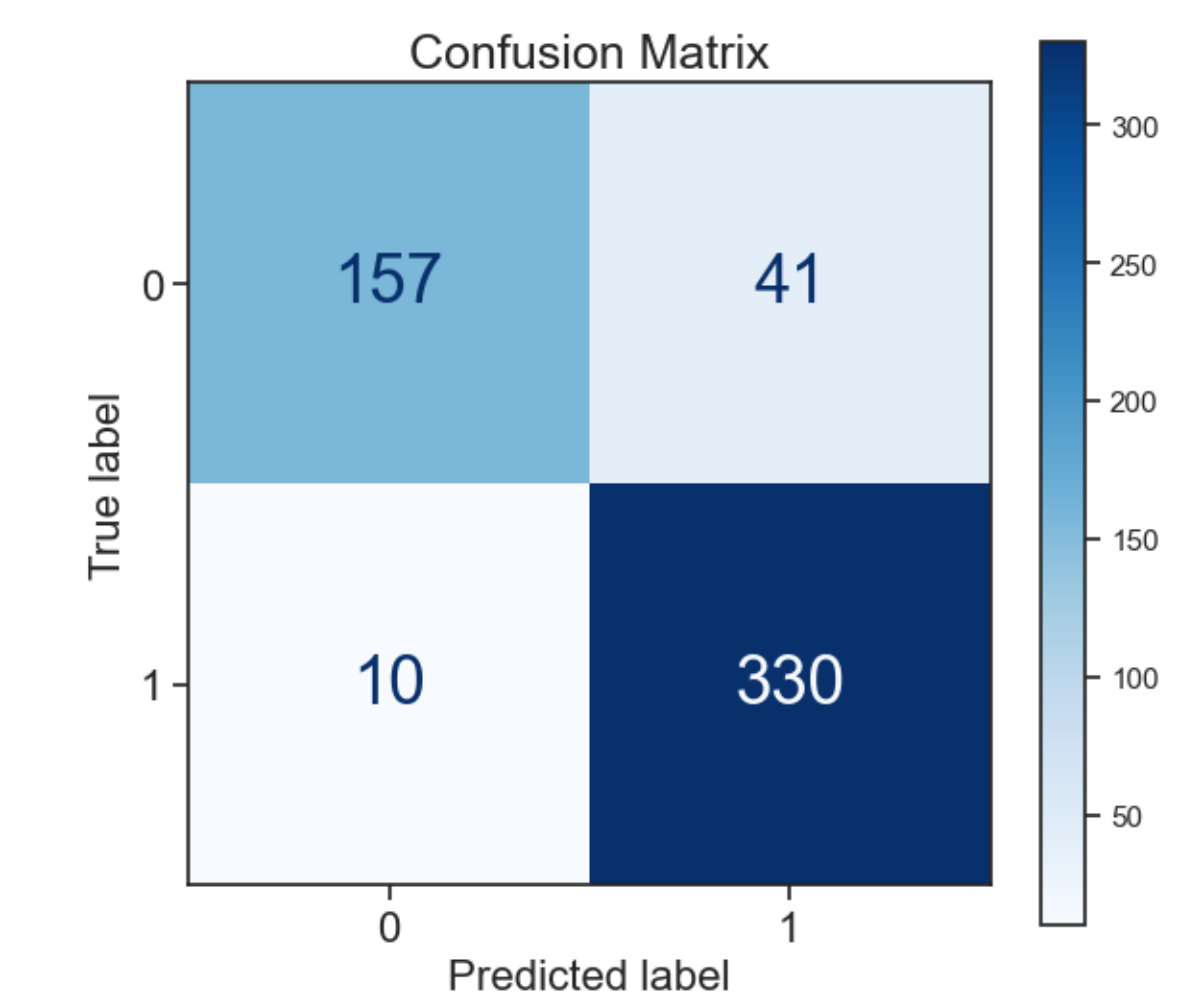
Metric	Train	Test
AUC	0.9995	0.9671
accuracy	0.9828	0.9052
balanced accuracy	0.9766	0.8818
recall_(sens)	1.0	0.9706
specificity	0.9533	0.7929
precision	0.9735	0.8895
f1_score	0.9827	0.9032

Protein: P03372 - metric: precision

Total (train): 2150

Inactive: 792
Active: 1358

ROC curve

AUC_train = 1.000
AUC_test = 0.964
Random

Pareto Chart with Highlighted Best Attempt

Trials
Pareto Front Trials
Best Trial (# 1870)

Confusion Matrix

	Predicted 0	Predicted 1
True 0	168	30
True 1	20	320

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.85	0.15
True 1	0.059	0.94

Calibration curve

train: 0.0089
test: 0.0741
perfectly_calibrated

Active/Inactive probability

Inactive zone / Active zone

type:
active (340)
inactive (198)
decoy (538)

XGBoost Parameters	Value
alpha	1.5539
colsample_bytree	0.7007
gamma	0.0113
lambda	0.9664
learning_rate	0.2098
max_depth	6.0
min_child_weight	3.0
n_estimators	720.0
reg_alpha	0.4591
reg_lambda	0.1916
scale_pos_weight	2.3446
subsample	0.5905
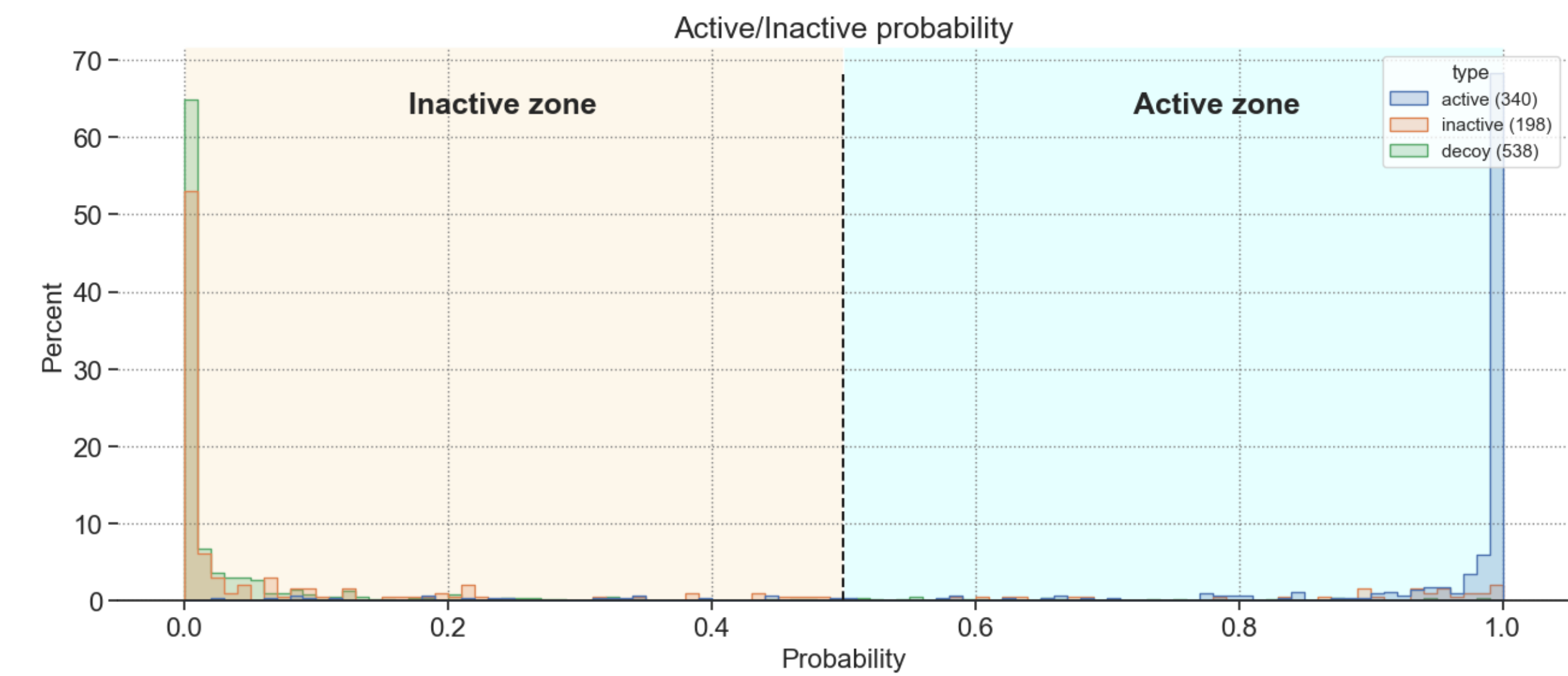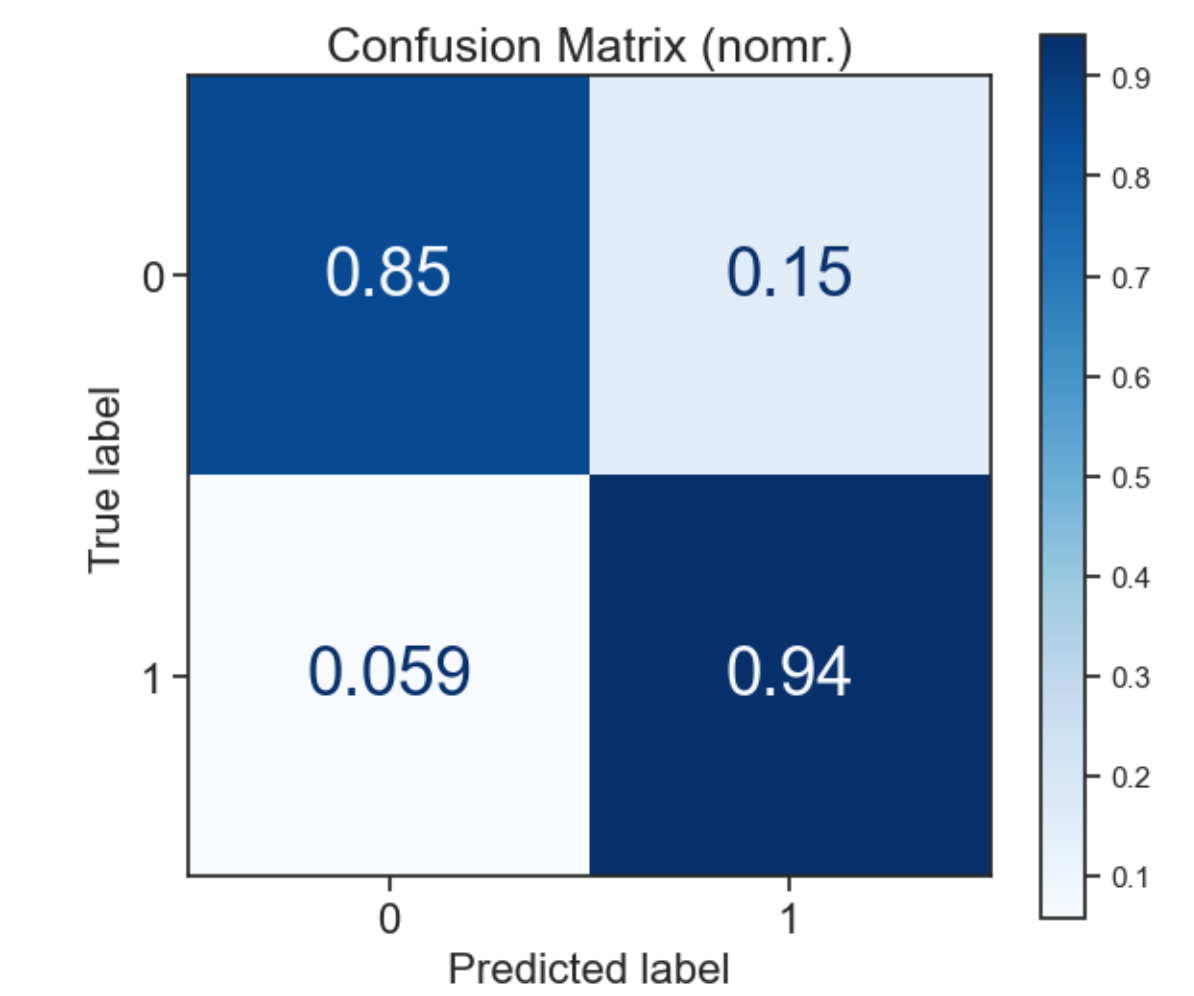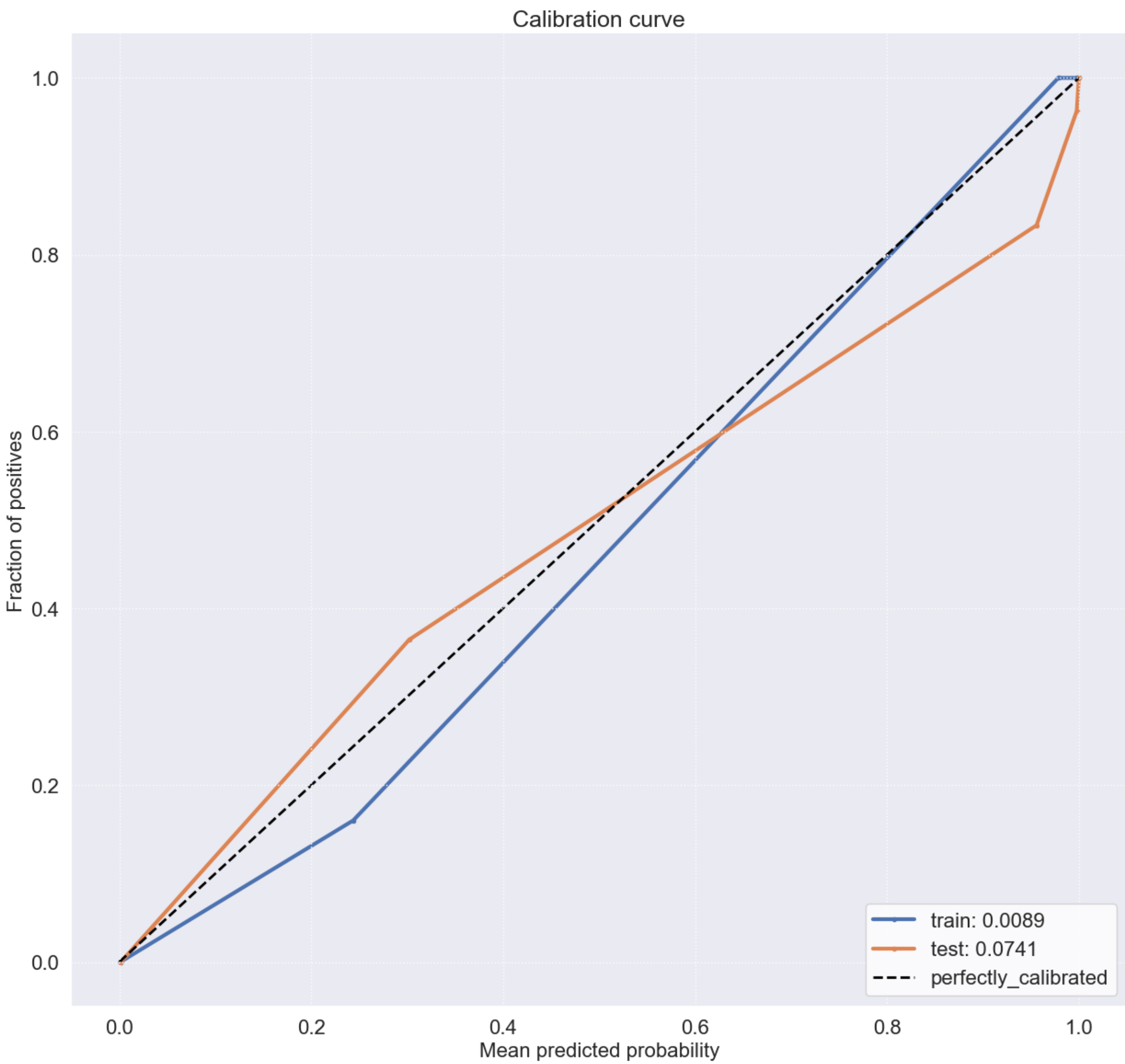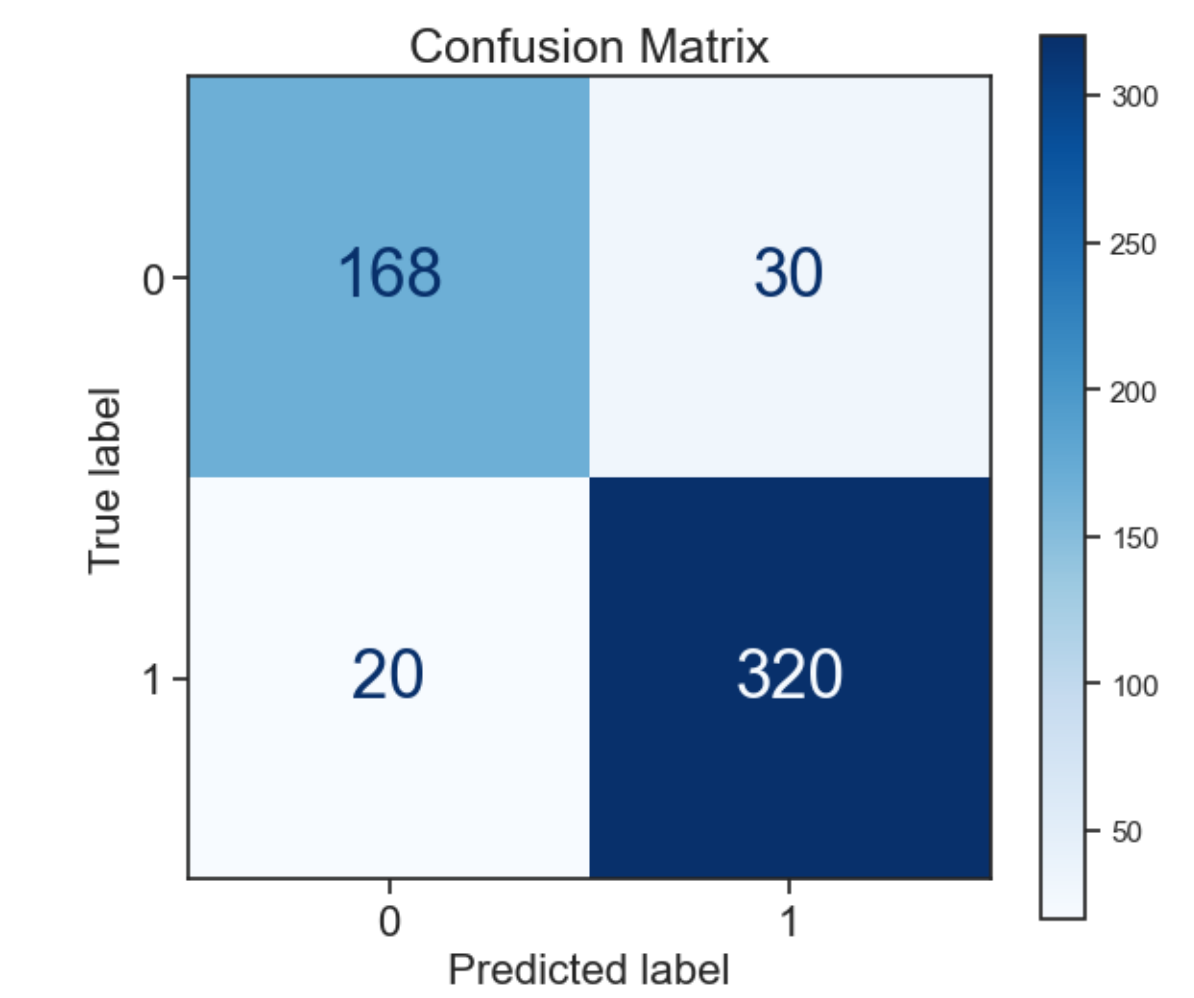
Metric	Train	Test
AUC	0.9999	0.9636
accuracy	0.9893	0.9071
balanced accuracy	0.9855	0.8948
recall_(sens)	1.0	0.9412
specificity	0.971	0.8485
precision	0.9833	0.9143
f1_score	0.9893	0.9065

Protein: P03372 - metric: SoftVotingXGBoostEnsemble
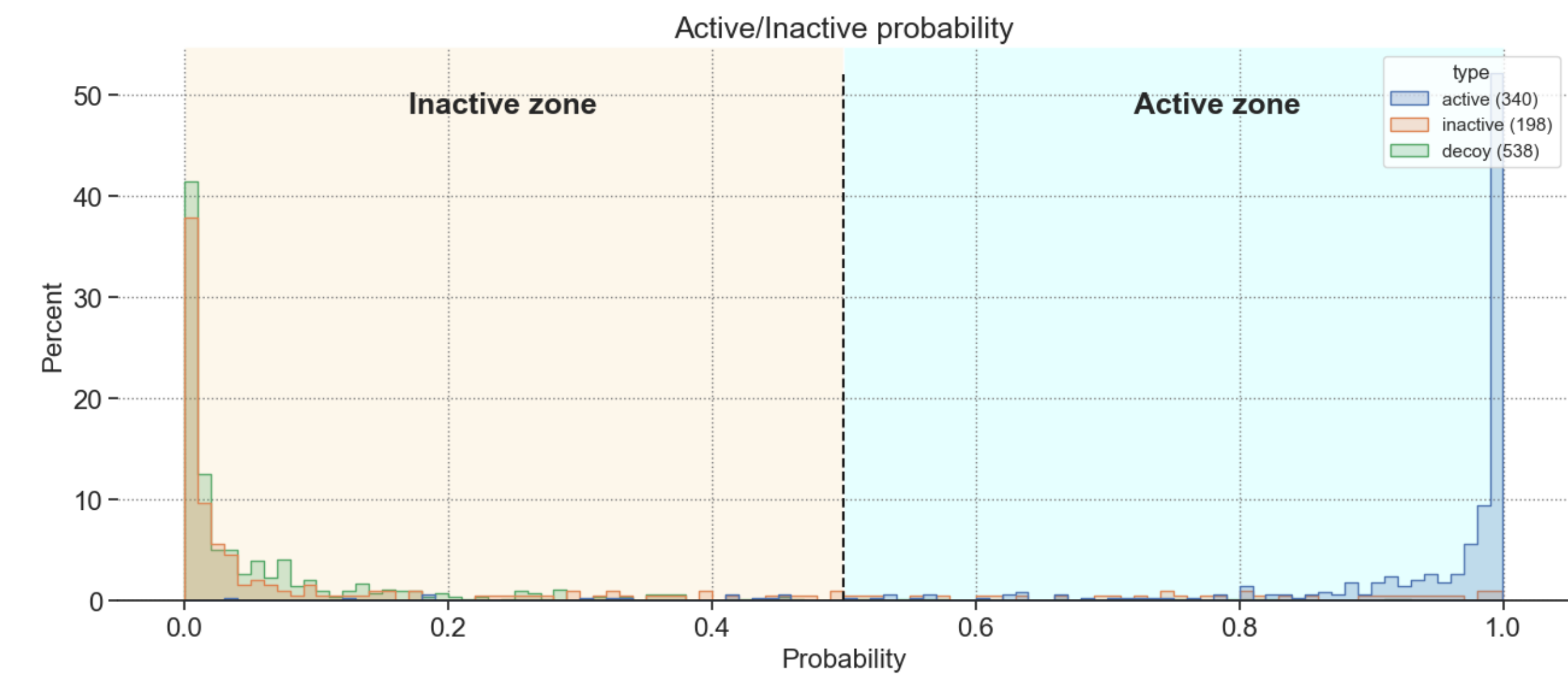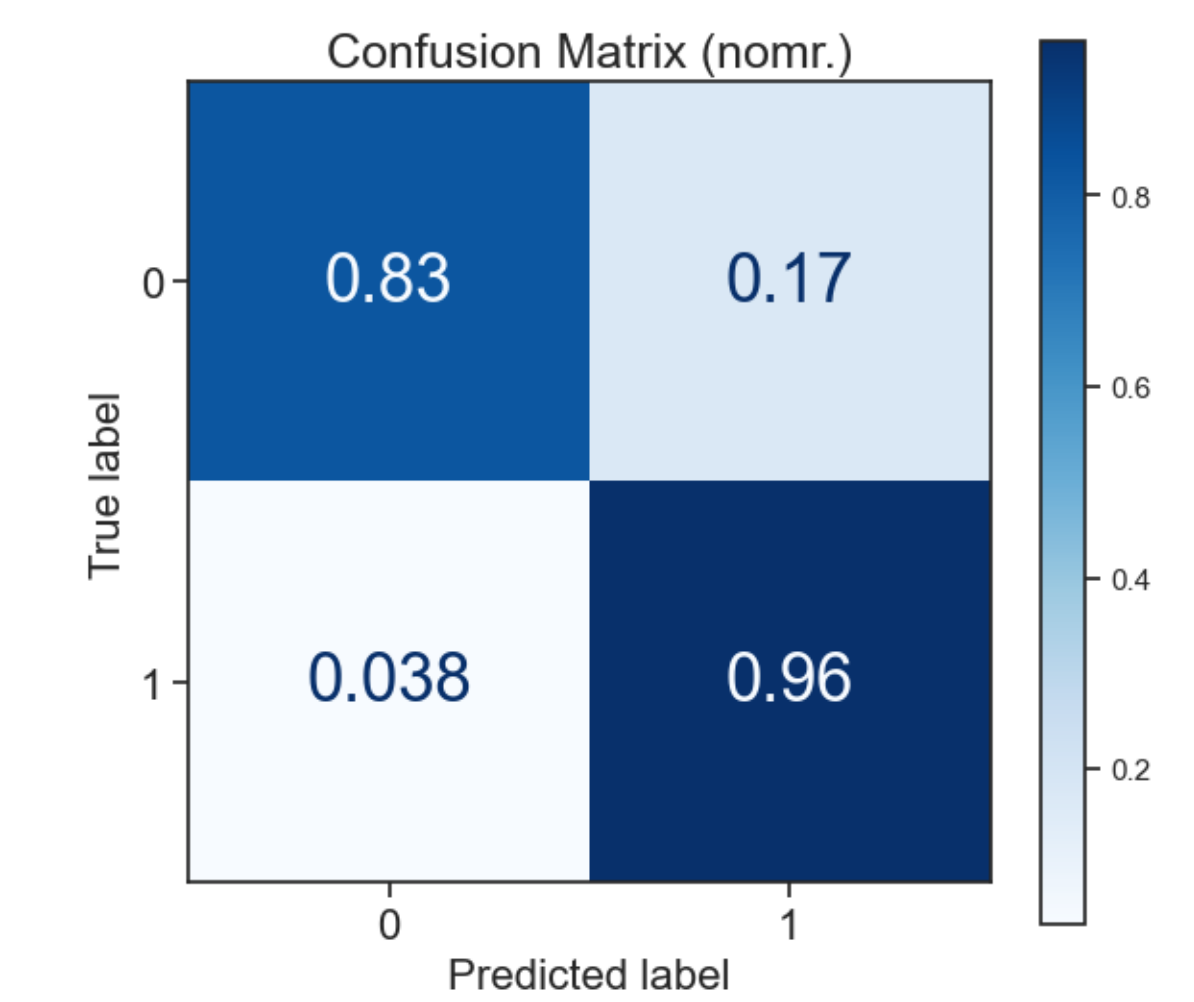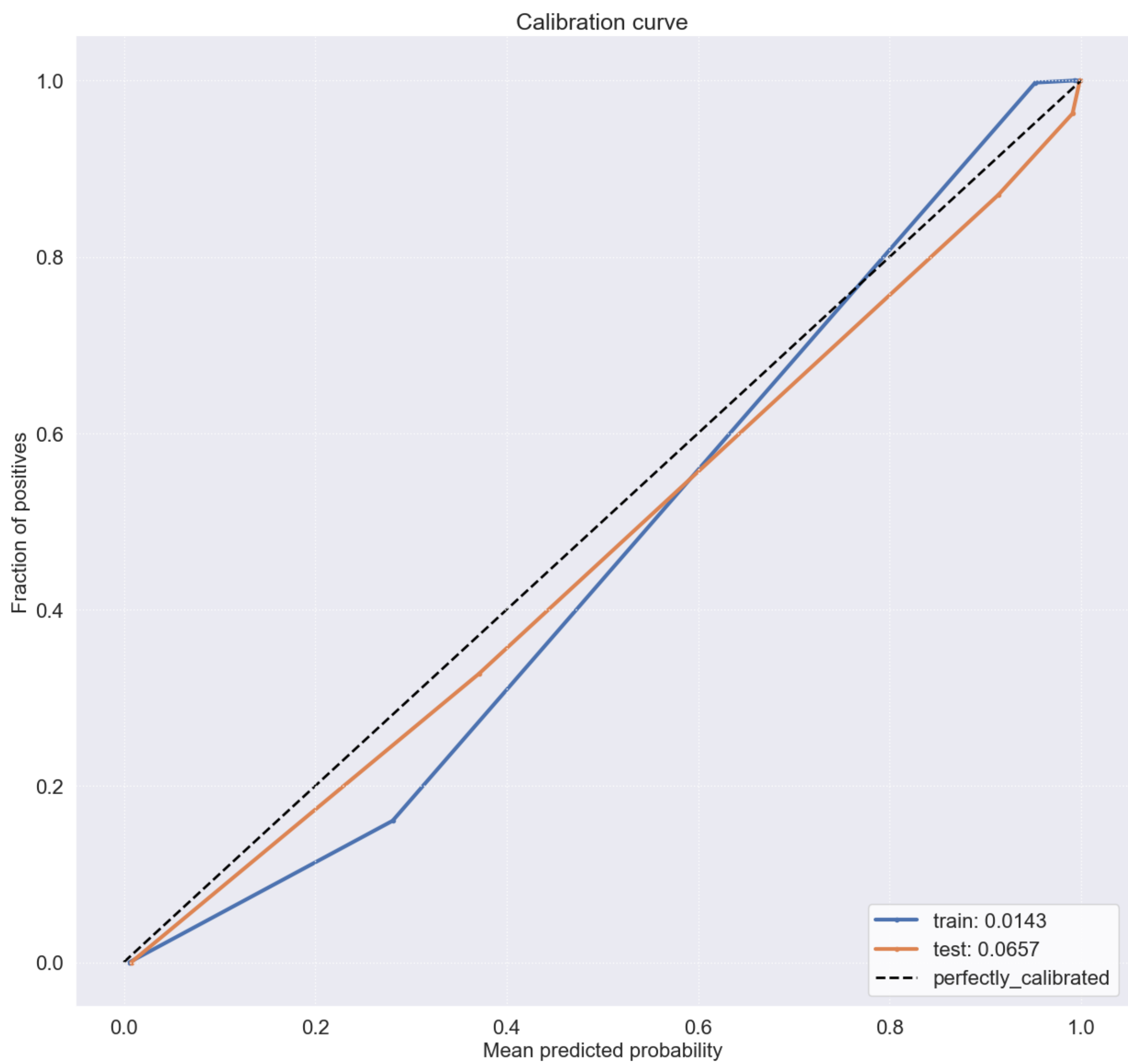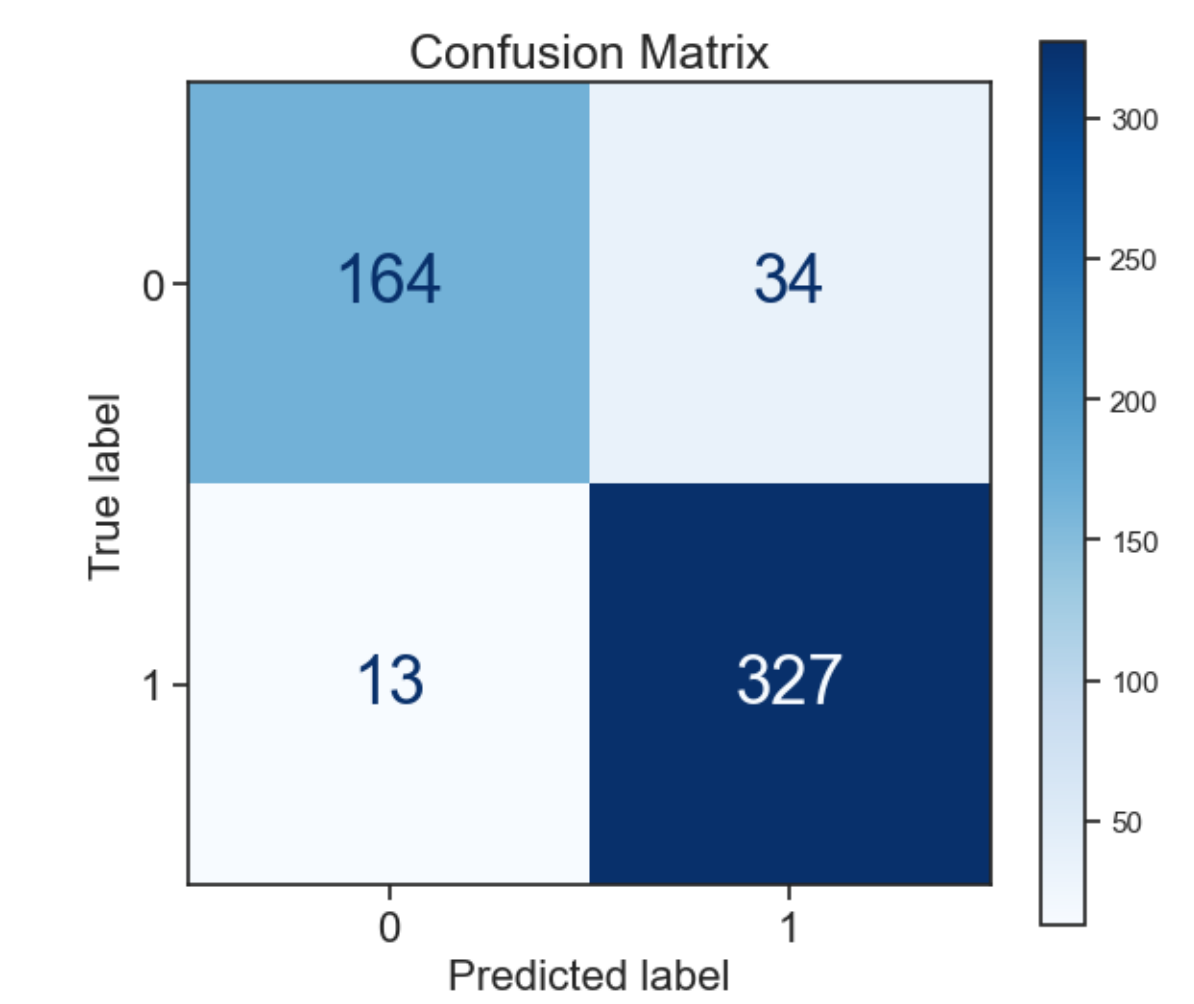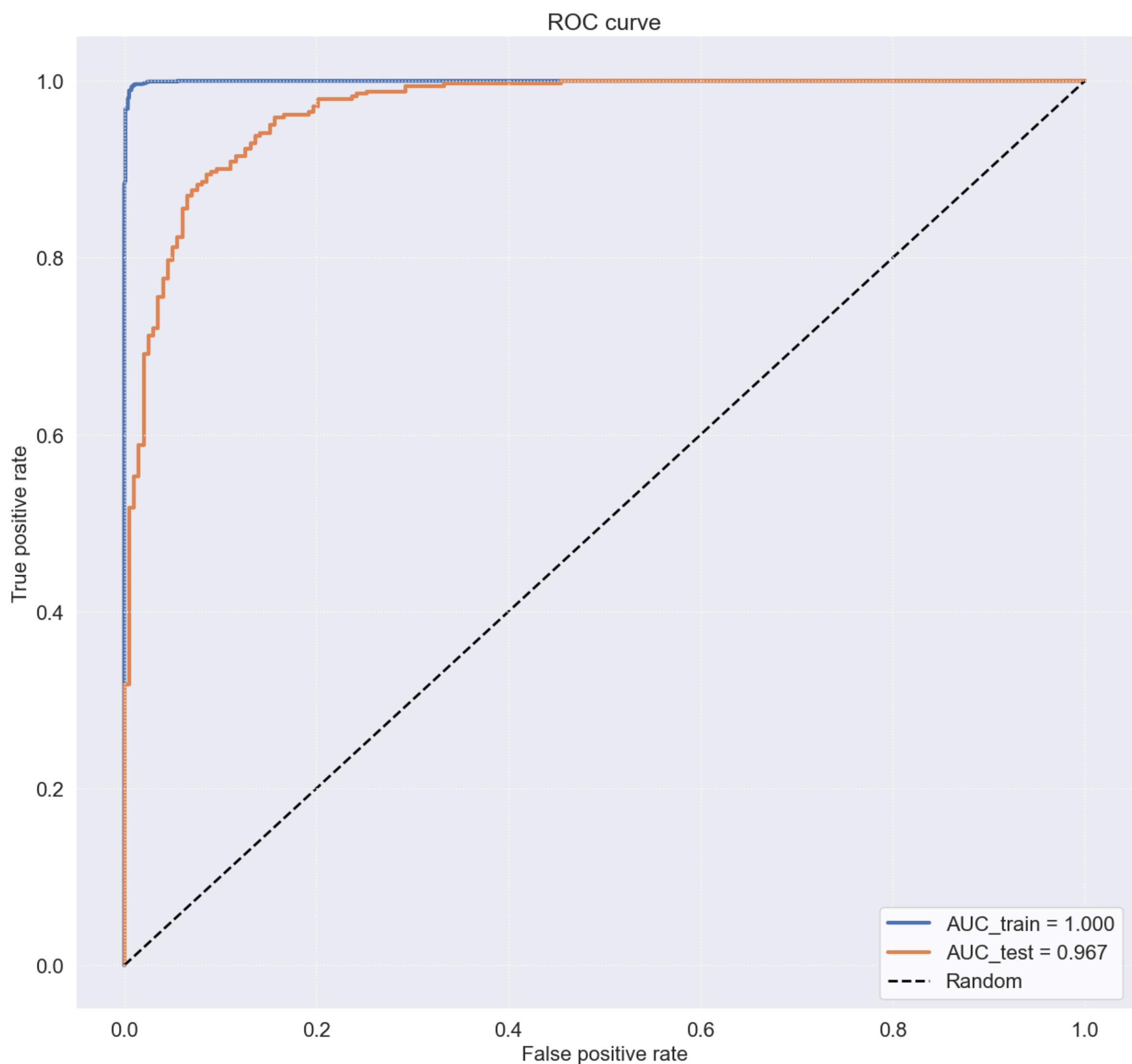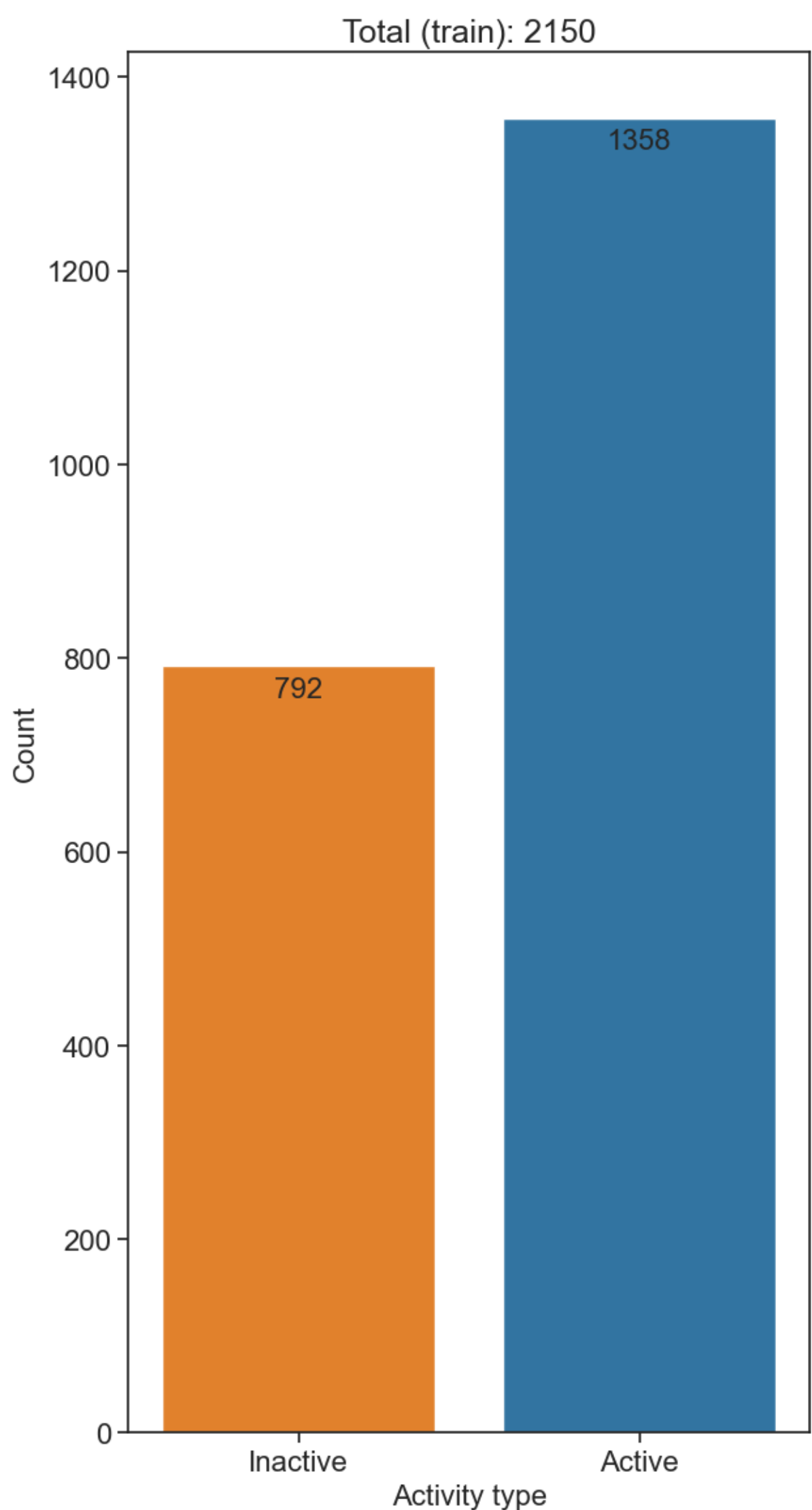
Metric	Train	Test
AUC	0.9996	0.9674
accuracy	0.9828	0.9126
balanced accuracy	0.9769	0.895
recall_(sens)	0.9993	0.9618
specificity	0.9545	0.8283
precision	0.9742	0.9058
f1_score	0.9827	0.9115