Protein: Q01959 - metric: balanced_accuracy

Total (train): 1601

Active: 731
Inactive: 870

x-axis: Activity type
y-axis: Count

ROC curve

- AUC_train = 0.998
- AUC_test = 0.966
- Random

x-axis: False positive rate
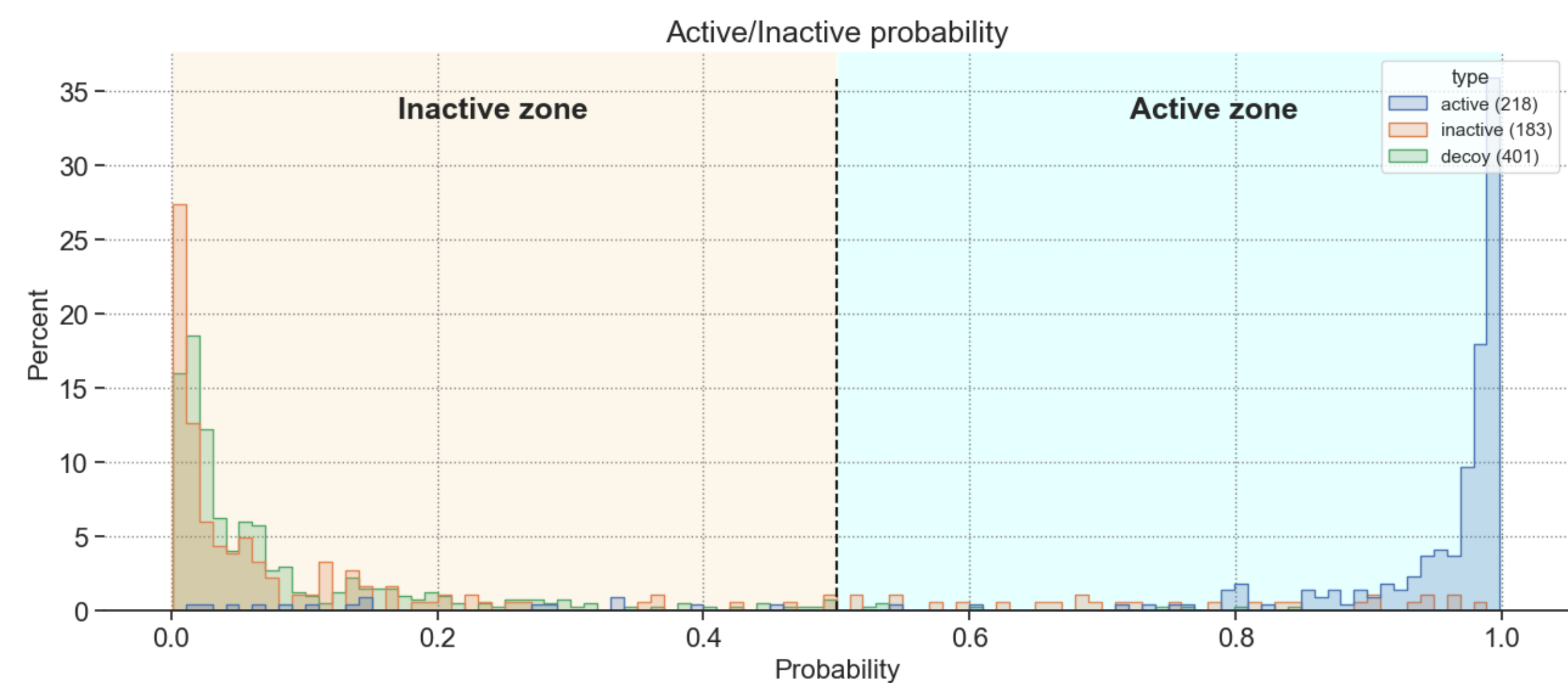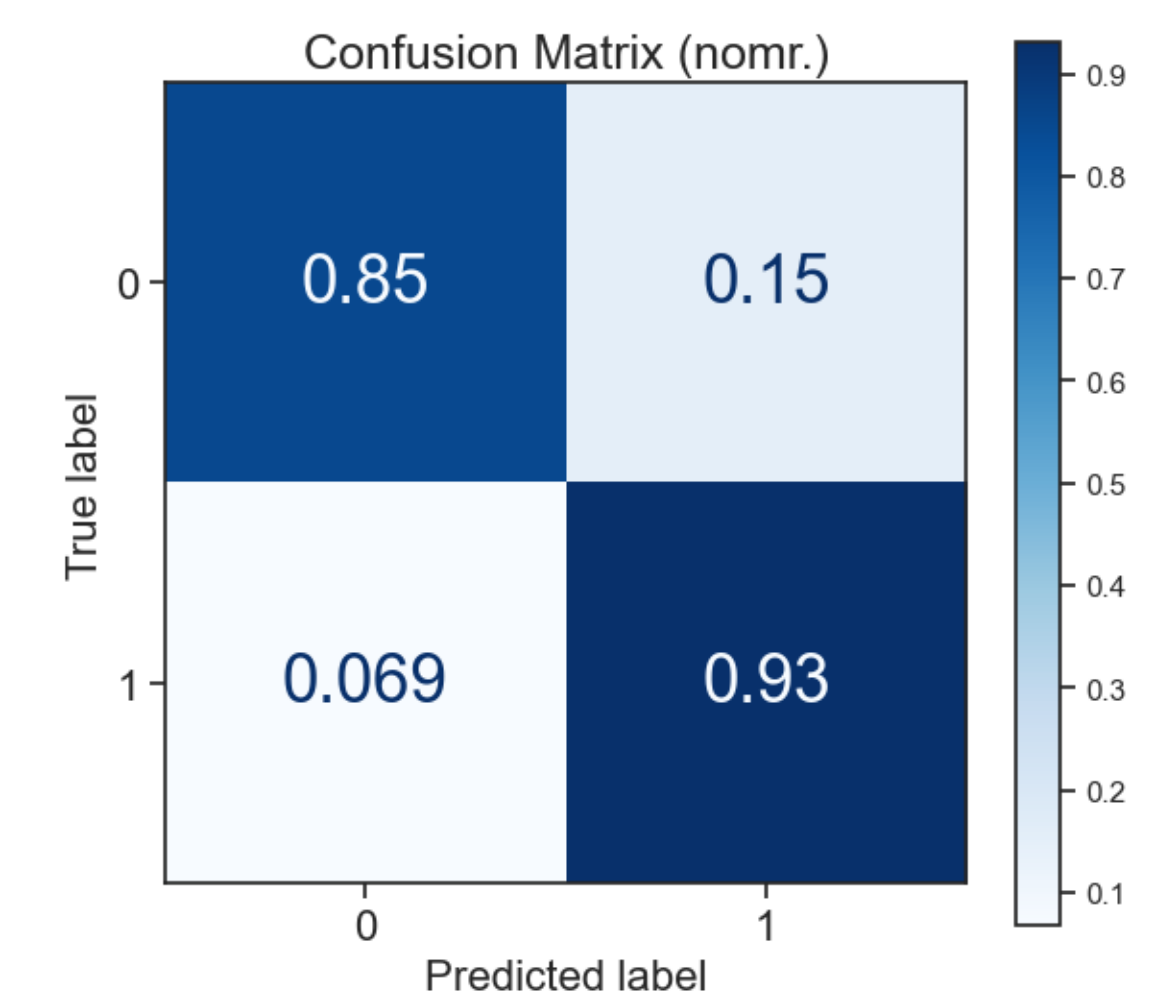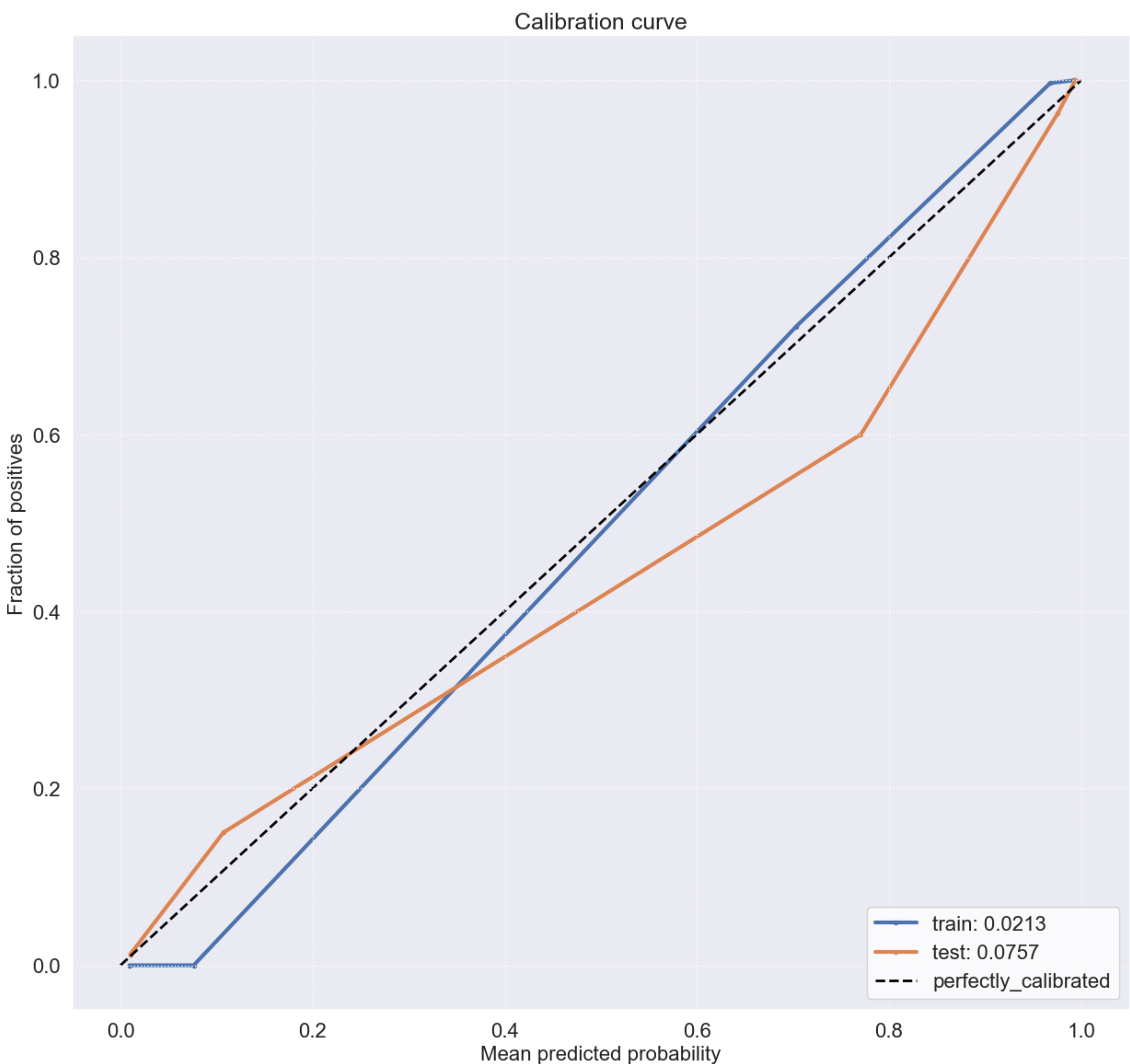y-axis: True positive rate

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2160)

x-axis: Score balanced_accuracy (Test)
y-axis: Difference in Train-Test balanced_accuracy
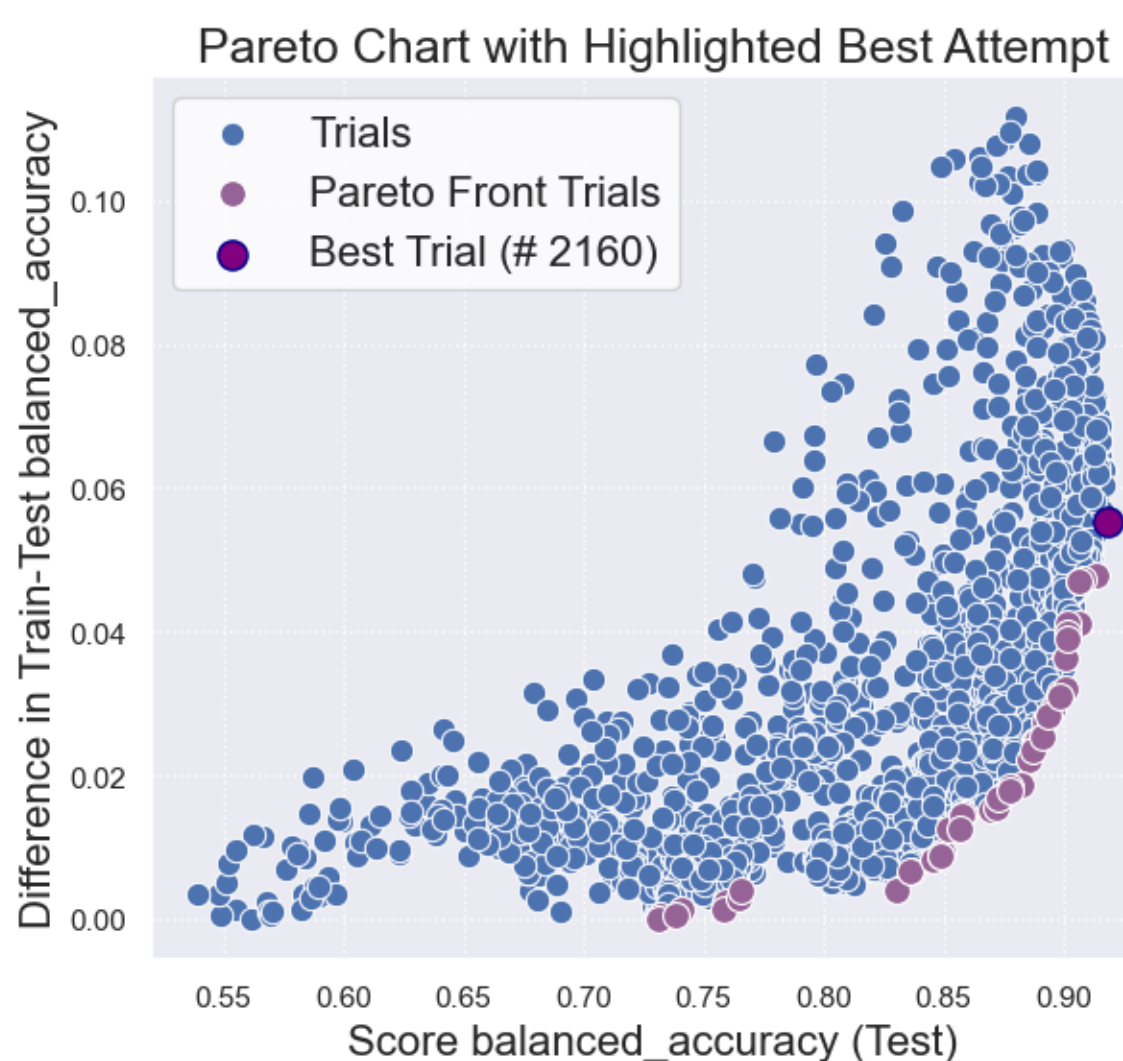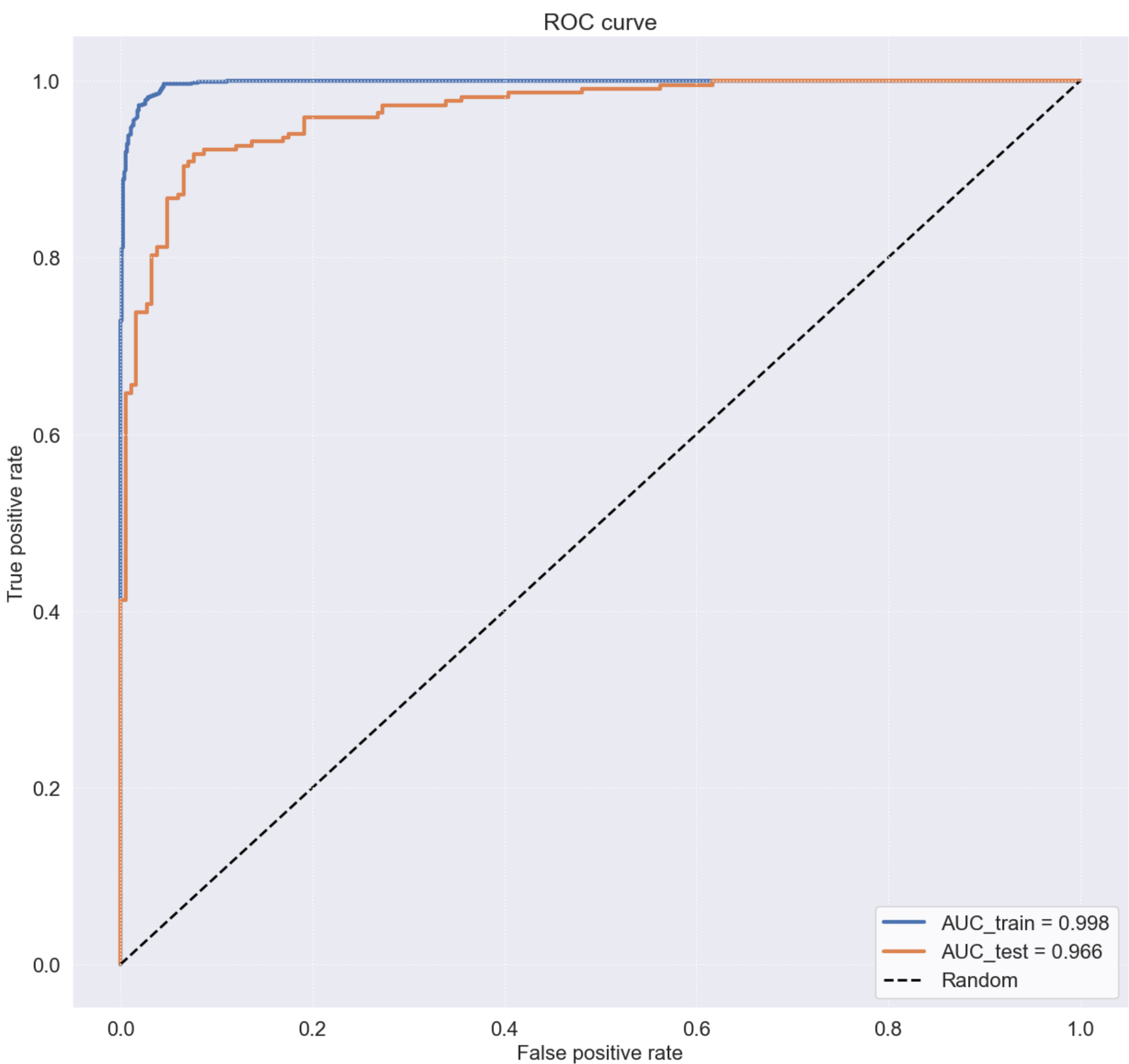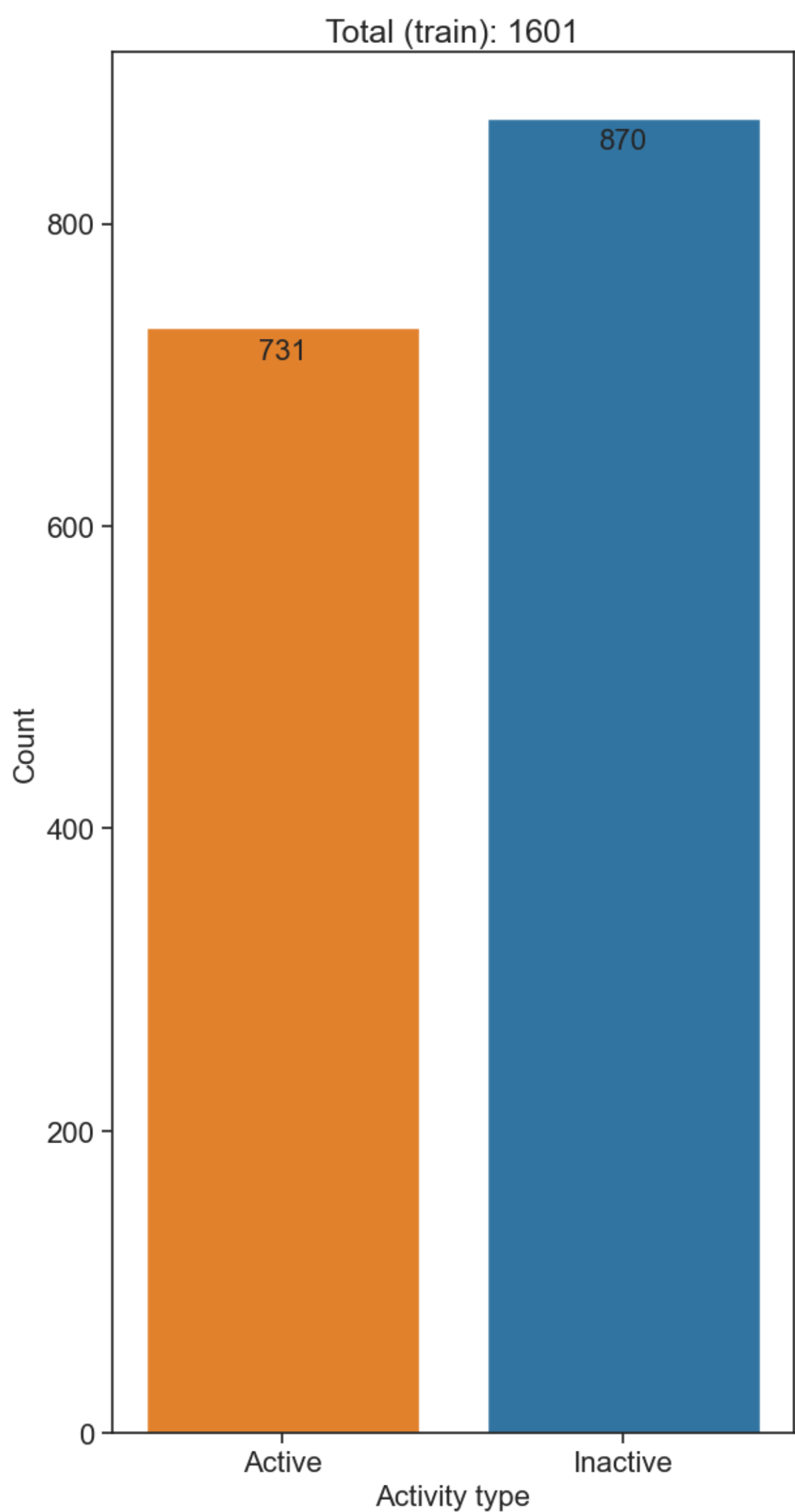
XGBoost Parameters	Value
alpha	0.8002
colsample_bytree	0.5899
gamma	0.7535
lambda	4.5497
learning_rate	0.0627
max_depth	9.0
min_child_weight	1.0
n_estimators	501.0
reg_alpha	0.6631
reg_lambda	0.1098
scale_pos_weight	1.2071
subsample	0.5866

Confusion Matrix

	0	1
0	155	28
1	15	203

True label / Predicted label

Confusion Matrix (nomr.)

	0	1
0	0.85	0.15
1	0.069	0.93

True label / Predicted label

Calibration curve

- train: 0.0213
- test: 0.0757
- perfectly_calibrated

x-axis: Mean predicted probability
y-axis: Fraction of positives

Active/Inactive probability

- active (218)
- inactive (183)
- decoy (401)

Inactive zone / Active zone

x-axis: Probability
y-axis: Percent

Metric	Train	Test
AUC	0.9977	0.9657
accuracy	0.9738	0.8928
balanced accuracy	0.9727	0.8891
recall_(sens)	0.9851	0.9312
specificity	0.9603	0.847
precision	0.9673	0.8788
f1_score	0.9737	0.8923

Protein: Q01959 - metric: f1

Total (train): 1601

ROC curve

	AUC_train = 0.998
	AUC_test = 0.967
	Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 1329)
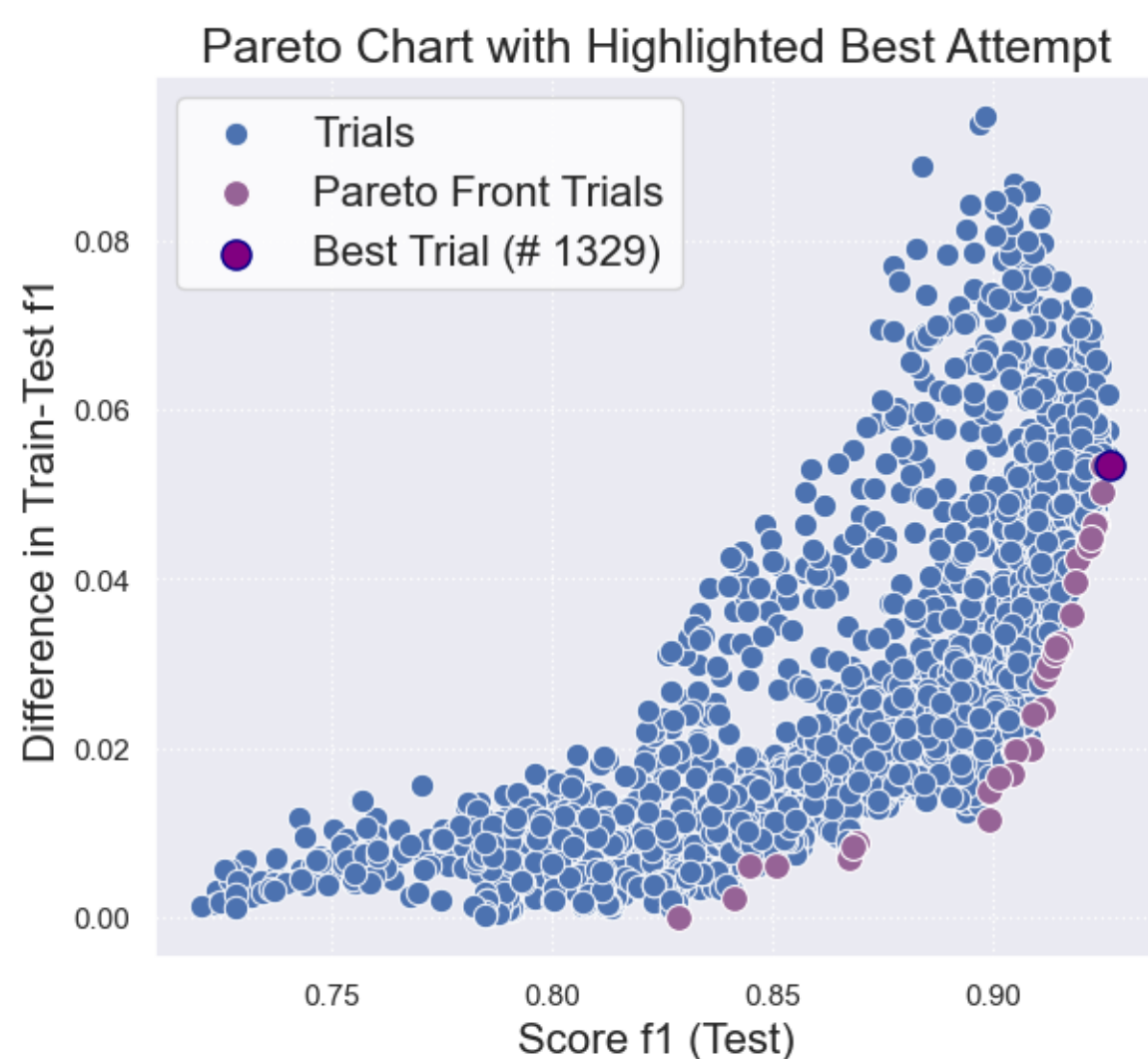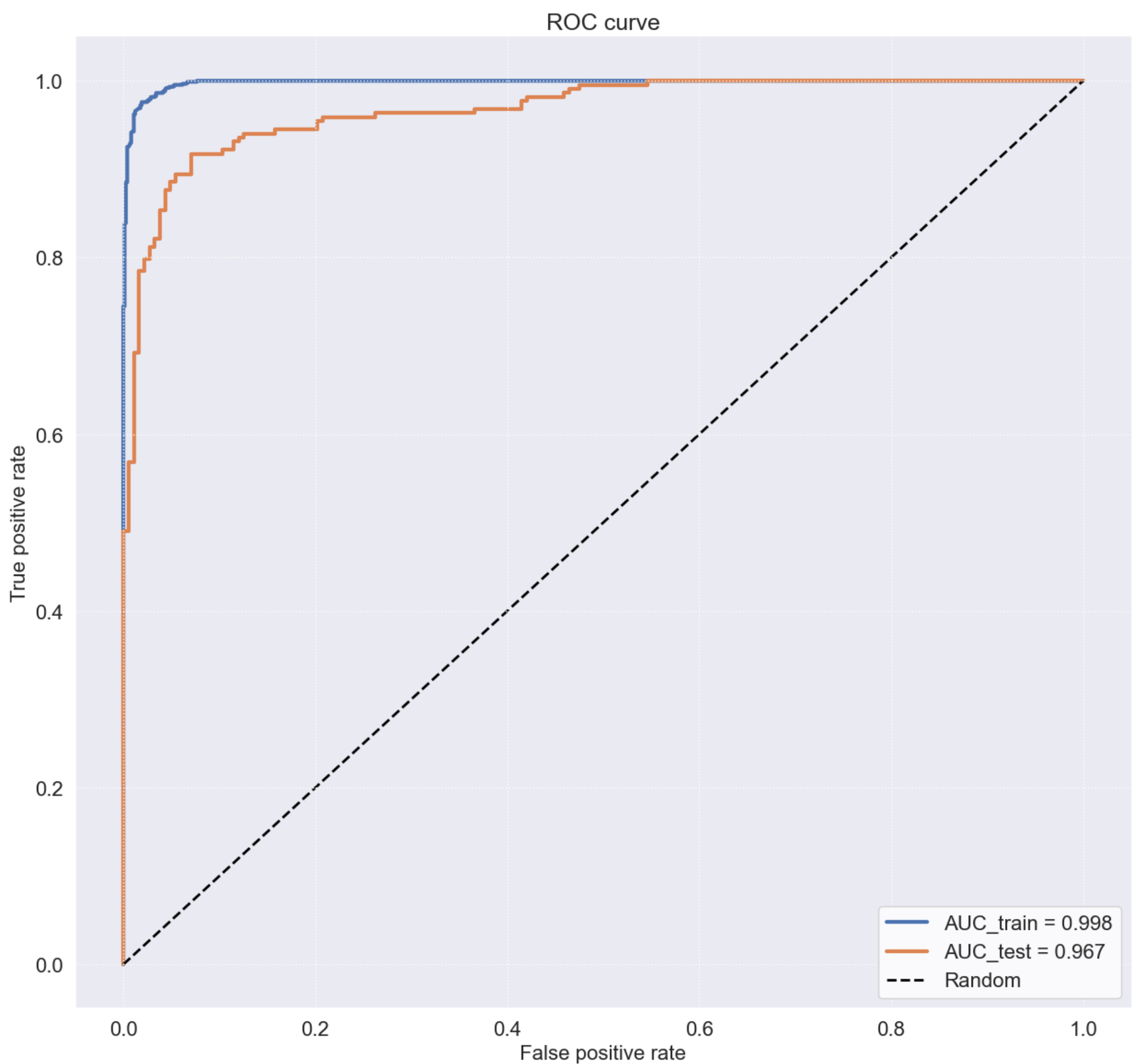
XGBoost Parameters	Value
alpha	4.8055
colsample_bytree	0.5345
gamma	0.159
lambda	3.1865
learning_rate	0.0183
max_depth	10.0
min_child_weight	3.0
n_estimators	948.0
reg_alpha	0.0923
reg_lambda	0.6445
scale_pos_weight	1.1028
subsample	0.9159

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

	train: 0.0207
	test: 0.0721
	perfectly_calibrated

Active/Inactive probability

- active (218)
- inactive (183)
- decoy (401)
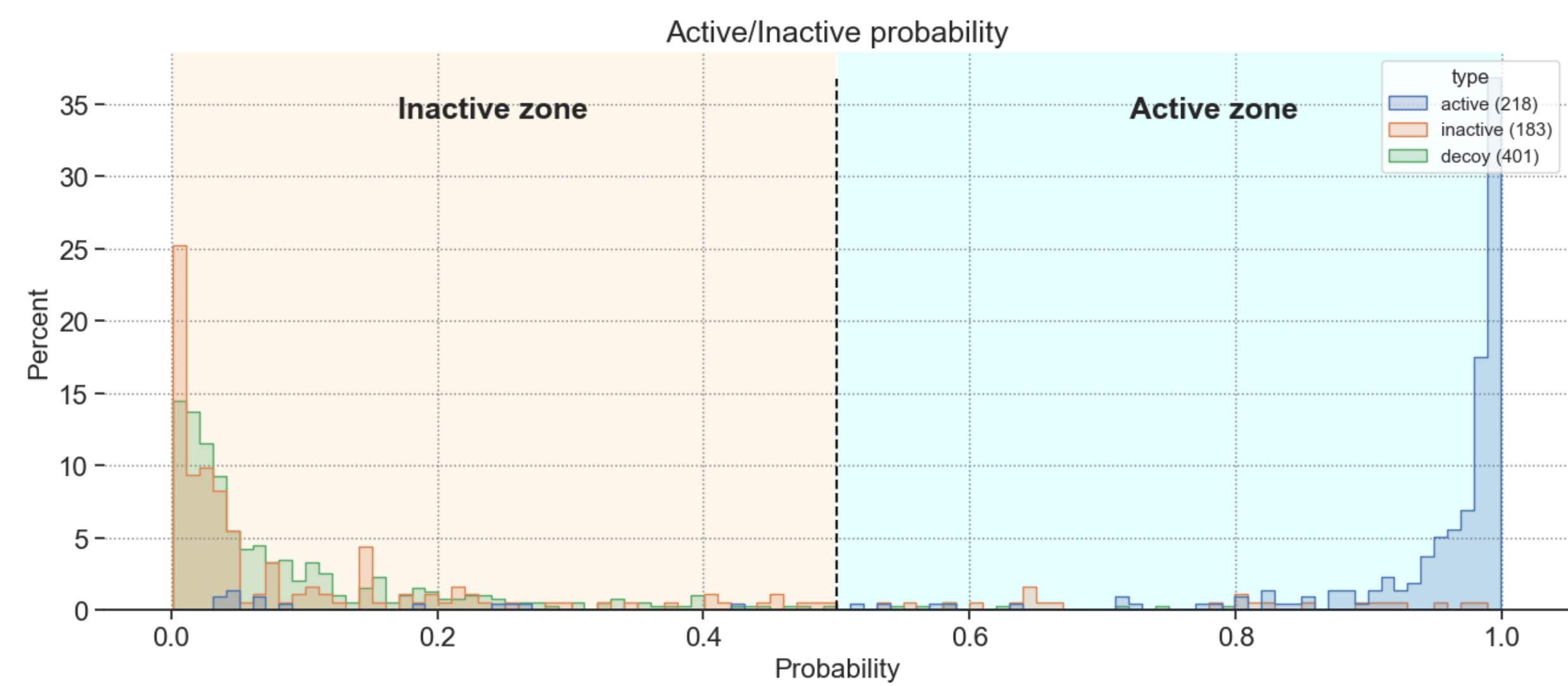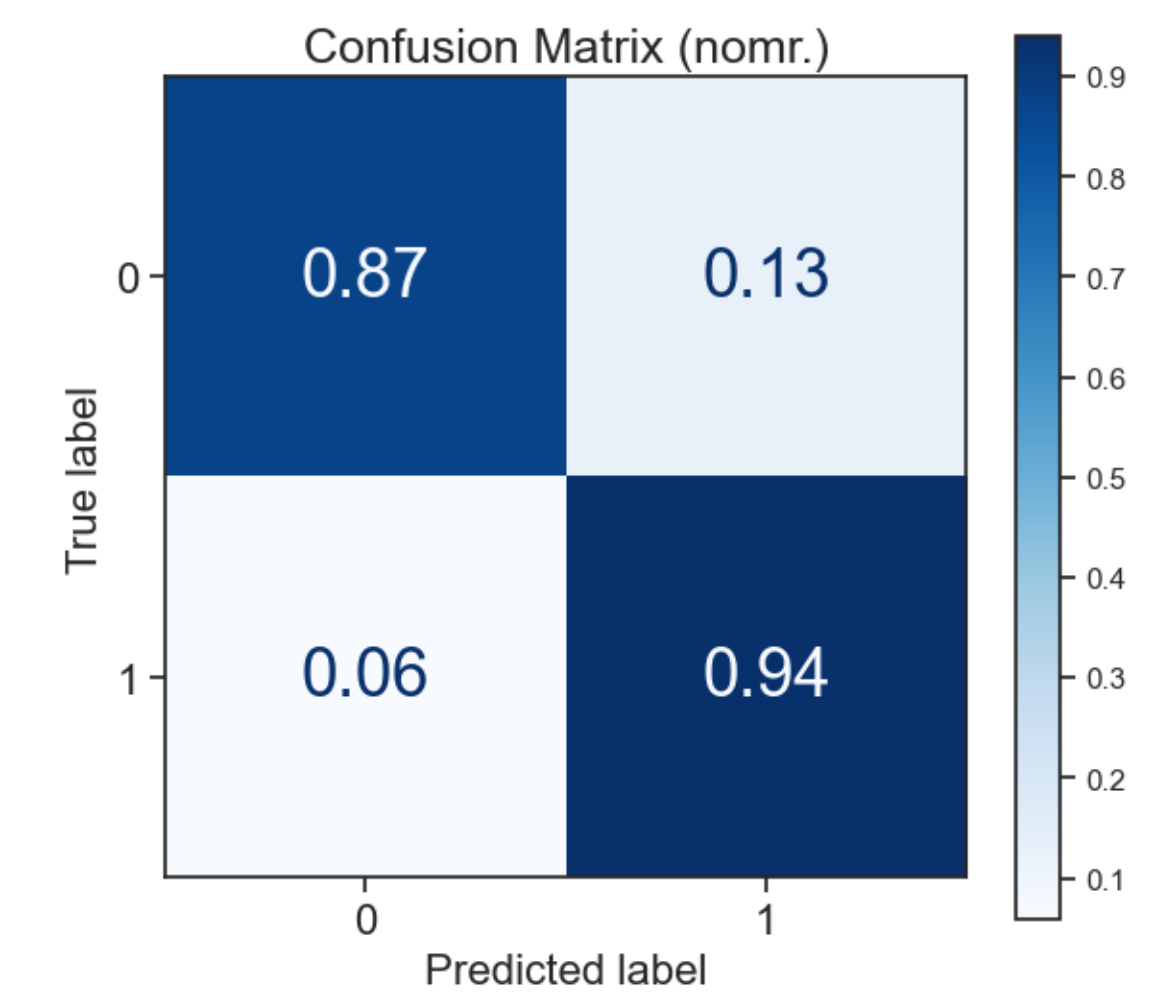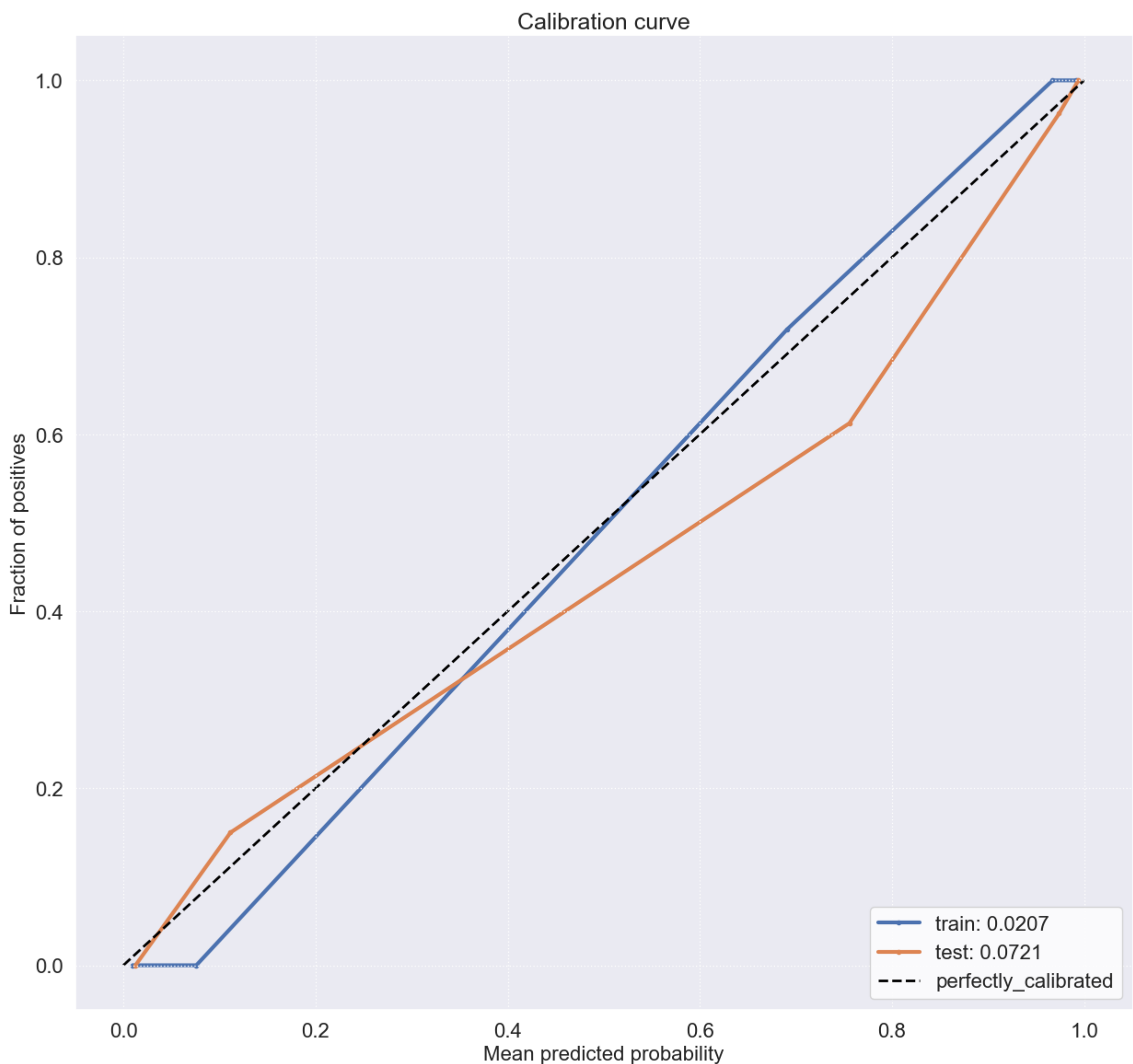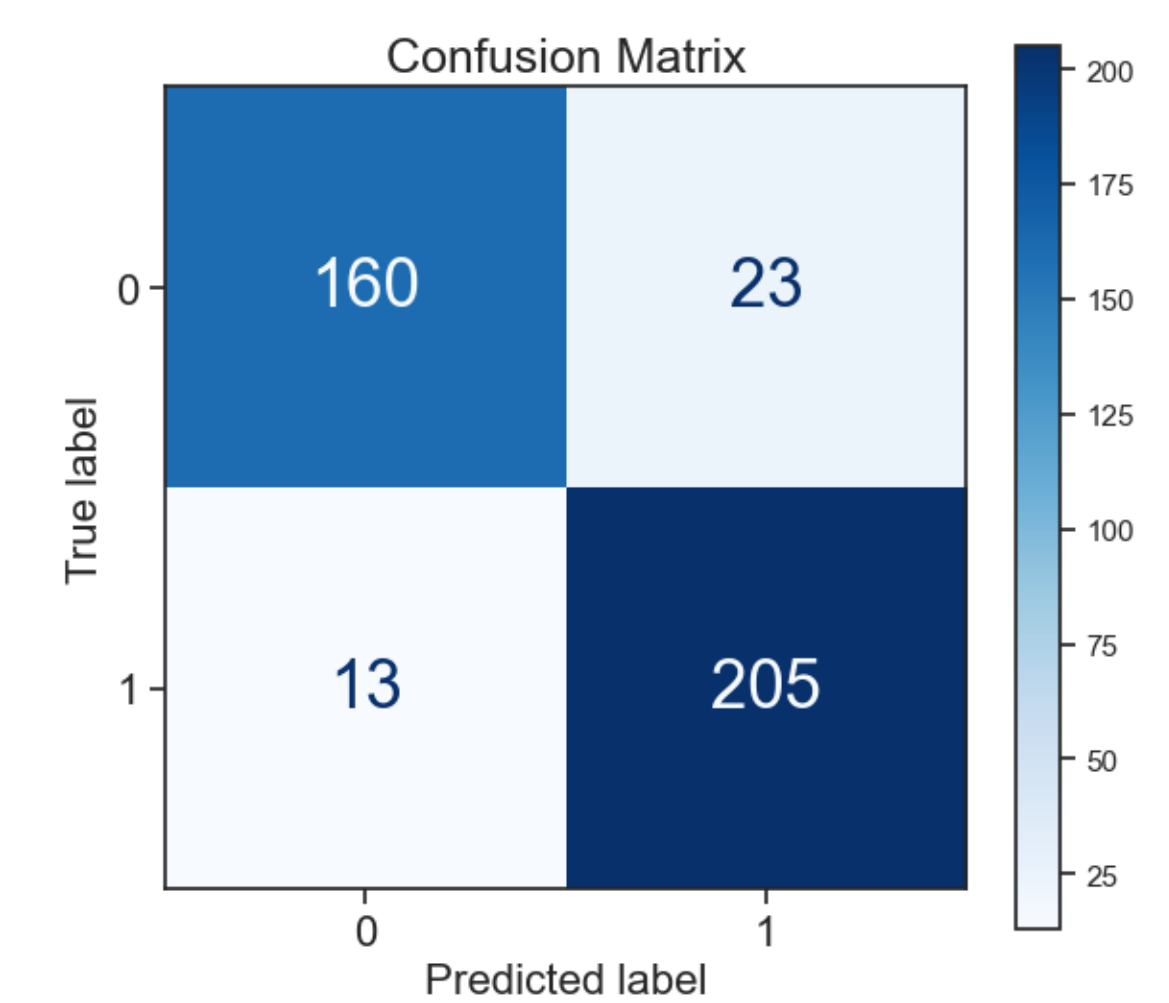
Metric	Train	Test
AUC	0.998	0.9672
accuracy	0.9756	0.9102
balanced accuracy	0.9749	0.9073
recall_(sens)	0.9839	0.9404
specificity	0.9658	0.8743
precision	0.9716	0.8991
f1_score	0.9756	0.91

Protein: Q01959 - metric: precision

Total (train): 1601

Active: 731
Inactive: 870

ROC curve

AUC_train = 0.998
AUC_test = 0.968
Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2149)

X-axis: Score precision (Test)
Y-axis: Difference in Train-Test precision
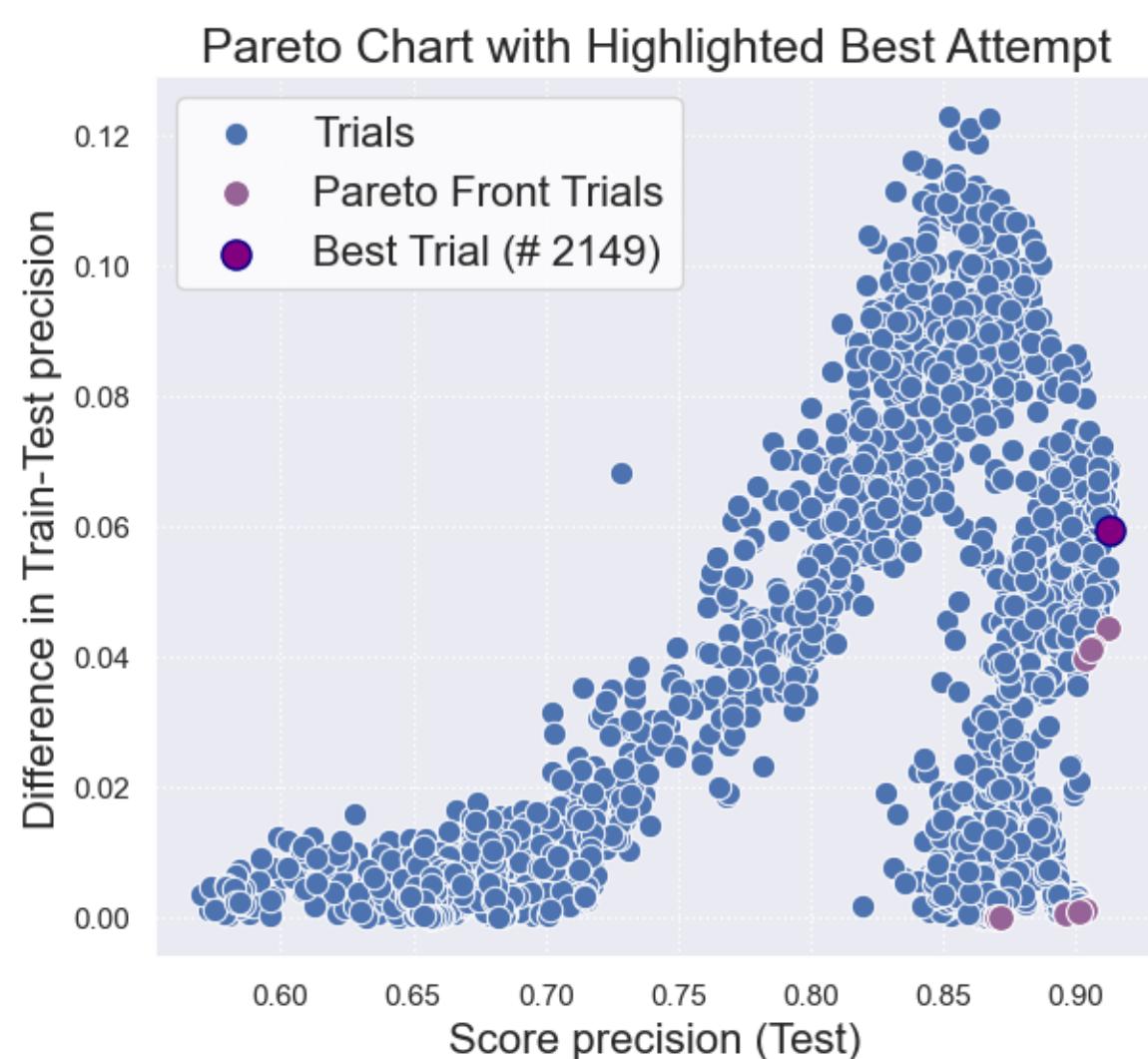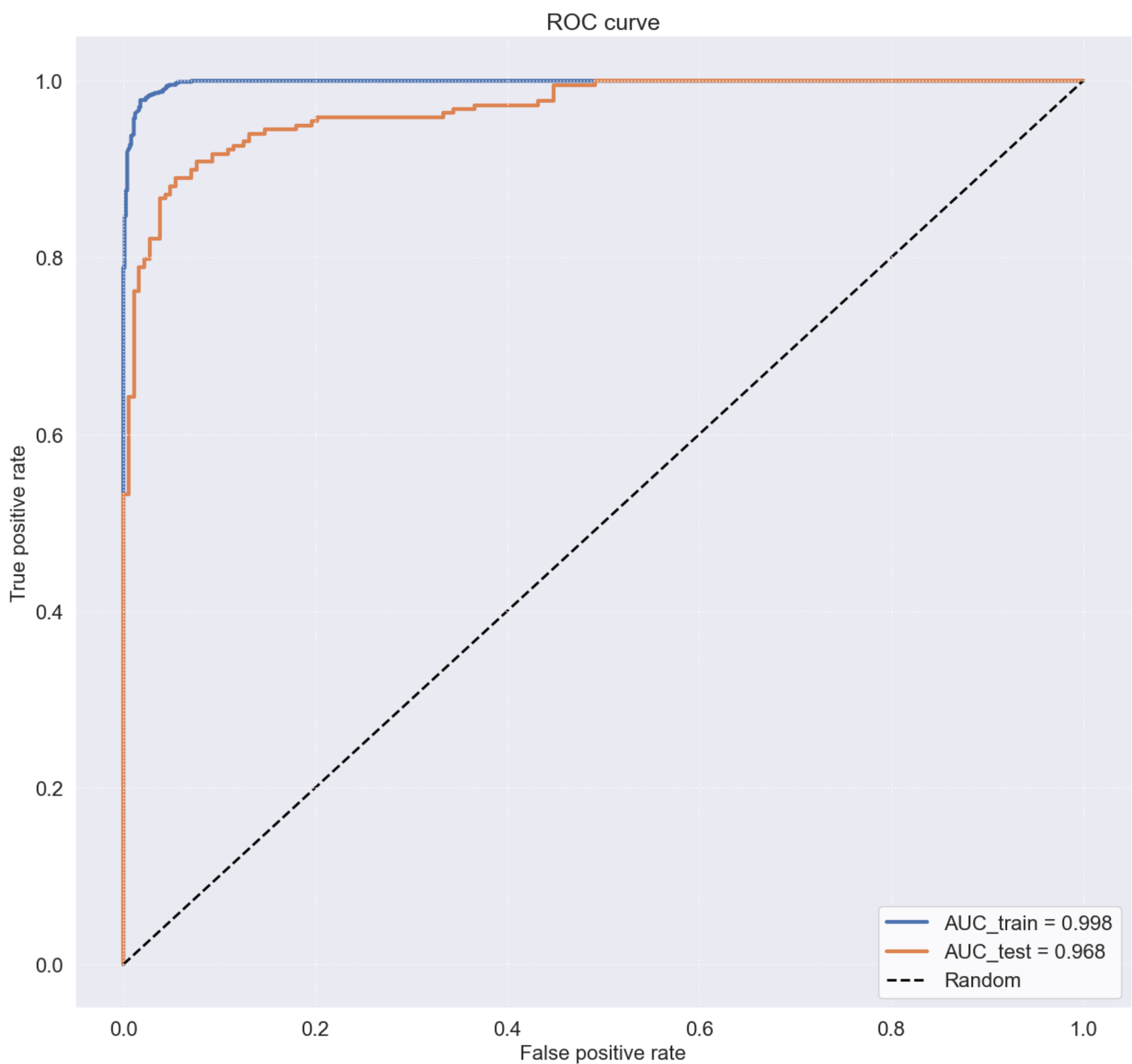
XGBoost Parameters	Value
alpha	3.4157
colsample_bytree	0.7817
gamma	0.261
lambda	3.5311
learning_rate	0.0425
max_depth	10.0
min_child_weight	3.0
n_estimators	602.0
reg_alpha	0.2882
reg_lambda	0.3216
scale_pos_weight	1.2392
subsample	0.9787

Confusion Matrix

	Predicted 0	Predicted 1
True 0	156	27
True 1	12	206

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.85	0.15
True 1	0.055	0.94

Calibration curve

train: 0.0199
test: 0.0738
perfectly_calibrated

Active/Inactive probability

Inactive zone | Active zone

type:
- active (218)
- inactive (183)
- decoy (401)
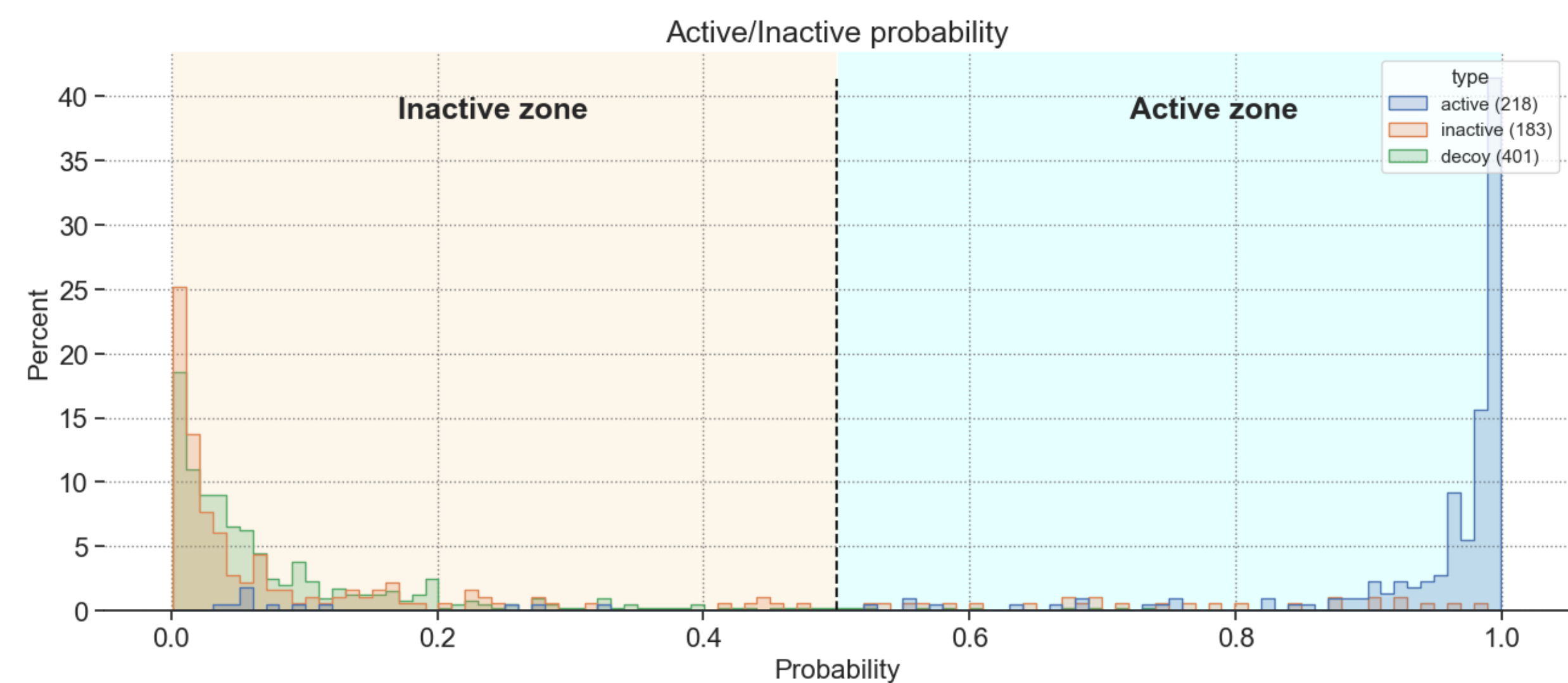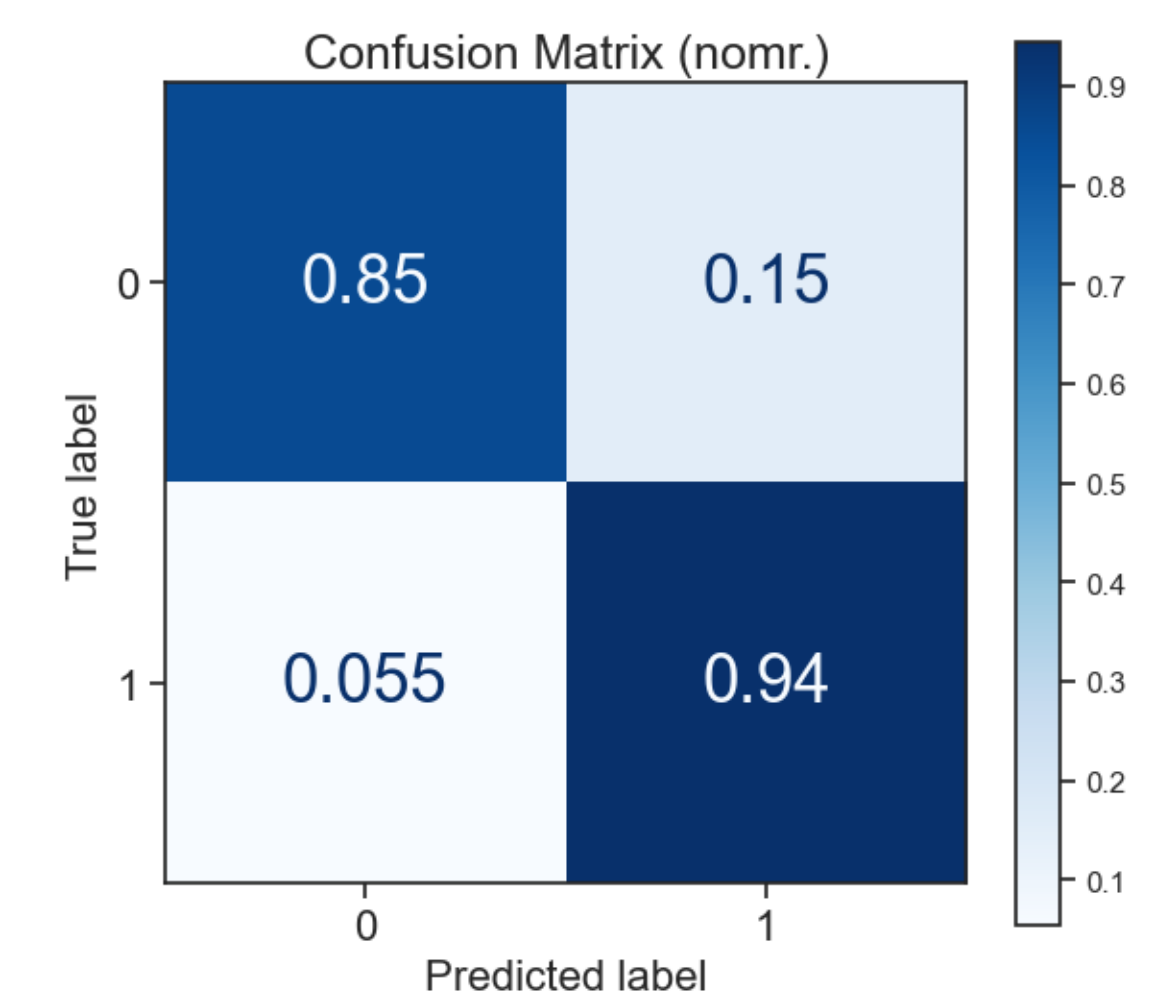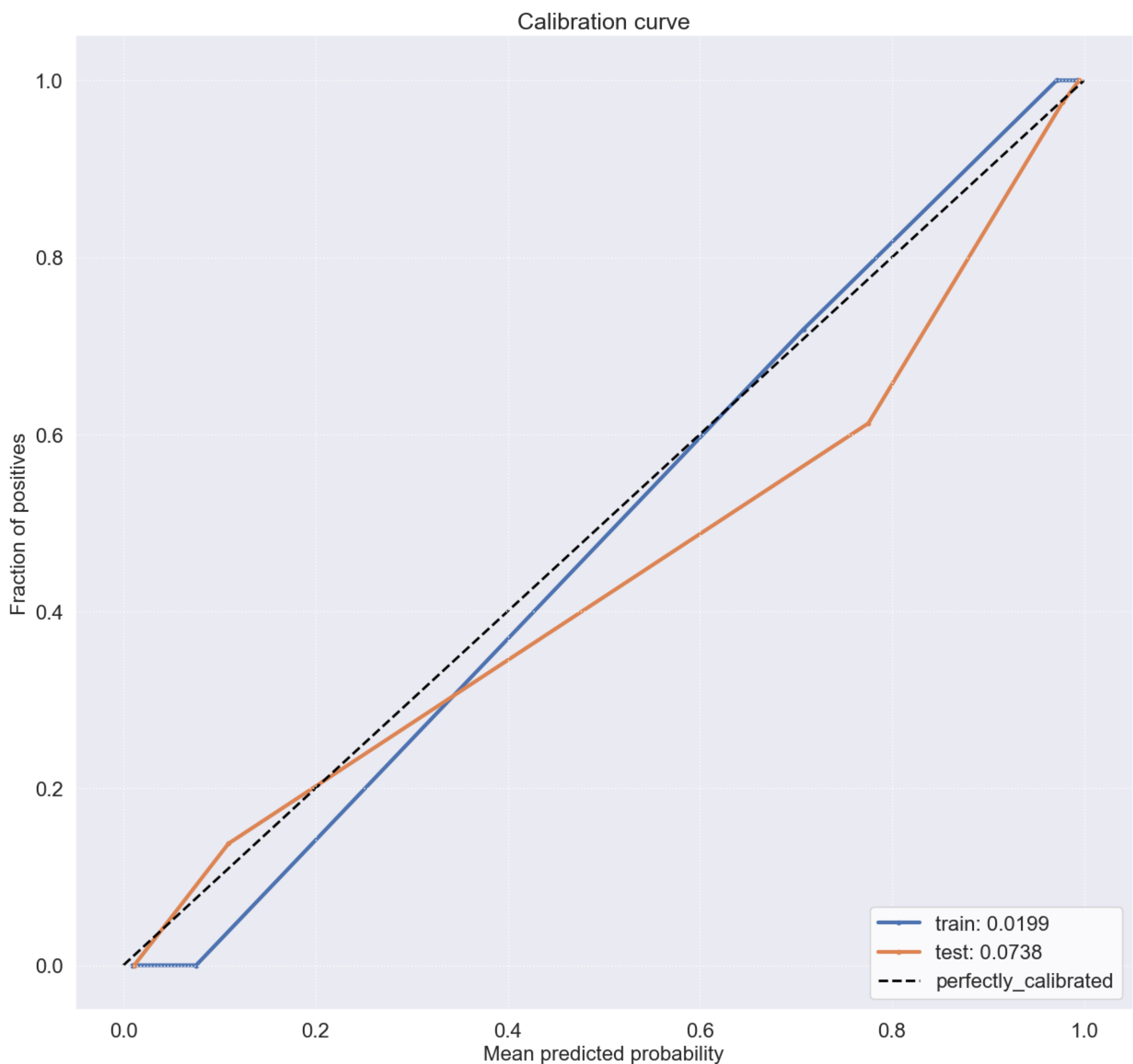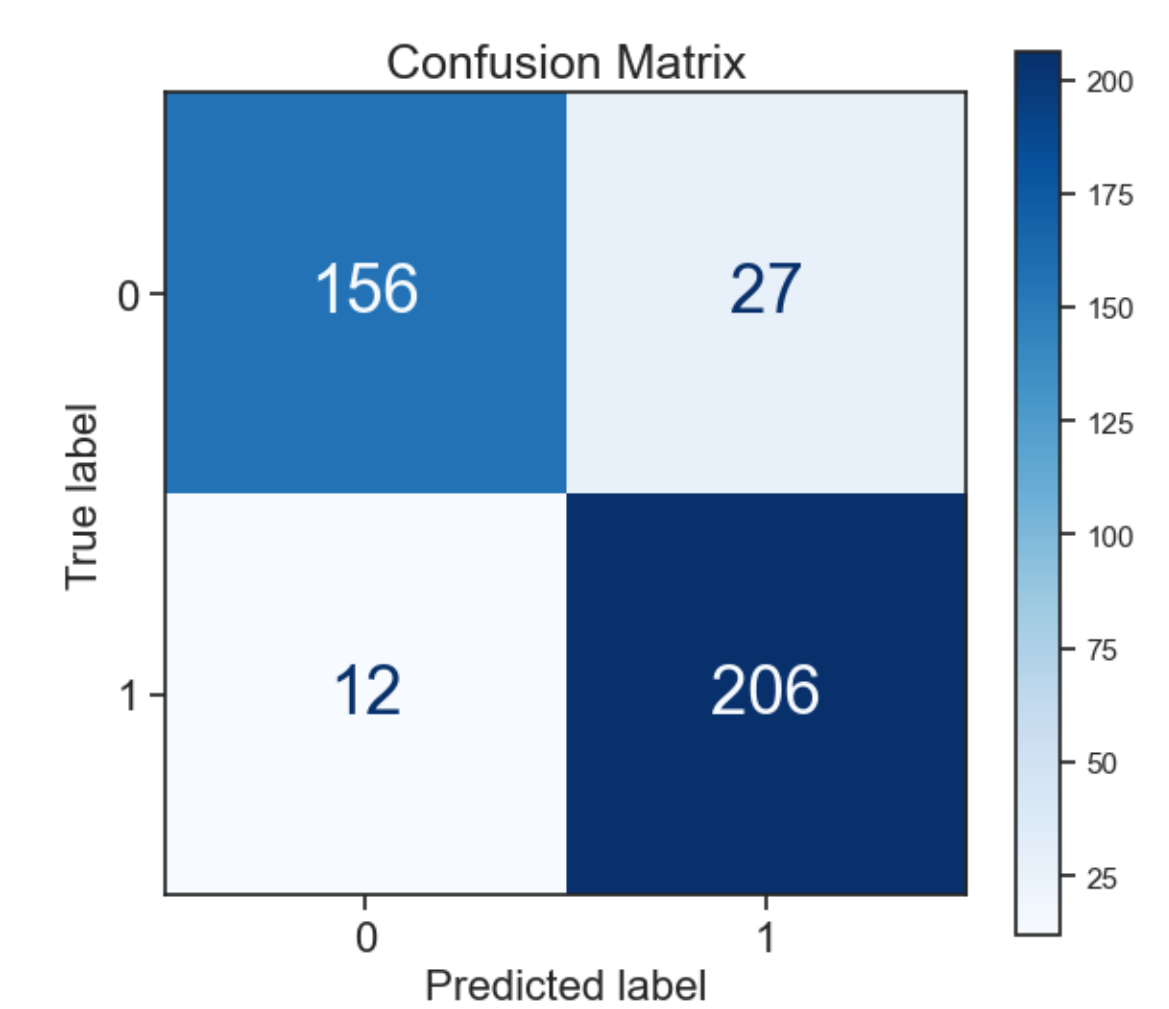
Metric	Train	Test
AUC	0.9981	0.9683
accuracy	0.9775	0.9027
balanced accuracy	0.9767	0.8987
recall_(sens)	0.9862	0.945
specificity	0.9672	0.8525
precision	0.9728	0.8841
f1_score	0.9775	0.9023

Protein: Q01959 - metric: SoftVotingXGBoostEnsemble
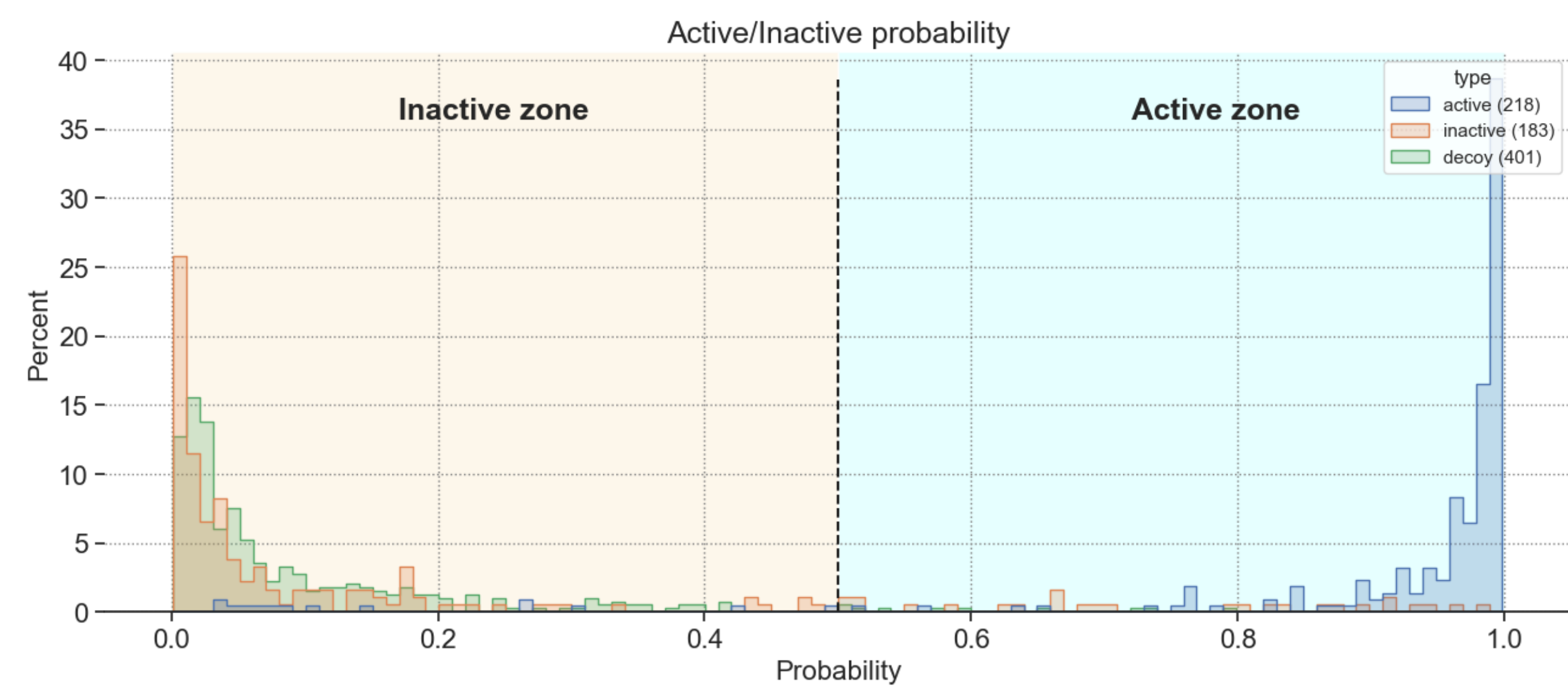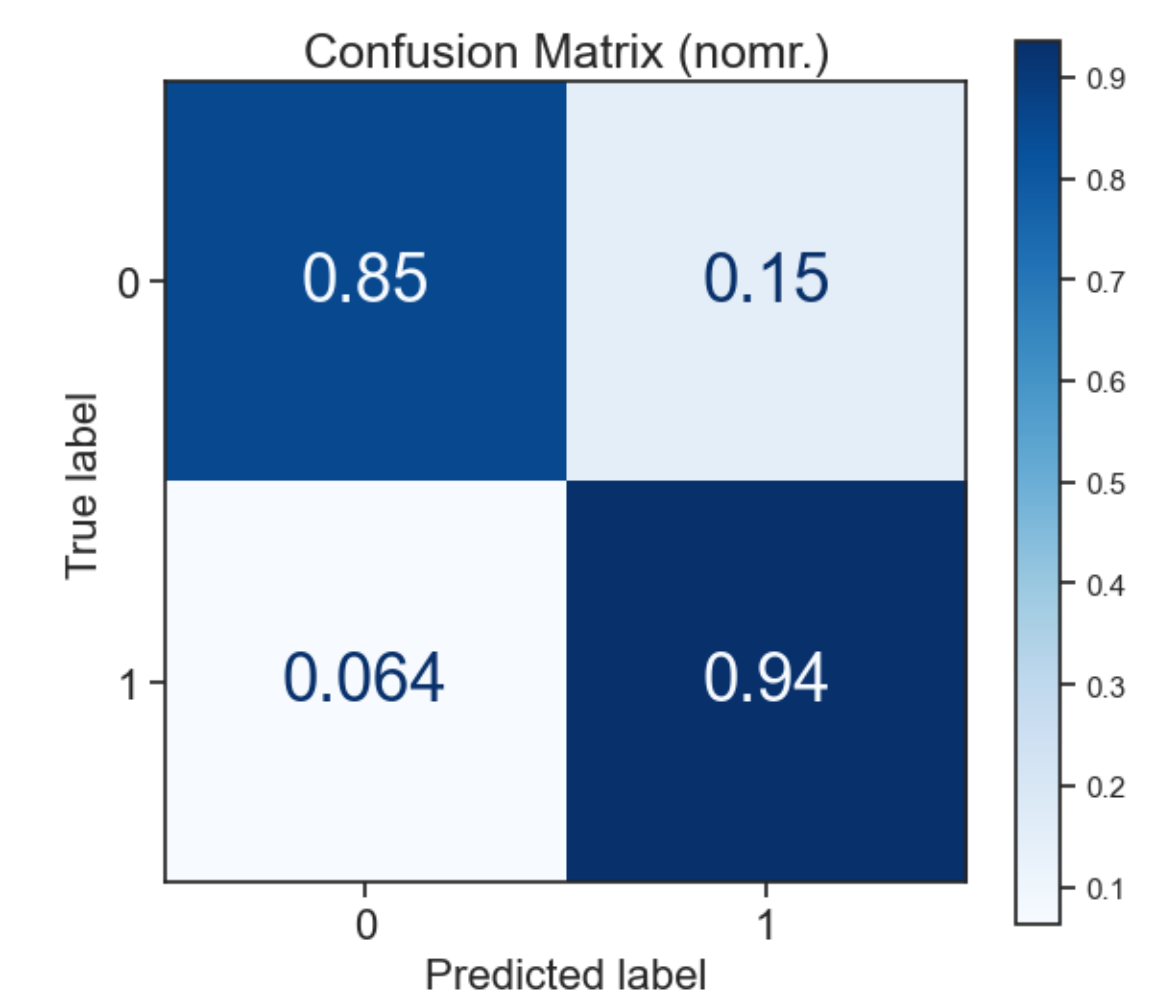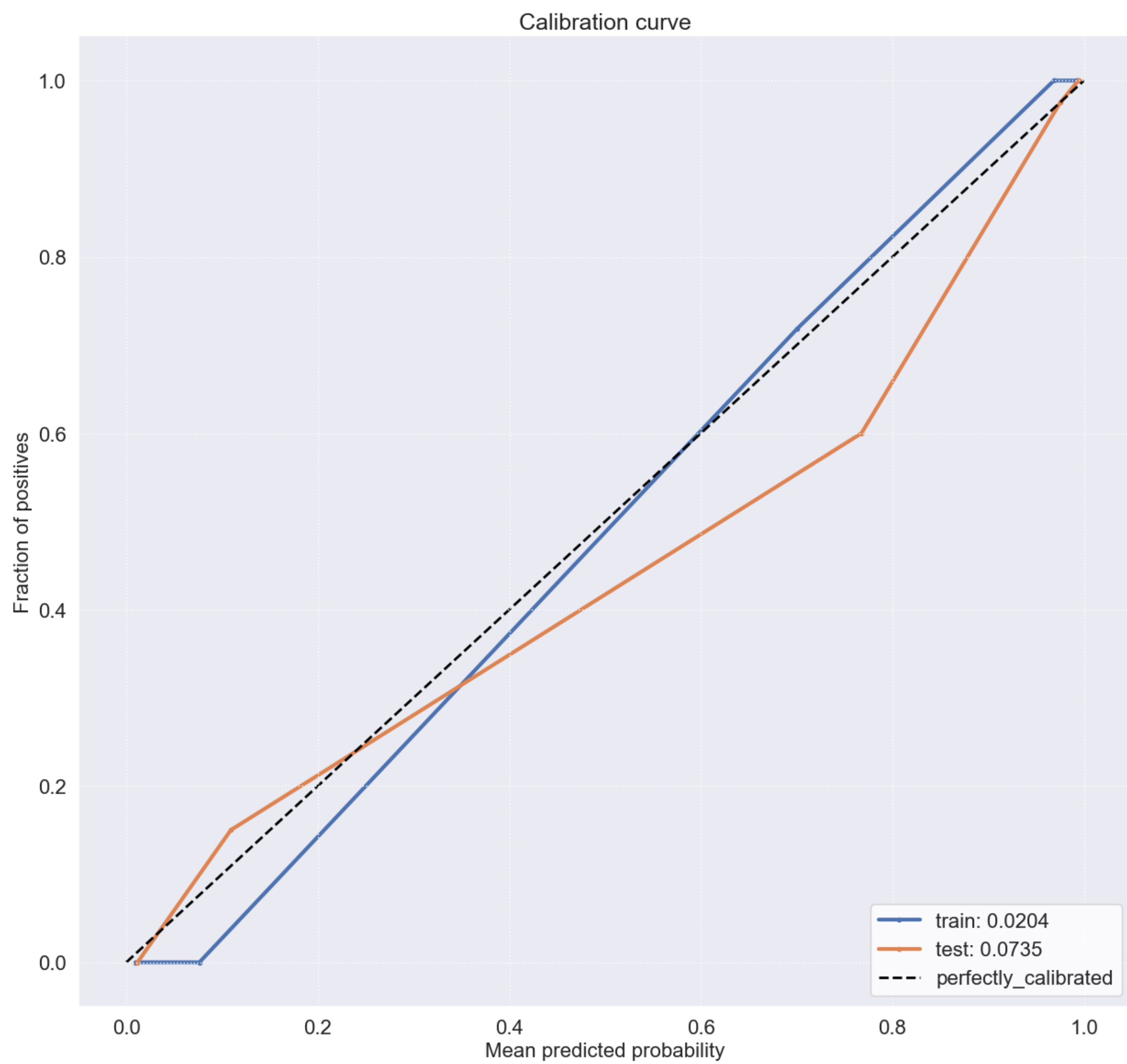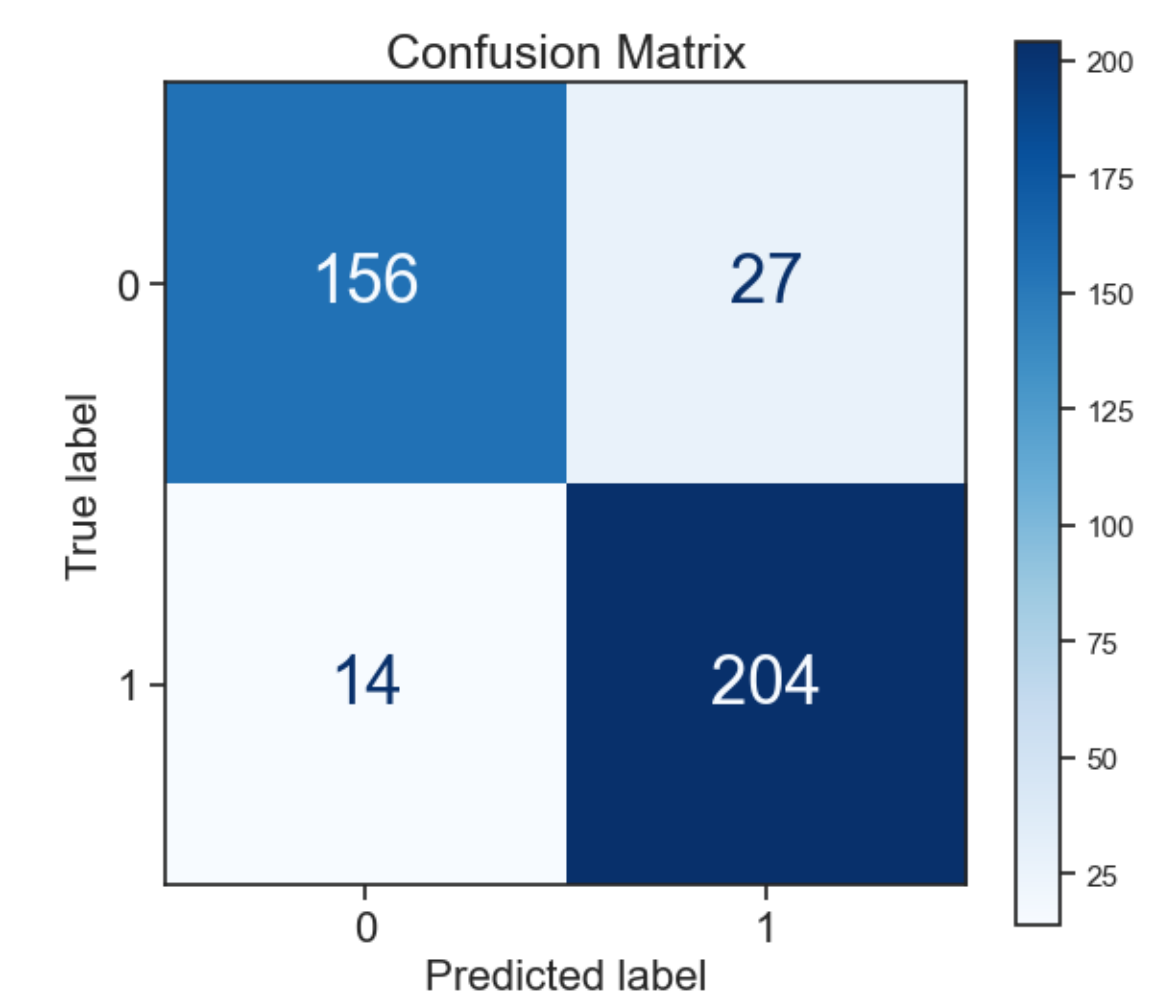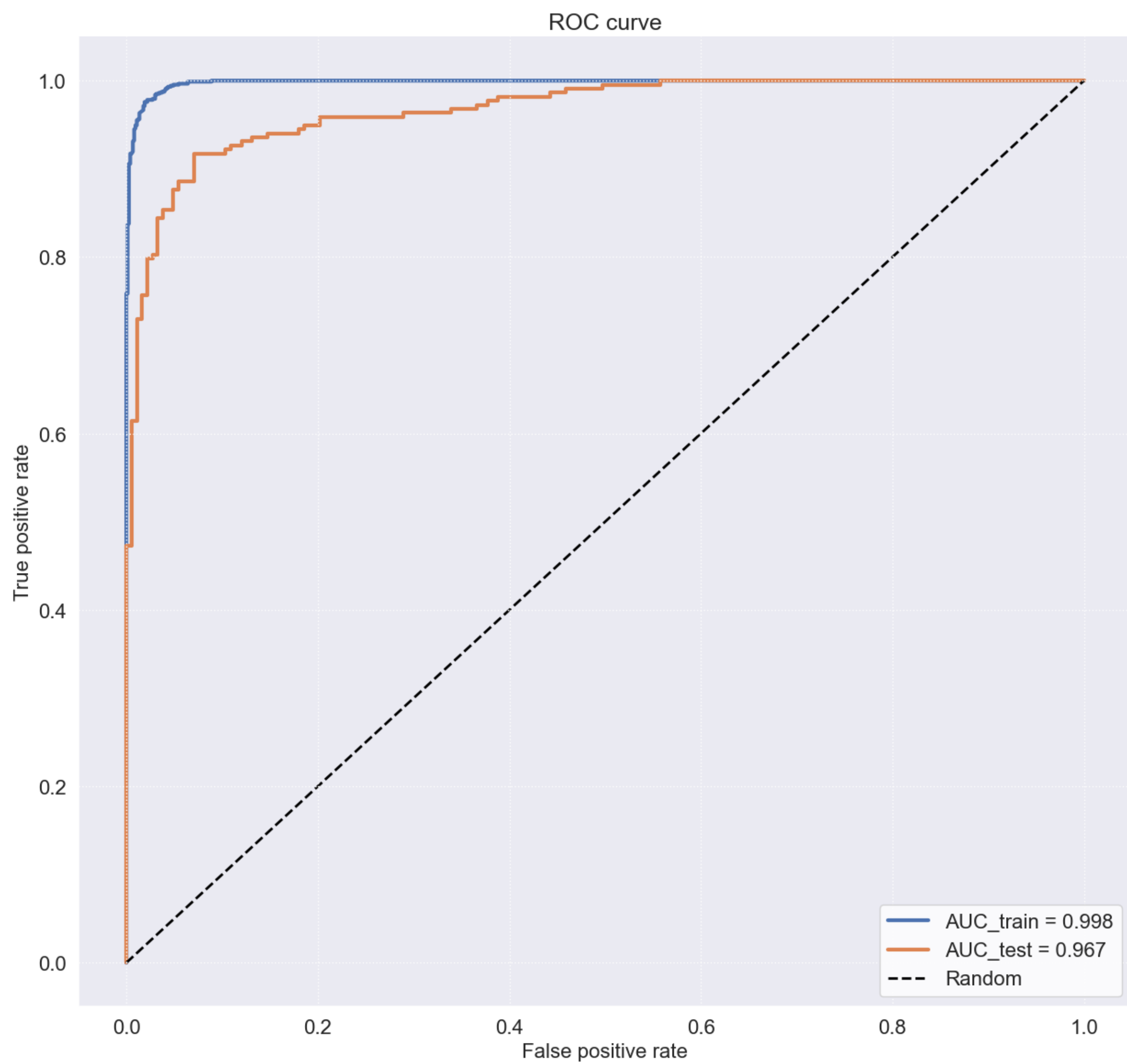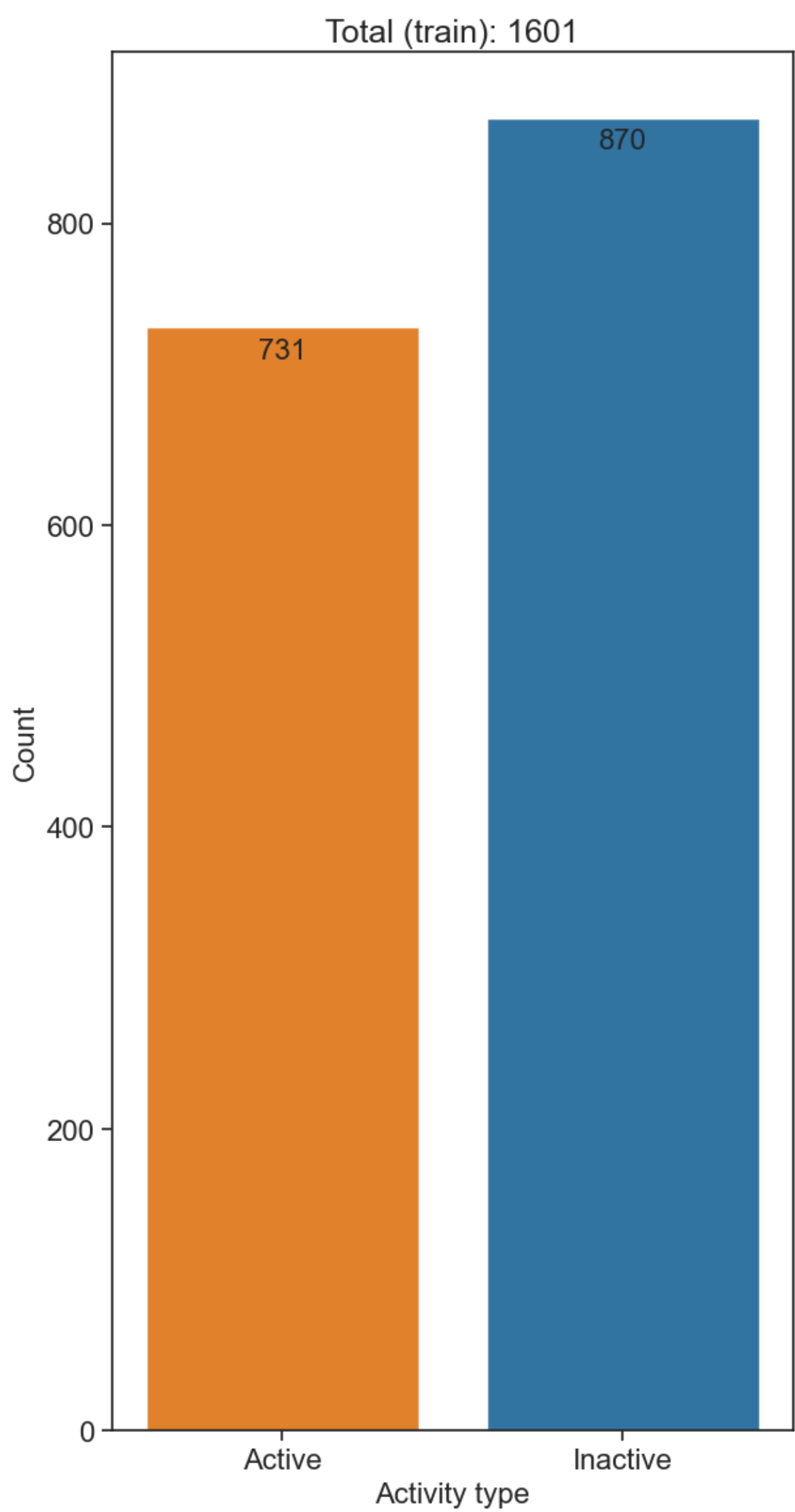
Metric	Train	Test
AUC	0.998	0.9675
accuracy	0.9769	0.8978
balanced accuracy	0.976	0.8941
recall_(sens)	0.9862	0.9358
specificity	0.9658	0.8525
precision	0.9717	0.8831
f1_score	0.9769	0.8974