Protein: P37231 - metric: balanced_accuracy

Total (train): 2288

Count by Activity type: Inactive 1204, Active 1084

ROC curve

AUC_train = 0.997
AUC_test = 0.961
Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2532)

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

Confusion Matrix

	Predicted 0	Predicted 1
True 0	264	38
True 1	20	251

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.87	0.13
True 1	0.074	0.93

Calibration curve

train: 0.0253
test: 0.0752
perfectly_calibrated

Active/Inactive probability

Inactive zone | Active zone

type:
- active (271)
- inactive (302)
- decoy (569)

XGBoost Parameters	Value
alpha	2.8469
colsample_bytree	0.557
gamma	0.5936
lambda	1.3031
learning_rate	0.0391
max_depth	10.0
min_child_weight	1.0
n_estimators	362.0
reg_alpha	0.6912
reg_lambda	0.7201
scale_pos_weight	1.4318
subsample	0.8992
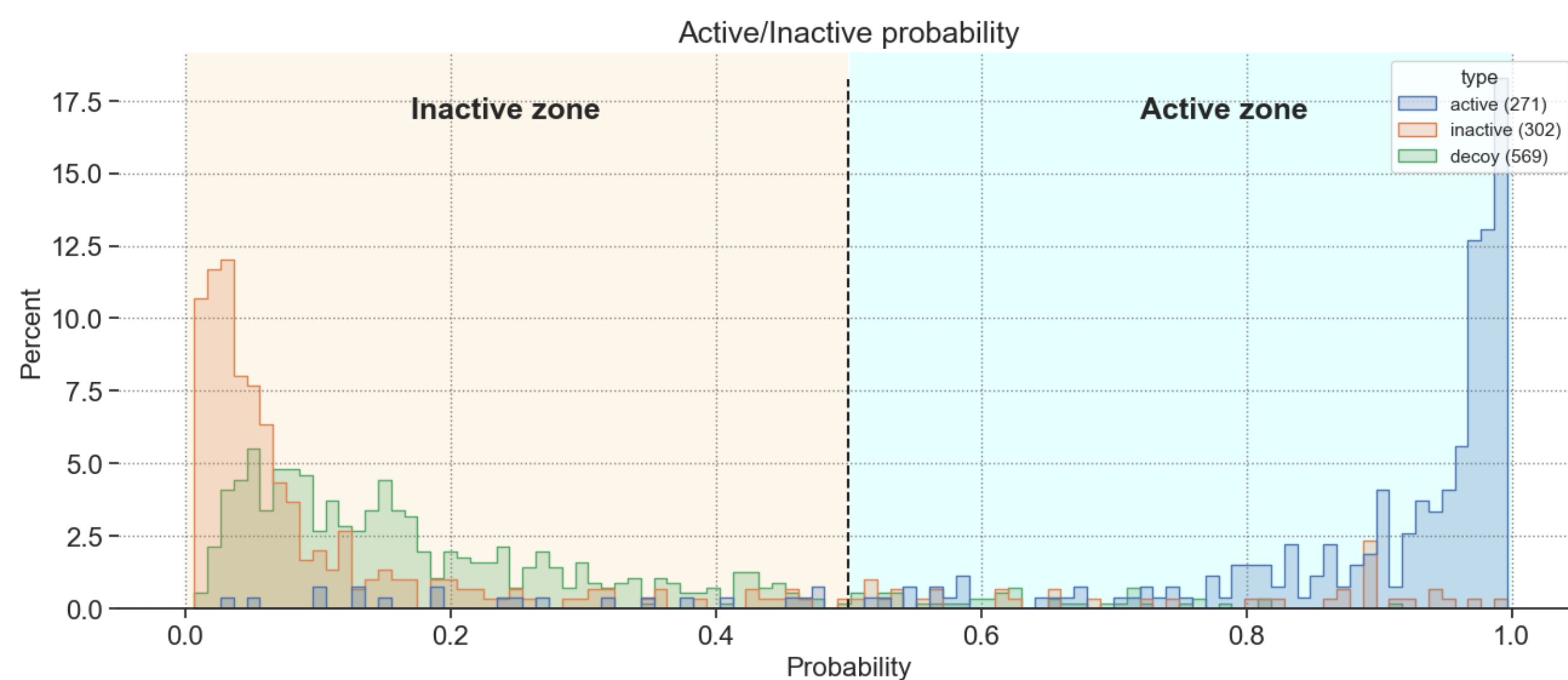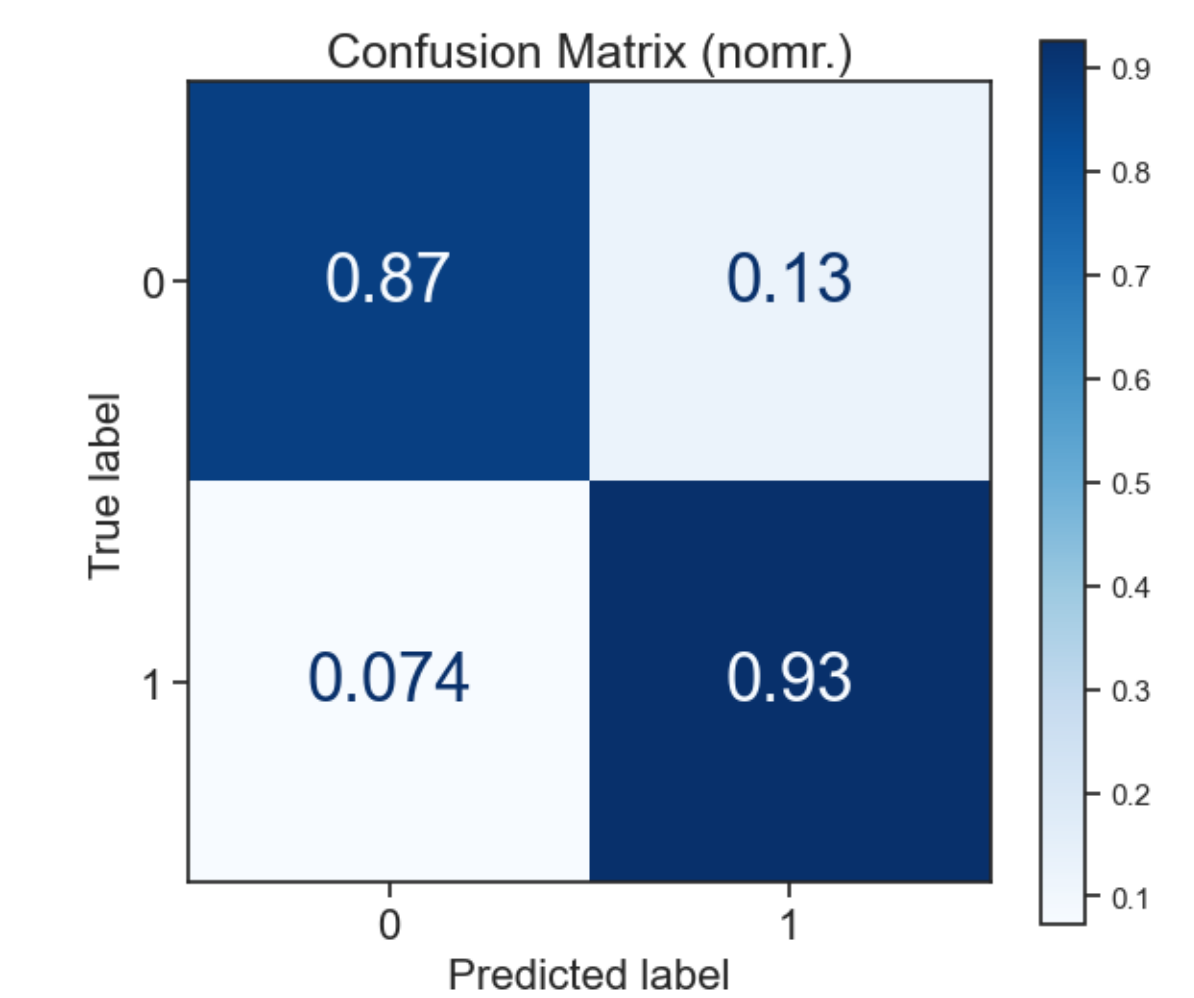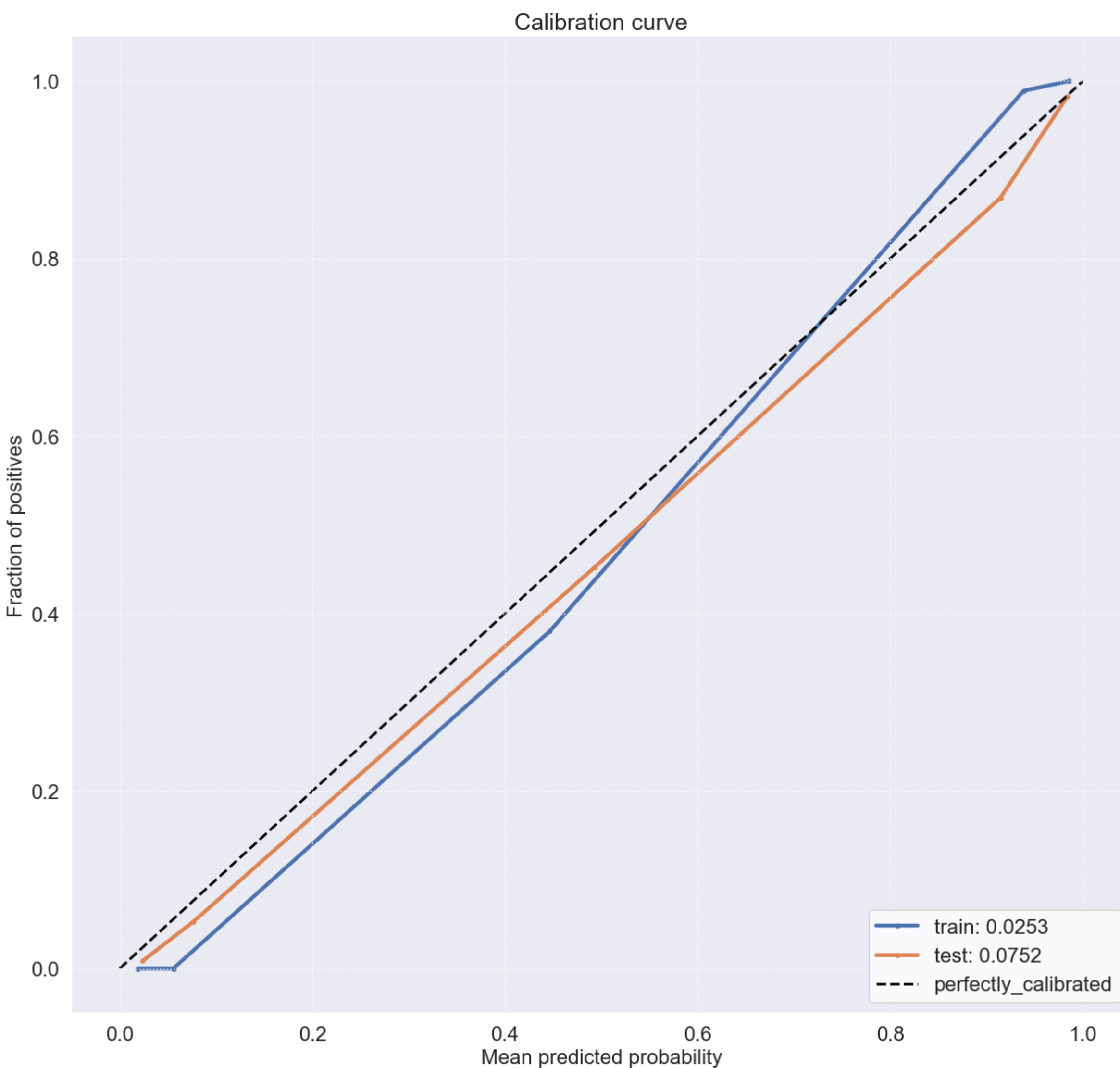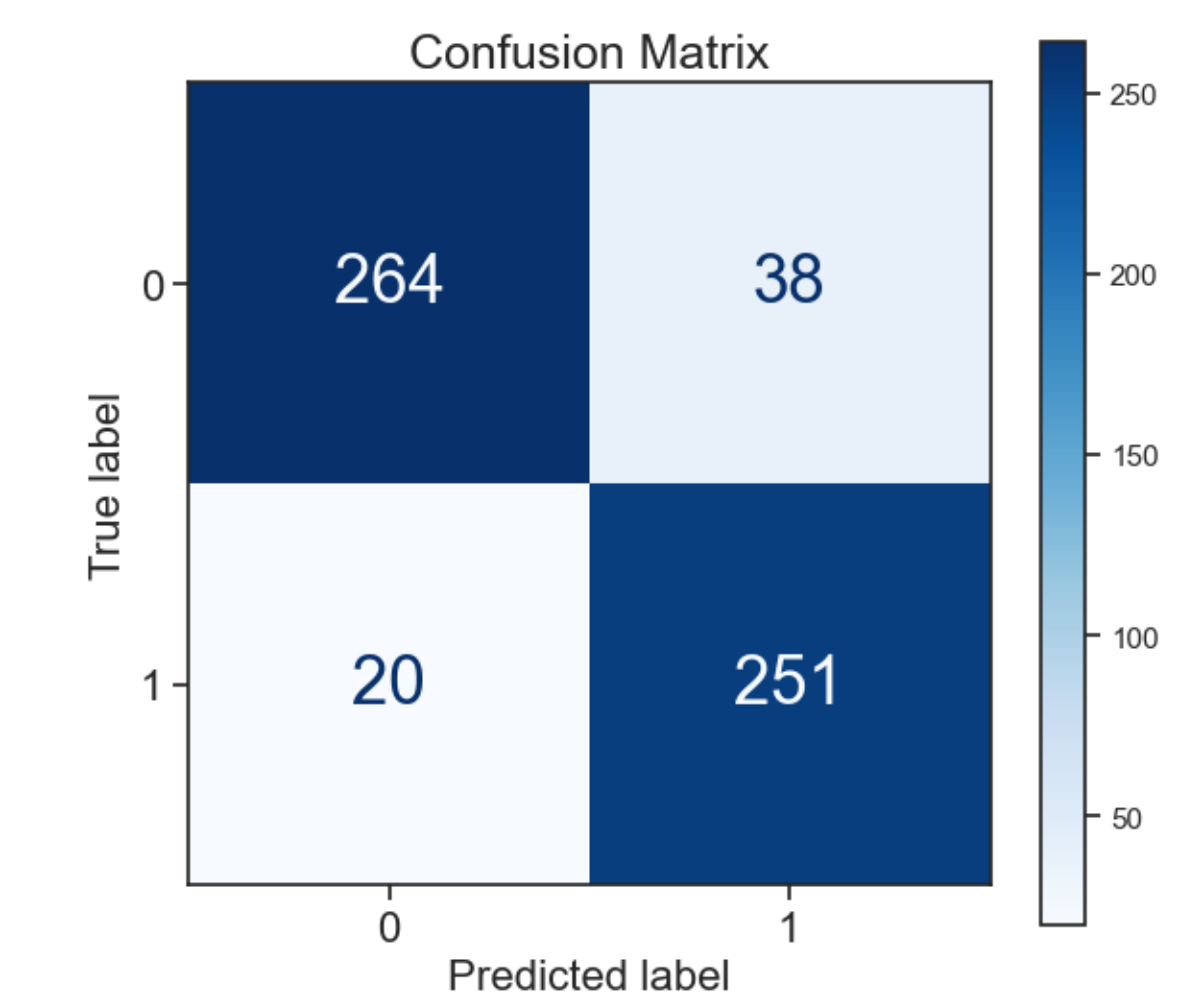
Metric	Train	Test
AUC	0.9967	0.9608
accuracy	0.9707	0.8988
balanced accuracy	0.9716	0.9002
recall_(sens)	0.988	0.9262
specificity	0.9551	0.8742
precision	0.952	0.8685
f1_score	0.9707	0.8989

Protein: P37231 - metric: f1

Total (train): 2288

ROC curve

	AUC_train = 0.999
	AUC_test = 0.965
	Random

Pareto Chart with Highlighted Best Attempt

	Trials
	Pareto Front Trials
	Best Trial (# 2920)
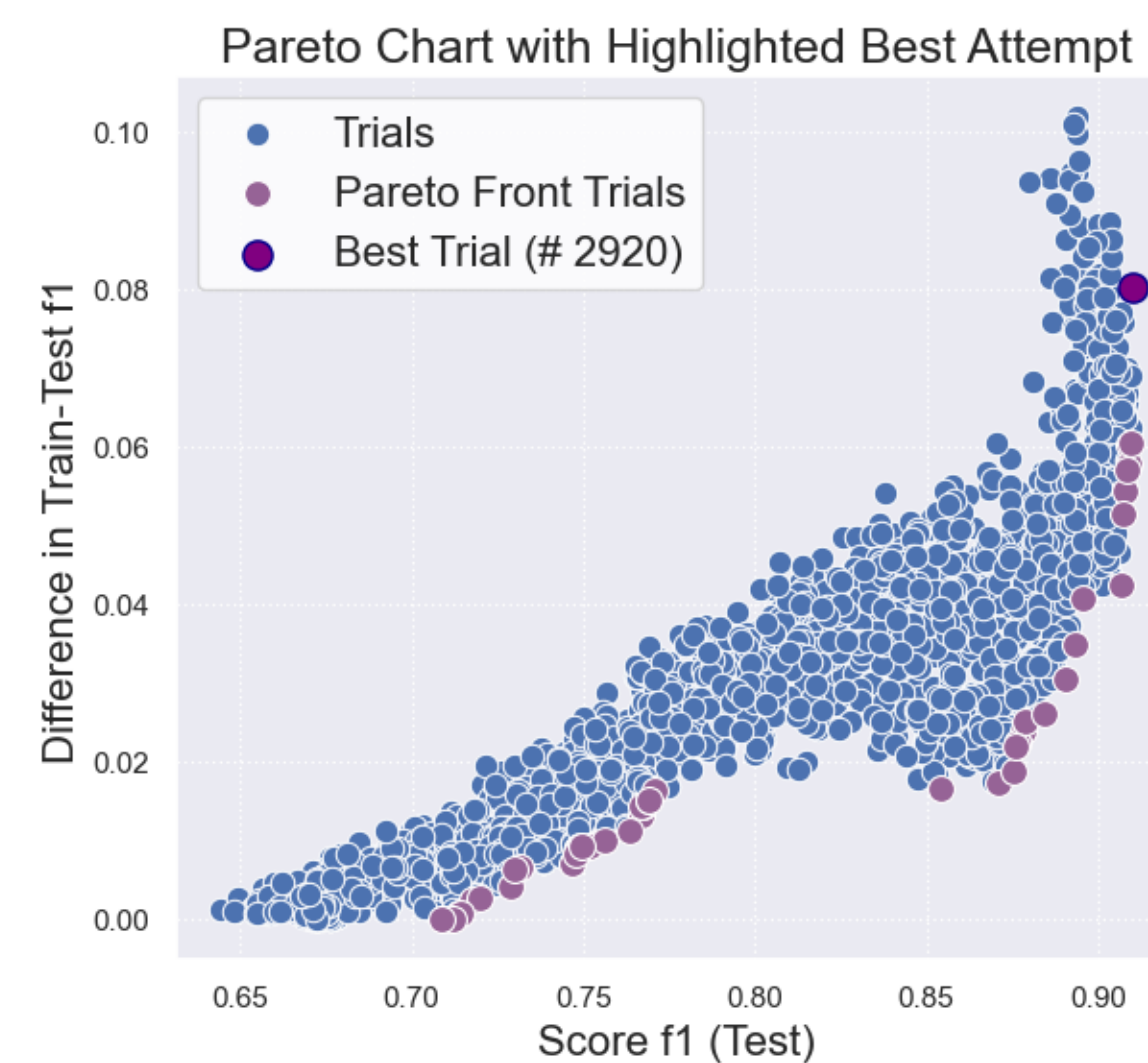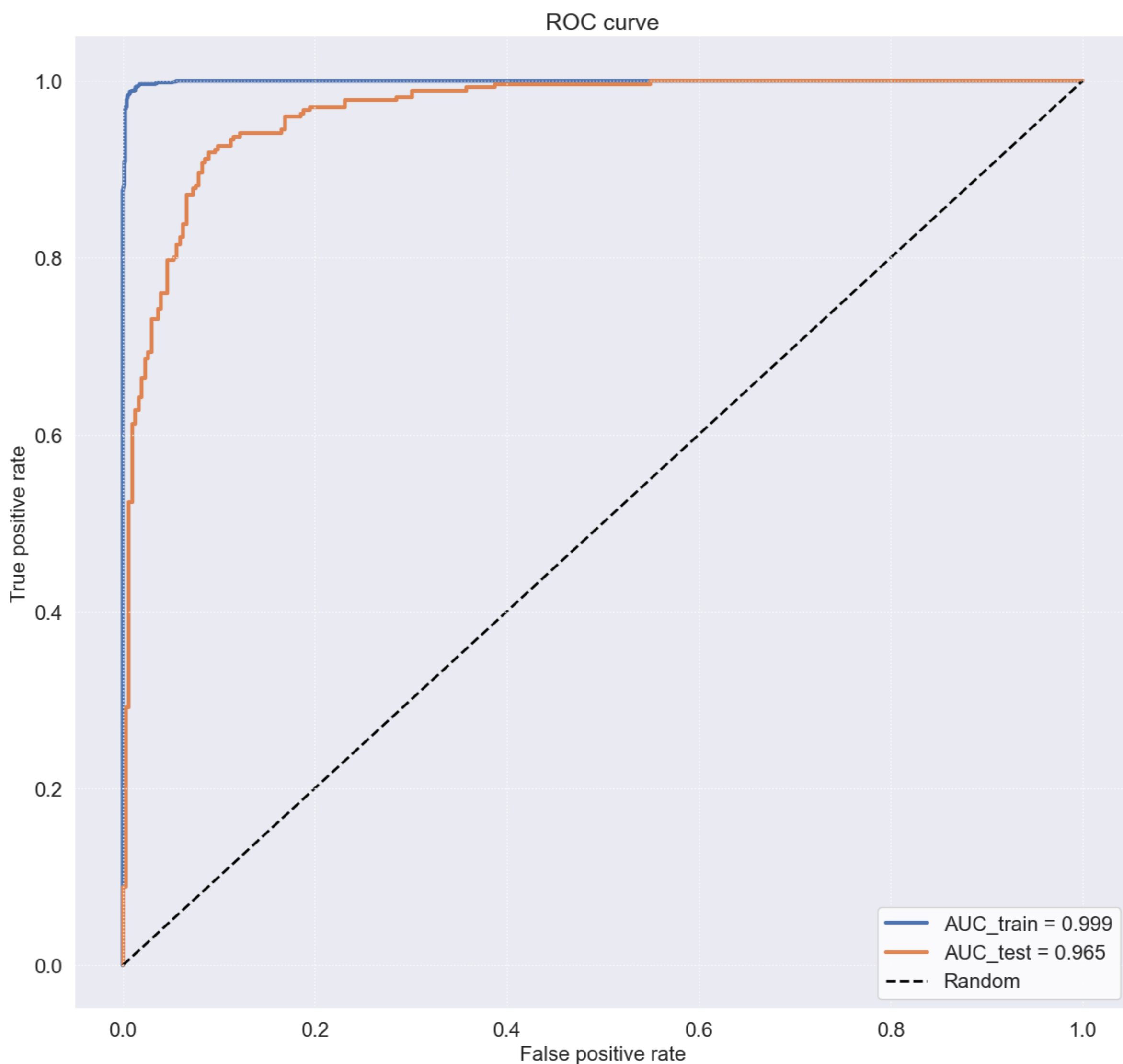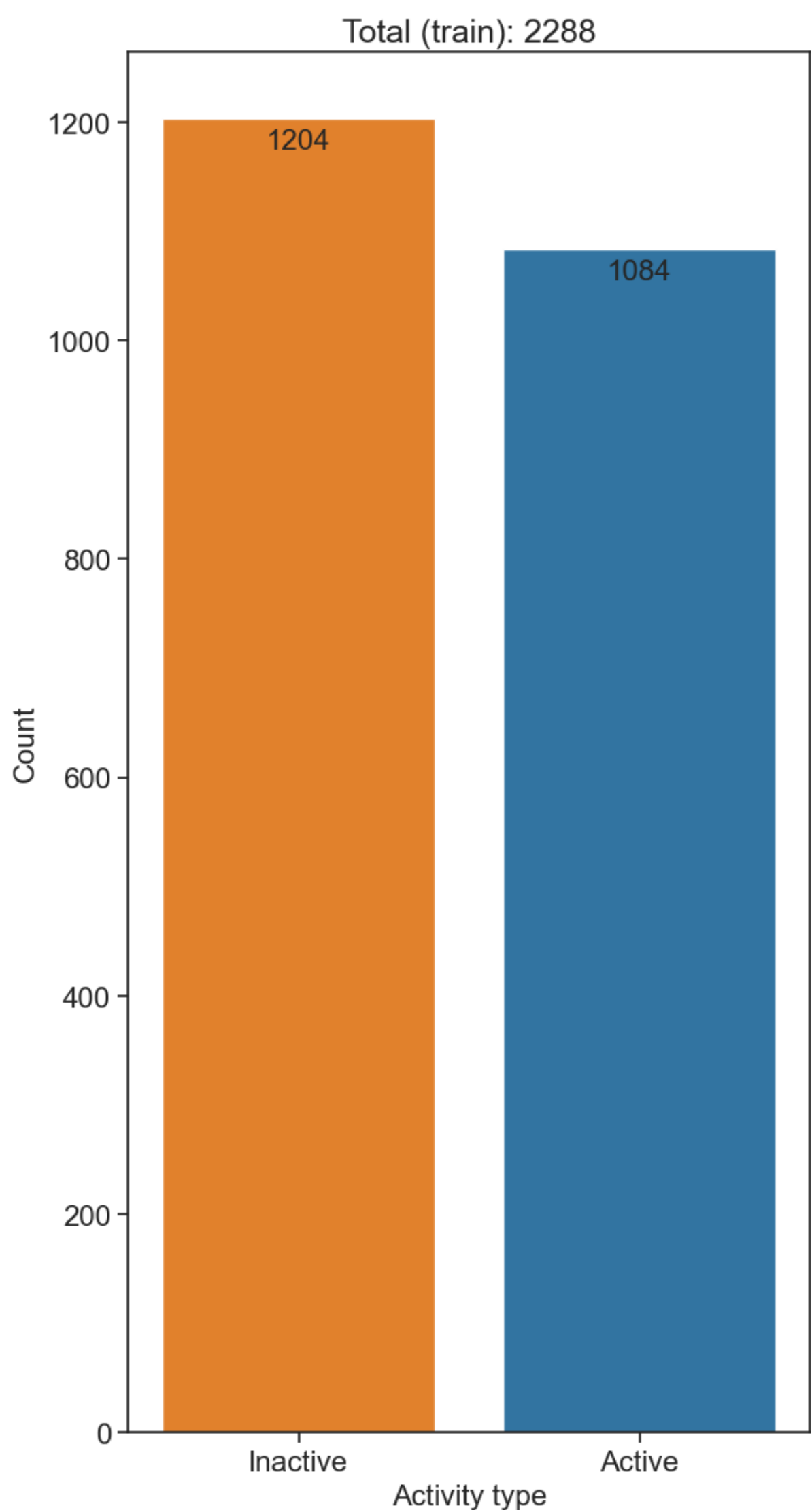
XGBoost Parameters	Value
alpha	3.8796
colsample_bytree	0.8156
gamma	0.2472
lambda	3.2935
learning_rate	0.0277
max_depth	10.0
min_child_weight	1.0
n_estimators	631.0
reg_alpha	0.3832
reg_lambda	0.98
scale_pos_weight	1.2139
subsample	0.831

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

	train: 0.0135
	test: 0.0713
	perfectly_calibrated

Active/Inactive probability
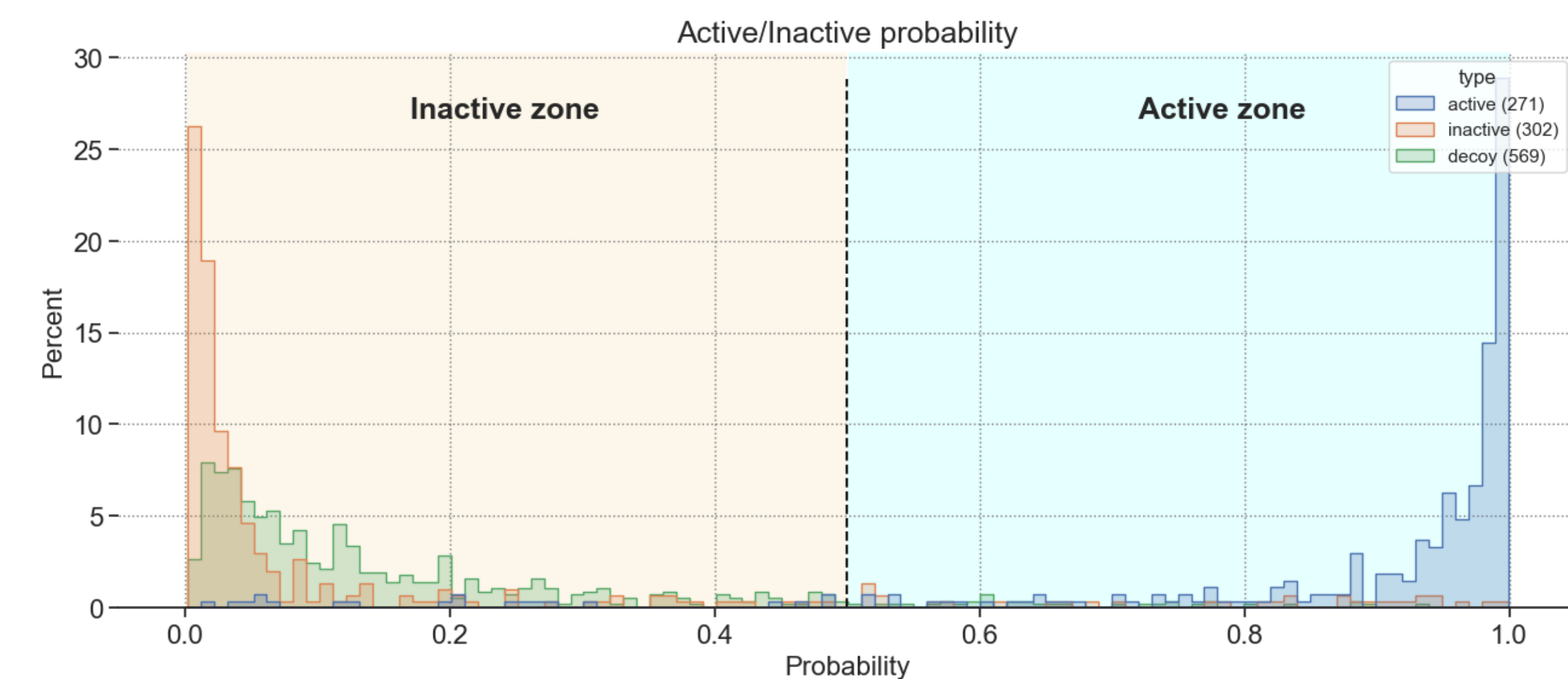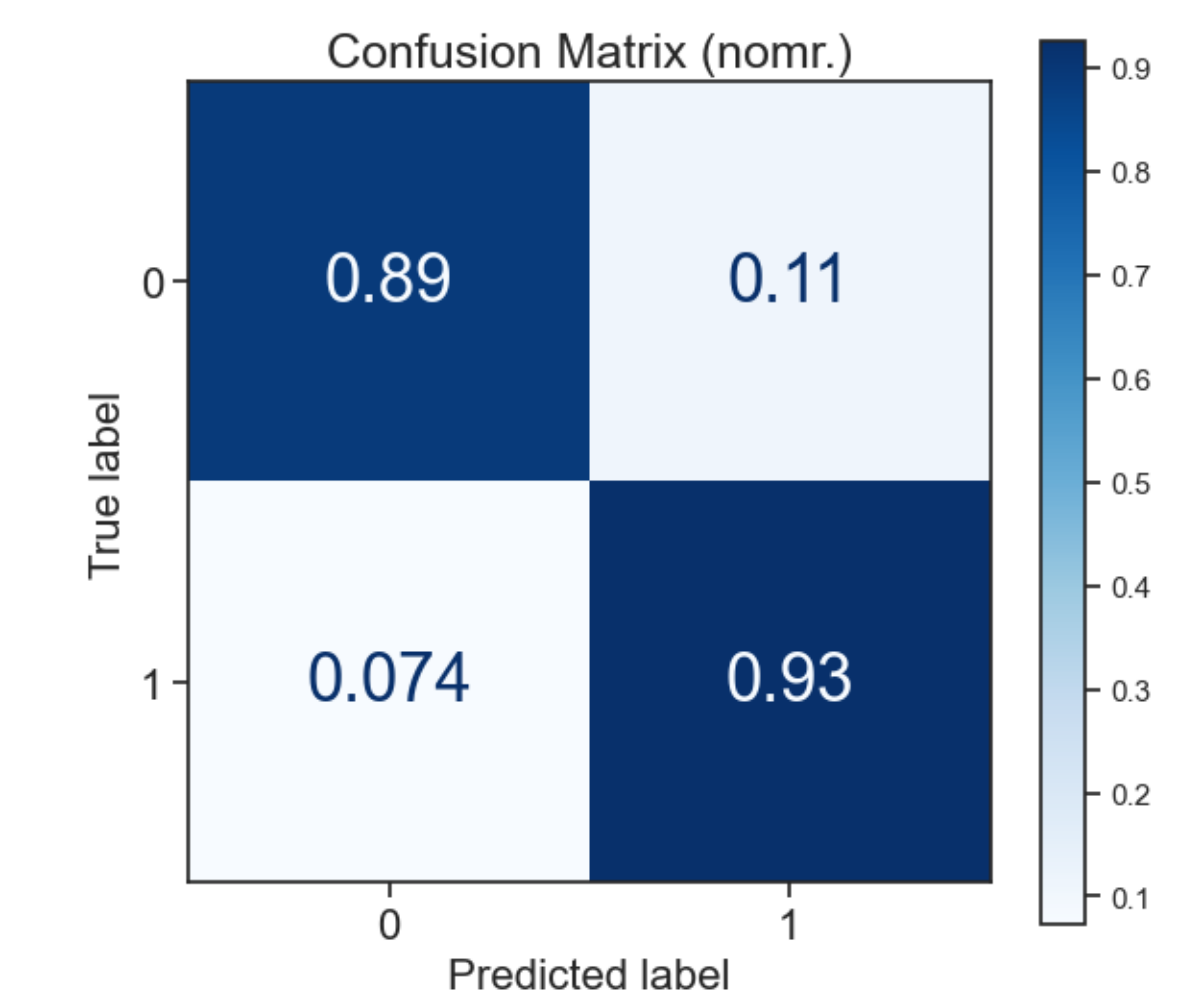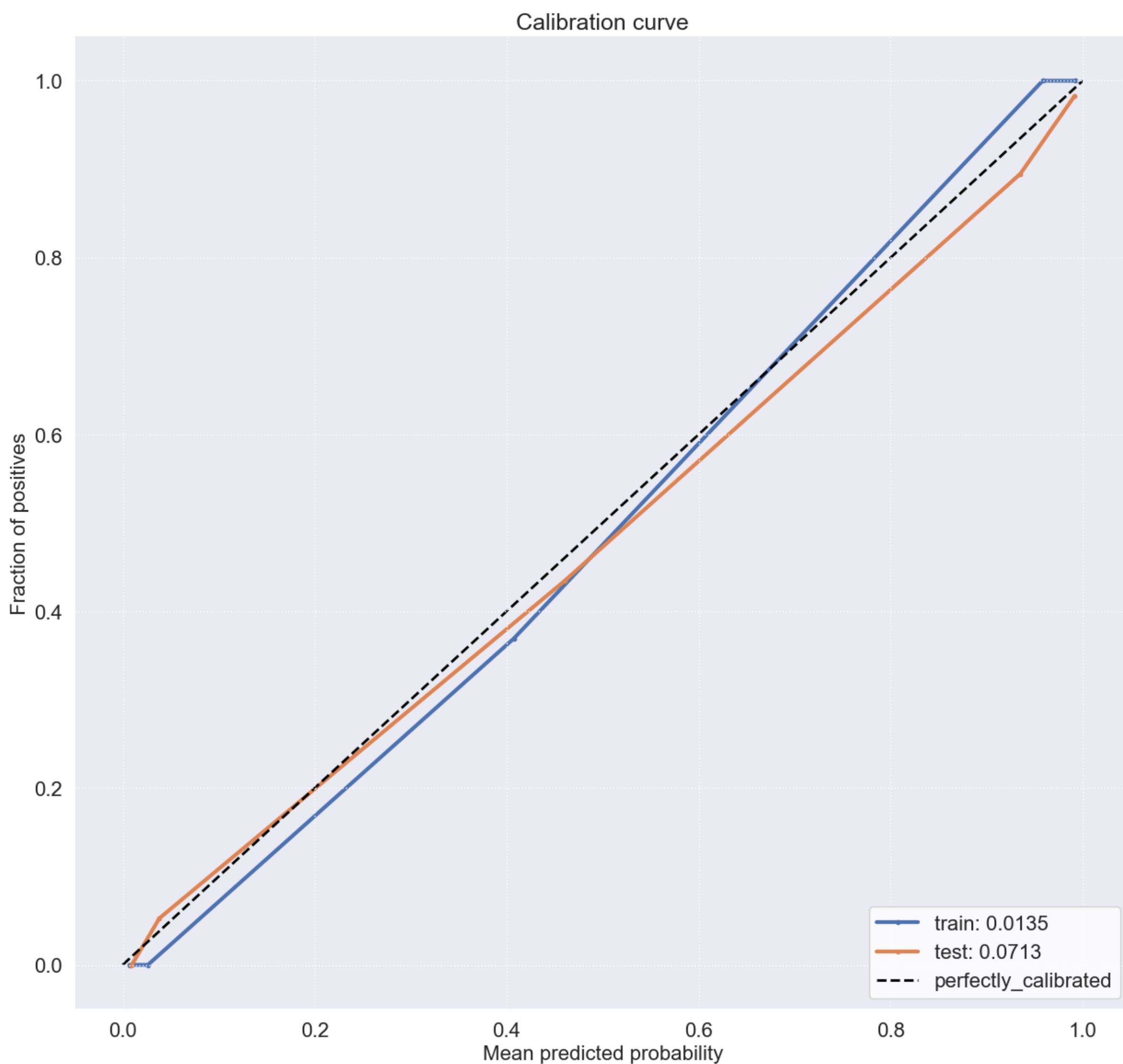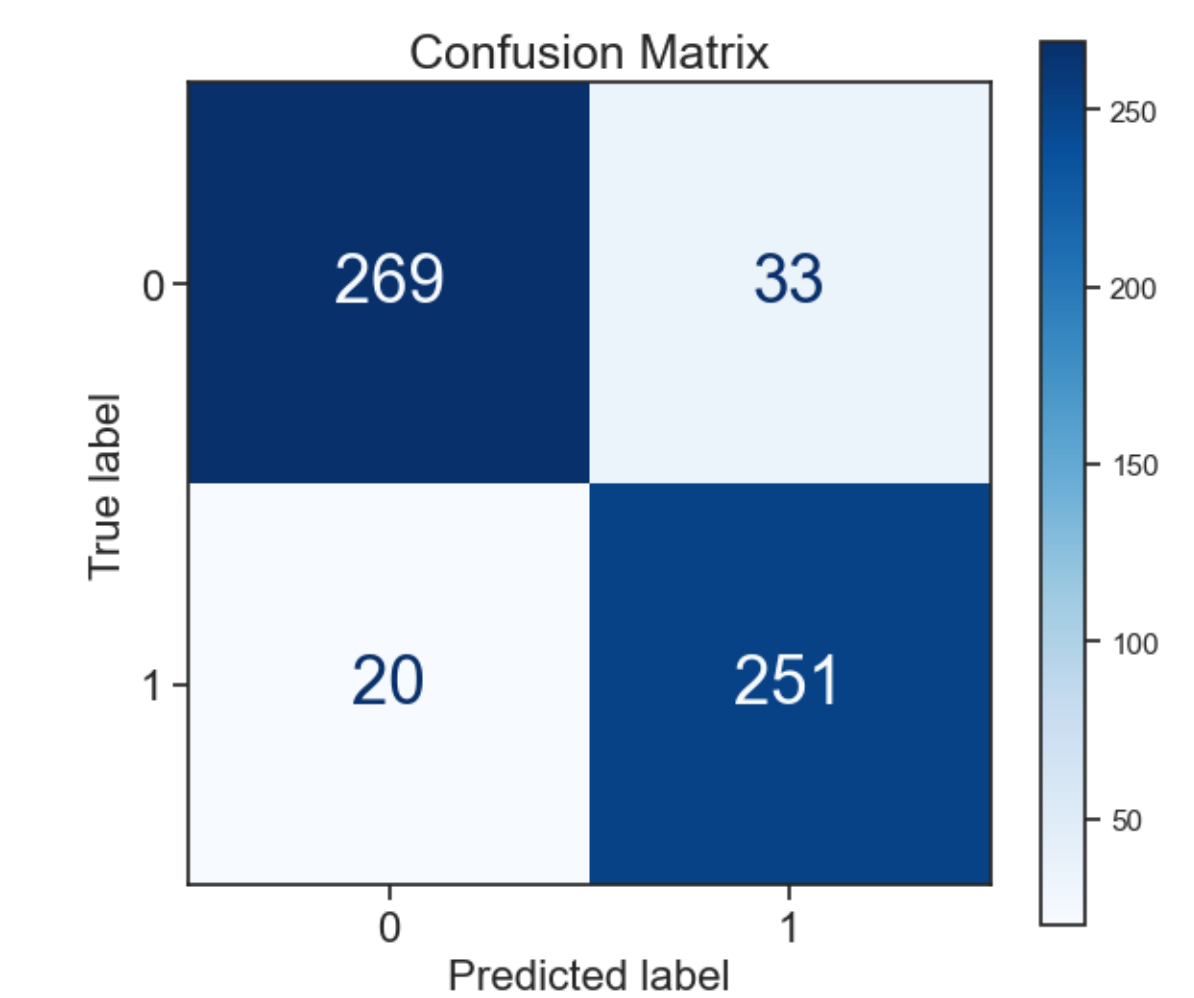
Metric	Train	Test
AUC	0.9994	0.9655
accuracy	0.9895	0.9075
balanced accuracy	0.9897	0.9085
recall_(sens)	0.9935	0.9262
specificity	0.9859	0.8907
precision	0.9845	0.8838
f1_score	0.9895	0.9076

Protein: P37231 - metric: precision

Total (train): 2288

Bar chart — Inactive: 1204, Active: 1084

ROC curve

- AUC_train = 1.000
- AUC_test = 0.963
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2864)

Confusion Matrix

	Pred 0	Pred 1
True 0	266	36
True 1	21	250

Confusion Matrix (nomr.)

	Pred 0	Pred 1
True 0	0.88	0.12
True 1	0.077	0.92

Calibration curve

- train: 0.0053
- test: 0.0756
- perfectly_calibrated

Active/Inactive probability

- active (271)
- inactive (302)
- decoy (569)

Inactive zone / Active zone

XGBoost Parameters	Value
alpha	4.211
colsample_bytree	0.7365
gamma	0.1026
lambda	2.2111
learning_rate	0.0836
max_depth	9.0
min_child_weight	1.0
n_estimators	654.0
reg_alpha	0.204
reg_lambda	0.6349
scale_pos_weight	1.5778
subsample	0.8924
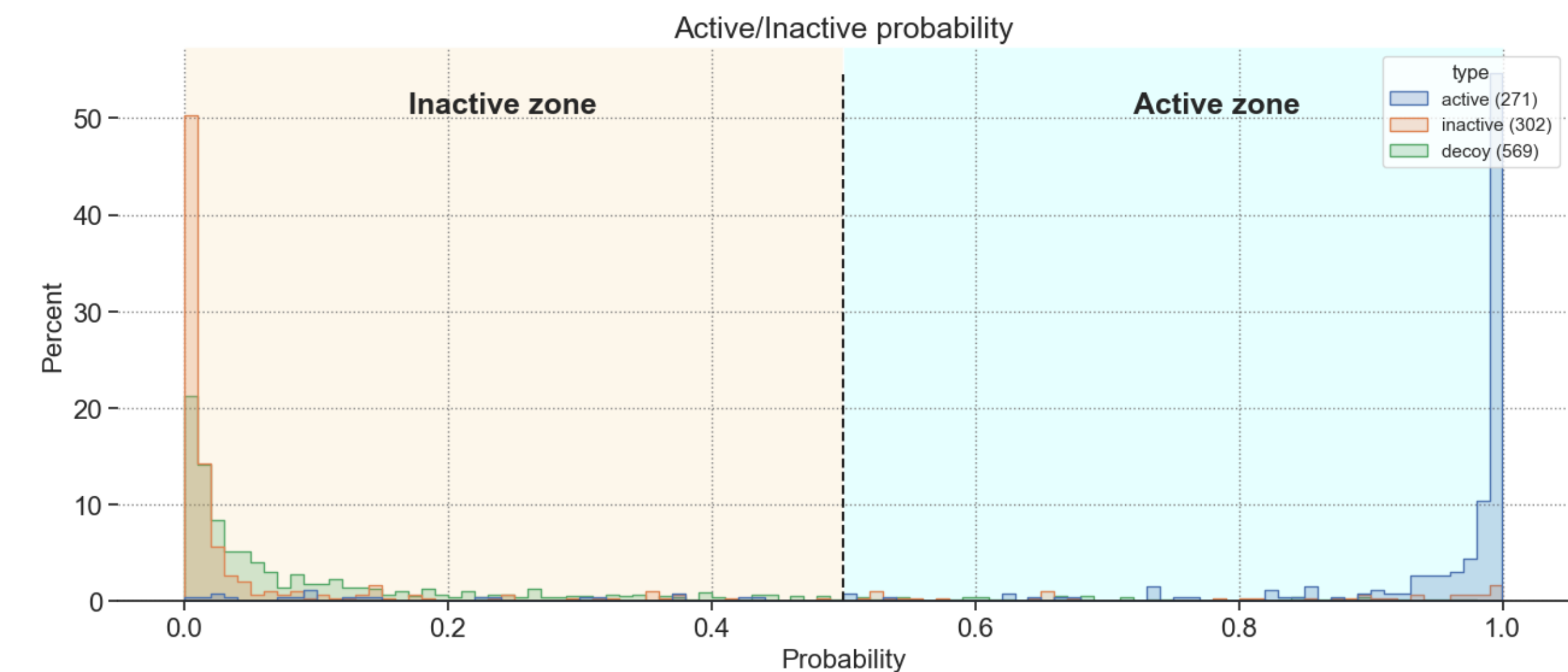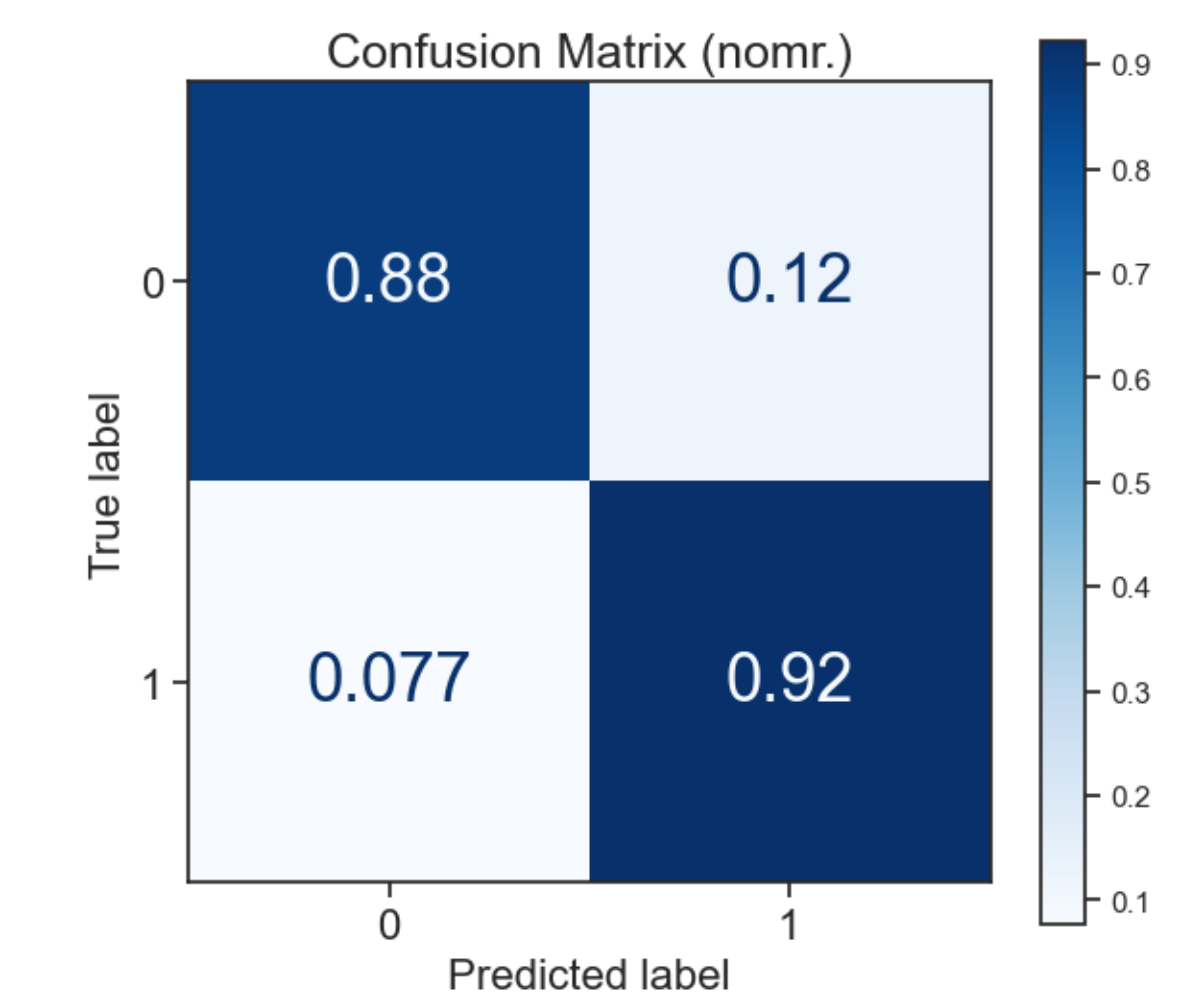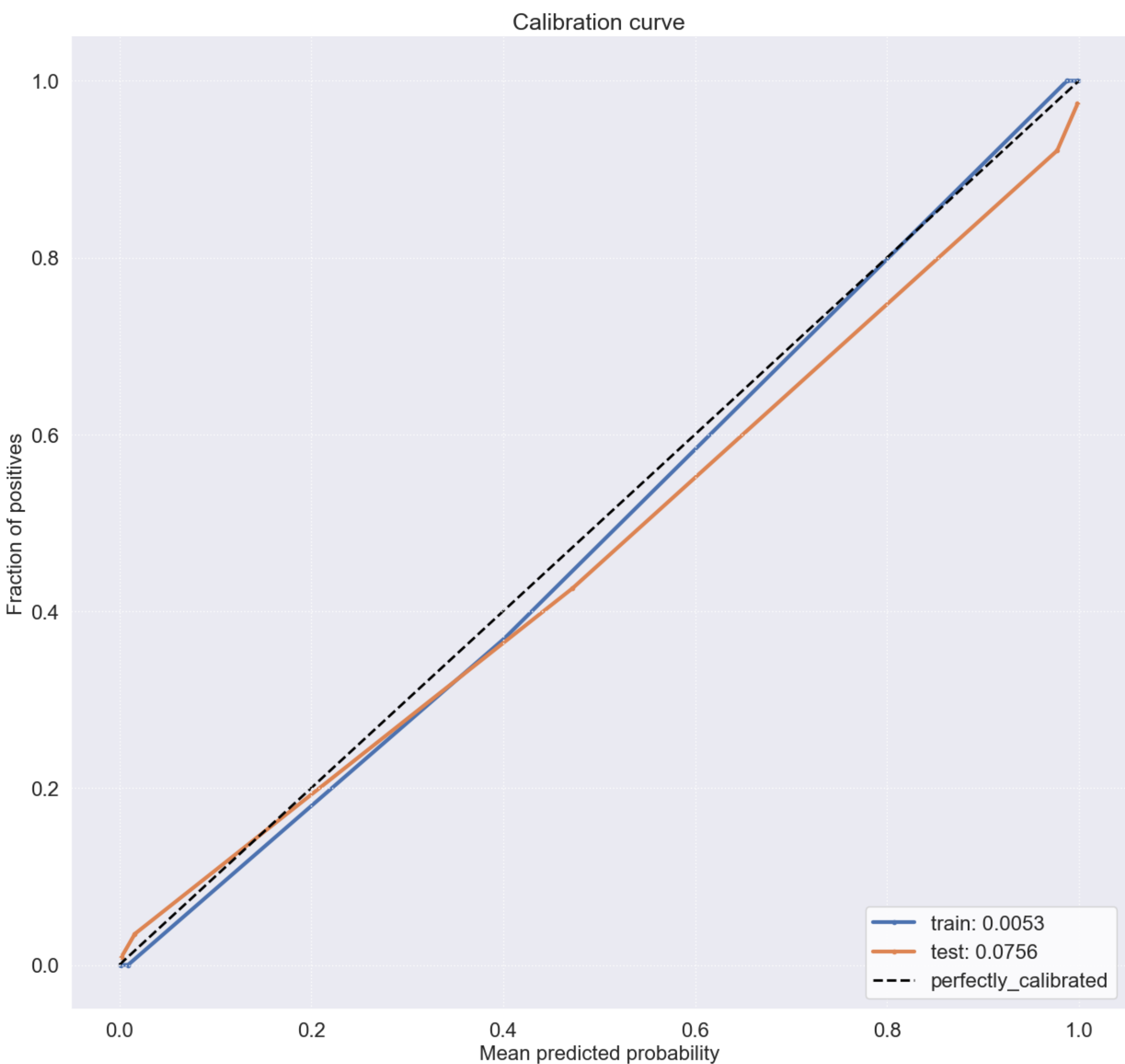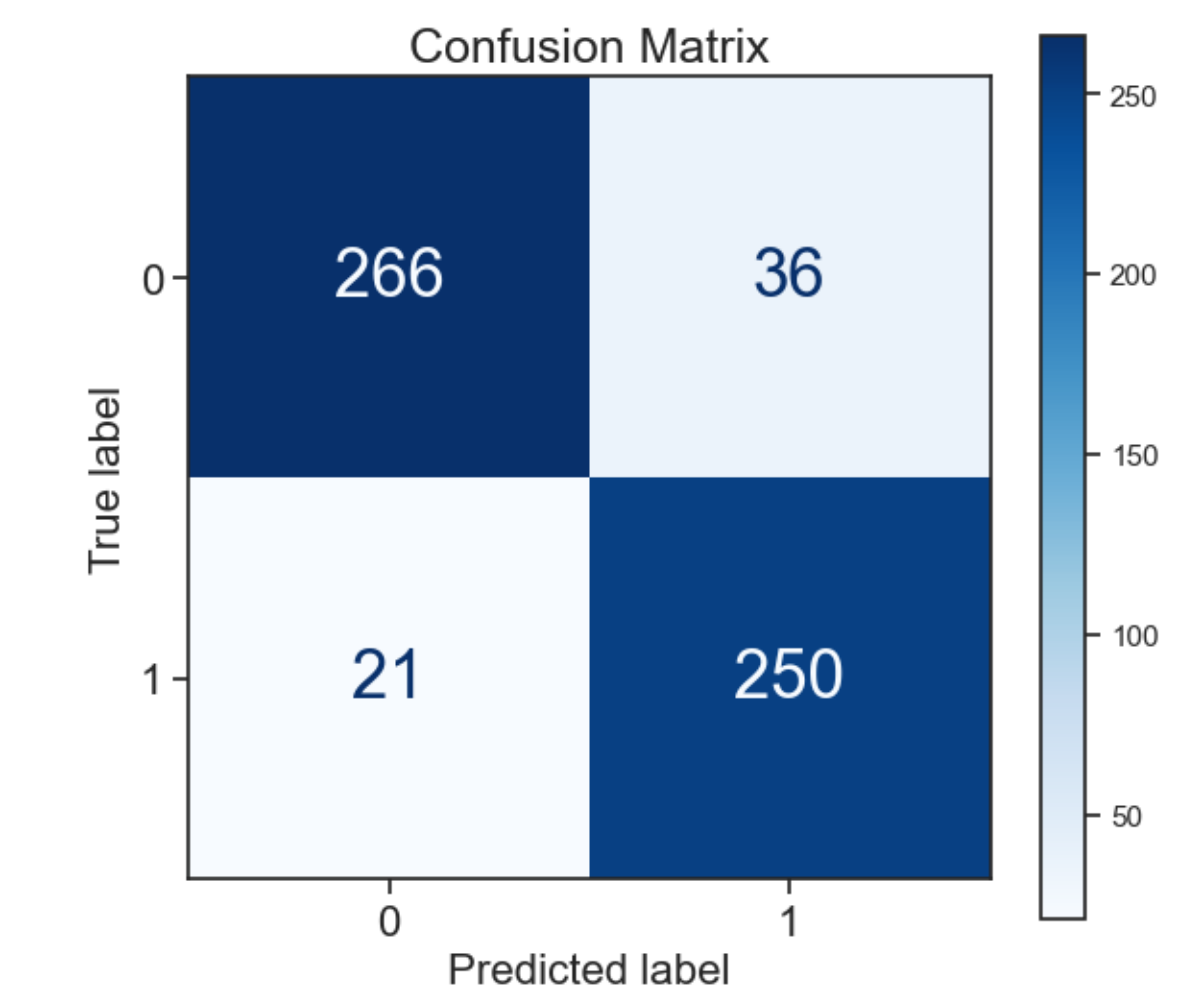
Metric	Train	Test
AUC	0.9999	0.963
accuracy	0.9952	0.9005
balanced accuracy	0.9954	0.9017
recall_(sens)	0.9991	0.9225
specificity	0.9917	0.8808
precision	0.9909	0.8741
f1_score	0.9952	0.9006

Protein: P37231 - metric: SoftVotingXGBoostEnsemble
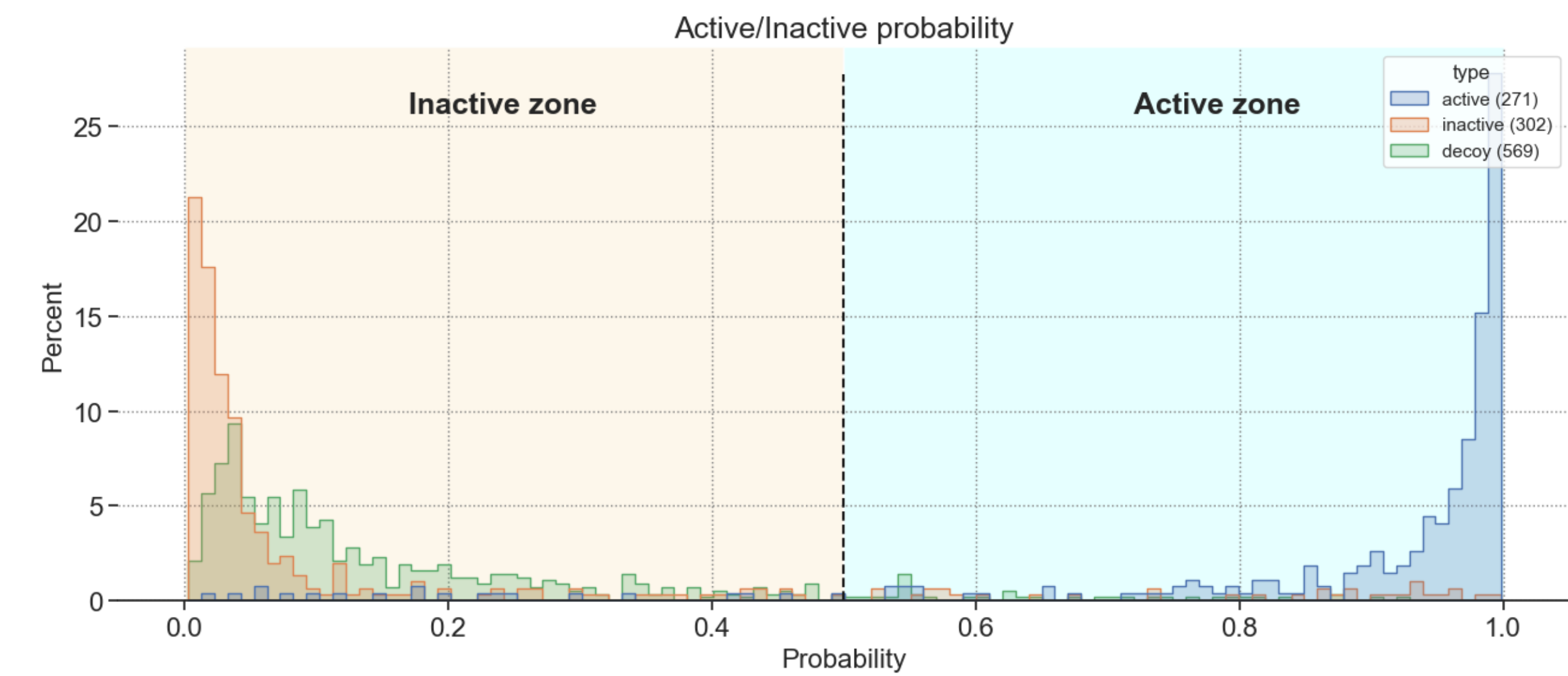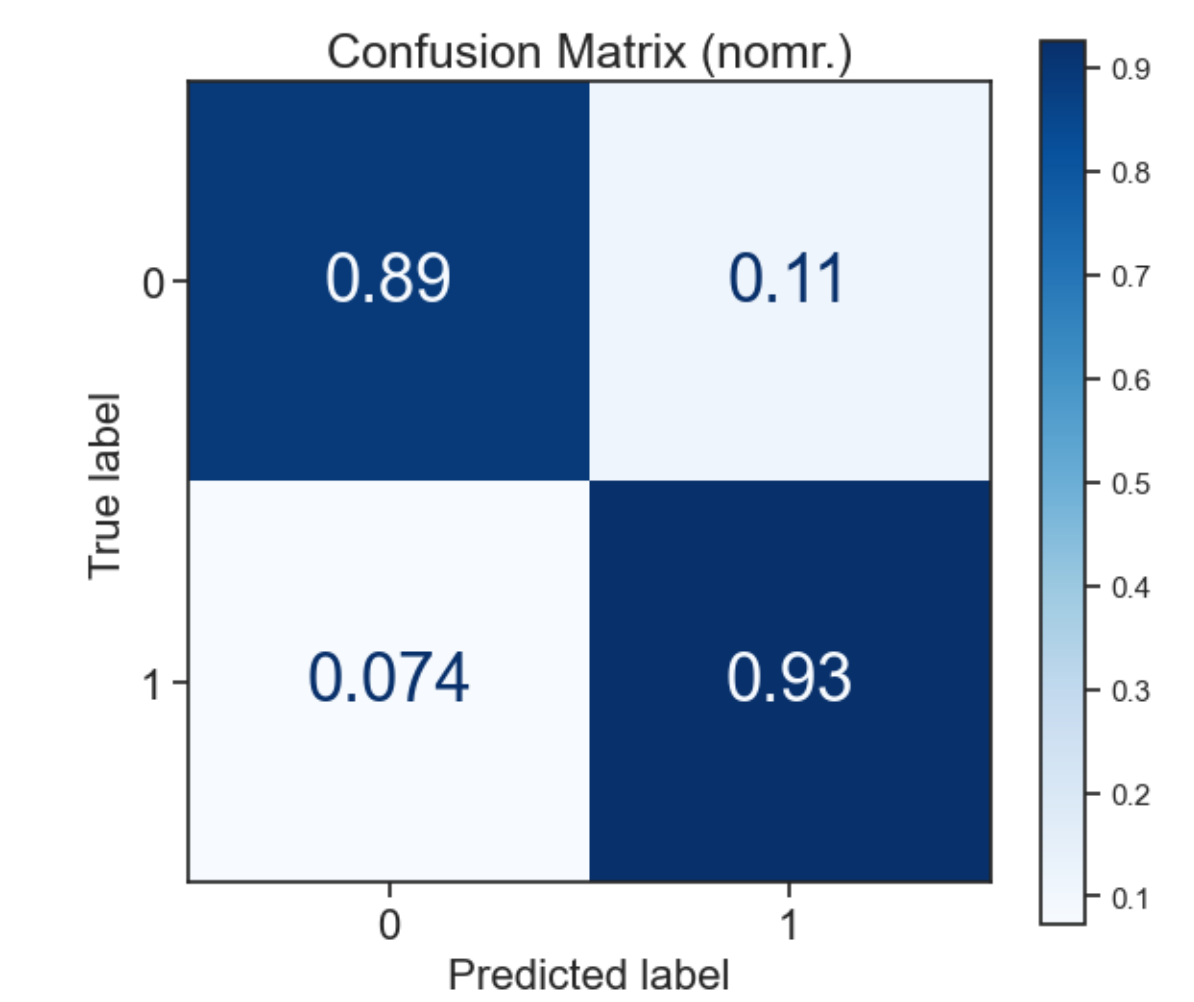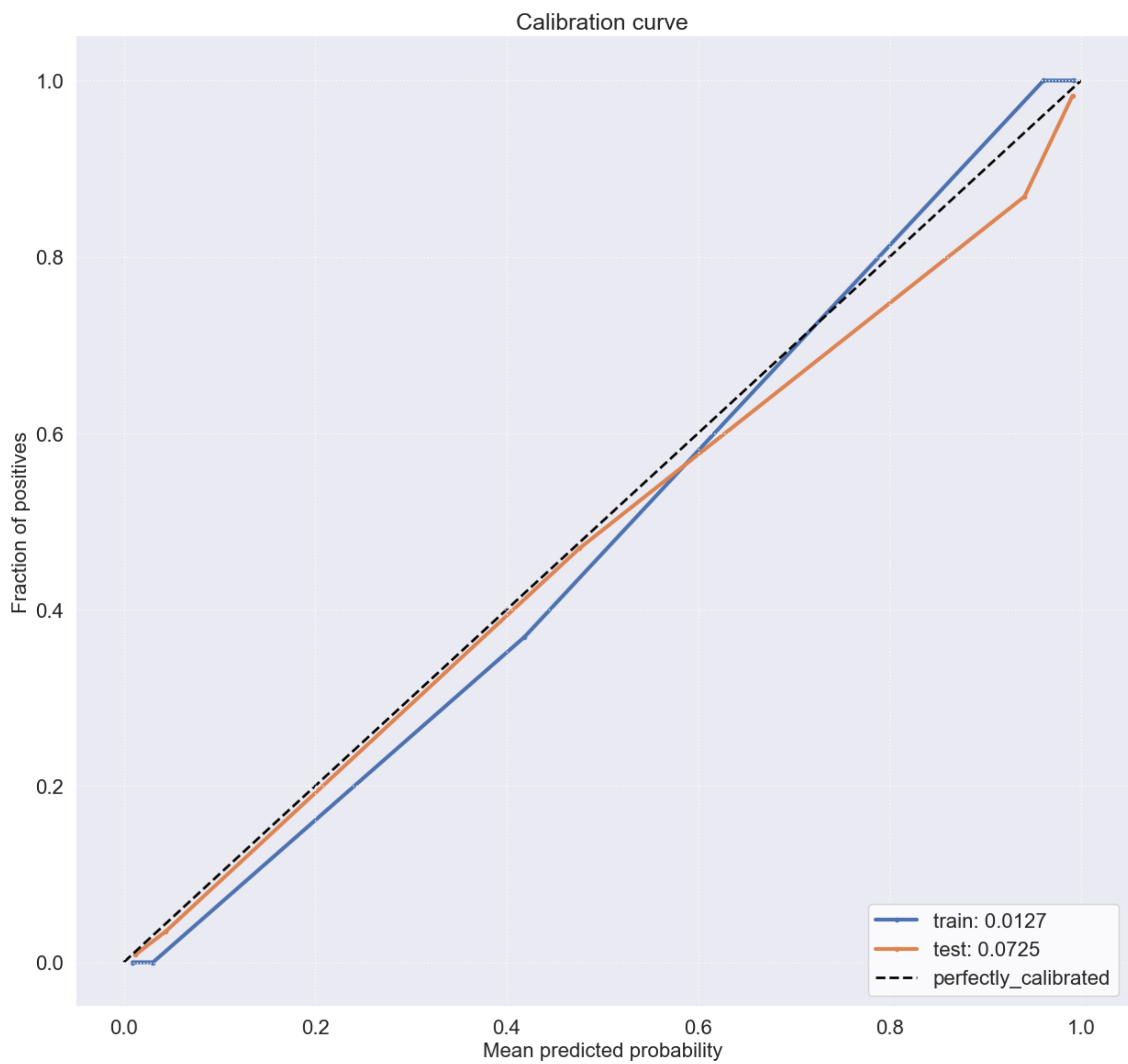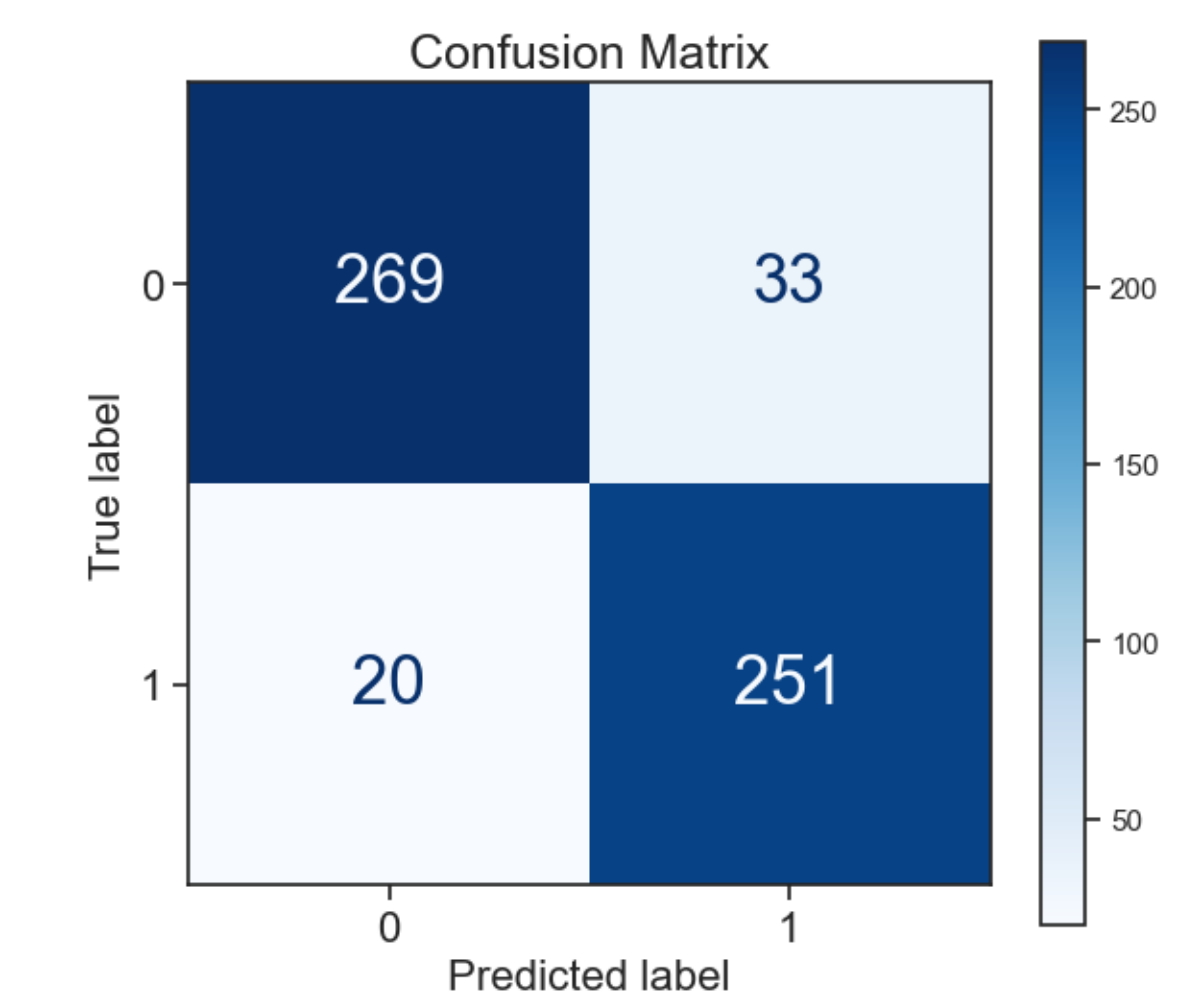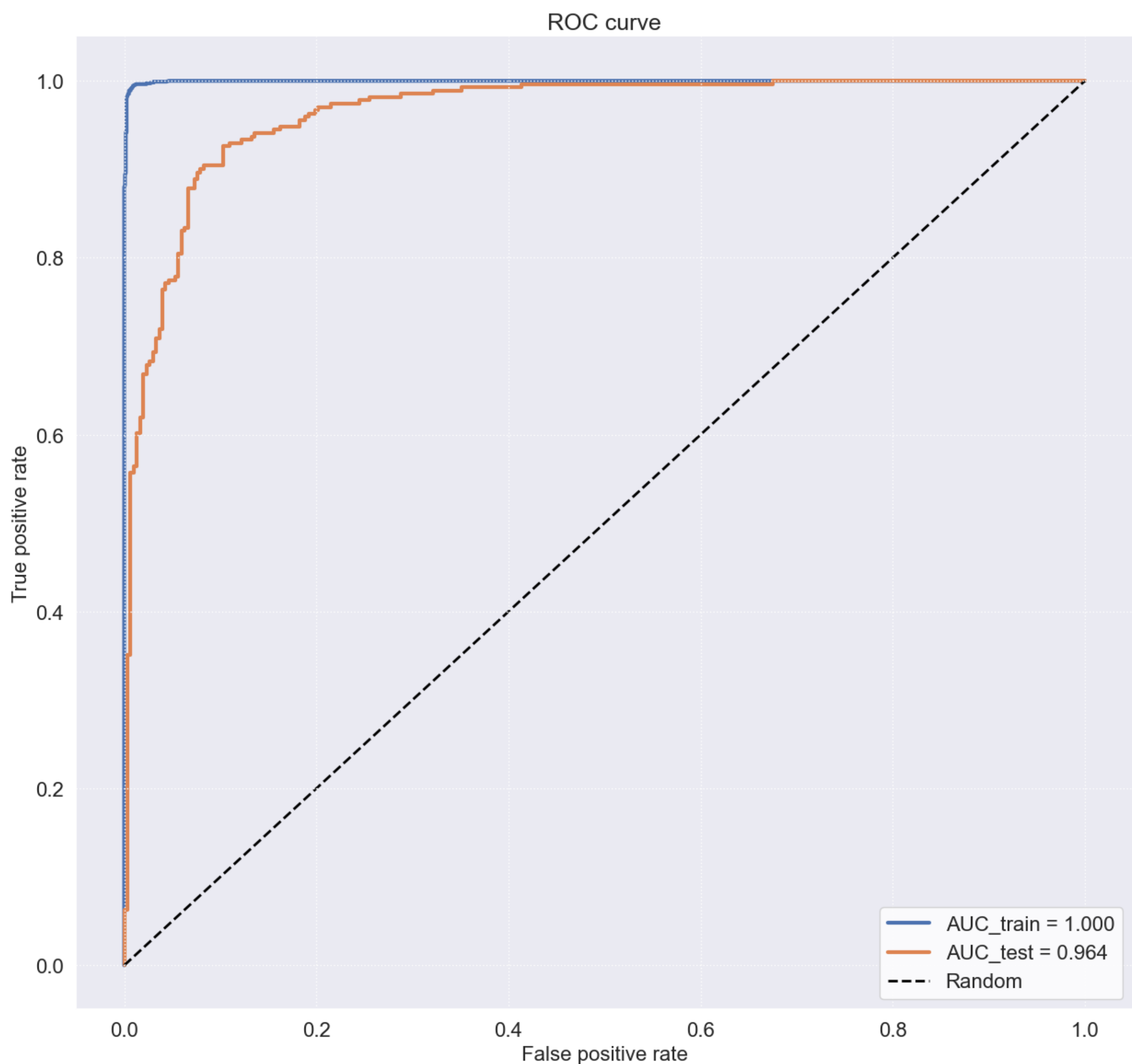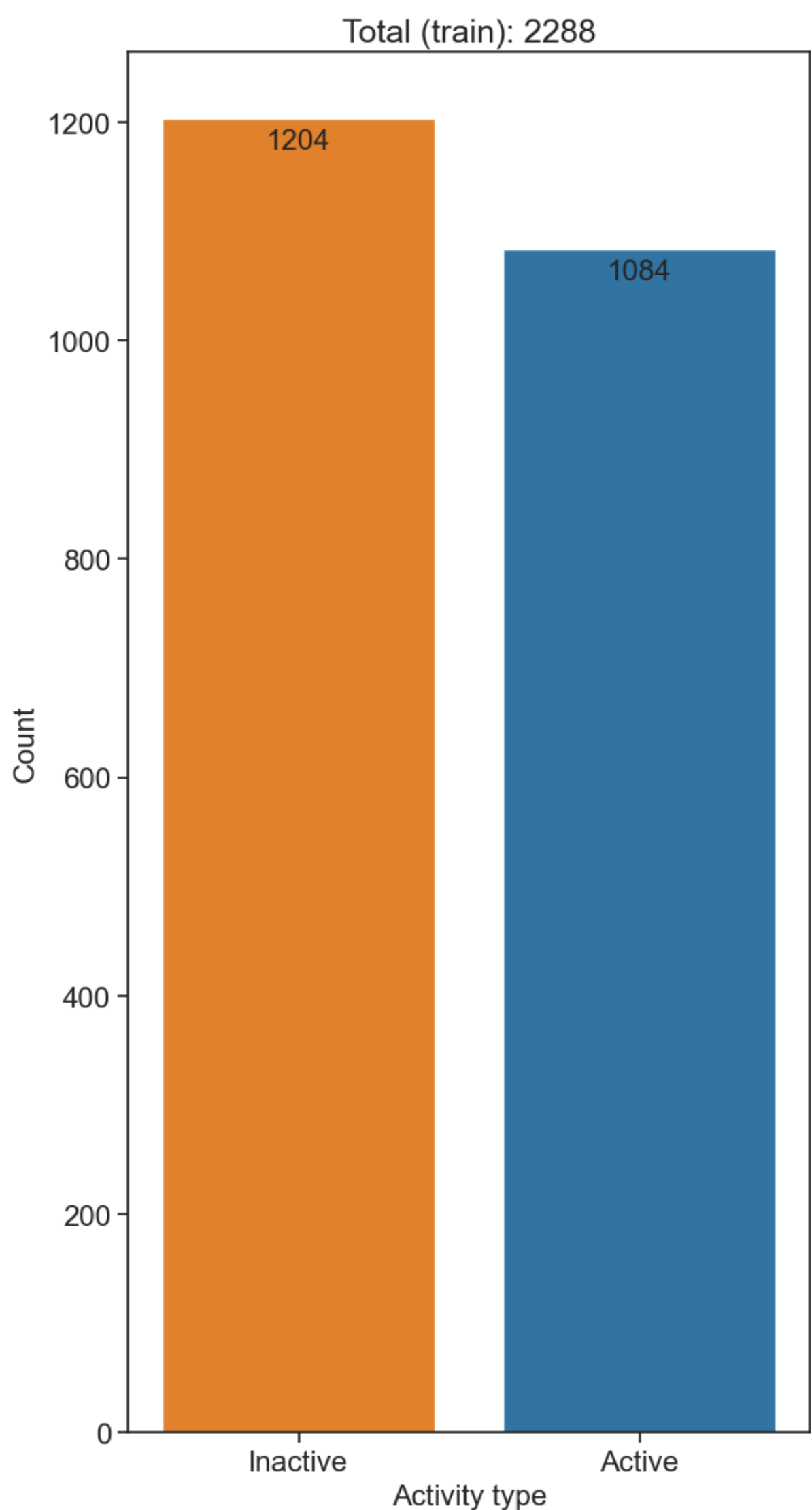
Metric	Train	Test
AUC	0.9996	0.964
accuracy	0.9904	0.9075
balanced accuracy	0.9907	0.9085
recall_(sens)	0.9963	0.9262
specificity	0.985	0.8907
precision	0.9836	0.8838
f1_score	0.9904	0.9076