Protein: P06213 - metric: balanced_accuracy

Total (train): 411

ROC curve

AUC_train = 0.998
AUC_test = 0.925
Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2788)

Confusion Matrix

	0	1
0	30	8
1	7	58

Confusion Matrix (nomr.)

	0	1
0	0.79	0.21
1	0.11	0.89

Calibration curve

train: 0.0193
test: 0.103
perfectly_calibrated

Active/Inactive probability
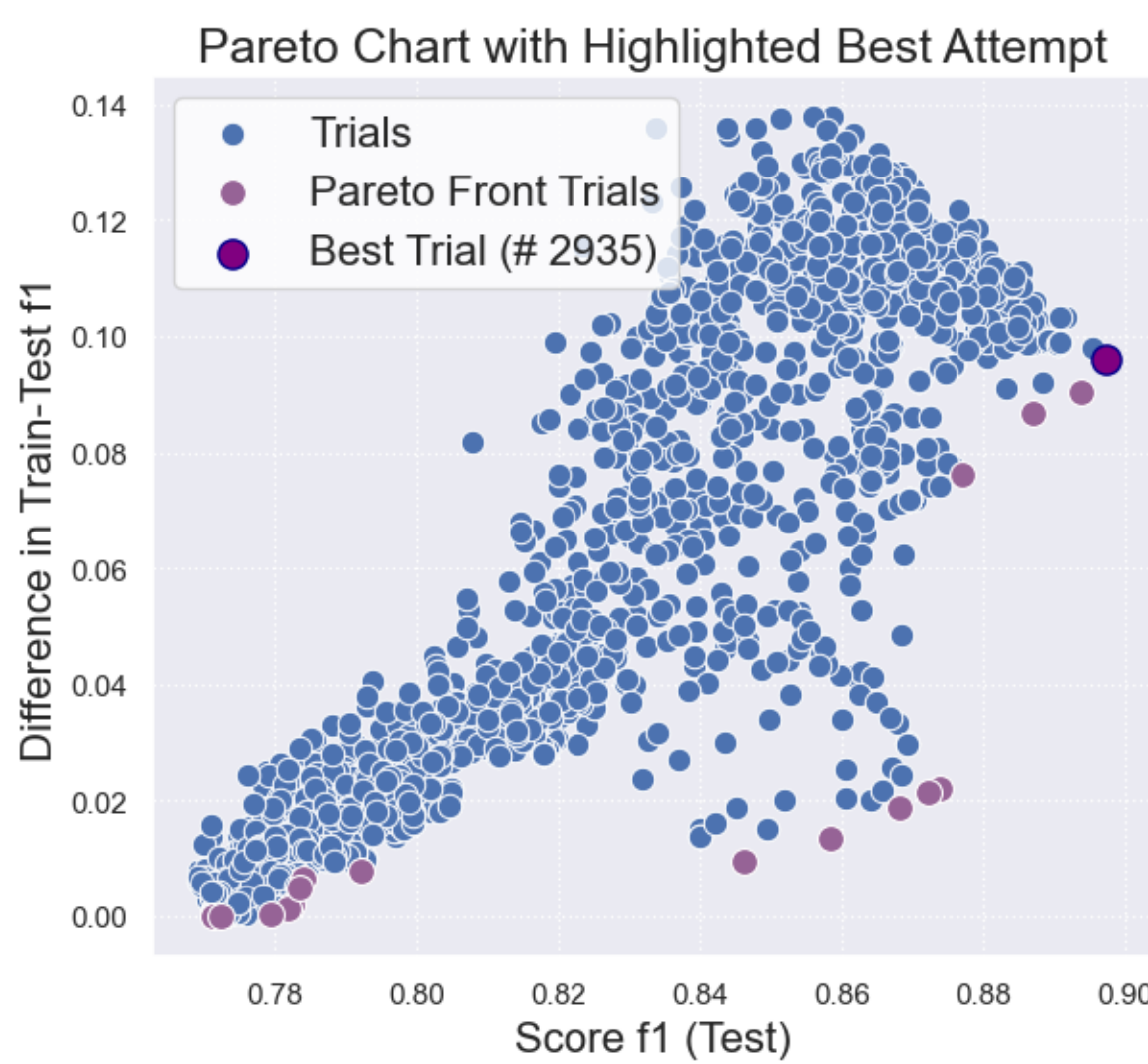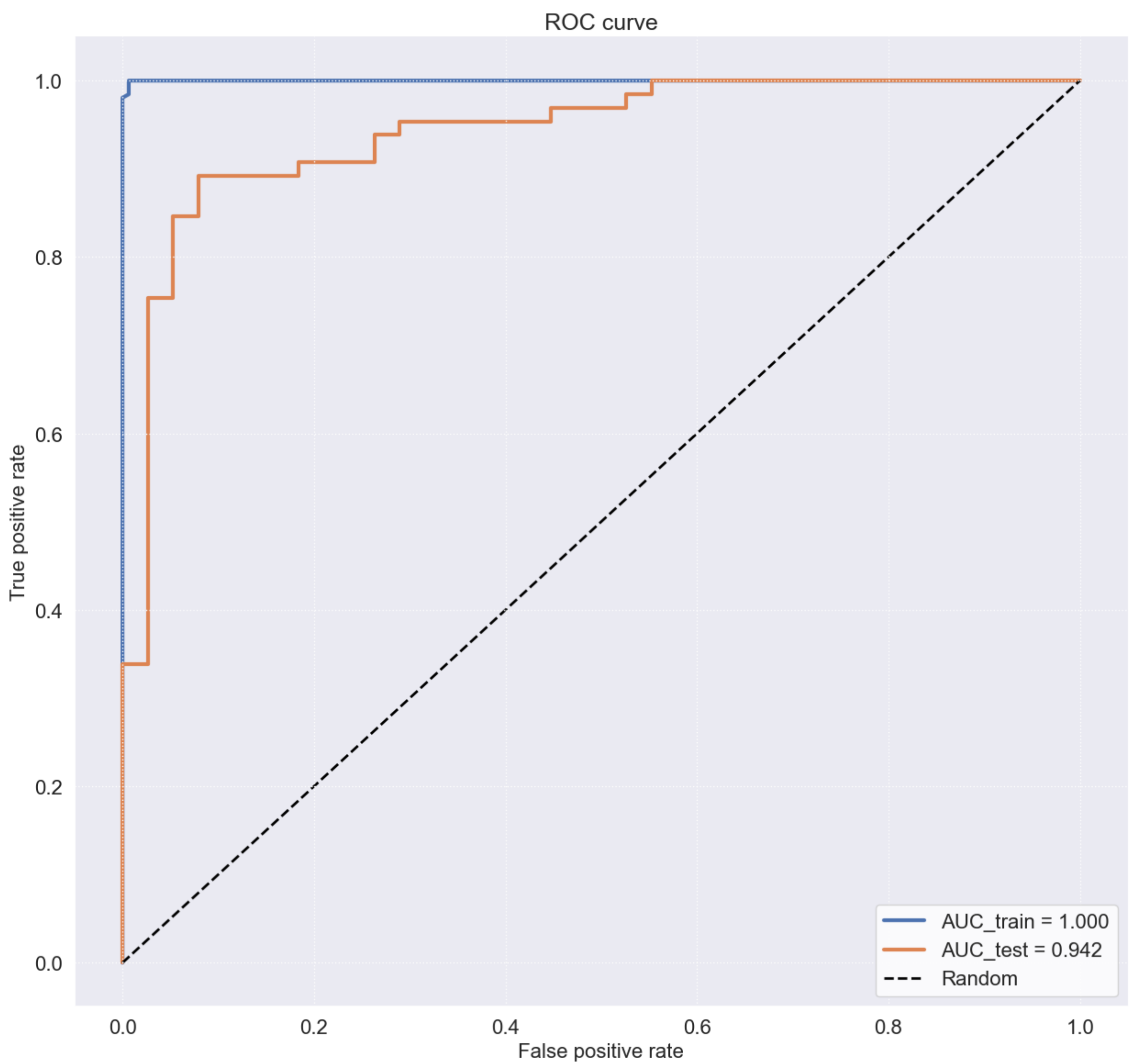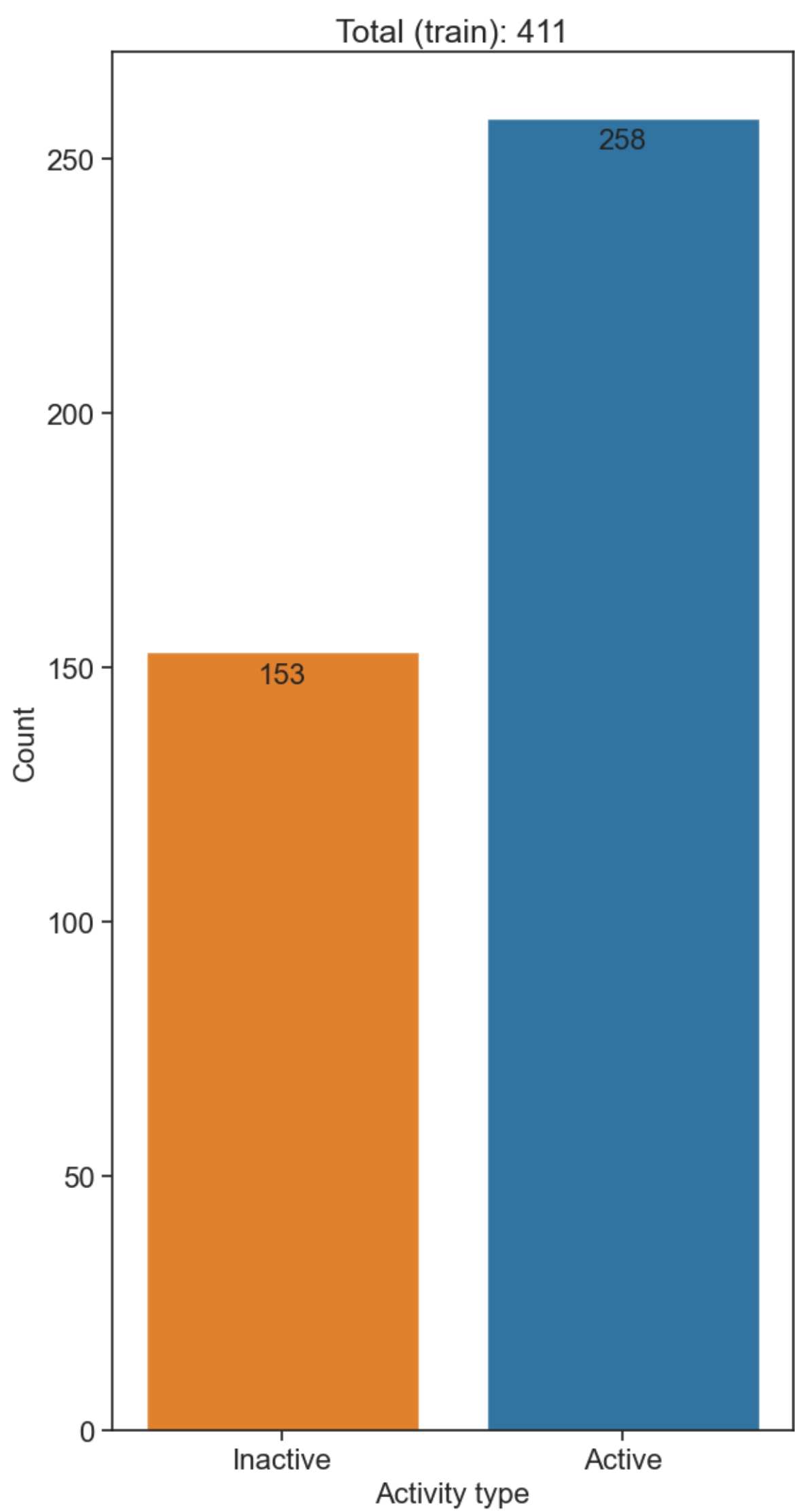
Inactive zone | Active zone

type
- active (65)
- inactive (38)
- decoy (103)
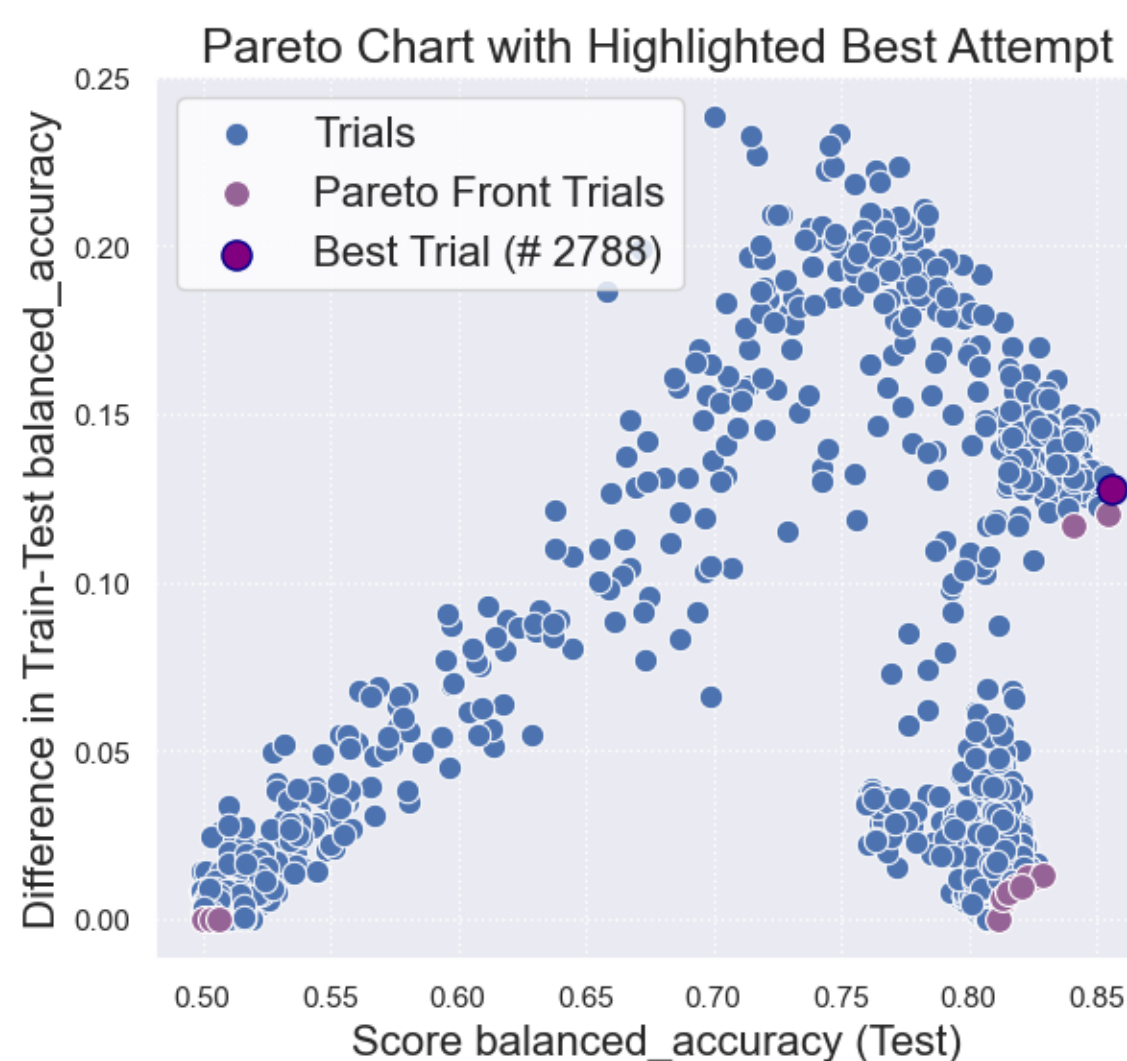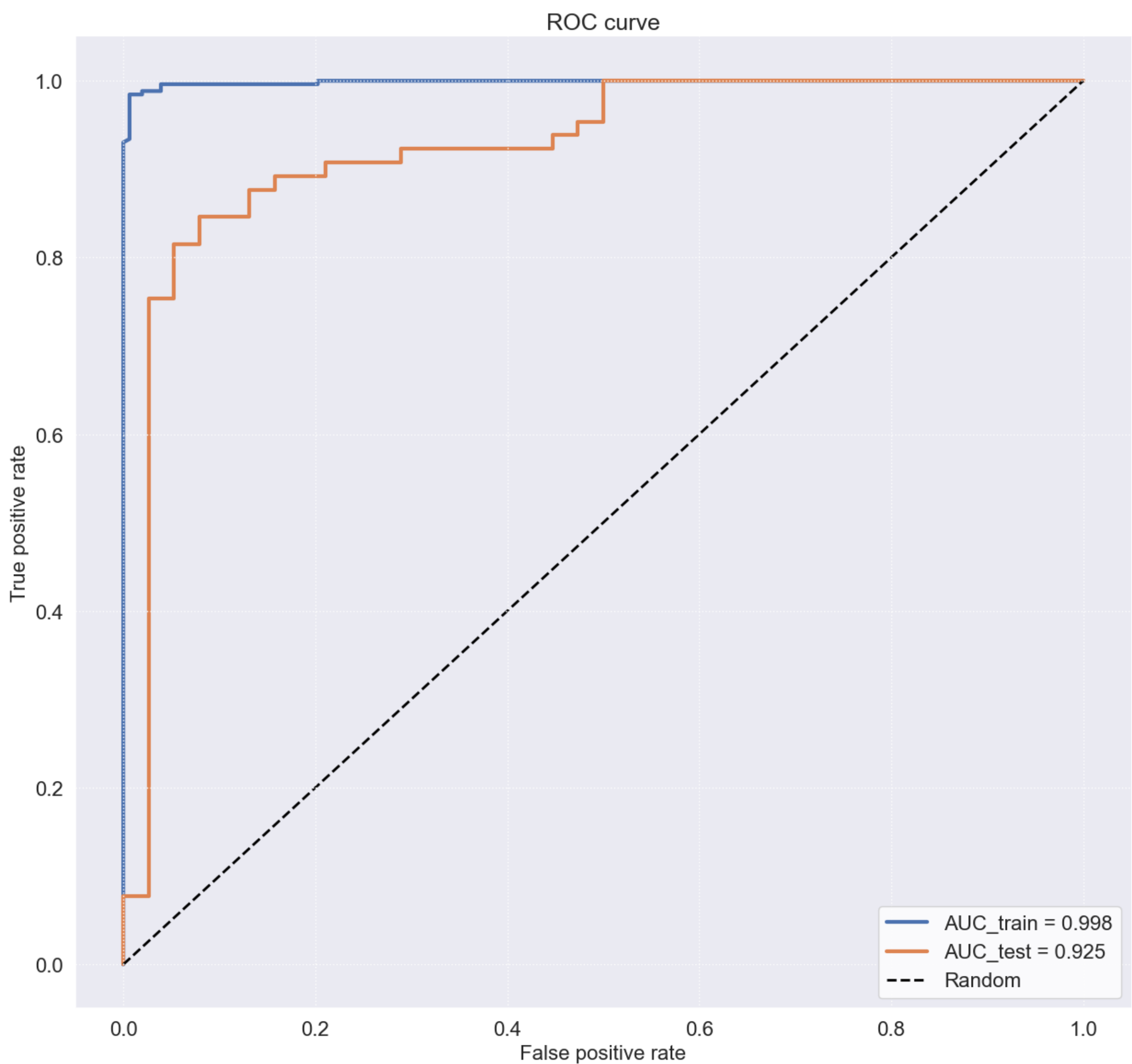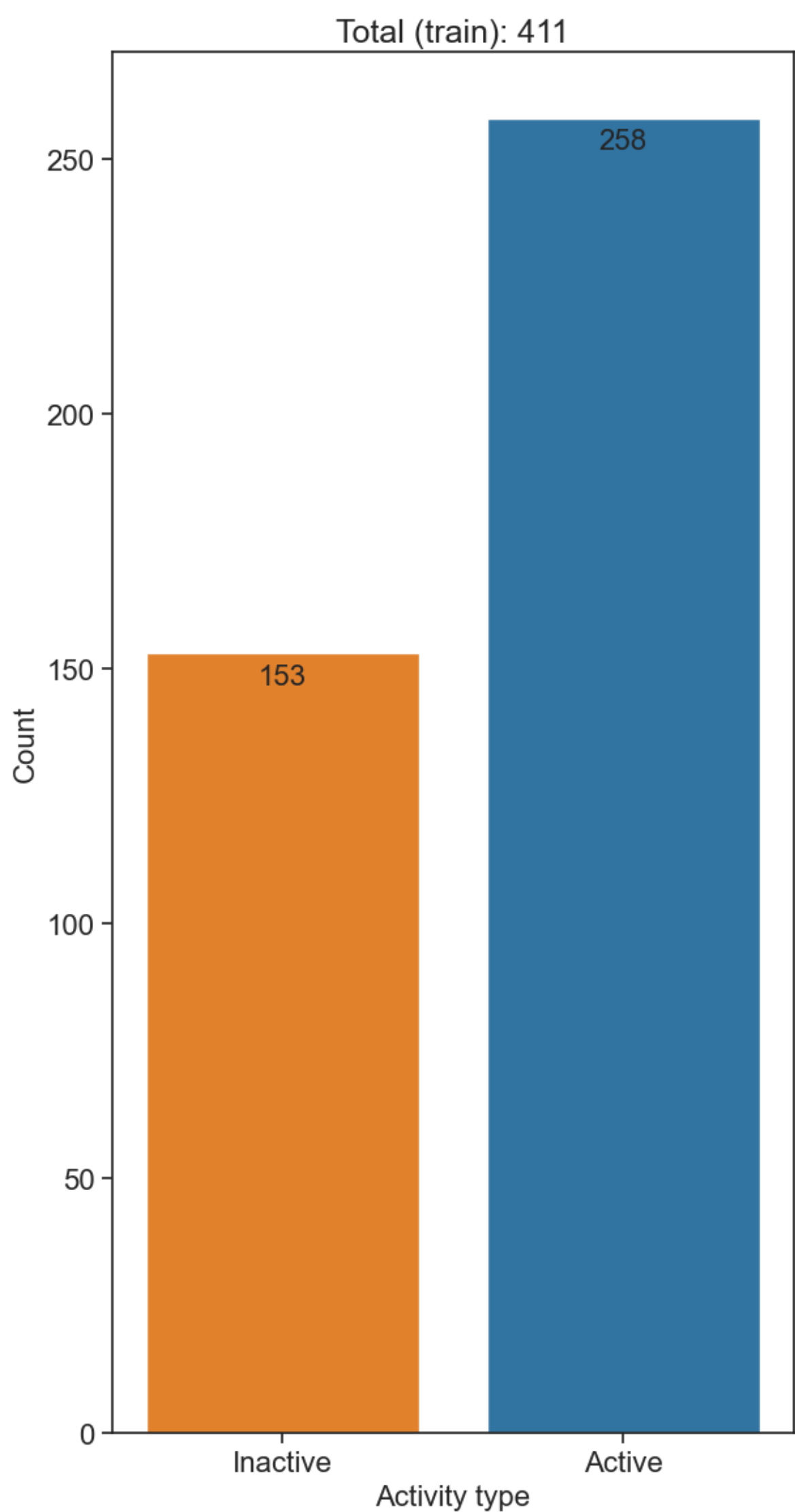
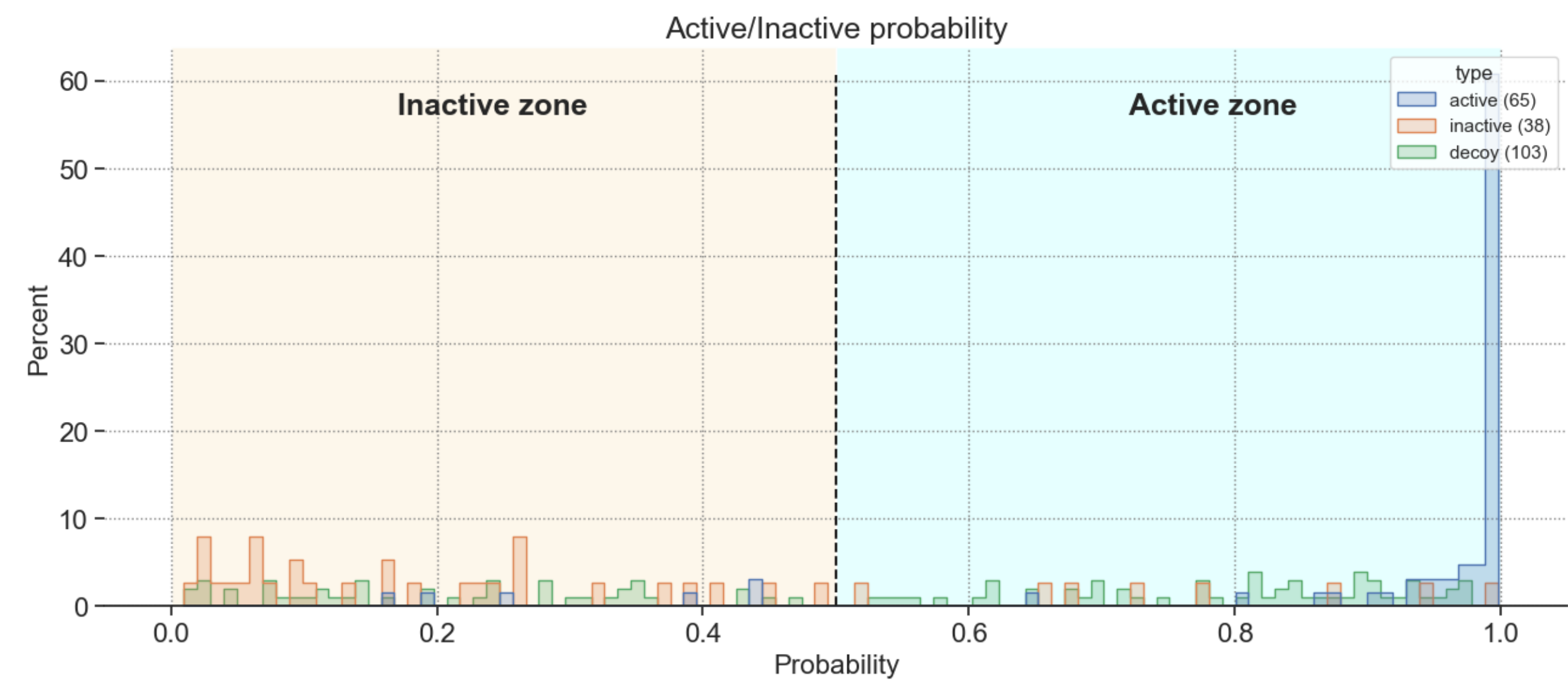| XGBoost Parameters	Value
alpha	1.4574
colsample_bytree	0.7944
gamma	0.963
lambda	0.0386
learning_rate	0.0998
max_depth	10.0
min_child_weight	2.0
n_estimators	878.0
reg_alpha	0.0532
reg_lambda	0.4472
scale_pos_weight	1.0672
subsample	0.608
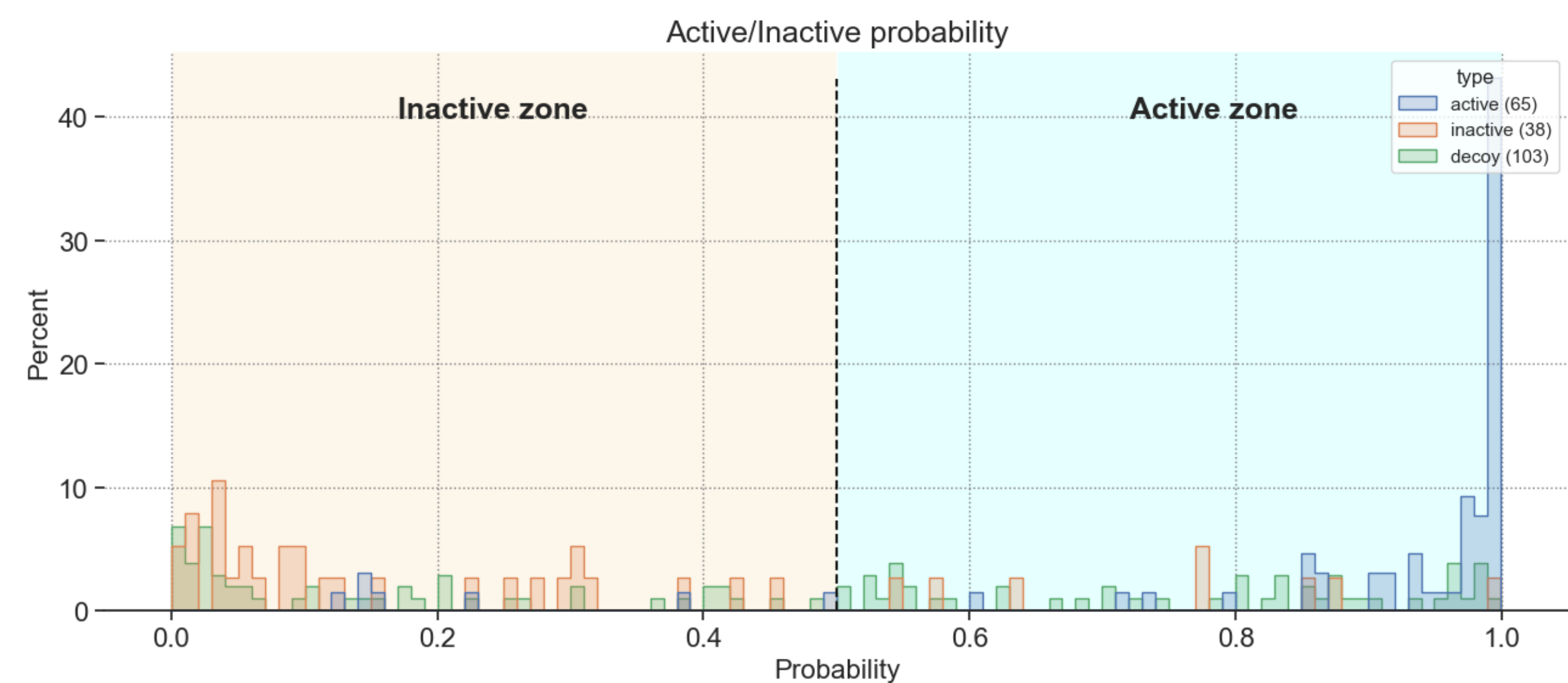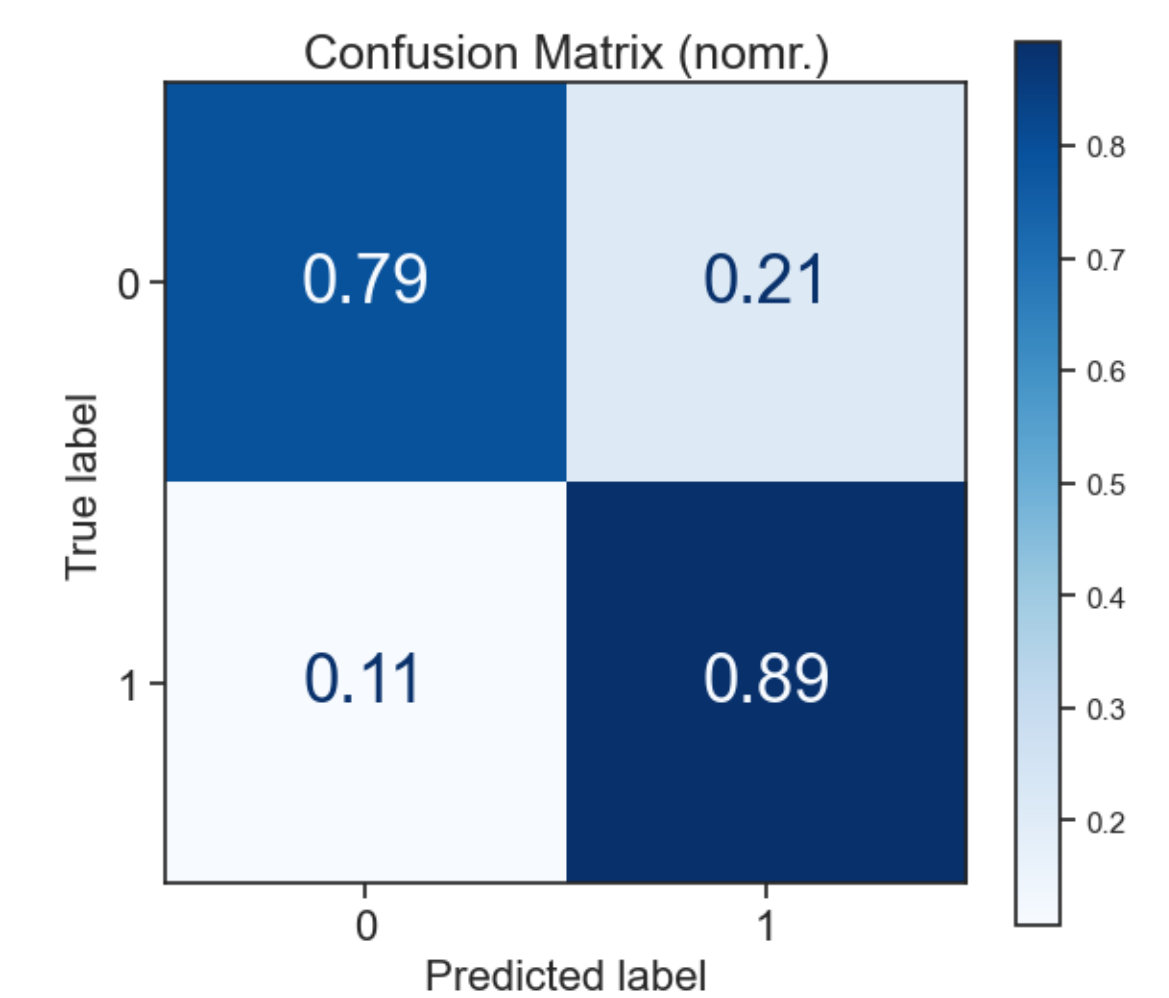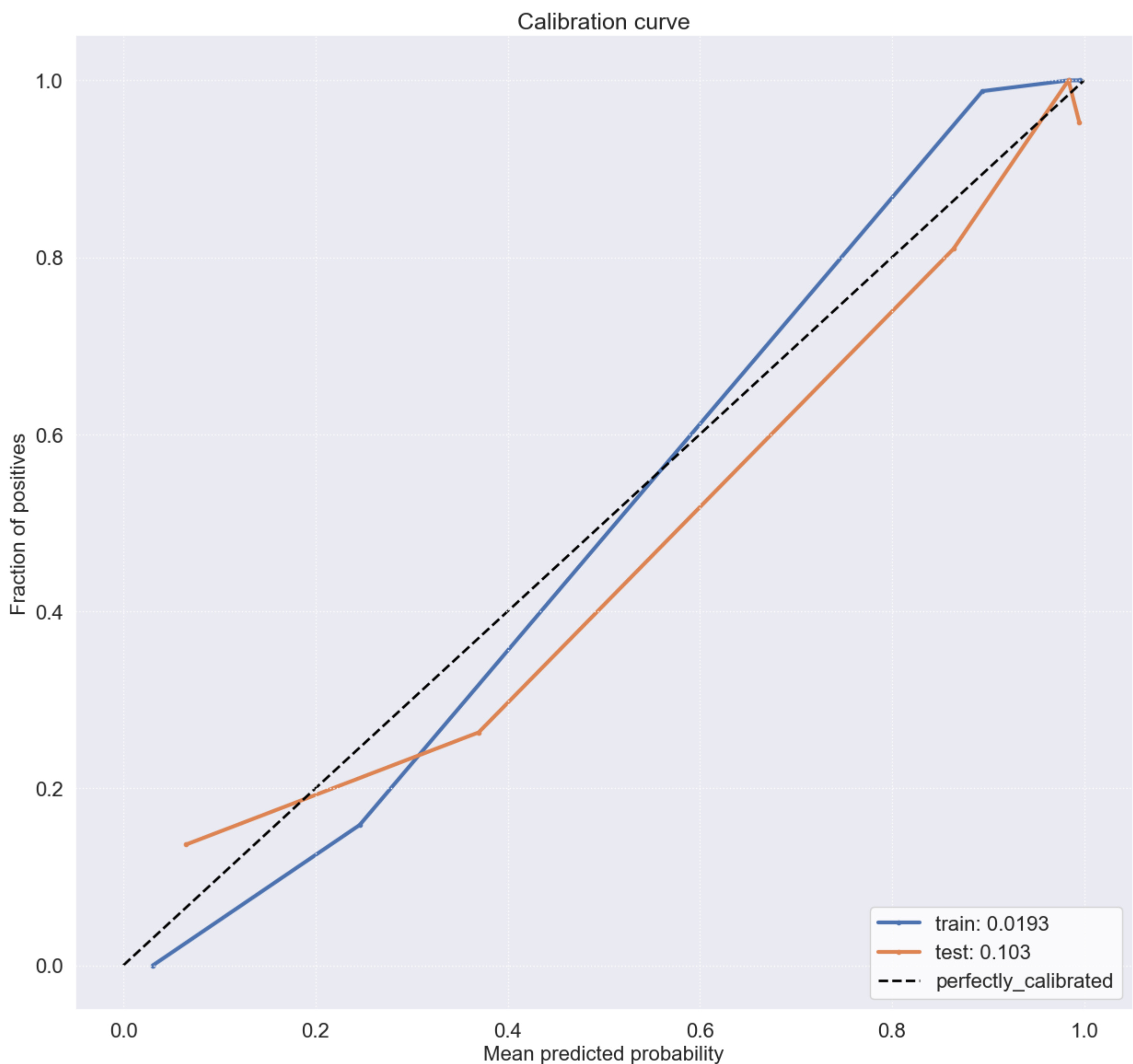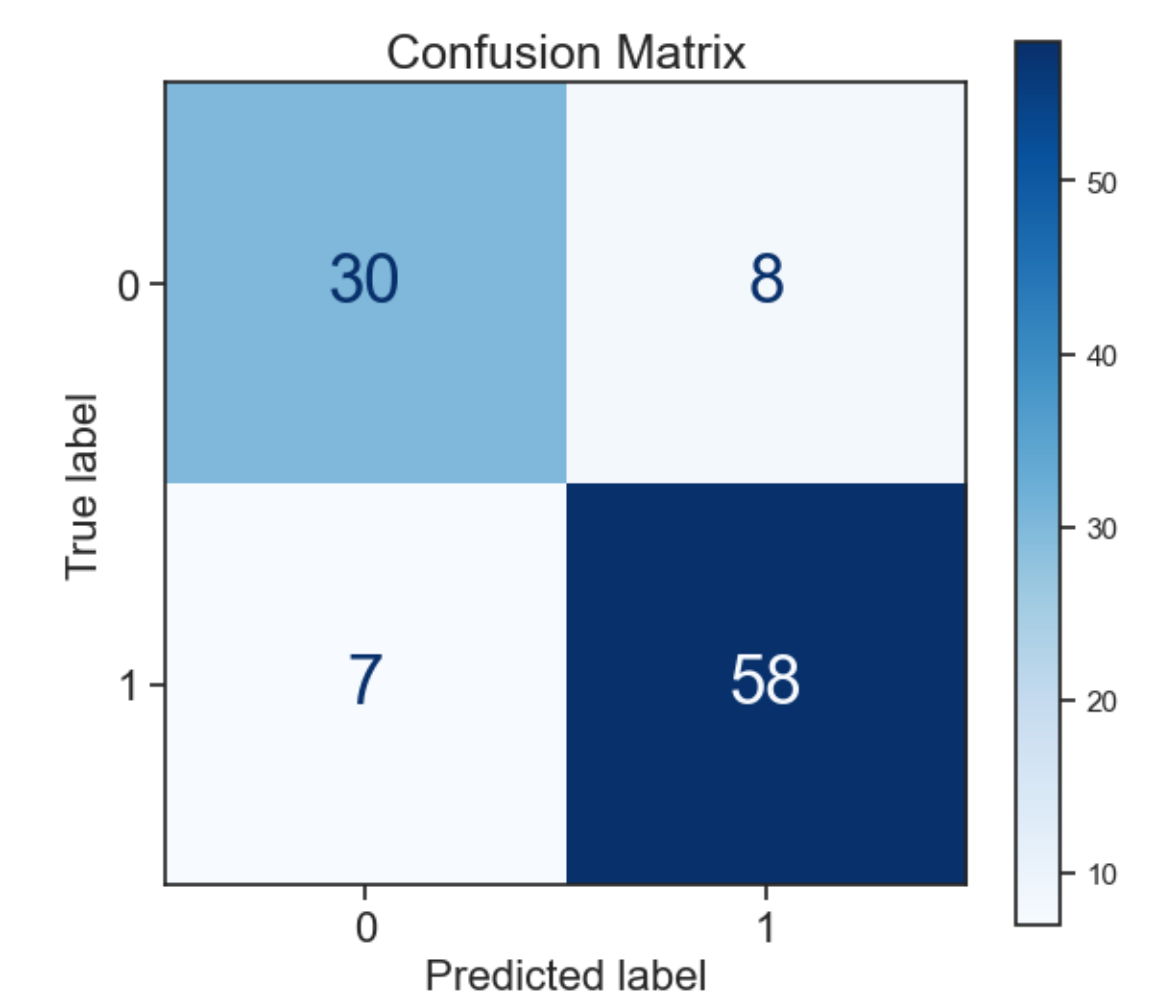
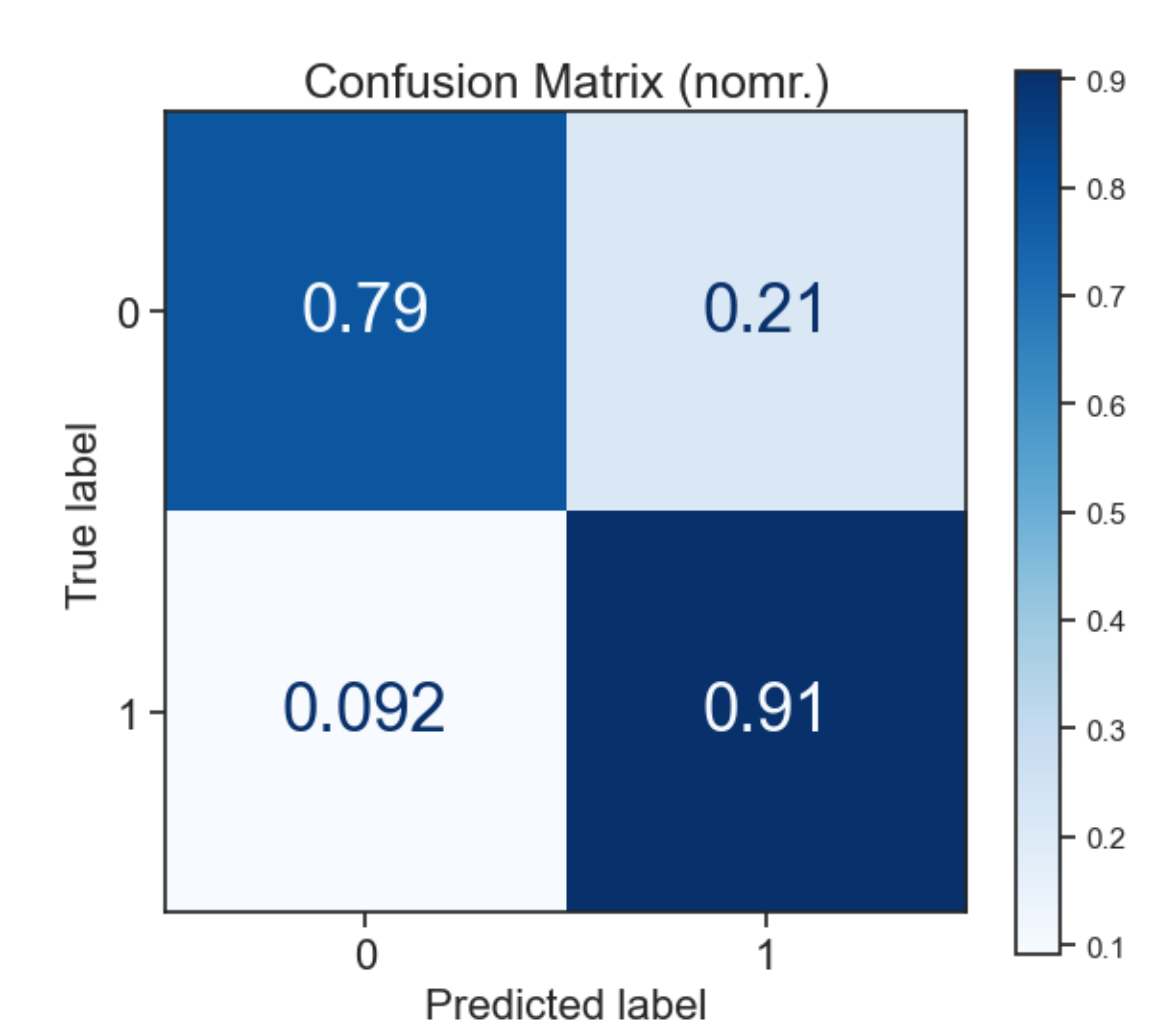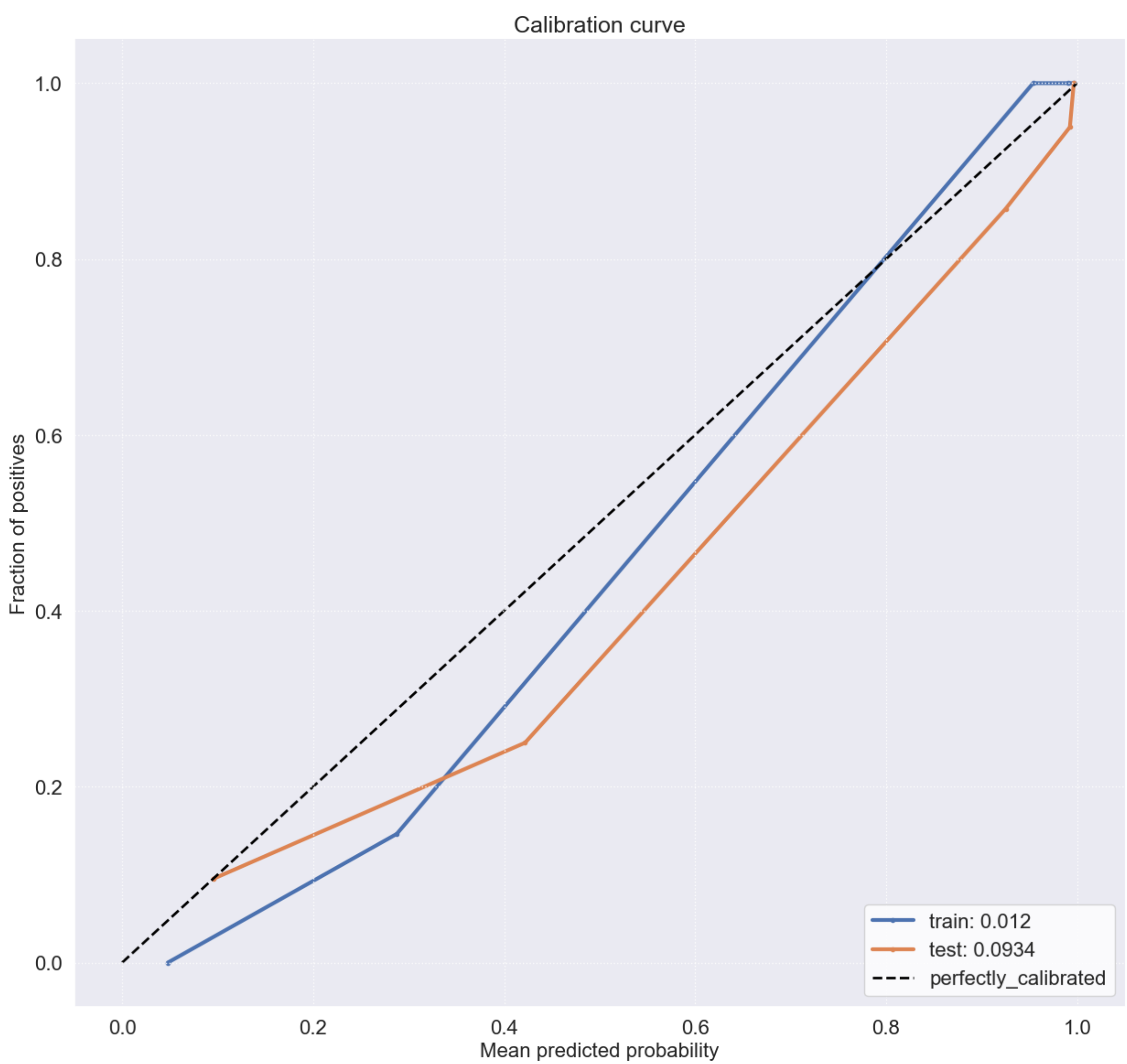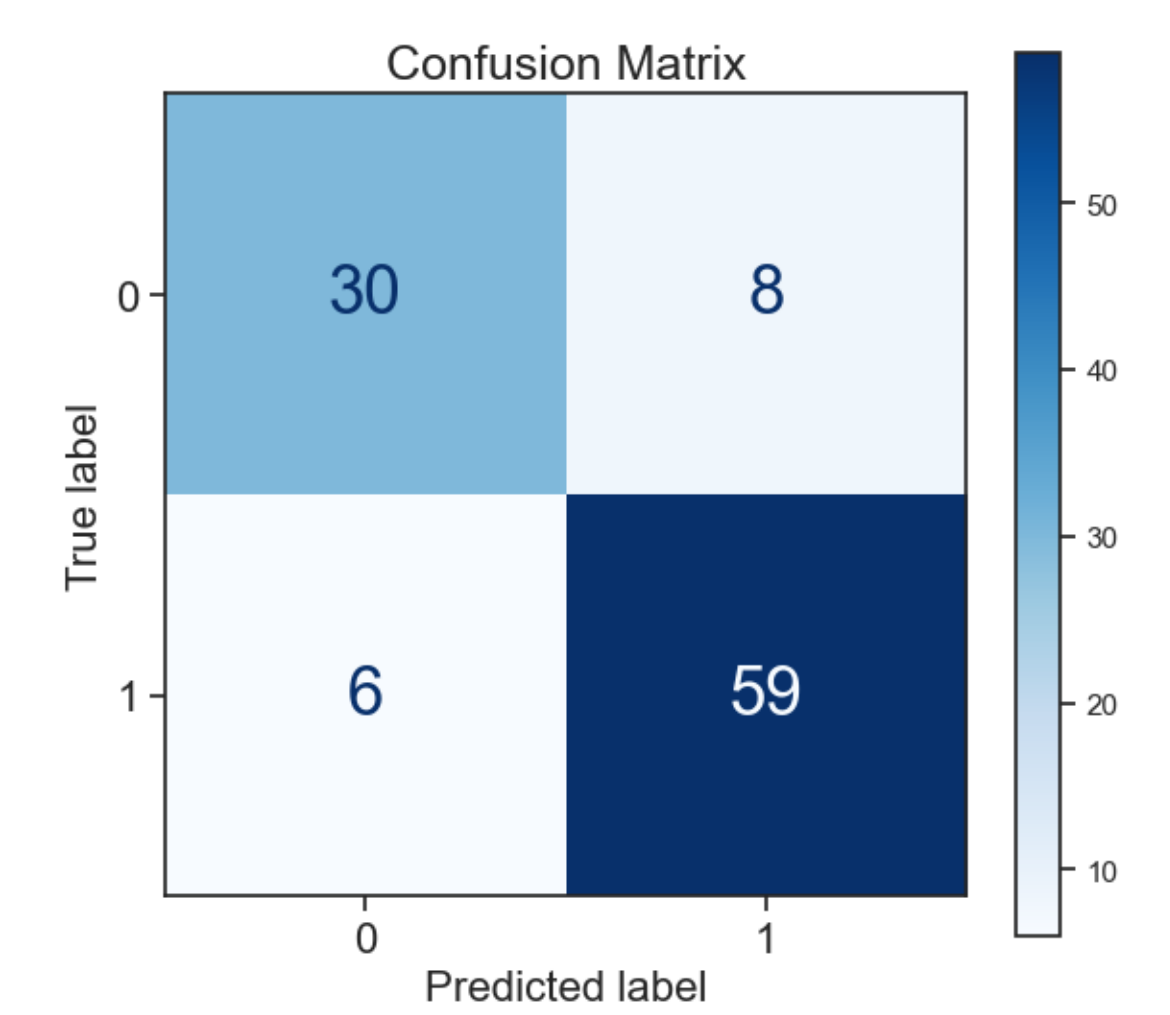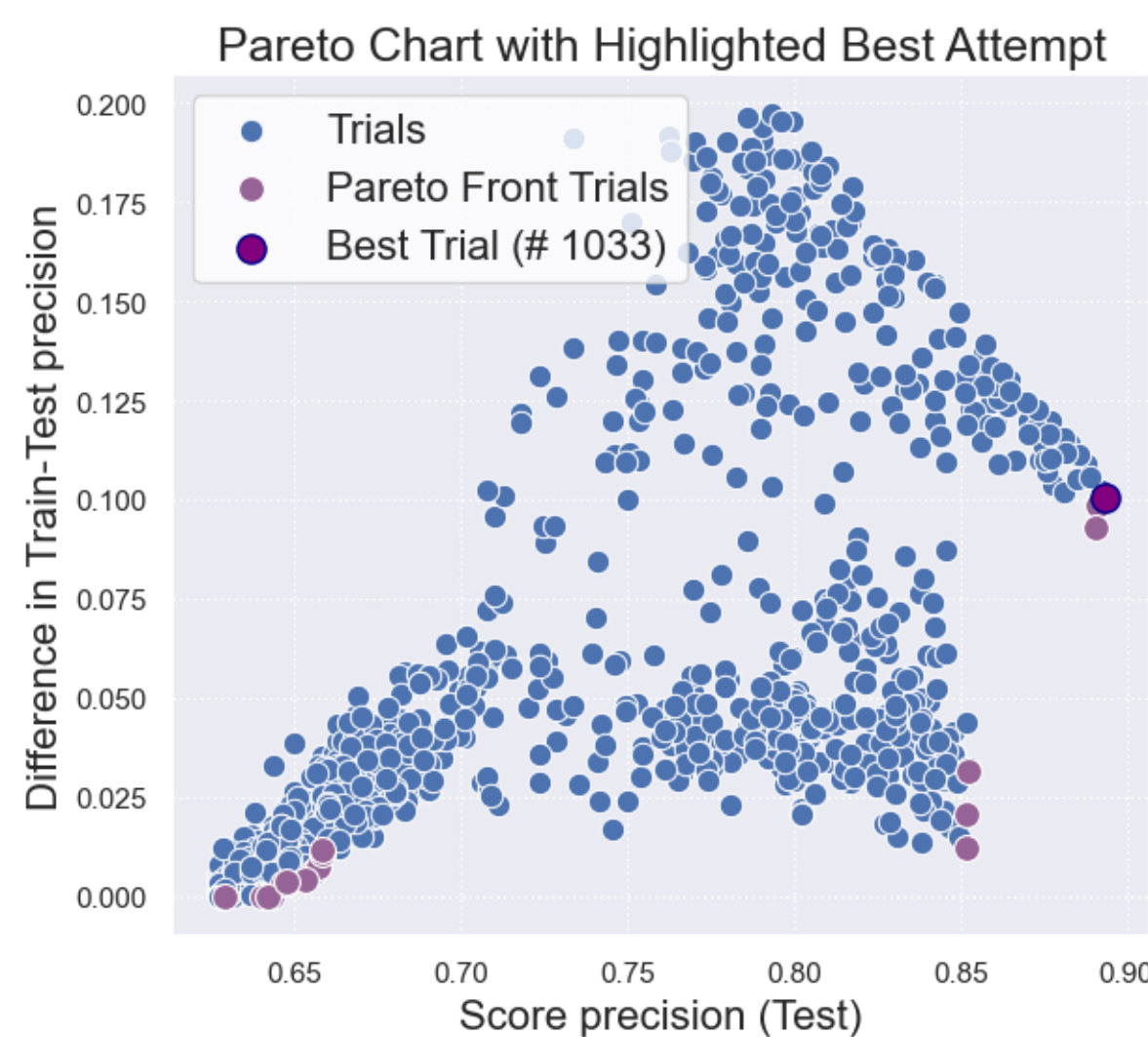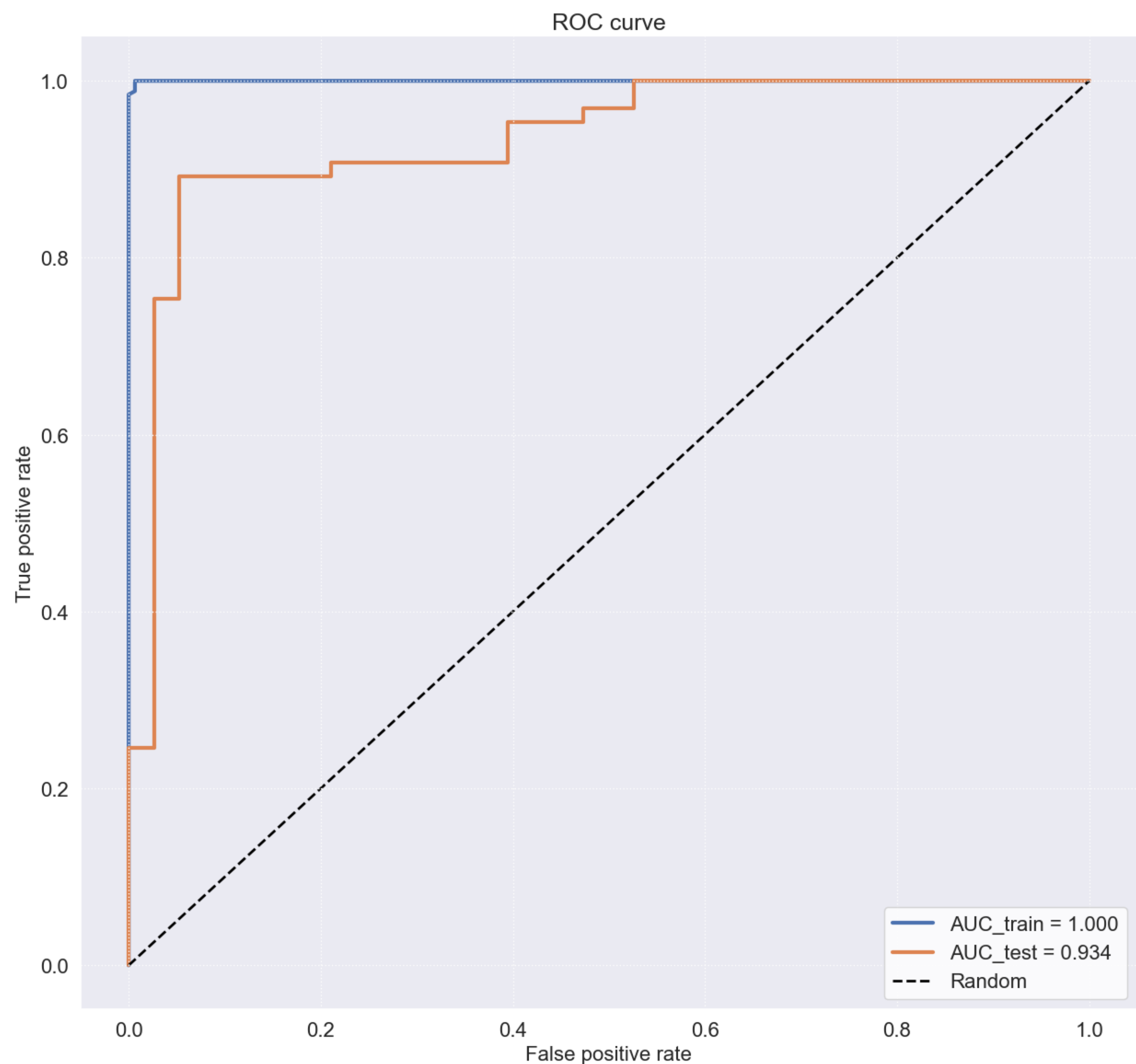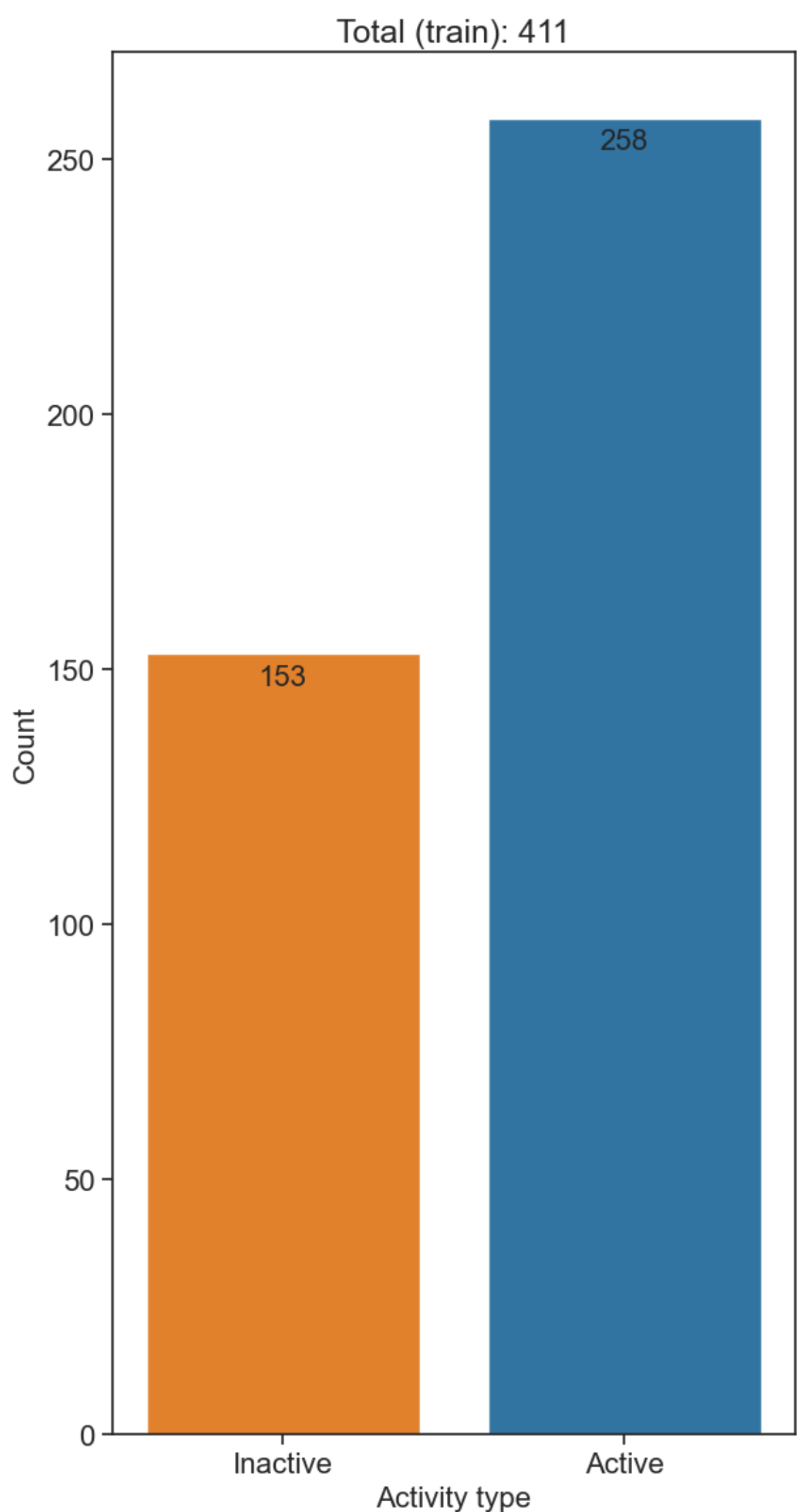| Metric	Train	Test
AUC	0.9985	0.9251
accuracy	0.9805	0.8544
balanced accuracy	0.9778	0.8409
recall_(sens)	0.9884	0.8923
specificity	0.9673	0.7895
precision	0.9808	0.8788
f1_score	0.9805	0.854

Protein: P06213 - metric: f1

Total (train): 411

Activity type	Count
Inactive	153
Active	258

ROC curve

- AUC_train = 1.000
- AUC_test = 0.942
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2935)

X-axis: Score f1 (Test)
Y-axis: Difference in Train-Test f1

Confusion Matrix

	Predicted 0	Predicted 1
True 0	30	8
True 1	6	59

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.79	0.21
True 1	0.092	0.91

Calibration curve

- train: 0.012
- test: 0.0934
- perfectly_calibrated

Active/Inactive probability

- Inactive zone
- Active zone

type:
- active (65)
- inactive (38)
- decoy (103)

| XGBoost Parameters	Value
alpha	3.533
colsample_bytree	0.7193
gamma	0.7841
lambda	4.5179
learning_rate	0.023
max_depth	10.0
min_child_weight	1.0
n_estimators	502.0
reg_alpha	0.1448
reg_lambda	0.0253
scale_pos_weight	2.6219
subsample	0.7166

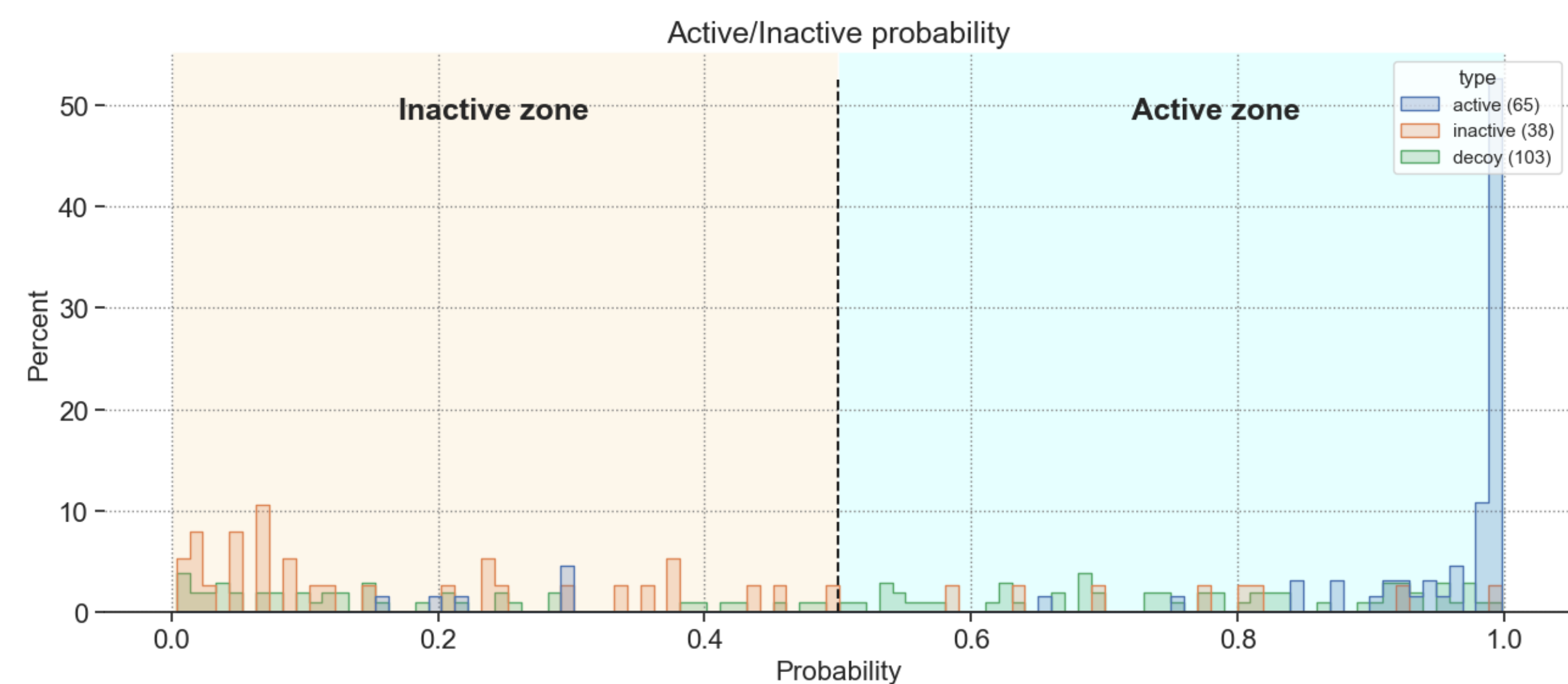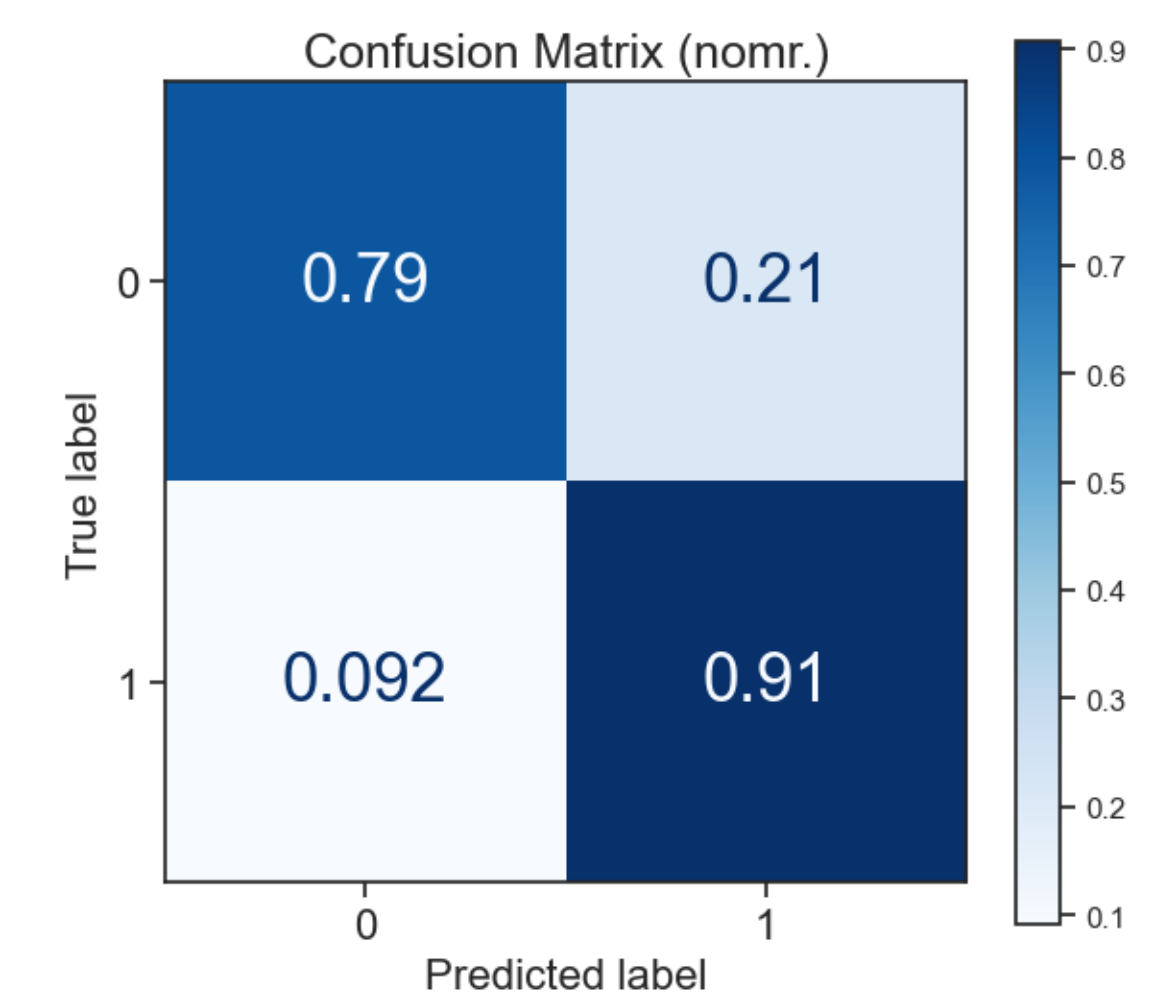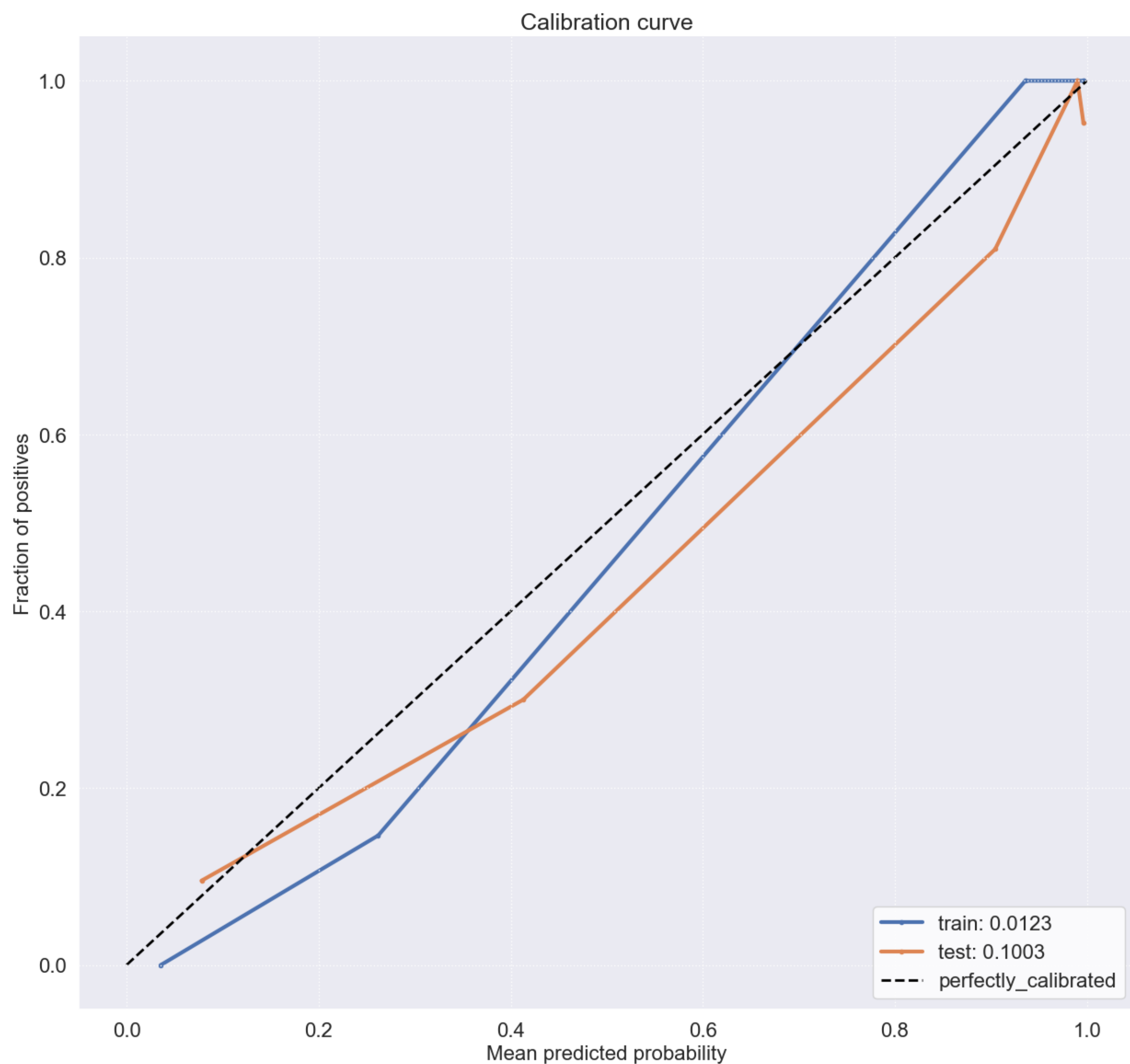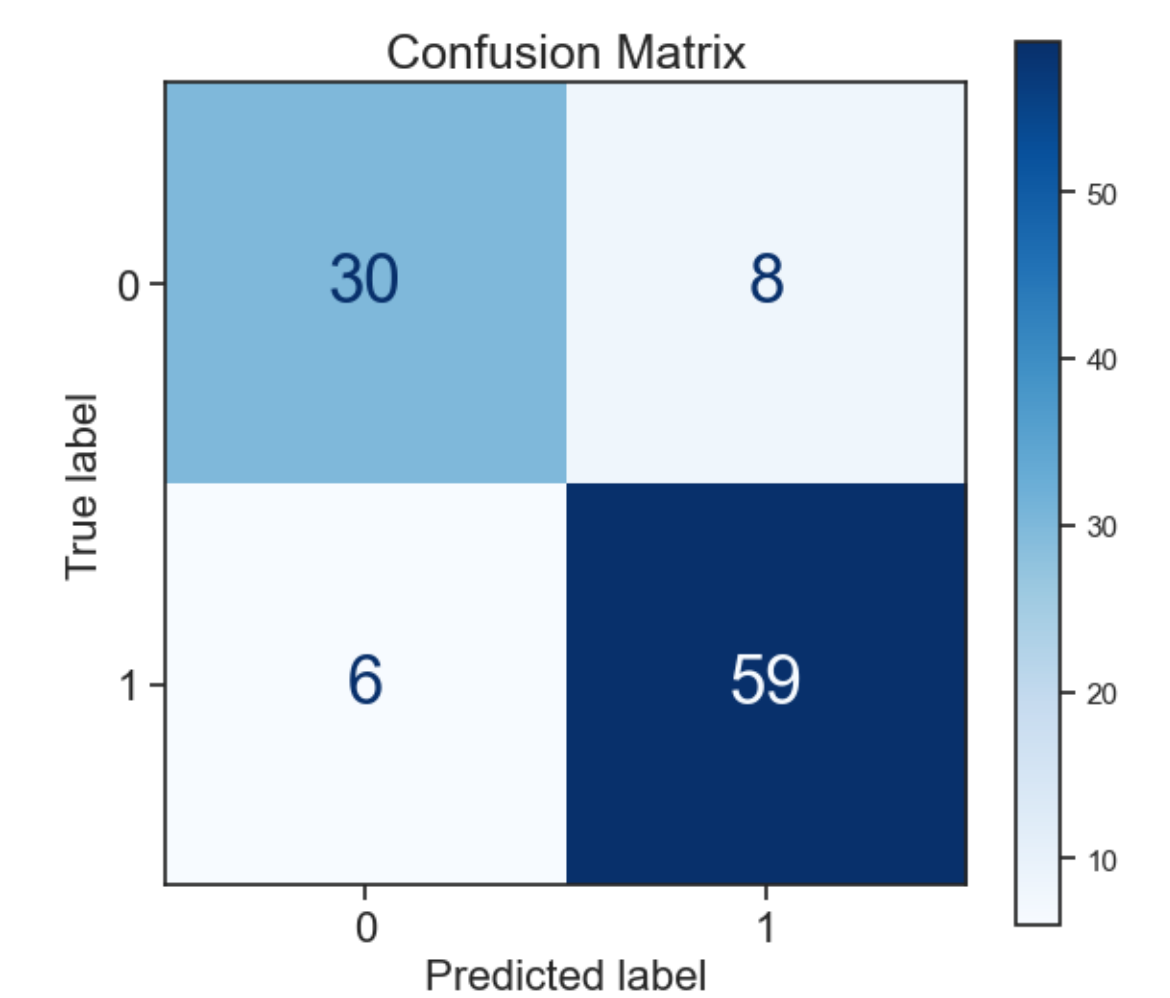| Metric	Train	Test
AUC	0.9999	0.9417
accuracy	0.9927	0.8641
balanced accuracy	0.9902	0.8486
recall_(sens)	1.0	0.9077
specificity	0.9804	0.7895
precision	0.9885	0.8806
f1_score	0.9927	0.8633

Protein: P06213 - metric: precision

Total (train): 411

ROC curve

AUC_train = 1.000
AUC_test = 0.934
Random

Pareto Chart with Highlighted Best Attempt

Trials
Pareto Front Trials
Best Trial (# 1033)

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

train: 0.0092
test: 0.1158
perfectly_calibrated

Active/Inactive probability


Inactive zone | Active zone

type
active (65)
inactive (38)
decoy (103)

XGBoost Parameters	Value
alpha	2.728
colsample_bytree	0.7487
gamma	0.5311
lambda	3.6587
learning_rate	0.1589
max_depth	5.0
min_child_weight	1.0
n_estimators	975.0
reg_alpha	0.3405
reg_lambda	0.8213
scale_pos_weight	2.2065
subsample	0.6879
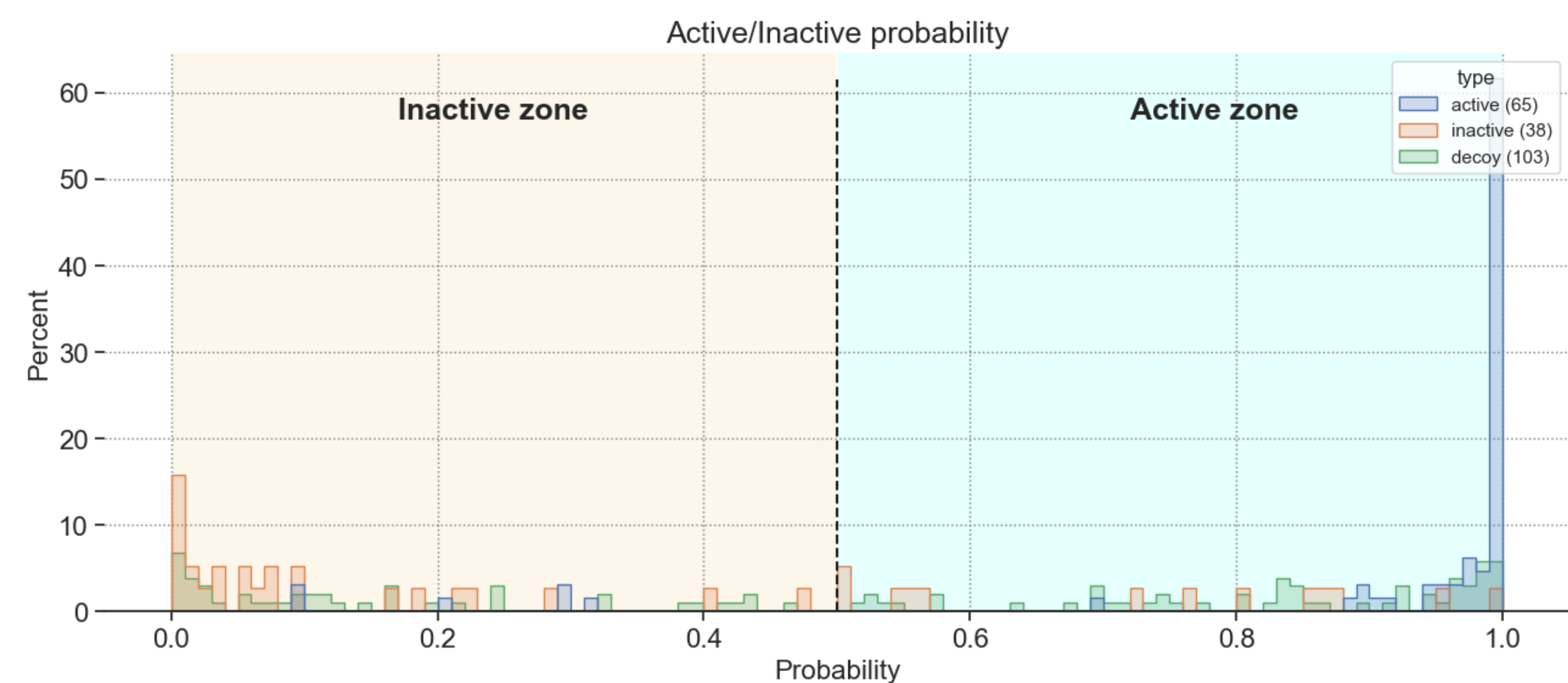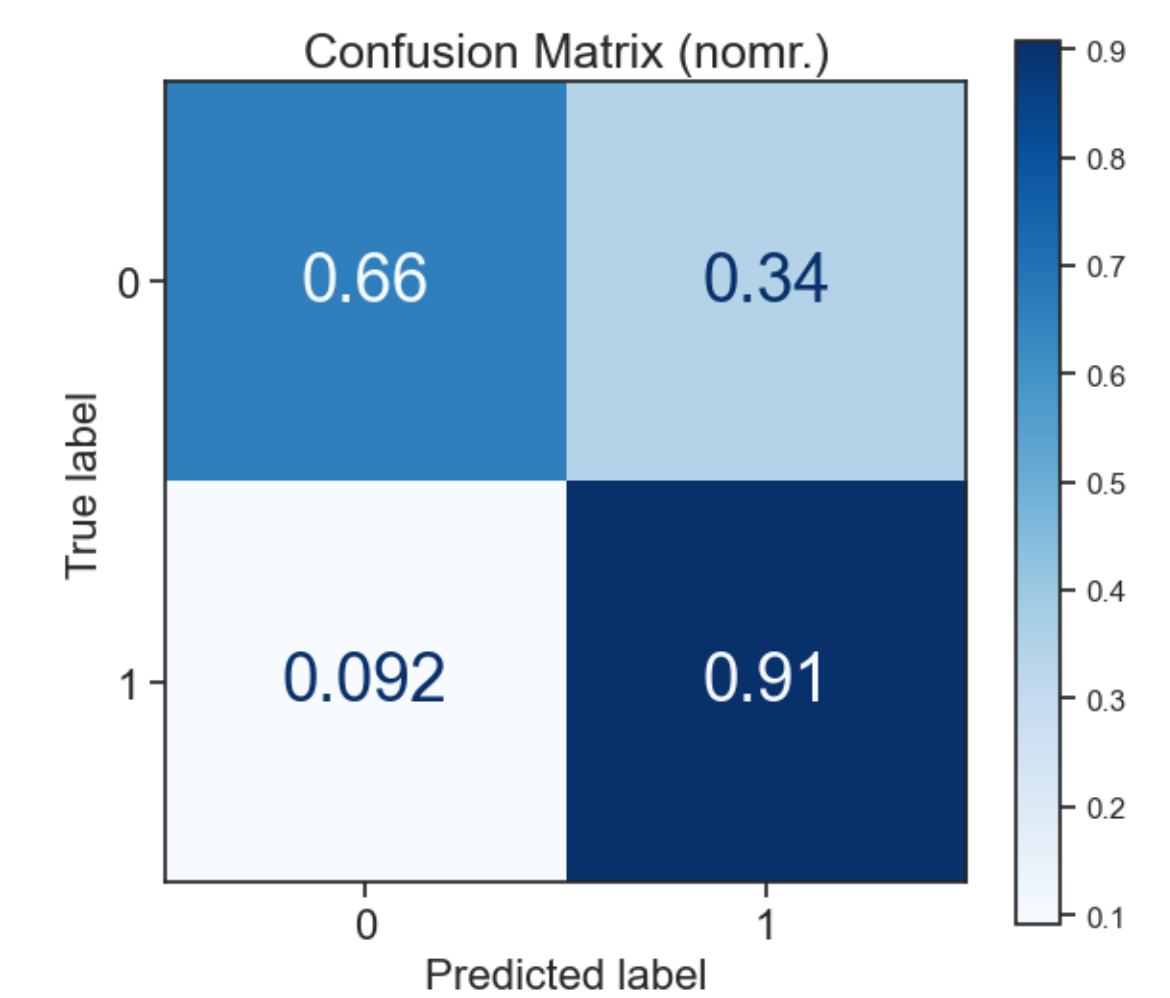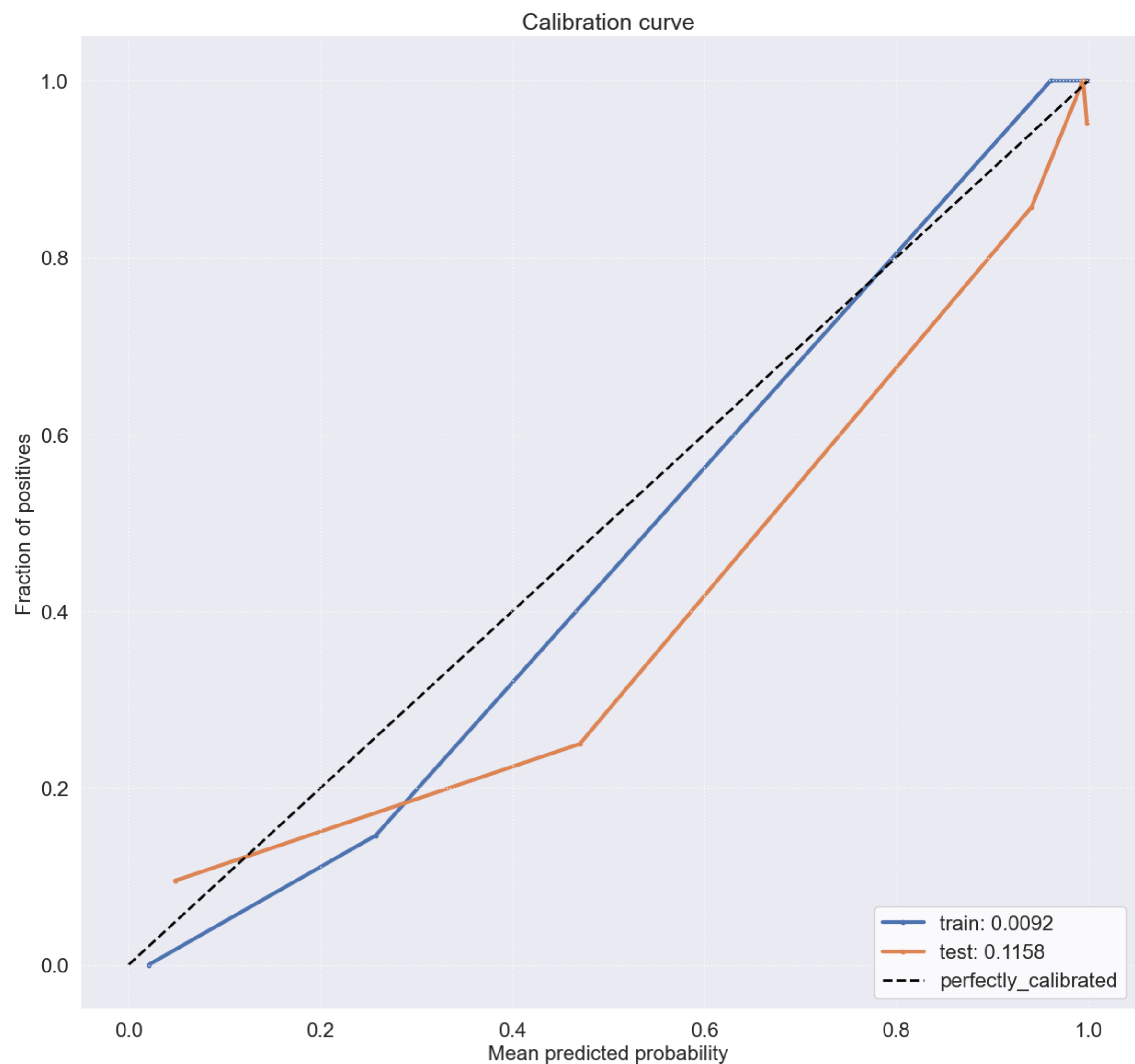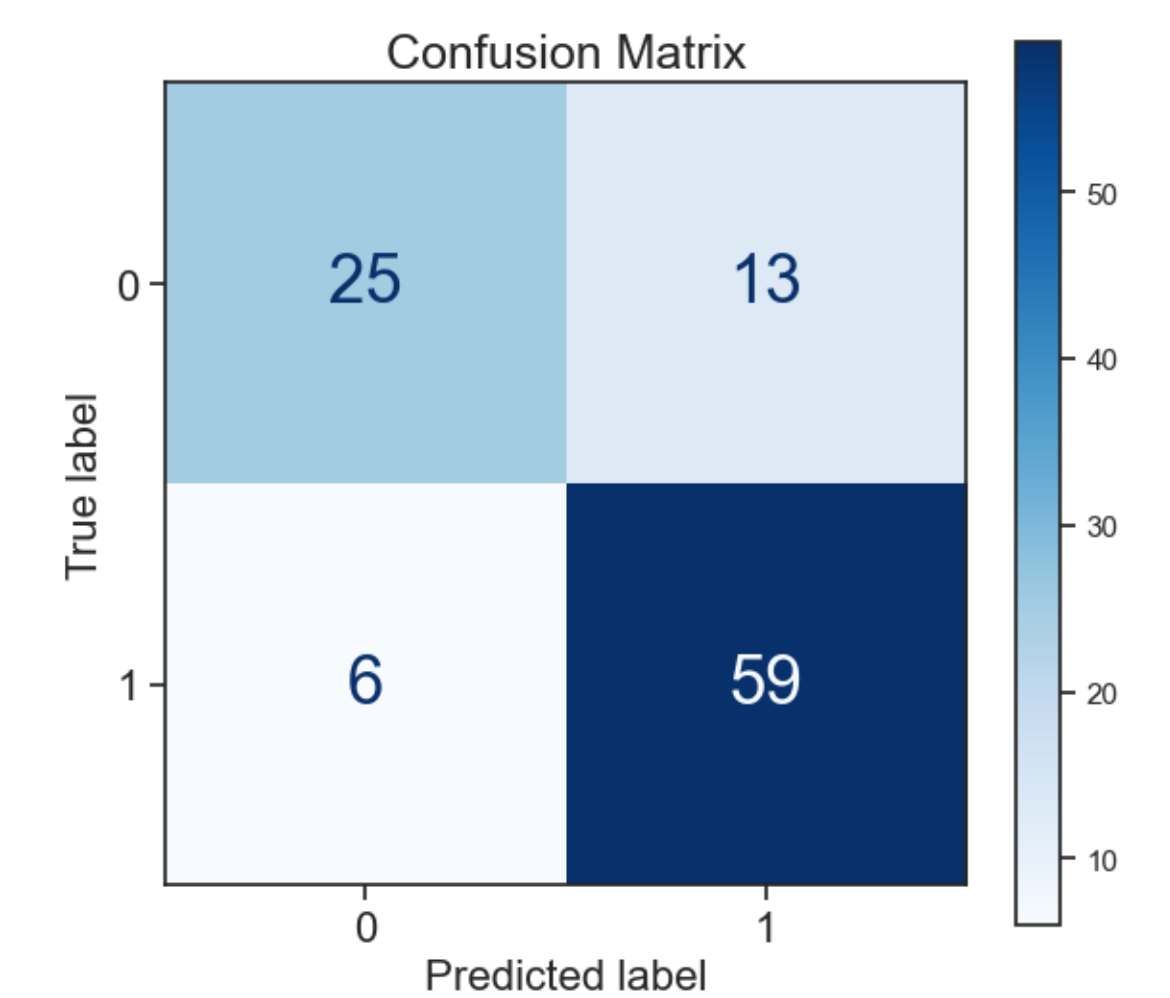
Metric	Train	Test
AUC	0.9999	0.9344
accuracy	0.9927	0.8155
balanced accuracy	0.9902	0.7828
recall_(sens)	1.0	0.9077
specificity	0.9804	0.6579
precision	0.9885	0.8194
f1_score	0.9927	0.8109

Protein: P06213 - metric: SoftVotingXGBoostEnsemble
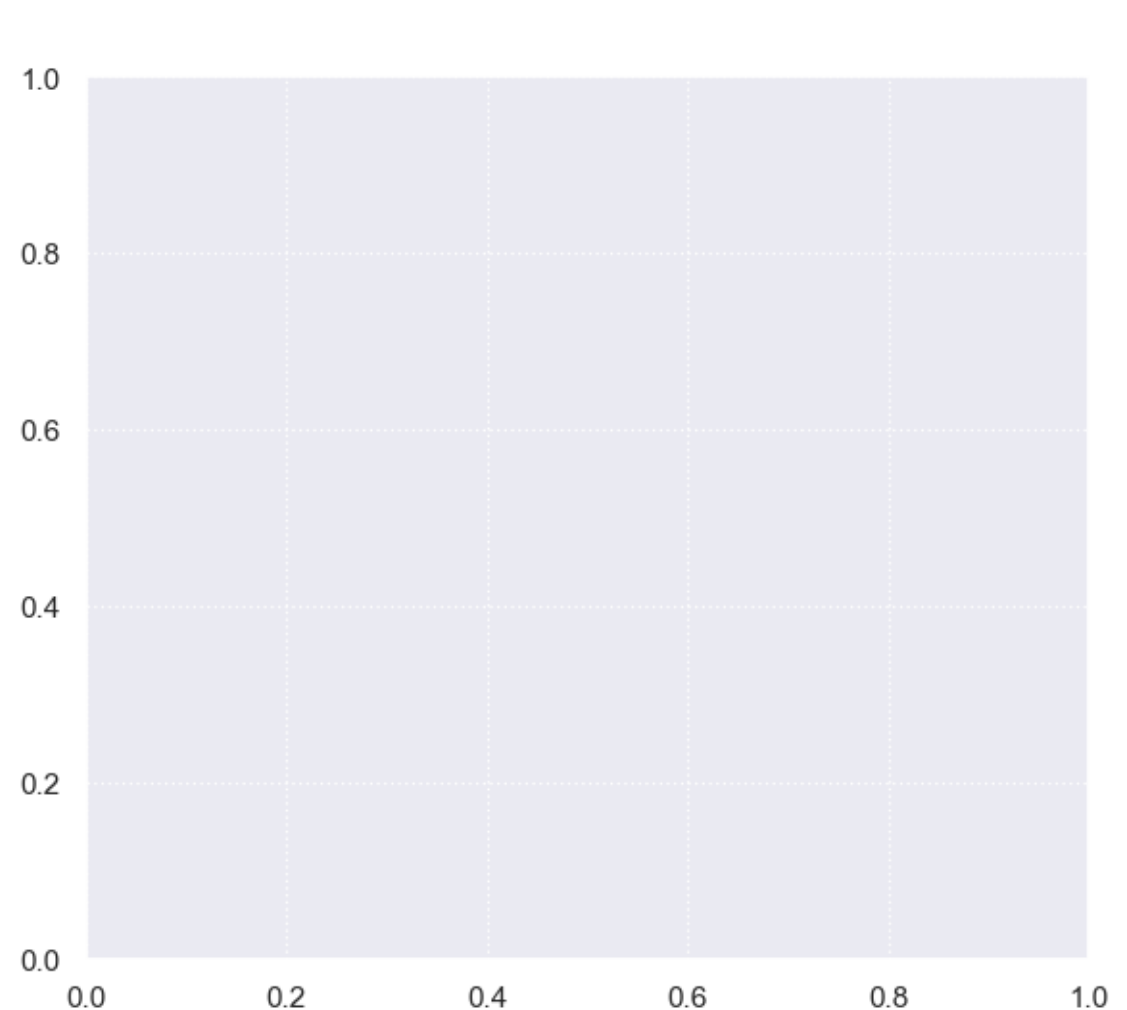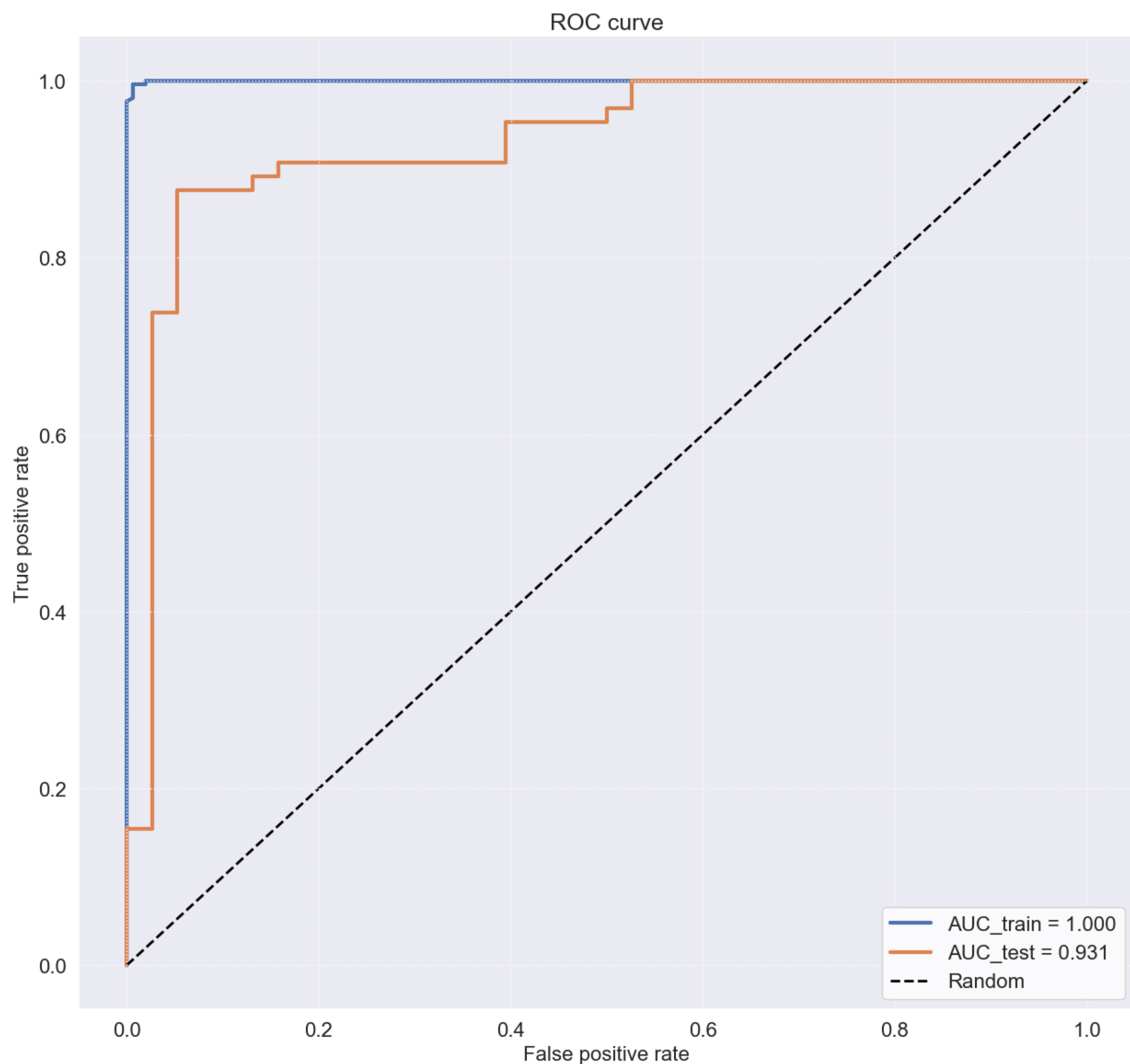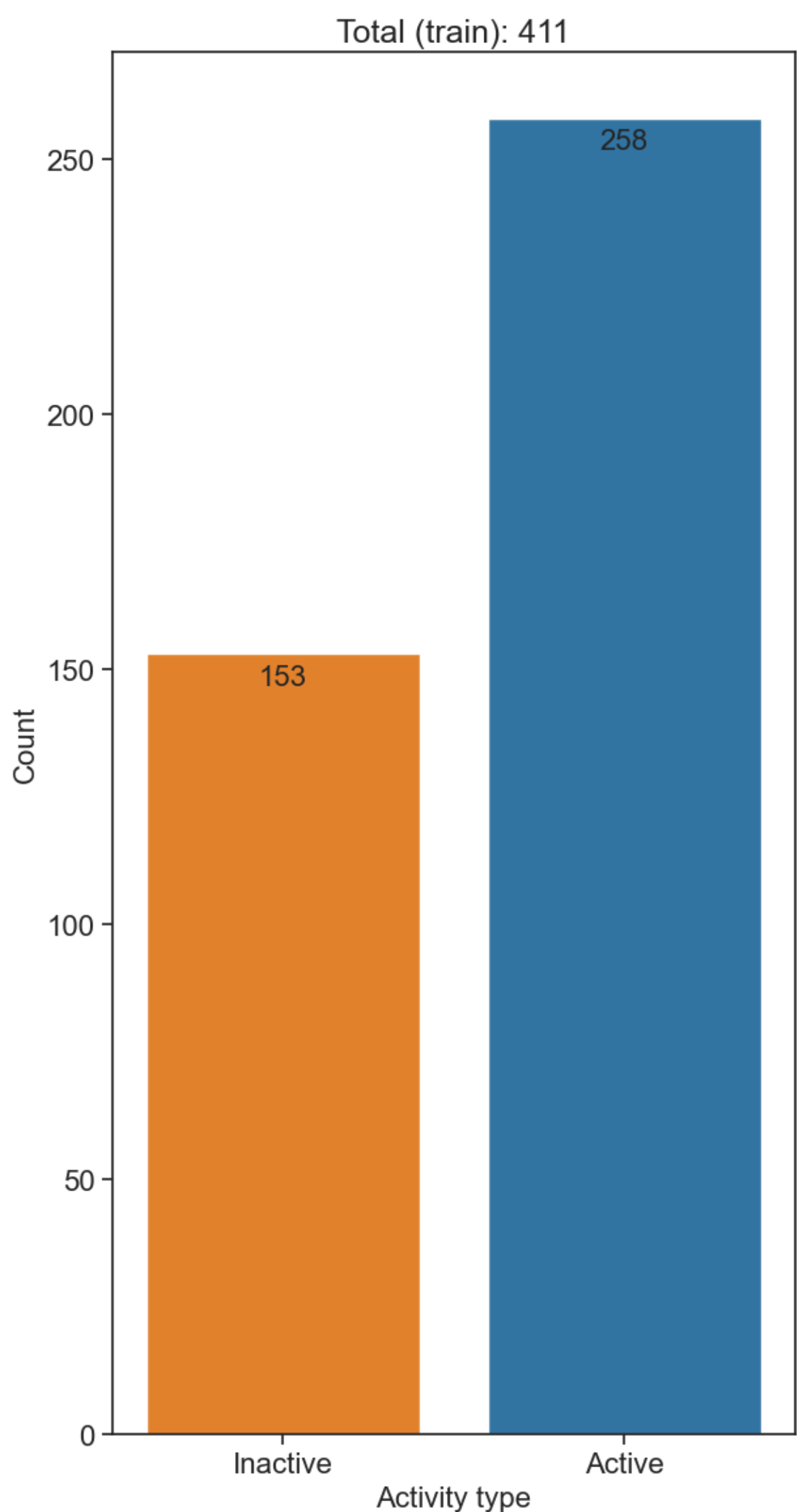
Metric	Train	Test
AUC	0.9998	0.9308
accuracy	0.9903	0.8641
balanced accuracy	0.9869	0.8486
recall_(sens)	1.0	0.9077
specificity	0.9739	0.7895
precision	0.9847	0.8806
f1_score	0.9902	0.8633