Protein: P27338 - metric: balanced_accuracy

Total (train): 2088

Active: 728
Inactive: 1360

ROC curve

- AUC_train = 1.000
- AUC_test = 0.943
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2457)

x-axis: Score balanced_accuracy (Test)
y-axis: Difference in Train-Test balanced_accuracy

Confusion Matrix

	Predicted 0	Predicted 1
True 0	139	43
True 1	28	312

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.76	0.24
True 1	0.082	0.92

Calibration curve

- train: 0.007
- test: 0.0949
- perfectly_calibrated

Active/Inactive probability

Inactive zone | Active zone

type:
- active (340)
- inactive (182)
- decoy (520)

XGBoost Parameters	Value
alpha	2.9067
colsample_bytree	0.5045
gamma	0.0659
lambda	3.5222
learning_rate	0.0743
max_depth	9.0
min_child_weight	1.0
n_estimators	992.0
reg_alpha	0.6193
reg_lambda	0.7244
scale_pos_weight	1.2094
subsample	0.7433
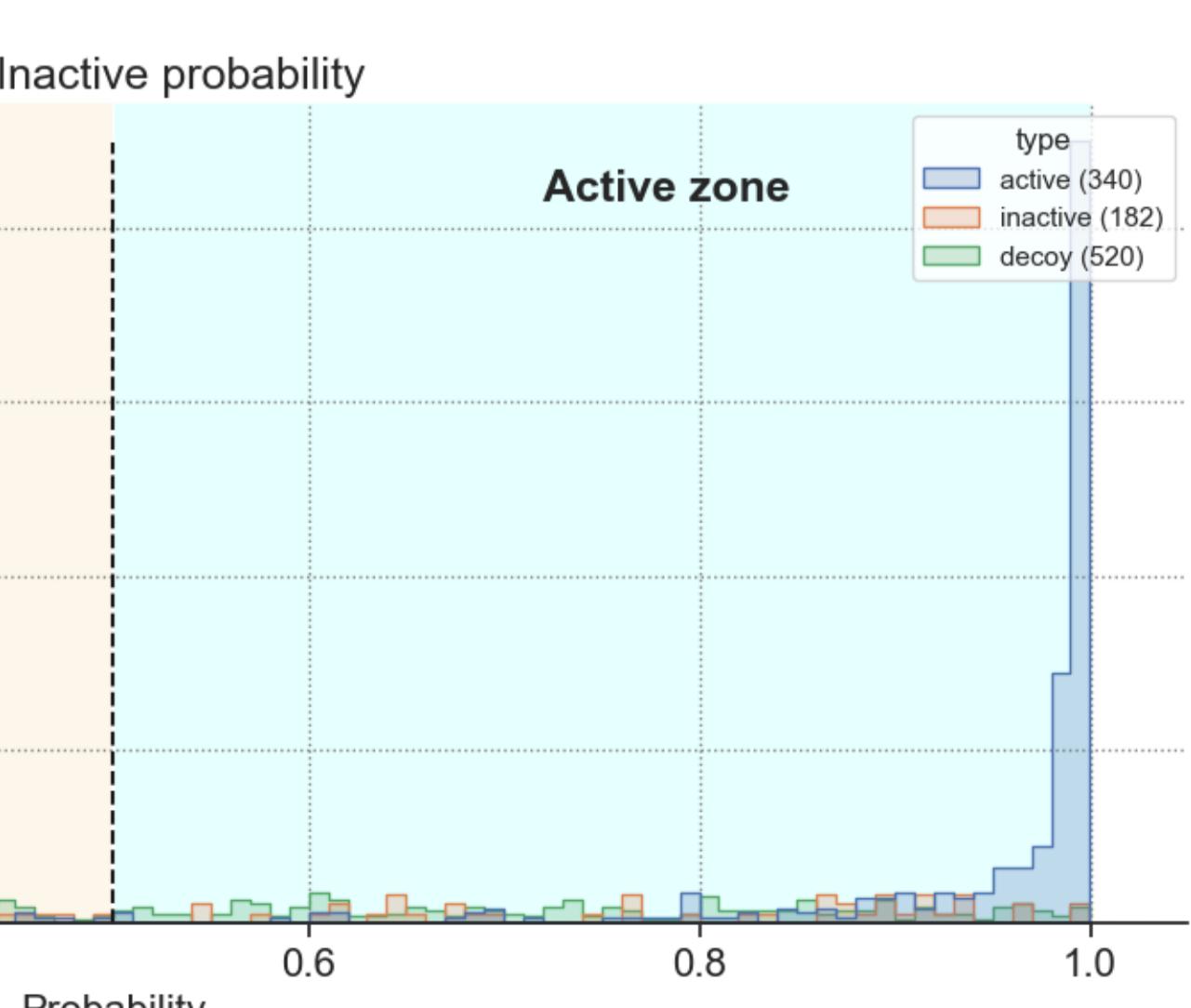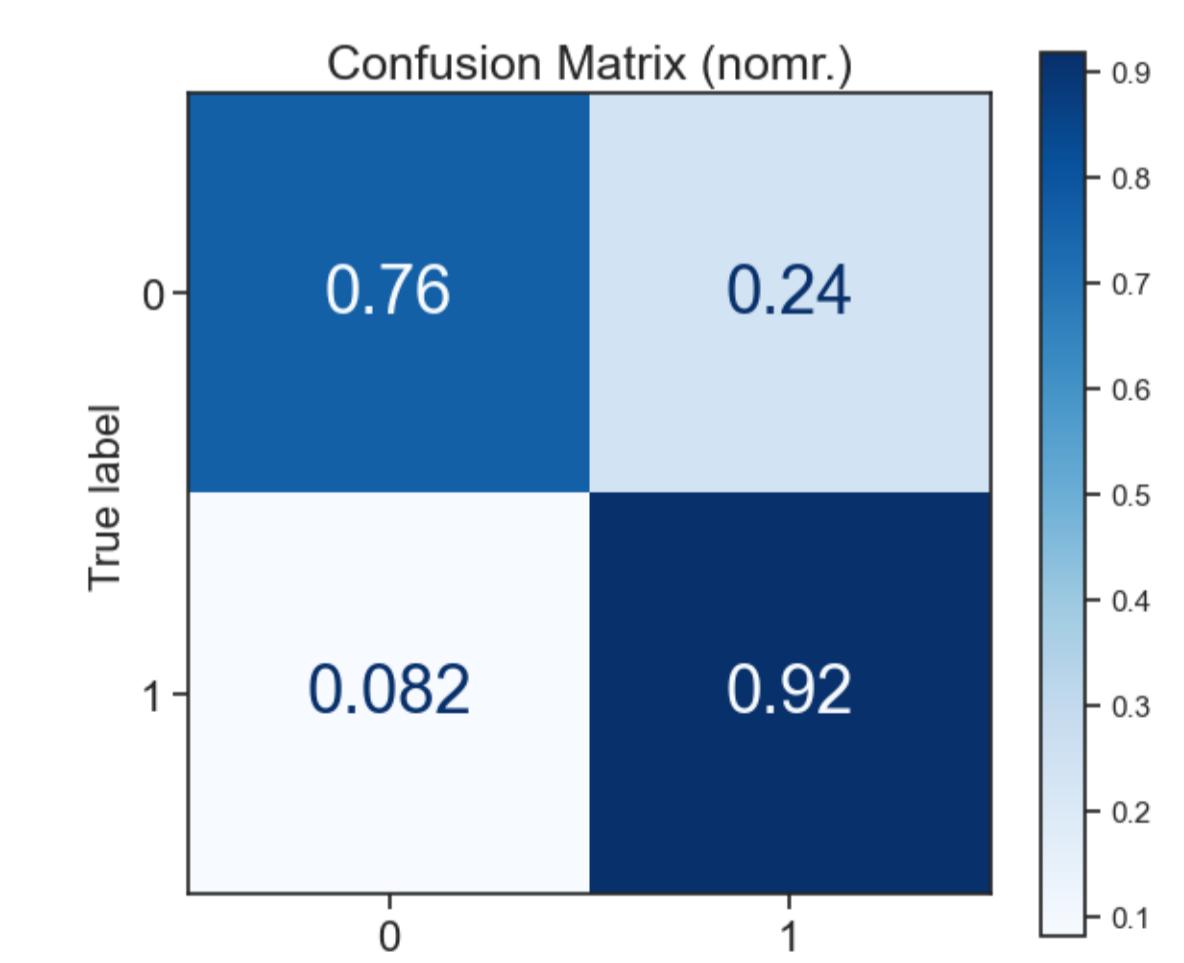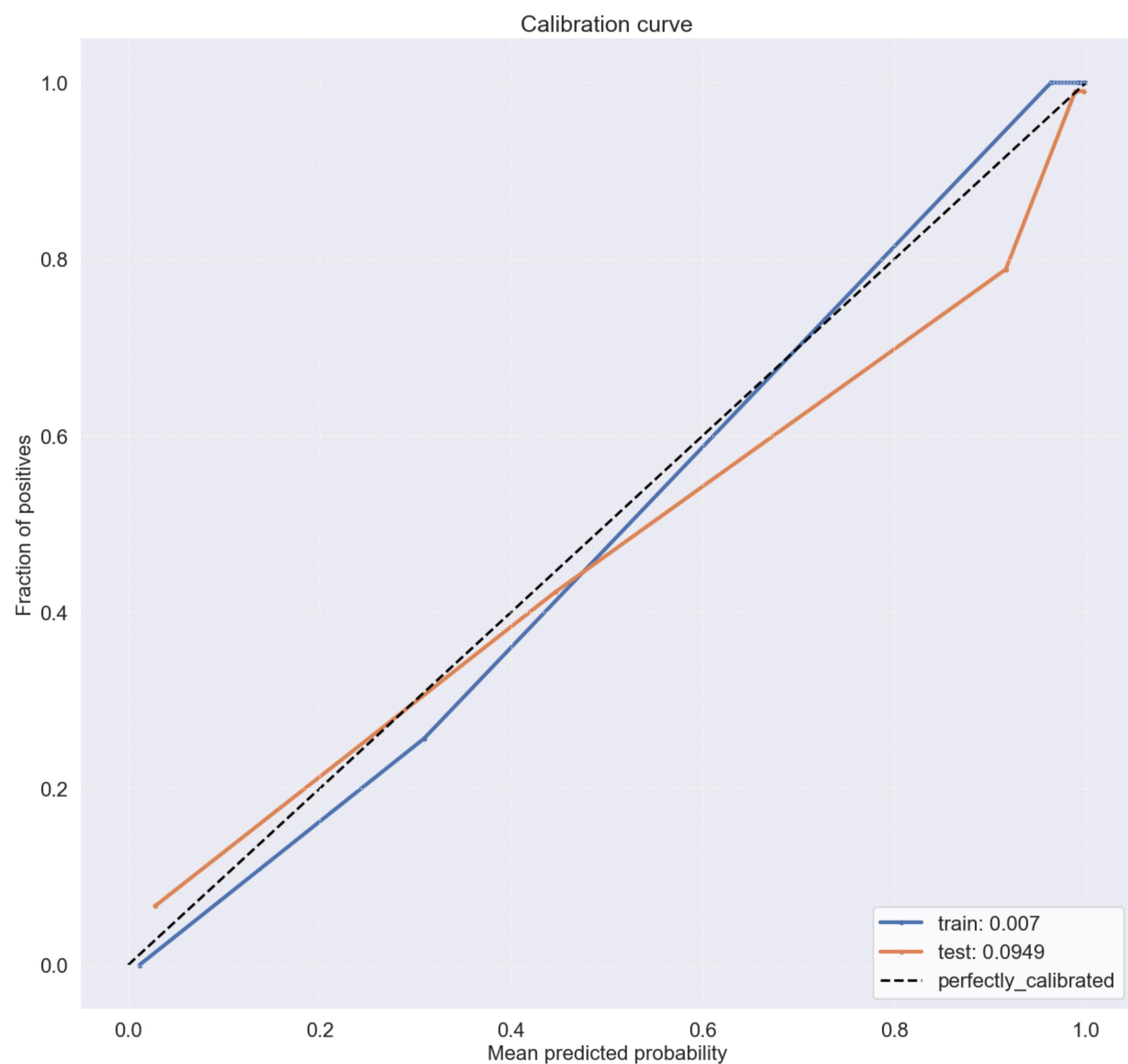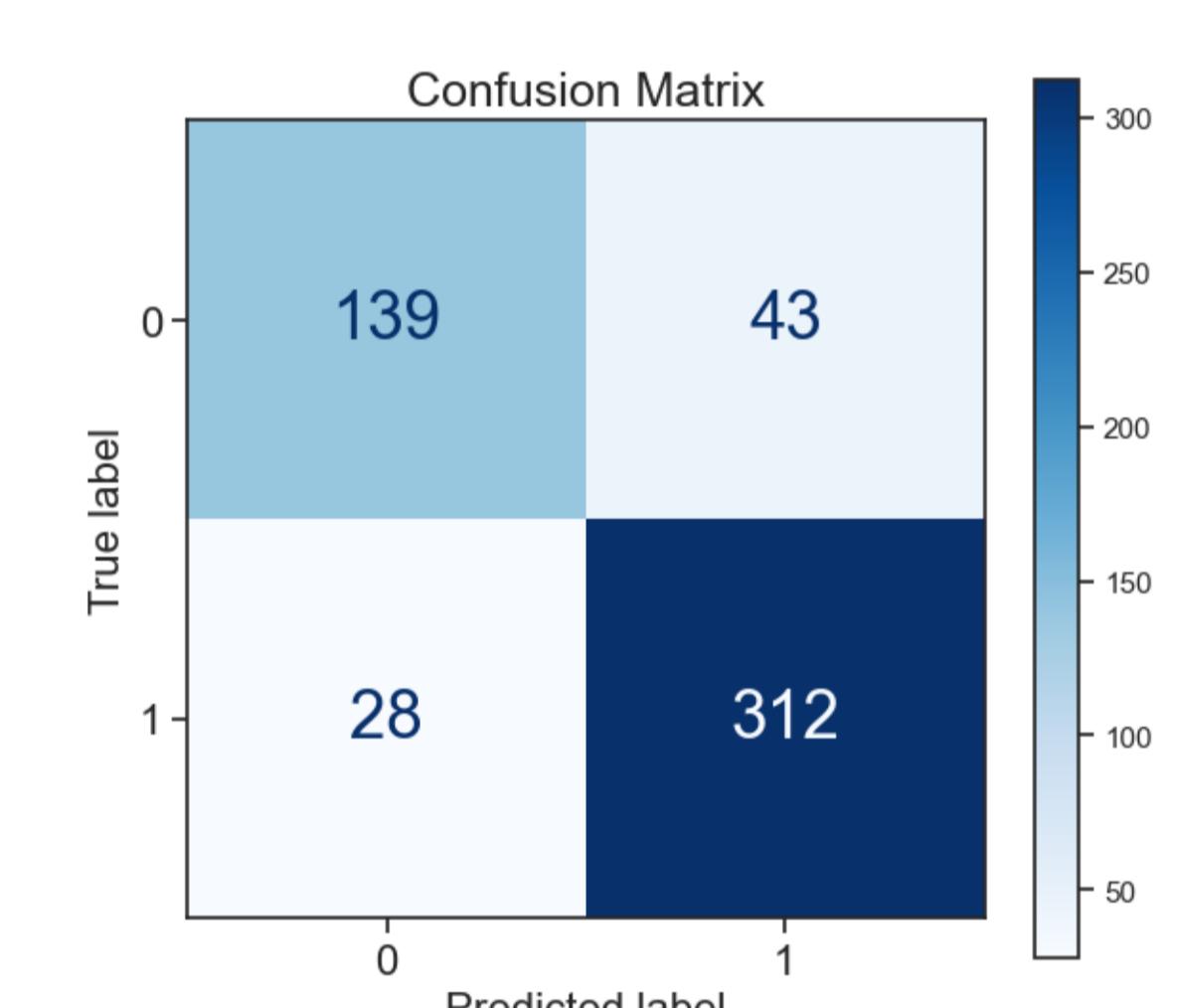
Metric	Train	Test
AUC	0.9999	0.9434
accuracy	0.9957	0.864
balanced accuracy	0.9945	0.8407
recall_(sens)	0.9985	0.9176
specificity	0.9904	0.7637
precision	0.9949	0.8789
f1_score	0.9957	0.8625

Protein: P27338 - metric: f1

Total (train): 2088

Active: 728
Inactive: 1360

ROC curve

- AUC_train = 1.000
- AUC_test = 0.943
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2601)
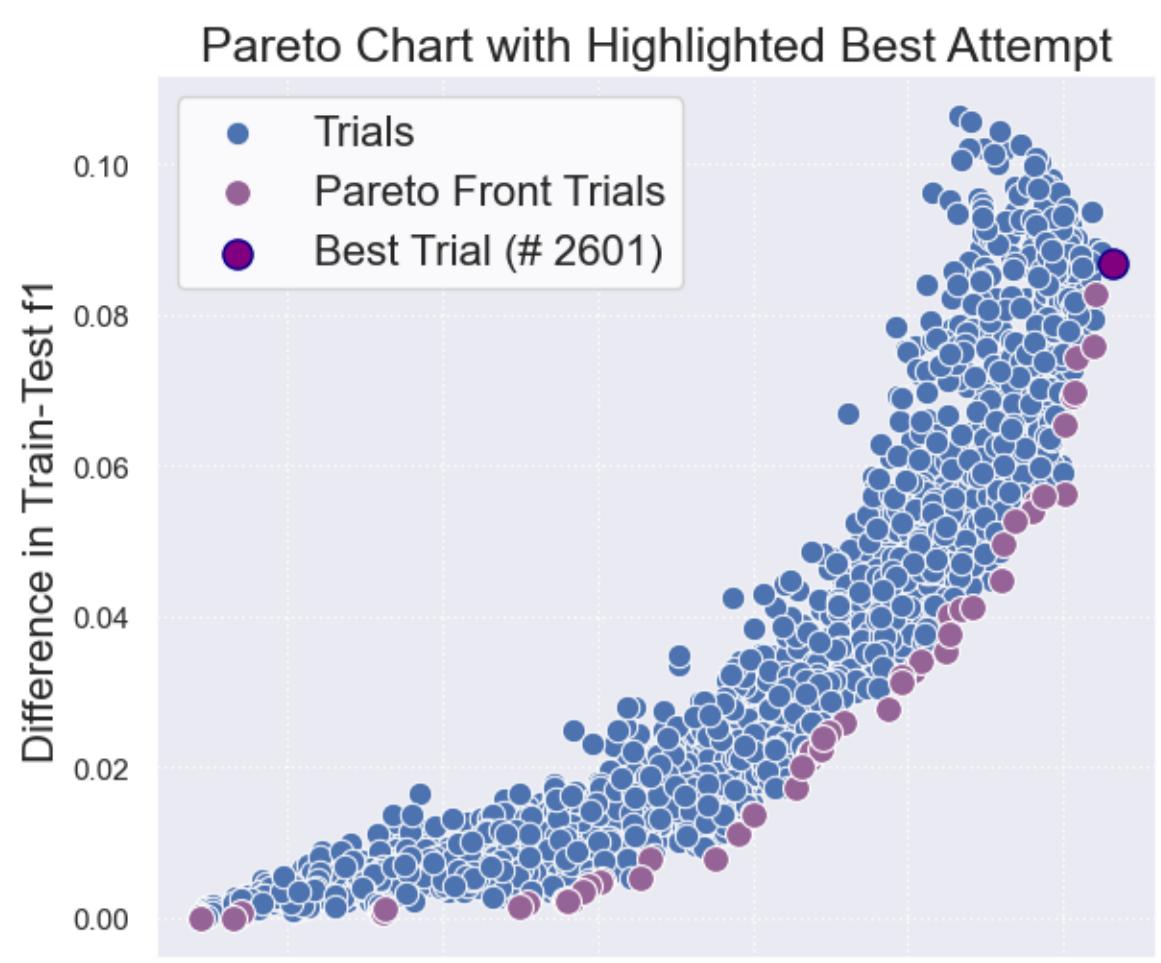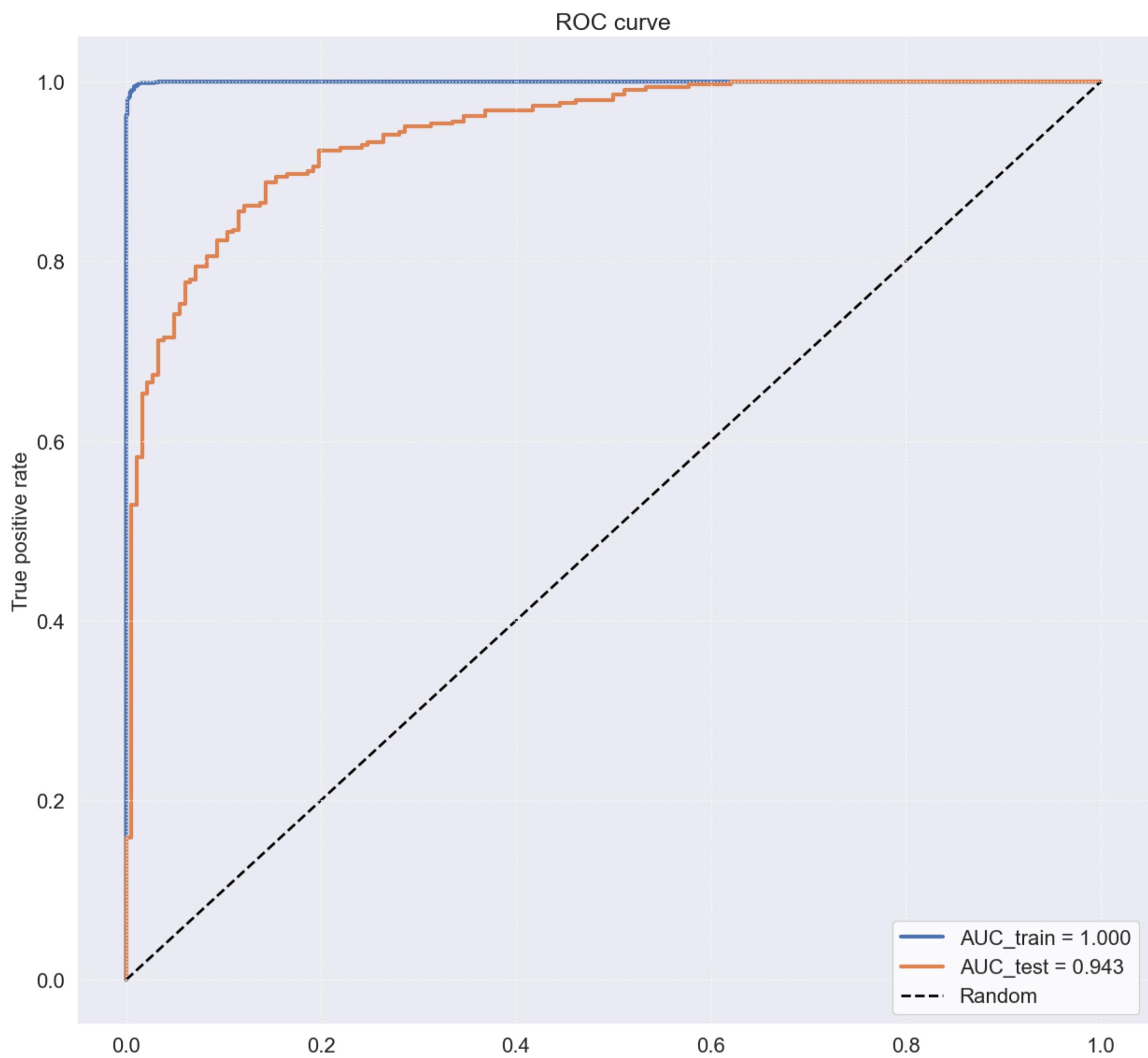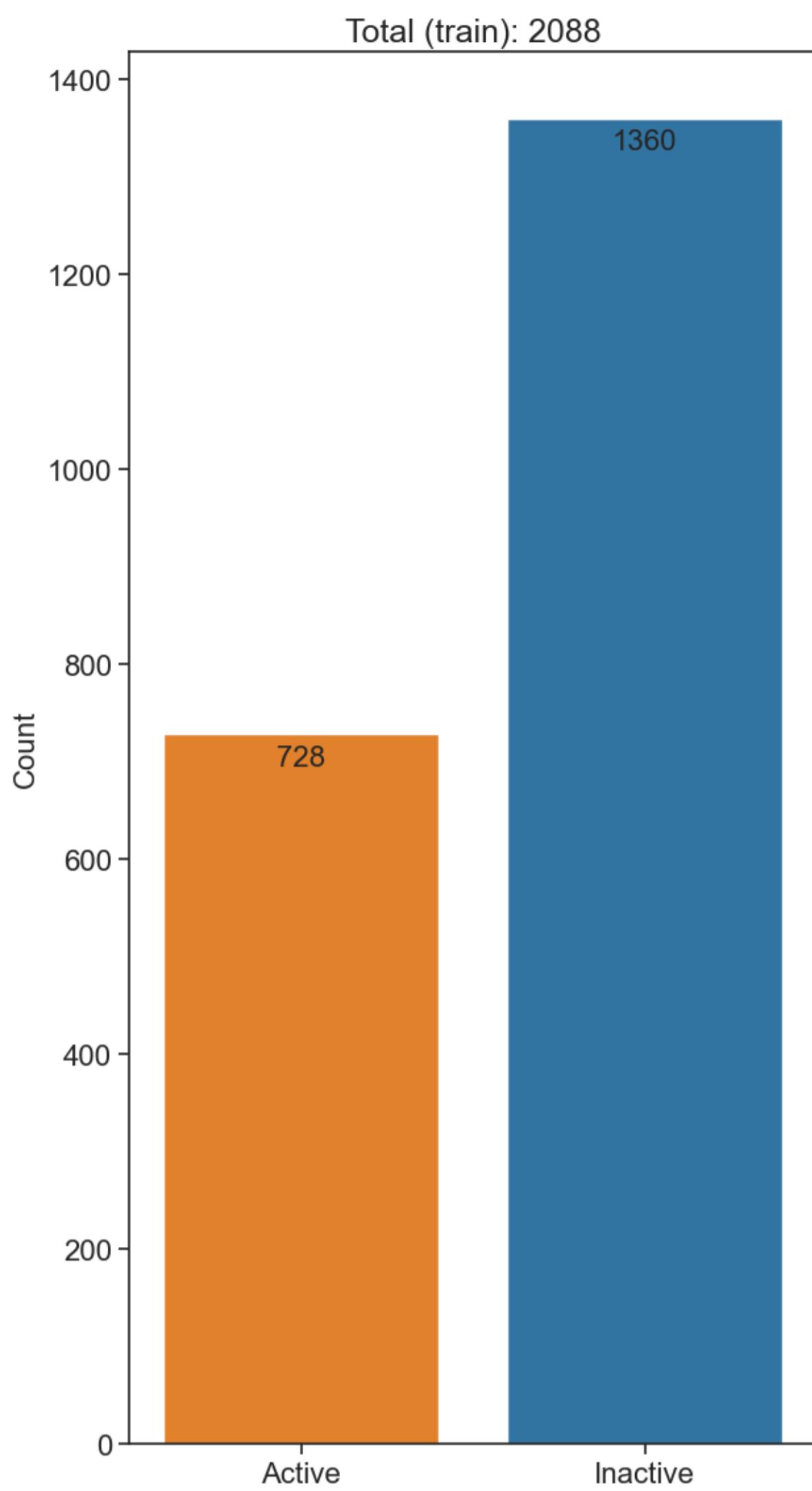
XGBoost Parameters	Value
alpha	1.5326
colsample_bytree	0.6602
gamma	0.3019
lambda	3.1133
learning_rate	0.0318
max_depth	10.0
min_child_weight	1.0
n_estimators	823.0
reg_alpha	0.1337
reg_lambda	0.8143
scale_pos_weight	1.5198
subsample	0.7908

Confusion Matrix

	0	1
0	136	46
1	23	317

Confusion Matrix (nomr.)

	0	1
0	0.75	0.25
1	0.068	0.93

Calibration curve

- train: 0.0126
- test: 0.0932
- perfectly_calibrated

Active/Inactive probability

Inactive zone | Active zone

type:
- active (340)
- inactive (182)
- decoy (520)
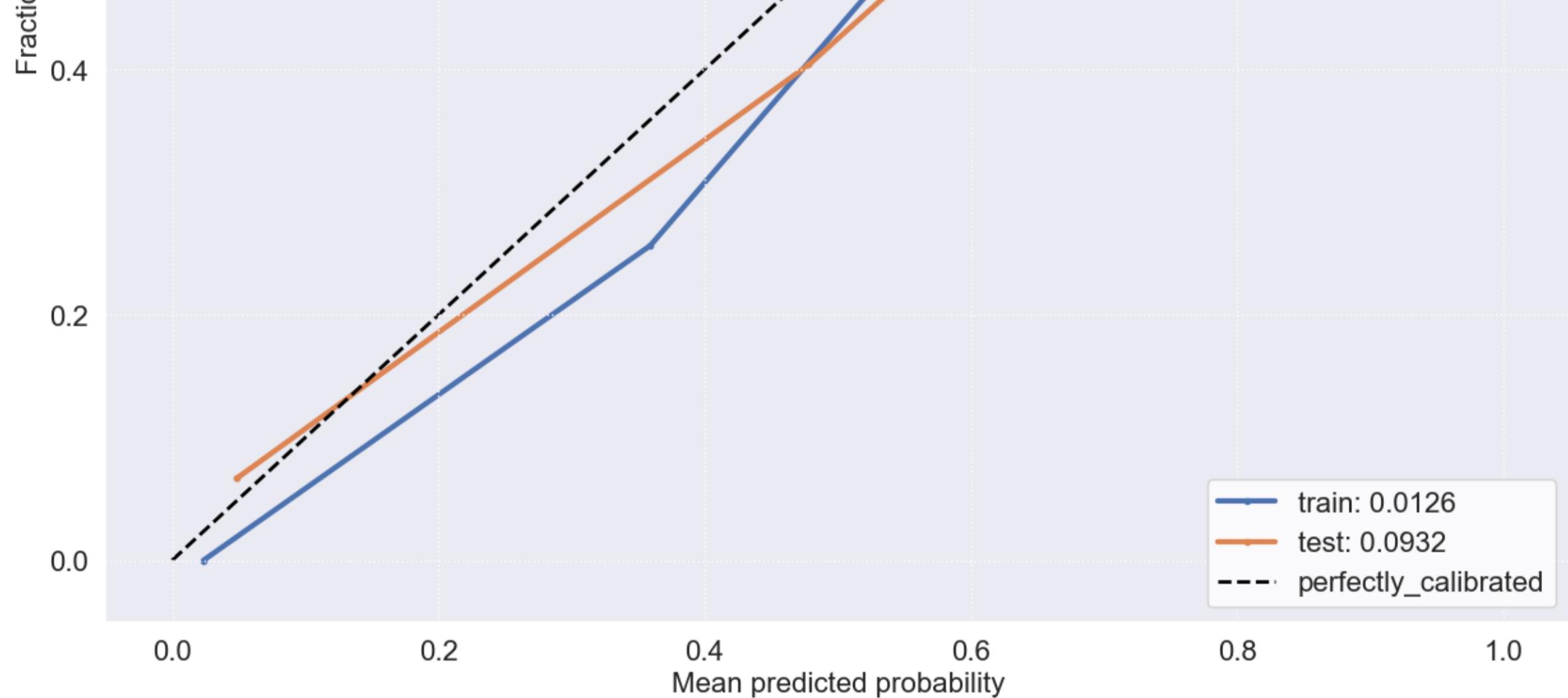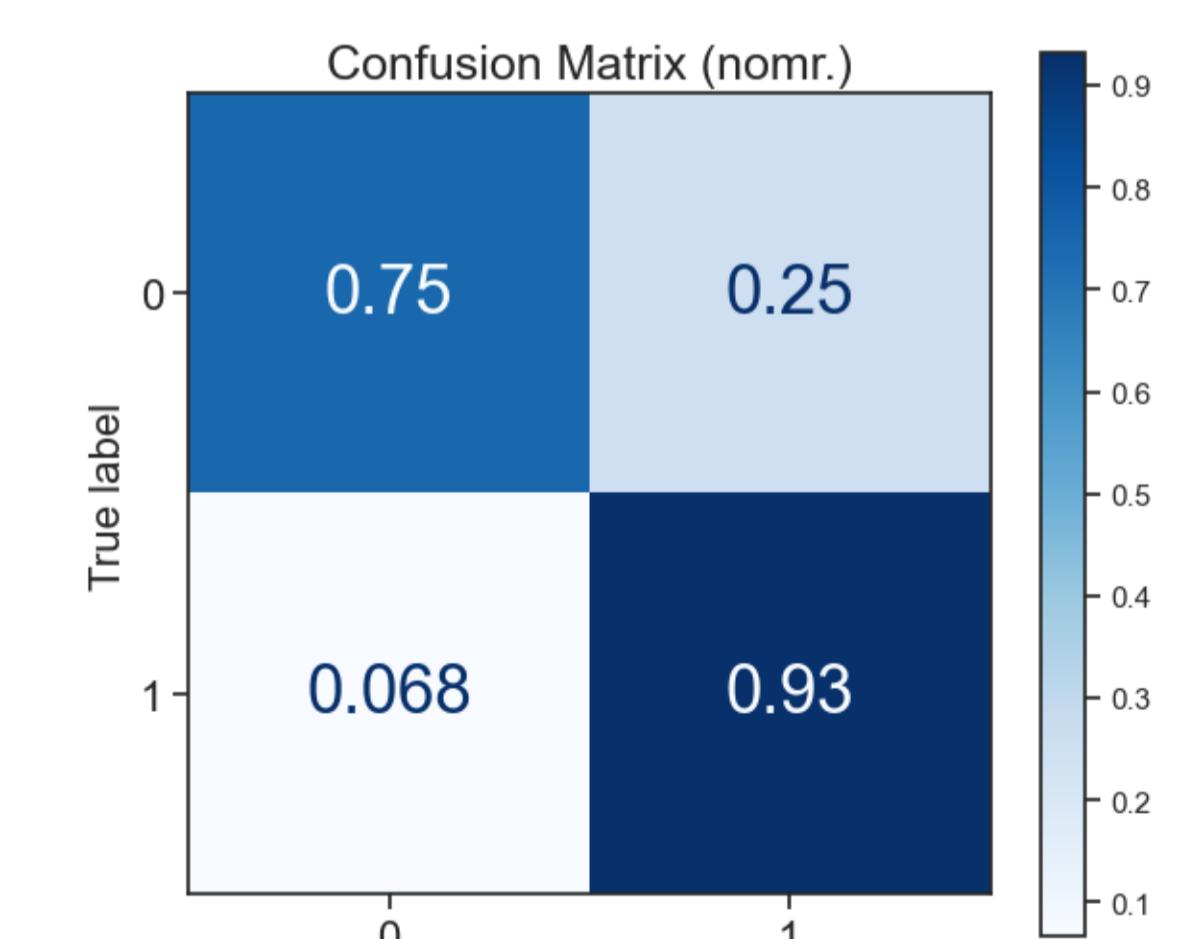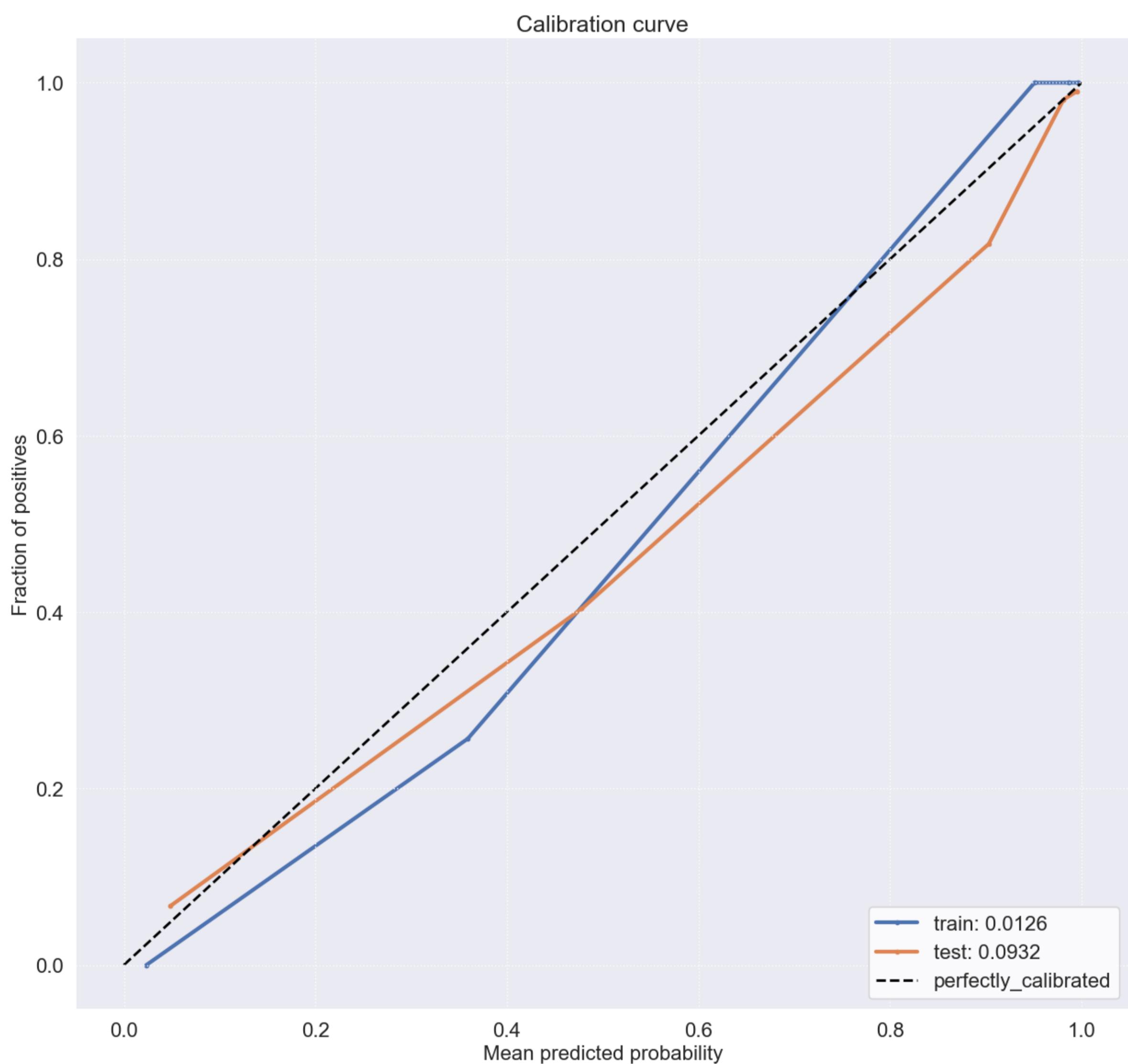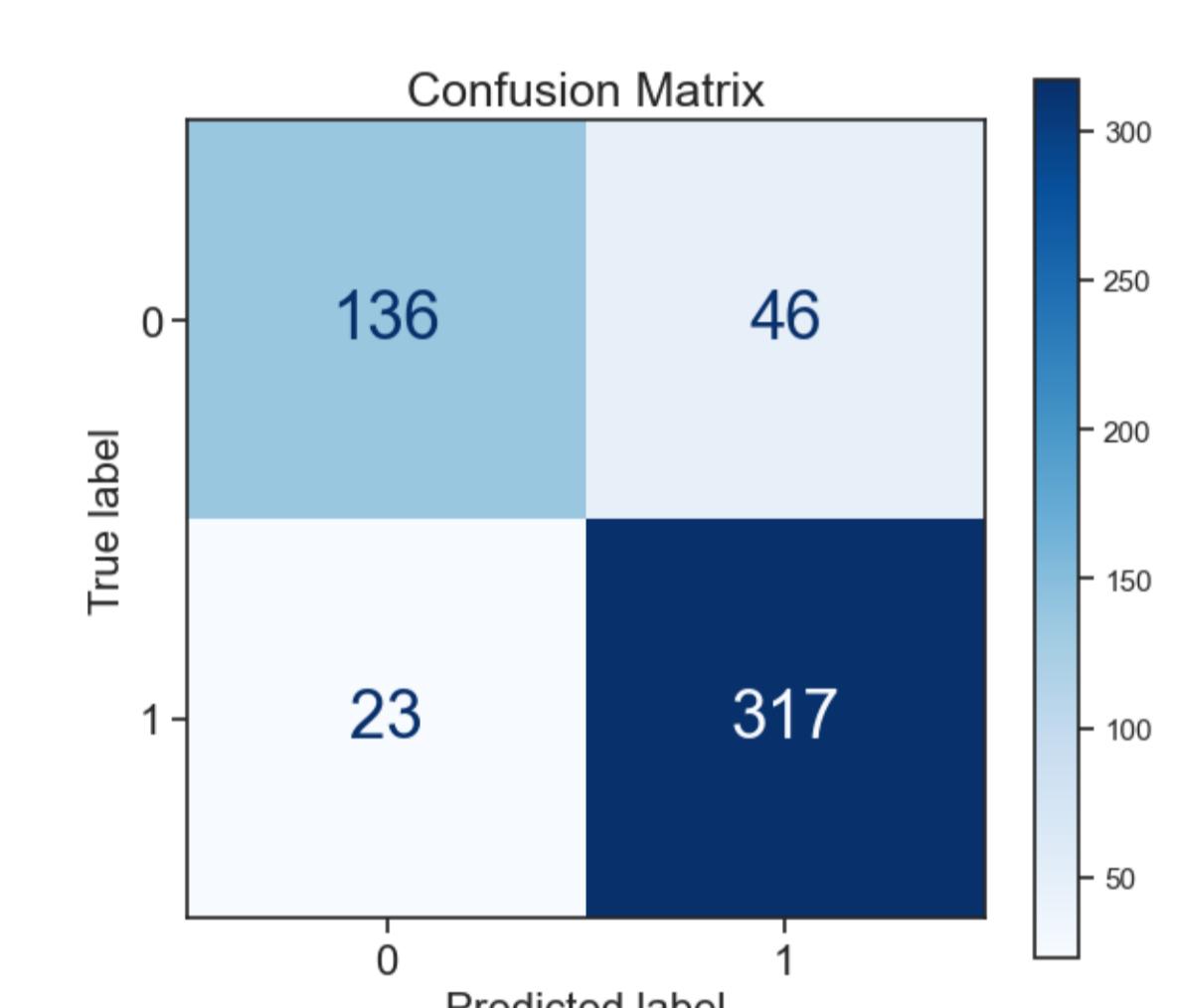
Metric	Train	Test
AUC	0.9998	0.9433
accuracy	0.9909	0.8678
balanced accuracy	0.9876	0.8398
recall_(sens)	0.9985	0.9324
specificity	0.9766	0.7473
precision	0.9876	0.8733
f1_score	0.9909	0.8655

Protein: P27338 - metric: precision

Total (train): 2088

Bar chart showing: Active = 728, Inactive = 1360

ROC curve

- AUC_train = 1.000
- AUC_test = 0.950
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2589)
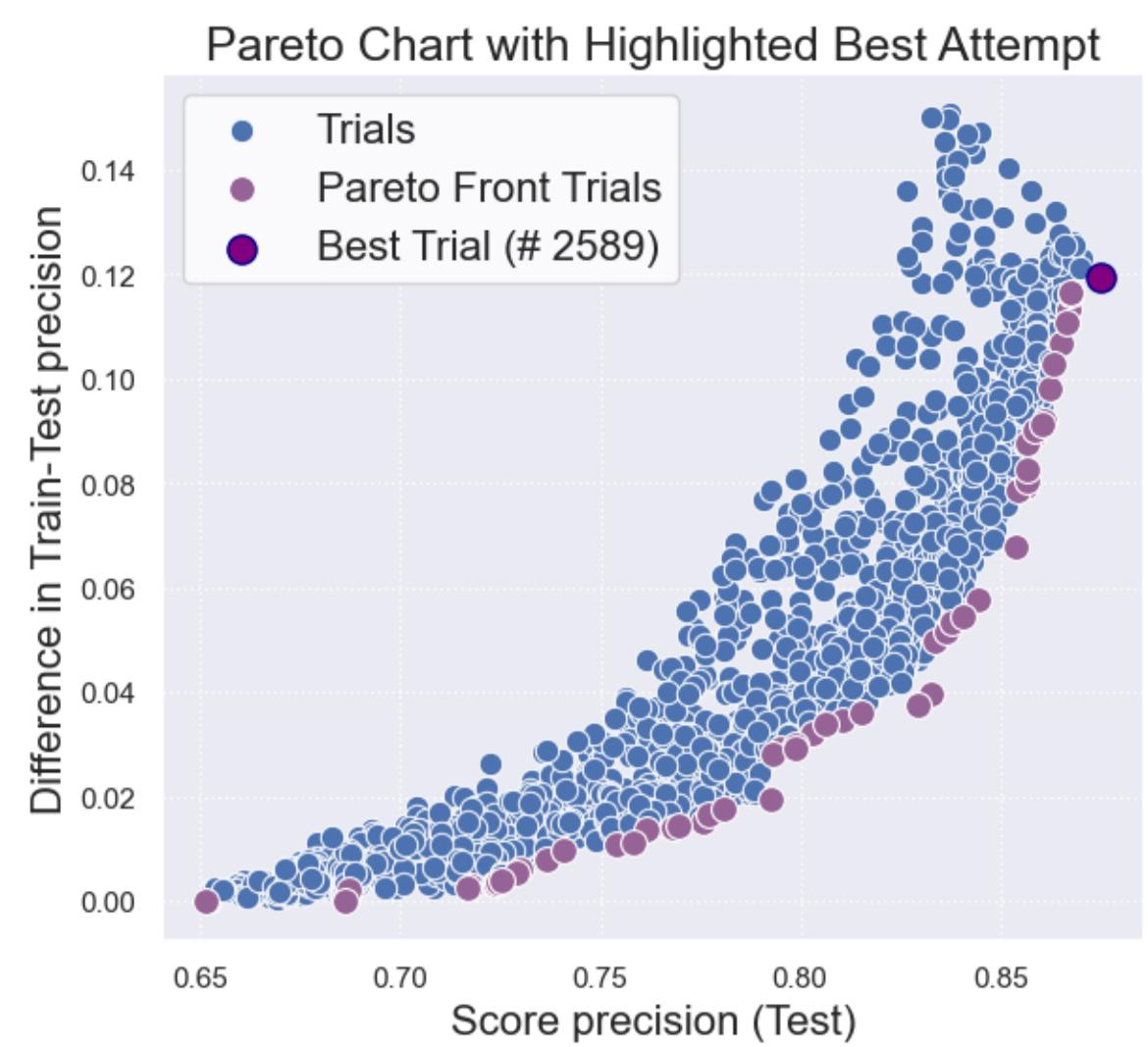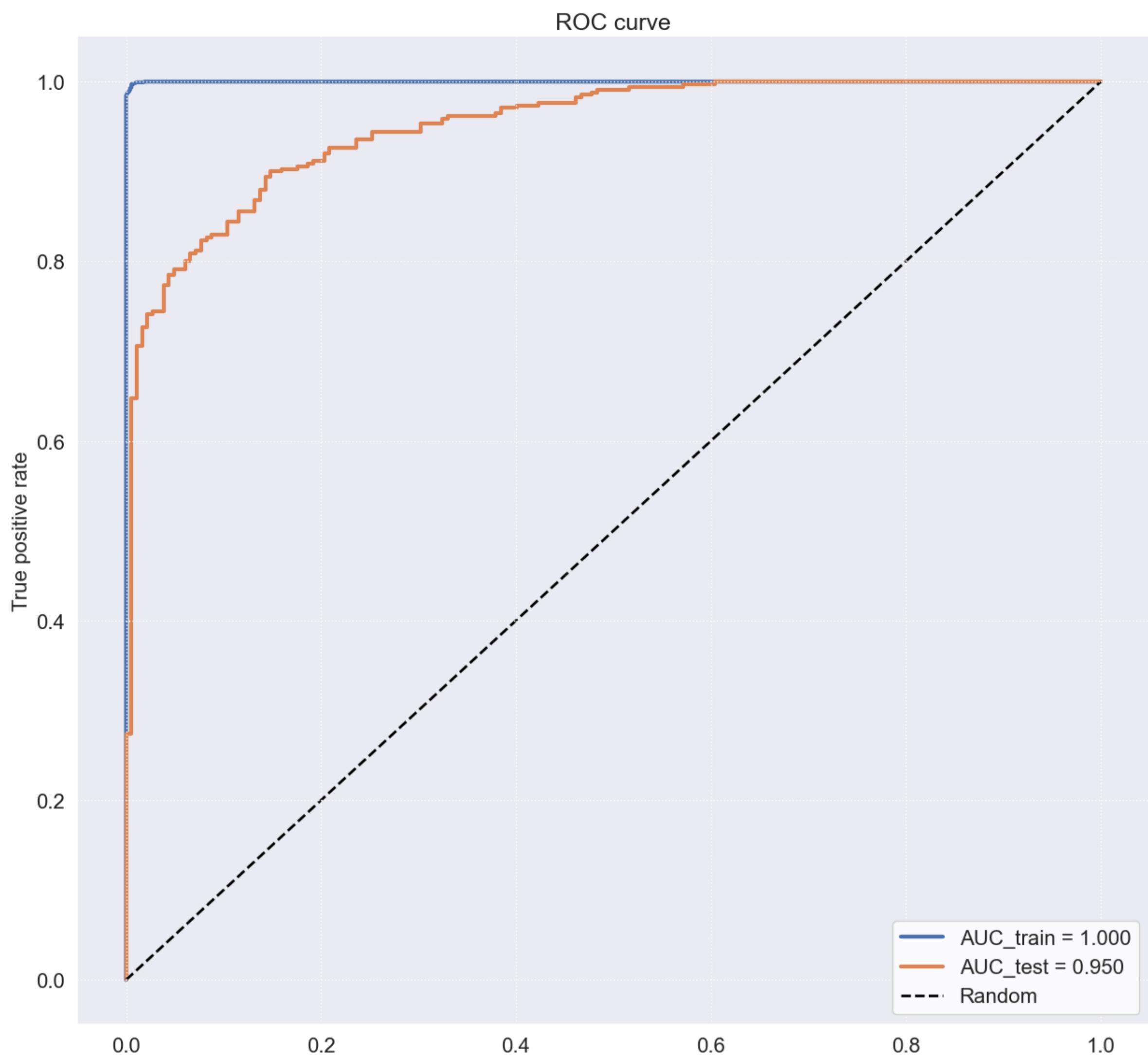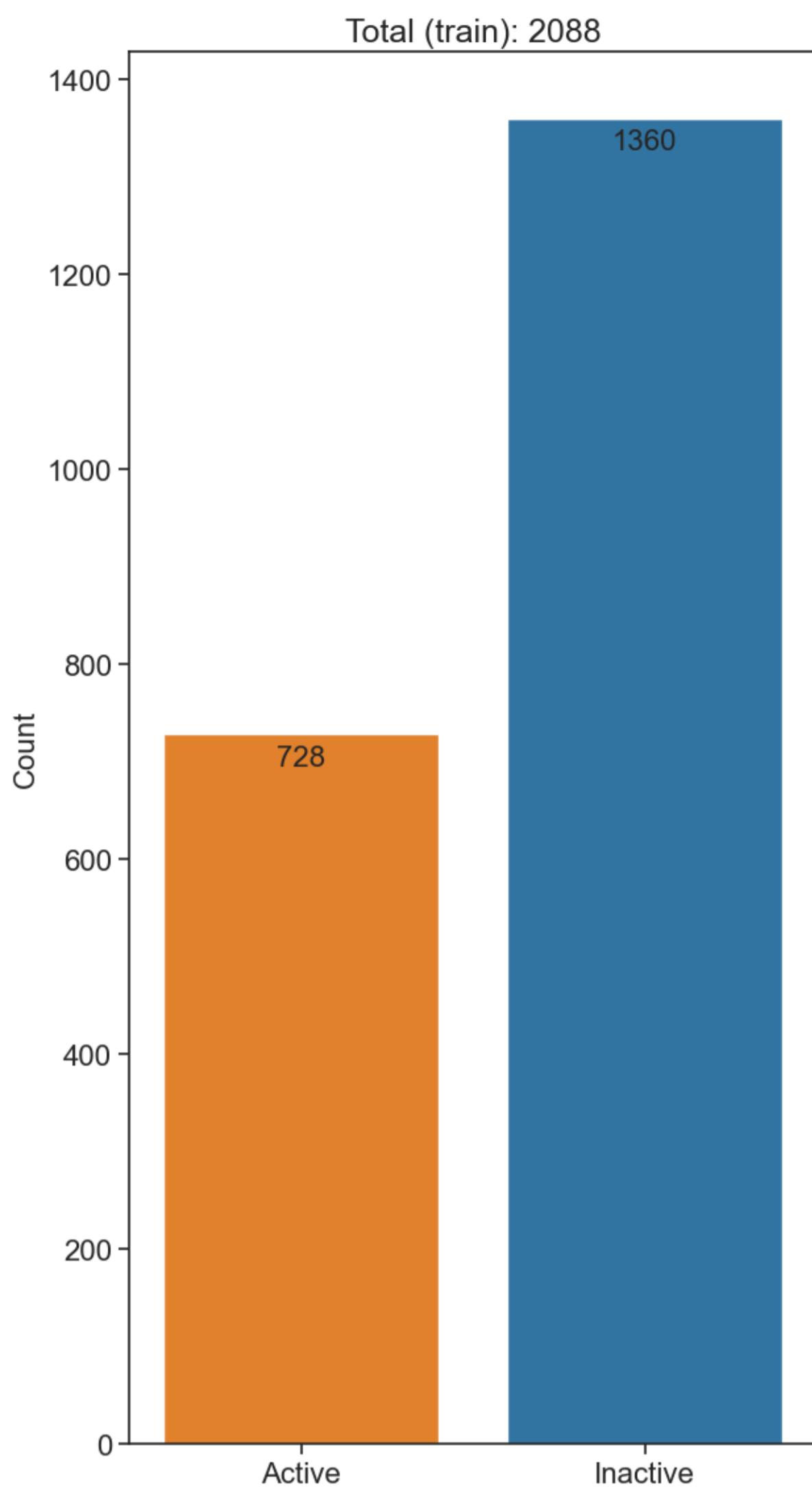
XGBoost Parameters	Value
alpha	4.5009
colsample_bytree	0.7063
gamma	0.1131
lambda	0.0241
learning_rate	0.0639
max_depth	8.0
min_child_weight	1.0
n_estimators	675.0
reg_alpha	0.1111
reg_lambda	0.6137
scale_pos_weight	1.4019
subsample	0.944

Confusion Matrix

	Predicted 0	Predicted 1
True 0	140	42
True 1	25	315

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.77	0.23
True 1	0.074	0.93

Calibration curve

- train: 0.0088
- test: 0.0902
- perfectly_calibrated

Active/Inactive probability

- Inactive zone / Active zone
- type
  - active (340)
  - inactive (182)
  - decoy (520)
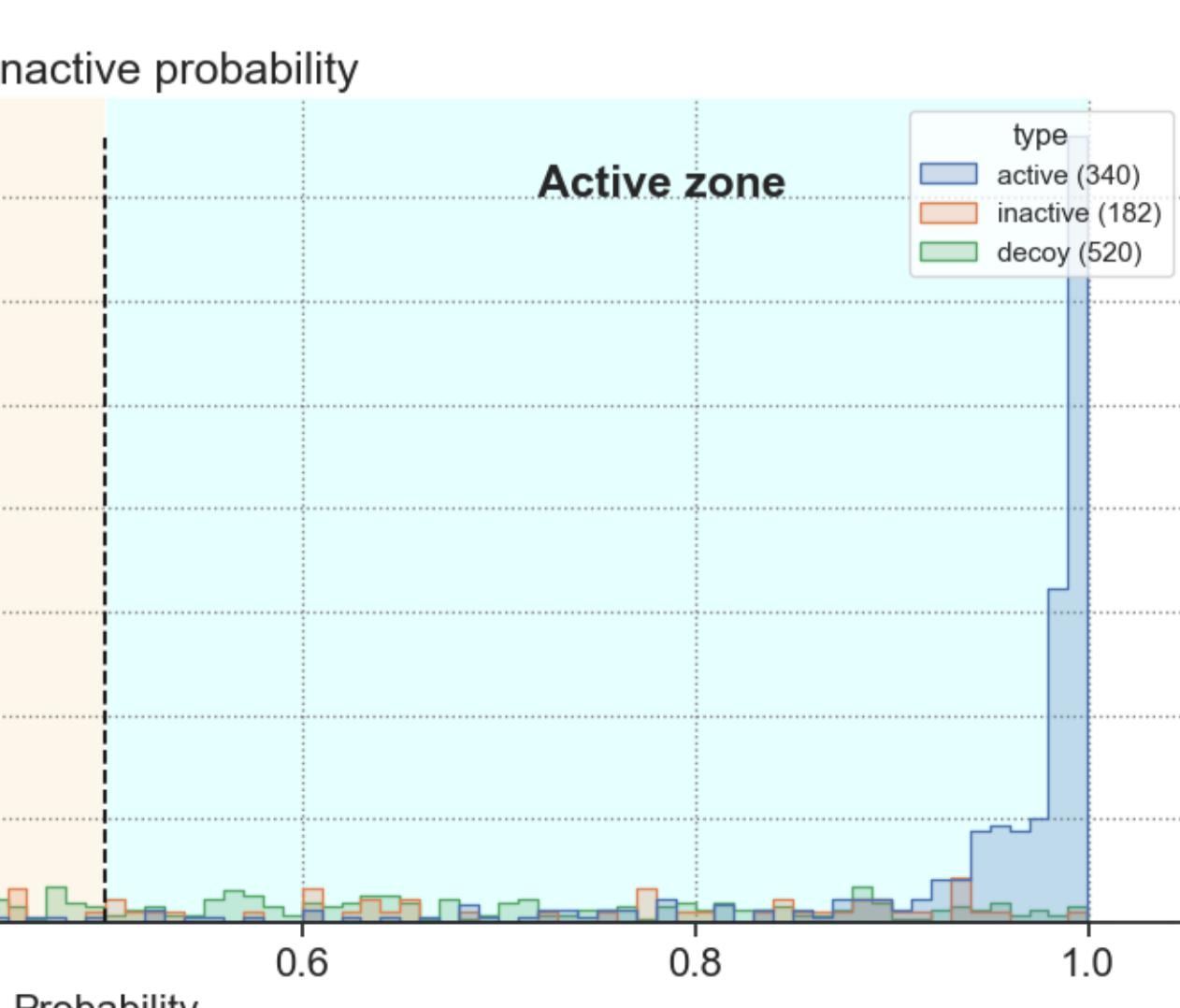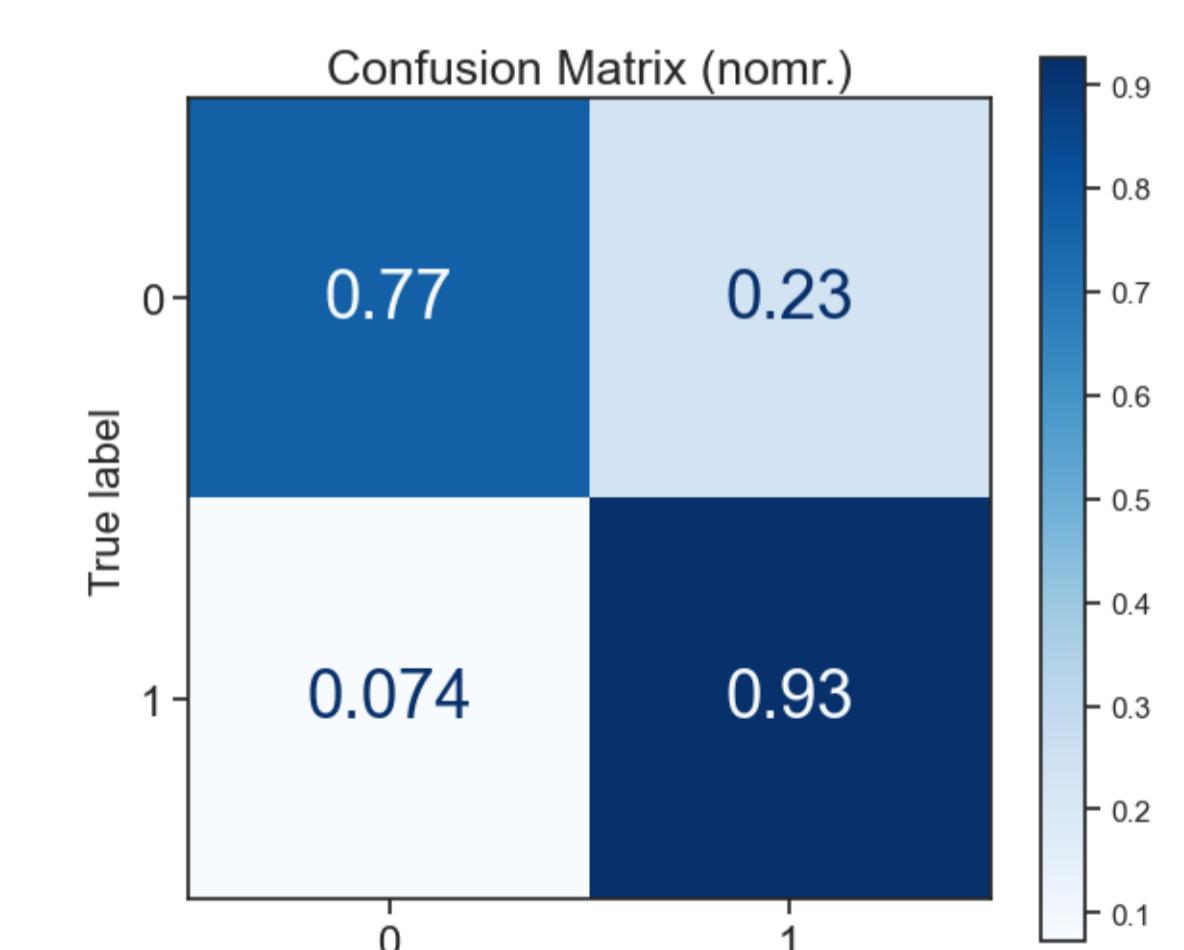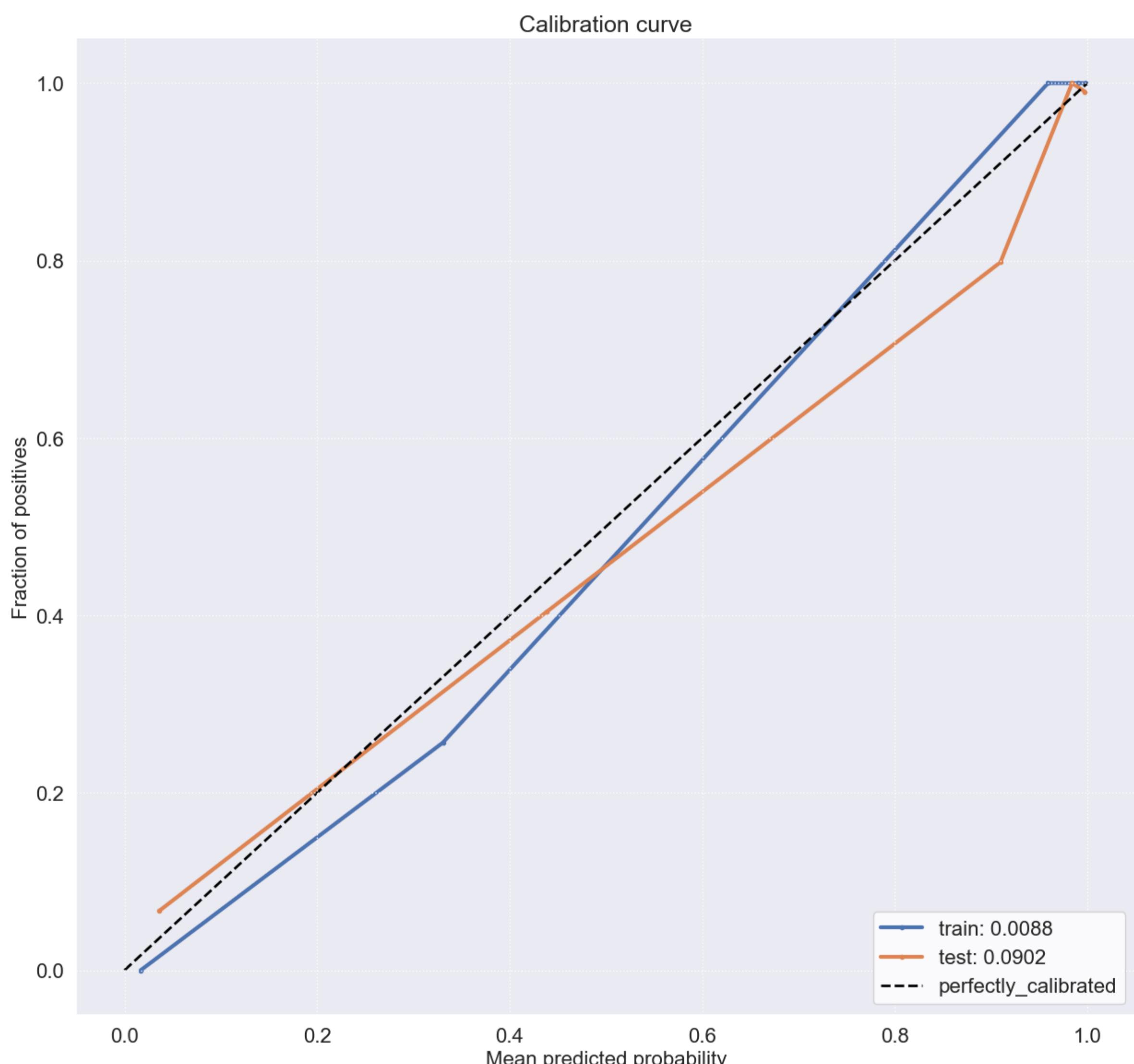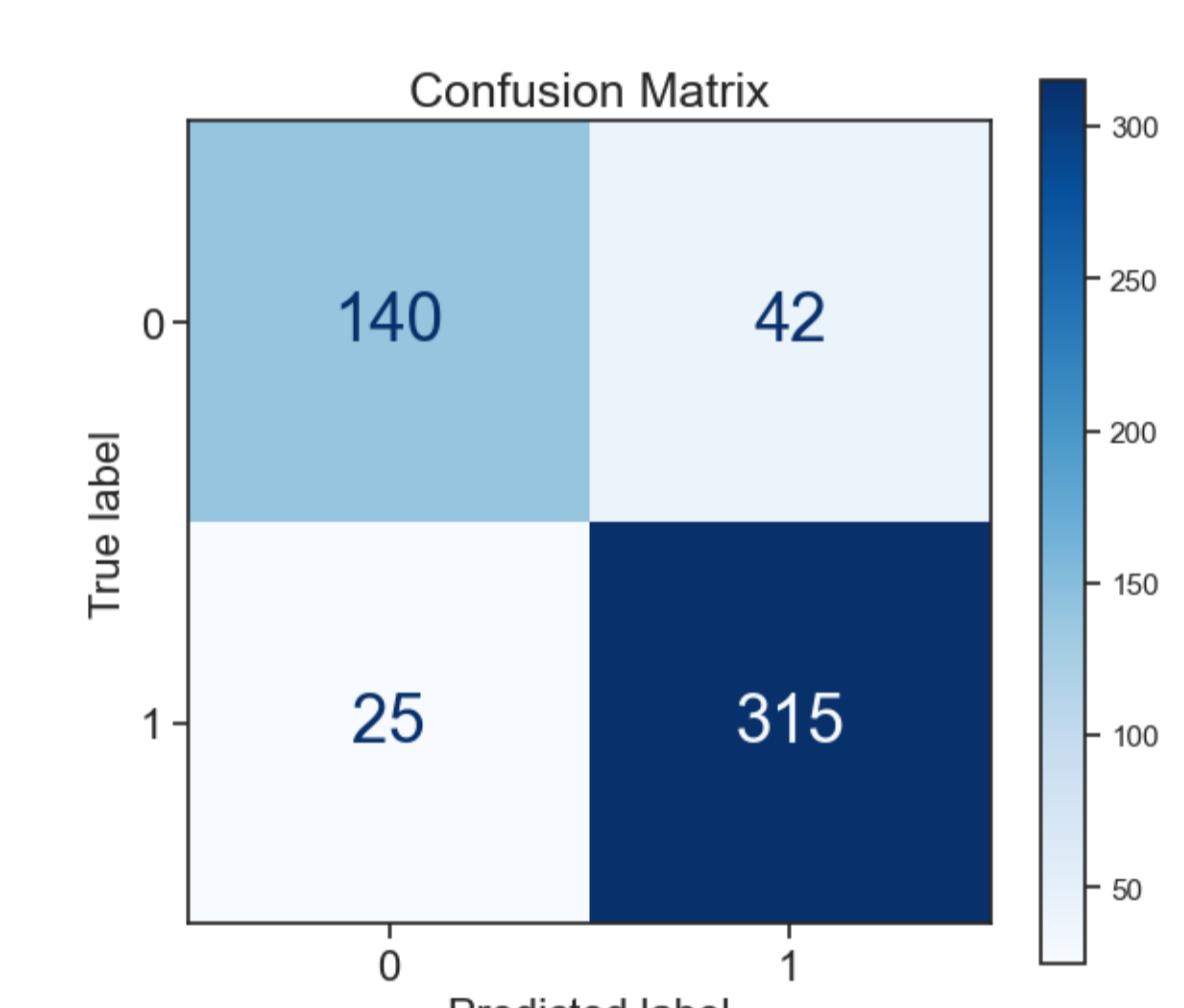
Metric	Train	Test
AUC	0.9999	0.9502
accuracy	0.9943	0.8716
balanced accuracy	0.9921	0.8479
recall_(sens)	0.9993	0.9265
specificity	0.9849	0.7692
precision	0.992	0.8824
f1_score	0.9942	0.8701

Protein: P27338 - metric: SoftVotingXGBoostEnsemble
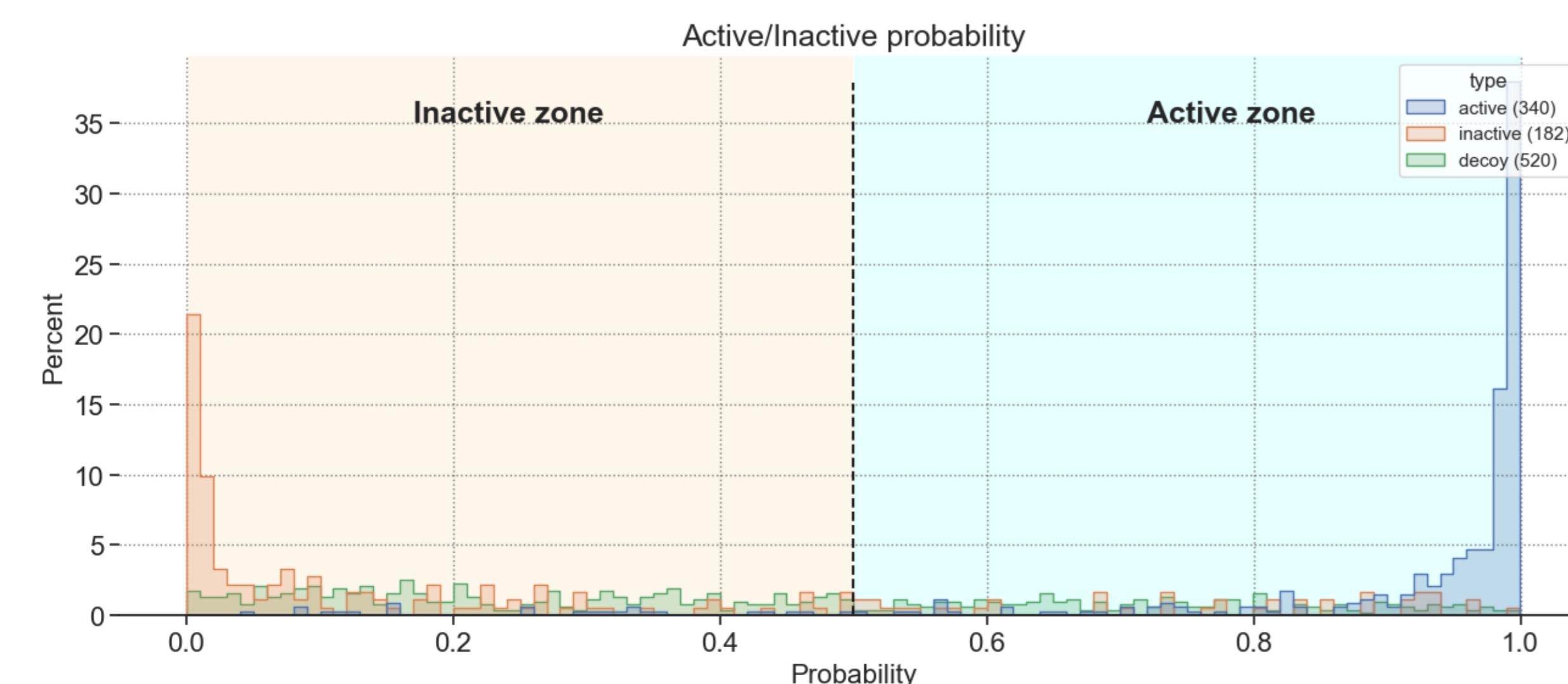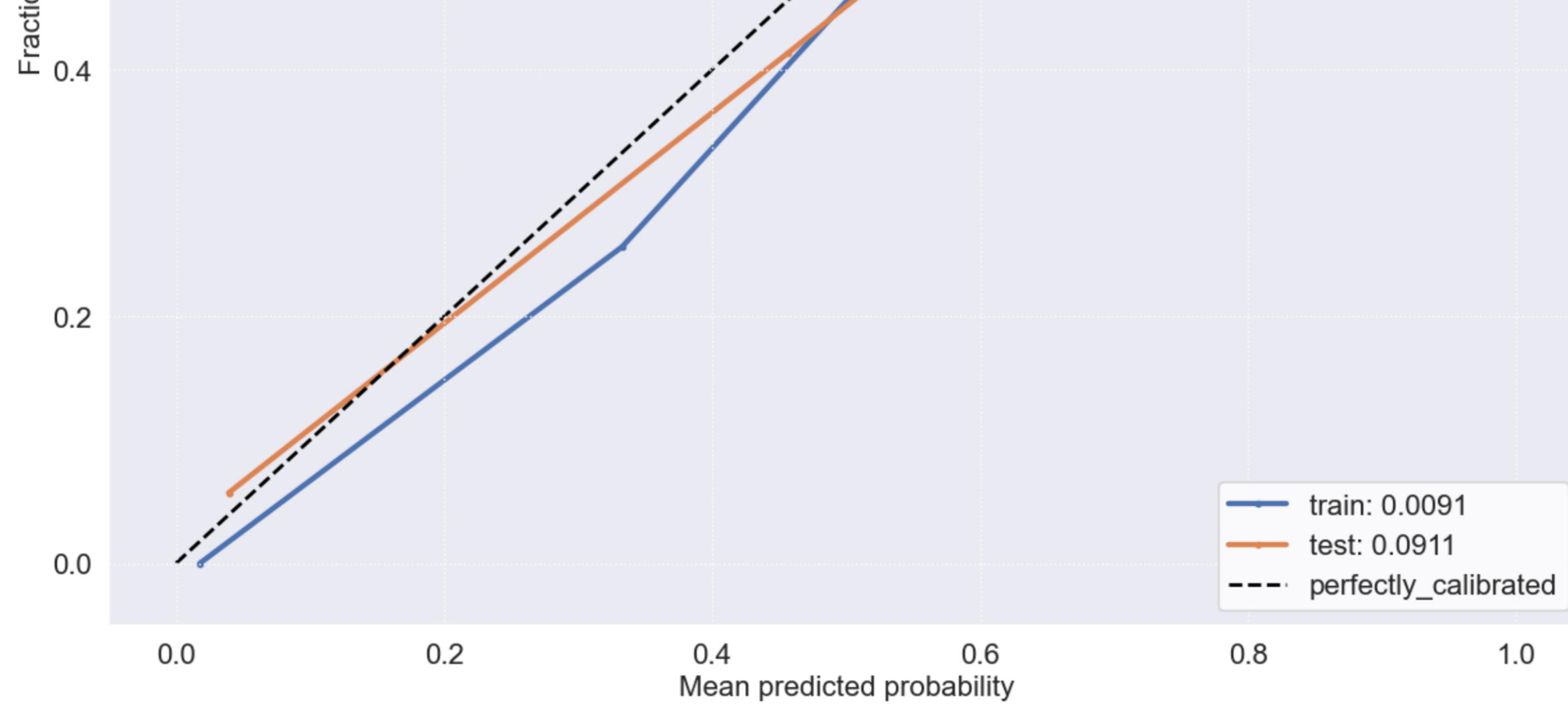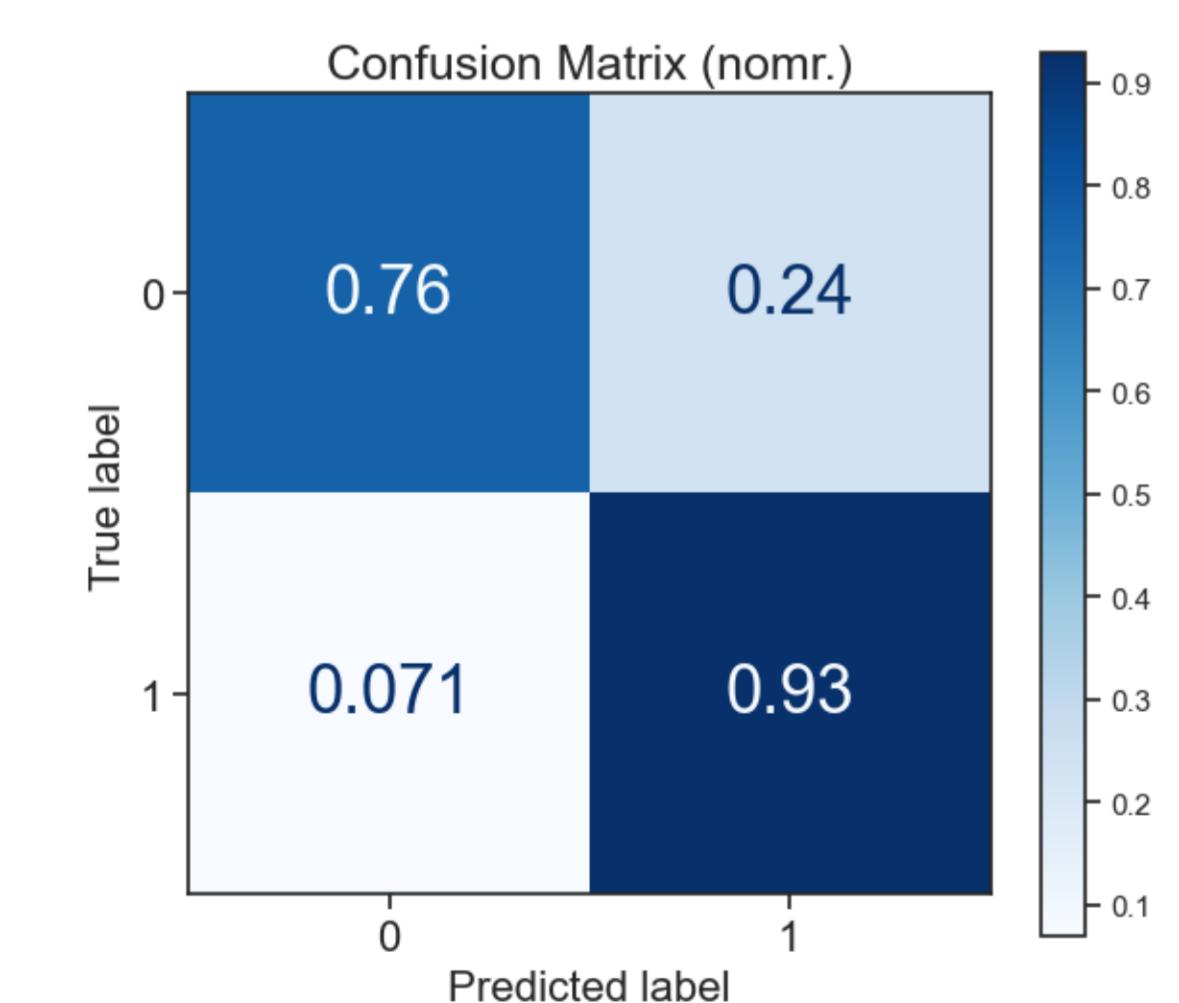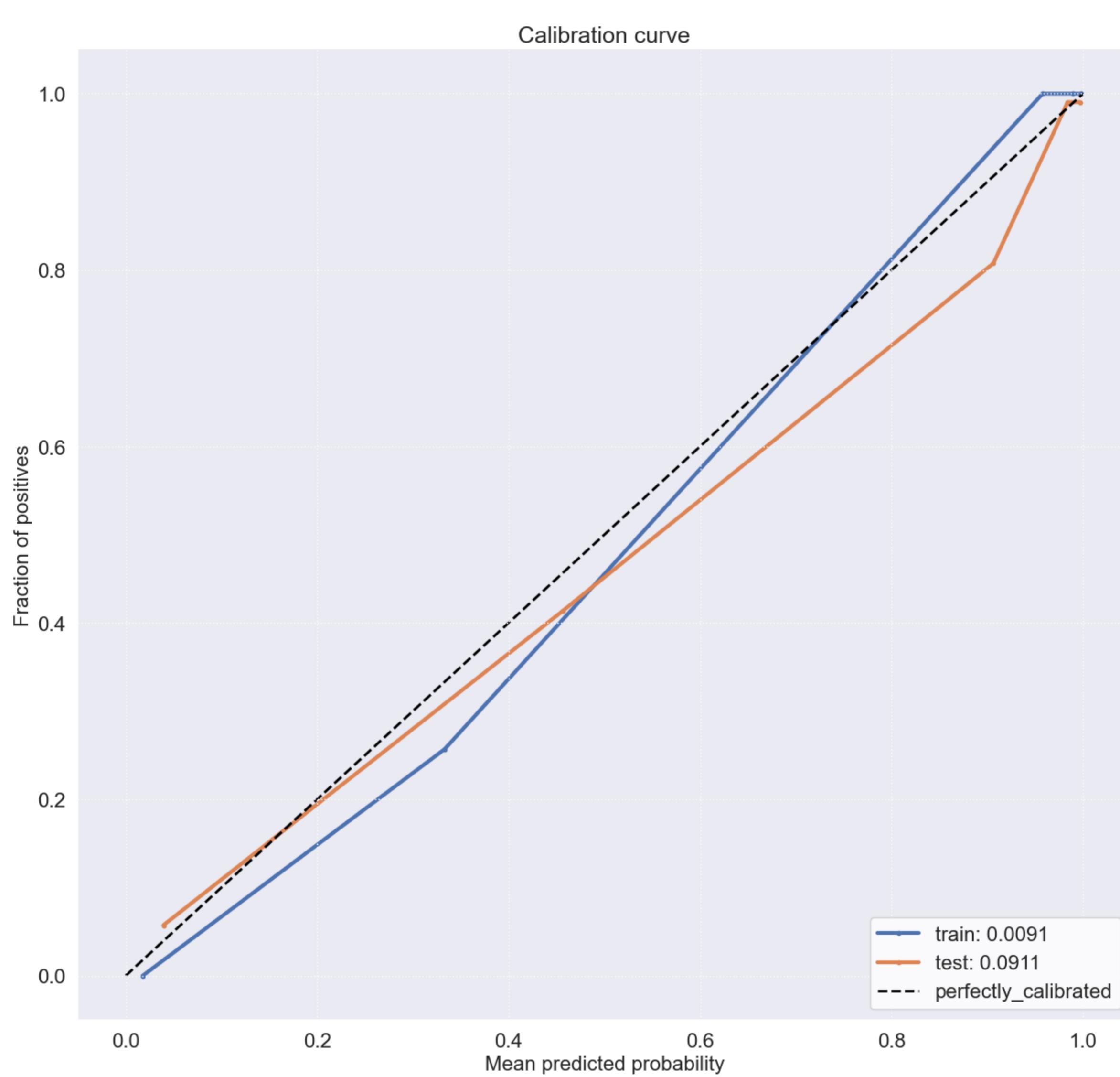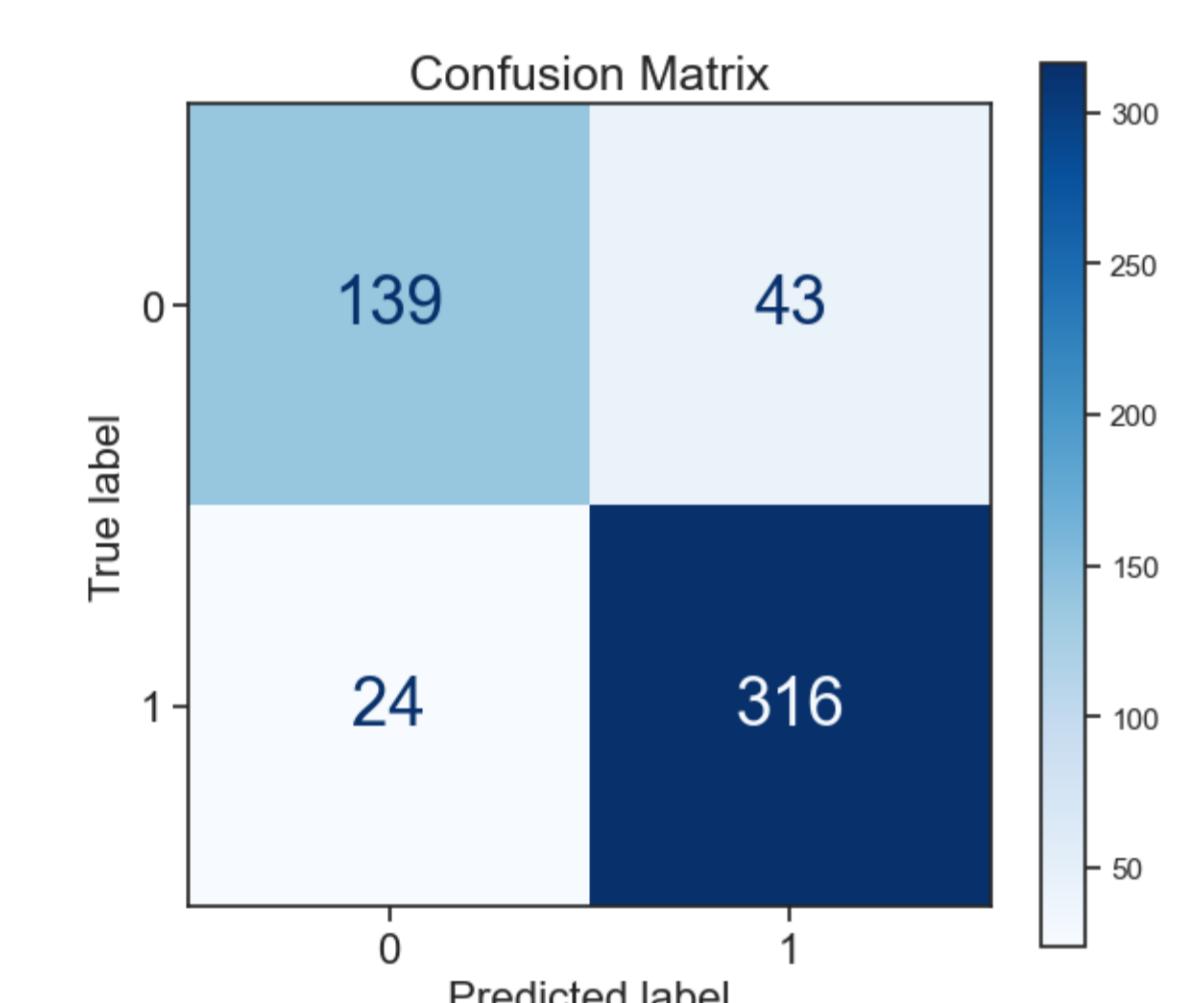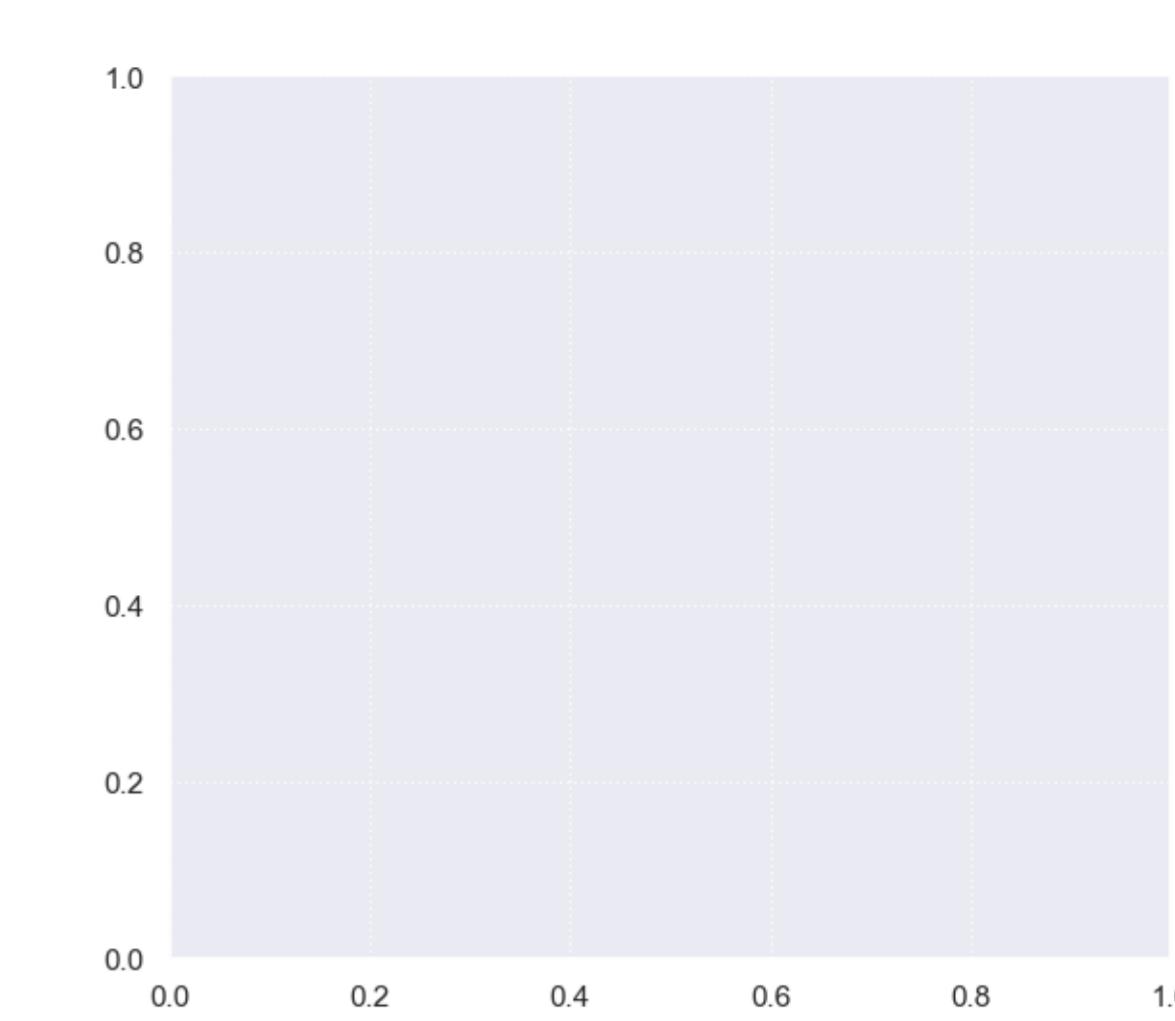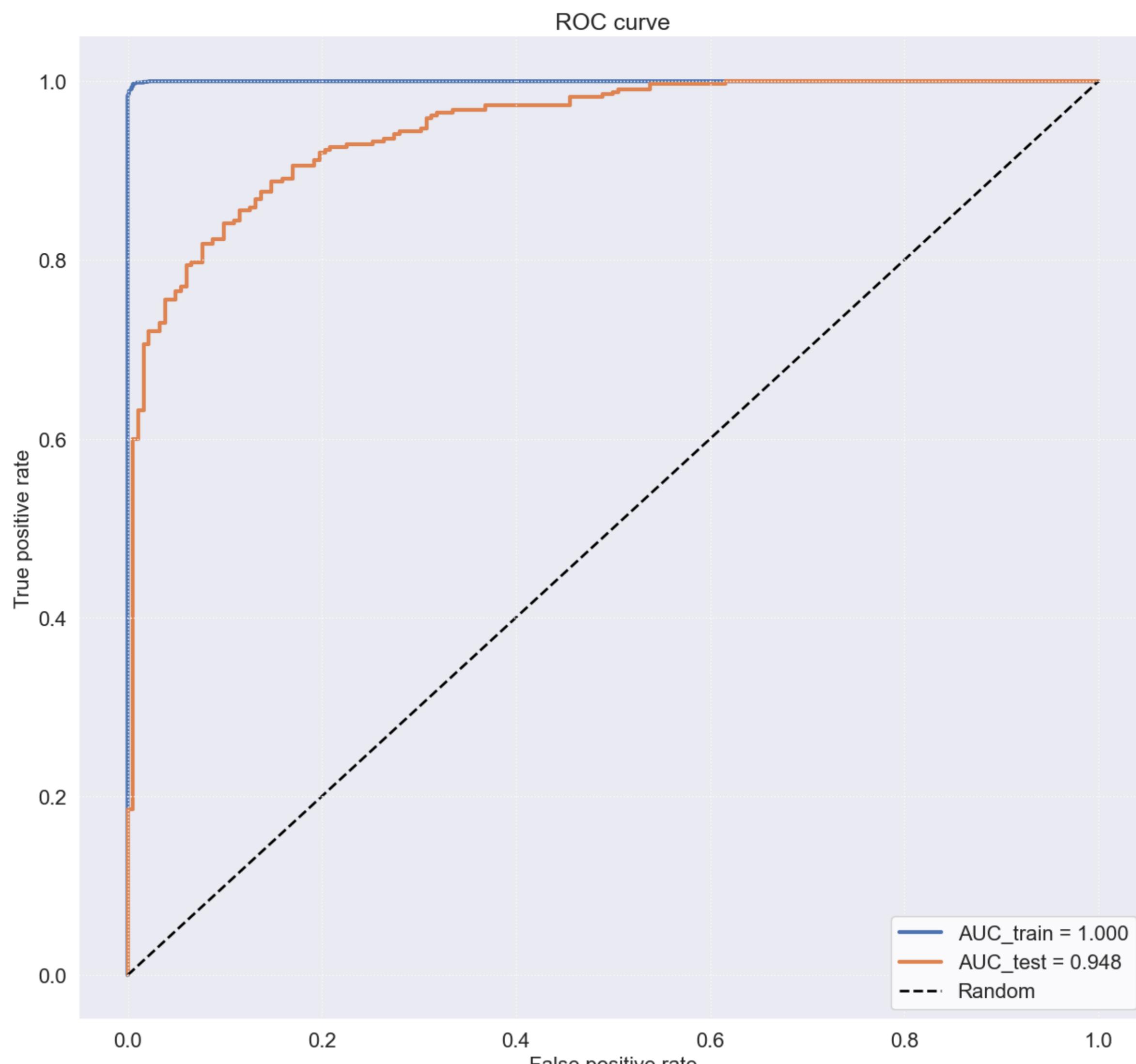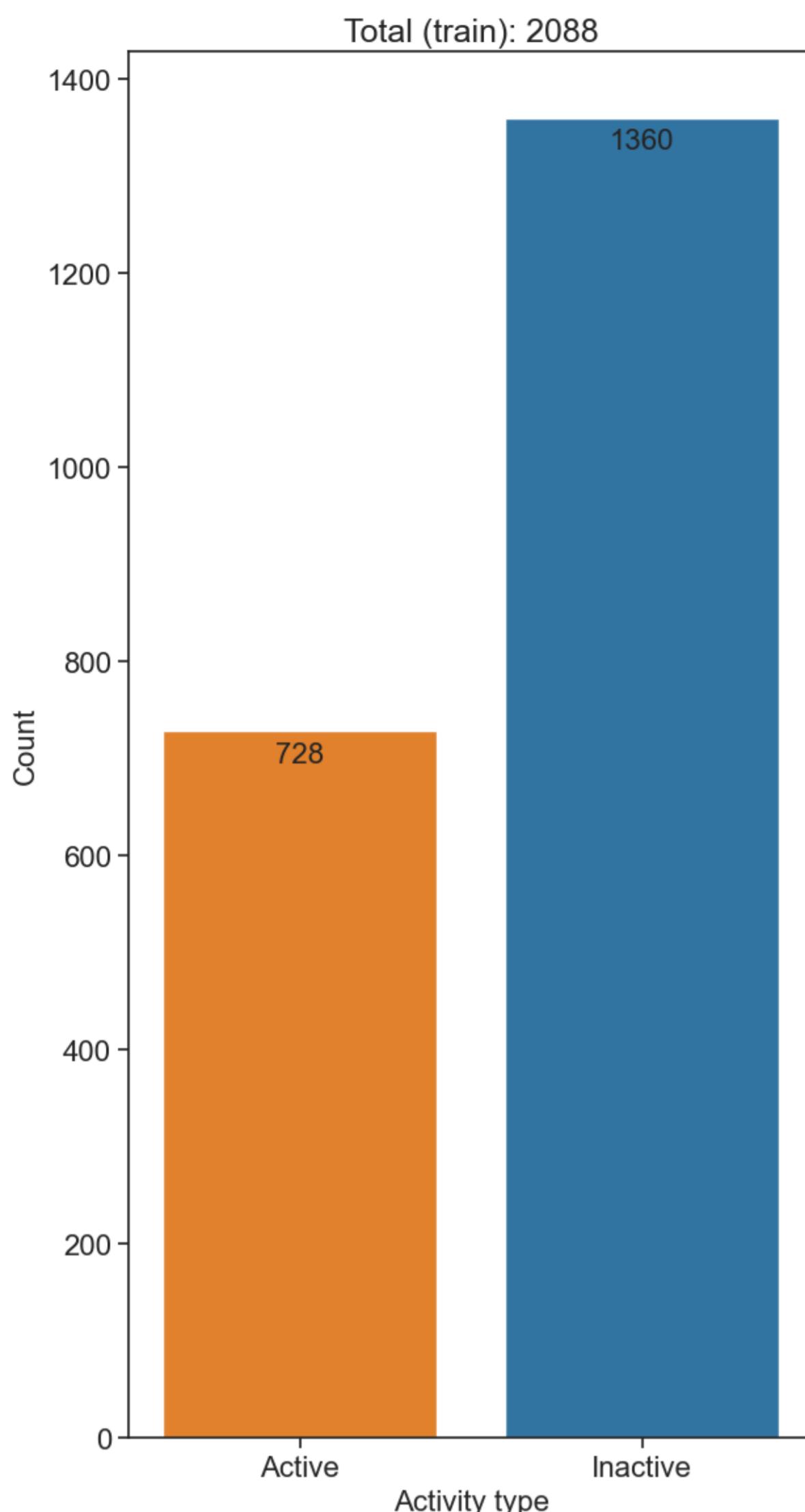
Metric	Train	Test
AUC	0.9999	0.9475
accuracy	0.9933	0.8716
balanced accuracy	0.991	0.8466
recall_(sens)	0.9985	0.9294
specificity	0.9835	0.7637
precision	0.9912	0.8802
f1_score	0.9933	0.8699