Protein: P19793 - metric: balanced_accuracy

Total (train): 352

Active: 185
Inactive: 167

Count / Activity type

ROC curve

True positive rate / False positive rate

AUC_train = 0.995
AUC_test = 0.969
Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 1619)

Difference in Train-Test balanced_accuracy / Score balanced_accuracy (Test)

Confusion Matrix

	0	1
0	42	5
1	1	41

True label / Predicted label

Confusion Matrix (nomr.)

	0	1
0	0.89	0.11
1	0.024	0.98

True label / Predicted label

Calibration curve

Fraction of positives / Mean predicted probability

train: 0.0316
test: 0.0649
perfectly_calibrated

Active/Inactive probability

Inactive zone / Active zone

type
active (42)
inactive (47)
decoy (89)

Percent / Probability

XGBoost Parameters	Value
alpha	4.2429
colsample_bytree	0.963
gamma	0.5315
lambda	2.9912
learning_rate	0.1242
max_depth	3.0
min_child_weight	3.0
n_estimators	692.0
reg_alpha	0.1892
reg_lambda	0.0183
scale_pos_weight	50.1312
subsample	0.5217
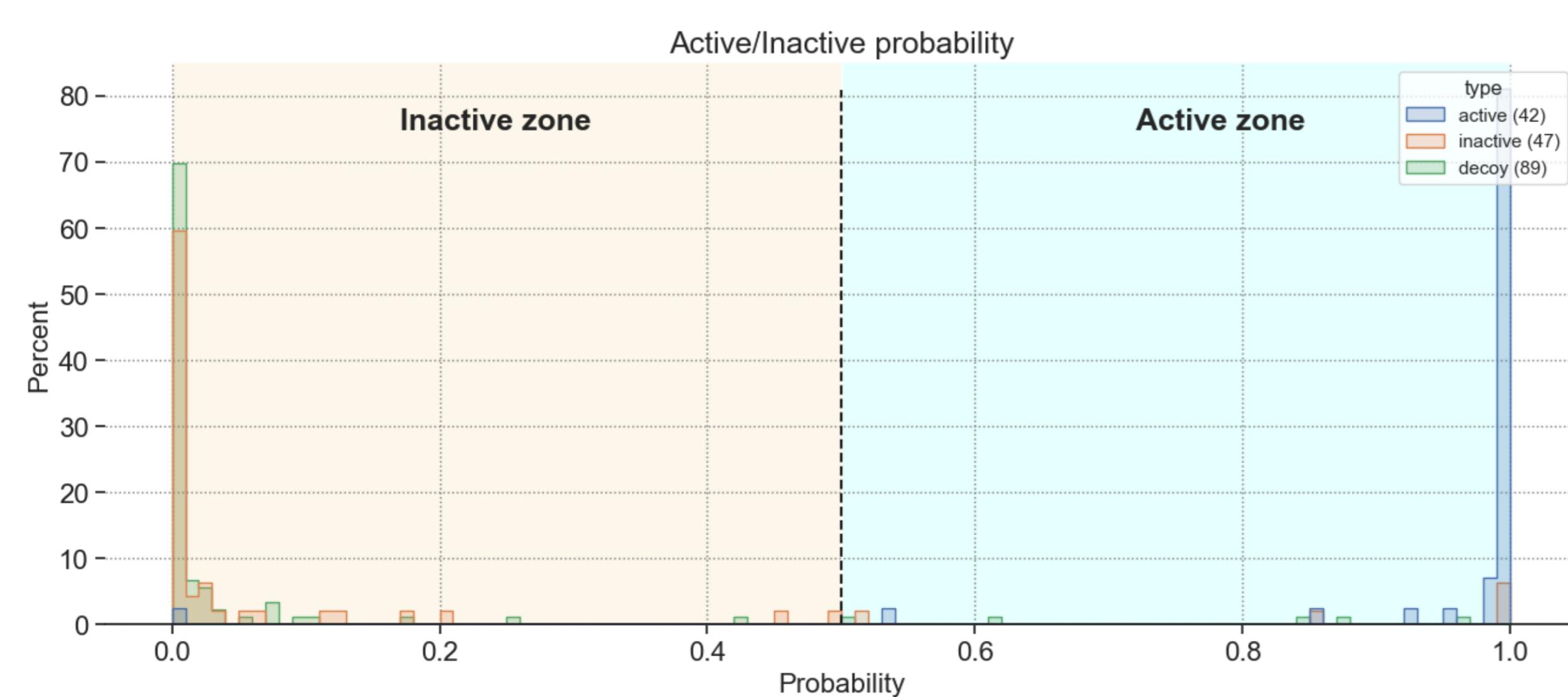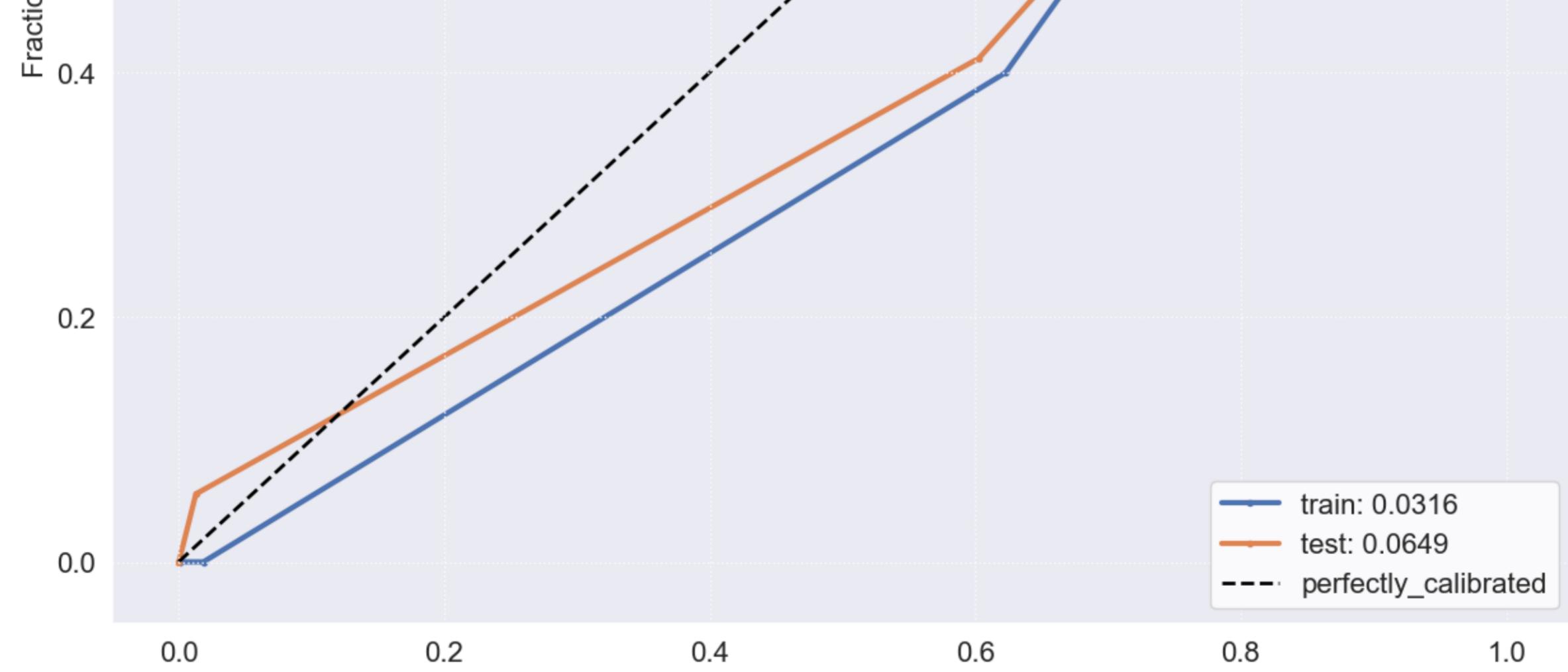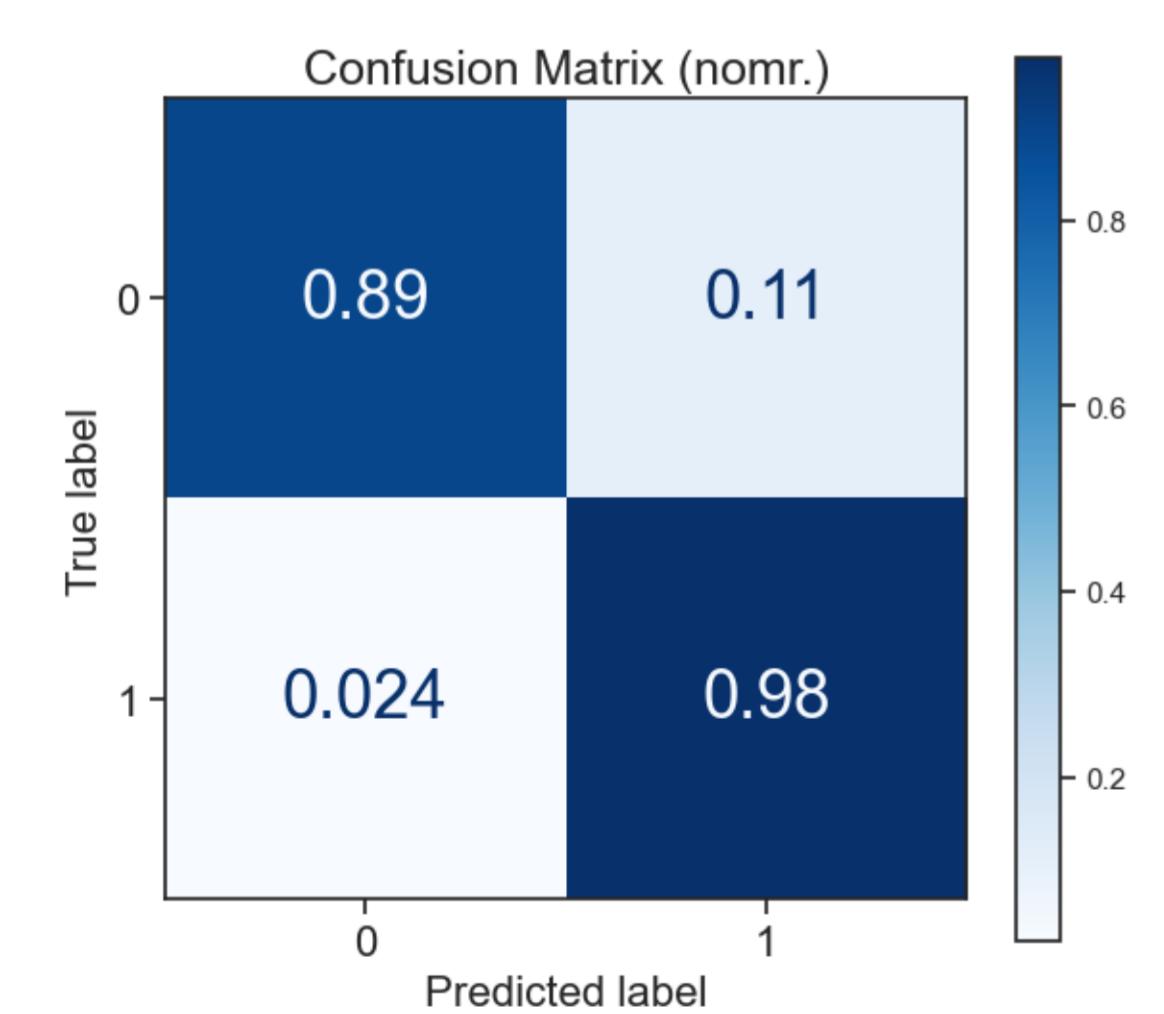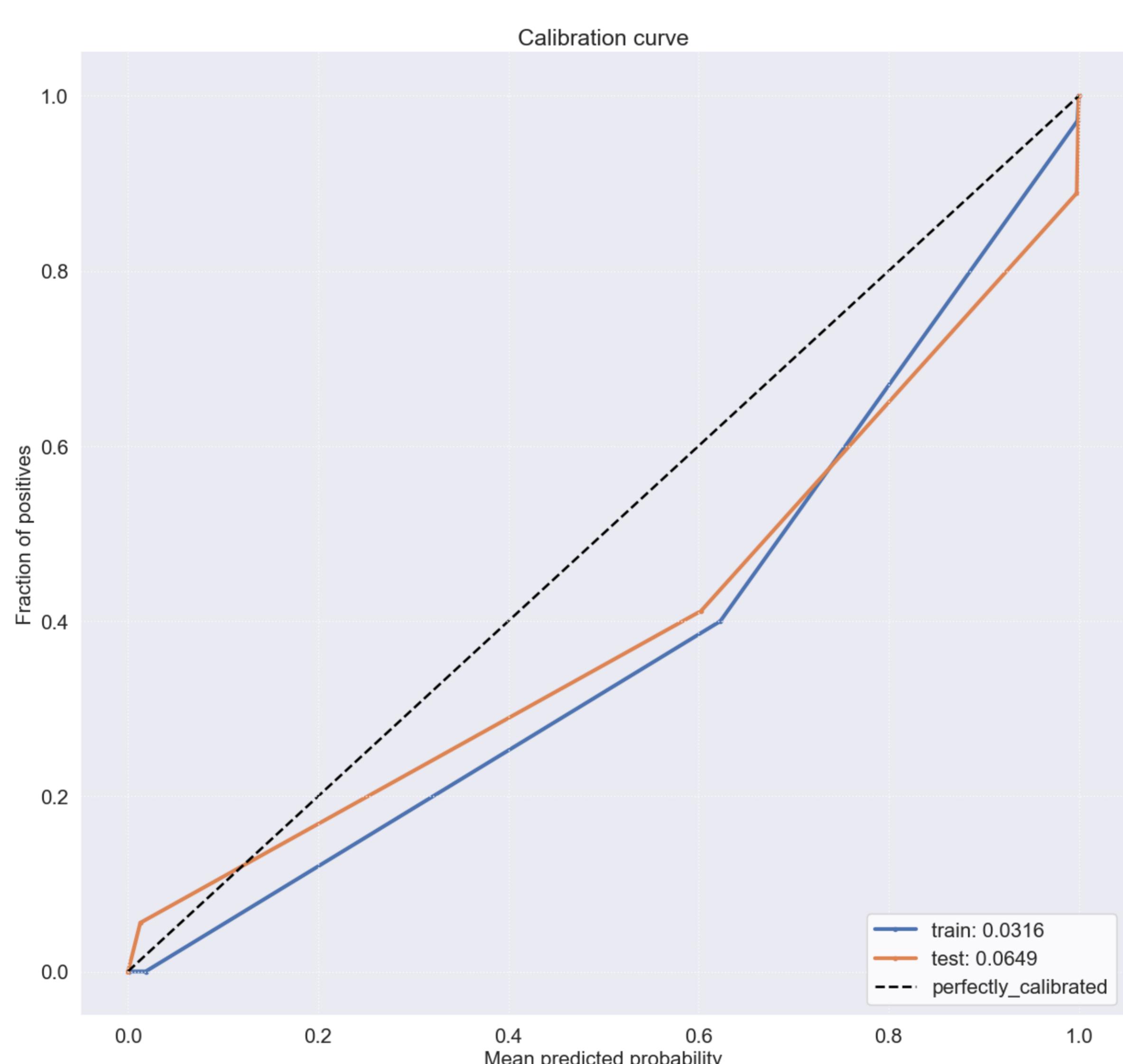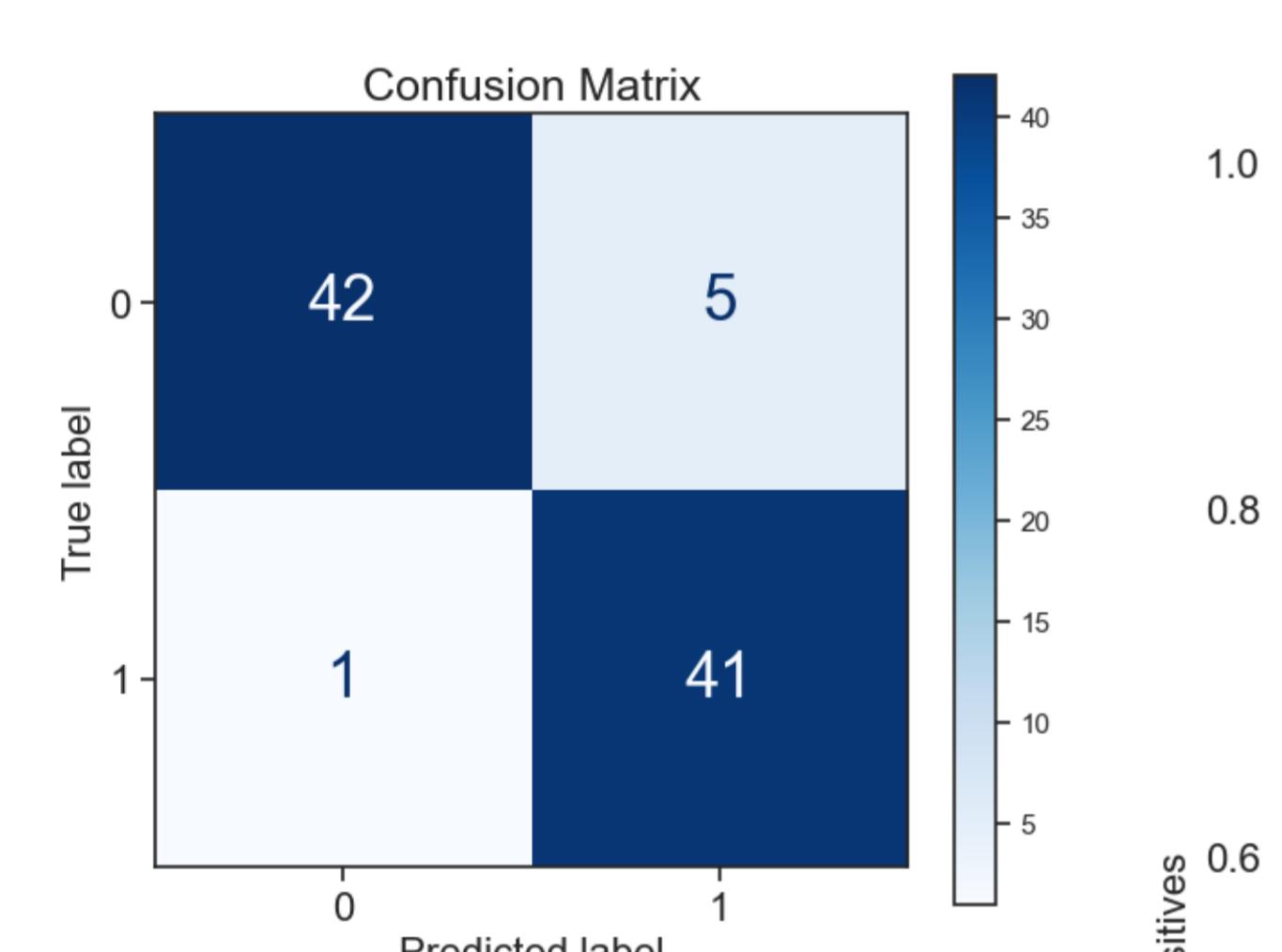
Metric	Train	Test
AUC	0.995	0.9686
accuracy	0.9659	0.9326
balanced accuracy	0.9676	0.9349
recall_(sens)	1.0	0.9762
specificity	0.9351	0.8936
precision	0.933	0.8913
f1_score	0.9659	0.9326

Protein: P19793 - metric: f1

Total (train): 352

ROC curve

- AUC_train = 1.000
- AUC_test = 0.983
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 1673)
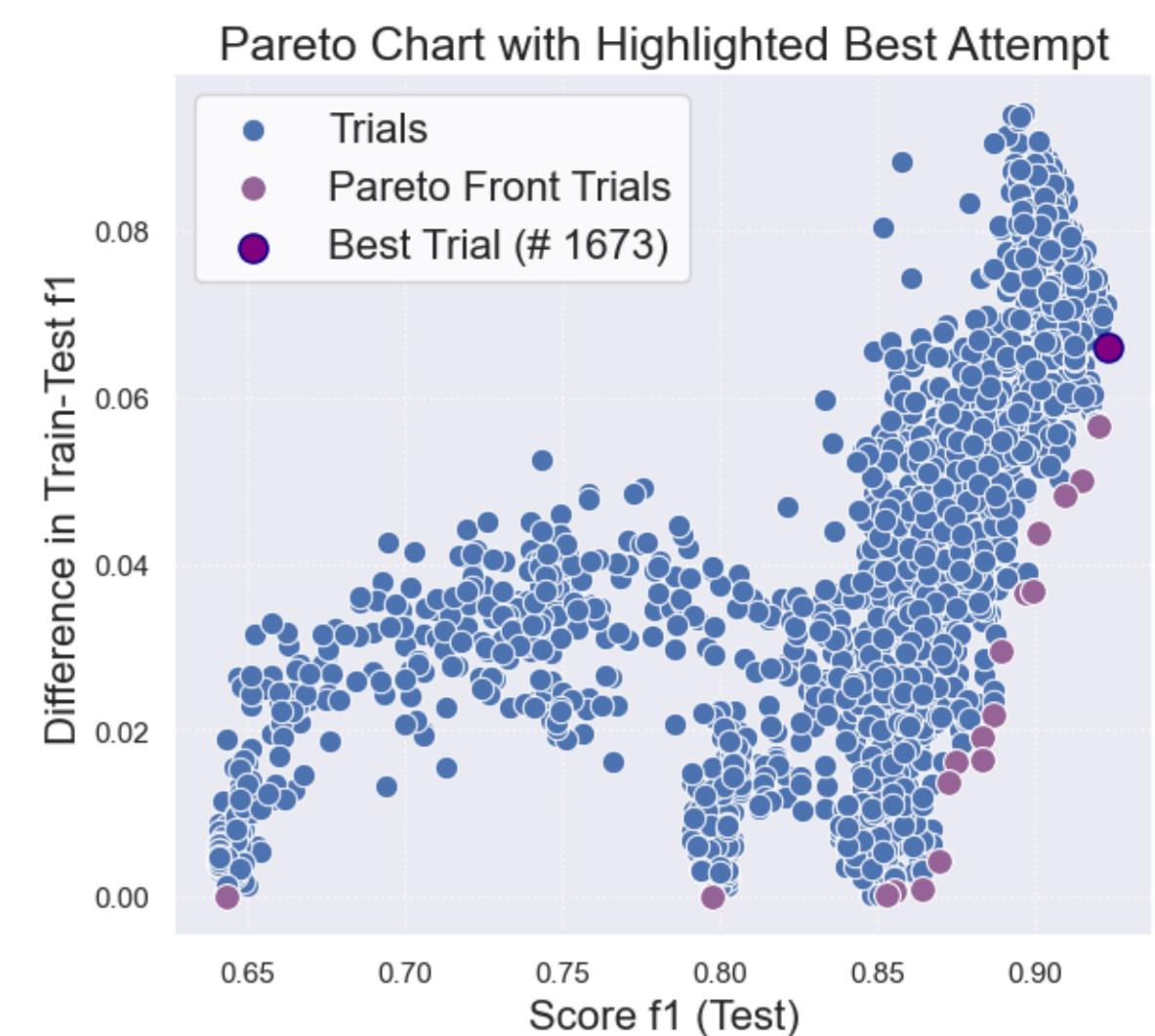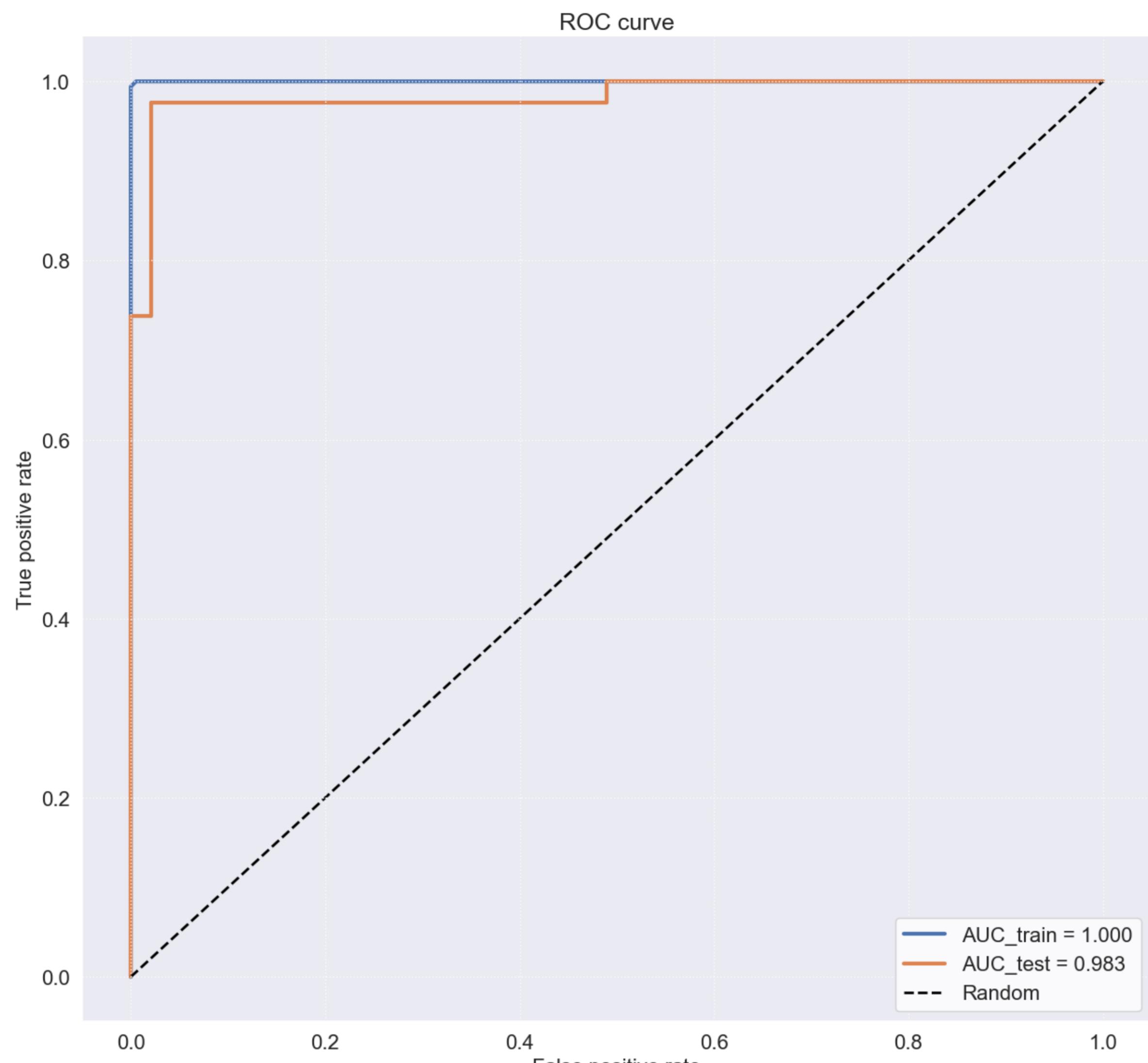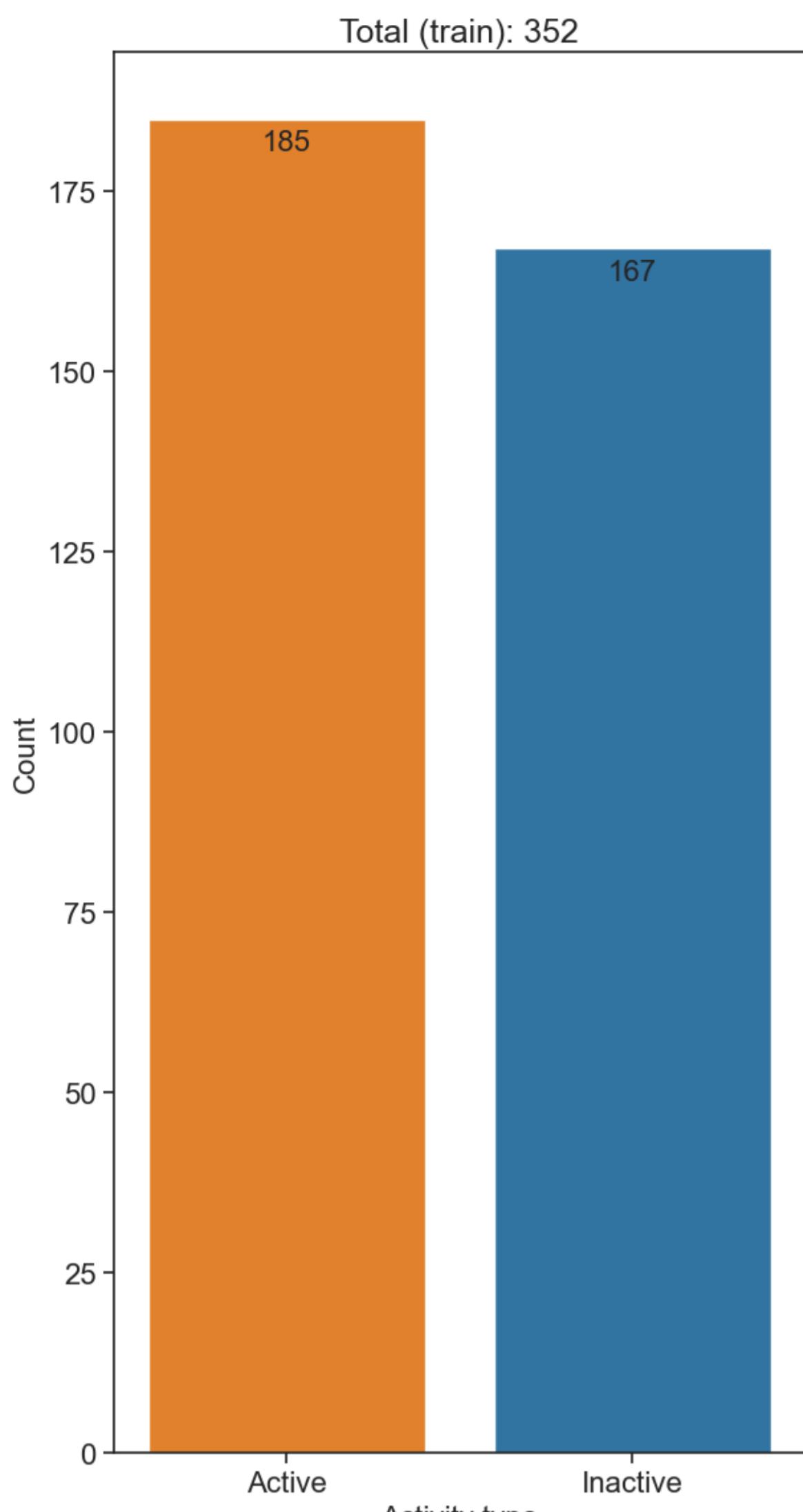
XGBoost Parameters	Value
alpha	4.4634
colsample_bytree	0.6164
gamma	0.086
lambda	1.8294
learning_rate	0.3562
max_depth	5.0
min_child_weight	1.0
n_estimators	313.0
reg_alpha	0.5757
reg_lambda	0.4371
scale_pos_weight	25.2867
subsample	0.6867

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

- train: 0.0085
- test: 0.0483
- perfectly_calibrated

Active/Inactive probability

Inactive zone

Active zone

type
- active (42)
- inactive (47)
- decoy (89)
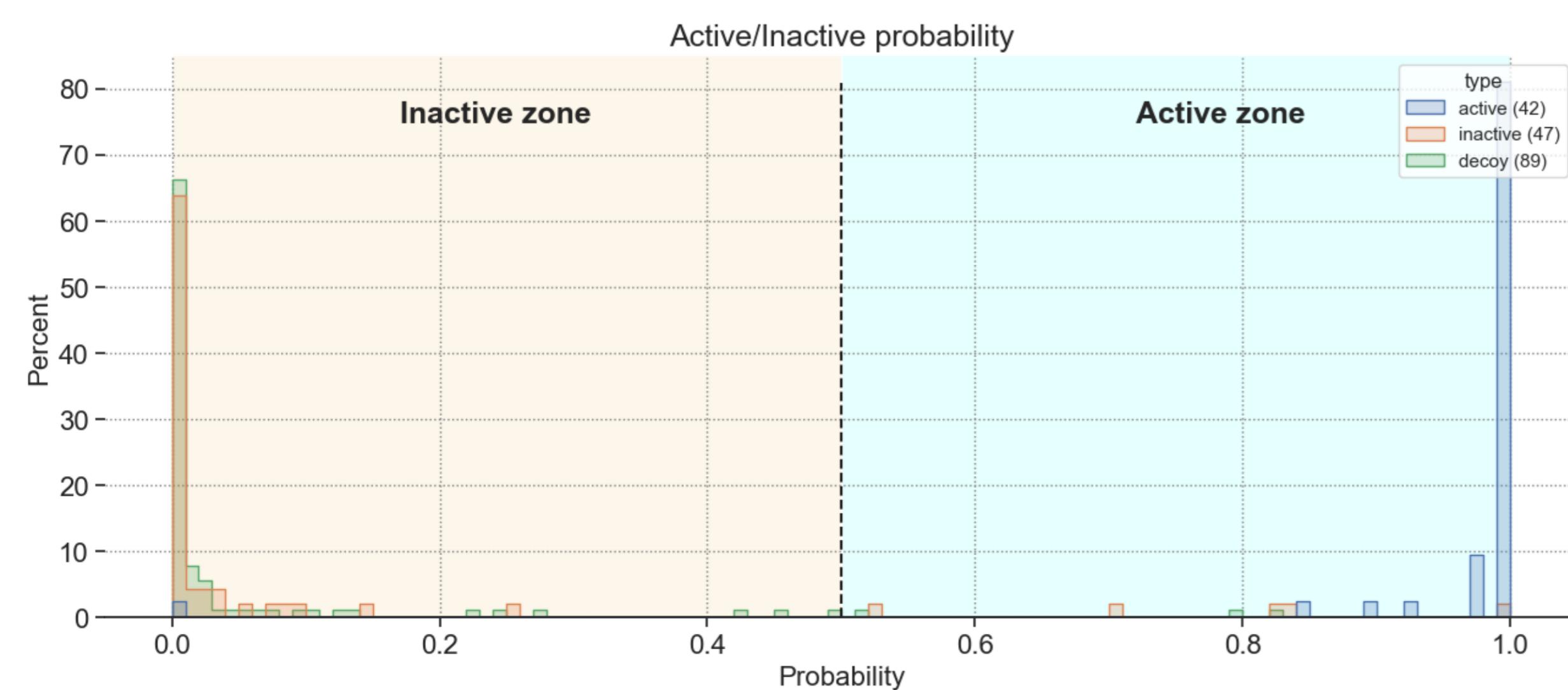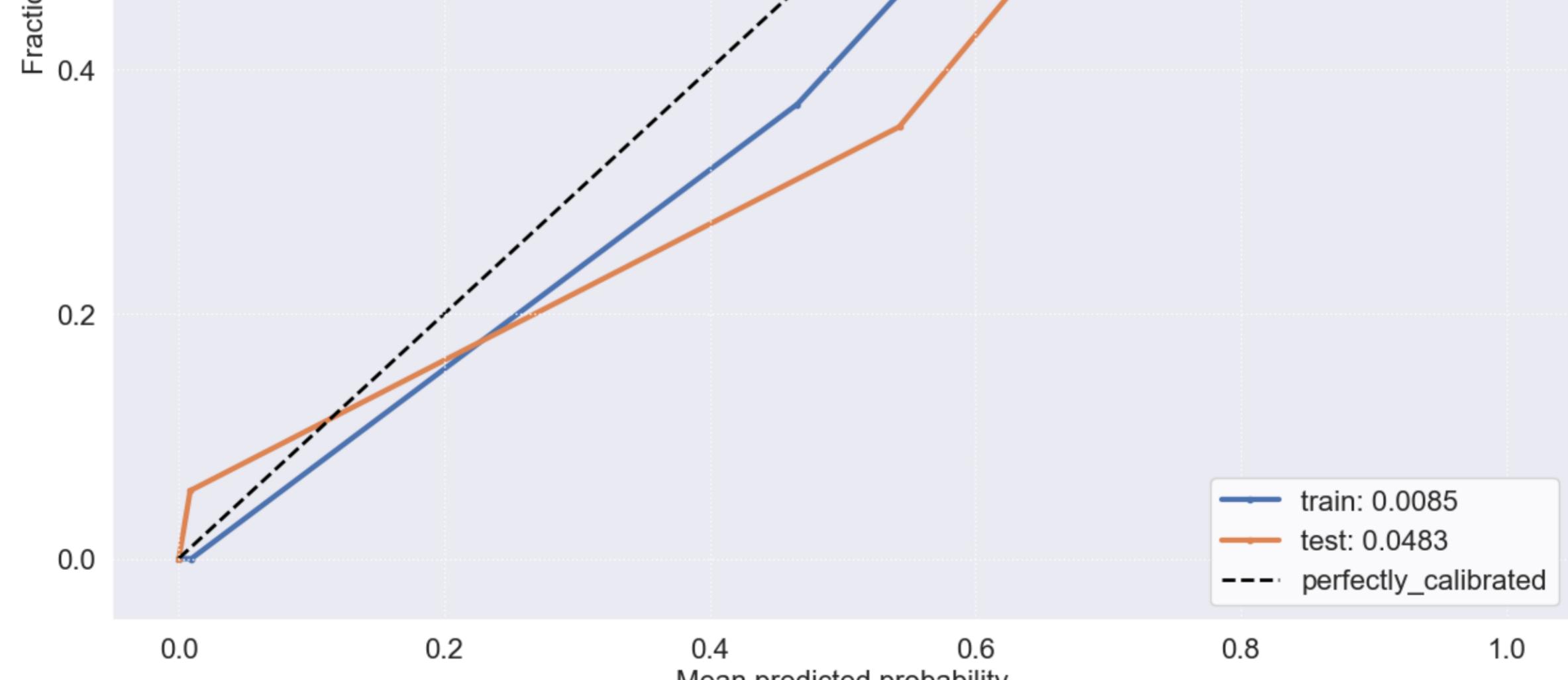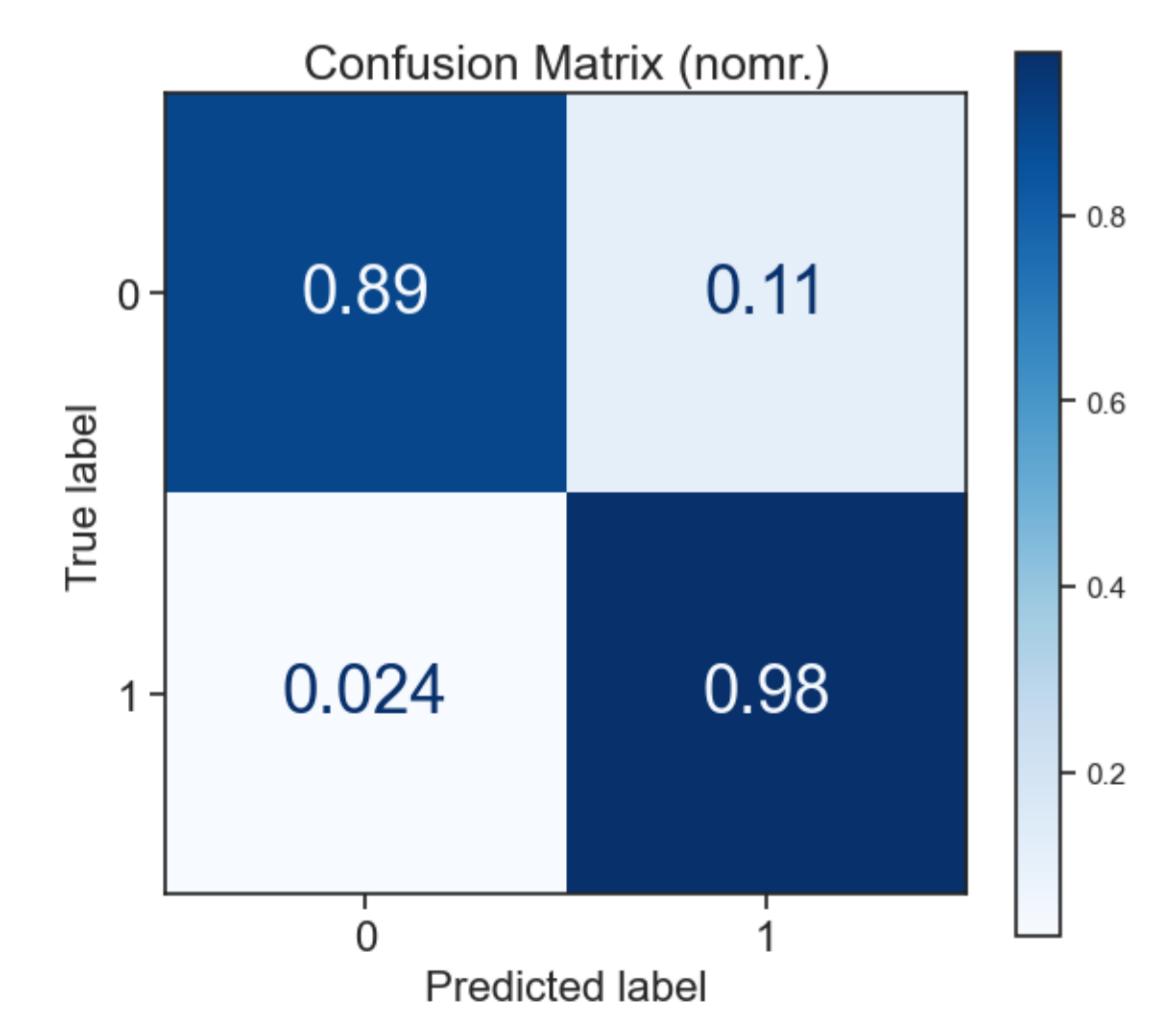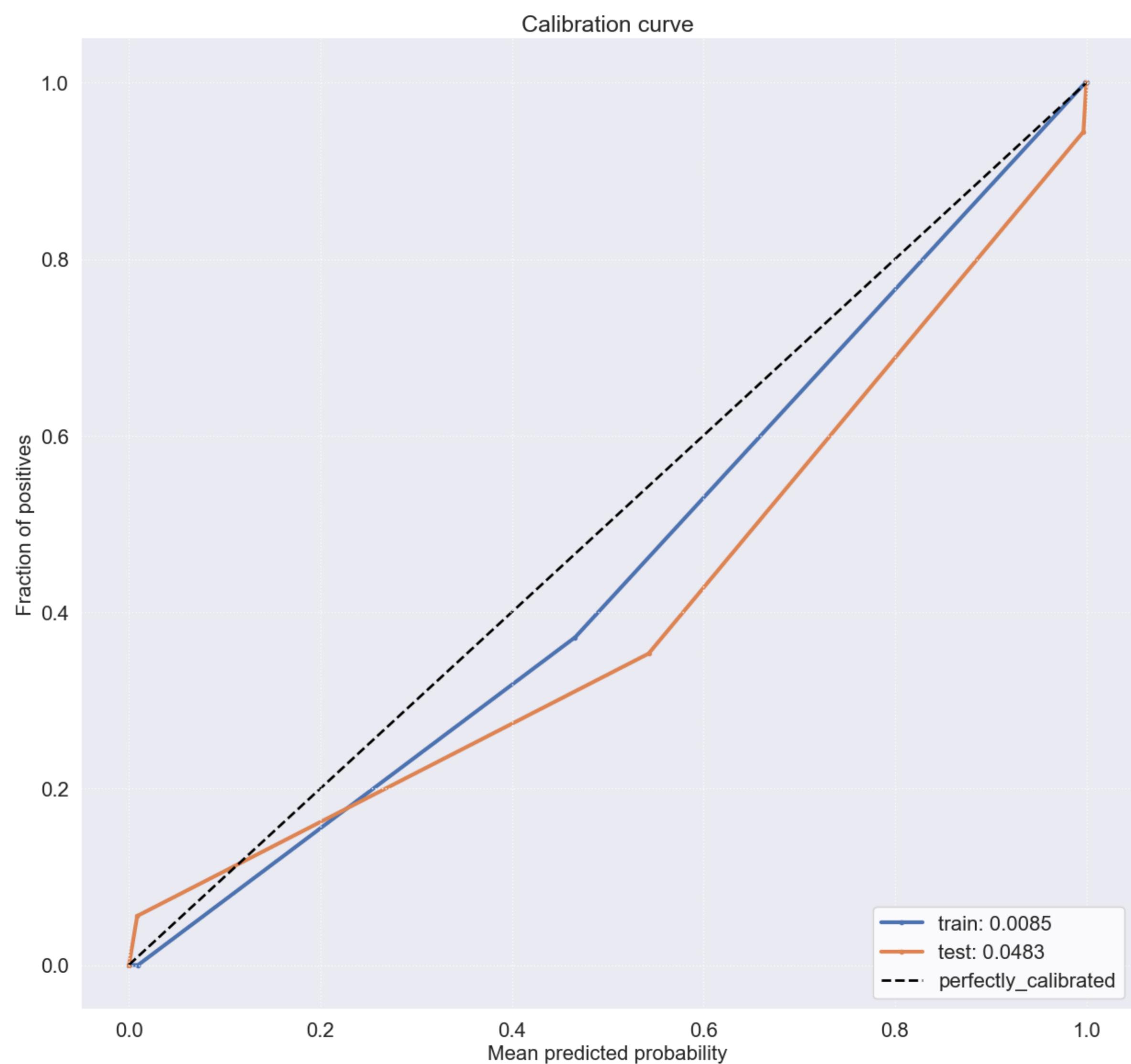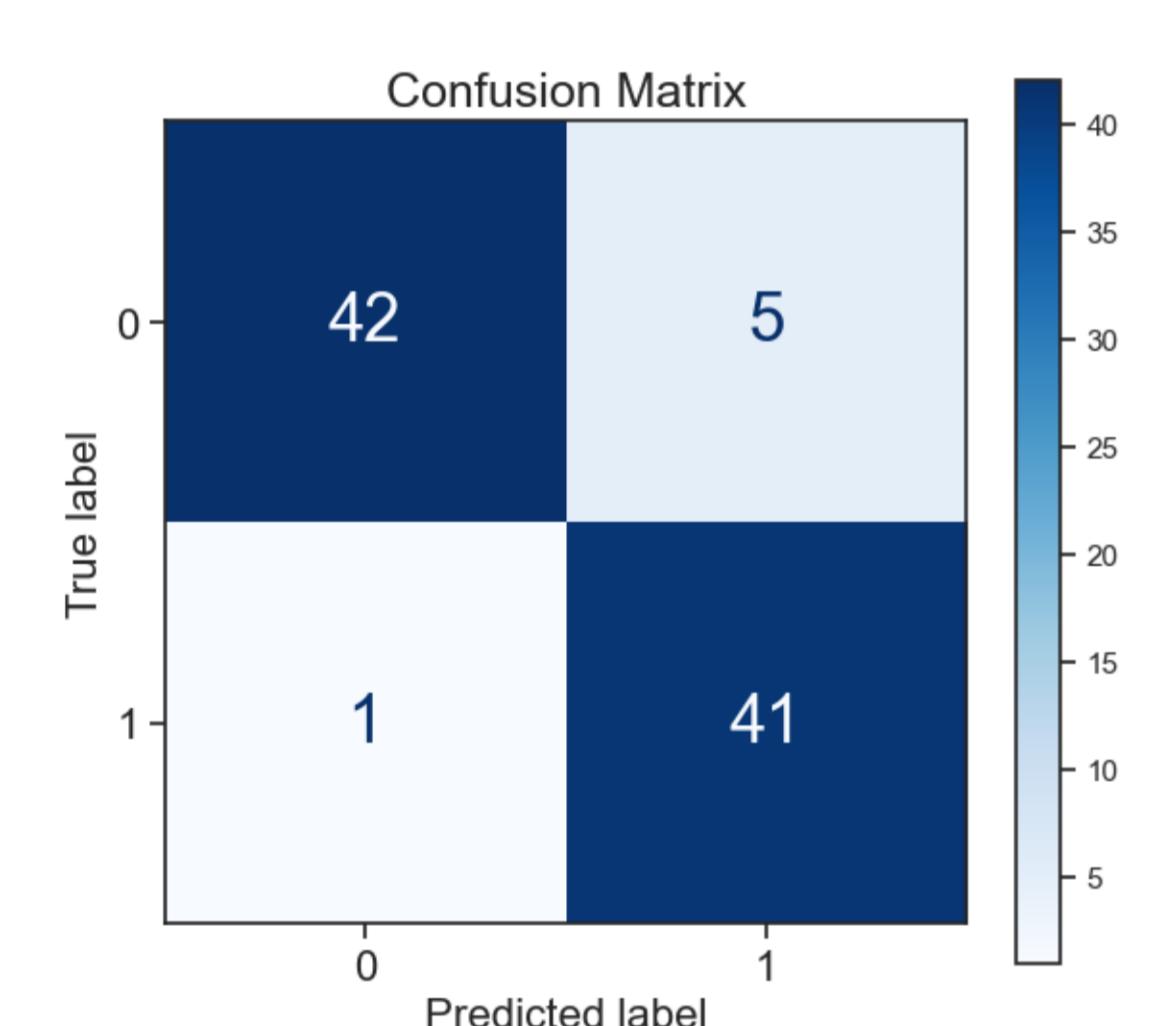
Metric	Train	Test
AUC	1.0	0.9833
accuracy	0.9886	0.9326
balanced accuracy	0.9892	0.9349
recall_(sens)	1.0	0.9762
specificity	0.9784	0.8936
precision	0.9766	0.8913
f1_score	0.9886	0.9326

Protein: P19793 - metric: precision

Total (train): 352

ROC curve

AUC_train = 0.998
AUC_test = 0.978
Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 1719)
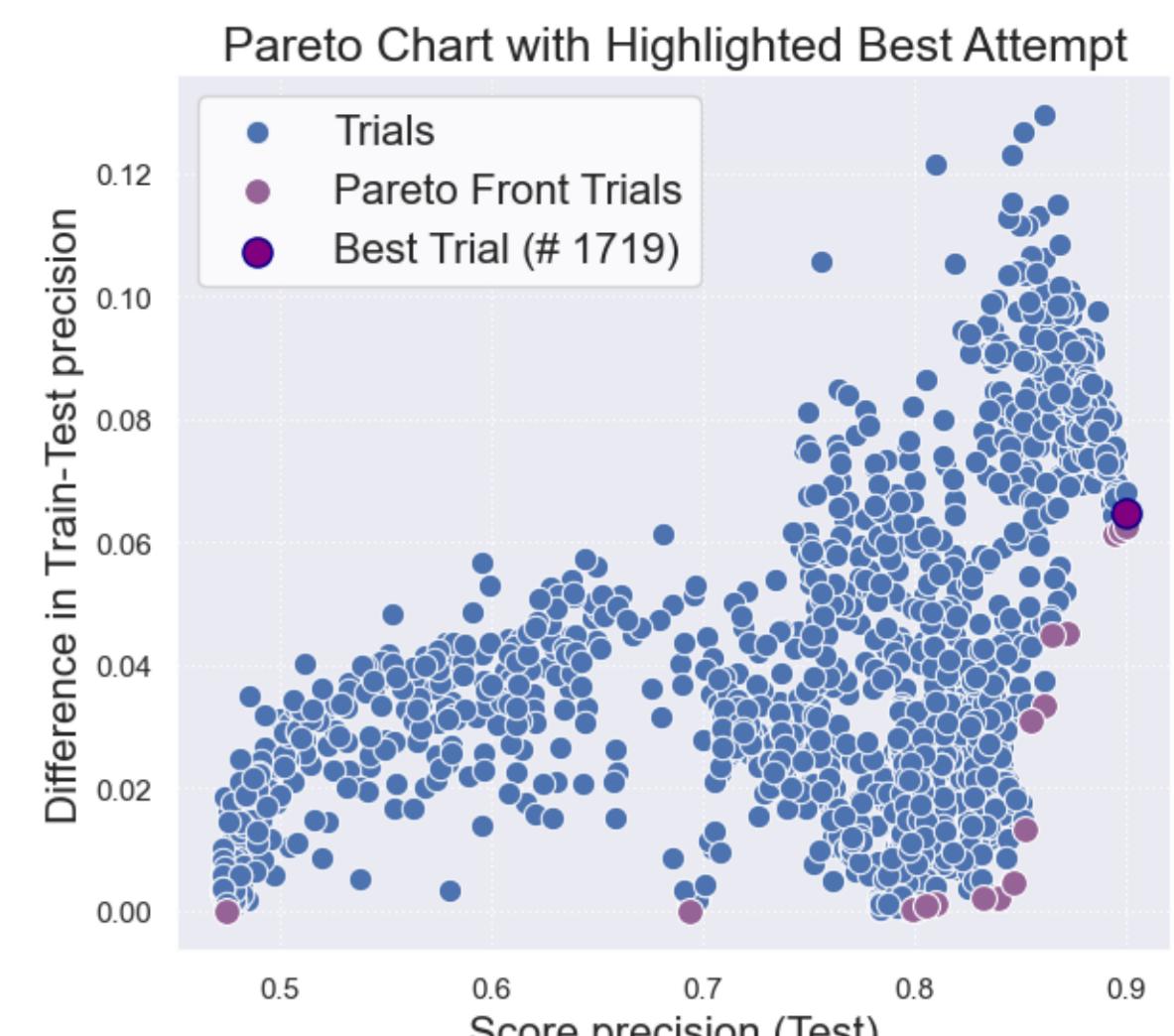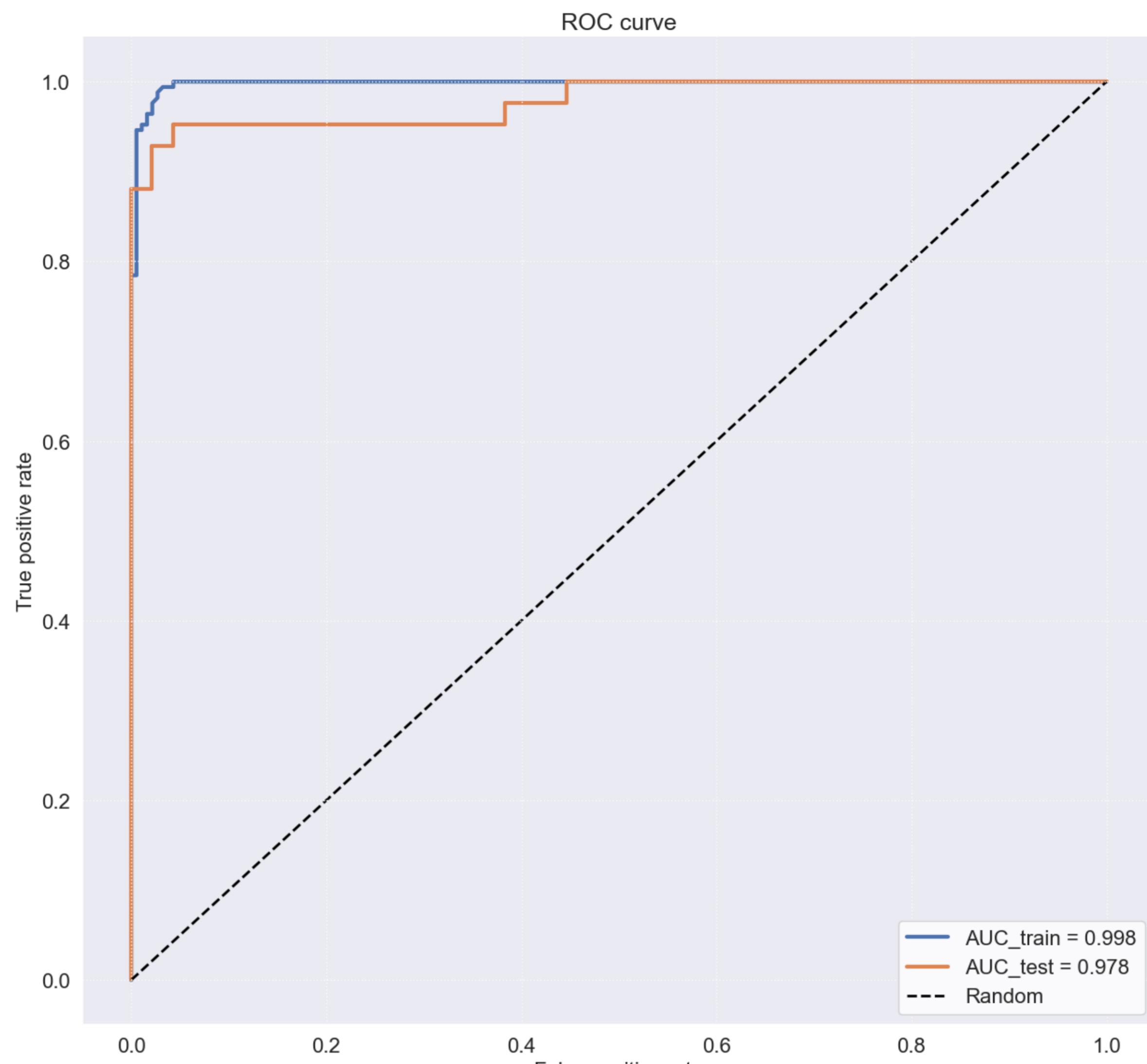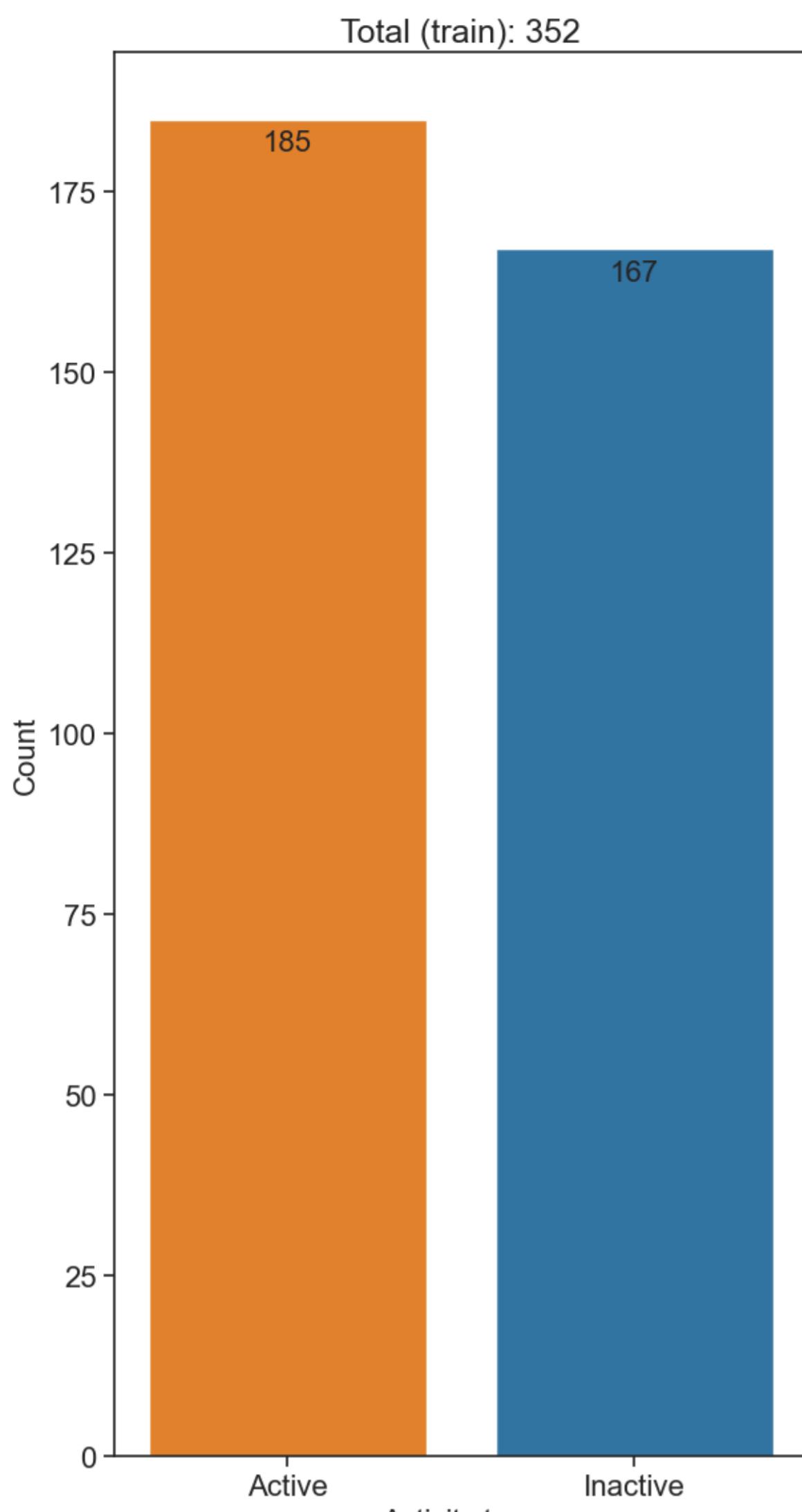
XGBoost Parameters	Value
alpha	1.7722
colsample_bytree	0.7288
gamma	0.119
lambda	4.6043
learning_rate	0.1714
max_depth	9.0
min_child_weight	2.0
n_estimators	335.0
reg_alpha	0.2269
reg_lambda	0.8956
scale_pos_weight	2.4632
subsample	0.5541

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

train: 0.0199
test: 0.0581
perfectly_calibrated

Active/Inactive probability

type
active (42)
inactive (47)
decoy (89)
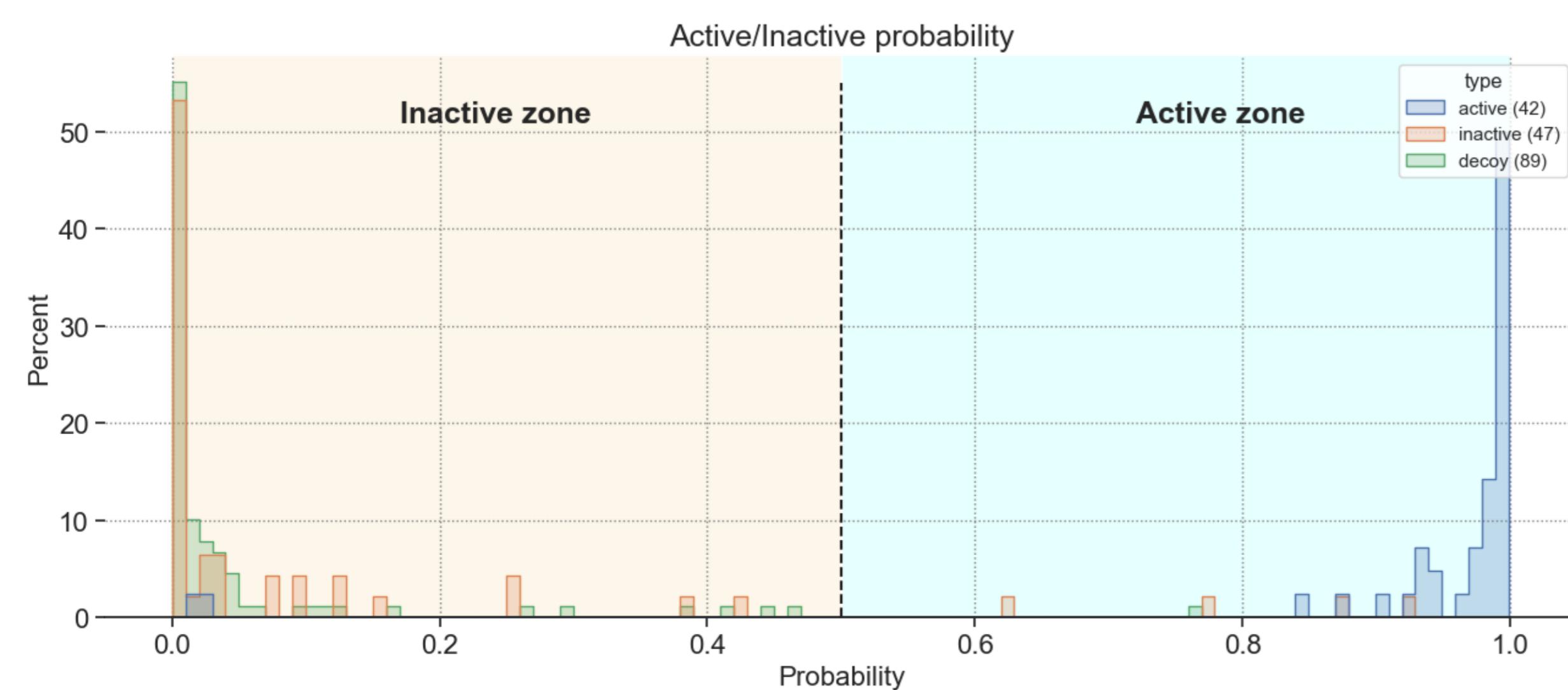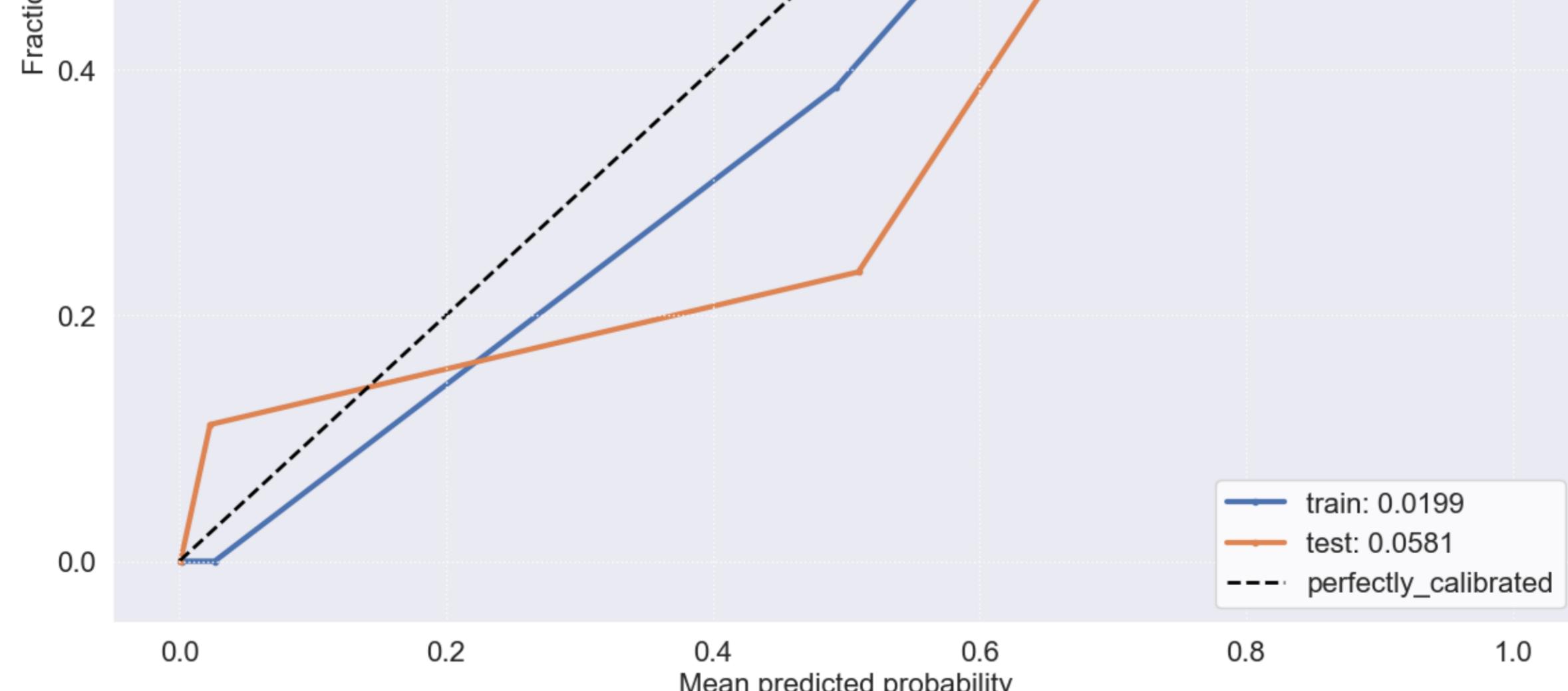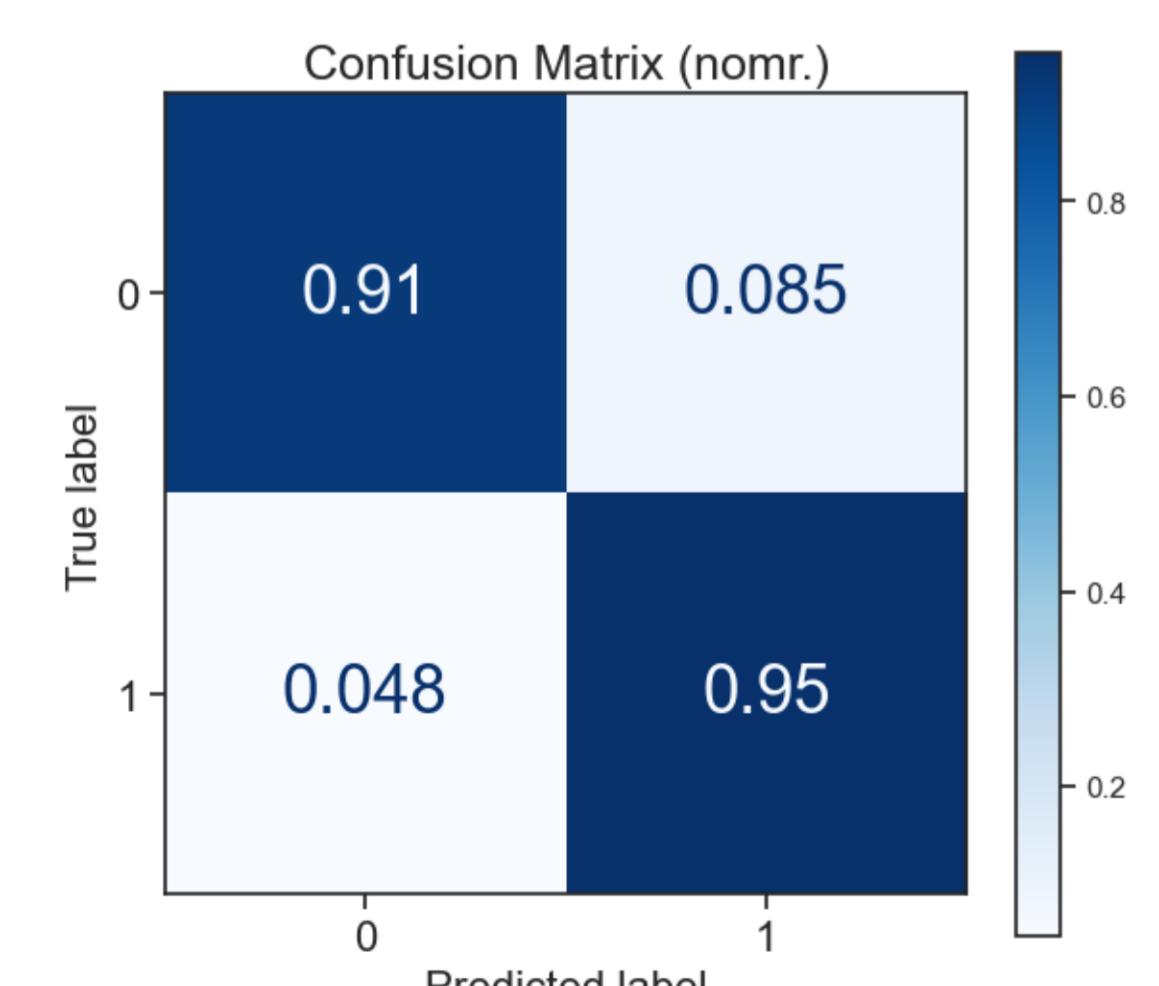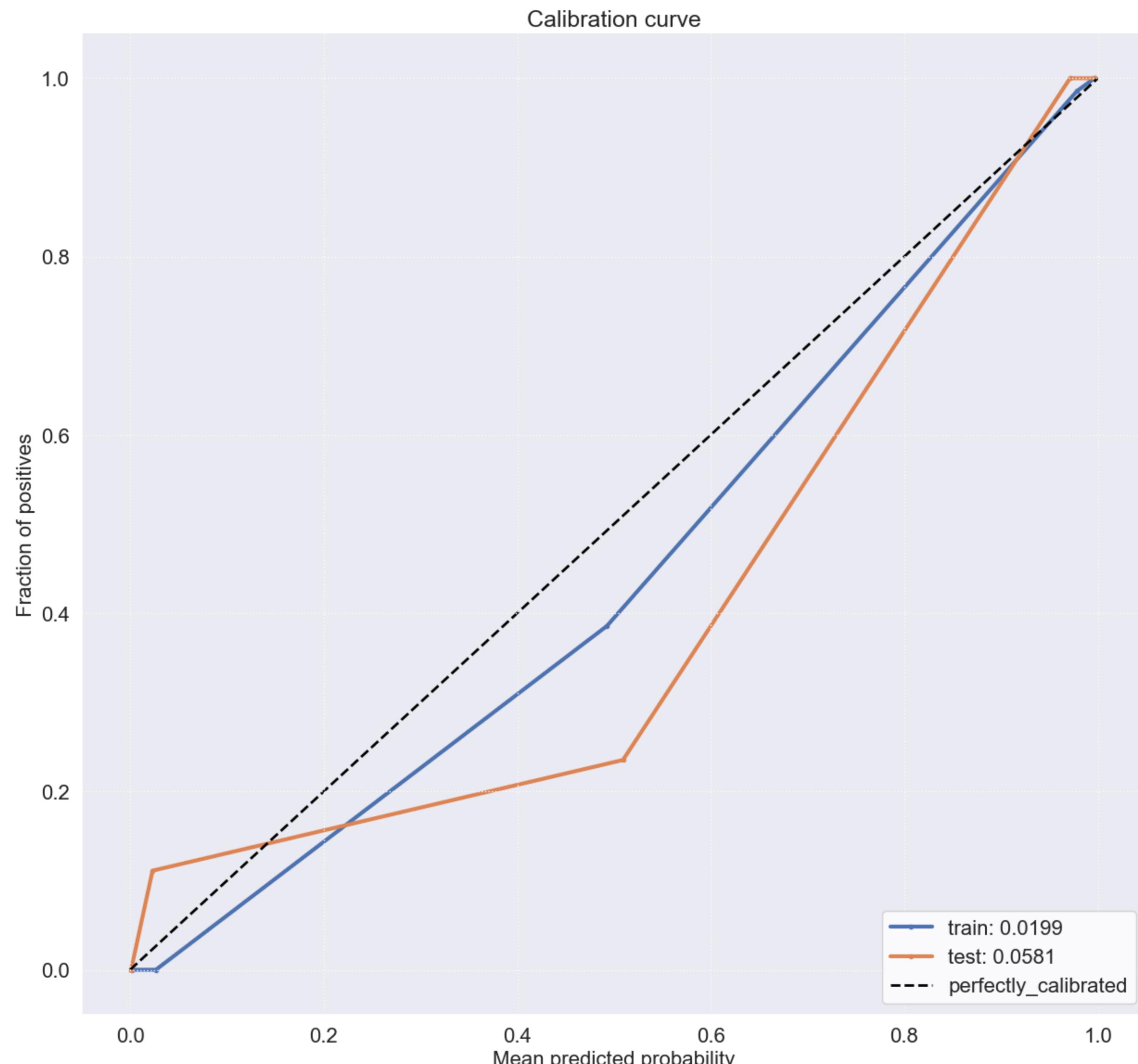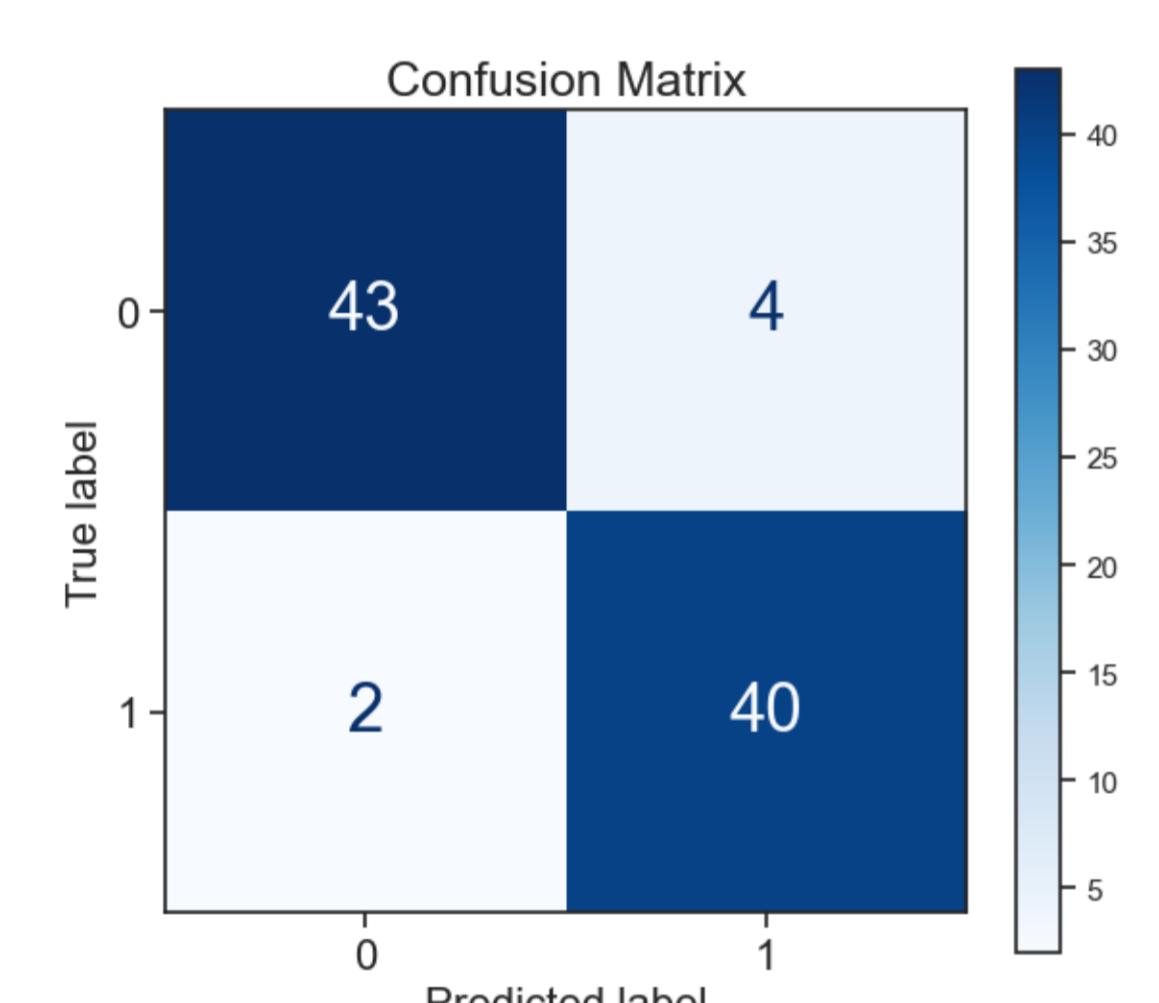
Metric	Train	Test
AUC	0.9979	0.9782
accuracy	0.9801	0.9326
balanced accuracy	0.9808	0.9336
recall_(sens)	0.994	0.9524
specificity	0.9676	0.9149
precision	0.9651	0.9091
f1_score	0.9801	0.9326

Protein: P19793 - metric: SoftVotingXGBoostEnsemble
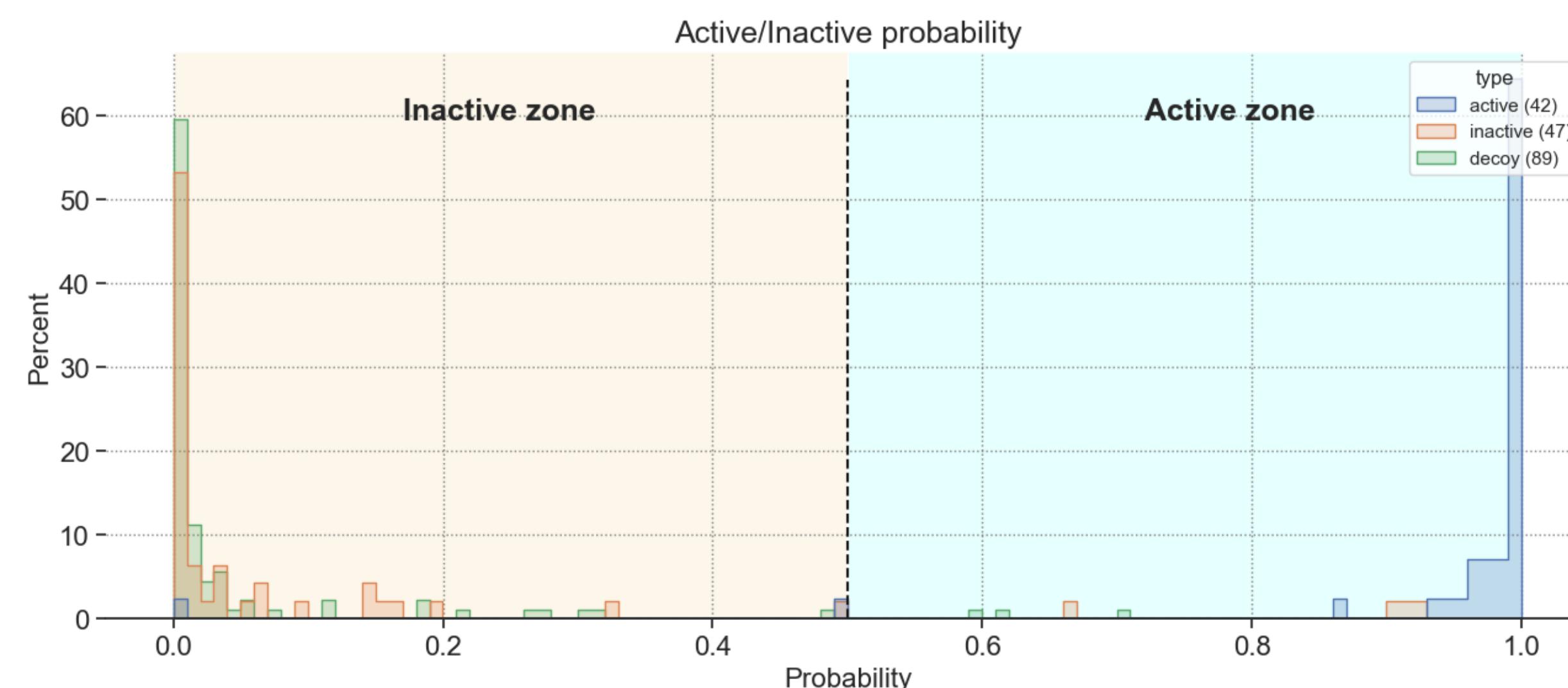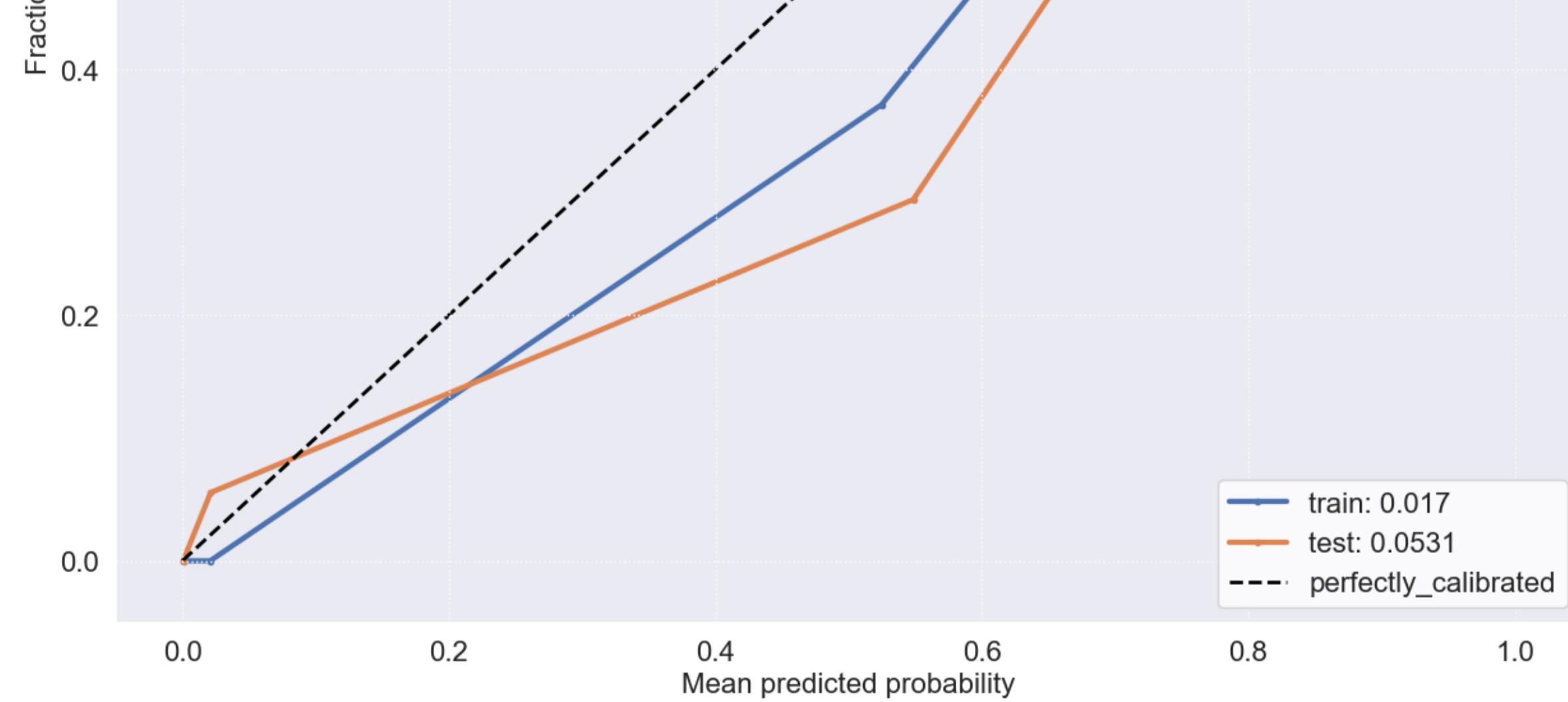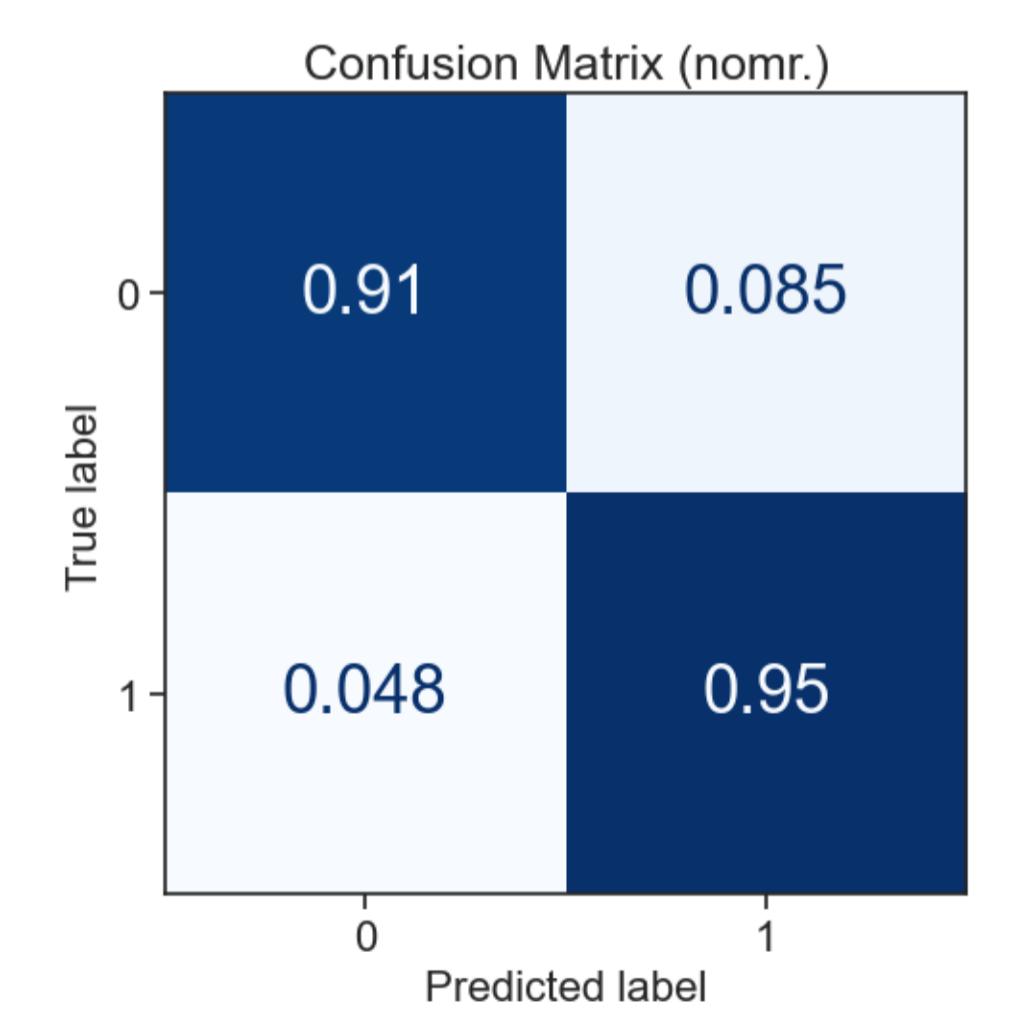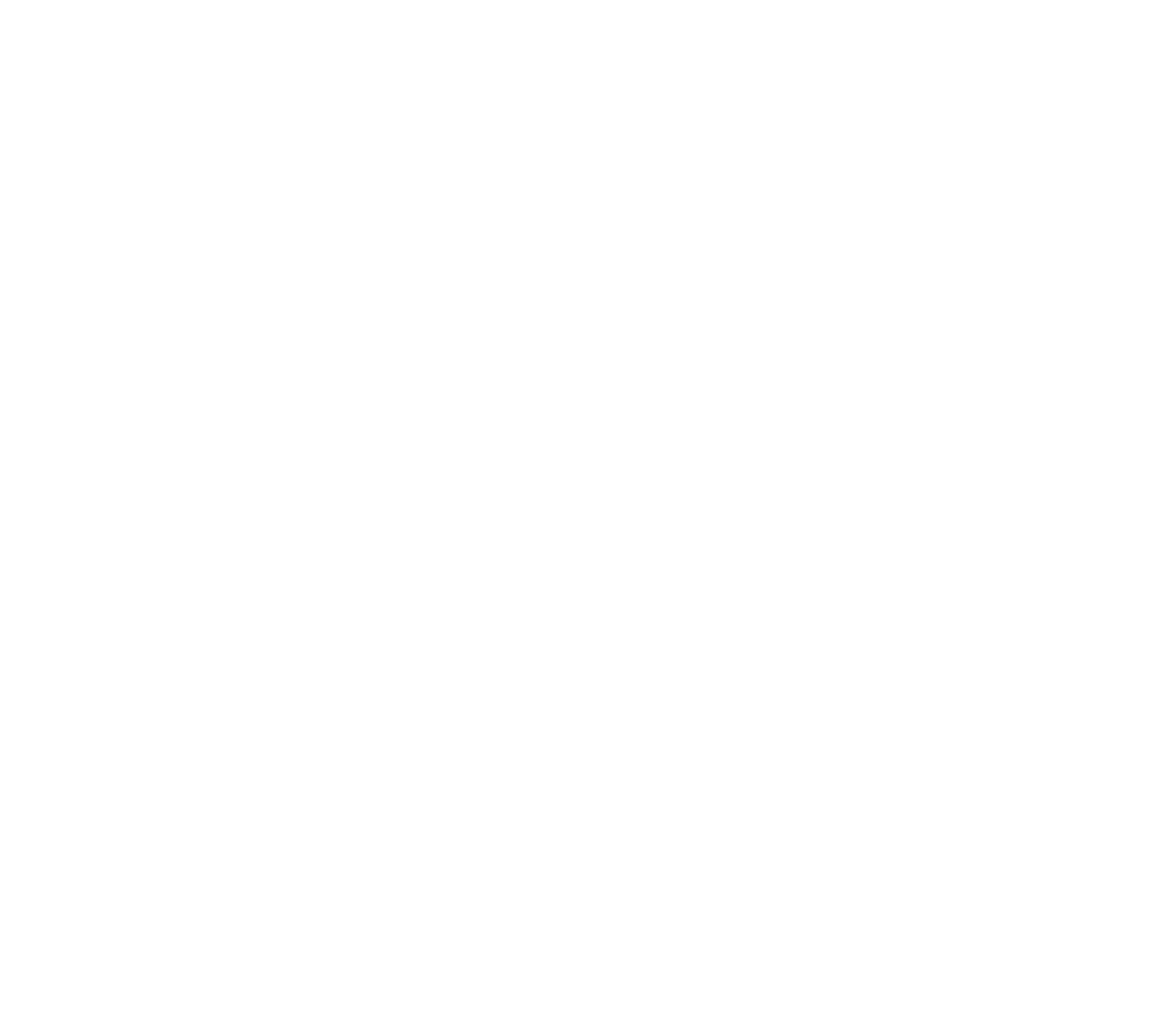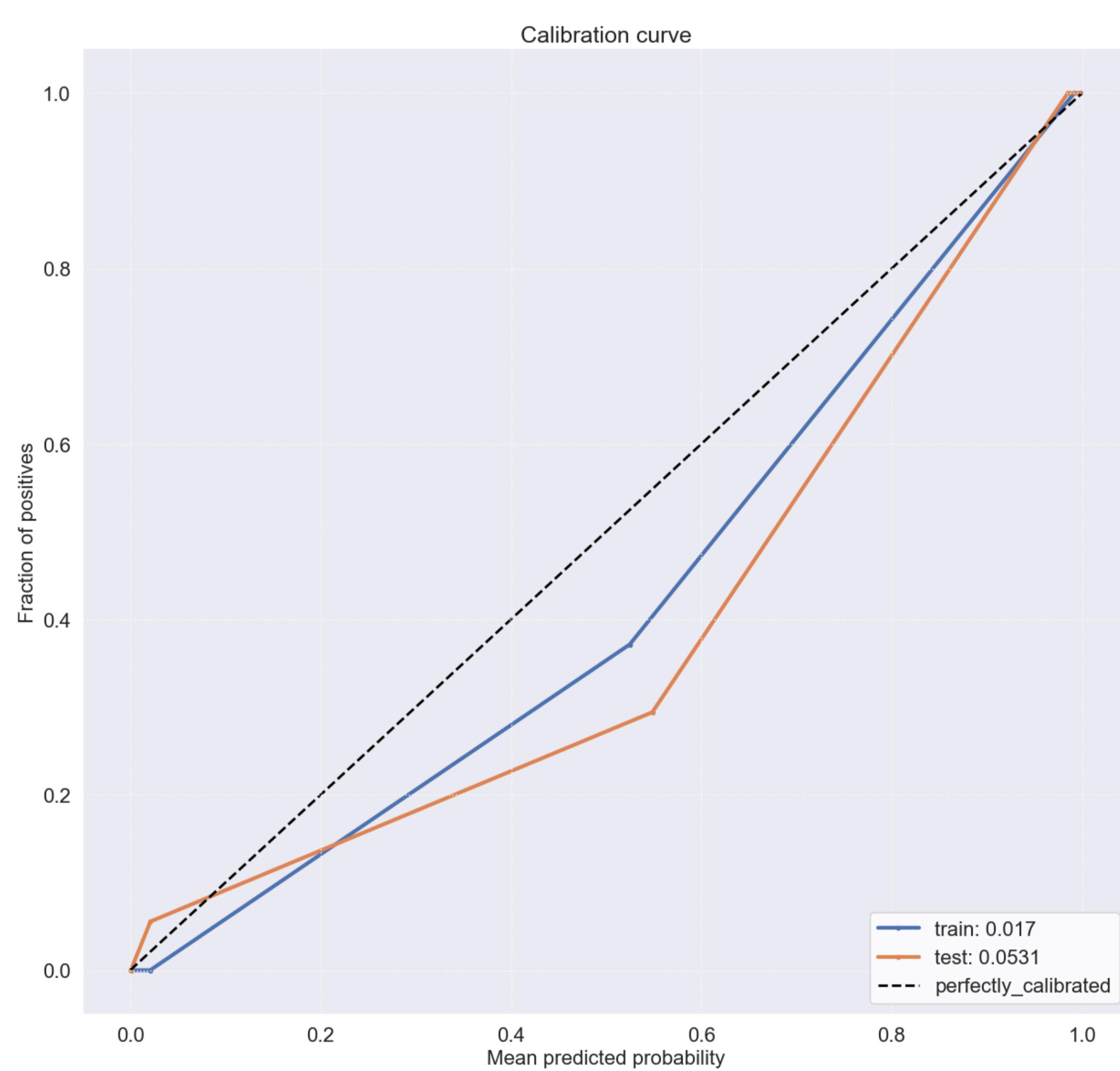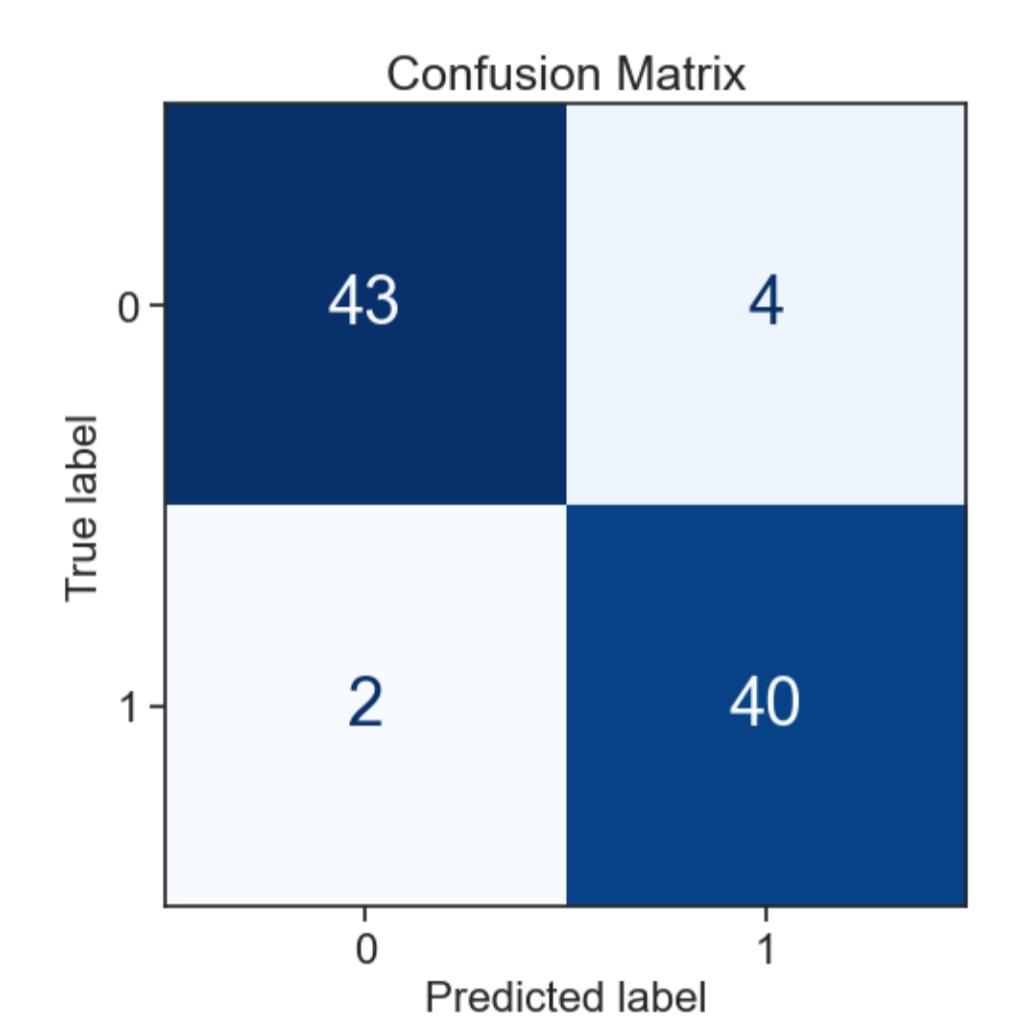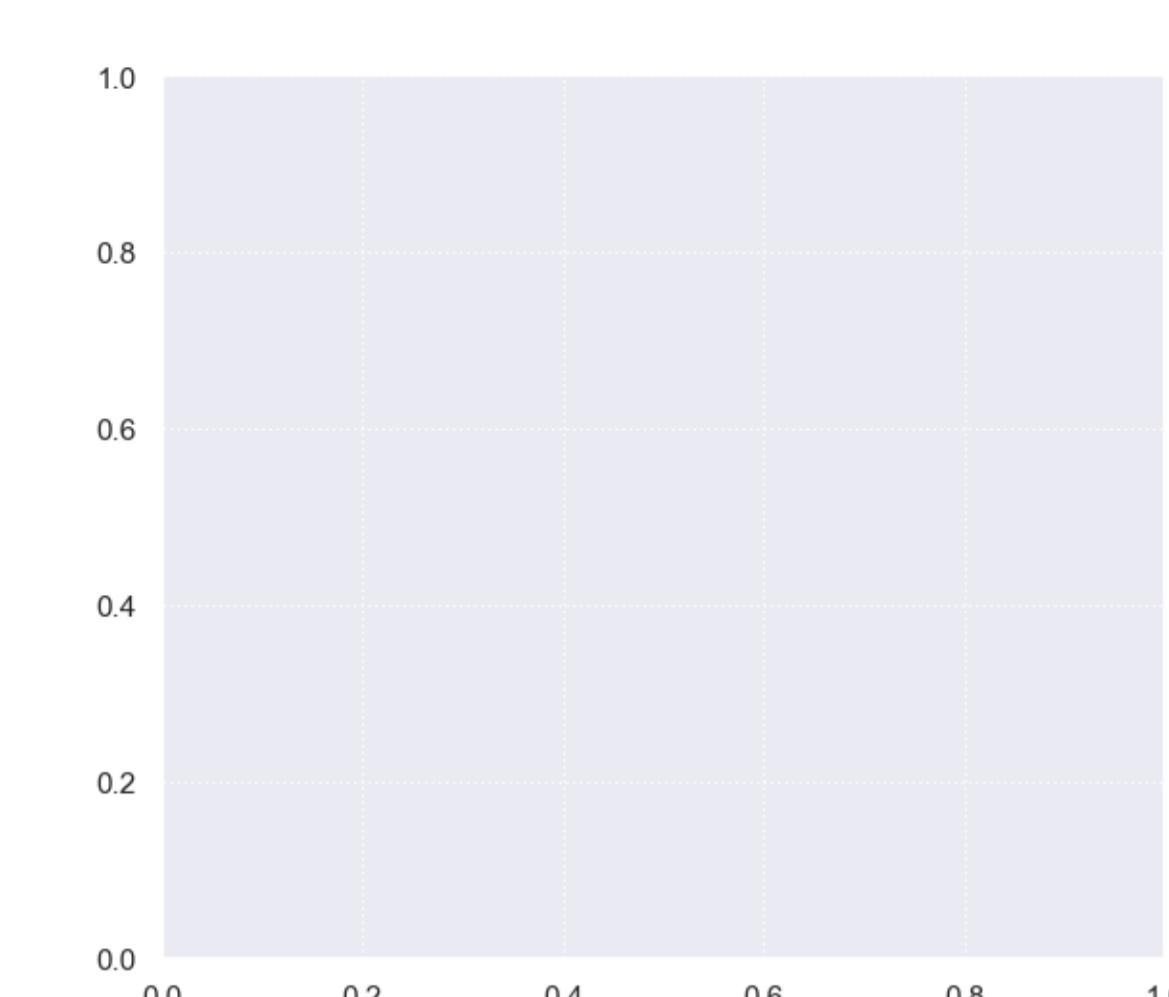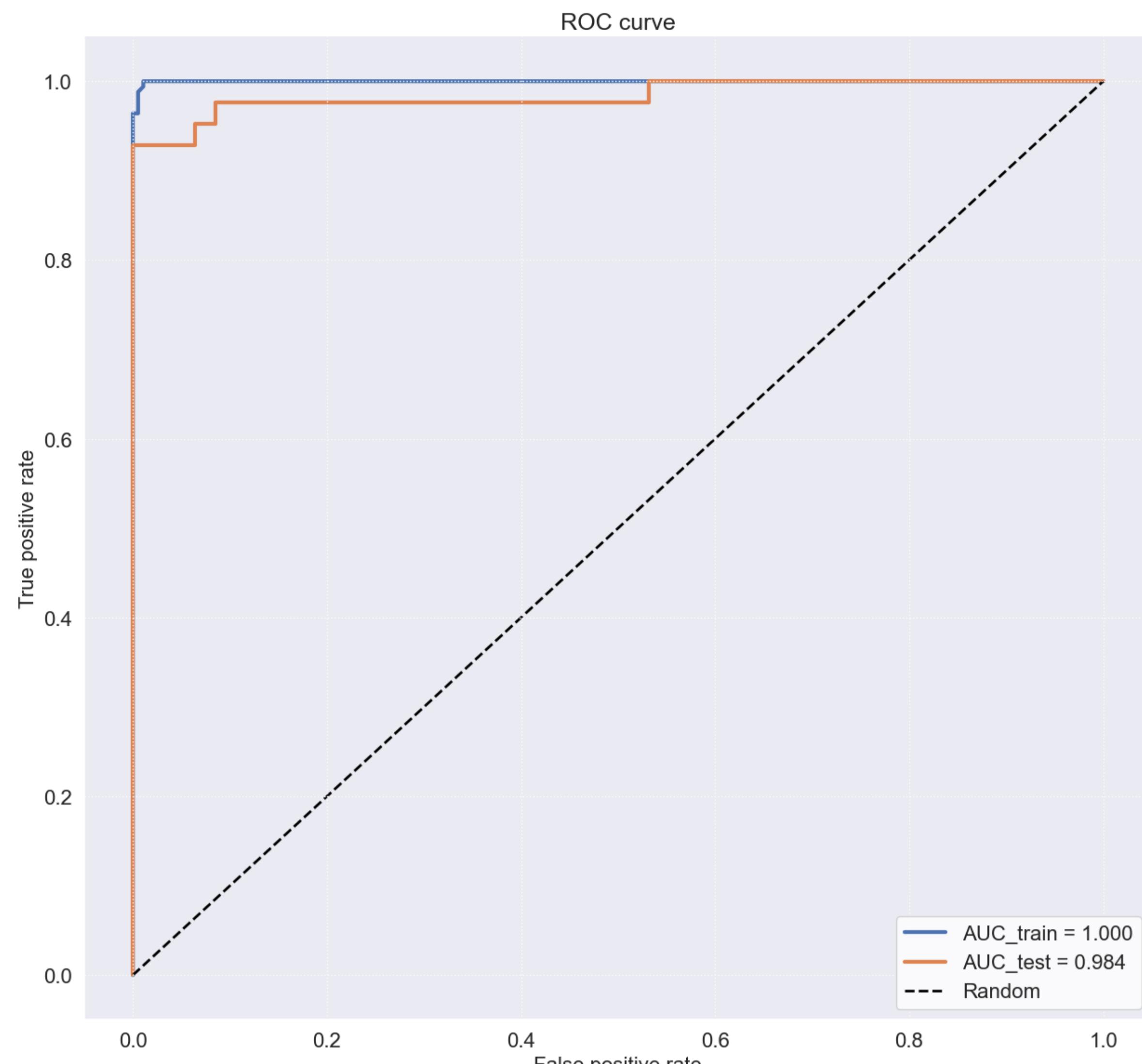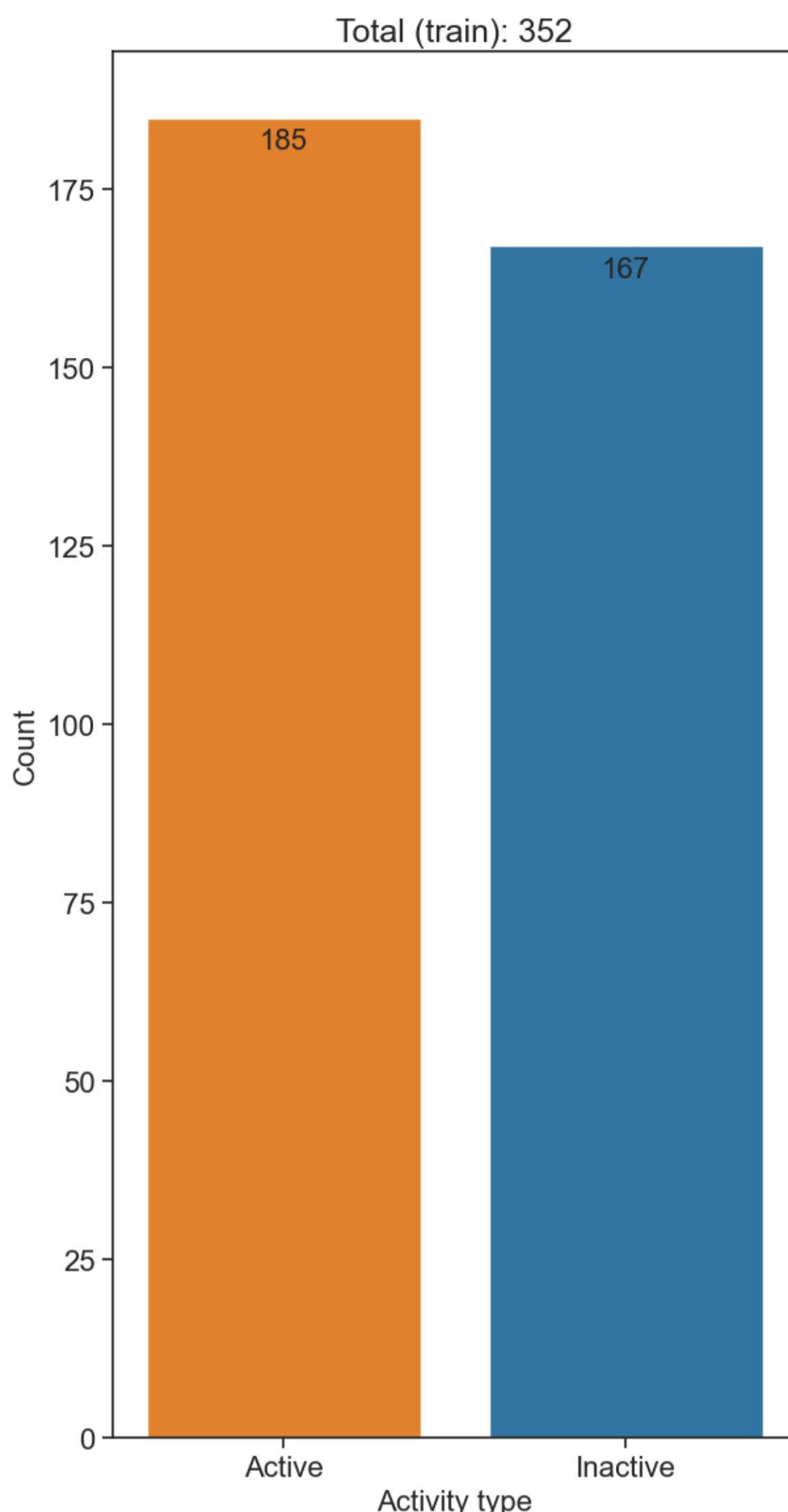
Metric	Train	Test
AUC	0.9998	0.9838
accuracy	0.9801	0.9326
balanced accuracy	0.9811	0.9336
recall_(sens)	1.0	0.9524
specificity	0.9622	0.9149
precision	0.9598	0.9091
f1_score	0.9801	0.9326