Protein: P30968 - metric: balanced_accuracy

Total (train): 869

ROC curve

- AUC_train = 0.999
- AUC_test = 0.944
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 1871)
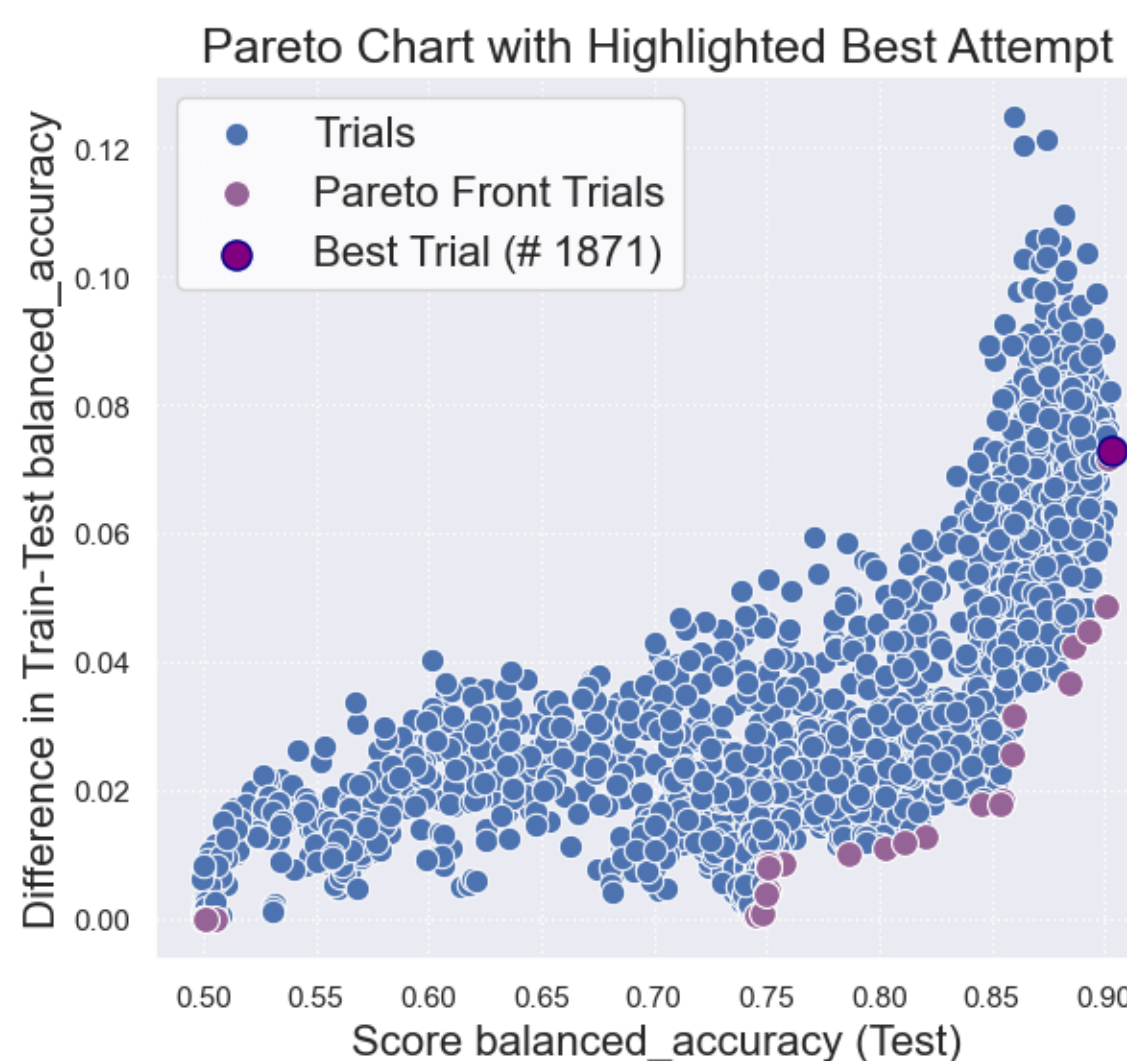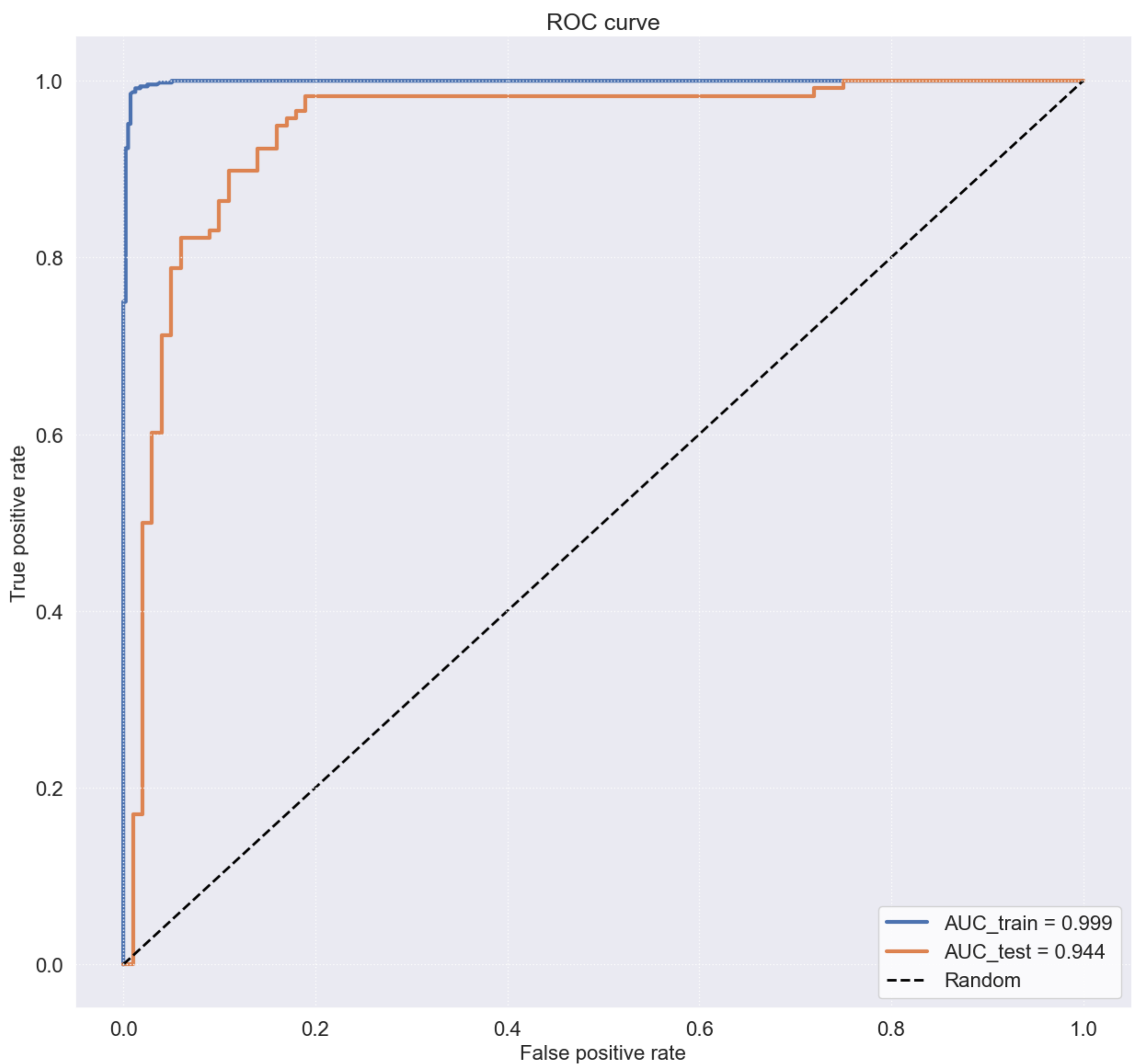
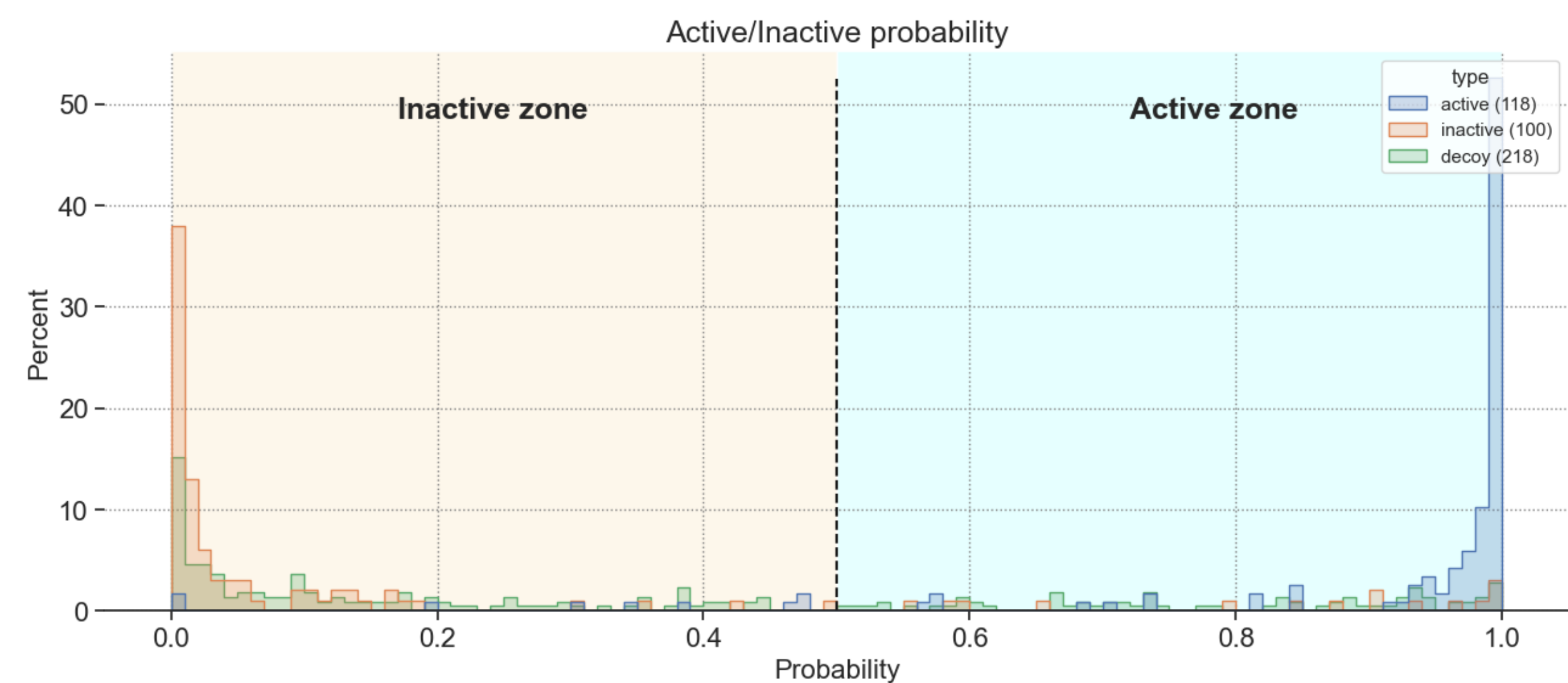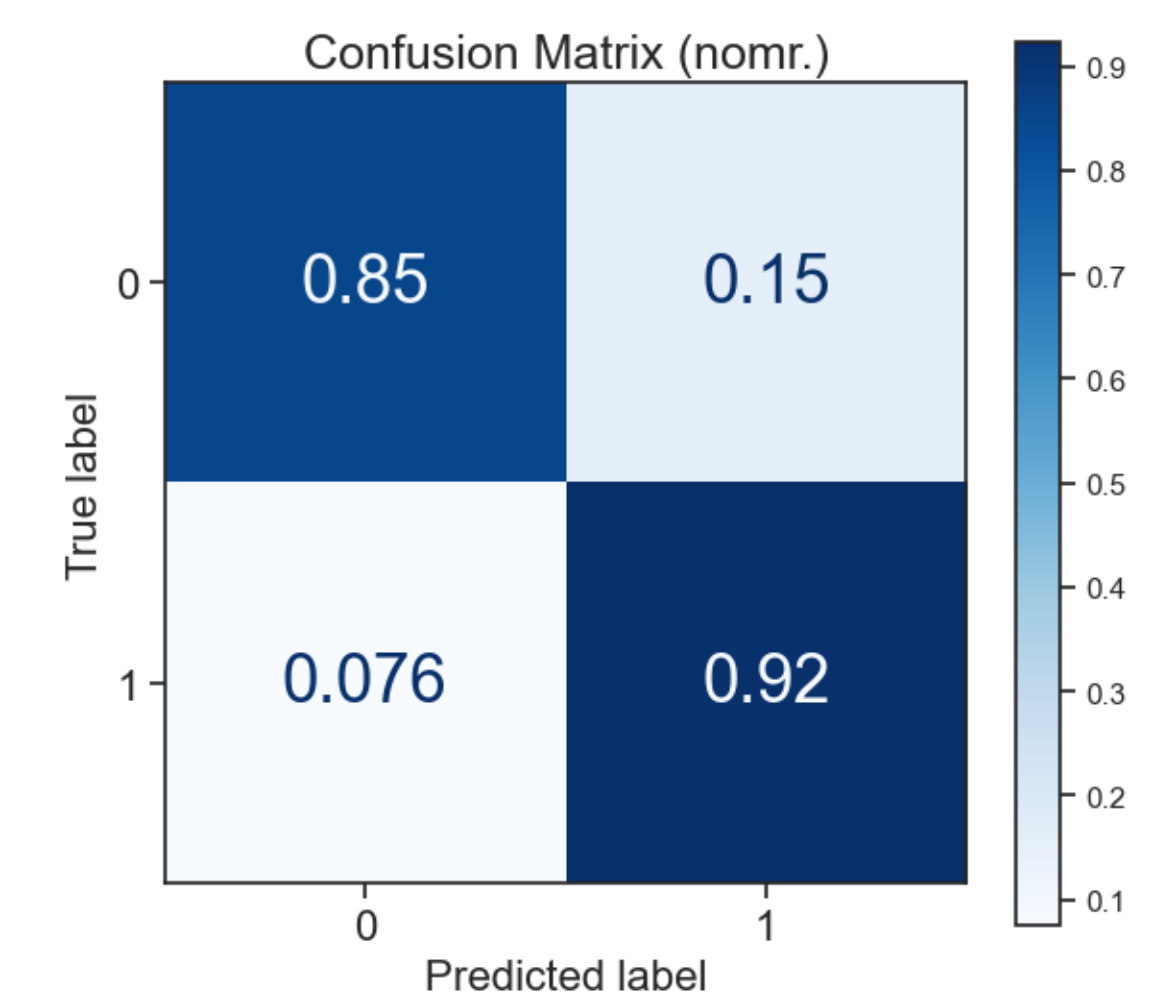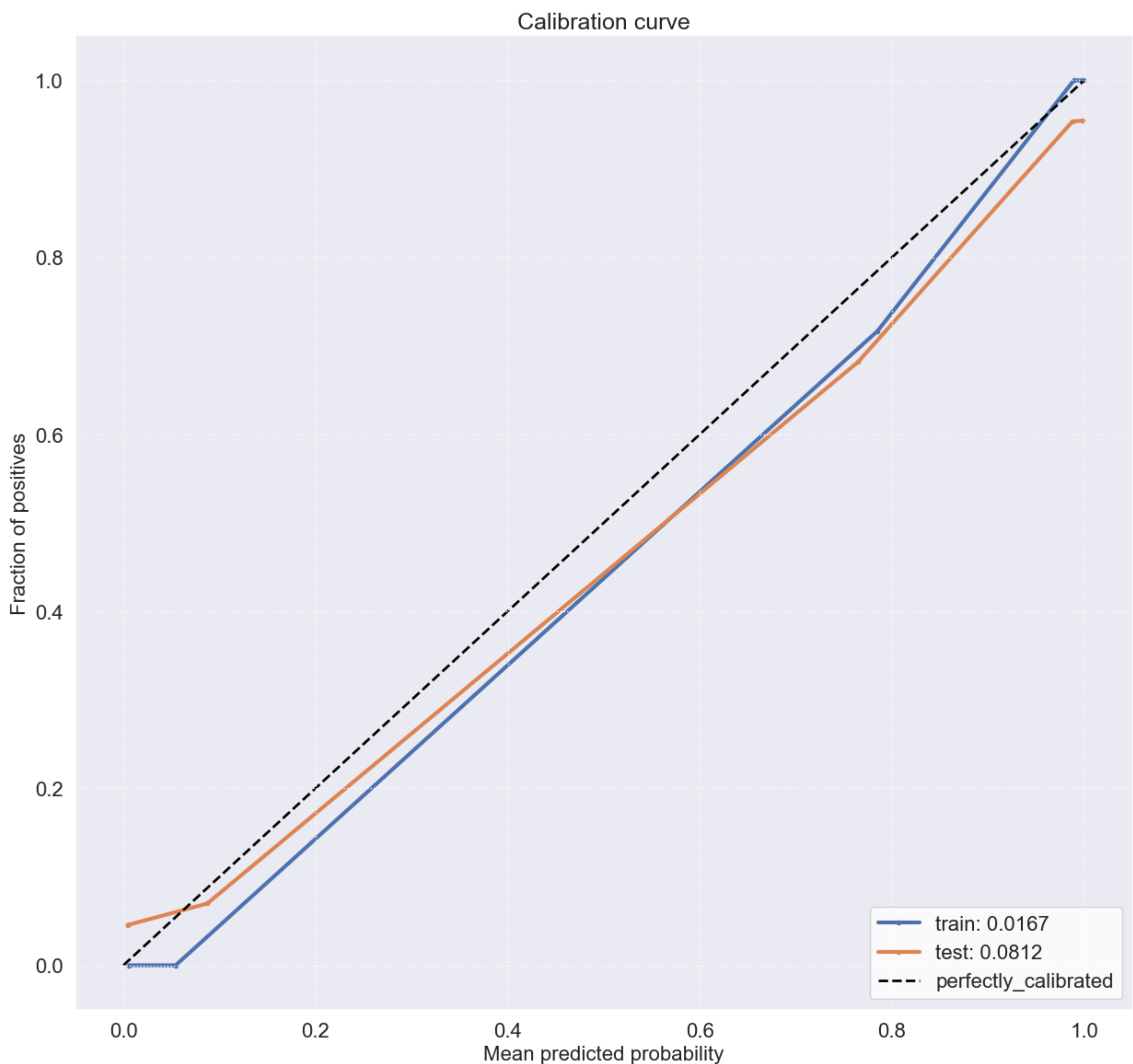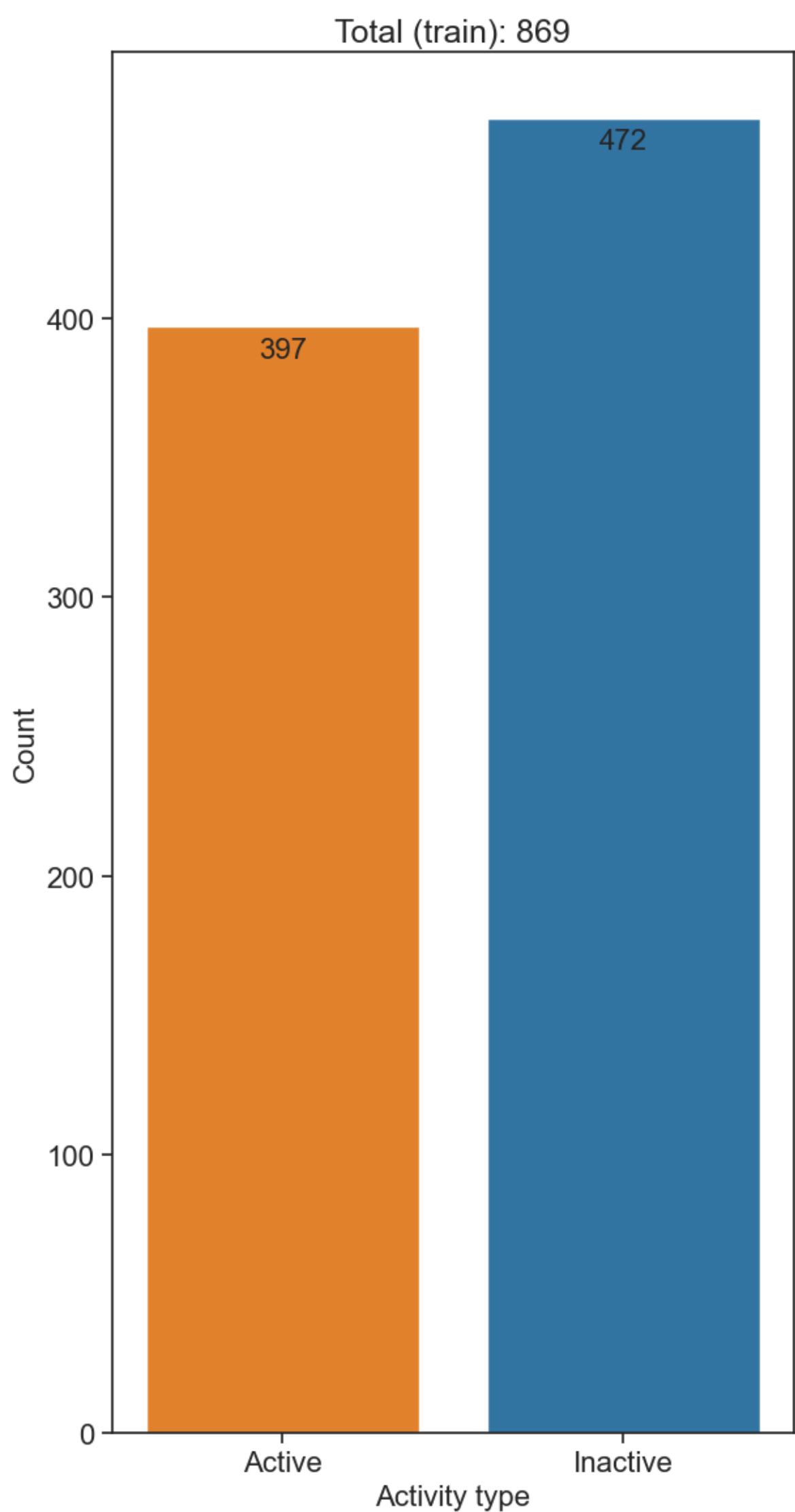| XGBoost Parameters	Value
alpha	2.0184
colsample_bytree	0.6806
gamma	0.4333
lambda	1.265
learning_rate	0.1182
max_depth	10.0
min_child_weight	4.0
n_estimators	754.0
reg_alpha	0.1122
reg_lambda	0.9539
scale_pos_weight	2.8235
subsample	0.6463

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

- train: 0.0167
- test: 0.0812
- perfectly_calibrated

Active/Inactive probability

- active (118)
- inactive (100)
- decoy (218)

Inactive zone

Active zone
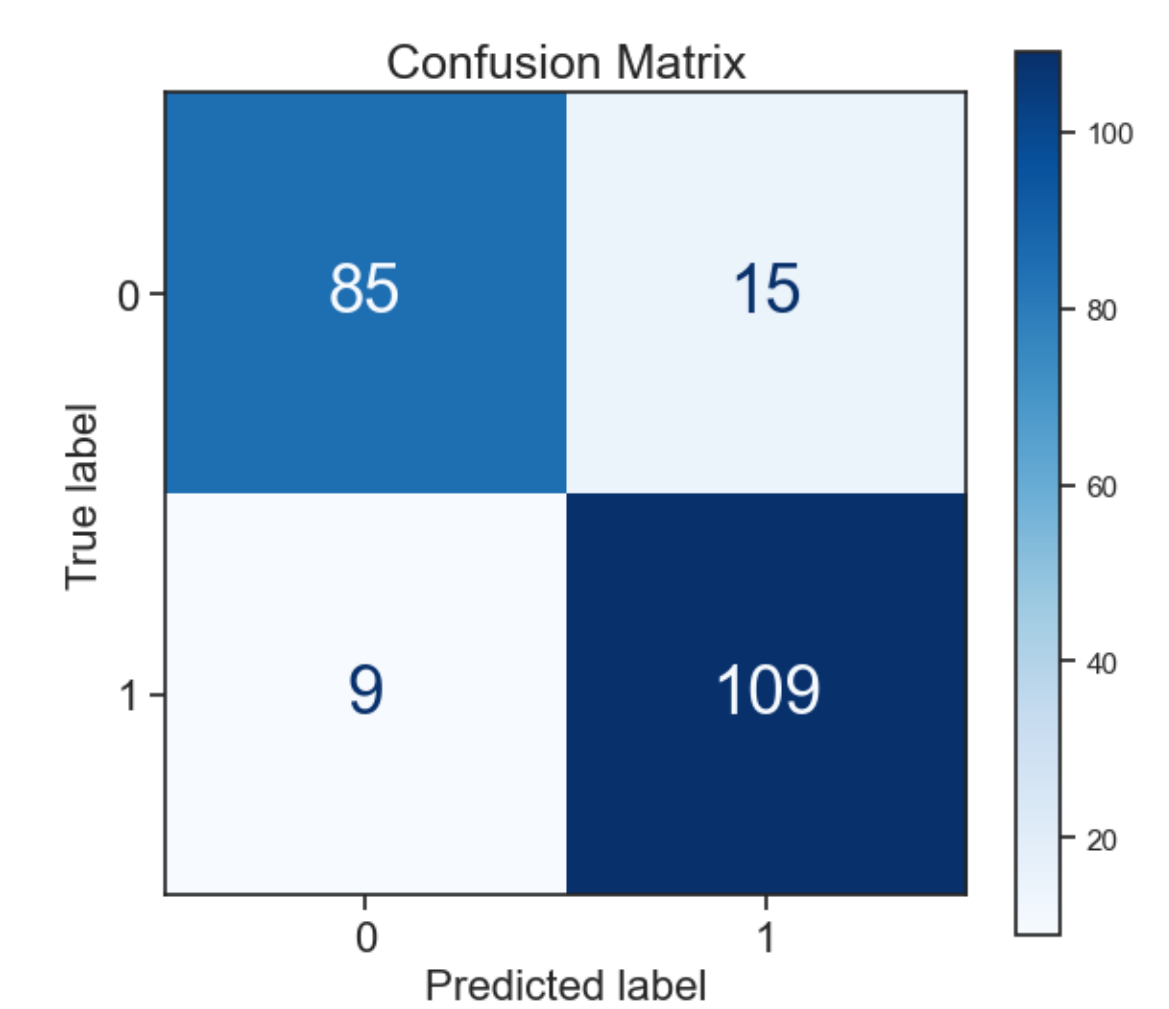
Metric	Train	Test
AUC	0.9988	0.9442
accuracy	0.9804	0.8899
balanced accuracy	0.9788	0.8869
recall_(sens)	0.9979	0.9237
specificity	0.9597	0.85
precision	0.9671	0.879
f1_score	0.9804	0.8896

Protein: P30968 - metric: f1

Total (train): 869

ROC curve

- AUC_train = 0.995
- AUC_test = 0.944
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 1687)

Confusion Matrix

	0	1
0	88	12
1	10	108

Confusion Matrix (nomr.)

	0	1
0	0.88	0.12
1	0.085	0.92

Calibration curve

- train: 0.0275
- test: 0.0836
- perfectly_calibrated
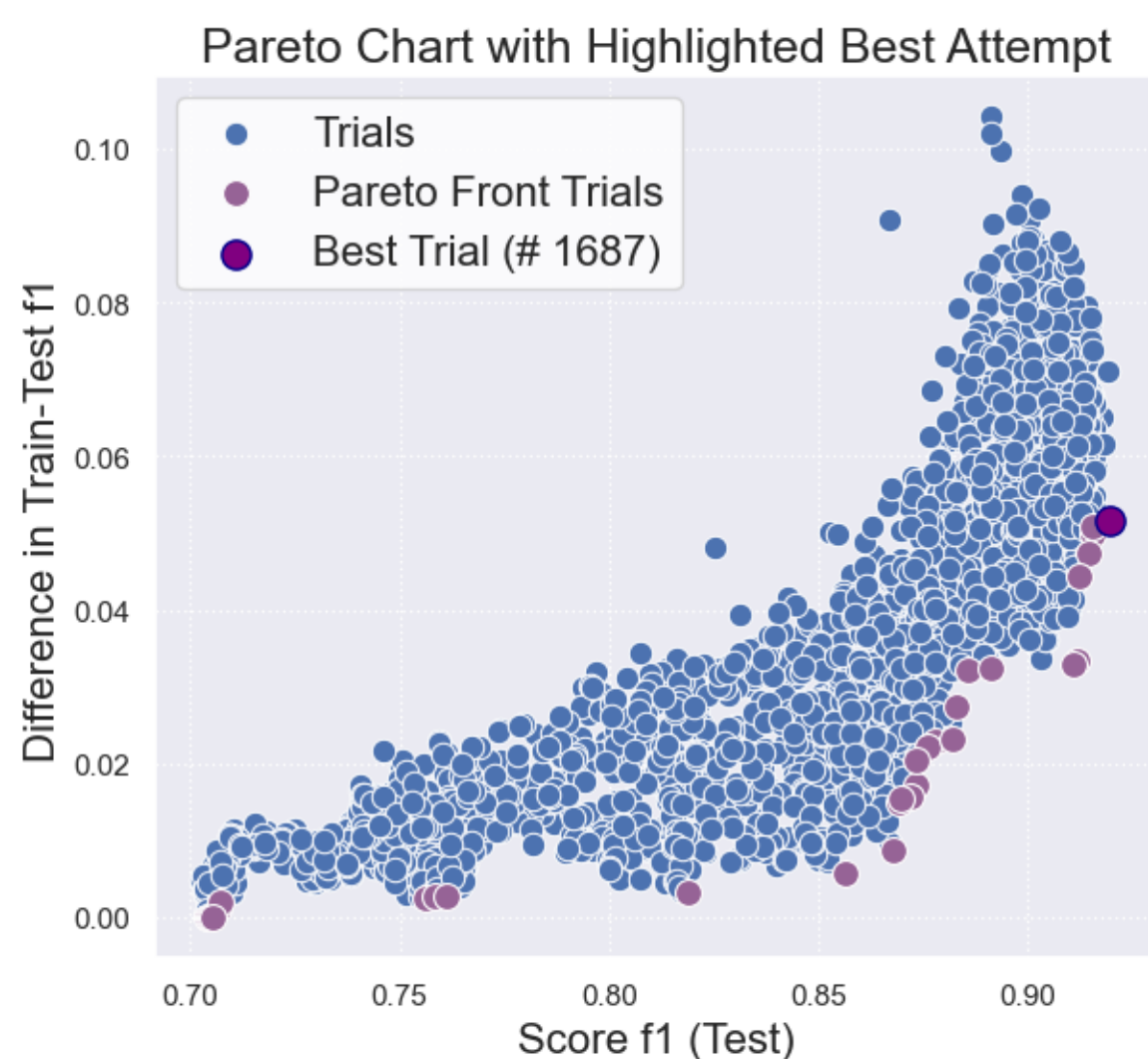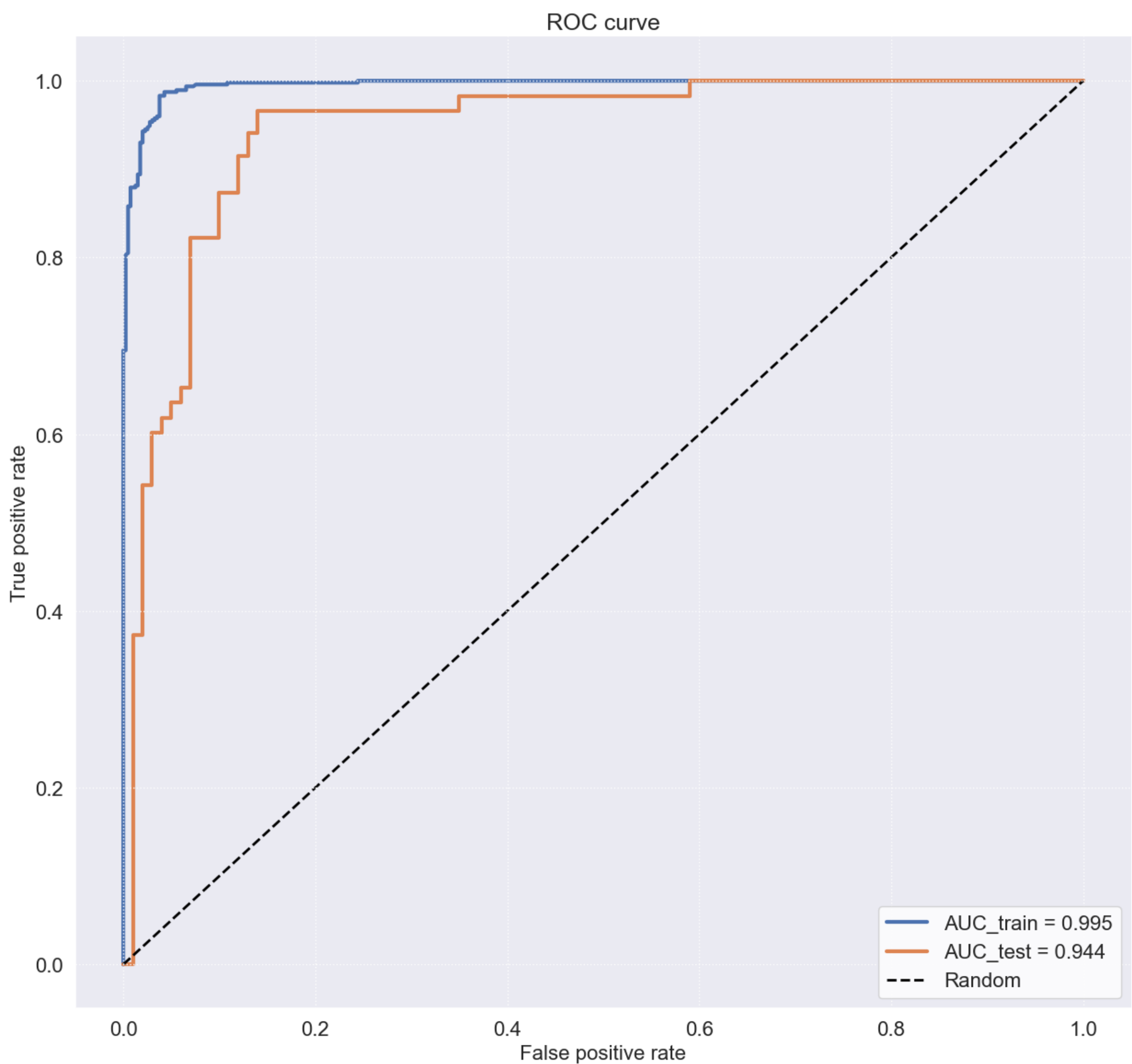
XGBoost Parameters	Value
alpha	1.1175
colsample_bytree	0.9541
gamma	0.6335
lambda	3.3643
learning_rate	0.2813
max_depth	5.0
min_child_weight	2.0
n_estimators	524.0
reg_alpha	0.623
reg_lambda	0.7736
scale_pos_weight	1.1844
subsample	0.8865

Active/Inactive probability

type
- active (118)
- inactive (100)
- decoy (218)
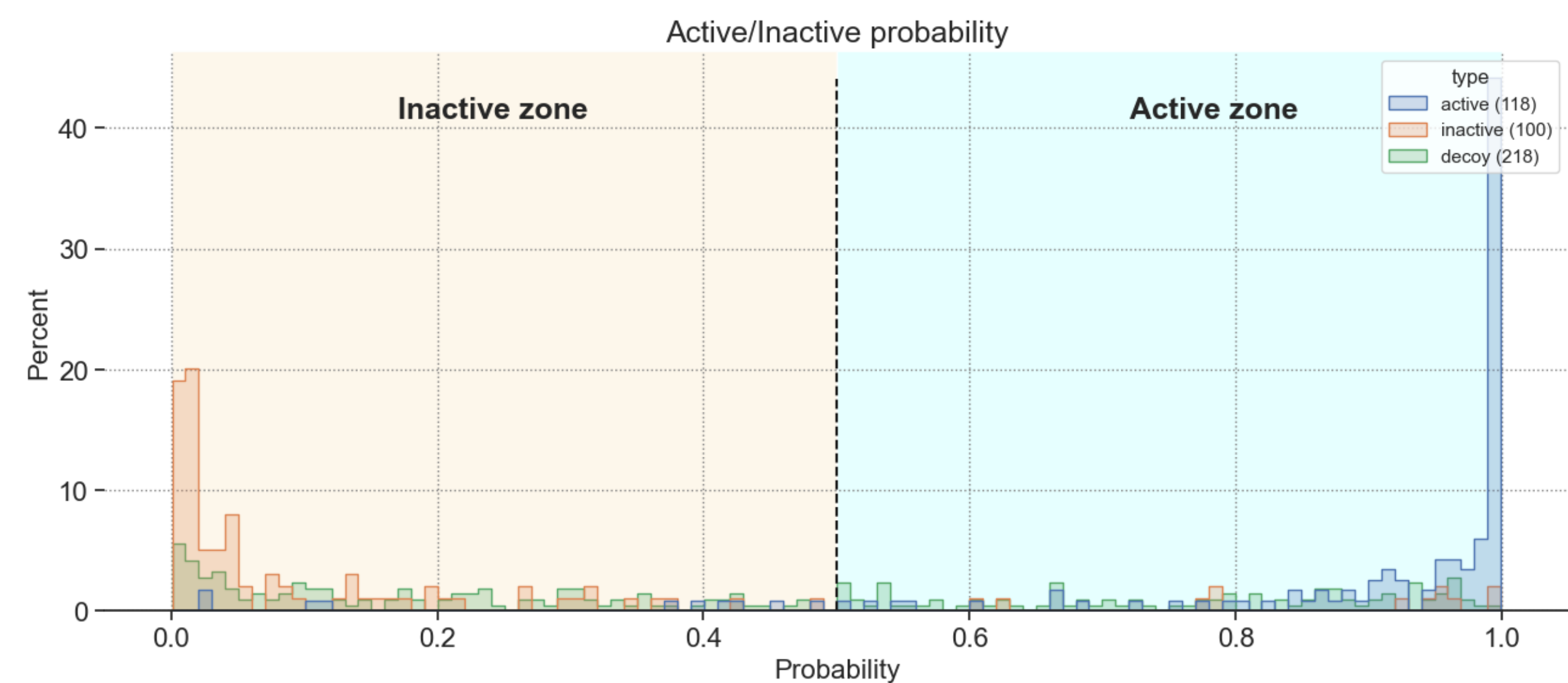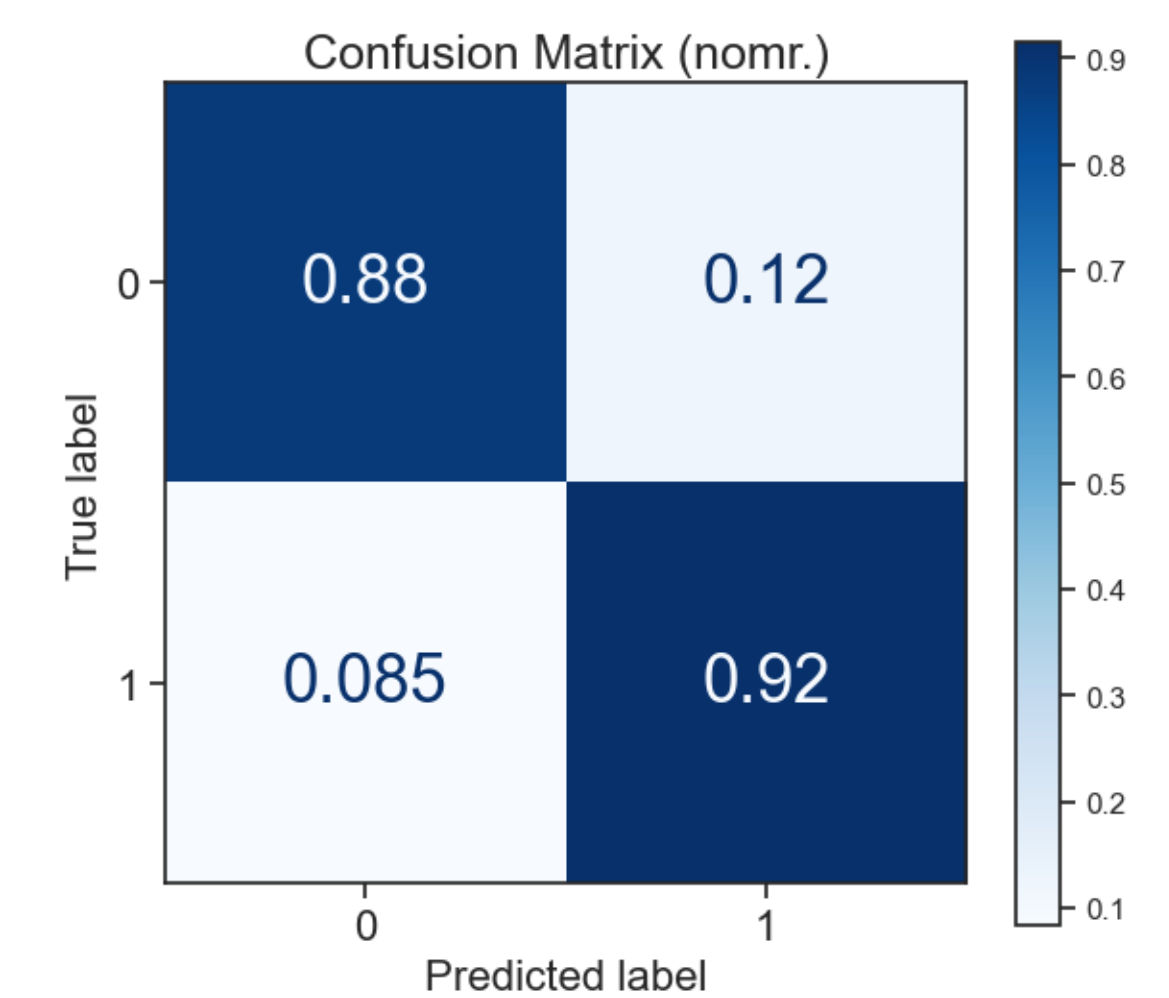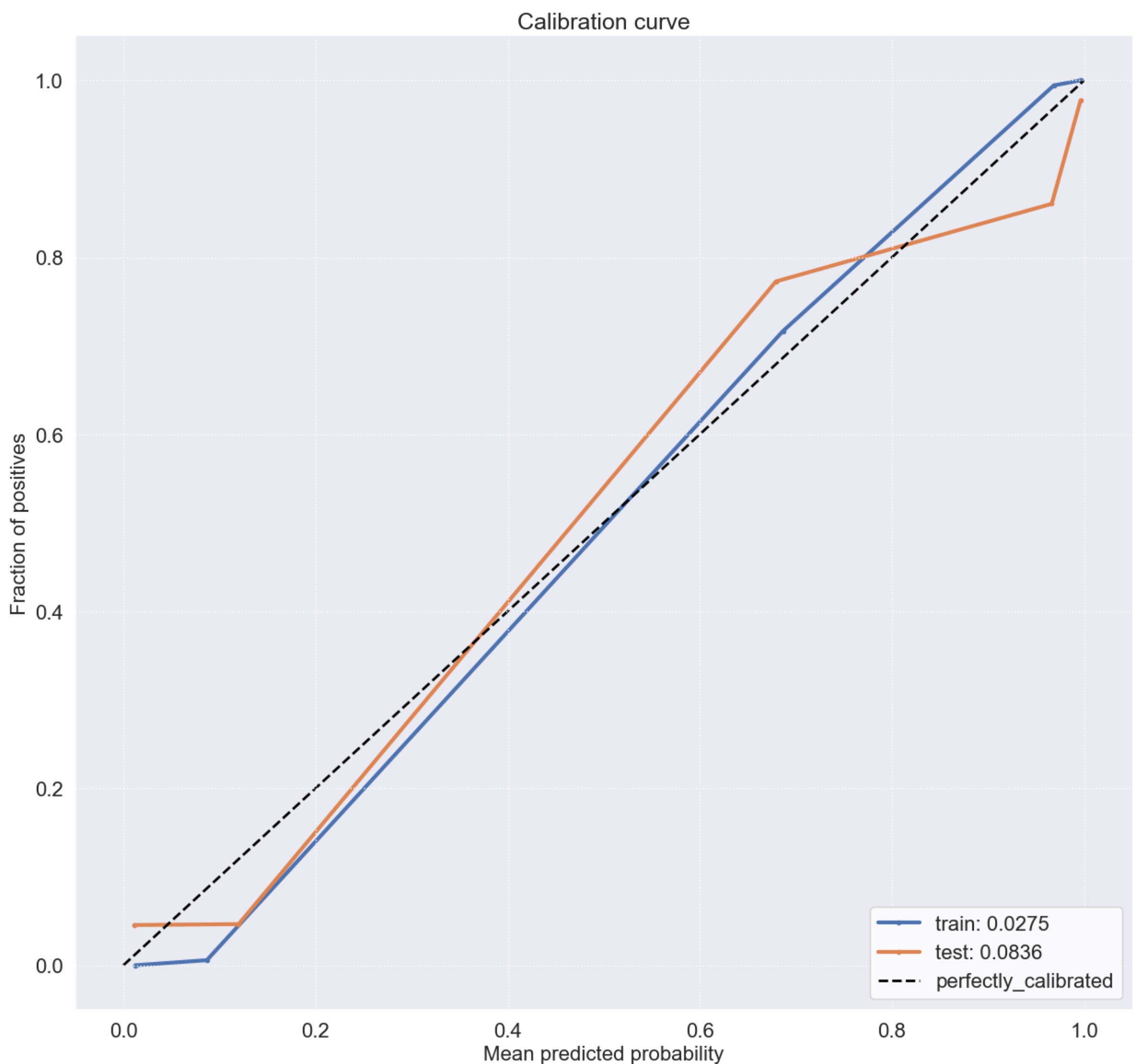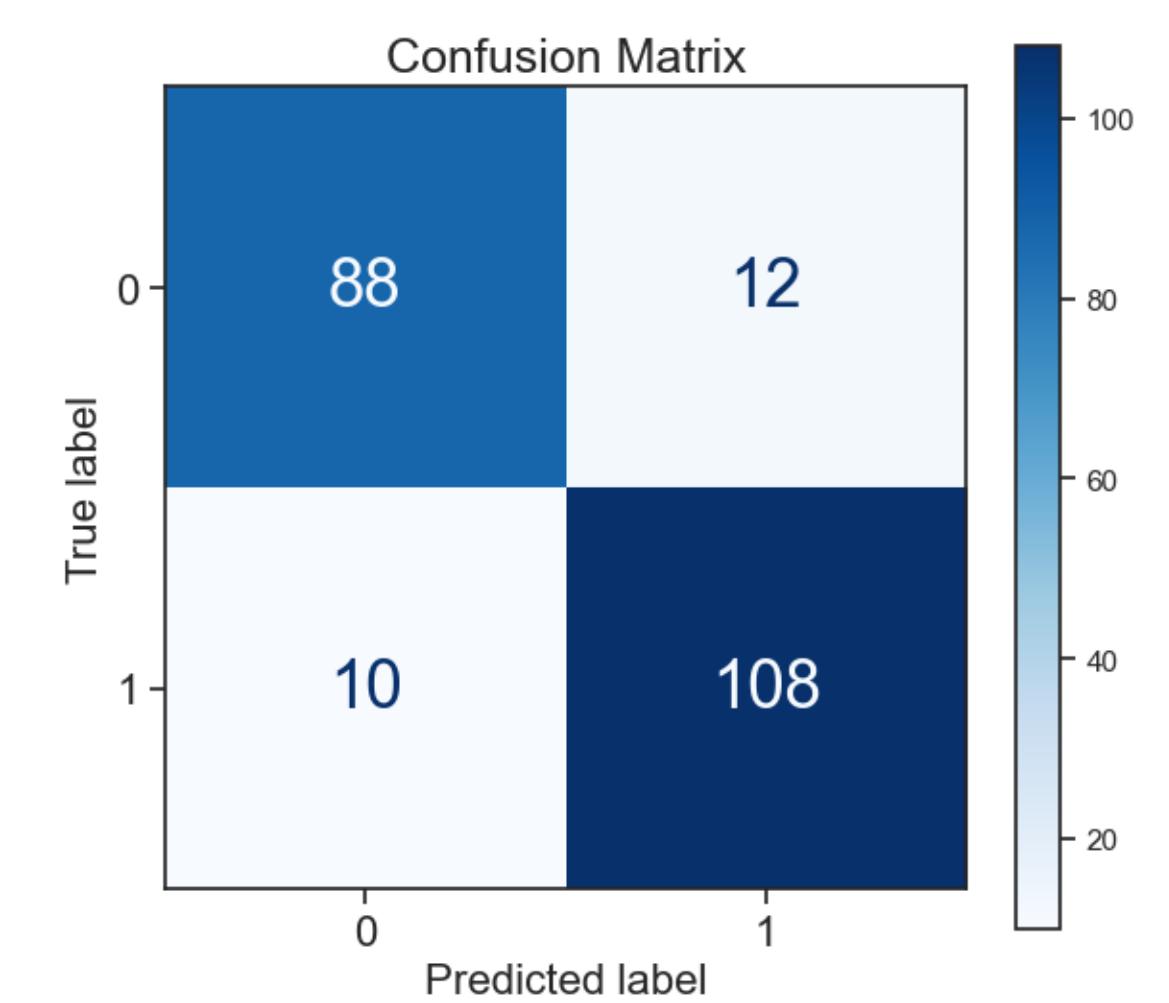
Metric	Train	Test
AUC	0.9953	0.9437
accuracy	0.9701	0.8991
balanced accuracy	0.9695	0.8976
recall_(sens)	0.9767	0.9153
specificity	0.9622	0.88
precision	0.9685	0.9
f1_score	0.9701	0.899

Protein: P30968 - metric: precision

Total (train): 869

Active: 397
Inactive: 472

ROC curve

- AUC_train = 0.999
- AUC_test = 0.943
- Random

X-axis: False positive rate
Y-axis: True positive rate

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 1600)

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

Confusion Matrix

	Predicted 0	Predicted 1
True 0	87	13
True 1	9	109

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.87	0.13
True 1	0.076	0.92

Calibration curve

- train: 0.0141
- test: 0.0803
- perfectly_calibrated

X-axis: Mean predicted probability
Y-axis: Fraction of positives

Active/Inactive probability

Inactive zone | Active zone

type:
- active (118)
- inactive (100)
- decoy (218)

X-axis: Probability
Y-axis: Percent

XGBoost Parameters	Value
alpha	3.7643
colsample_bytree	0.7896
gamma	0.0411
lambda	3.531
learning_rate	0.3124
max_depth	3.0
min_child_weight	3.0
n_estimators	544.0
reg_alpha	0.5778
reg_lambda	0.9007
scale_pos_weight	1.7727
subsample	0.5807
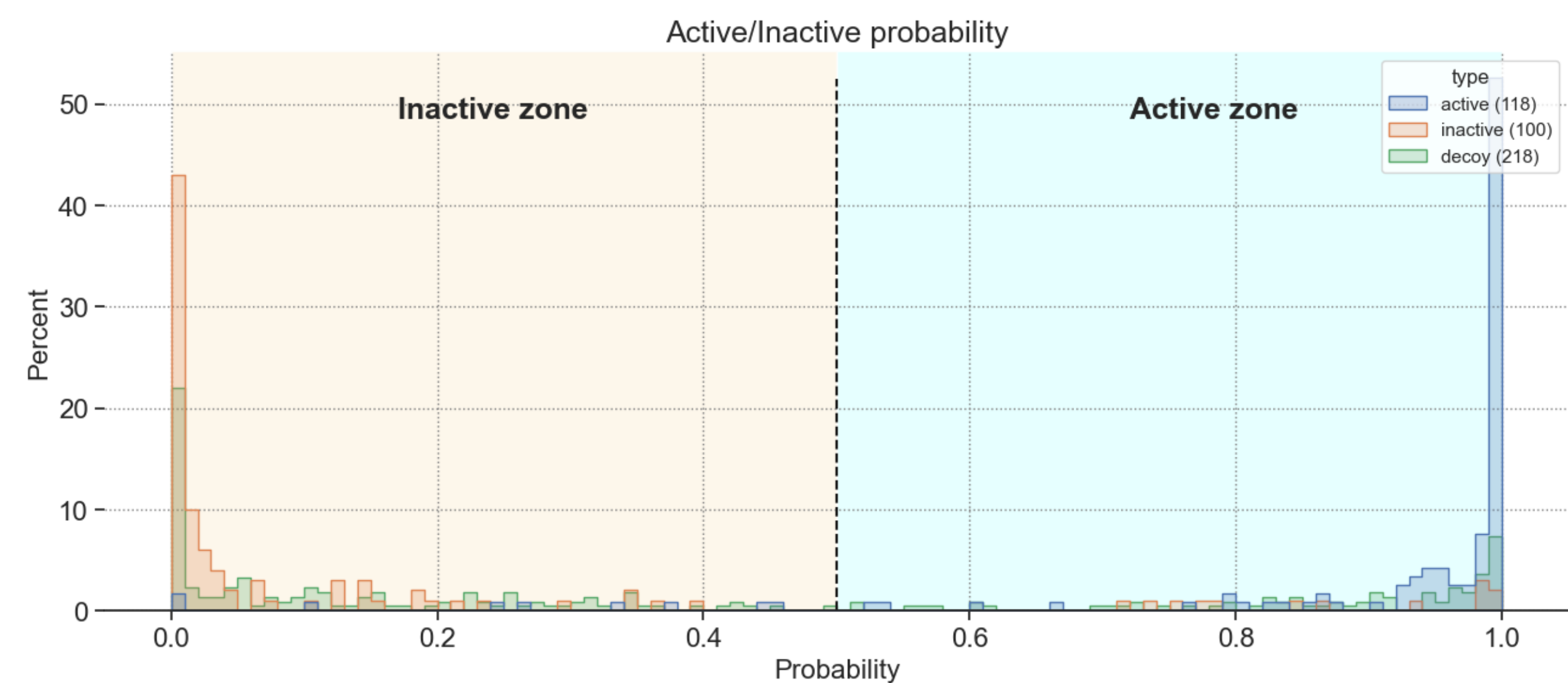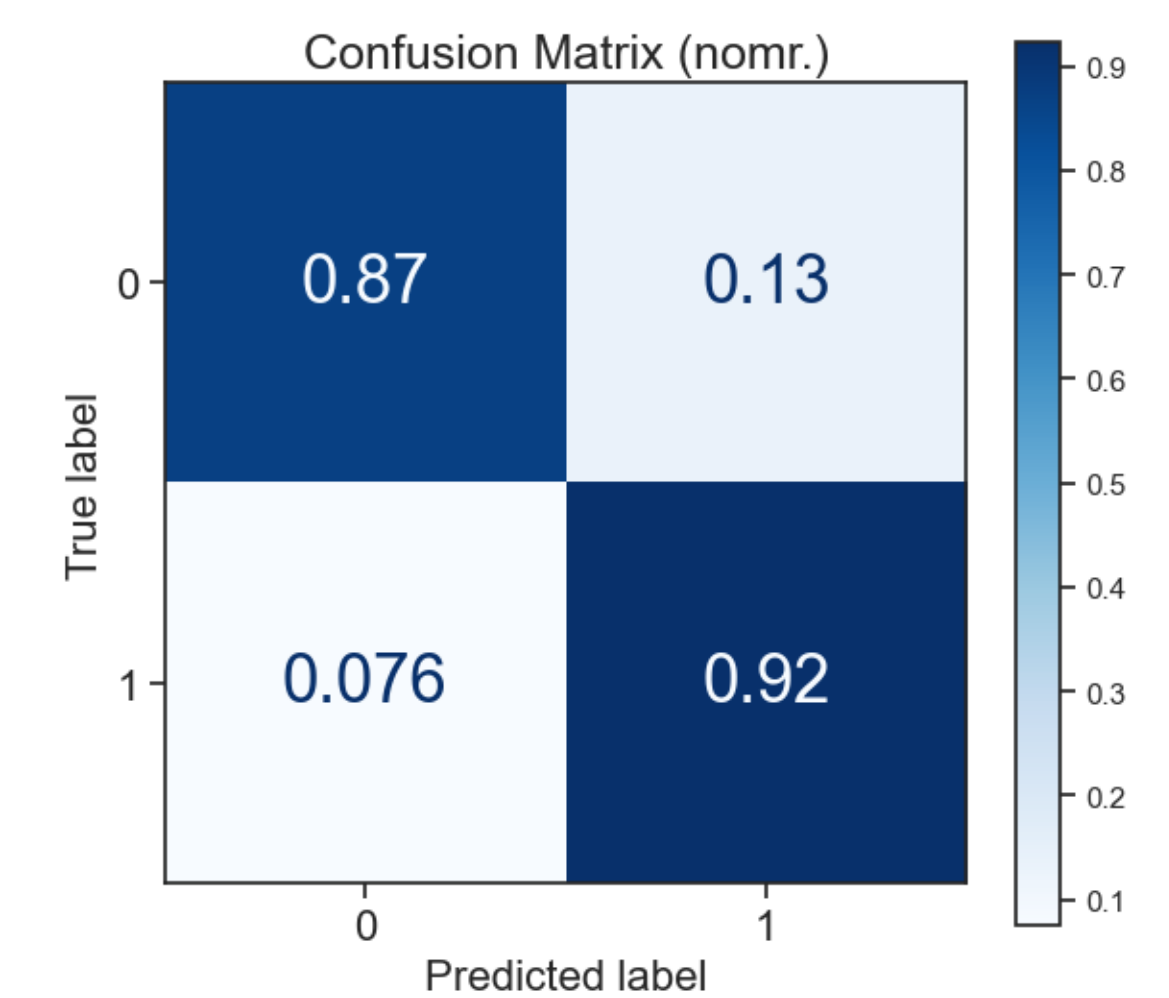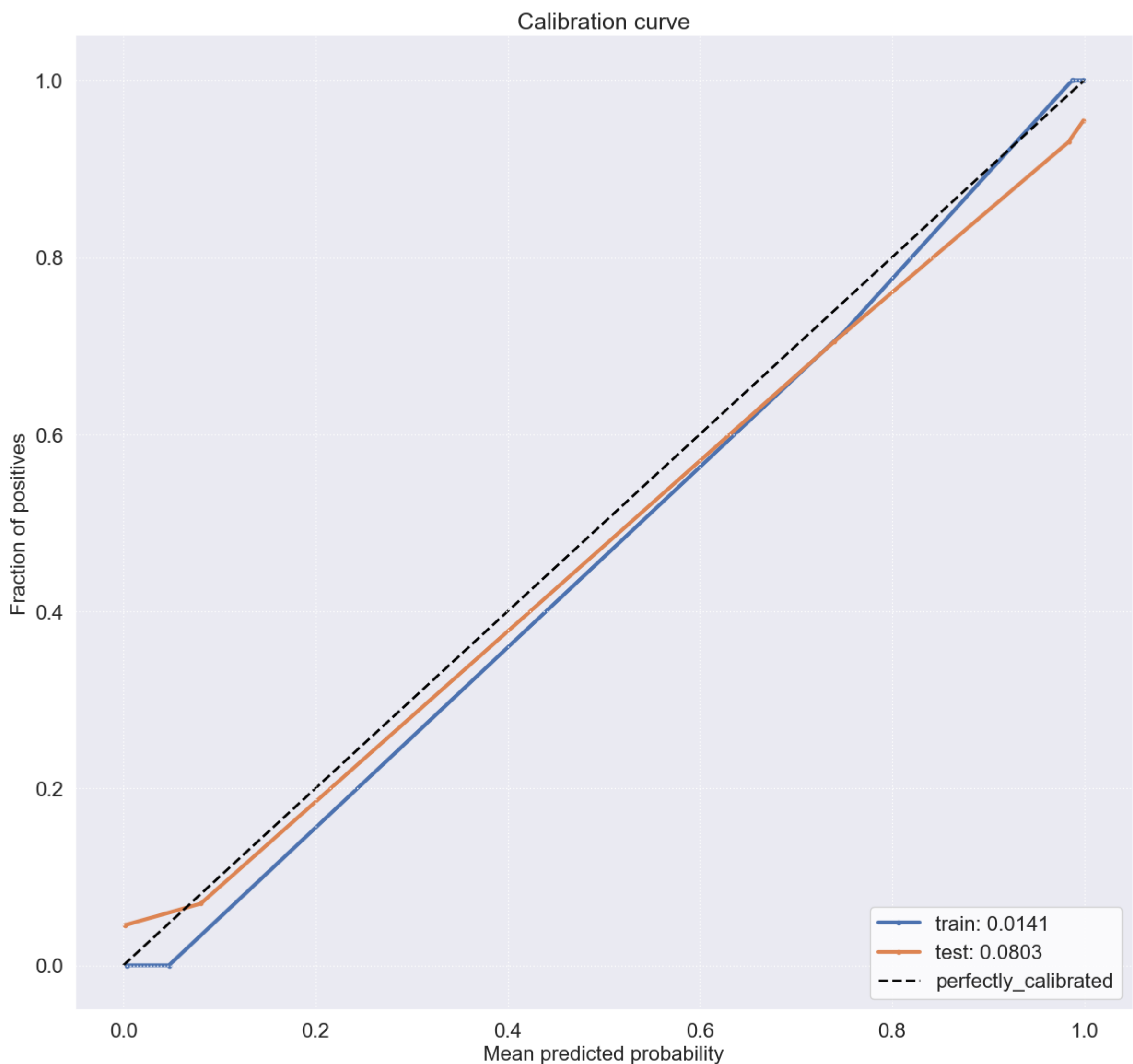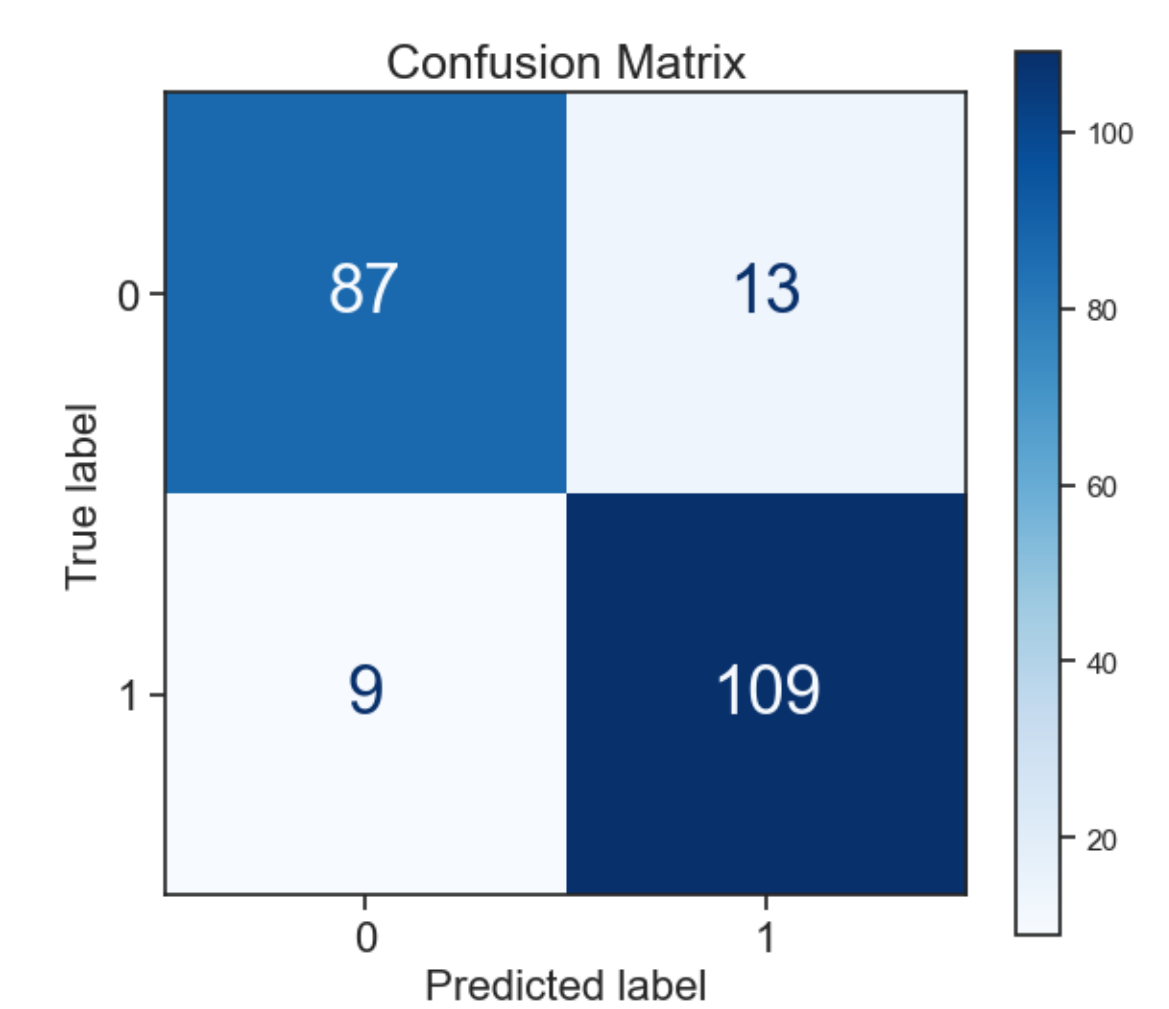
Metric	Train	Test
AUC	0.9992	0.9431
accuracy	0.9839	0.8991
balanced accuracy	0.9826	0.8969
recall_(sens)	0.9979	0.9237
specificity	0.9673	0.87
precision	0.9731	0.8934
f1_score	0.9839	0.8989

Protein: P30968 - metric: SoftVotingXGBoostEnsemble
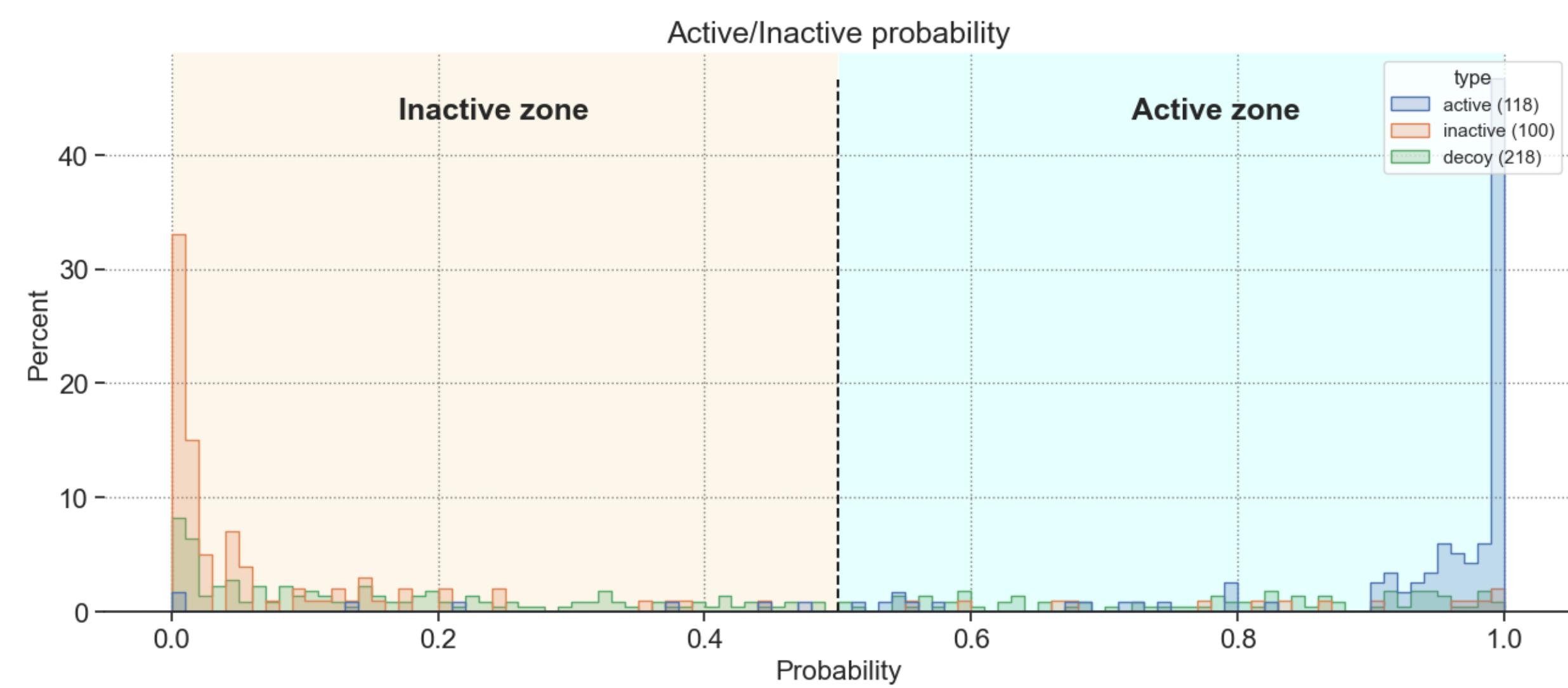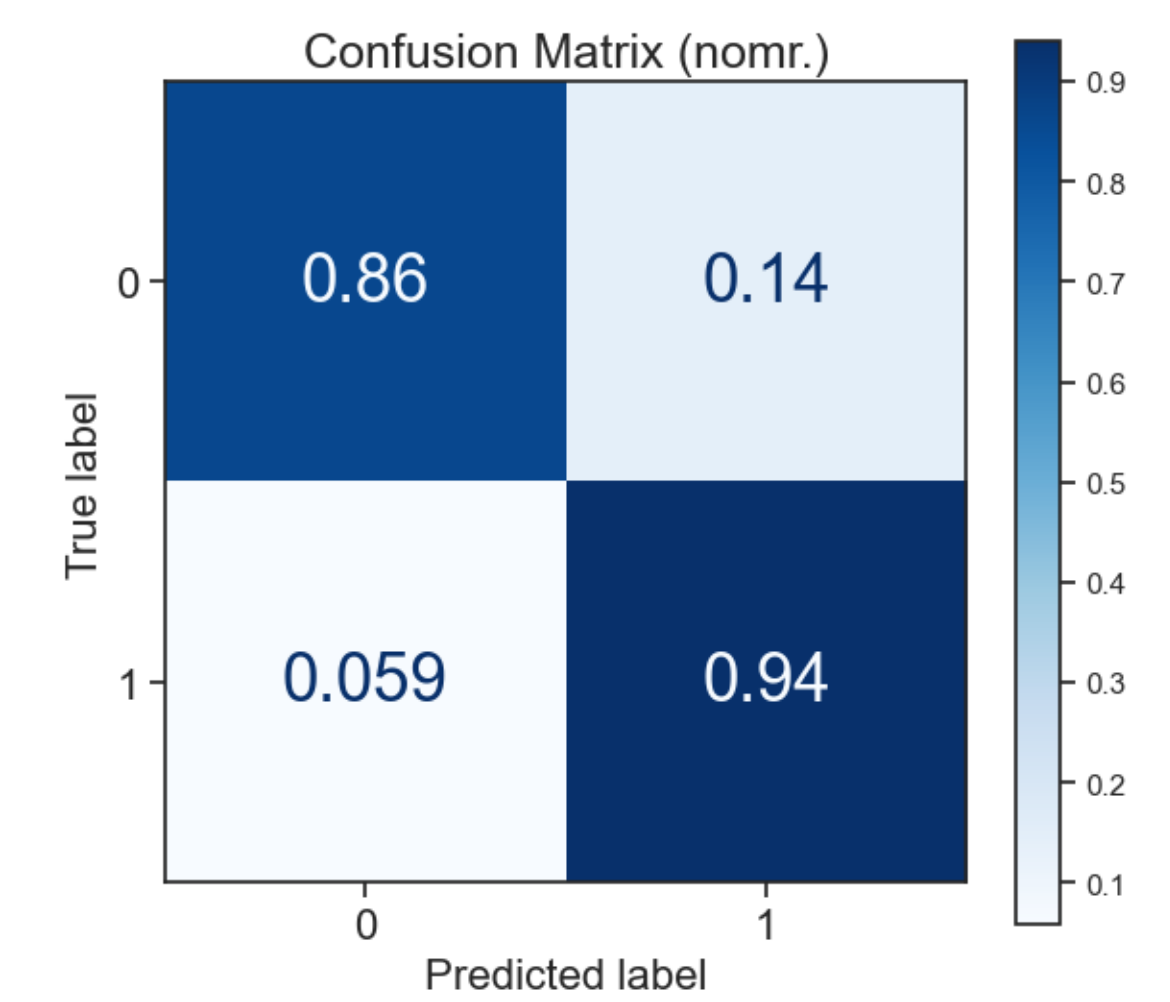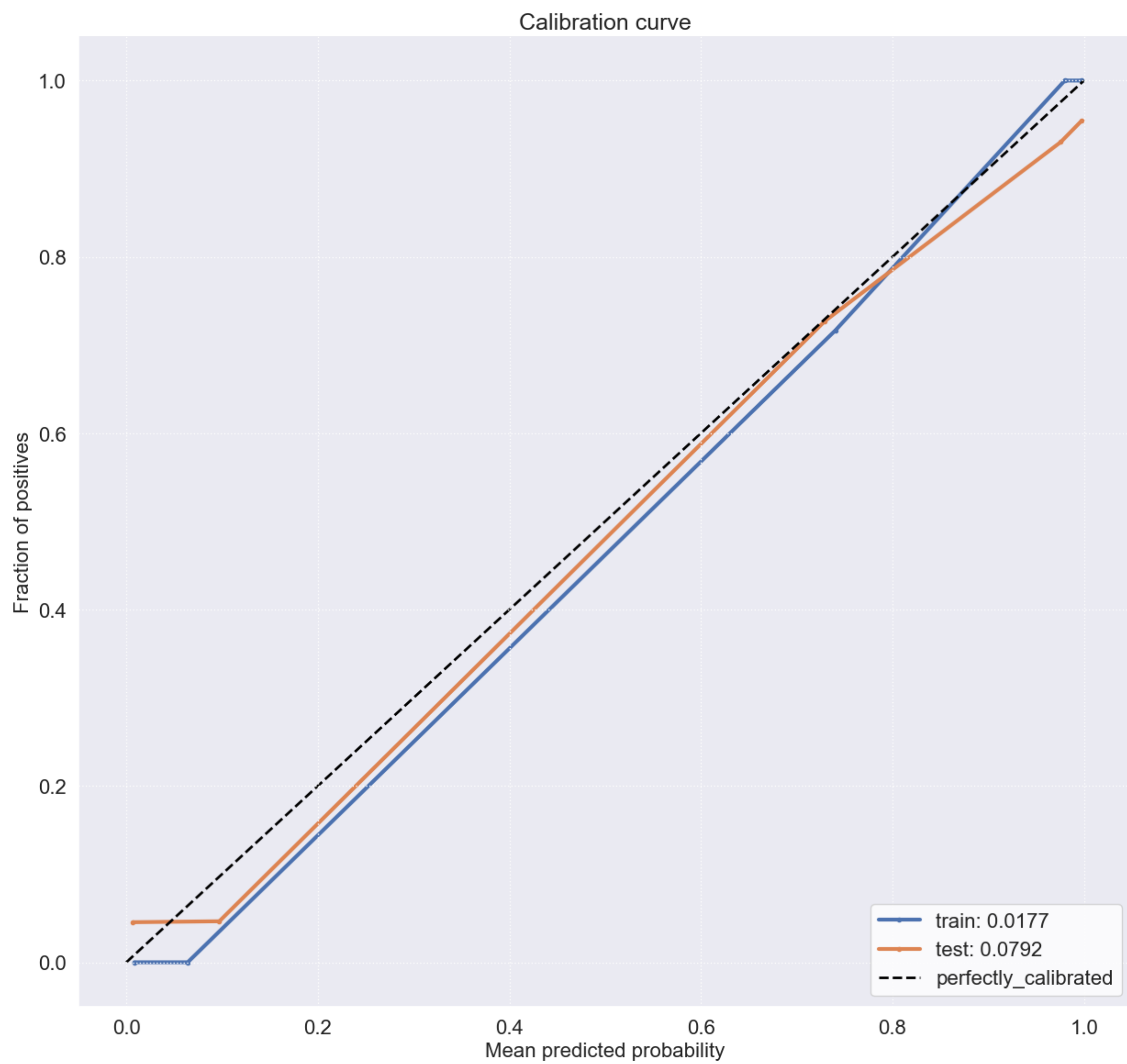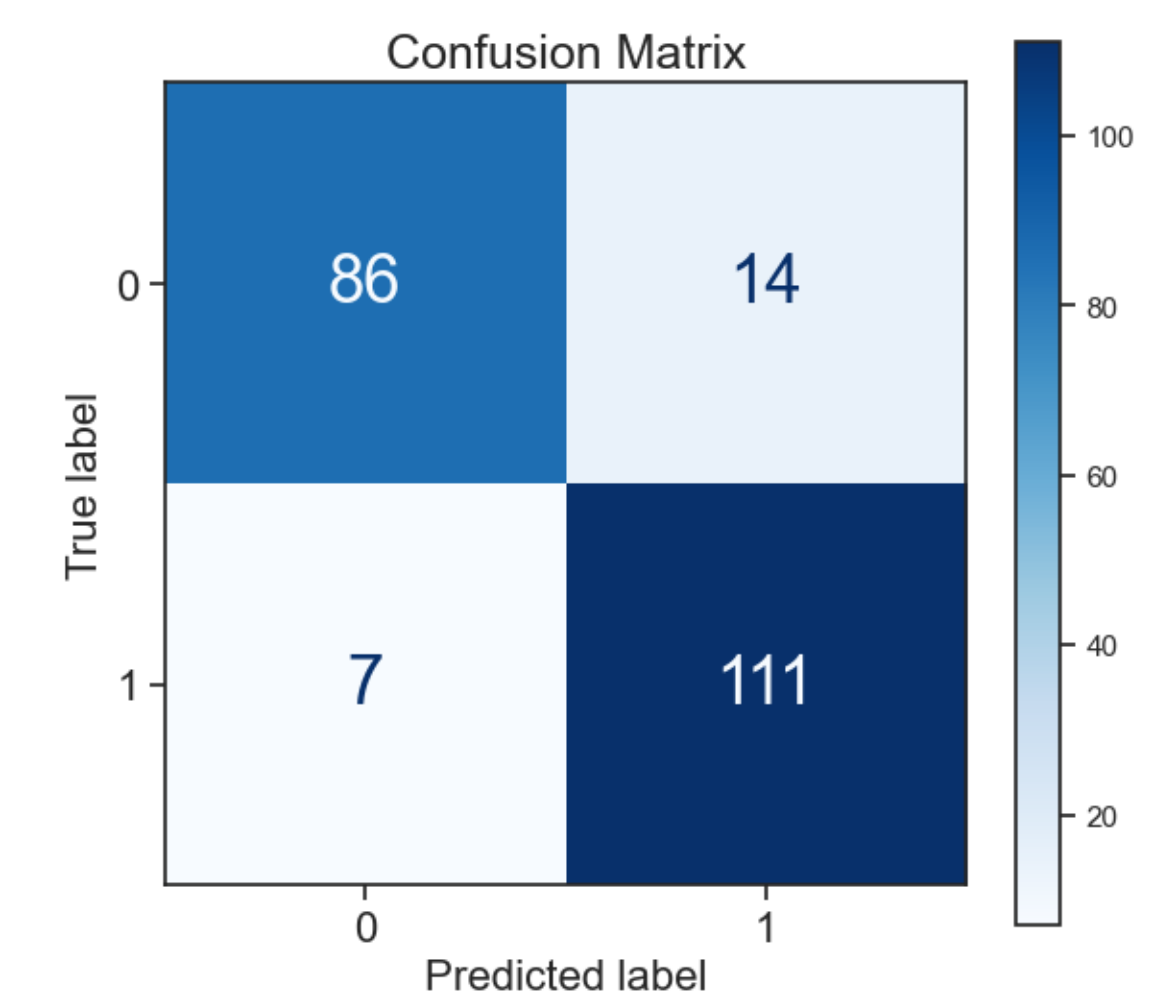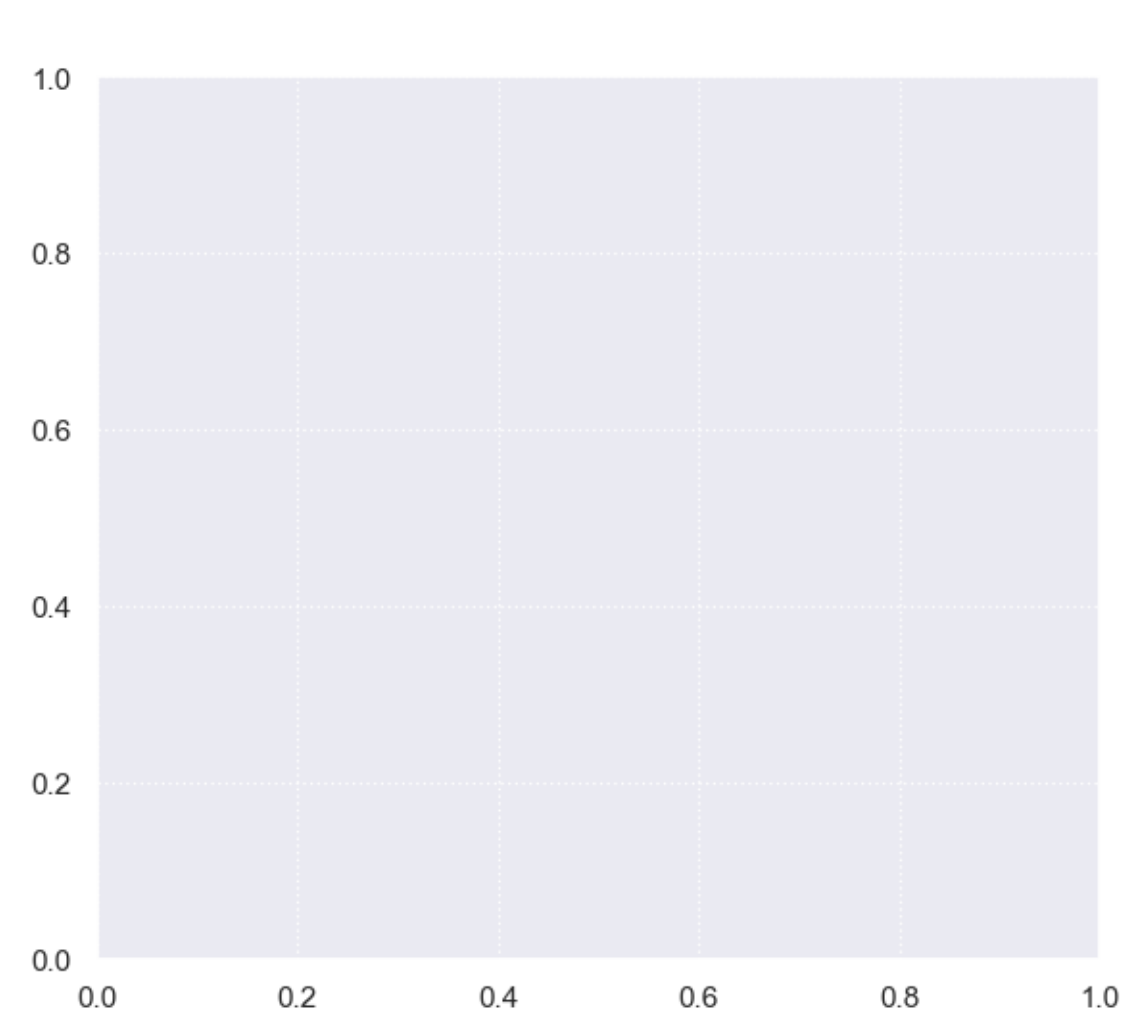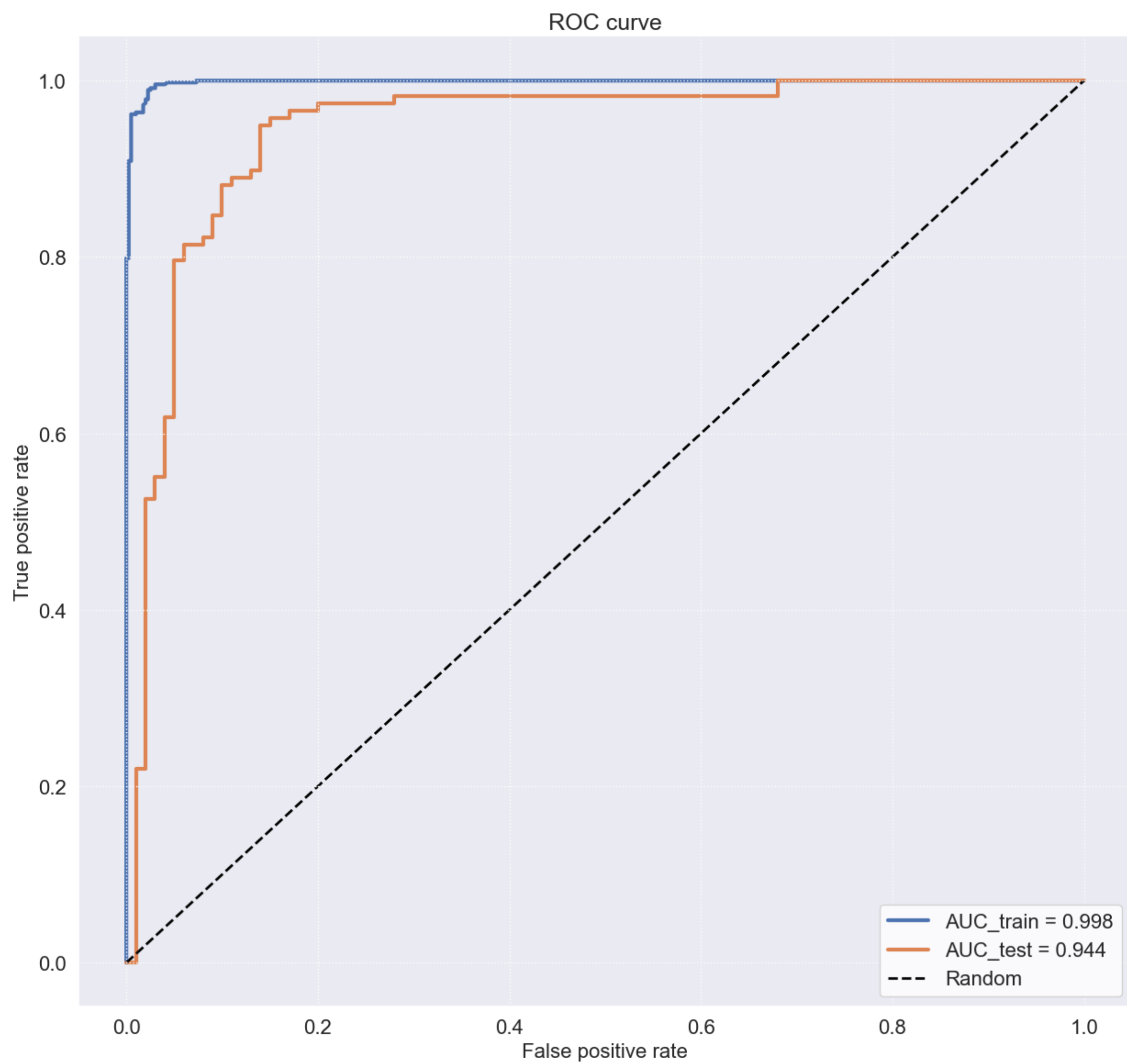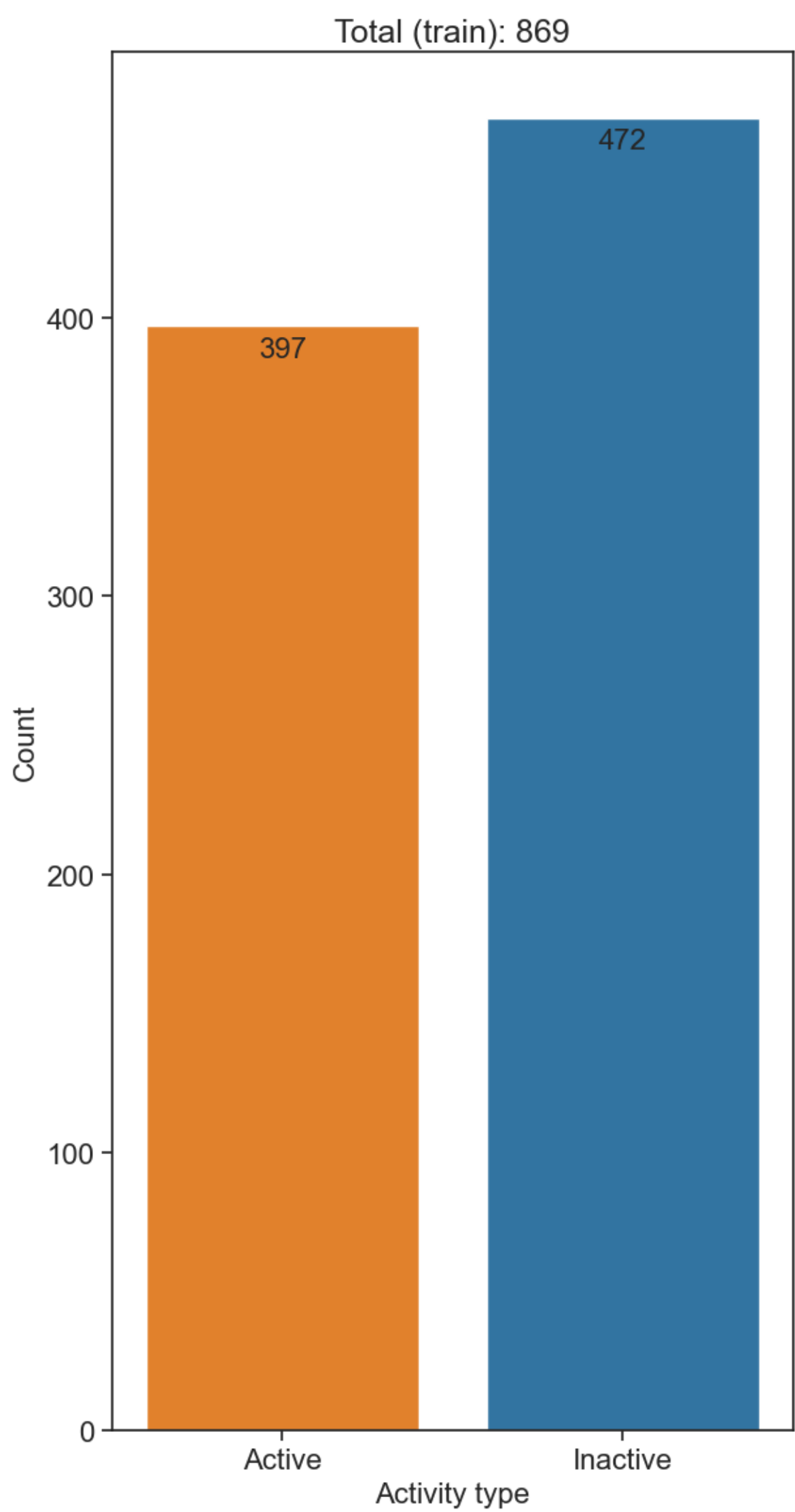
Metric	Train	Test
AUC	0.9985	0.9445
accuracy	0.9816	0.9037
balanced accuracy	0.9802	0.9003
recall_(sens)	0.9958	0.9407
specificity	0.9647	0.86
precision	0.9711	0.888
f1_score	0.9816	0.9033