Protein: P08588 - metric: balanced_accuracy

Total (train): 801

ROC curve

- AUC_train = 1.000
- AUC_test = 0.955
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
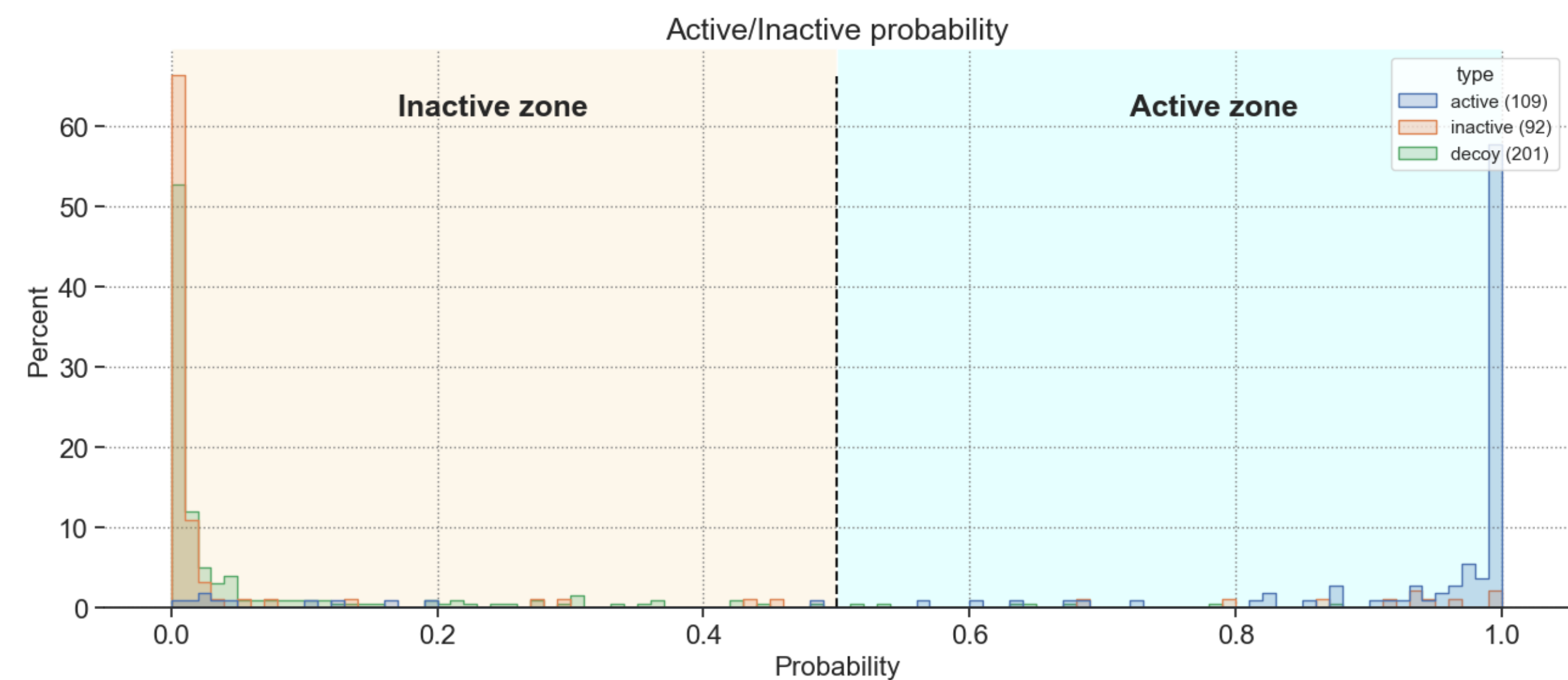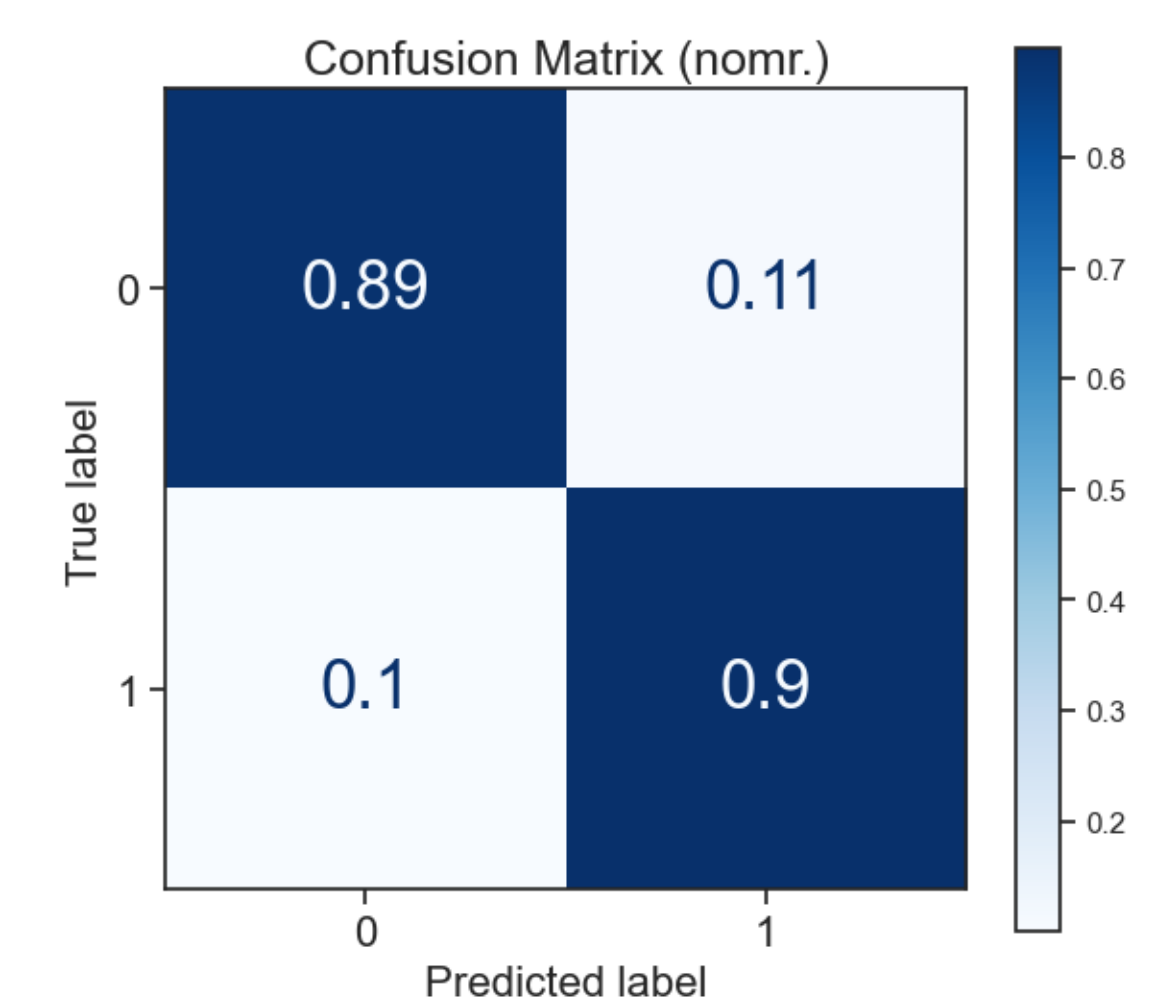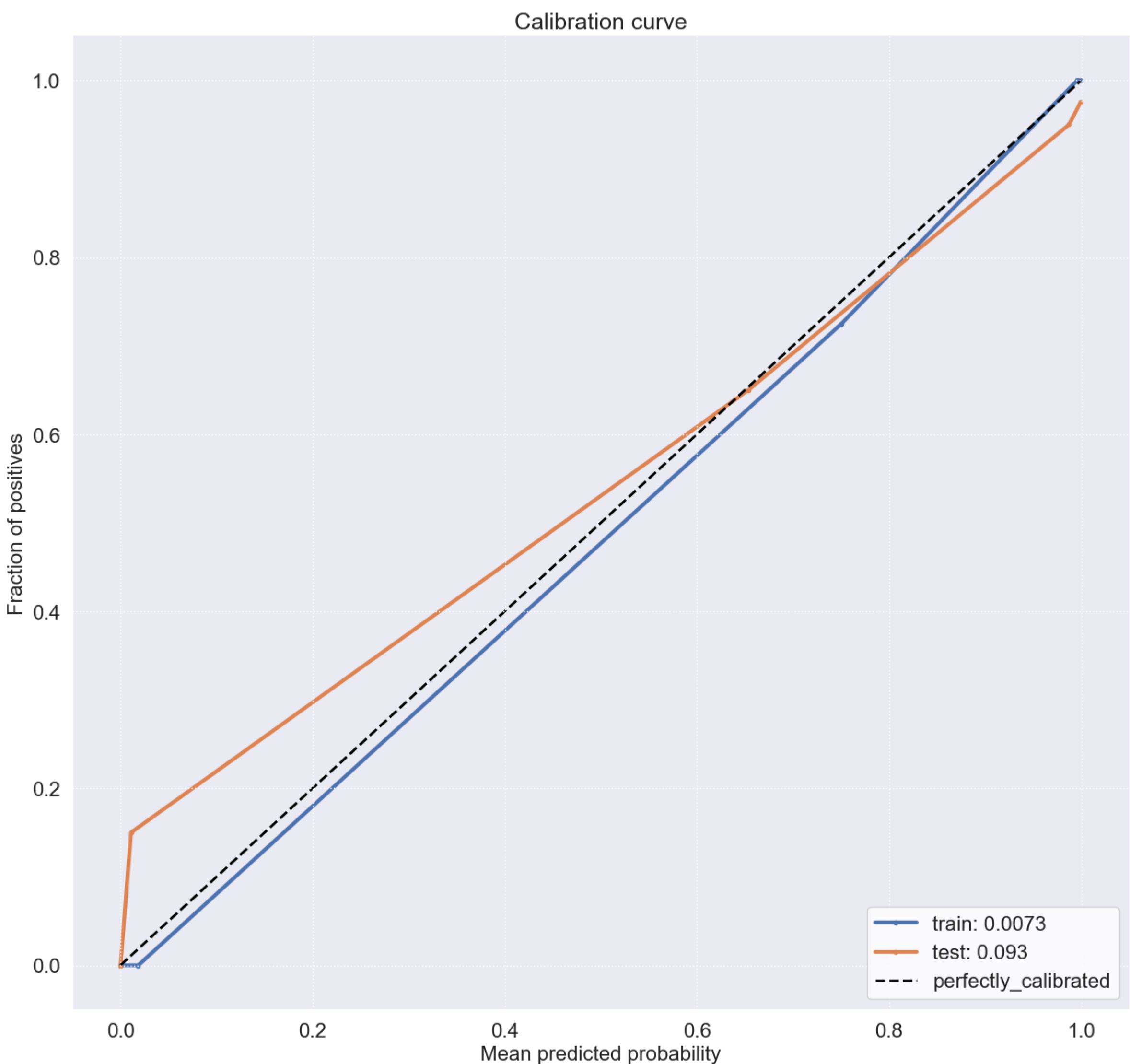- Best Trial (# 2107)
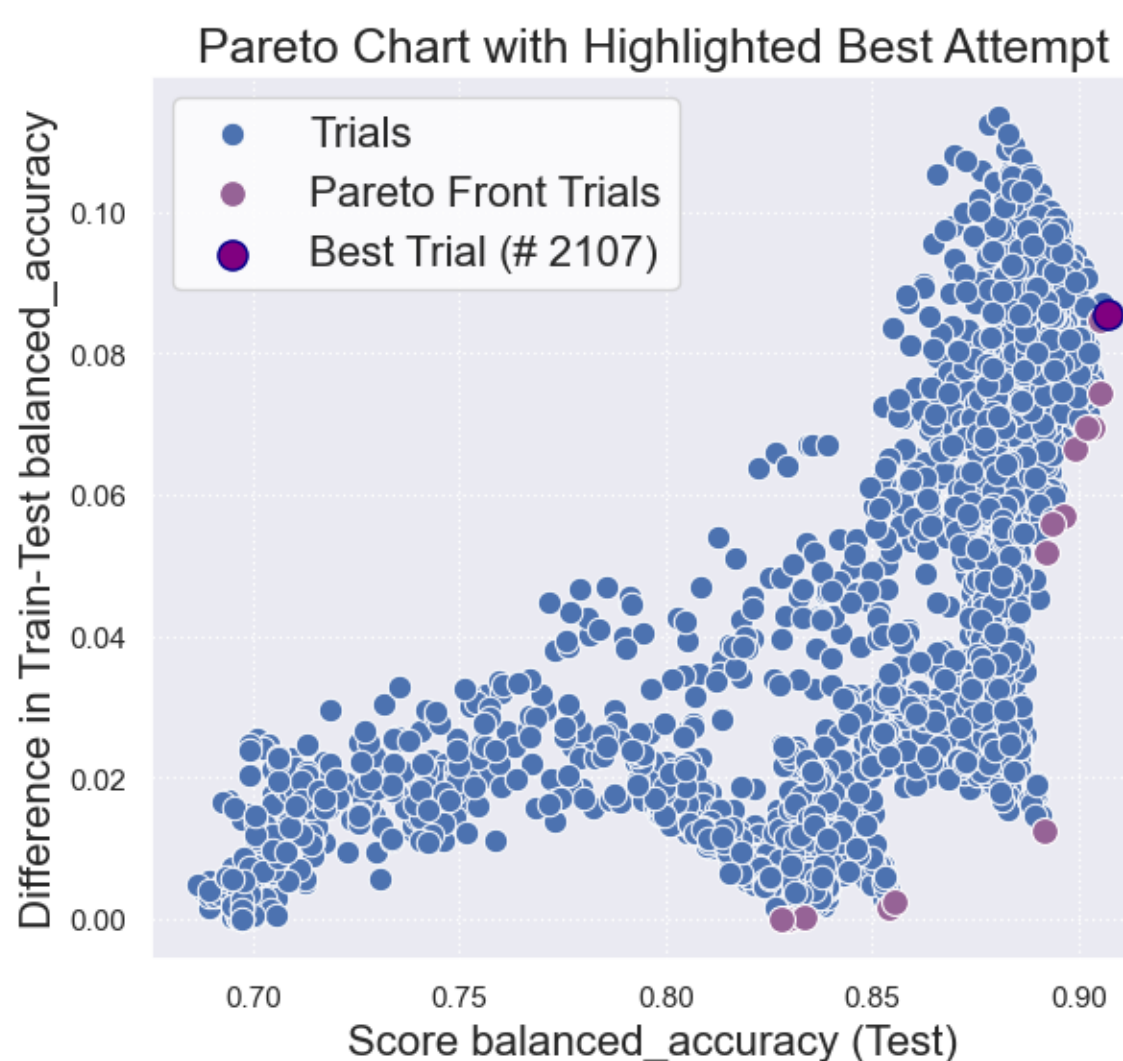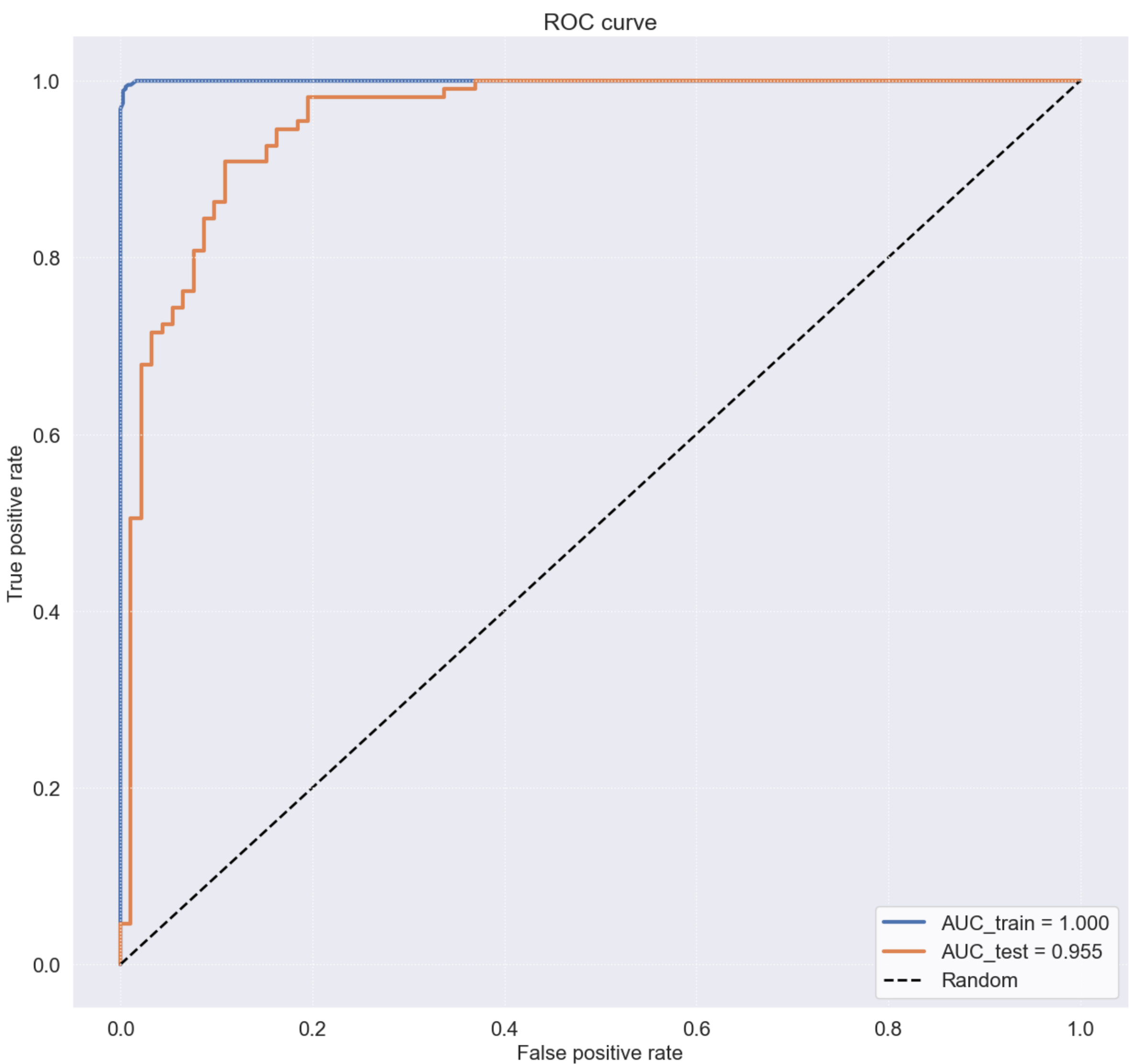
XGBoost Parameters	Value
alpha	2.5798
colsample_bytree	0.9548
gamma	0.1682
lambda	3.1338
learning_rate	0.2508
max_depth	10.0
min_child_weight	1.0
n_estimators	446.0
reg_alpha	0.8781
reg_lambda	0.1112
scale_pos_weight	1.8141
subsample	0.542

Confusion Matrix

Confusion Matrix (nomr.)

Calibration curve

- train: 0.0073
- test: 0.093
- perfectly_calibrated

Active/Inactive probability

- type
- active (109)
- inactive (92)
- decoy (201)
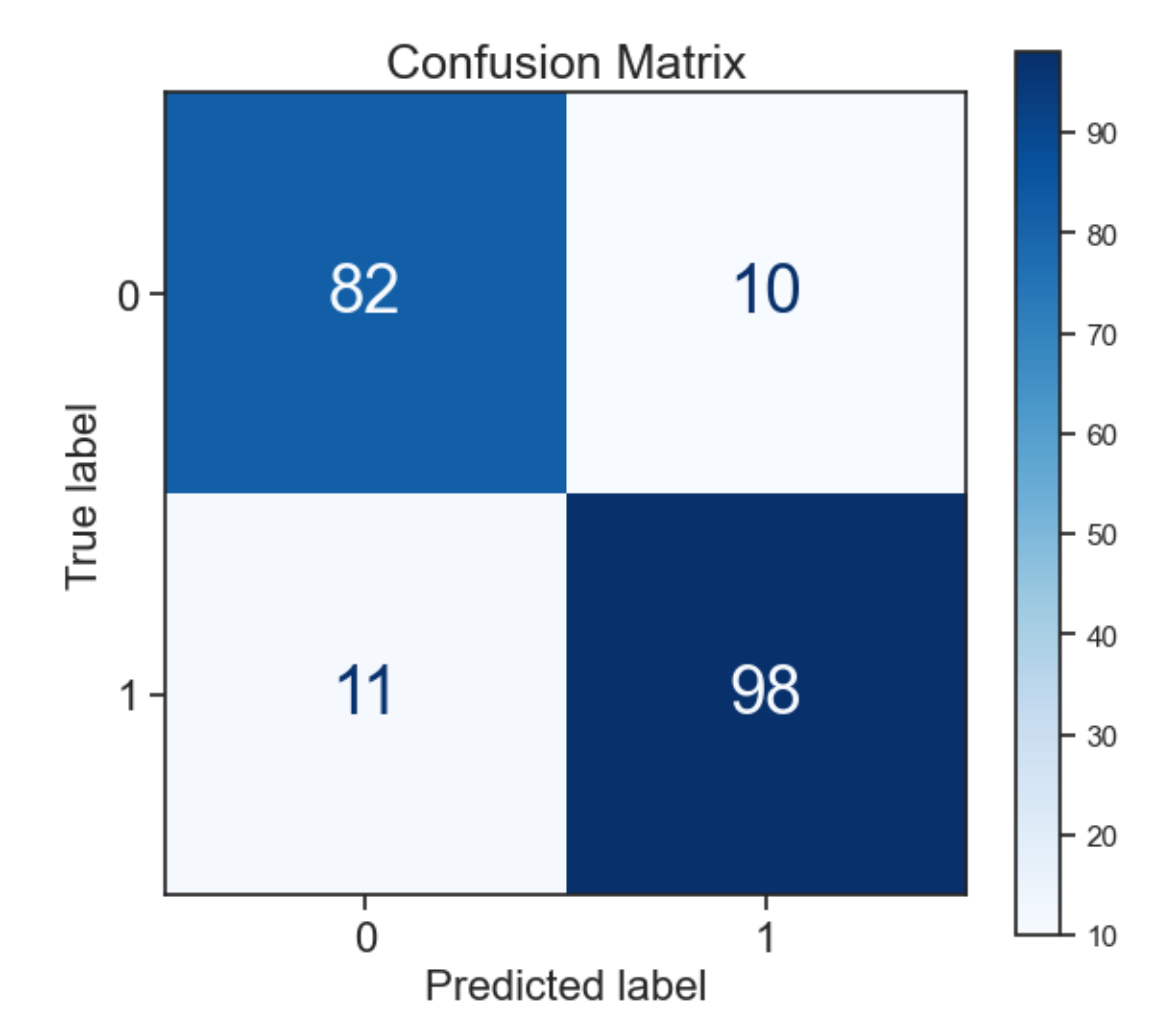
Metric	Train	Test
AUC	0.9999	0.9546
accuracy	0.9913	0.8955
balanced accuracy	0.9904	0.8952
recall_(sens)	1.0	0.8991
specificity	0.9808	0.8913
precision	0.9842	0.9074
f1_score	0.9913	0.8956

Protein: P08588 - metric: f1

Total (train): 801

Bar chart showing Activity type counts: Inactive = 365, Active = 436.

ROC curve

AUC_train = 0.999
AUC_test = 0.950
Random

Pareto Chart with Highlighted Best Attempt

Trials
Pareto Front Trials
Best Trial (# 2309)

X-axis: Score f1 (Test)
Y-axis: Difference in Train-Test f1
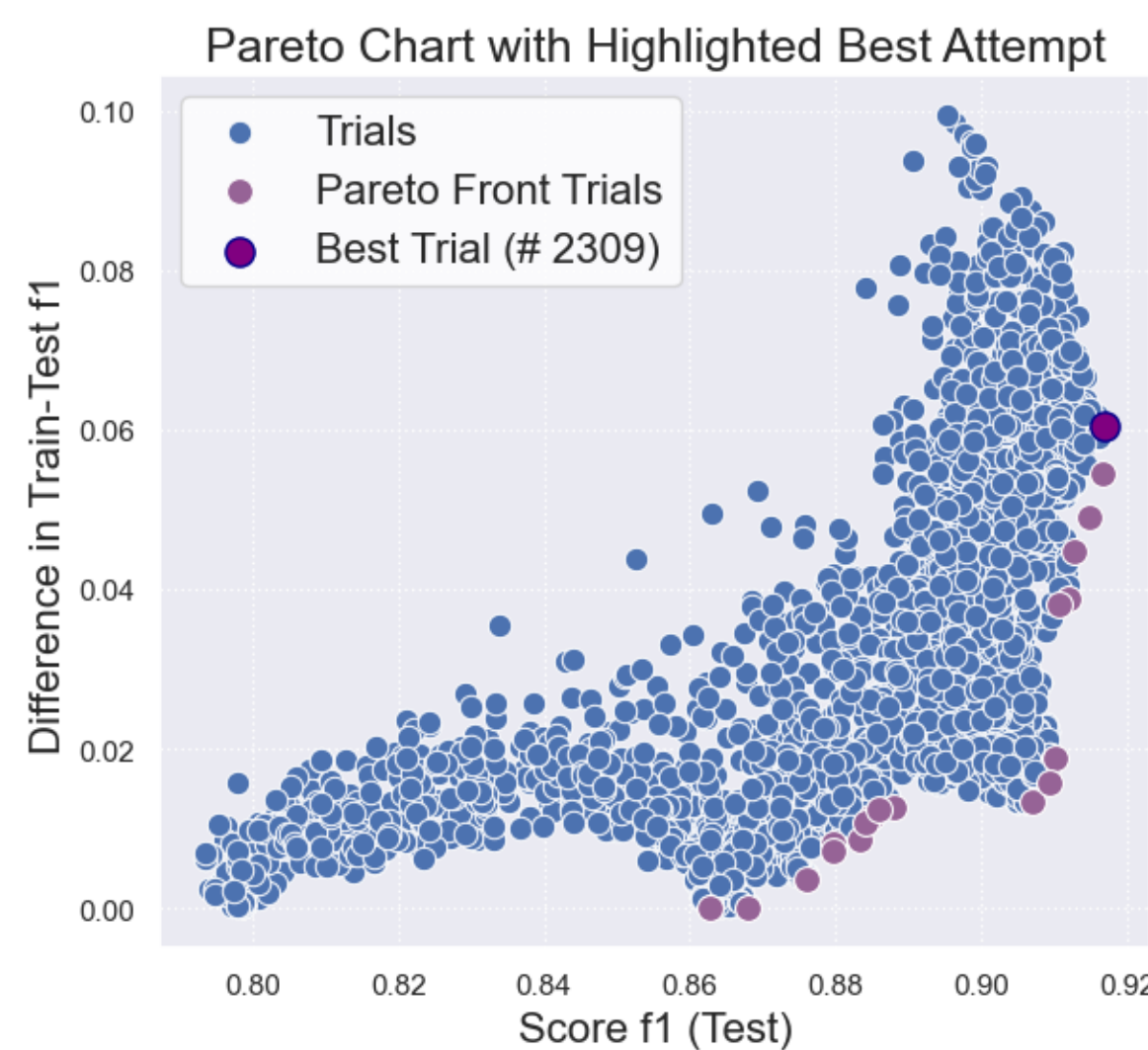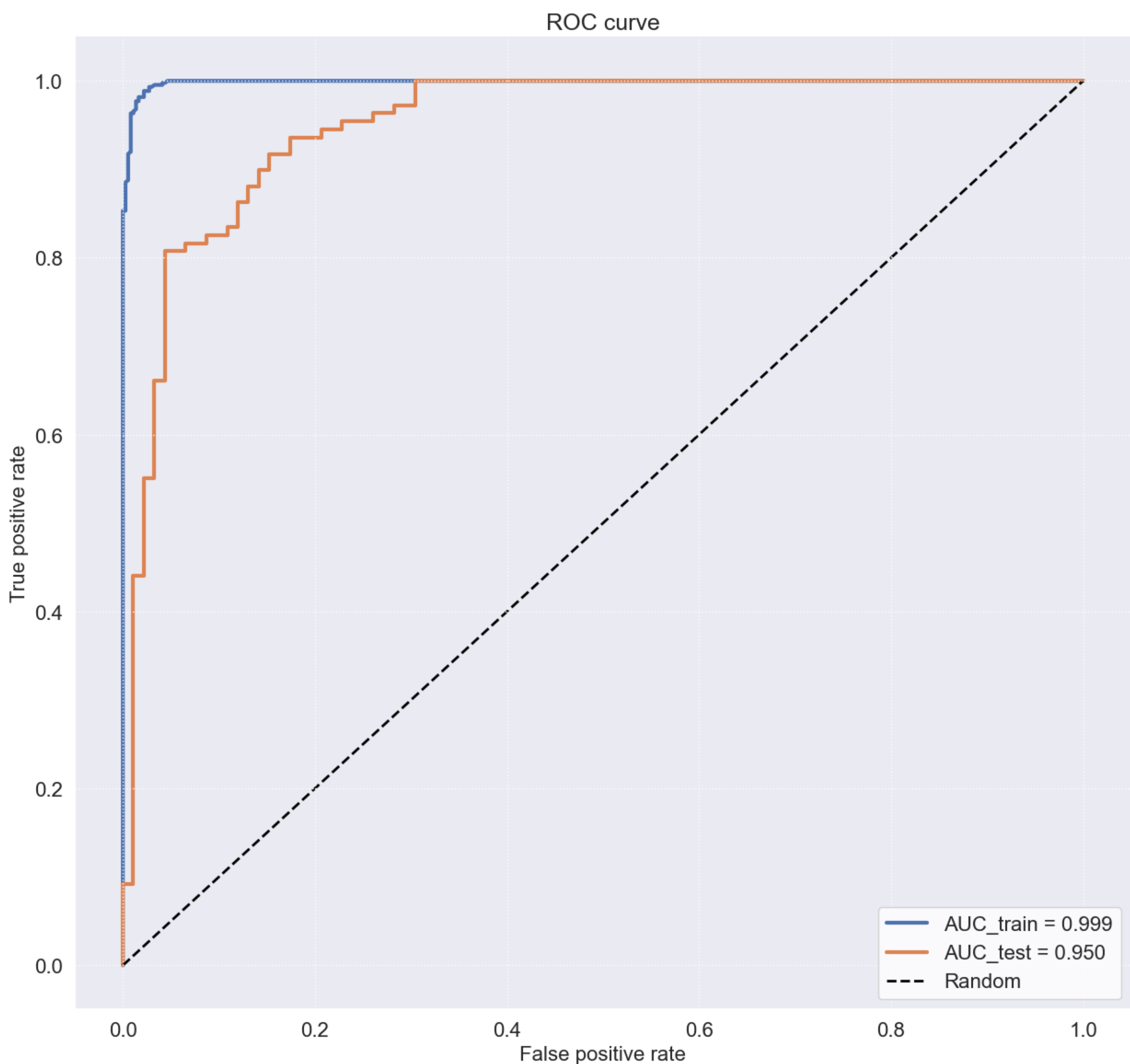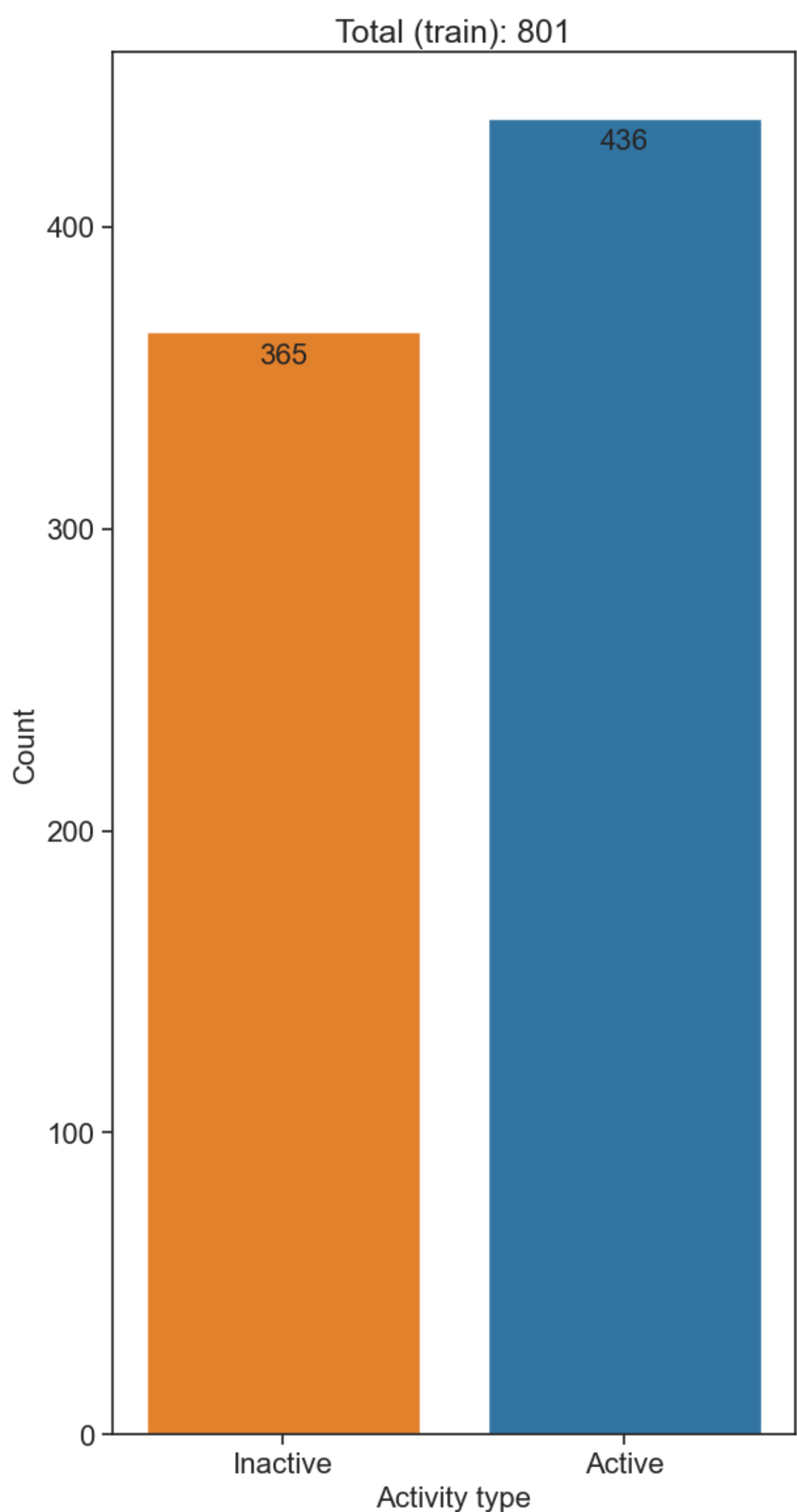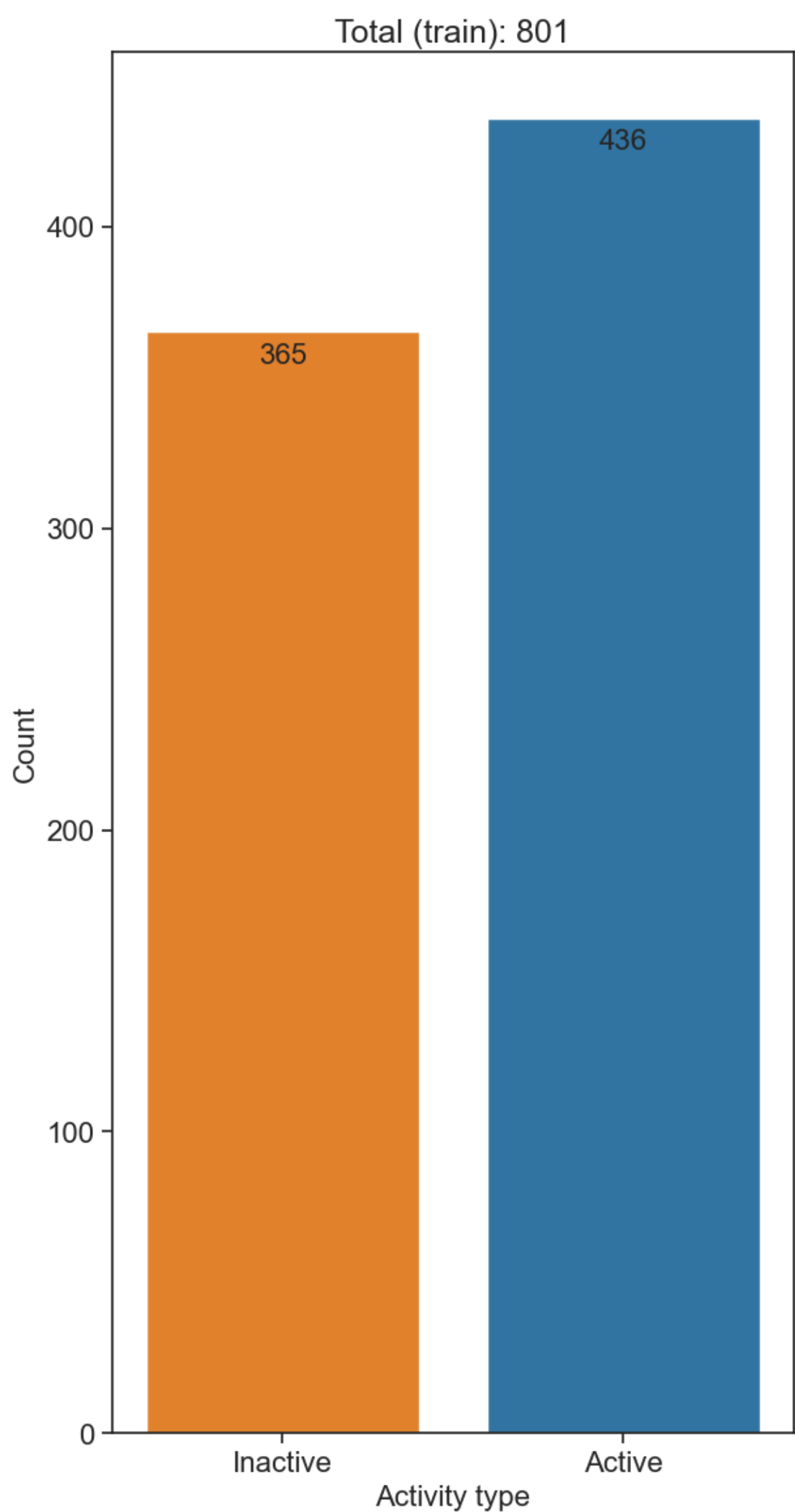
XGBoost Parameters	Value
alpha	2.4115
colsample_bytree	0.95
gamma	0.1309
lambda	0.4448
learning_rate	0.0146
max_depth	9.0
min_child_weight	2.0
n_estimators	575.0
reg_alpha	0.3426
reg_lambda	0.0272
scale_pos_weight	2.8993
subsample	0.7619

Confusion Matrix

	Predicted 0	Predicted 1
True 0	78	14
True 1	9	100

Confusion Matrix (nomr.)

	Predicted 0	Predicted 1
True 0	0.85	0.15
True 1	0.083	0.92

Calibration curve

train: 0.0221
test: 0.0942
perfectly_calibrated

Active/Inactive probability

Inactive zone   Active zone

type
active (109)
inactive (92)
decoy (201)
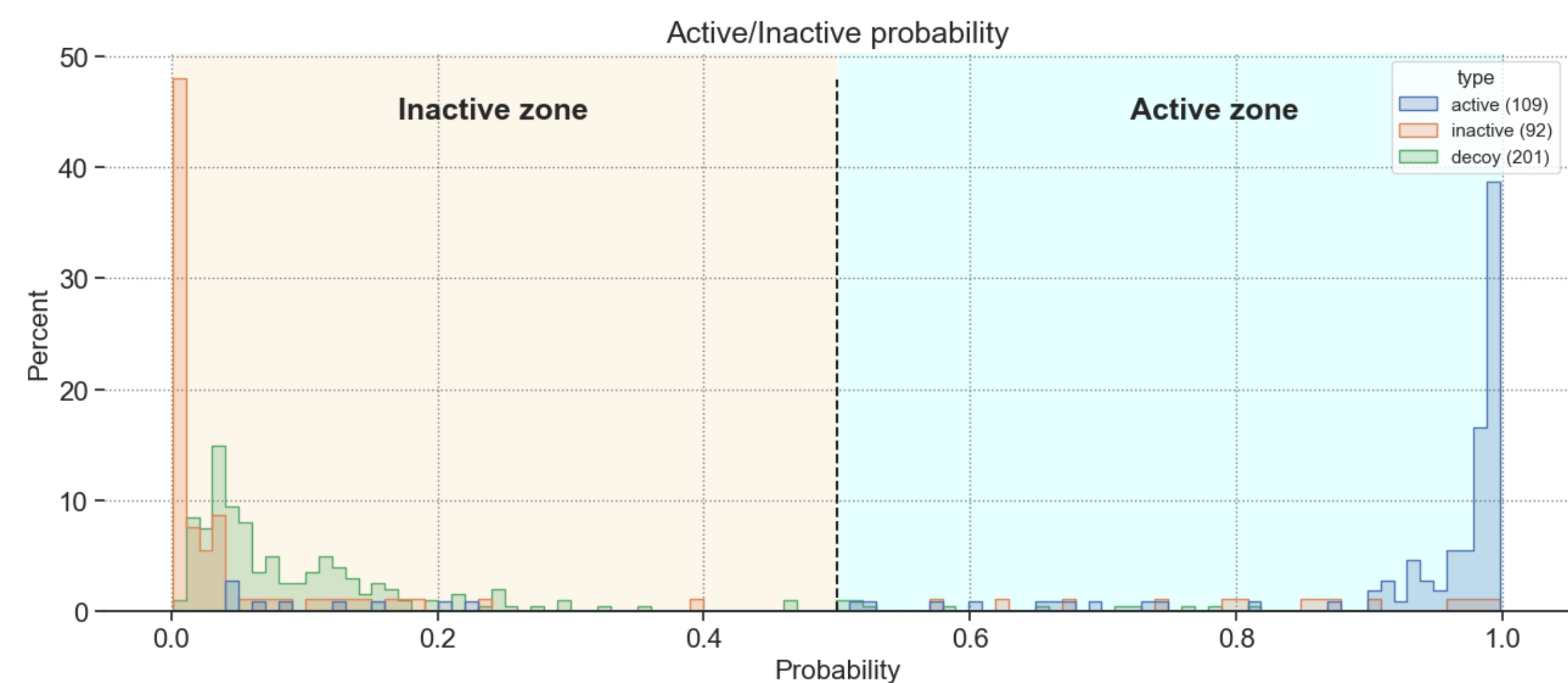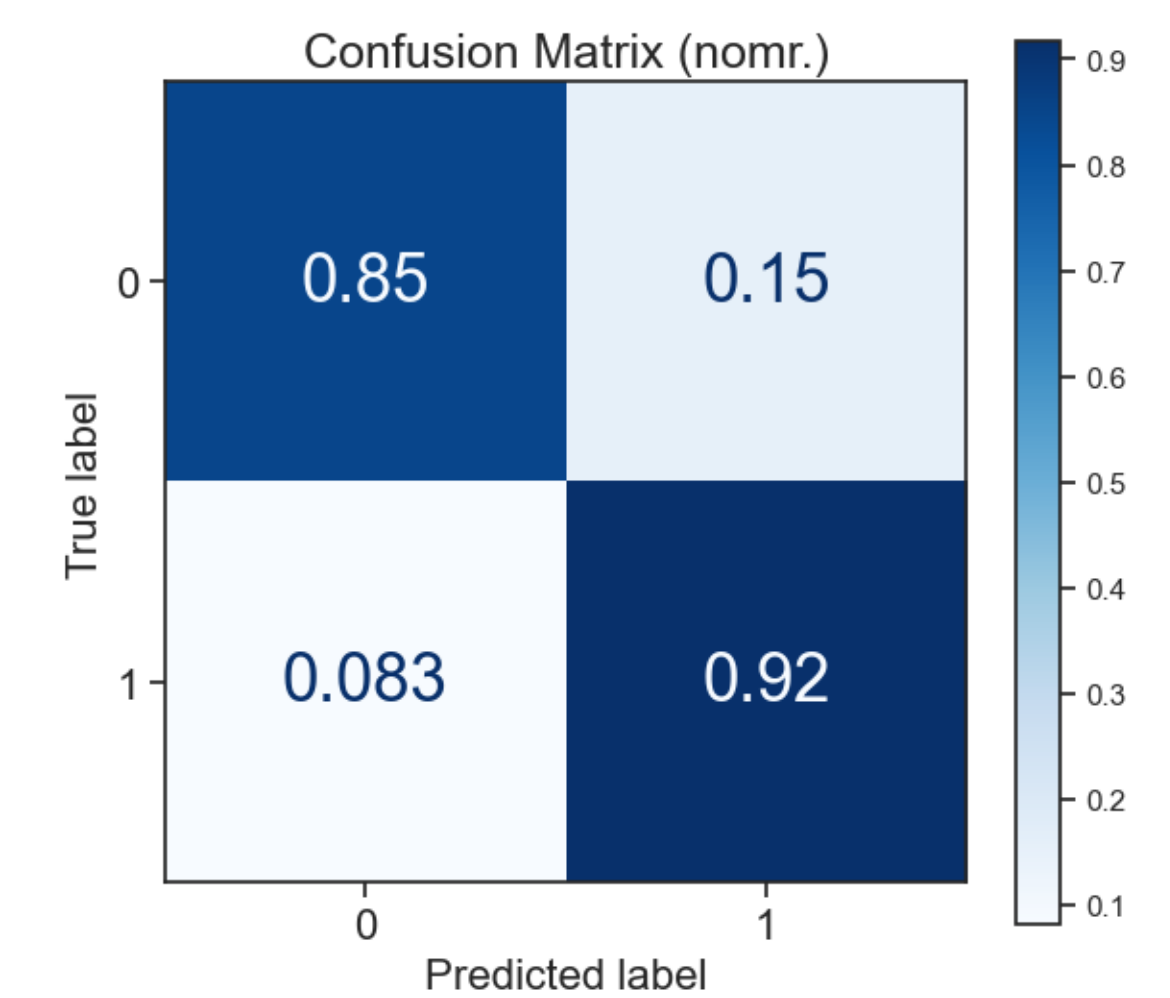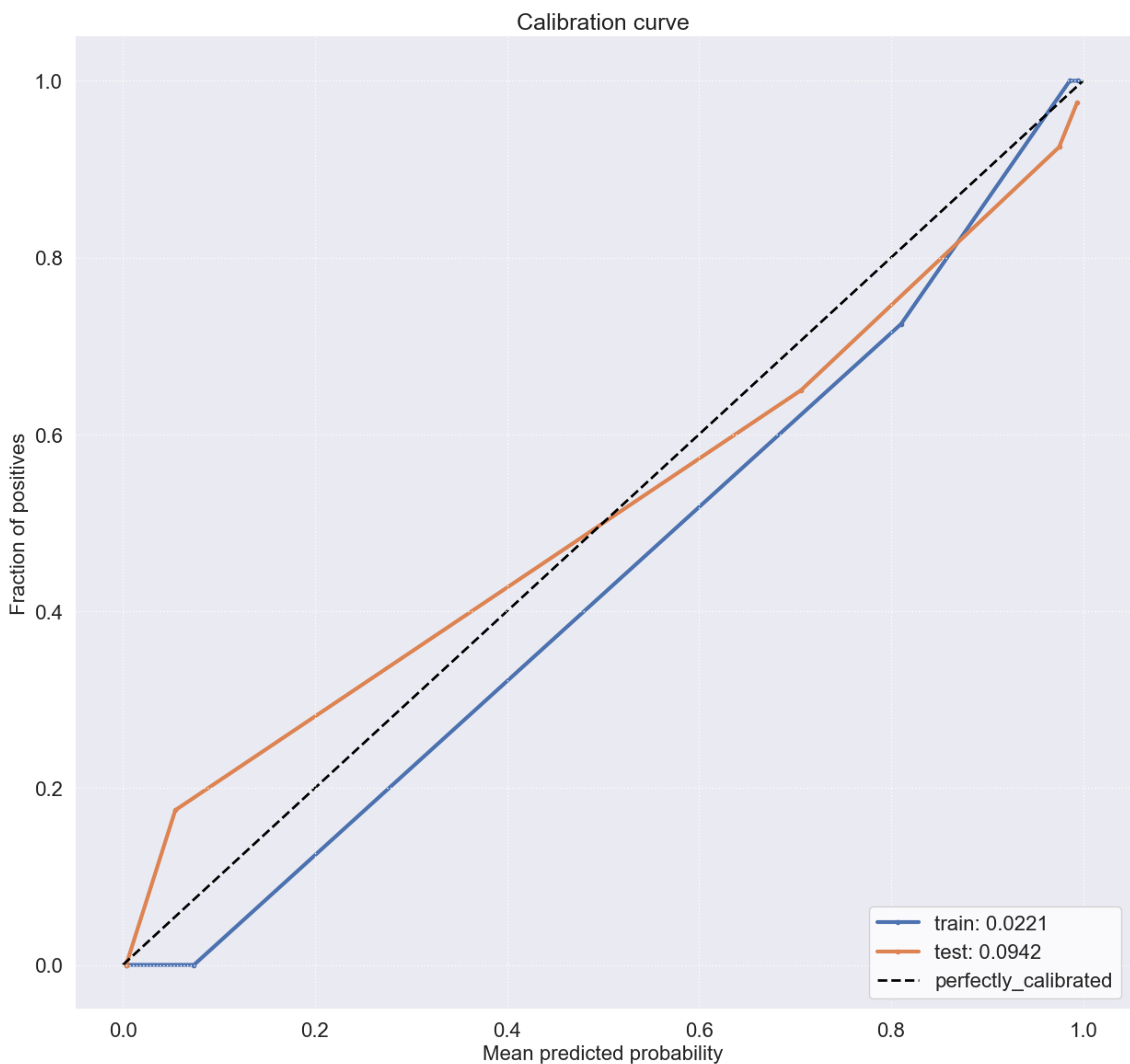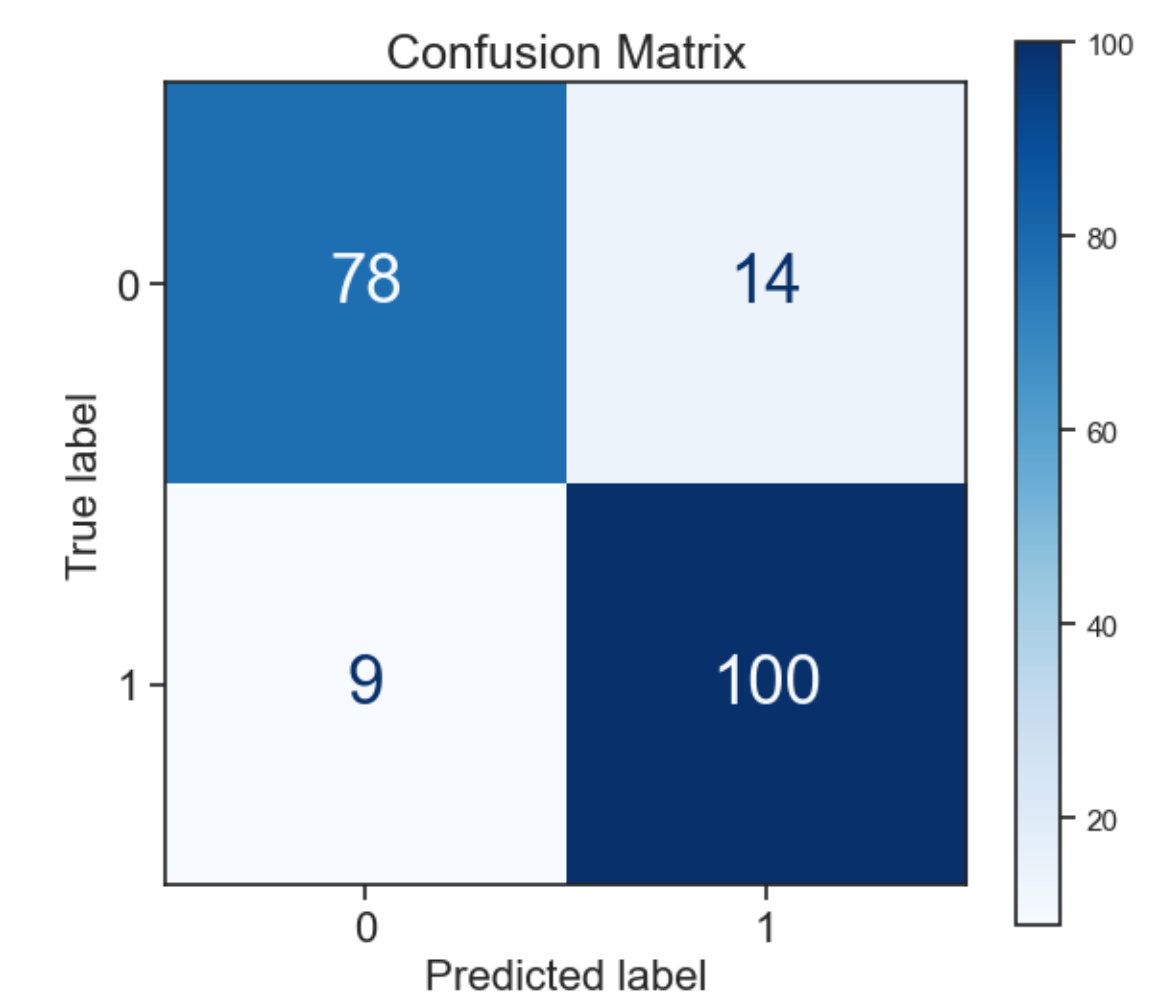
Metric	Train	Test
AUC	0.9986	0.9498
accuracy	0.9725	0.8856
balanced accuracy	0.9699	0.8826
recall_(sens)	1.0	0.9174
specificity	0.9397	0.8478
precision	0.952	0.8772
f1_score	0.9724	0.8853

Protein: P08588 - metric: precision

Total (train): 801

ROC curve

	AUC_train = 0.998
	AUC_test = 0.955
	Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2711)
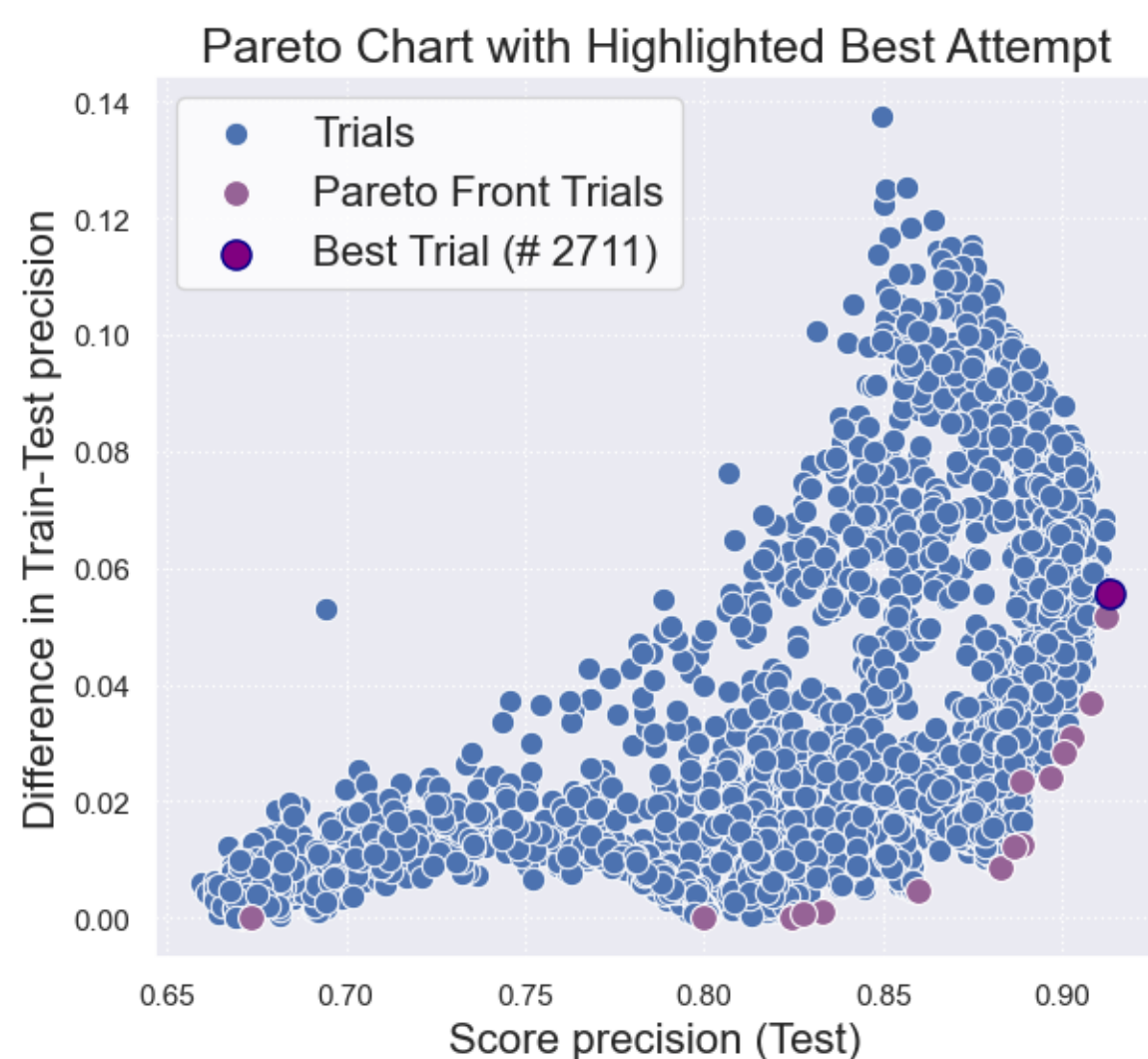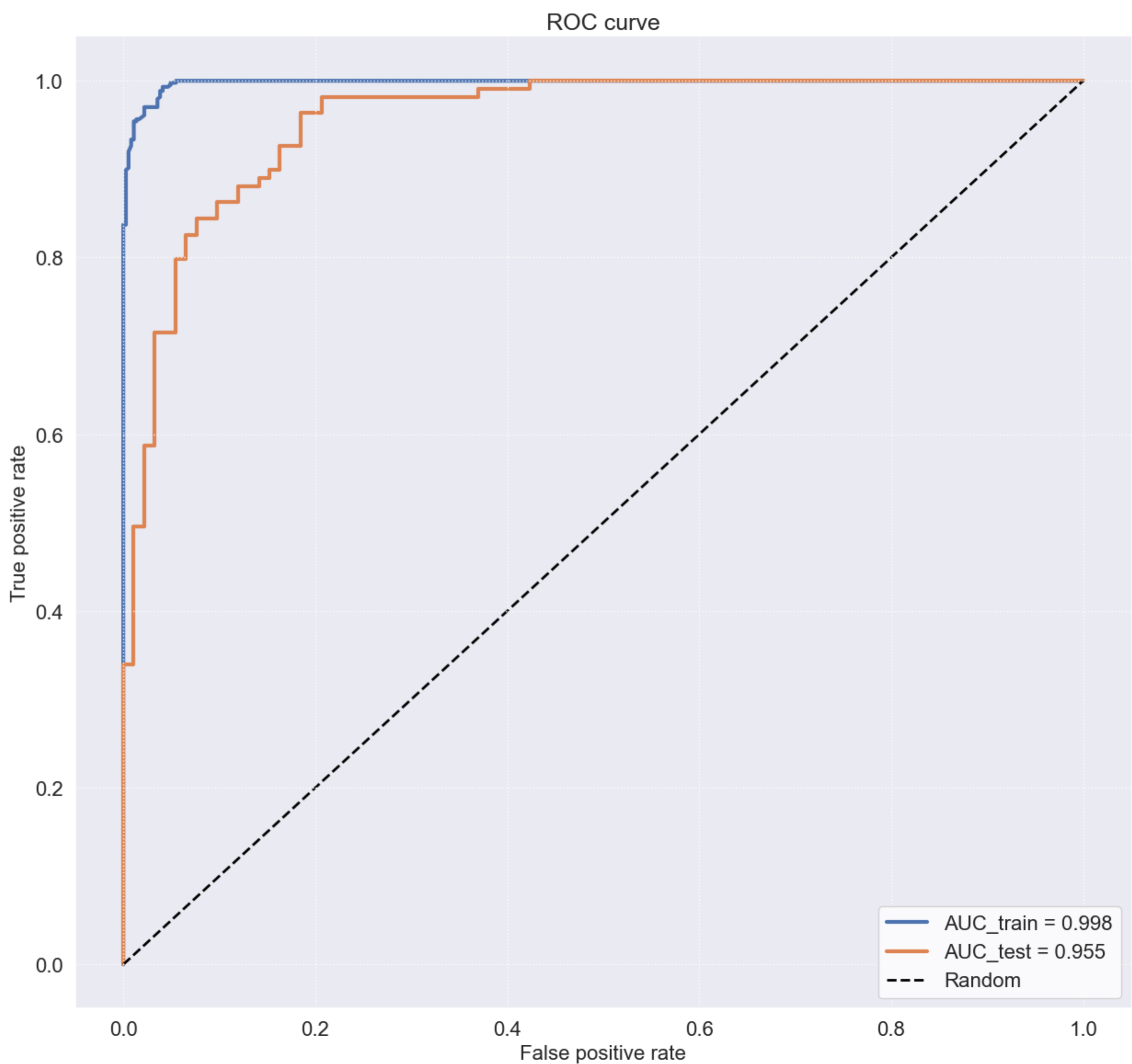
XGBoost Parameters	Value
alpha	2.3133
colsample_bytree	0.6094
gamma	0.511
lambda	4.1654
learning_rate	0.2238
max_depth	9.0
min_child_weight	2.0
n_estimators	379.0
reg_alpha	0.7061
reg_lambda	0.1606
scale_pos_weight	1.025
subsample	0.7457

Confusion Matrix

	0	1
0	81	11
1	15	94

Confusion Matrix (nomr.)

	0	1
0	0.88	0.12
1	0.14	0.86

Calibration curve

train: 0.0207	
test: 0.0921	
perfectly_calibrated	

Active/Inactive probability

Inactive zone   Active zone

type
- active (109)
- inactive (92)
- decoy (201)
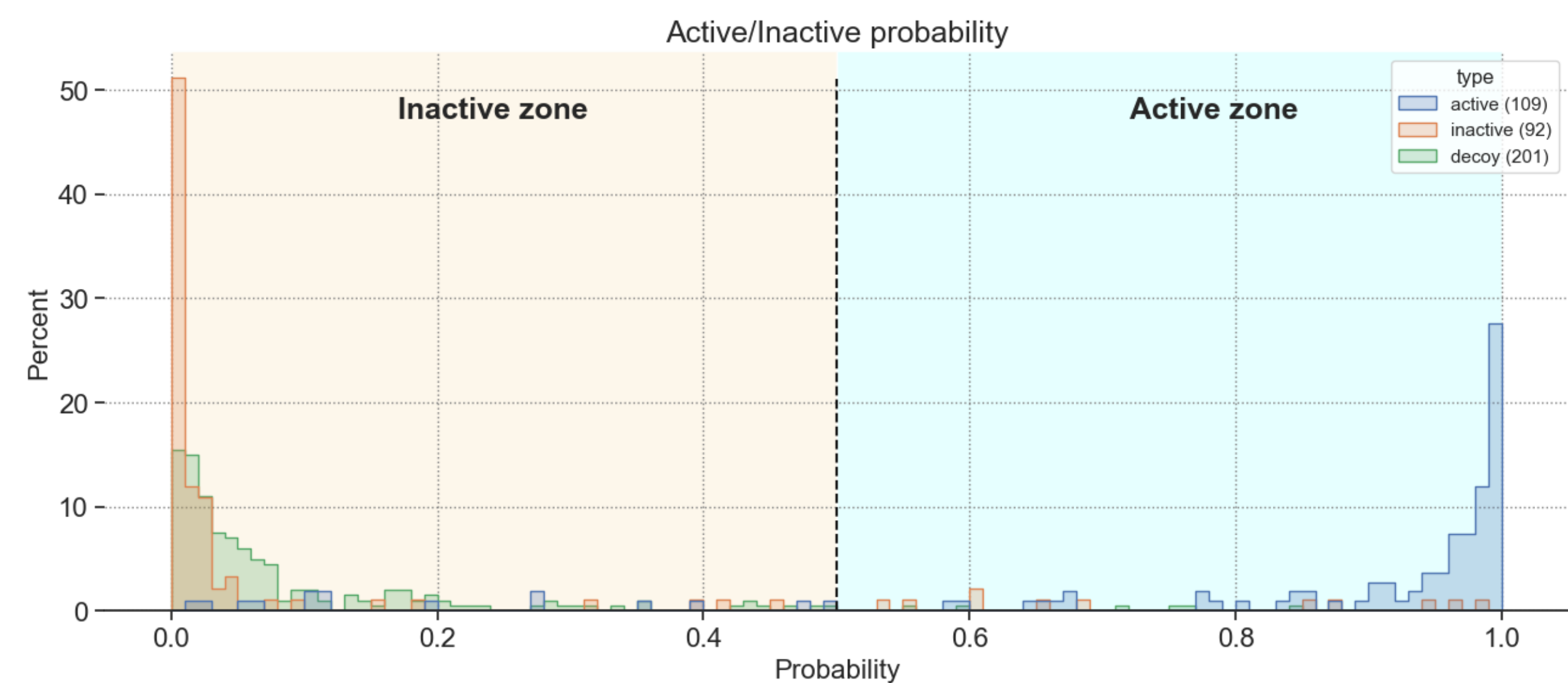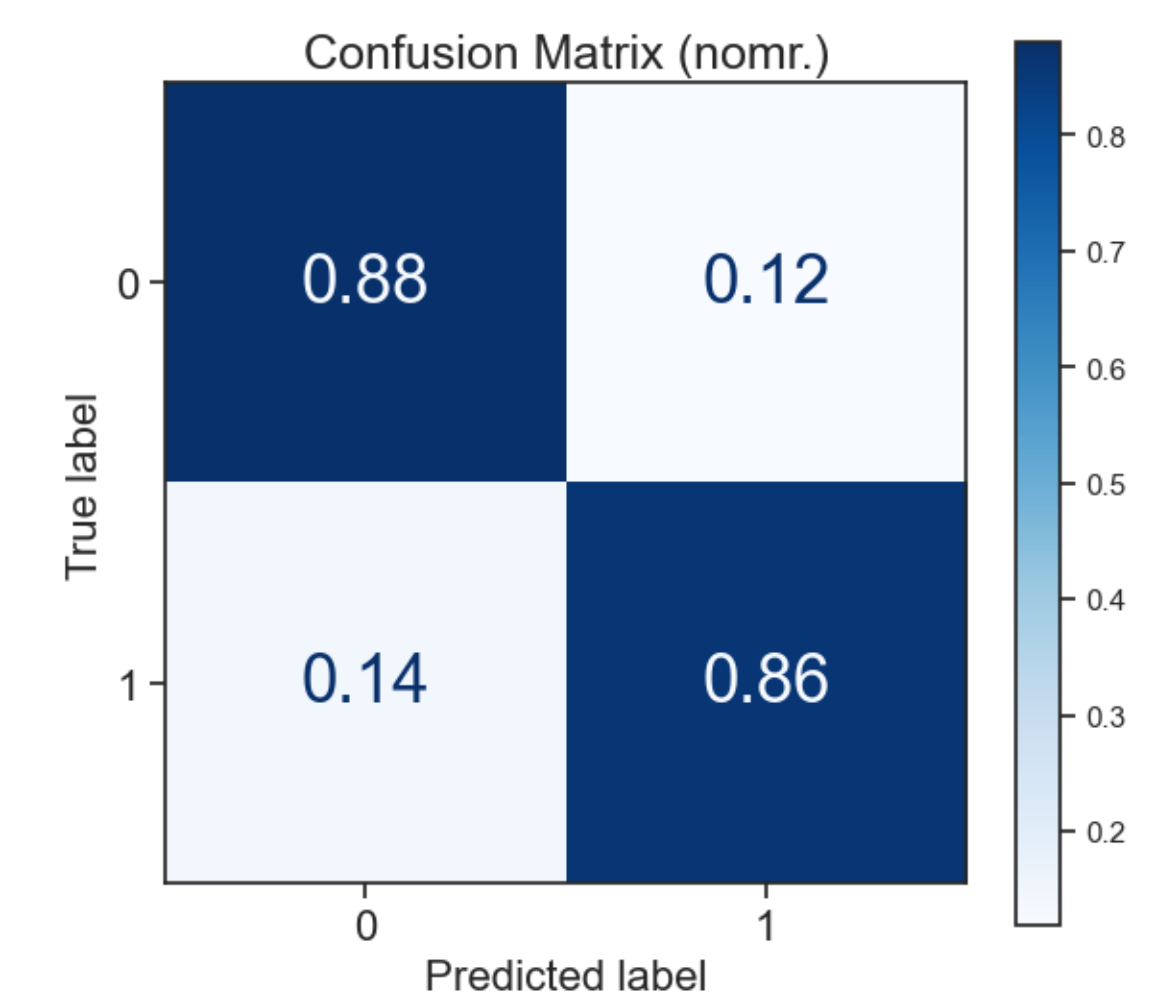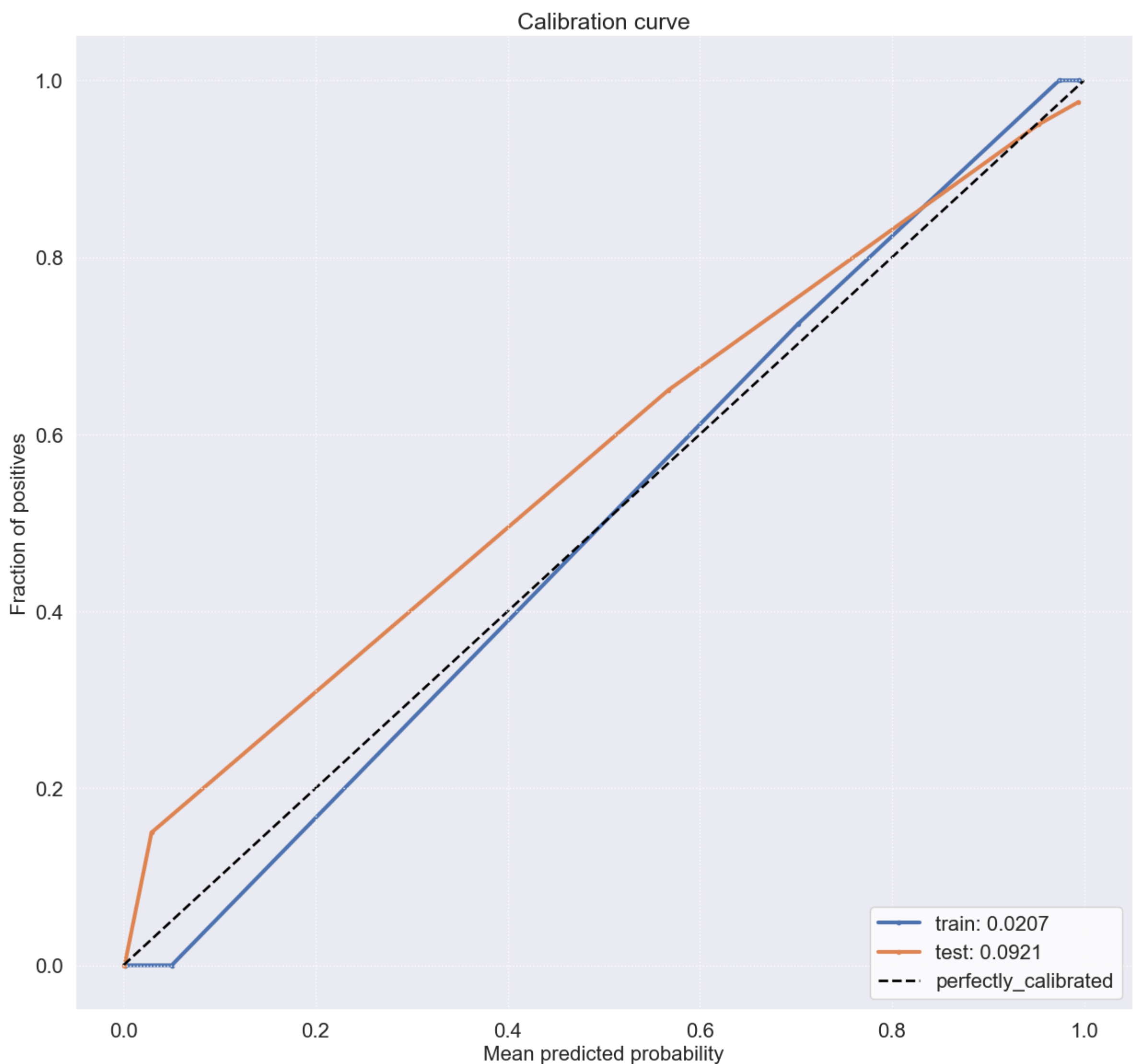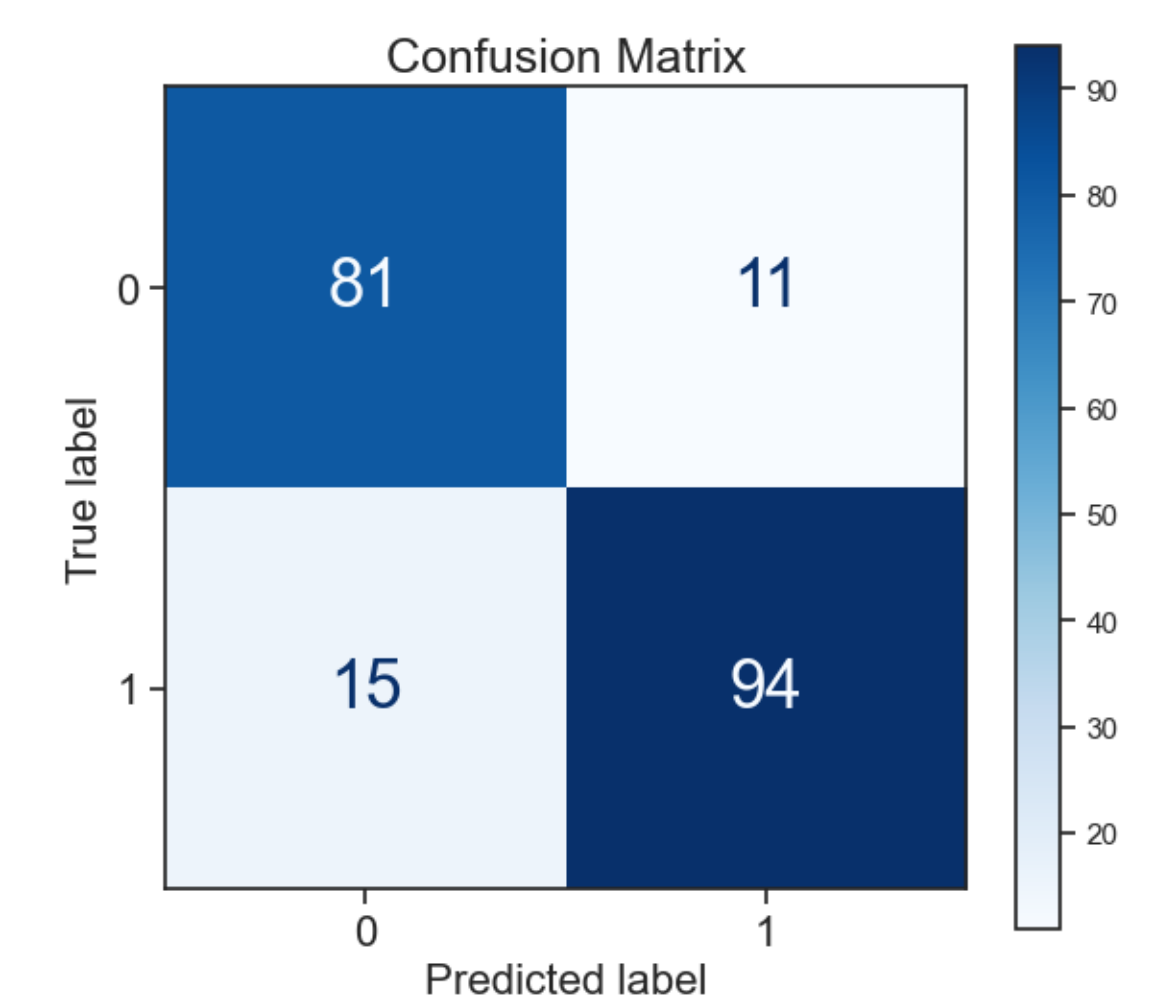
Metric	Train	Test
AUC	0.9979	0.9554
accuracy	0.9738	0.8706
balanced accuracy	0.9728	0.8714
recall_(sens)	0.9839	0.8624
specificity	0.9616	0.8804
precision	0.9684	0.8952
f1_score	0.9738	0.8708

Protein: P08588 - metric: SoftVotingXGBoostEnsemble
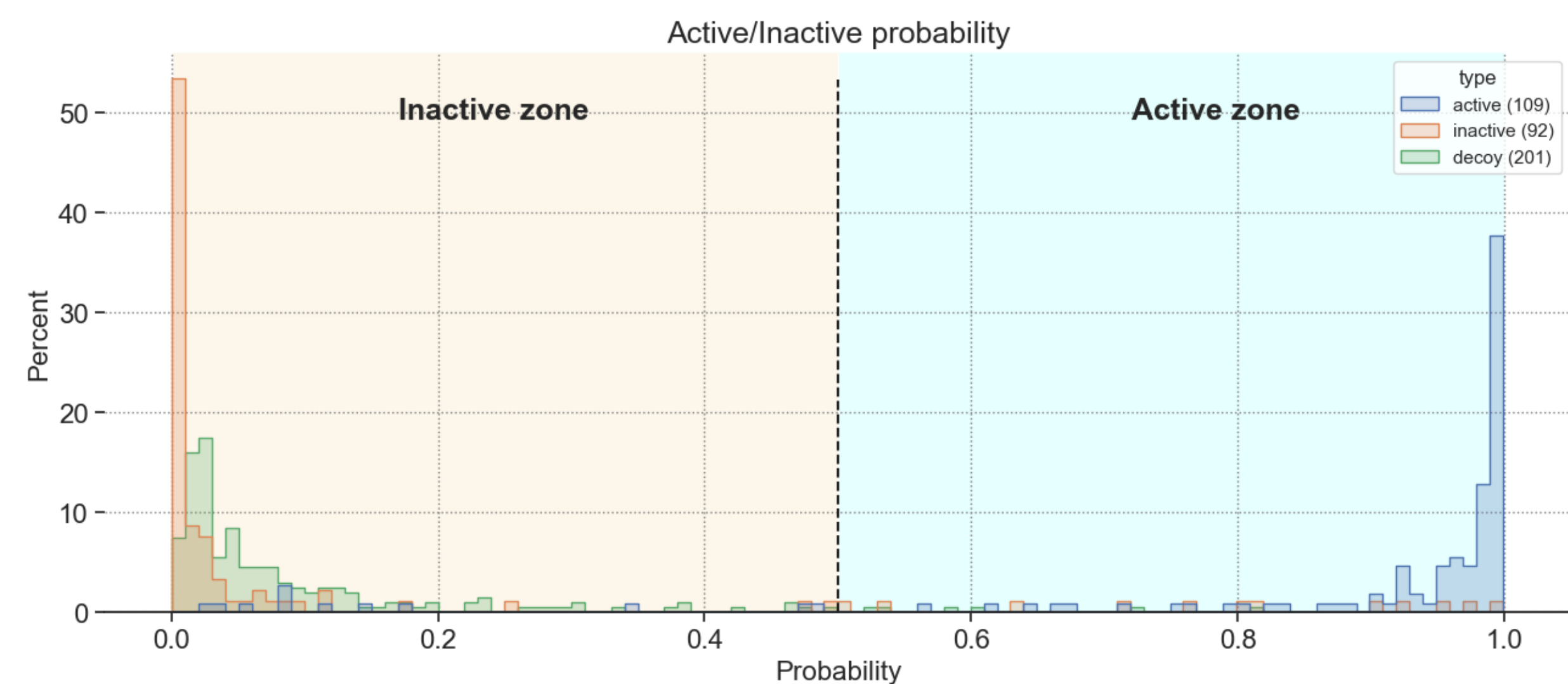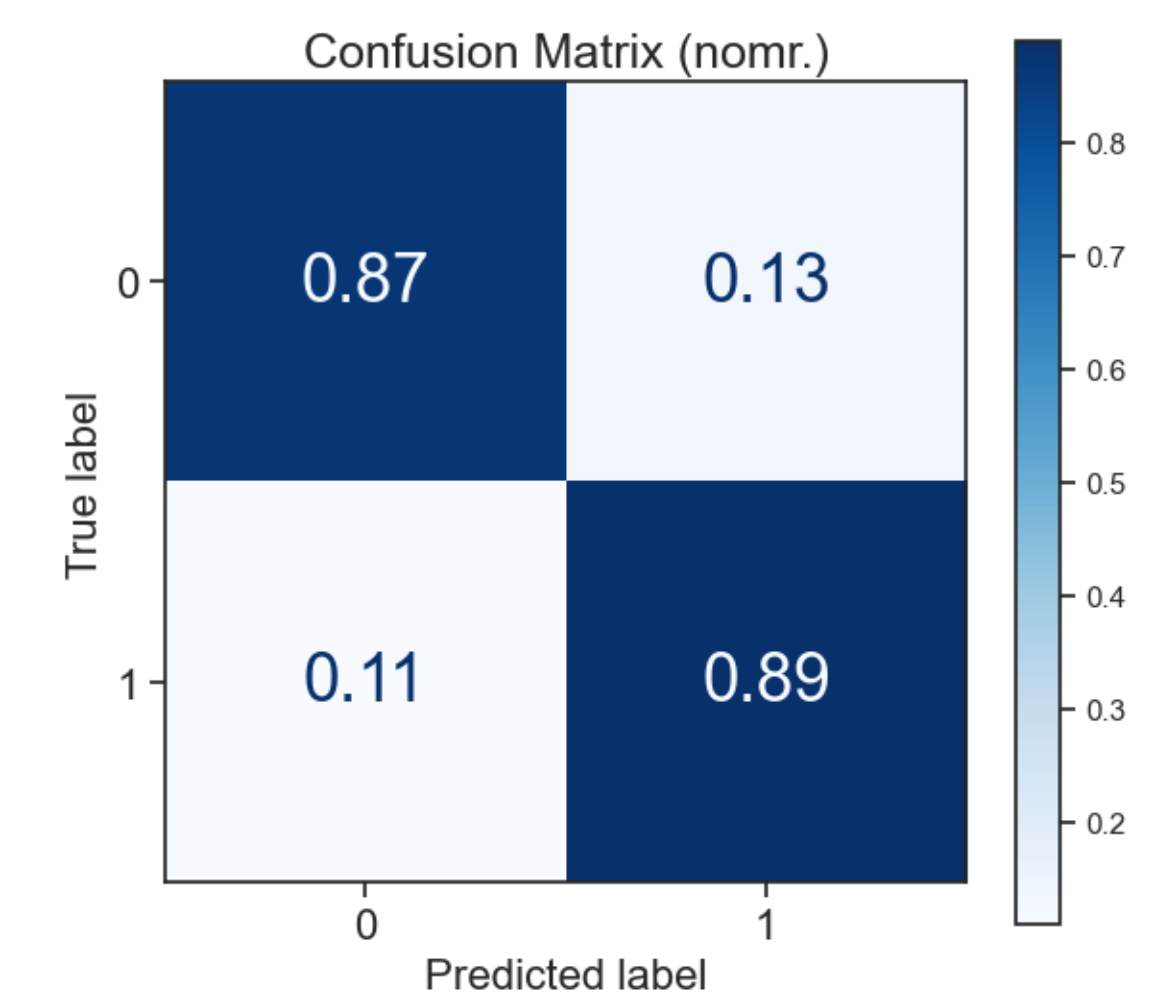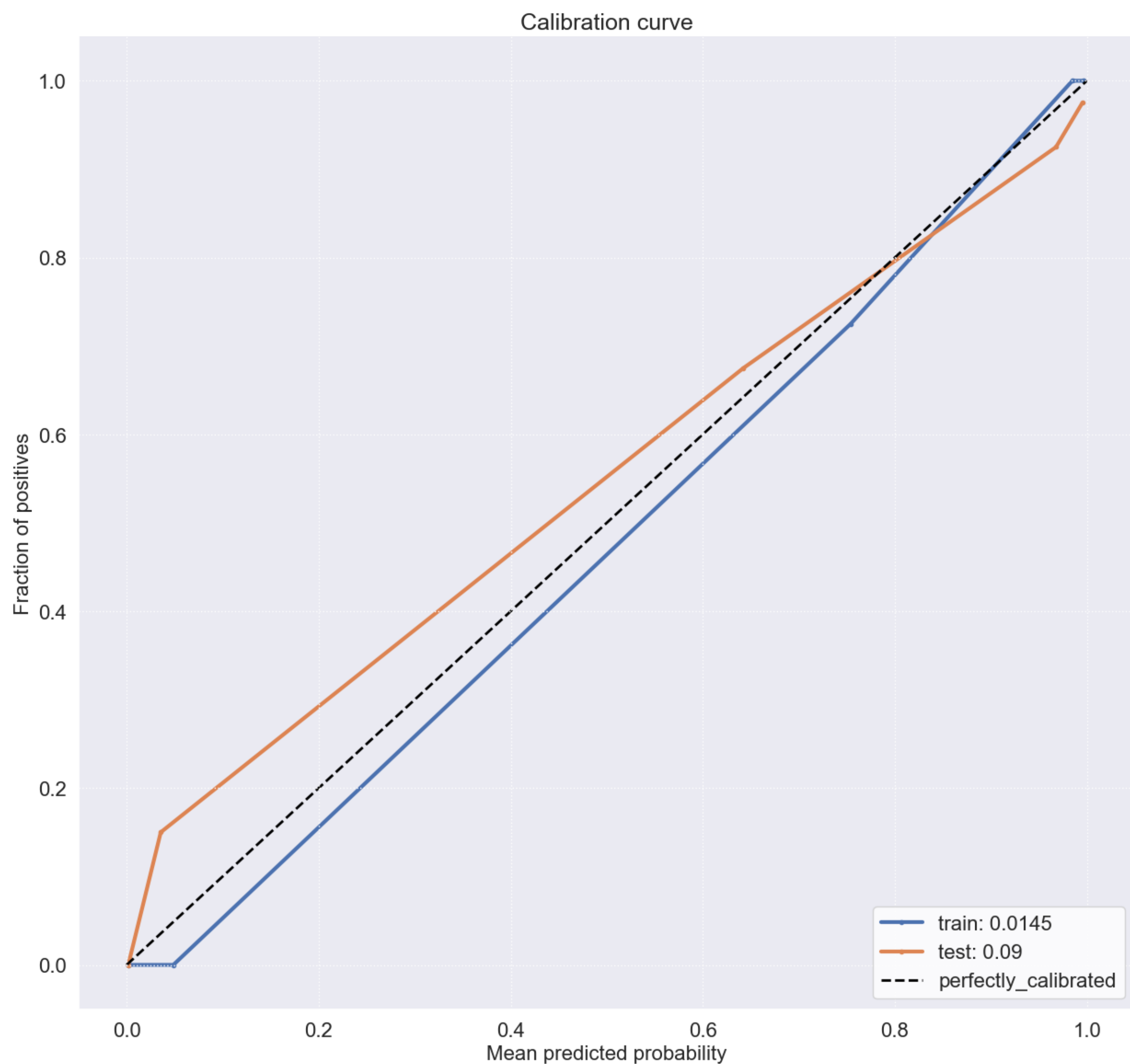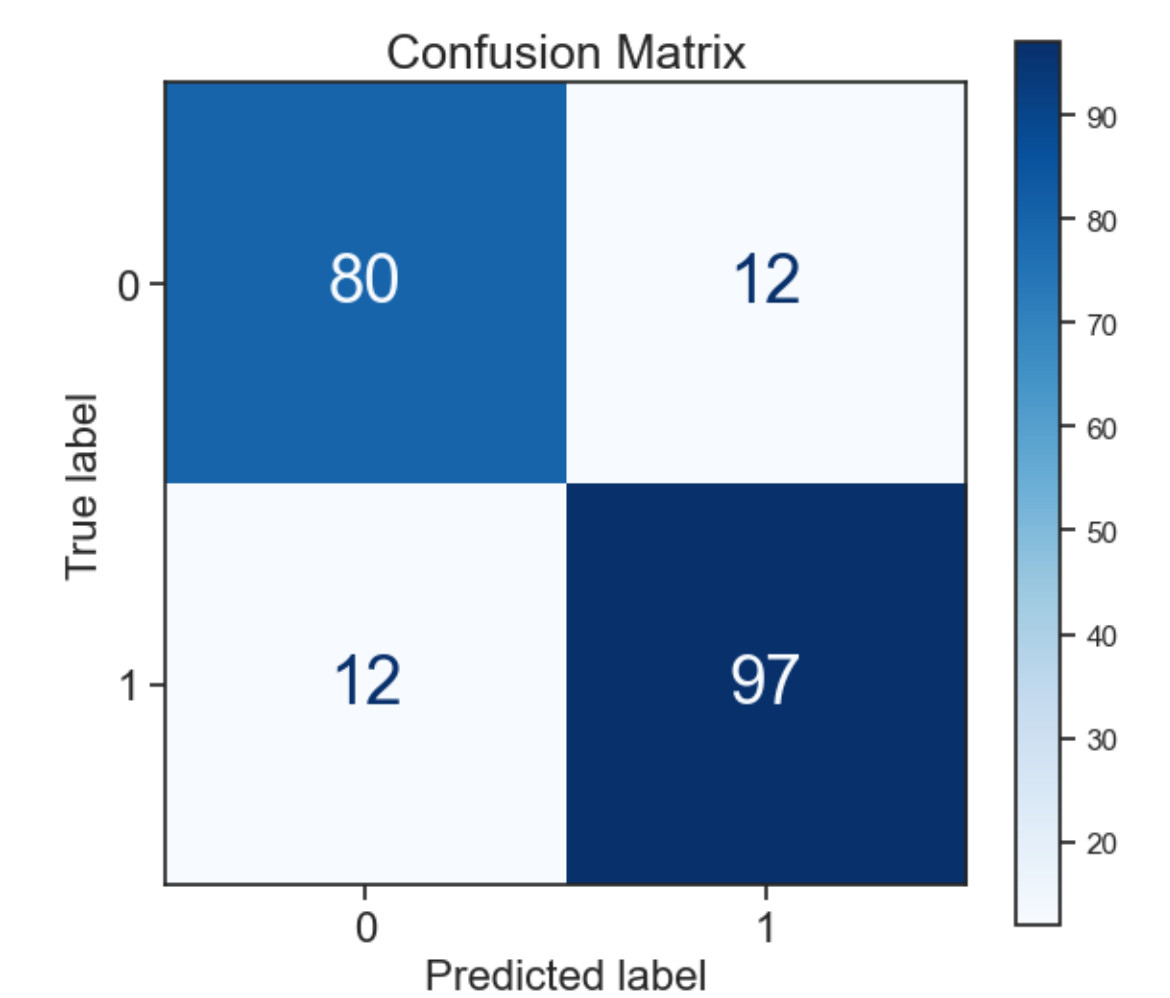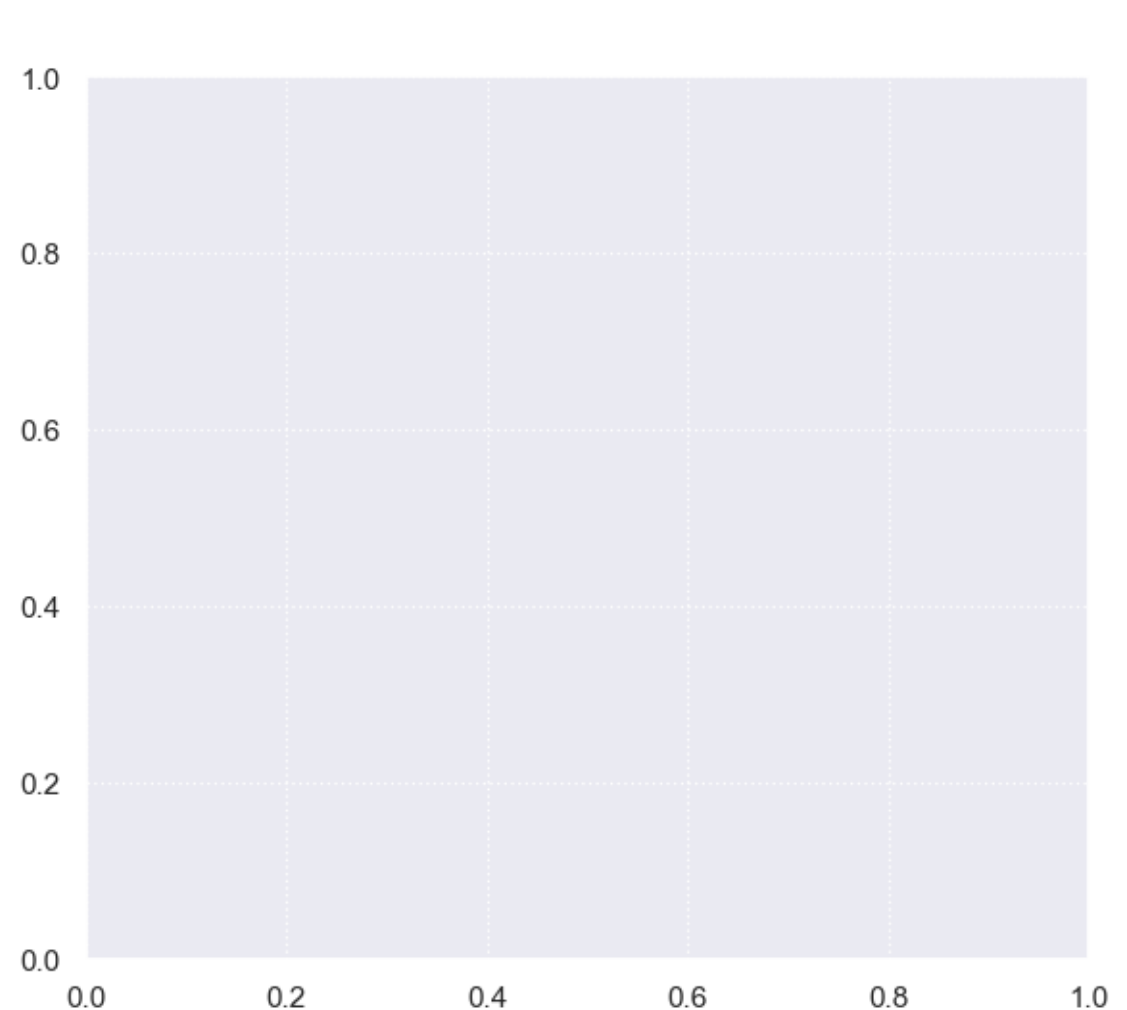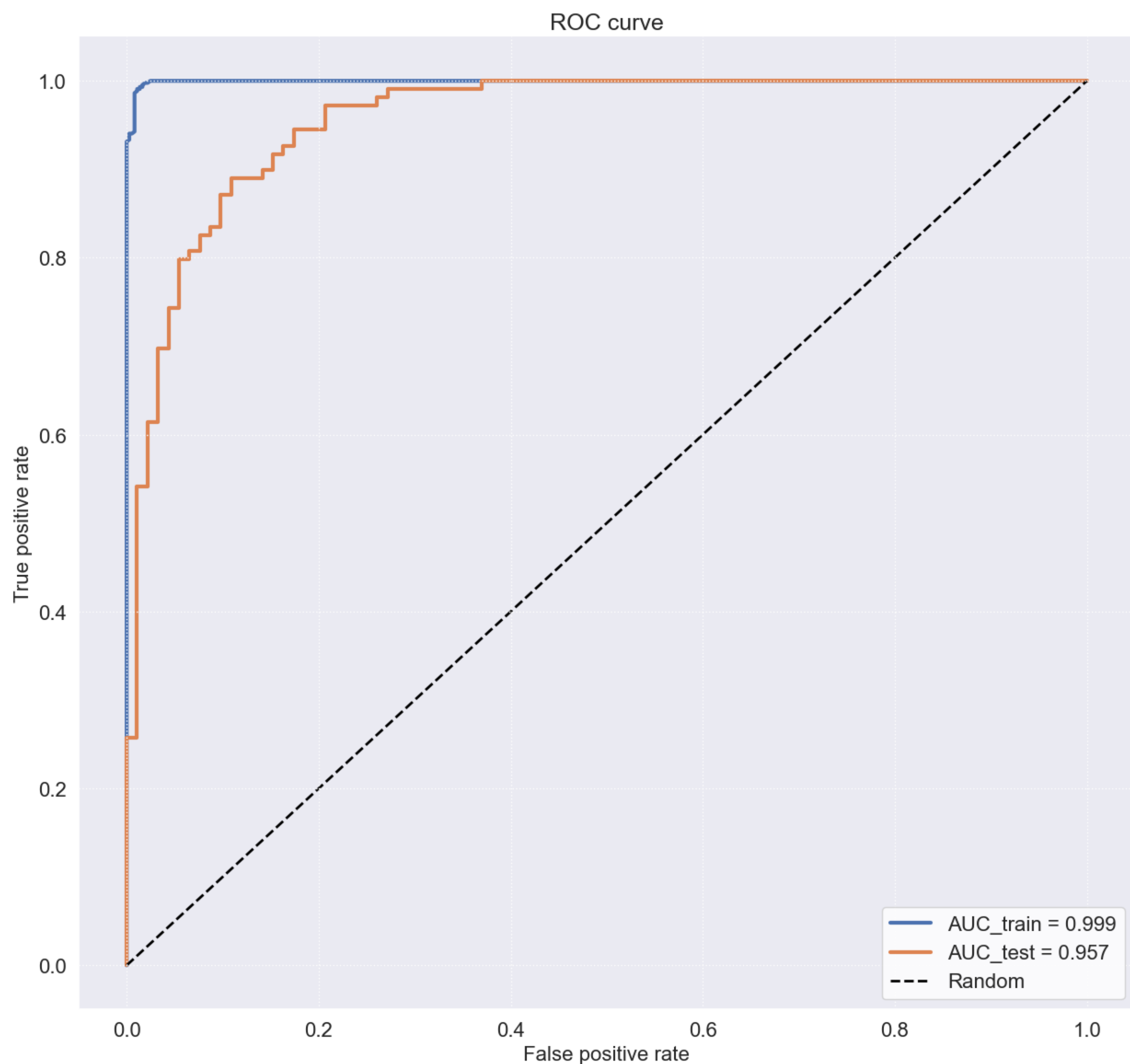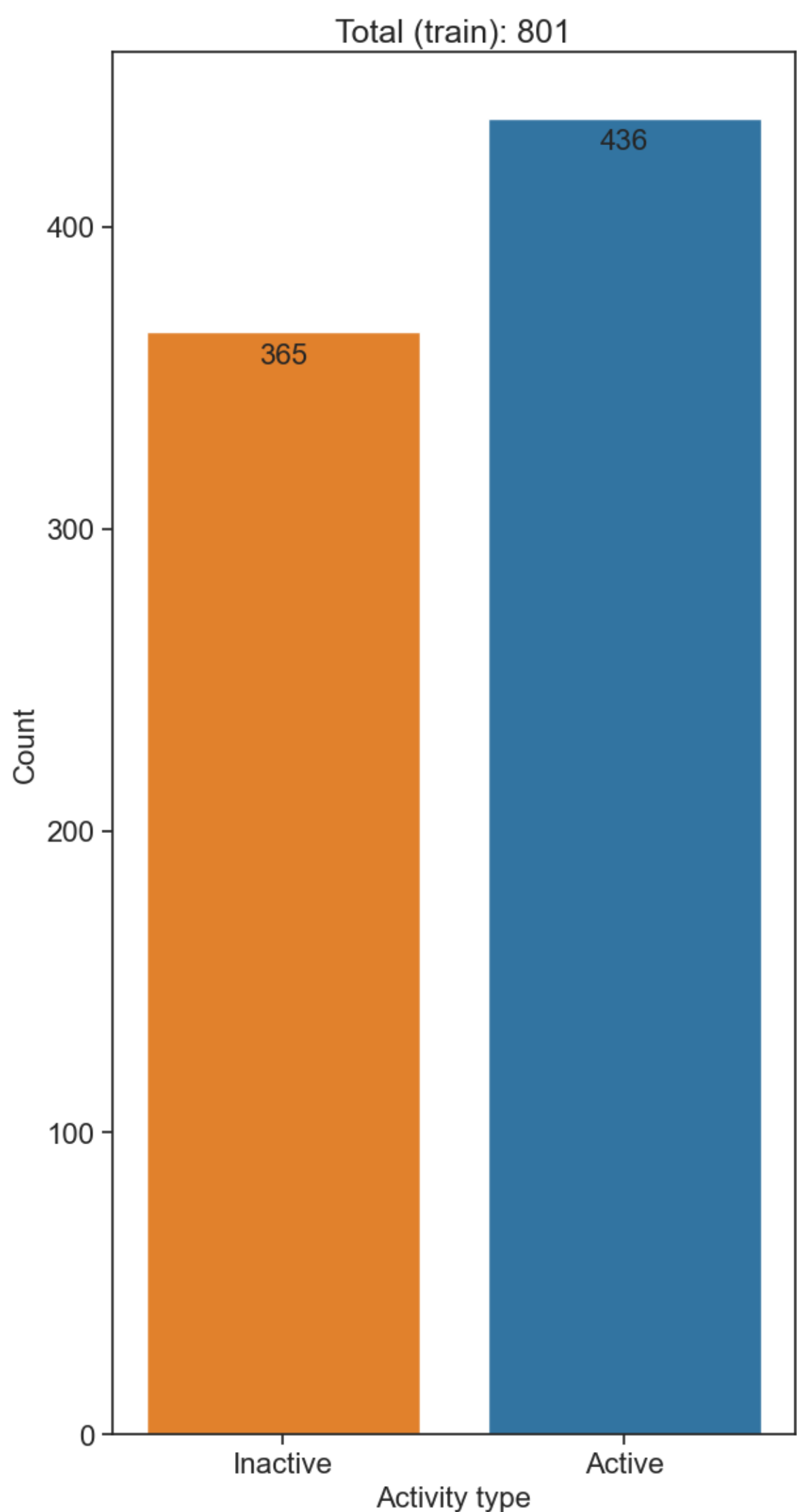
Metric	Train	Test
AUC	0.9994	0.9565
accuracy	0.9825	0.8806
balanced accuracy	0.9808	0.8797
recall_(sens)	1.0	0.8899
specificity	0.9616	0.8696
precision	0.9689	0.8899
f1_score	0.9825	0.8806