Protein: P18825 - metric: balanced_accuracy

Total (train): 375

Active: 102
Inactive: 273

ROC curve

- AUC_train = 1.000
- AUC_test = 0.943
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2704)

x-axis: Score balanced_accuracy (Test)
y-axis: Difference in Train-Test balanced_accuracy
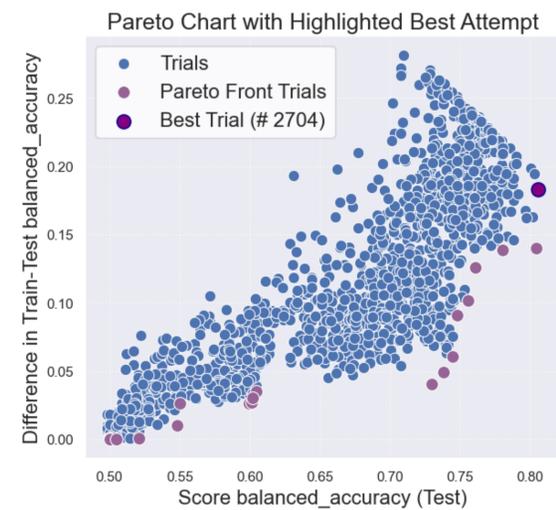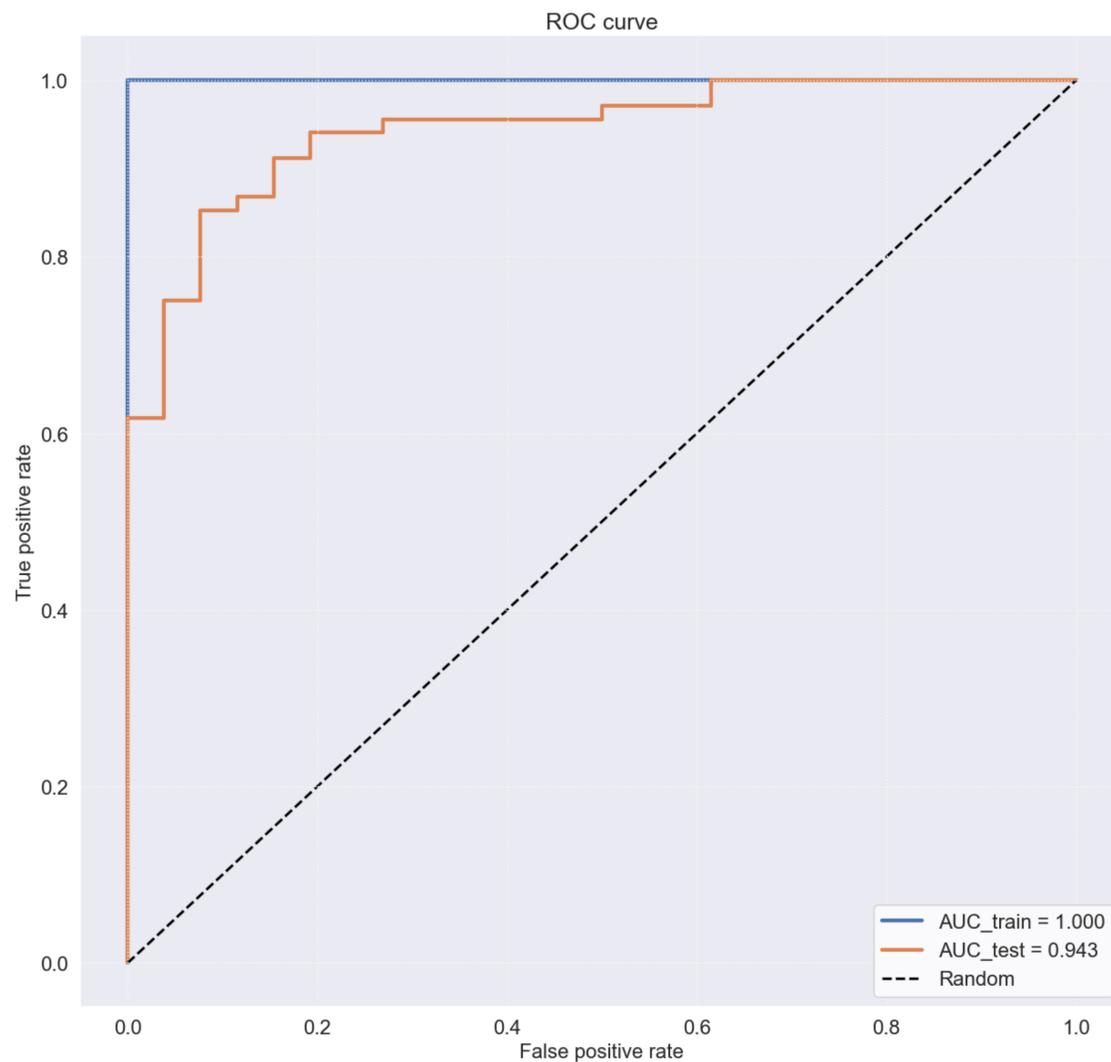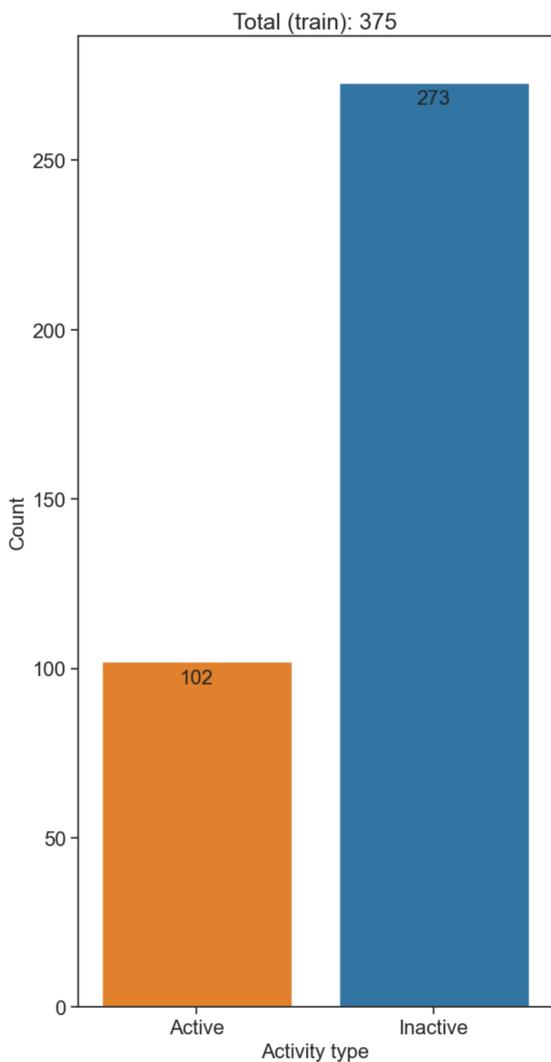
XGBoost Parameters	Value
alpha	2.7618
colsample_bytree	0.8182
gamma	0.9483
lambda	2.8891
learning_rate	0.9165
max_depth	9.0
min_child_weight	1.0
n_estimators	233.0
reg_alpha	0.3858
reg_lambda	0.1582
scale_pos_weight	4.3981
subsample	0.5694

Confusion Matrix

	0	1
0	19	7
1	4	64

Confusion Matrix (nomr.)

	0	1
0	0.73	0.27
1	0.059	0.94

Calibration curve

- train: 0.0061
- test: 0.0927
- perfectly_calibrated

Active/Inactive probability

Inactive zone | Active zone

type:
- active (68)
- inactive (26)
- decoy (94)
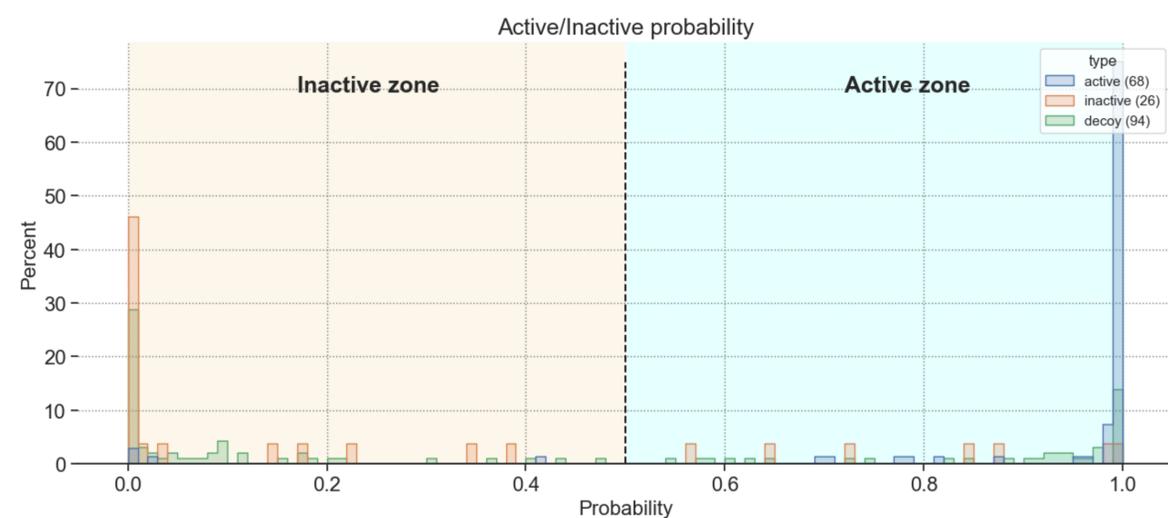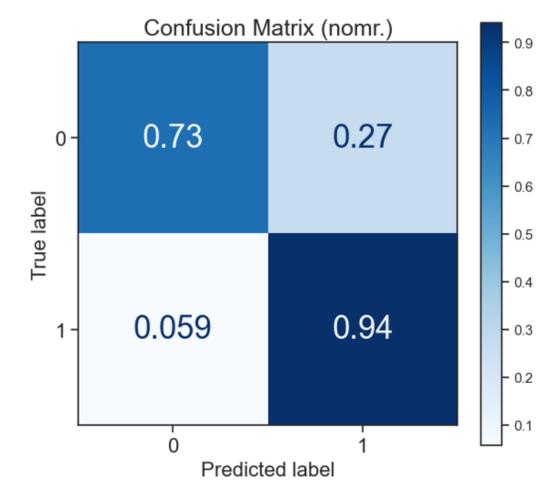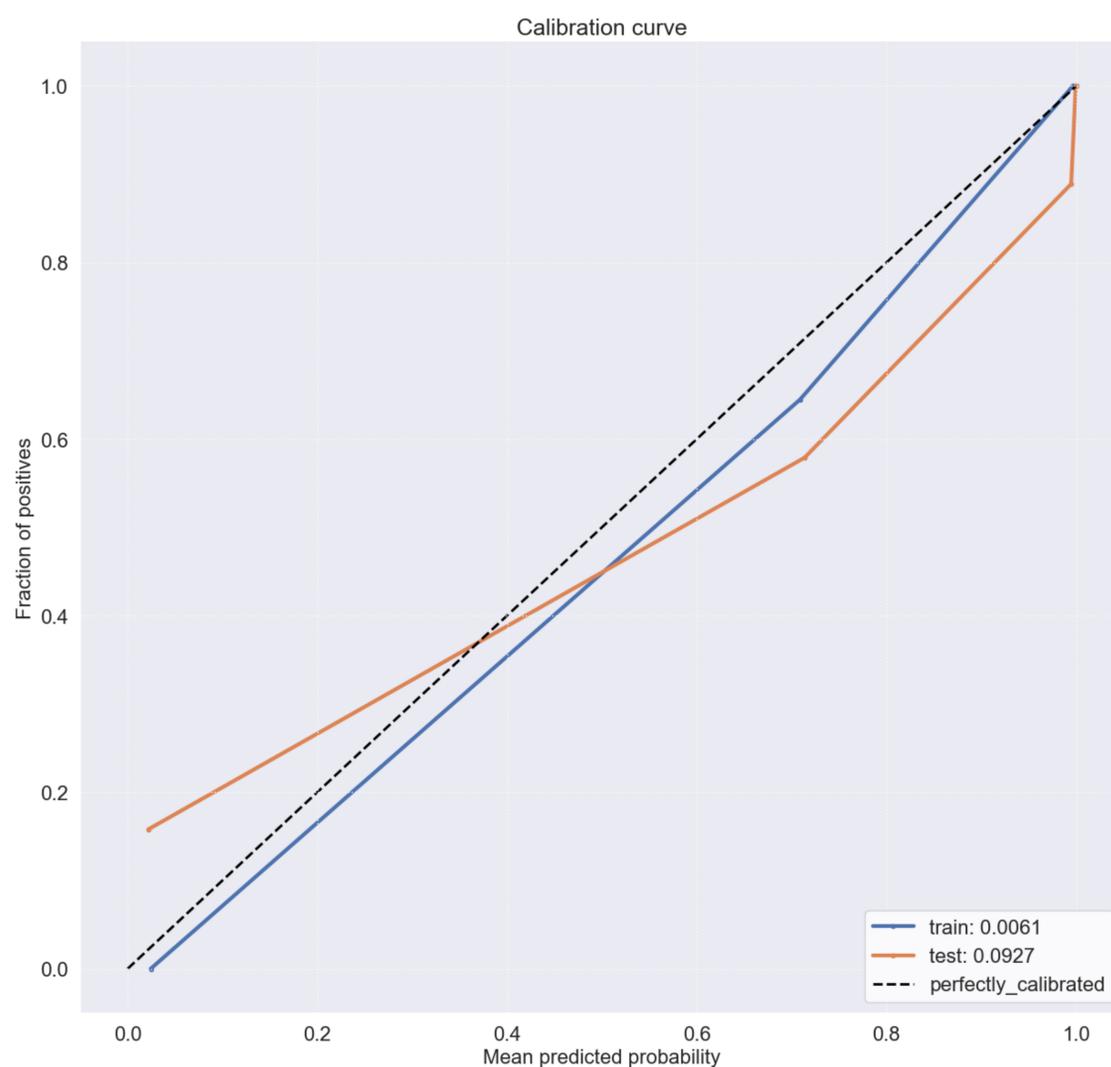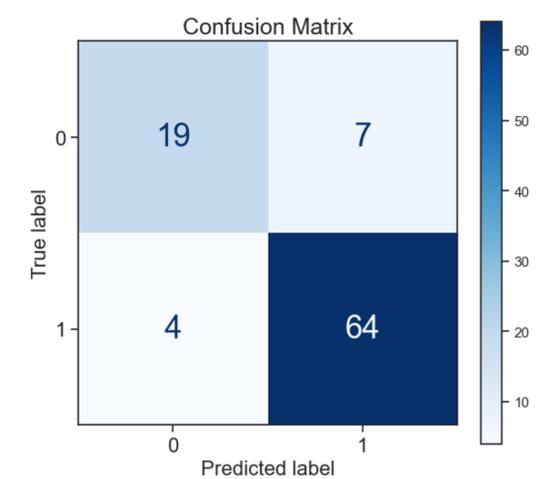
Metric	Train	Test
AUC	1.0	0.9434
accuracy	0.9947	0.883
balanced accuracy	0.9902	0.836
recall_(sens)	1.0	0.9412
specificity	0.9804	0.7308
precision	0.9927	0.9014
f1_score	0.9946	0.8807

Protein: P18825 - metric: f1

Total (train): 375

Active: 102
Inactive: 273

ROC curve

- AUC_train = 1.000
- AUC_test = 0.913
- Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 2986)
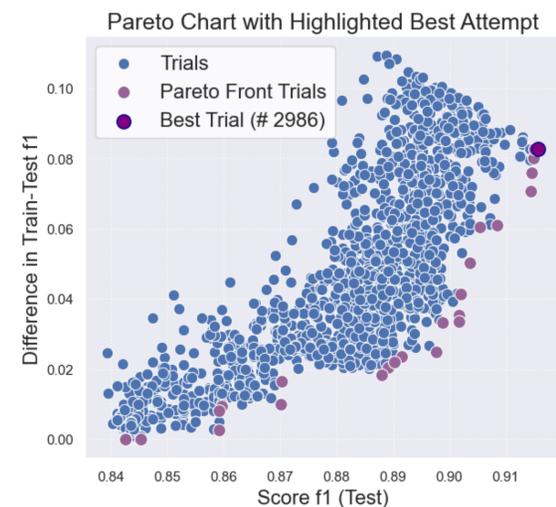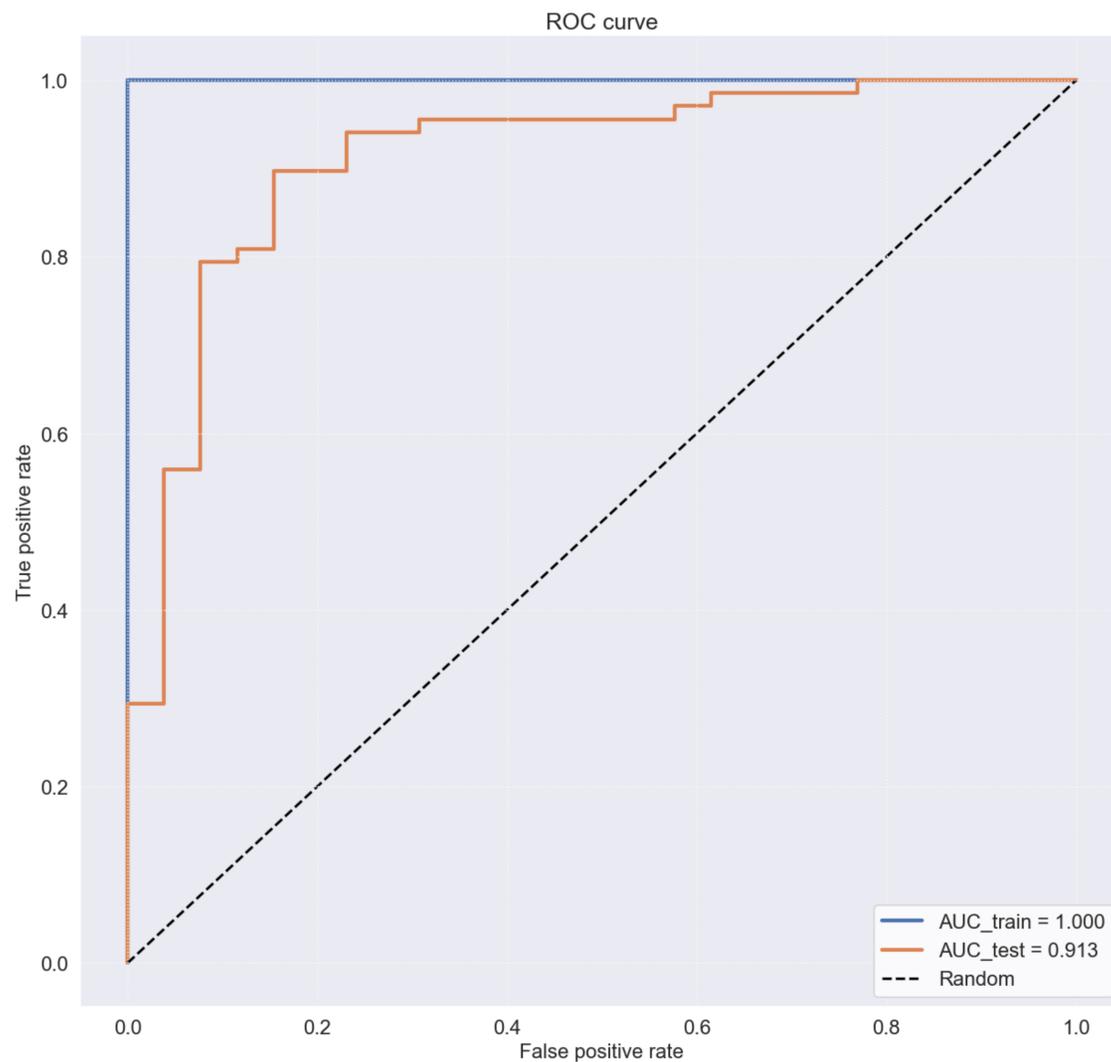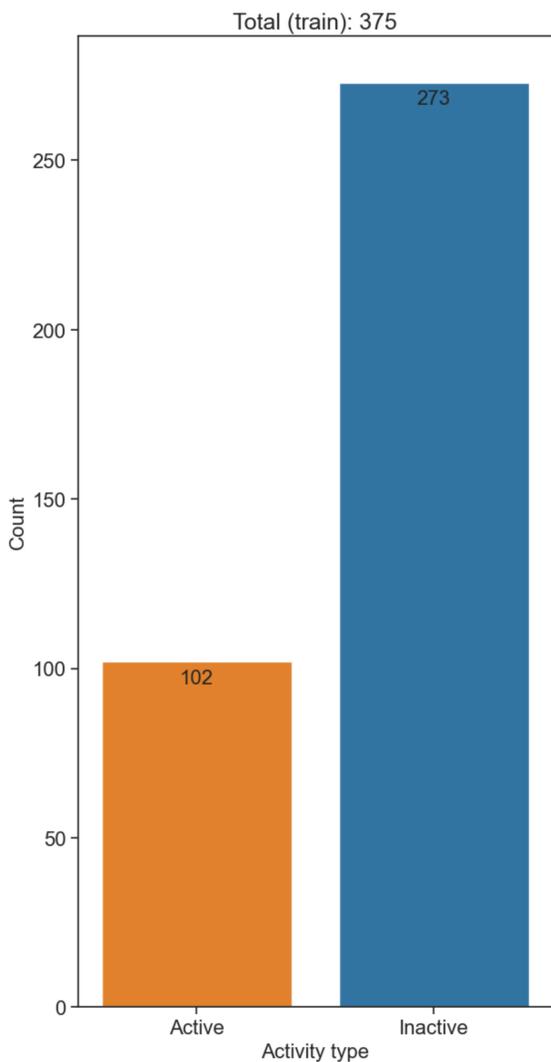
XGBoost Parameters	Value
alpha	0.5314
colsample_bytree	0.7708
gamma	0.5735
lambda	3.5624
learning_rate	0.446
max_depth	9.0
min_child_weight	1.0
n_estimators	778.0
reg_alpha	0.1799
reg_lambda	0.079
scale_pos_weight	61.435
subsample	0.5362

Confusion Matrix

	0	1
0	16	10
1	3	65

Confusion Matrix (nomr.)

	0	1
0	0.62	0.38
1	0.044	0.96

Calibration curve

- train: 0.004
- test: 0.1251
- perfectly_calibrated

Active/Inactive probability

- active (68)
- inactive (26)
- decoy (94)

Inactive zone — Active zone
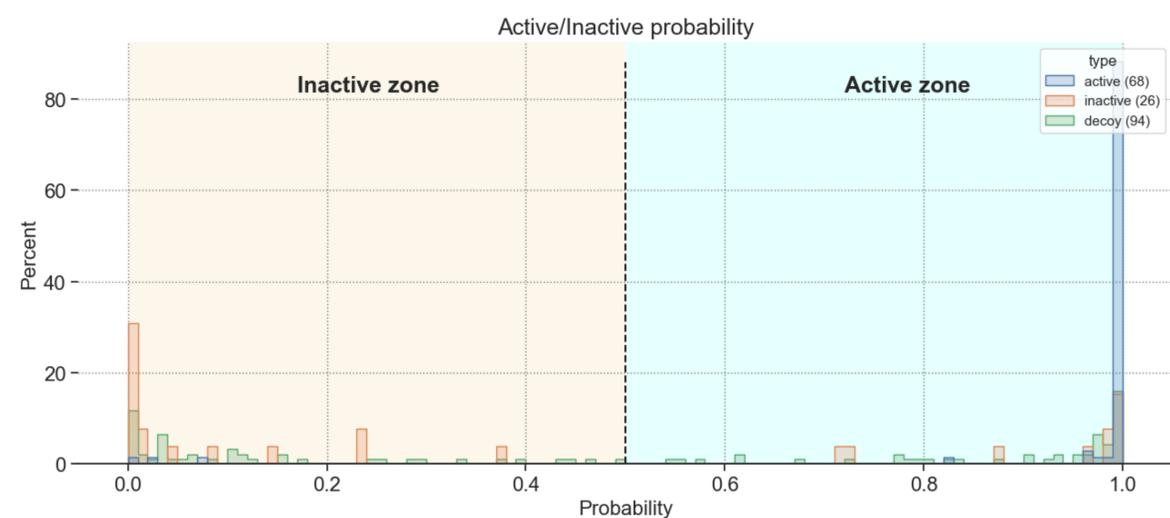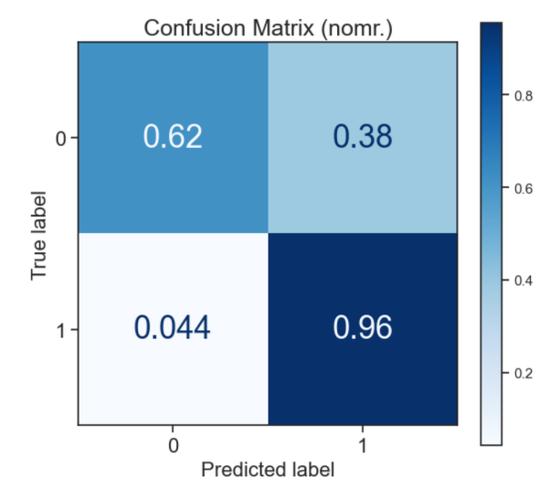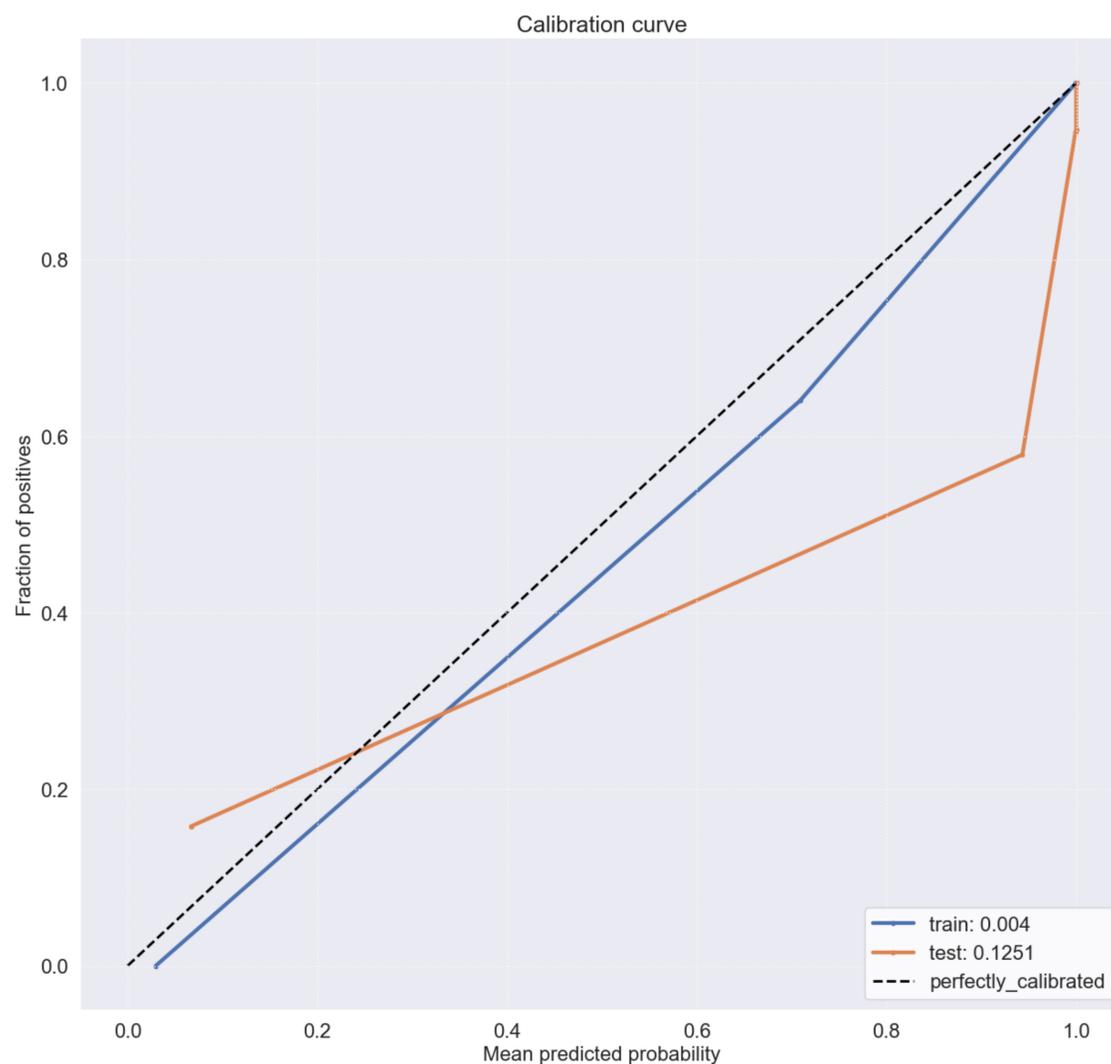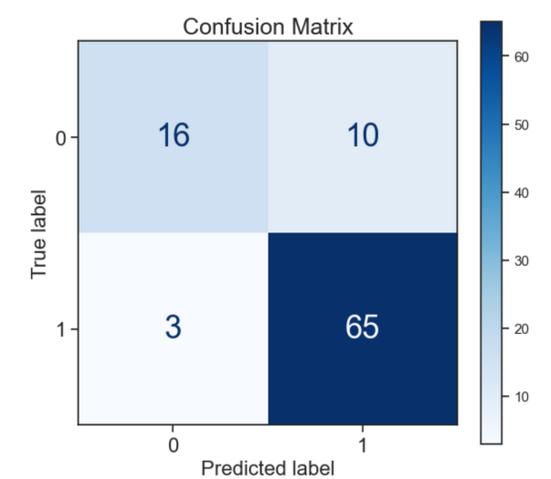
Metric	Train	Test
AUC	1.0	0.9129
accuracy	0.9973	0.8617
balanced accuracy	0.9951	0.7856
recall_(sens)	1.0	0.9559
specificity	0.9902	0.6154
precision	0.9964	0.8667
f1_score	0.9973	0.8543

Protein: P18825 - metric: precision

Total (train): 375

Active: 102
Inactive: 273

ROC curve

AUC_train = 1.000
AUC_test = 0.876
Random

Pareto Chart with Highlighted Best Attempt

- Trials
- Pareto Front Trials
- Best Trial (# 810)

x-axis: Score precision (Test)
y-axis: Difference in Train-Test precision
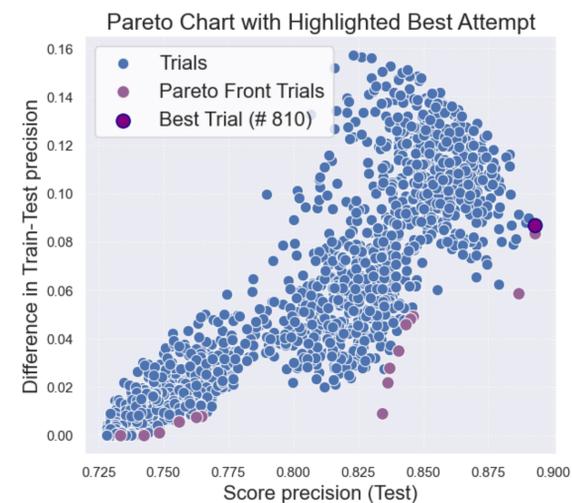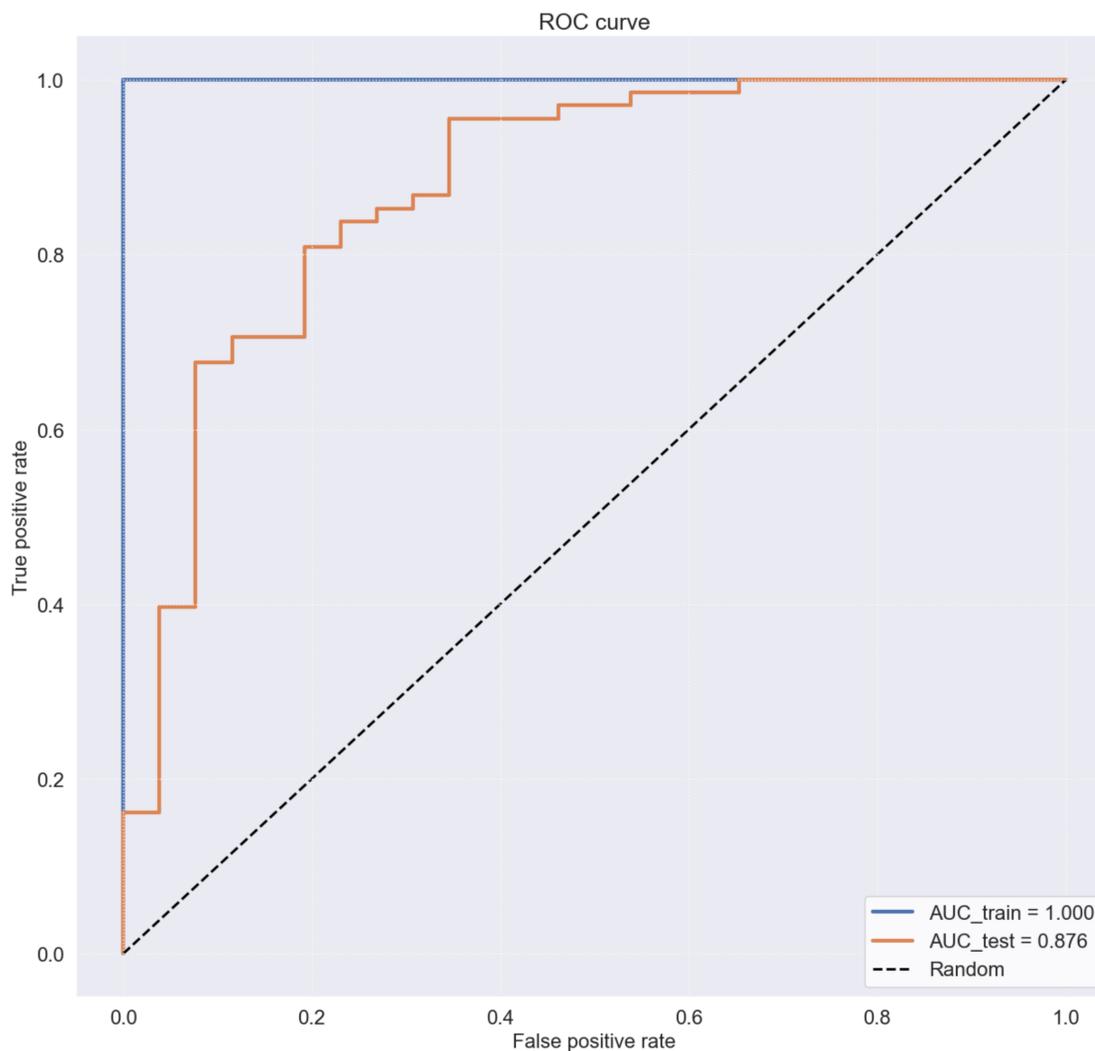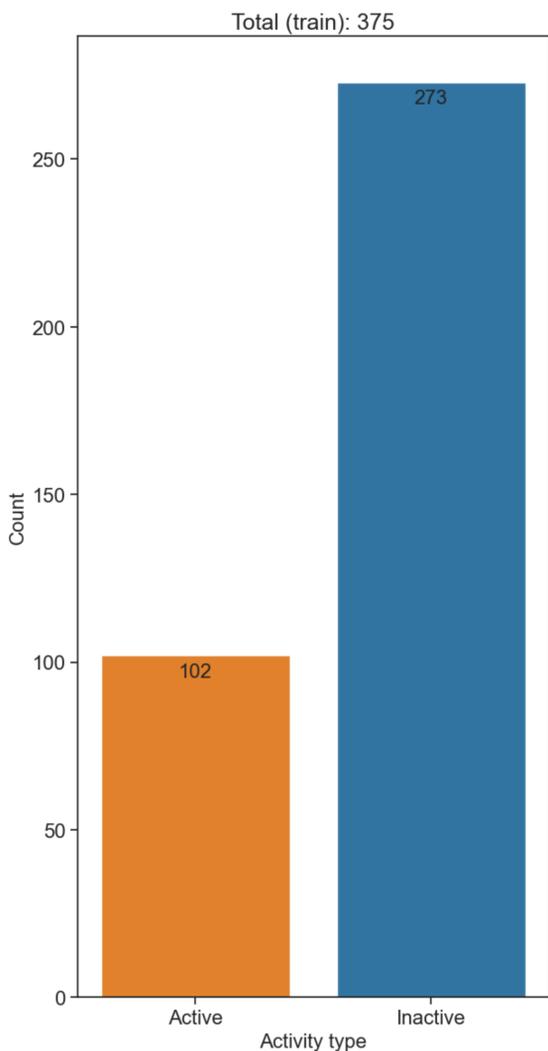
XGBoost Parameters	Value
alpha	0.4744
colsample_bytree	0.533
gamma	0.7575
lambda	1.0904
learning_rate	0.9313
max_depth	3.0
min_child_weight	2.0
n_estimators	977.0
reg_alpha	0.4535
reg_lambda	0.6041
scale_pos_weight	7.1899
subsample	0.5

Confusion Matrix

	0	1
0	17	9
1	3	65

Confusion Matrix (nomr.)

	0	1
0	0.65	0.35
1	0.044	0.96

Calibration curve

train: 0.011
test: 0.1196
perfectly_calibrated

Active/Inactive probability

type
active (68)
inactive (26)
decoy (94)

Inactive zone | Active zone
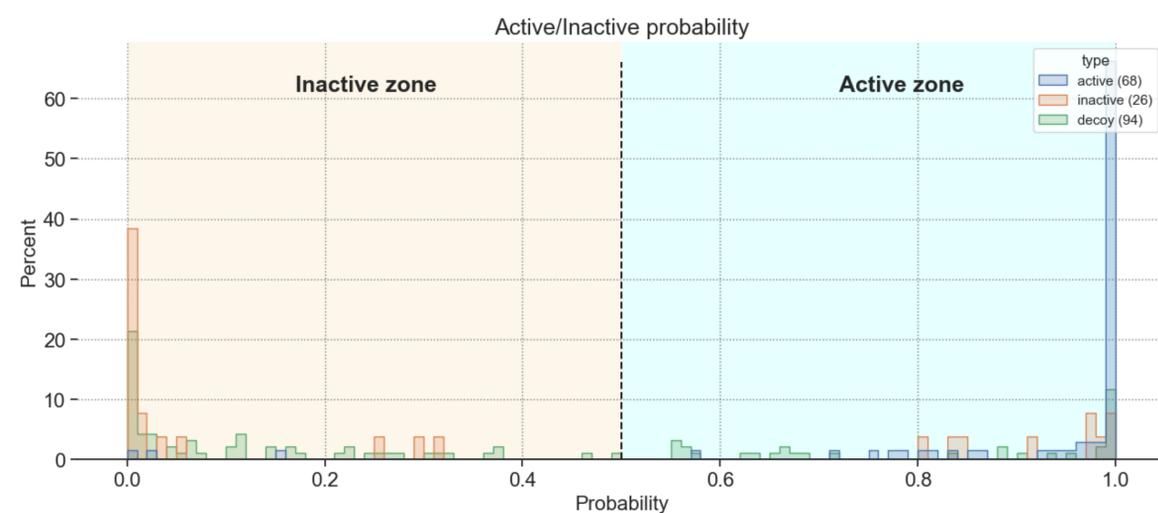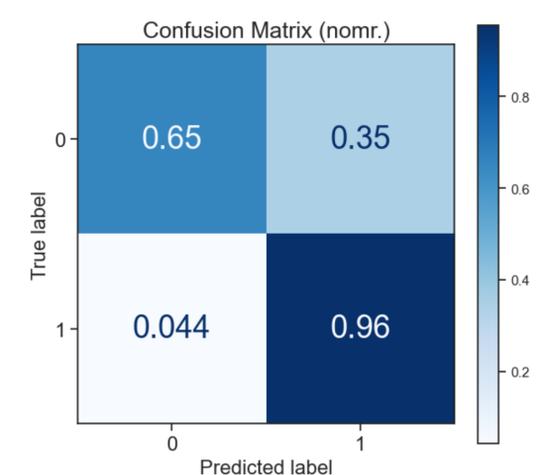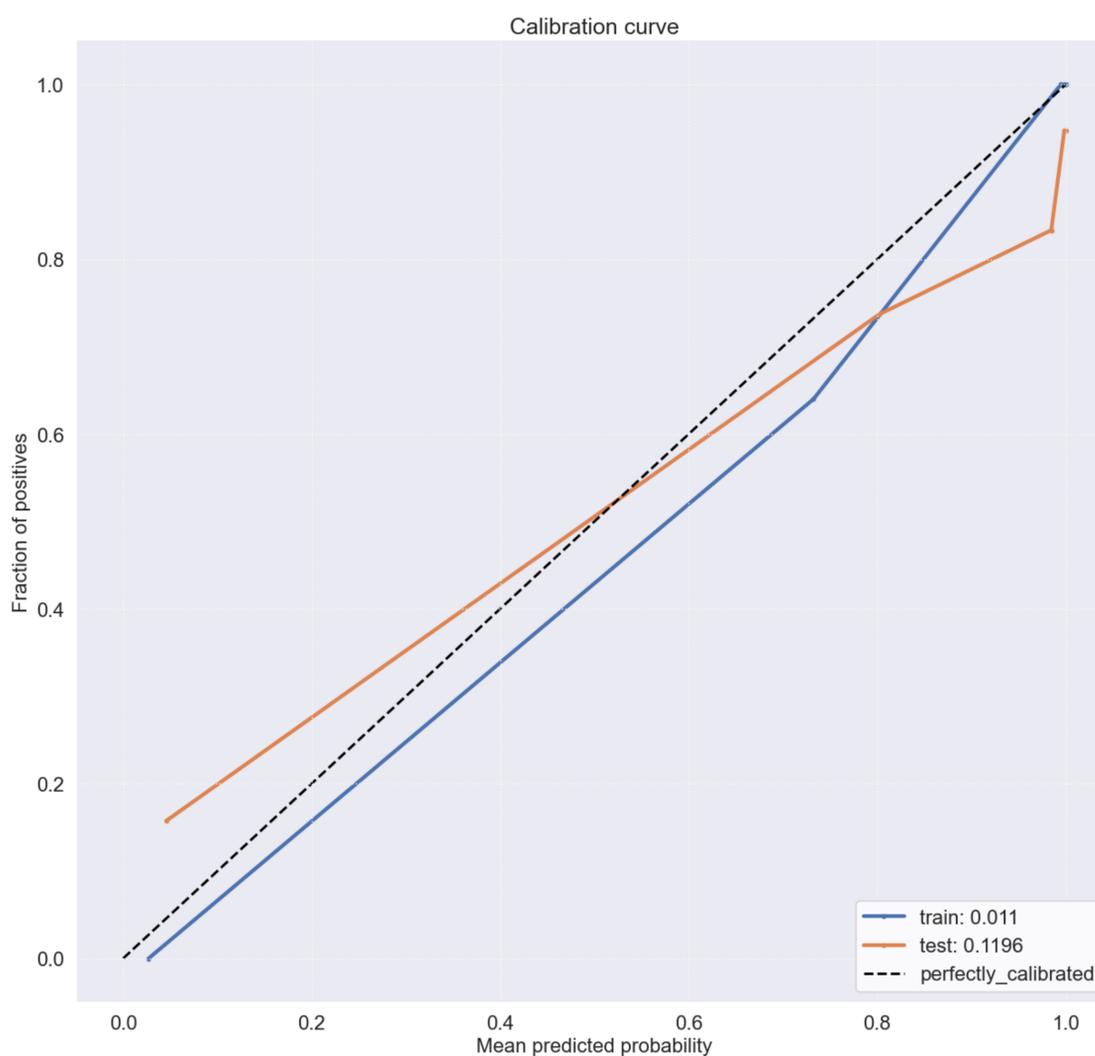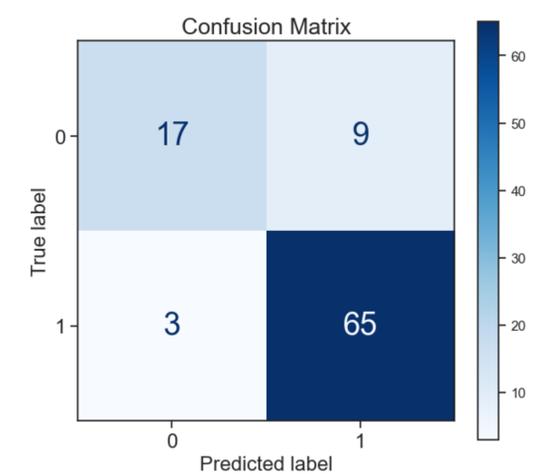
Metric	Train	Test
AUC	1.0	0.8761
accuracy	0.9893	0.8723
balanced accuracy	0.9804	0.8049
recall_(sens)	1.0	0.9559
specificity	0.9608	0.6538
precision	0.9856	0.8784
f1_score	0.9893	0.8667

Protein: P18825 - metric: SoftVotingXGBoostEnsemble
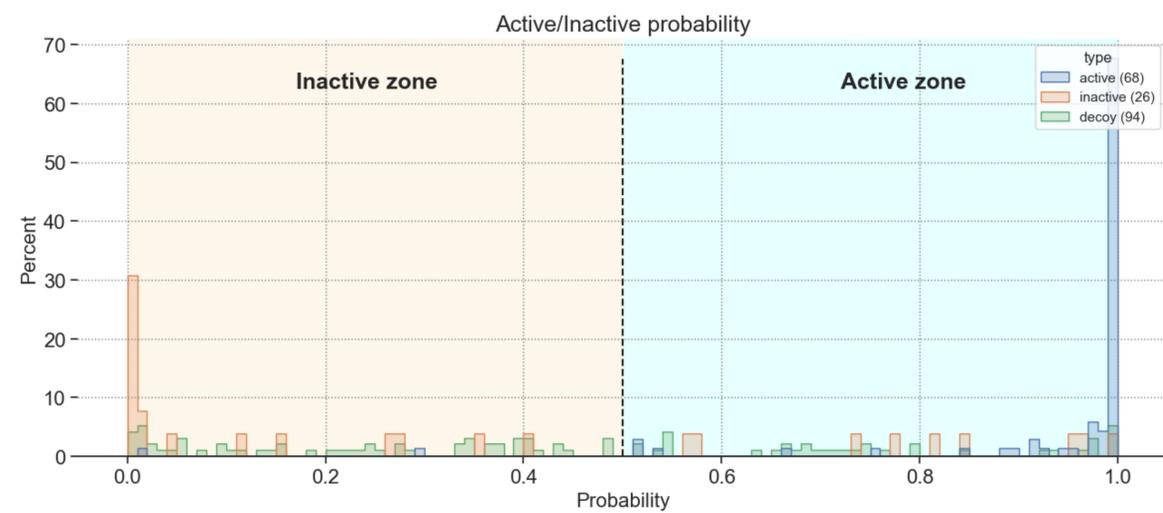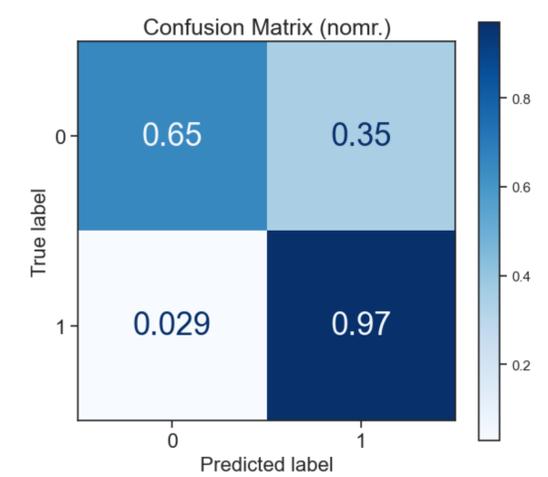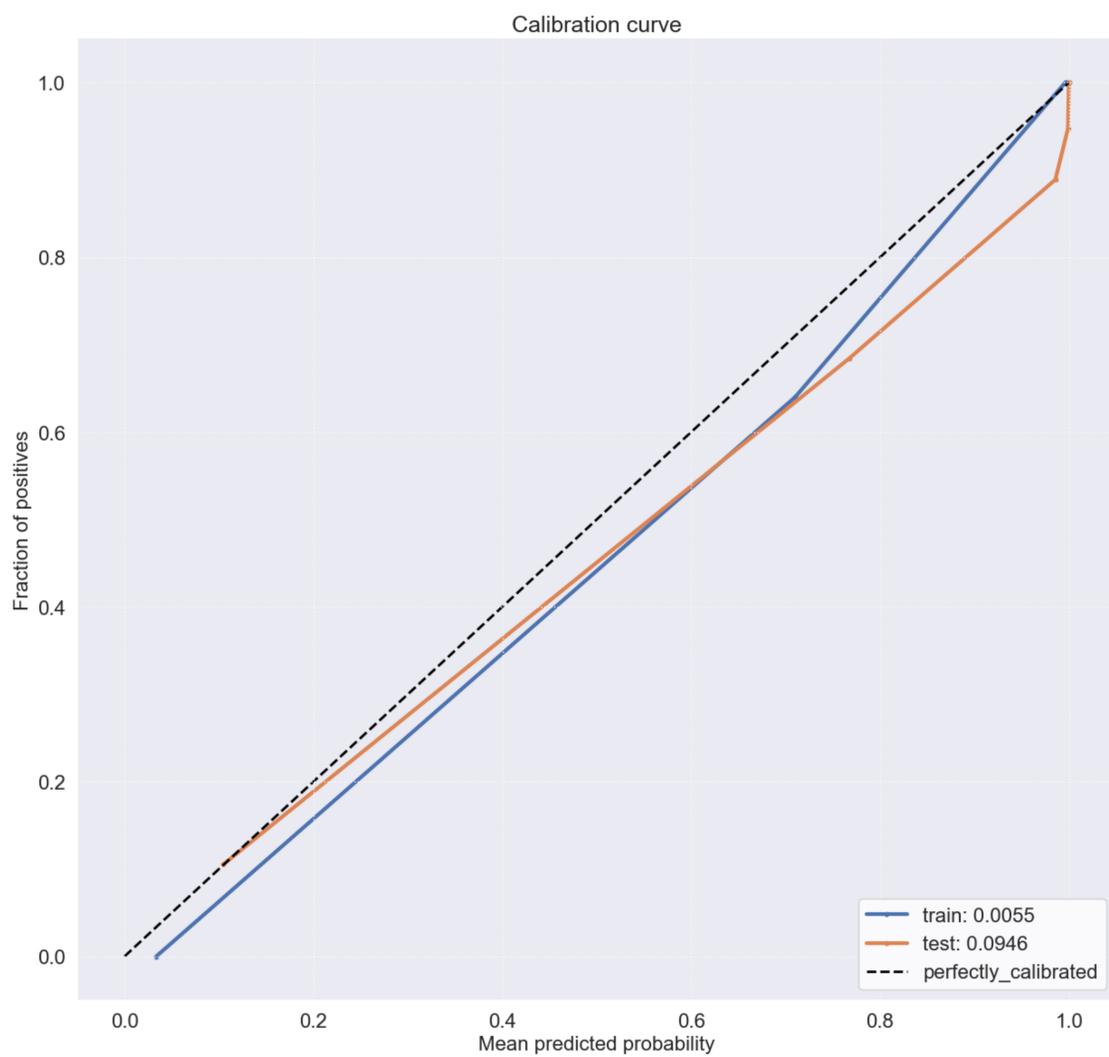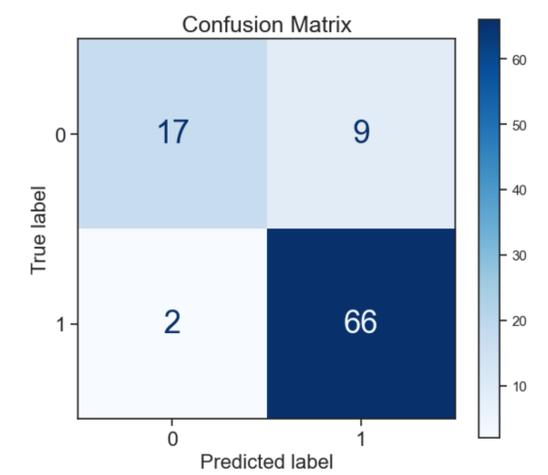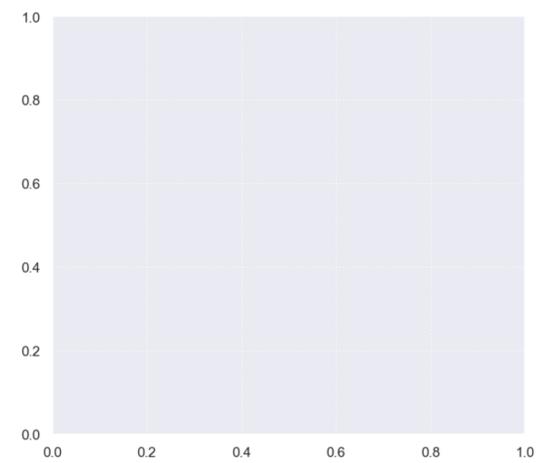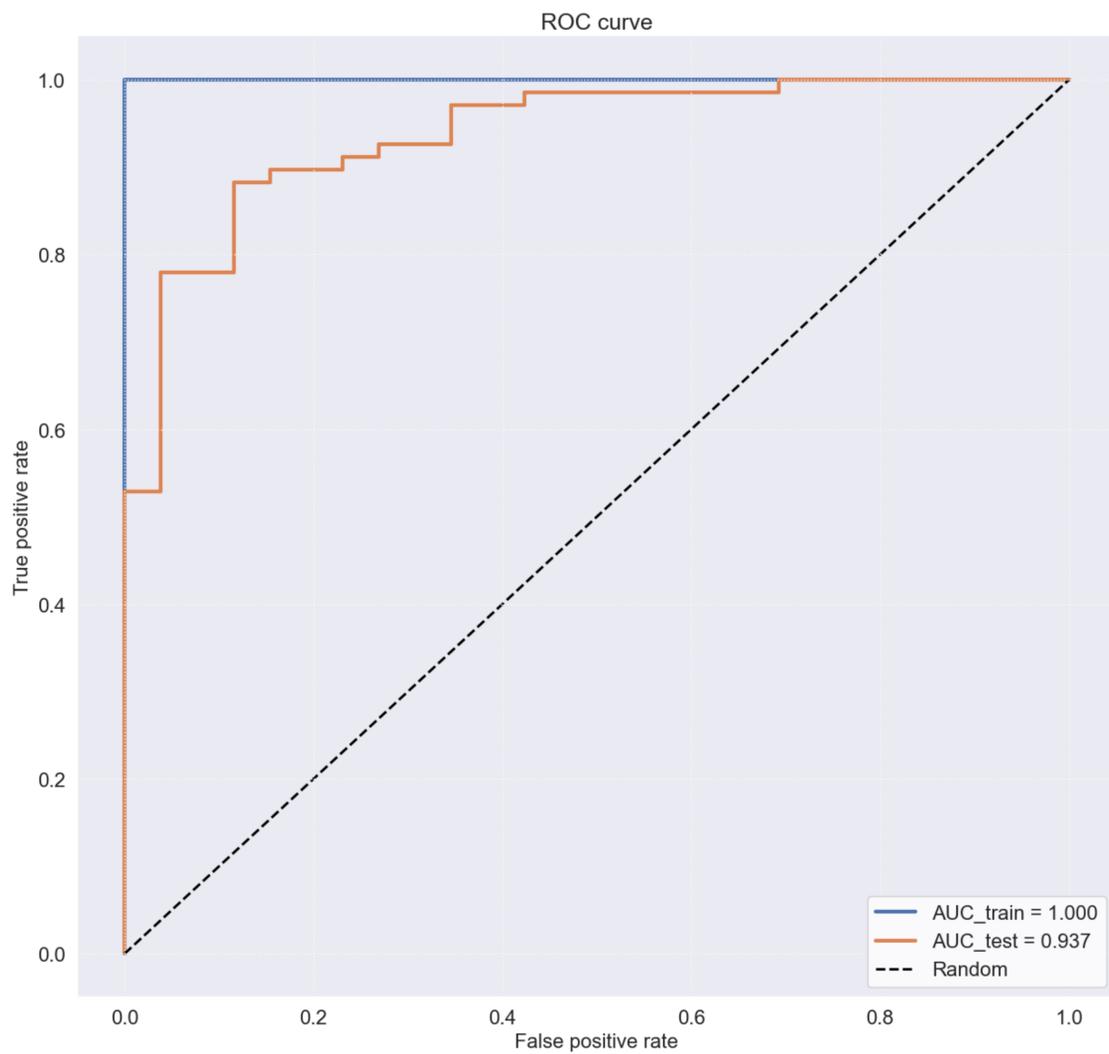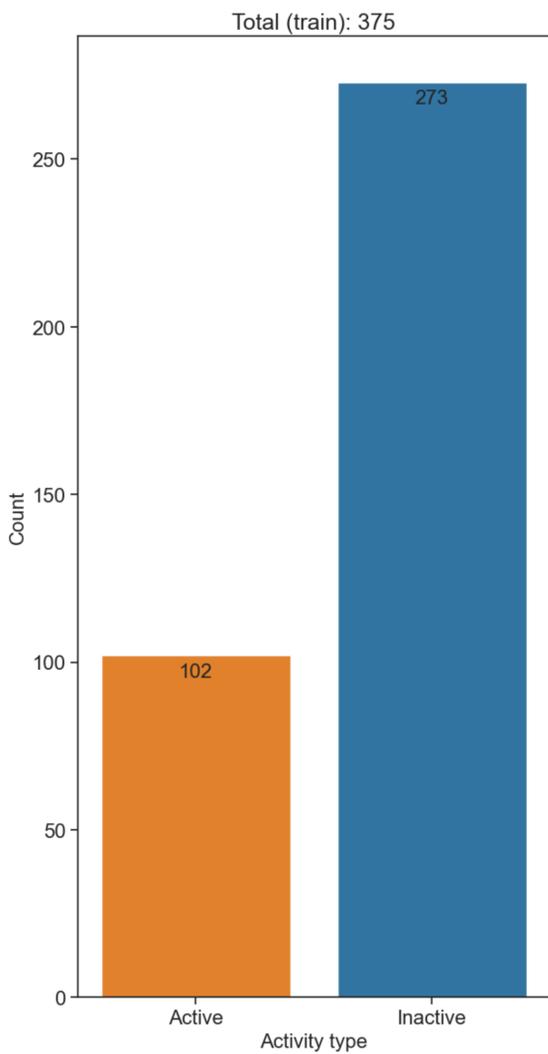
Metric	Train	Test
AUC	1.0	0.9372
accuracy	0.9947	0.883
balanced accuracy	0.9902	0.8122
recall_(sens)	1.0	0.9706
specificity	0.9804	0.6538
precision	0.9927	0.88
f1_score	0.9946	0.8767