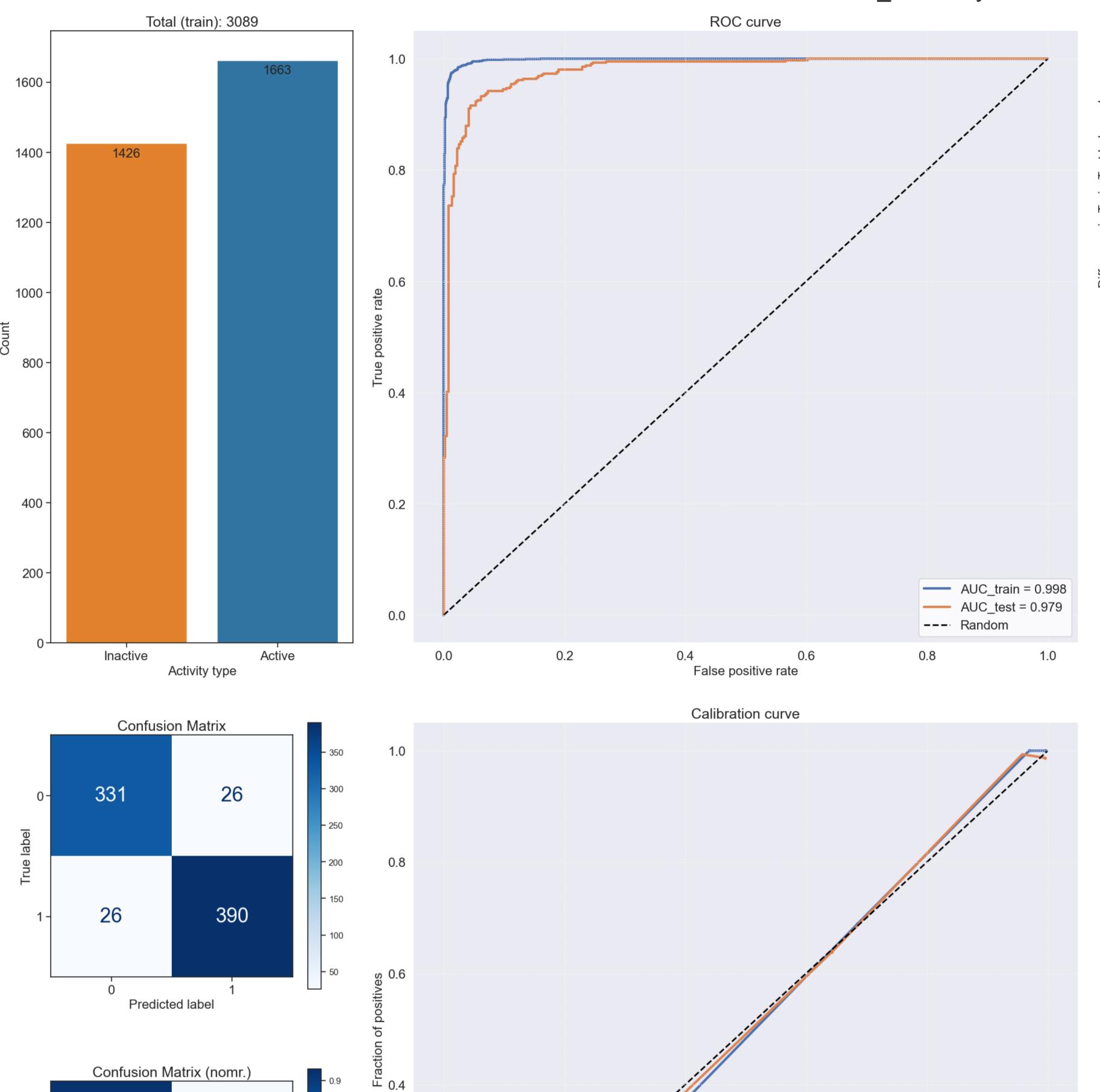
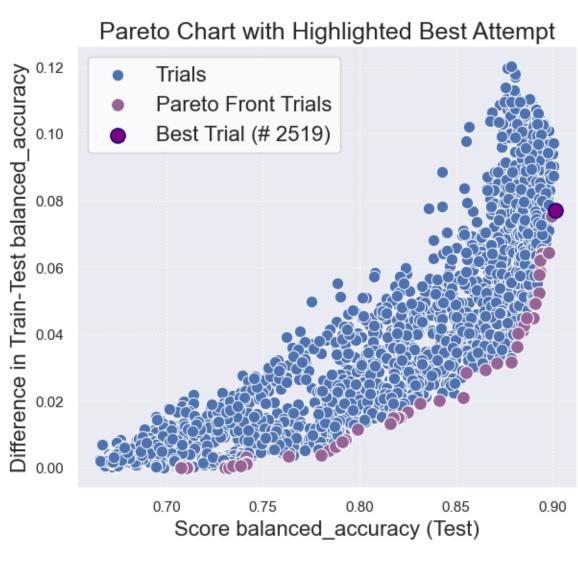
Protein: Q15858 - metric: balanced_accuracy





XGBoost Parameters	Value
alpha	4.456
colsample_bytree	0.9151
gamma	0.5433
lambda	3.1746
learning_rate	0.038
max_depth	9.0
min_child_weight	4.0
n_estimators	943.0
reg_alpha	0.5882
reg_lambda	0.2135
scale_pos_weight	1.1813
subsample	0.5998

	Active/Inactive probability					
40 	IIId			Active	zone ac	/pe tive (416) active (357) coy (765)
30						
Percent 50						
15						
5						
0	0.0	.2 0	0.4 Probability	0.6).8	1.0
	1 Tobability					

- 0.8

0.6

- 0.5

0.4

0.3

0.2

0.2

0.0

0.0

0.2

Confusion Matrix (nomr.)

Predicted label

0.073

0.94

0.93

0.062

0-

True label

Metric	Train	Test
AUC	0.9982	0.9786
accuracy	0.9783	0.9327
balanced accuracy	0.9775	0.9323
recall_(sens)	0.9886	0.9375
specificity	0.9663	0.9272
precision	0.9716	0.9375
f1_score	0.9783	0.9327

--- train: 0.021

0.8

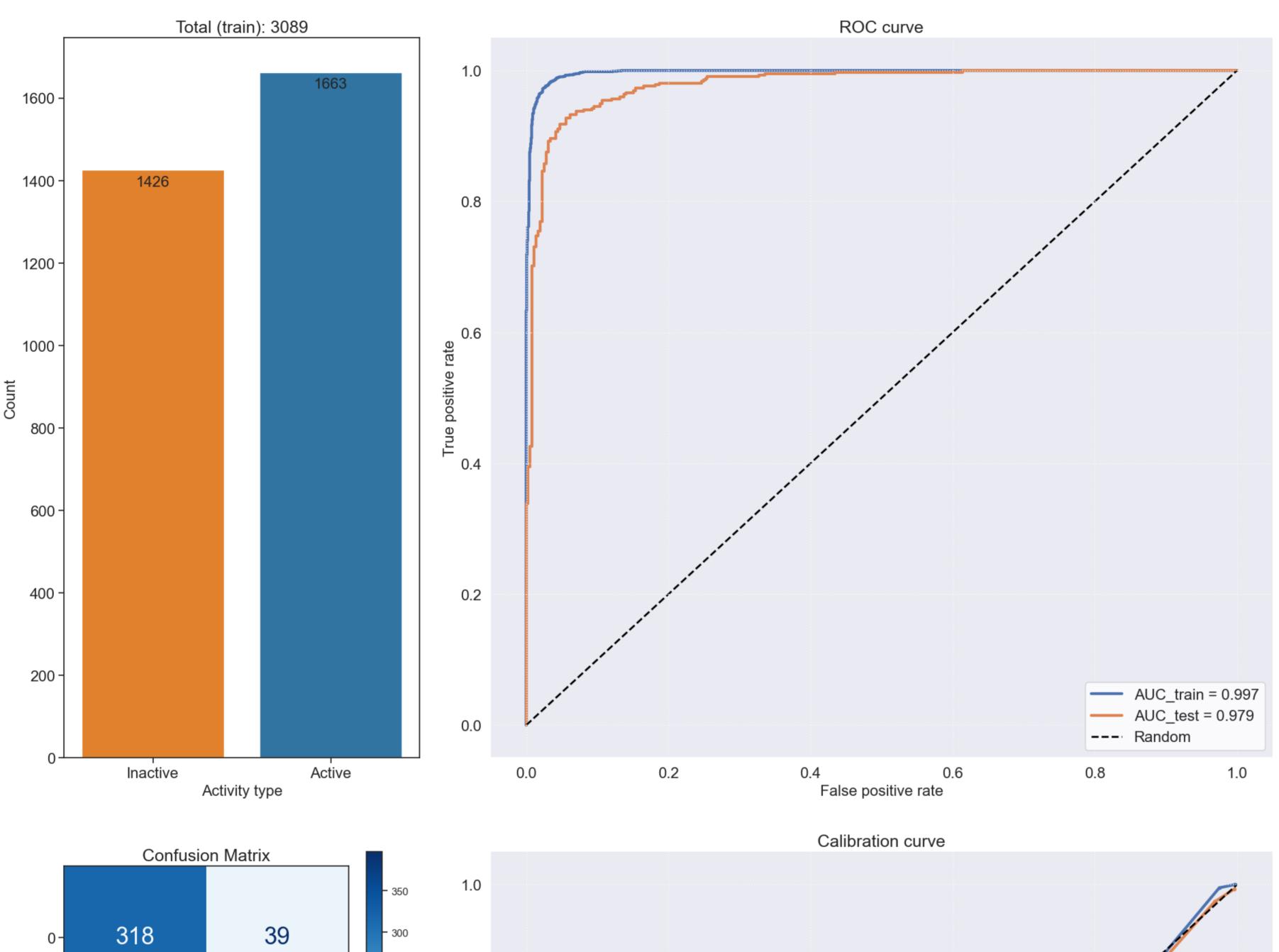
0.4 0.6 Mean predicted probability

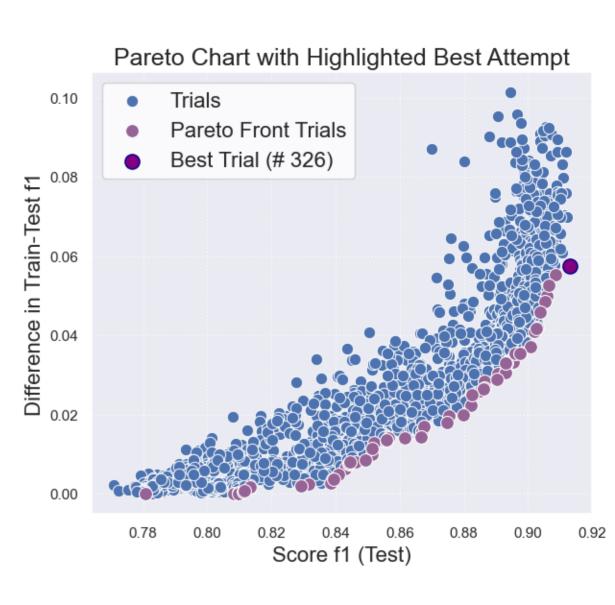
test: 0.0529

--- perfectly_calibrated

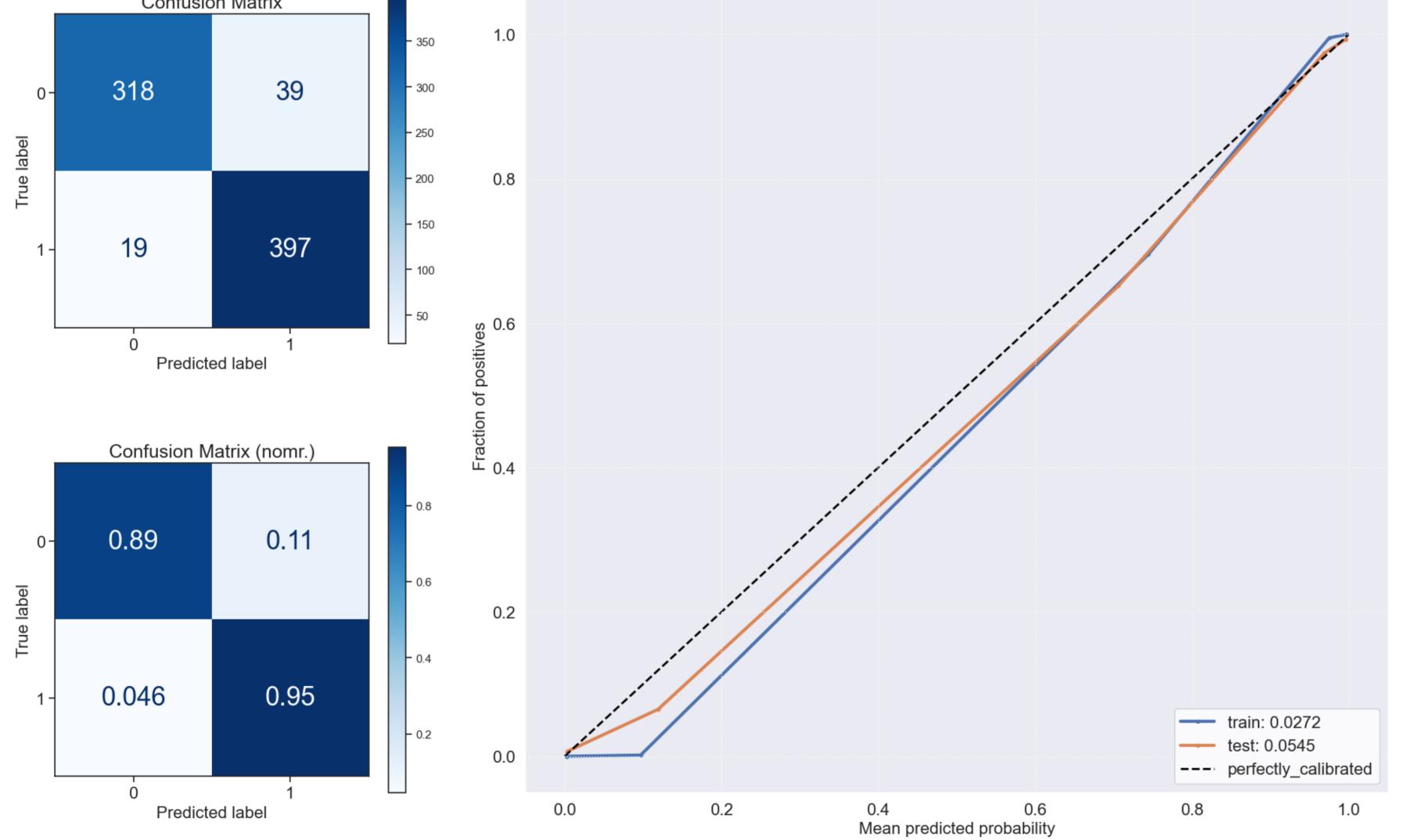
1.0

Protein: Q15858 - metric: f1





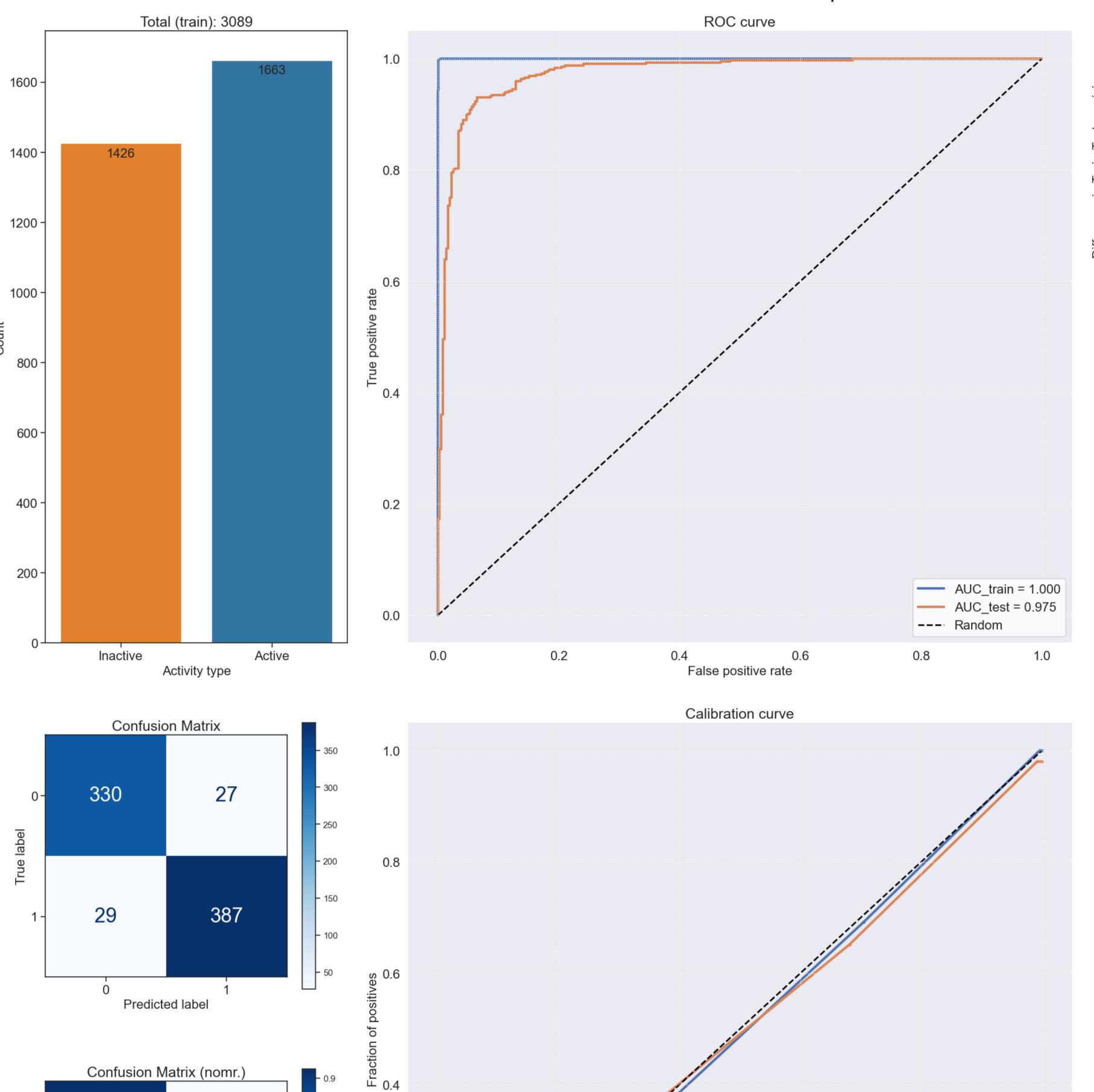
XGBoost Parameters	Value
alpha	0.5619
colsample_bytree	0.586
gamma	0.7775
lambda	0.0852
learning_rate	0.1151
max_depth	8.0
min_child_weight	7.0
n_estimators	522.0
reg_alpha	0.3414
reg_lambda	0.8676
scale_pos_weight	1.9112
subsample	0.7813

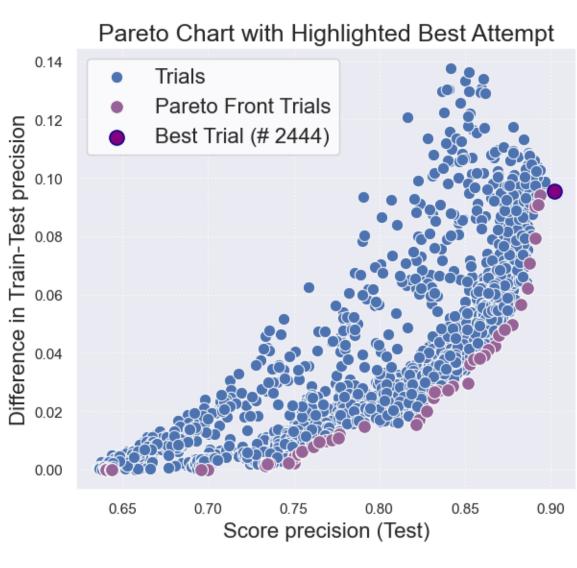


	Ac	tive/Inactive probab	ility		
40 -	Inactive zone		Active :	LOTTO	e re (416) tive (357)
30				decc	oy (765)
25					
Percent - 05					
15					
10 -					
0					
0.0	0 0.2 0	0.4 Probability	.6 0	.8	1.0

Metric	Train	Test
AUC	0.997	0.9789
accuracy	0.9657	0.925
balanced accuracy	0.9634	0.9225
recall_(sens)	0.9934	0.9543
specificity	0.9334	0.8908
precision	0.9456	0.9106
f1_score	0.9656	0.9248

Protein: Q15858 - metric: precision





XGBoost Parameters	Value
alpha	4.3526
colsample_bytree	0.6306
gamma	0.0343
lambda	3.3901
learning_rate	0.3818
max_depth	7.0
min_child_weight	2.0
n_estimators	699.0
reg_alpha	0.9155
reg_lambda	0.5837
scale_pos_weight	1.7113
subsample	0.5097

Active/Inactive probability						
60	Ina	ictive zone		Active	type active (41 inactive (41 decoy (76	357)
40						
Percent - 08						
20						
10						
0	0.0	.2 0.		0.6	0.8 1.0)
Probability						

- 0.8

- 0.5

0.4

- 0.3

0.2

0.1

0.2

0.0

0.0

0.2

Confusion Matrix (nomr.)

Predicted label

0.076

0.93

0.92

0.07

0-

True label

Metric	Train	Test
AUC	1.0	0.9745
accuracy	0.9981	0.9276
balanced accuracy	0.998	0.9273
recall_(sens)	0.9988	0.9303
specificity	0.9972	0.9244
precision	0.9976	0.9348
f1_score	0.9981	0.9276

--- train: 0.0033

0.8

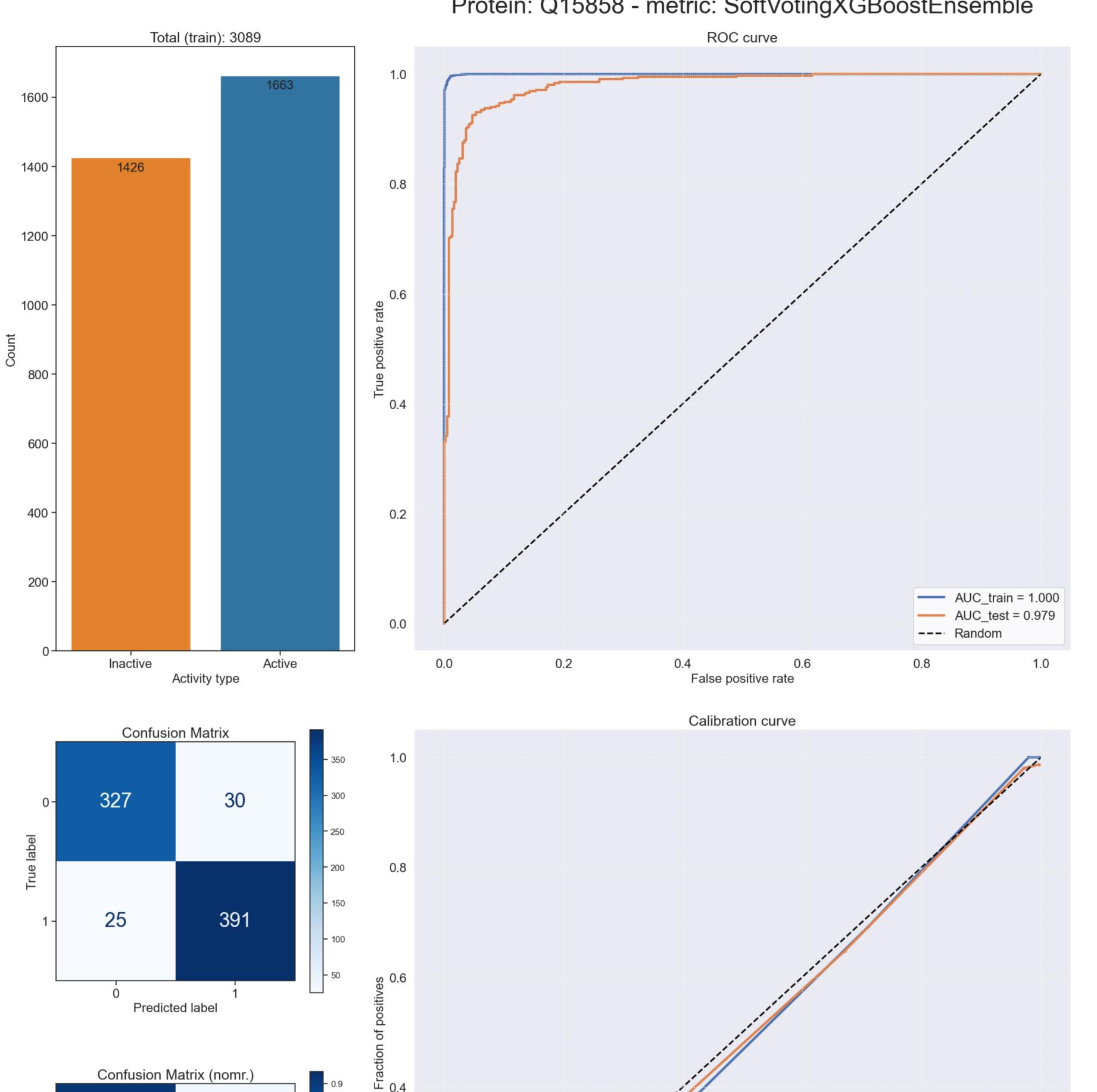
0.4 0.6 Mean predicted probability

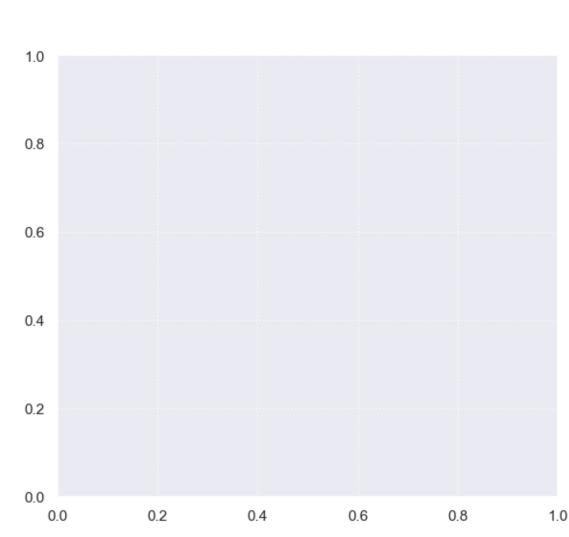
test: 0.0568

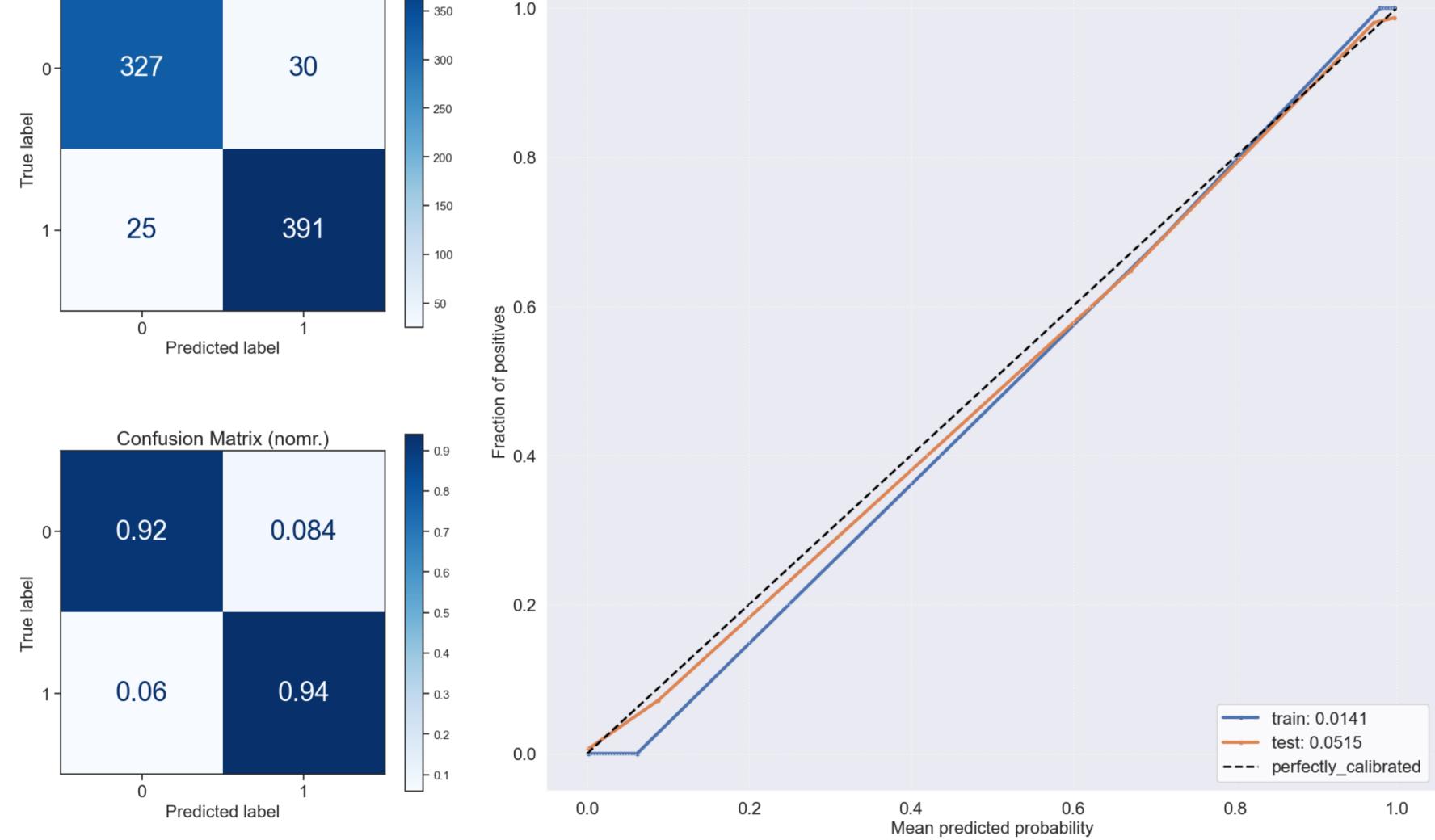
--- perfectly_calibrated

1.0

Protein: Q15858 - metric: SoftVotingXGBoostEnsemble







	Active/Inactive probability					
40	Ina	active zone			Active zone	type active (416) inactive (357) decoy (765)
30						
Percent 00						
10						
0(0.0	0.2	0.4 Probab	0.6 ility	0.8	1.0

Metric	Train	Test
AUC	0.9997	0.979
accuracy	0.9896	0.9288
balanced accuracy	0.9889	0.9279
recall_(sens)	0.9982	0.9399
specificity	0.9797	0.916
precision	0.9828	0.9287
f1_score	0.9896	0.9288