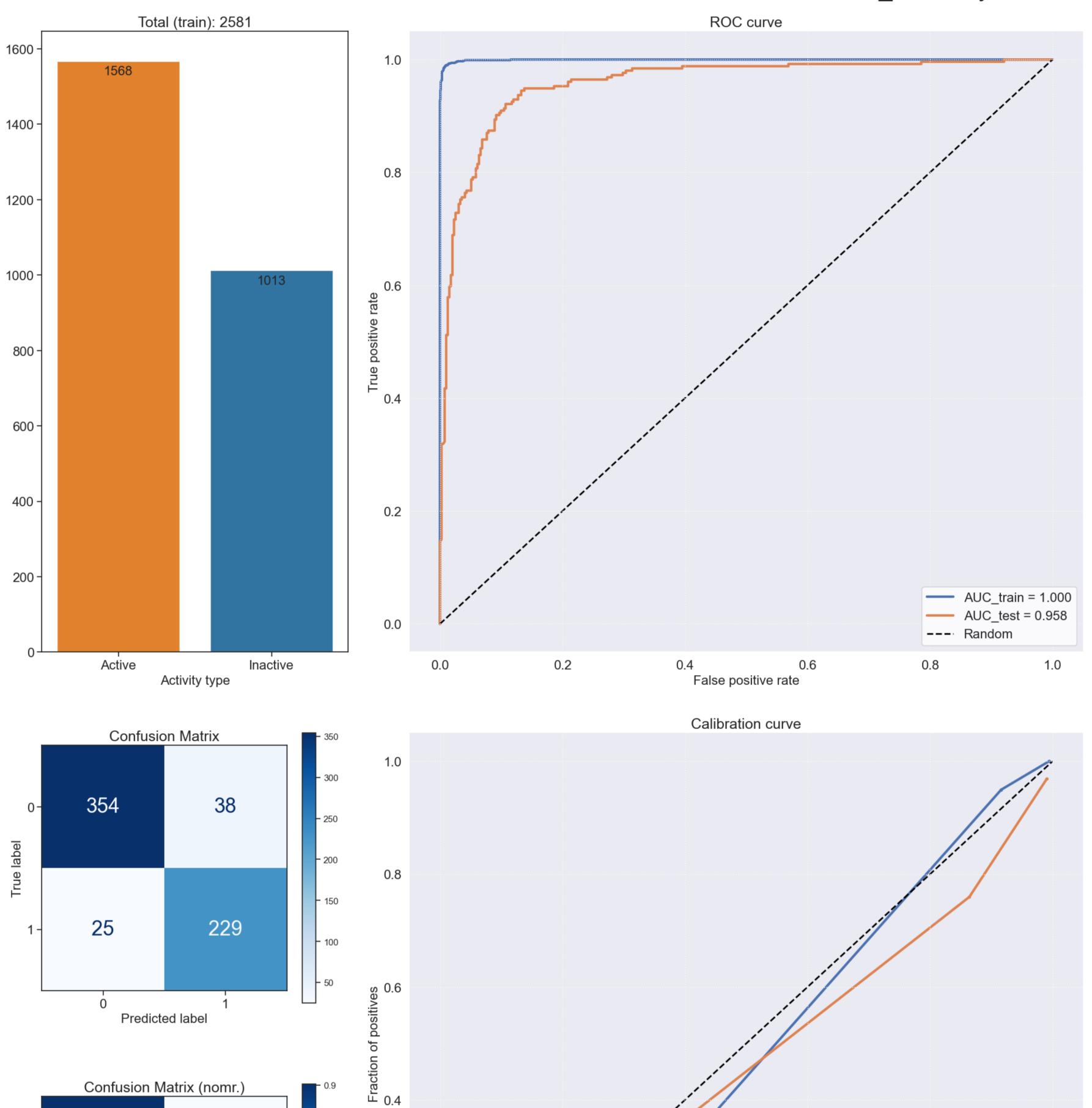
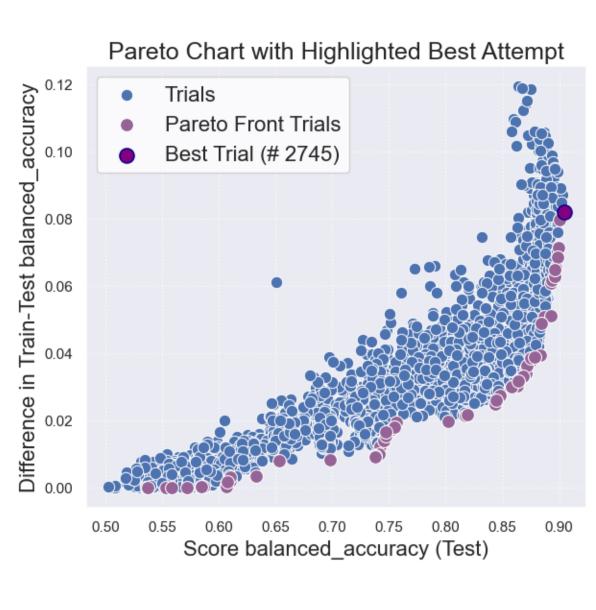
Protein: P31645 - metric: balanced_accuracy





XGBoost Parameters	Value
alpha	0.4346
colsample_bytree	0.5581
gamma	0.5752
lambda	1.3507
learning_rate	0.0995
max_depth	9.0
min_child_weight	1.0
n_estimators	904.0
reg_alpha	0.8875
reg_lambda	0.6738
scale_pos_weight	2.4352
subsample	0.6619

		Ac	tive/Inactive p	robability		
35	Ina	ctive zone			Active zone	type active (254) inactive (392) decoy (646)
30						
25						
ent						
Percent 05						
15						
10						
_						
5						
0		2	4	0.6	0.8	10
	0.0 0.2 0.4 0.6 0.8 1.0 Probability					

- 0.8

0.4

0.3

0.2

0.2

0.0

0.0

0.2

Confusion Matrix (nomr.)

Predicted label

0.097

0.9

0.9

0.098

0-

True label

Metric	Train	Test
AUC	0.9995	0.9583
accuracy	0.9841	0.9025
balanced accuracy	0.9859	0.9023
recall_(sens)	0.9941	0.9016
specificity	0.9777	0.9031
precision	0.9664	0.8577
f1_score	0.9842	0.9029

--- train: 0.0148

0.8

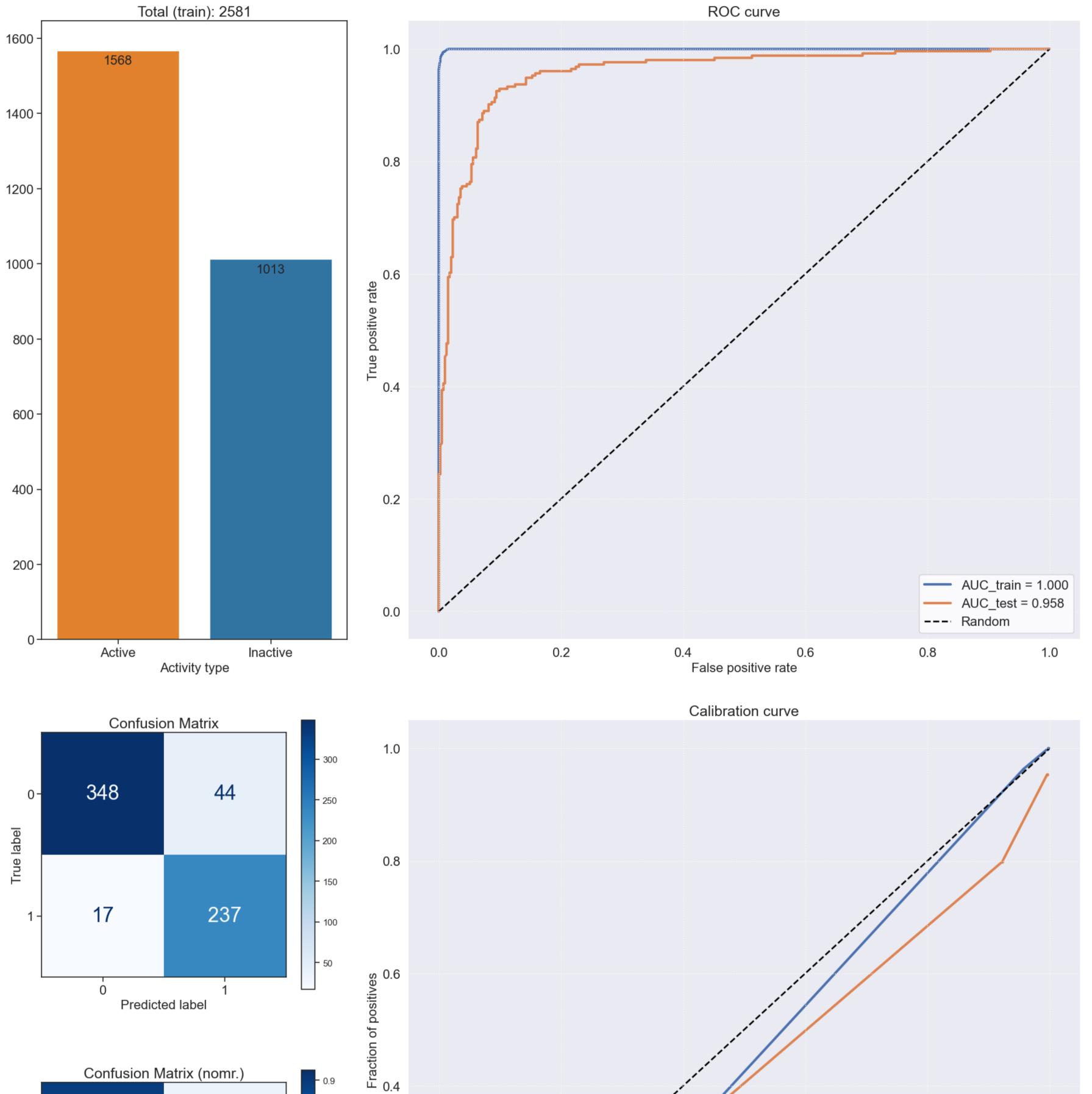
0.4 0.6 Mean predicted probability

test: 0.0749

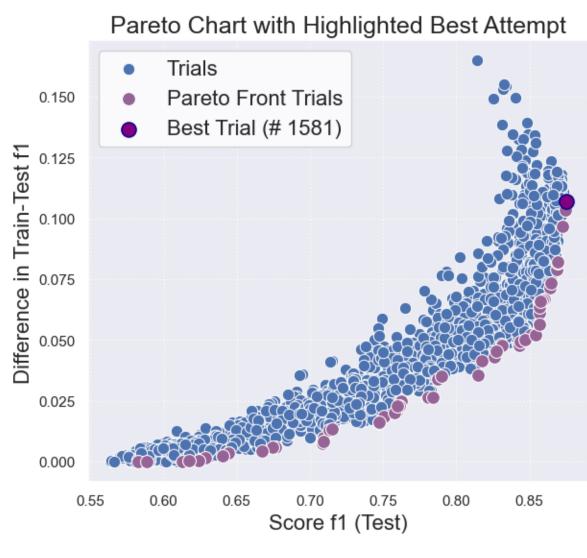
--- perfectly_calibrated

1.0

Protein: P31645 - metric: f1



0.2



XGBoost Parameters	Value
alpha	2.2783
colsample_bytree	0.5659
gamma	0.4523
lambda	4.9896
learning_rate	0.0599
max_depth	9.0
min_child_weight	1.0
n_estimators	763.0
reg_alpha	0.514
reg_lambda	0.2206
scale_pos_weight	5.131
subsample	0.6189

	Active/Inactive probability					
40	lna	ctive zone		Active :	zone a	type ctive (254) nactive (392) ecoy (646)
Percent 30						
20						
10						
0	0.0 0	.2 0	.4 (Probability	0.6 0	.8	1.0

- 0.8

- 0.5

0.4

0.3

0.2

0.1

0.2

0.0

0.0

Confusion Matrix (nomr.)

Predicted label

0.11

0.93

0.89

0.067

0-

True label

Metric	Train	Test
AUC	0.9999	0.9583
accuracy	0.9841	0.9056
balanced accuracy	0.9869	0.9104
recall_(sens)	1.0	0.9331
specificity	0.9739	0.8878
precision	0.9611	0.8434
f1_score	0.9842	0.9063

--- train: 0.0129

0.8

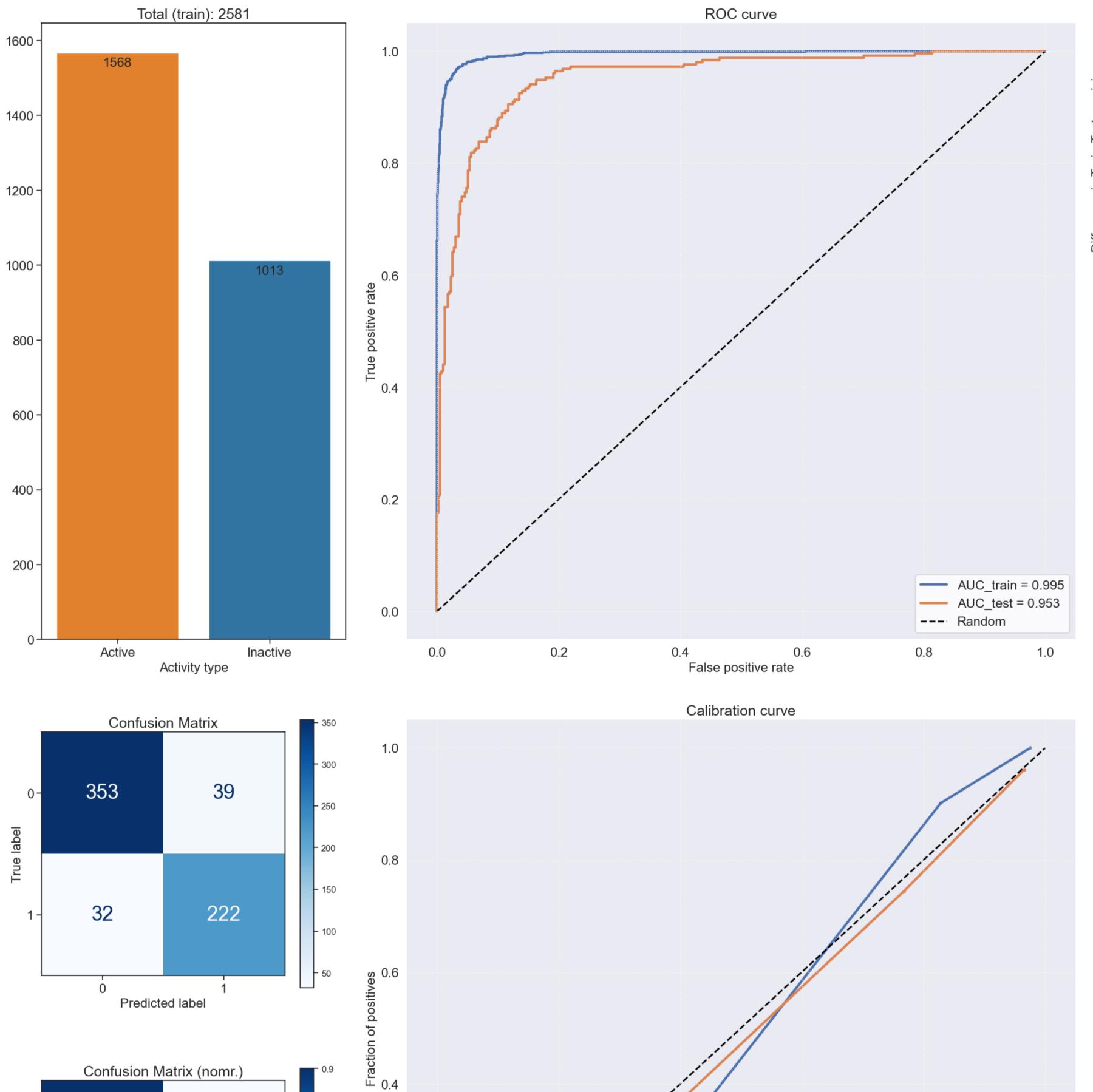
0.4 0.6 Mean predicted probability

test: 0.0768

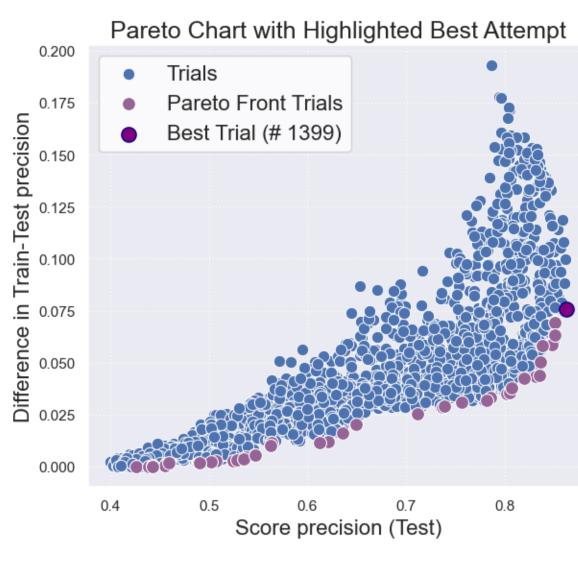
--- perfectly_calibrated

1.0

Protein: P31645 - metric: precision



0.2



XGBoost Parameters	Value
alpha	2.0954
colsample_bytree	0.6989
gamma	0.3671
lambda	3.3593
learning_rate	0.0204
max_depth	10.0
min_child_weight	3.0
n_estimators	604.0
reg_alpha	0.196
reg_lambda	0.2809
scale_pos_weight	1.6438
subsample	0.8528

		Active/Ina	ctive probability	:	:
		tive zone		Active zone	type active (254) inactive (392)
10					decoy (646)
Percent					
4					
2		_ 🗖 _			
0					10
	0.0 0.2	0.4 Pro	0.6 obability	0.8	1.0

- 0.8

0.4

0.3

0.2

0.2

0.0

0.0

Confusion Matrix (nomr.)

Predicted label

0.099

0.87

0.9

0.13

0-

True label

Metric	Train	Test
AUC	0.9951	0.9529
accuracy	0.964	0.8901
balanced accuracy	0.9663	0.8873
recall_(sens)	0.9773	0.874
specificity	0.9554	0.9005
precision	0.934	0.8506
f1_score	0.9641	0.8903

- train: 0.0325

0.8

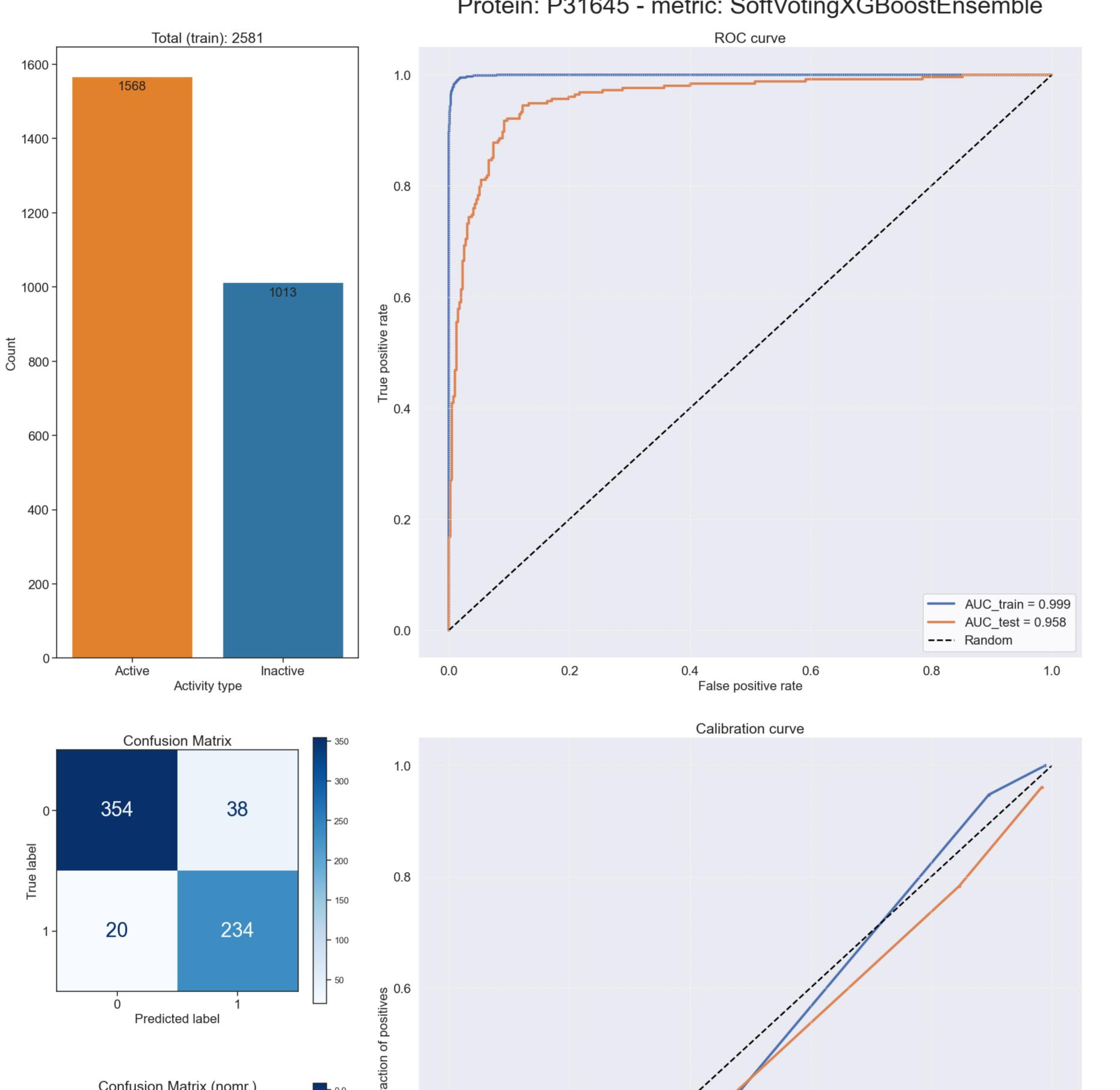
0.4 0.6 Mean predicted probability

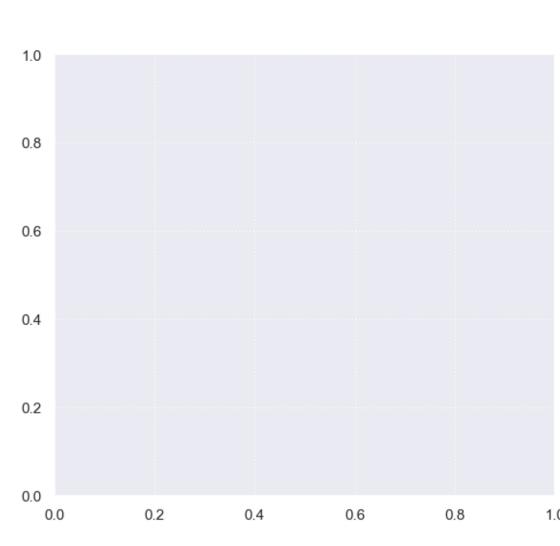
test: 0.0788

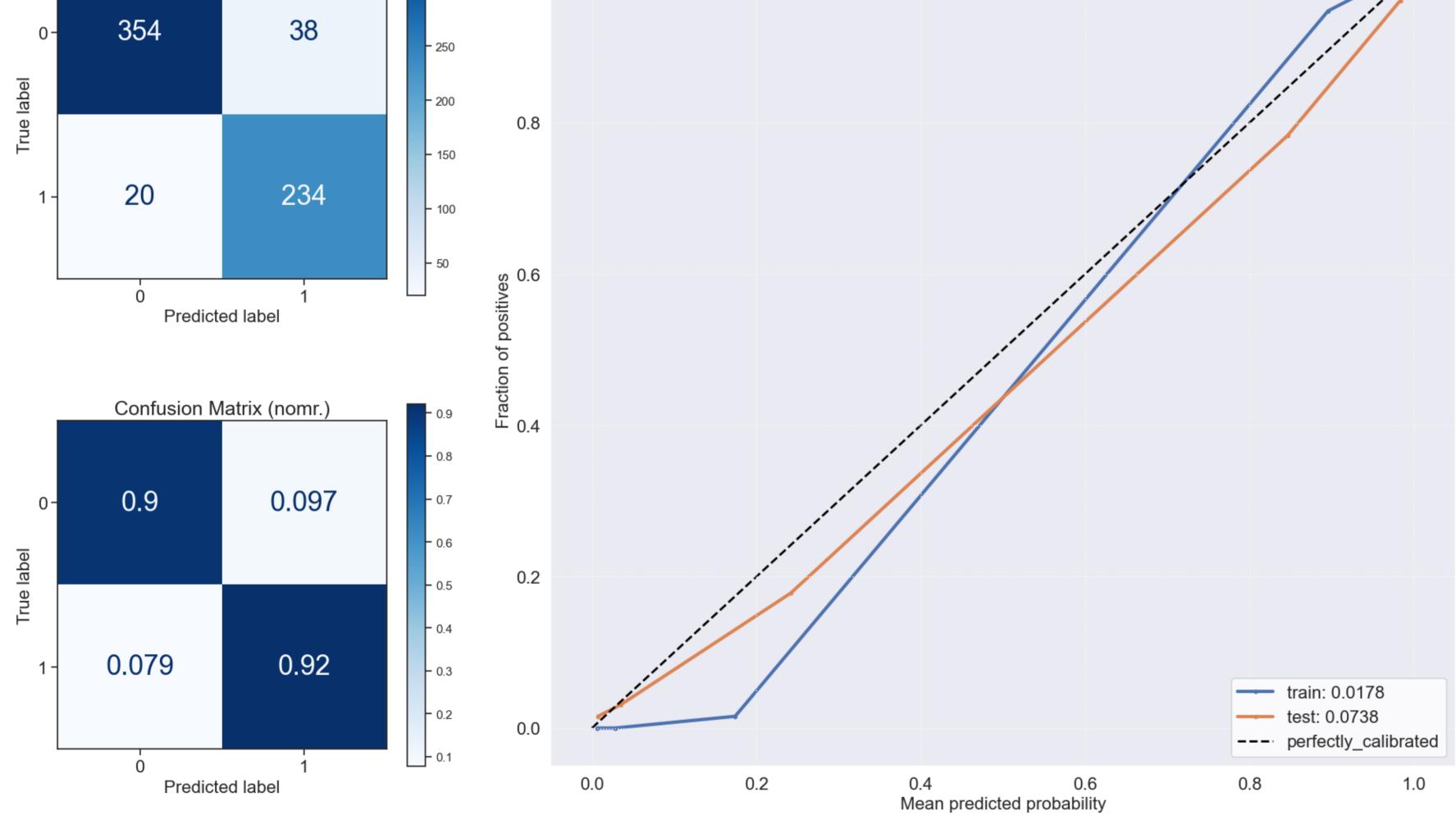
--- perfectly_calibrated

1.0

Protein: P31645 - metric: SoftVotingXGBoostEnsemble







25				Active		type active (254) inactive (392) decoy (646)
Percent 15						
10 -						
0	0.0 0.	.2 0	.4	0.6	0.8	1.0
			Probability			

Active/Inactive probability

Metric	Train	Test
AUC	0.9994	0.9581
accuracy	0.9837	0.9102
balanced accuracy	0.9857	0.9122
recall_(sens)	0.9951	0.9213
specificity	0.9764	0.9031
precision	0.9646	0.8603
f1_score	0.9838	0.9107