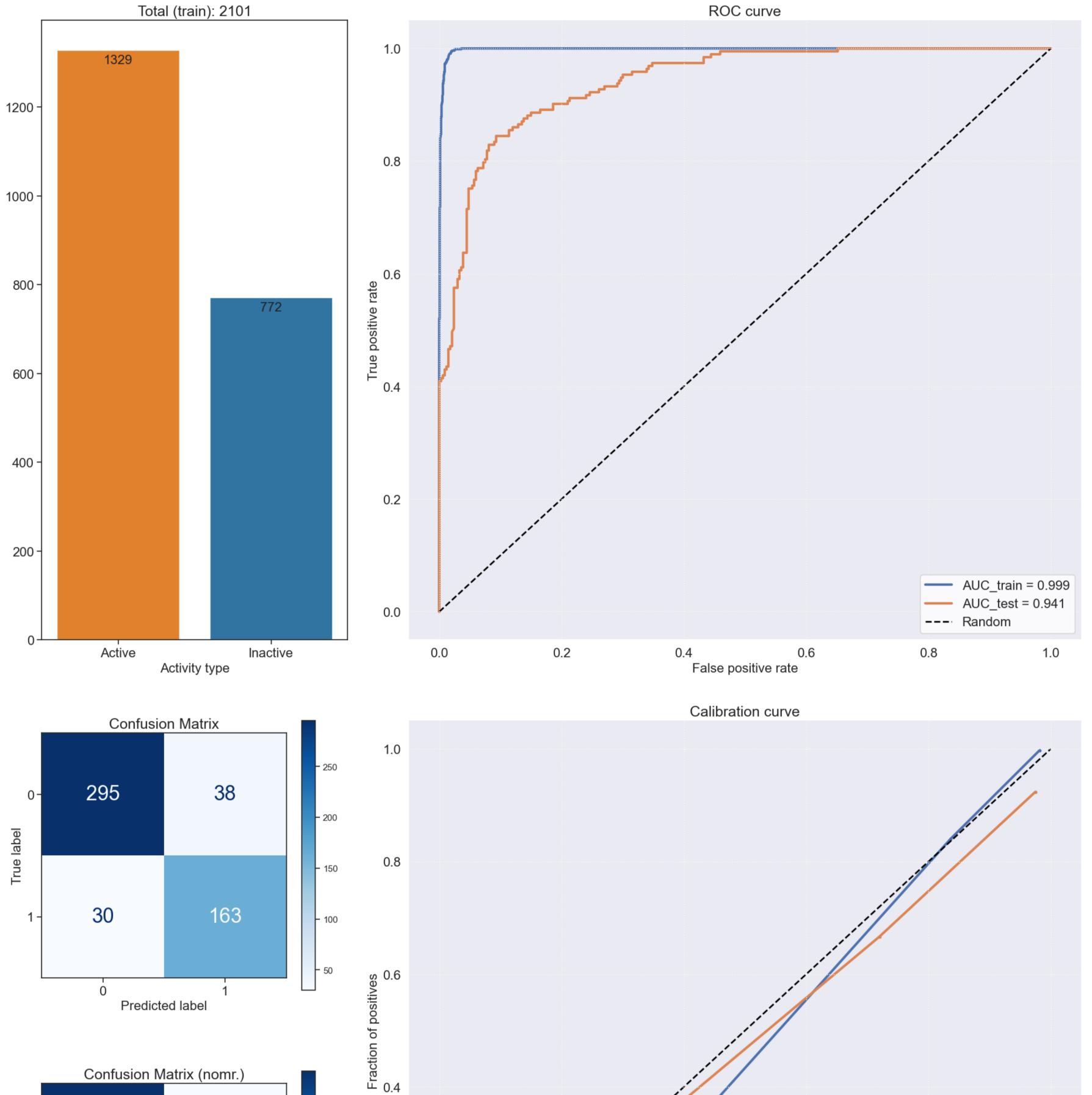
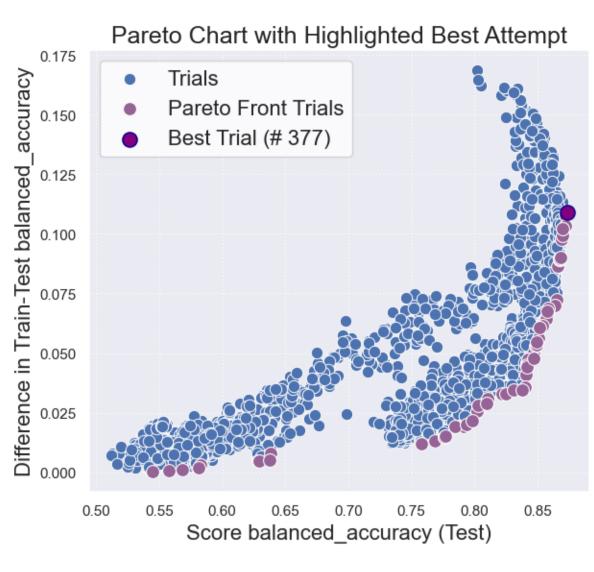
## Protein: P35354 - metric: balanced\_accuracy



0.2

0.4 0.6 Mean predicted probability



XGBoost Parameters	Value
alpha	1.5315
colsample_bytree	0.5726
gamma	0.6383
lambda	3.3615
learning_rate	0.1886
max_depth	9.0
min_child_weight	1.0
n_estimators	961.0
reg_alpha	0.6279
reg_lambda	0.4368
scale_pos_weight	2.3431
subsample	0.864

20.0			Ac		ve probabilit			
17.5		Ina	ctive zone			Active 2	zone	type active (193) inactive (333) decoy (524)
15.0							Ü	
12.5								
Percent								
7.5		<del></del>			 			
5.0								пП
2.5								
0.0	0.0	0.2	2 0	).4	0.6		8	1.0
				Prob	ability			

Confusion Matrix (nomr.)

Predicted label

0.11

0.84

0.6

0.4

0.3

0.2

0.2

0.0

0.0

0.89

0.16

0-

True label

Metric	Train	Test
AUC	0.9985	0.9413
accuracy	0.9781	0.8707
balanced accuracy	0.9824	0.8652
recall_(sens)	0.9987	0.8446
specificity	0.9661	0.8859
precision	0.9449	0.8109
f1_score	0.9782	0.8712

--- train: 0.0213

0.8

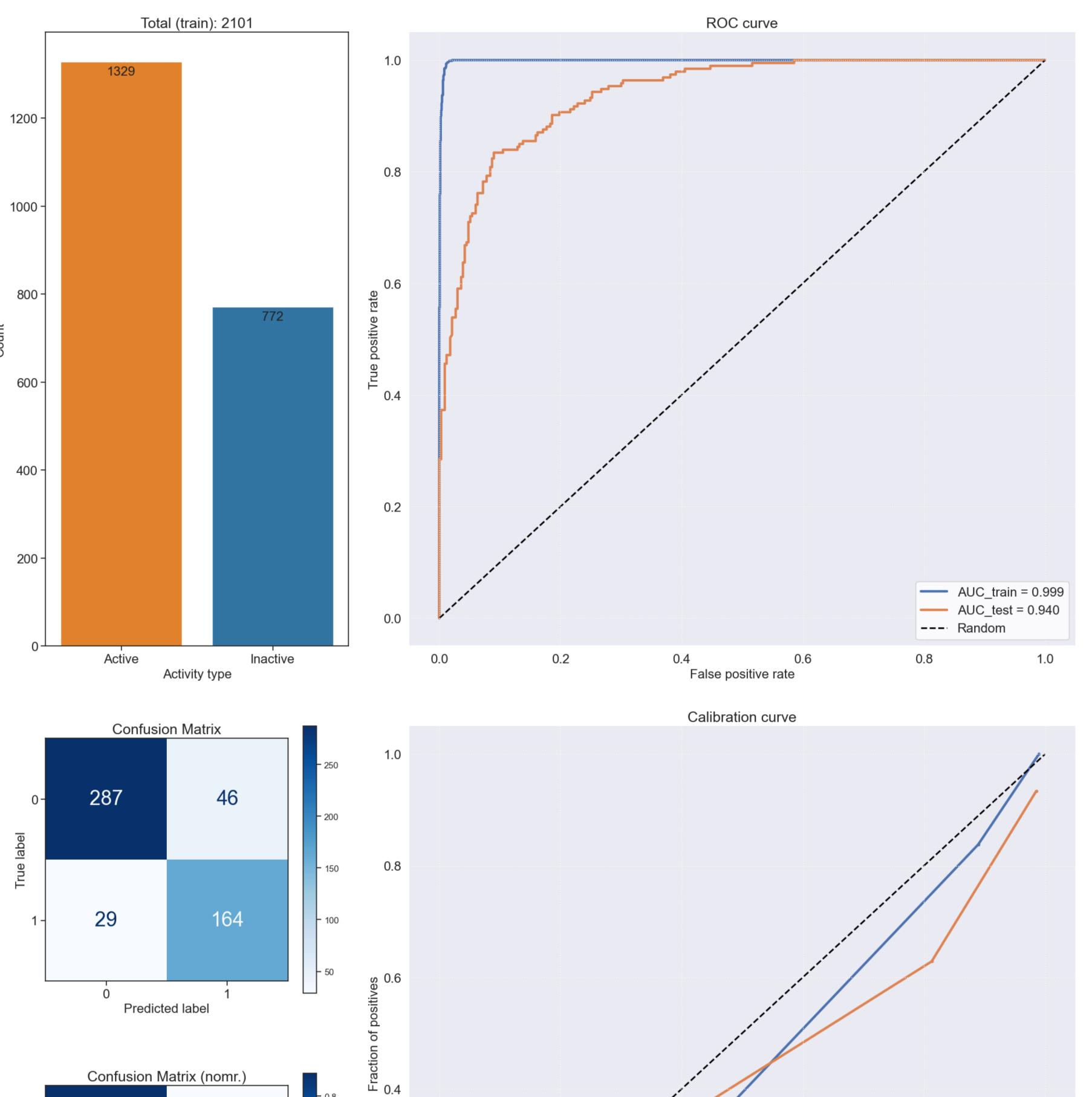
test: 0.0937

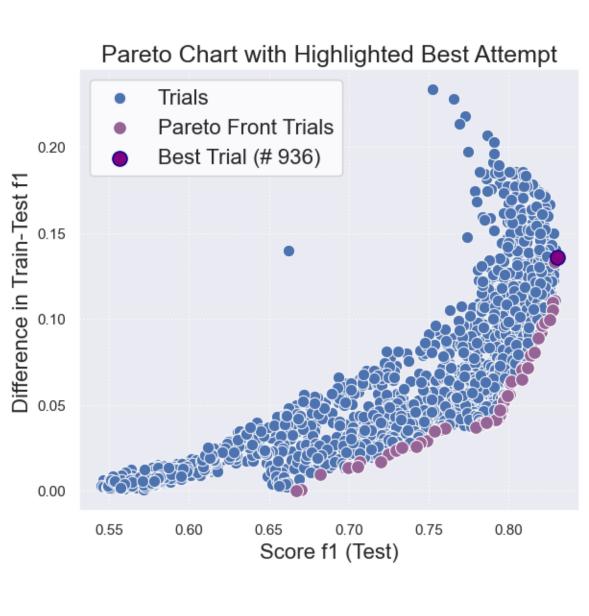
--- perfectly\_calibrated

1.0

20.0			Active/mactiv	c probability		
20.0		Inactive zone			Active zone	type active (193)
17.5		mactive zone				
15.0						
12.5						
Percent 0.01						
7.5						
3.0		<b>1</b>				
2.5						
0.0 —	0.0	0.2	0.4 Proba	0.6 ability	0.8	1.0

## Protein: P35354 - metric: f1





XGBoost Parameters	Value
alpha	1.9986
colsample_bytree	0.8616
gamma	0.8233
lambda	0.1898
learning_rate	0.1411
max_depth	10.0
min_child_weight	1.0
n_estimators	414.0
reg_alpha	0.4432
reg_lambda	0.6012
scale_pos_weight	4.653
subsample	0.8148

Active/Inactive probability								
25	Inactive zone		Active z	type active ( inactive decoy (	(333)			
20								
Percent								
10 <del></del>								
0	.0 0.2 0	.4 0.0	6 0.	8 1	0			
0	0.2	Probability	0.	1.	0			

0.2

0.0

0.0

0.2

0.4

0.3

0.2

Confusion Matrix (nomr.)

Predicted label

0.14

0.85

0.86

0.15

0-

True label

Metric	Train	Test
AUC	0.9989	0.9396
accuracy	0.9738	0.8574
balanced accuracy	0.9793	0.8558
recall_(sens)	1.0	0.8497
specificity	0.9586	0.8619
precision	0.9335	0.781
f1_score	0.974	0.8586

--- train: 0.0225

0.8

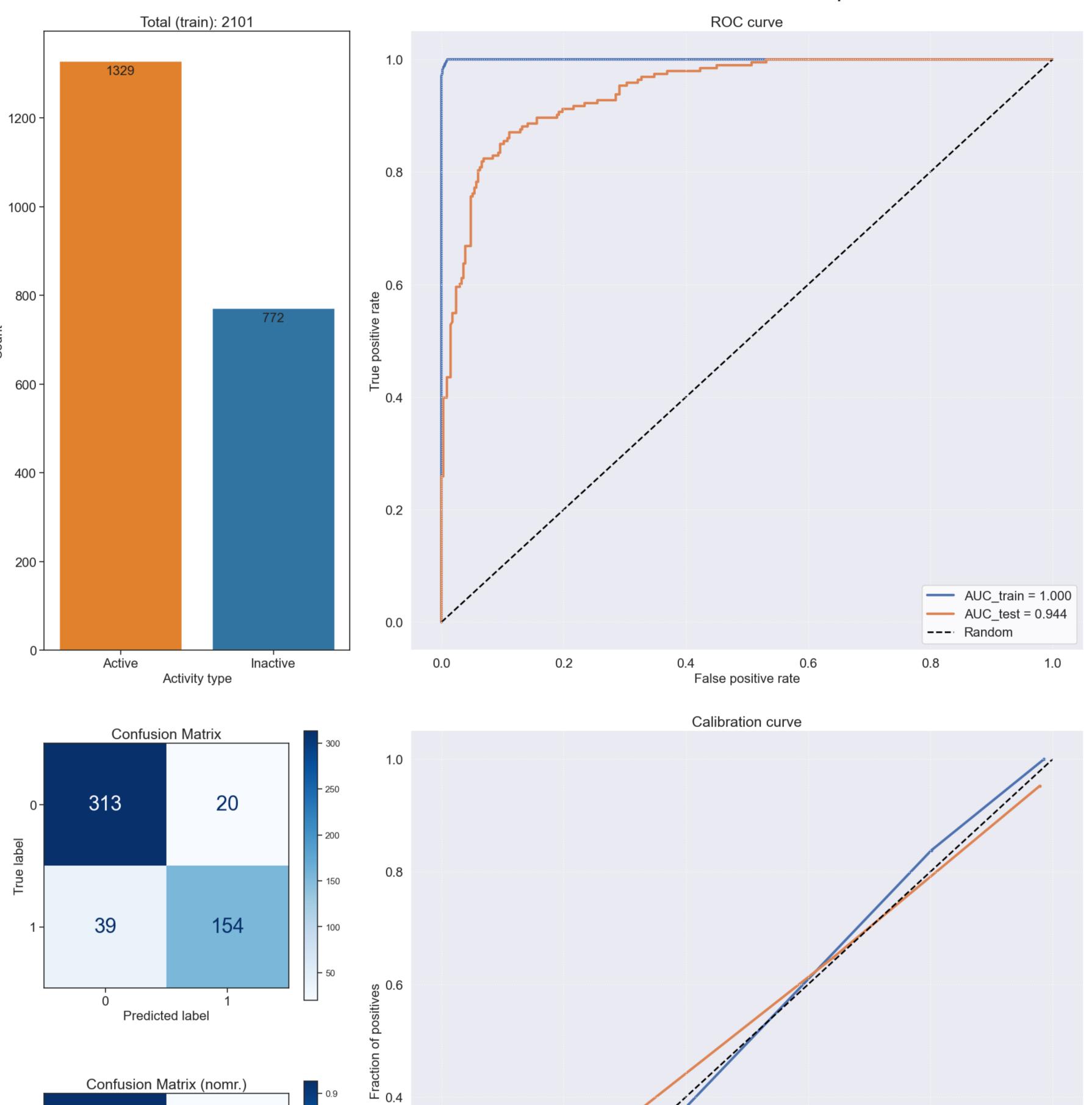
0.4 0.6 Mean predicted probability

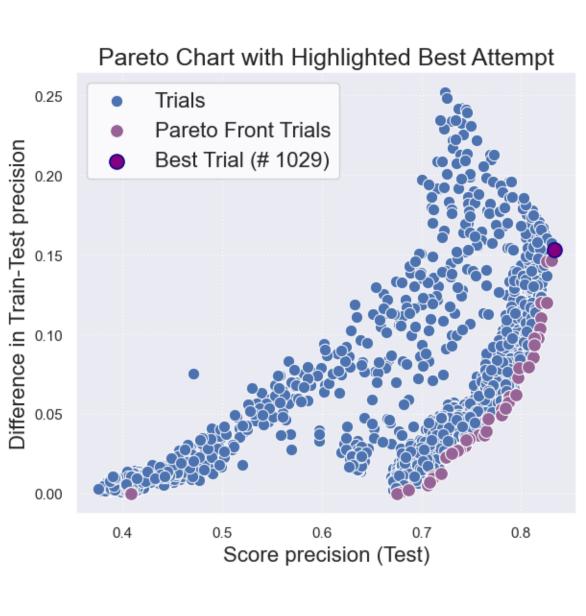
test: 0.1024

--- perfectly\_calibrated

1.0

## Protein: P35354 - metric: precision





XGBoost Parameters	Value
alpha	1.4547
colsample_bytree	0.5998
gamma	0.0015
lambda	2.3032
learning_rate	0.0442
max_depth	10.0
min_child_weight	1.0
n_estimators	922.0
reg_alpha	0.4385
reg_lambda	0.3836
scale_pos_weight	1.2925
subsample	0.7939

		Ac	tive/Inactive	e probability	У		
40 <del></del>		ctive zone			Active 2	inac	ve (193) tive (333) by (524)
30							
25							
10 <del></del>							
0	0.0 0	.2	).4 Probal	0.6	0.	.8	1.0
			110001	Jiirry			

- 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

0.06

8.0

0.94

0.2

0-

True label

Metric	Train	Test
AUC	0.9999	0.9444
accuracy	0.9938	0.8878
balanced accuracy	0.9951	0.8689
recall_(sens)	1.0	0.7979
specificity	0.9902	0.9399
precision	0.9834	0.8851
f1_score	0.9938	0.8865

--- train: 0.0086

0.8

0.4 0.6 Mean predicted probability

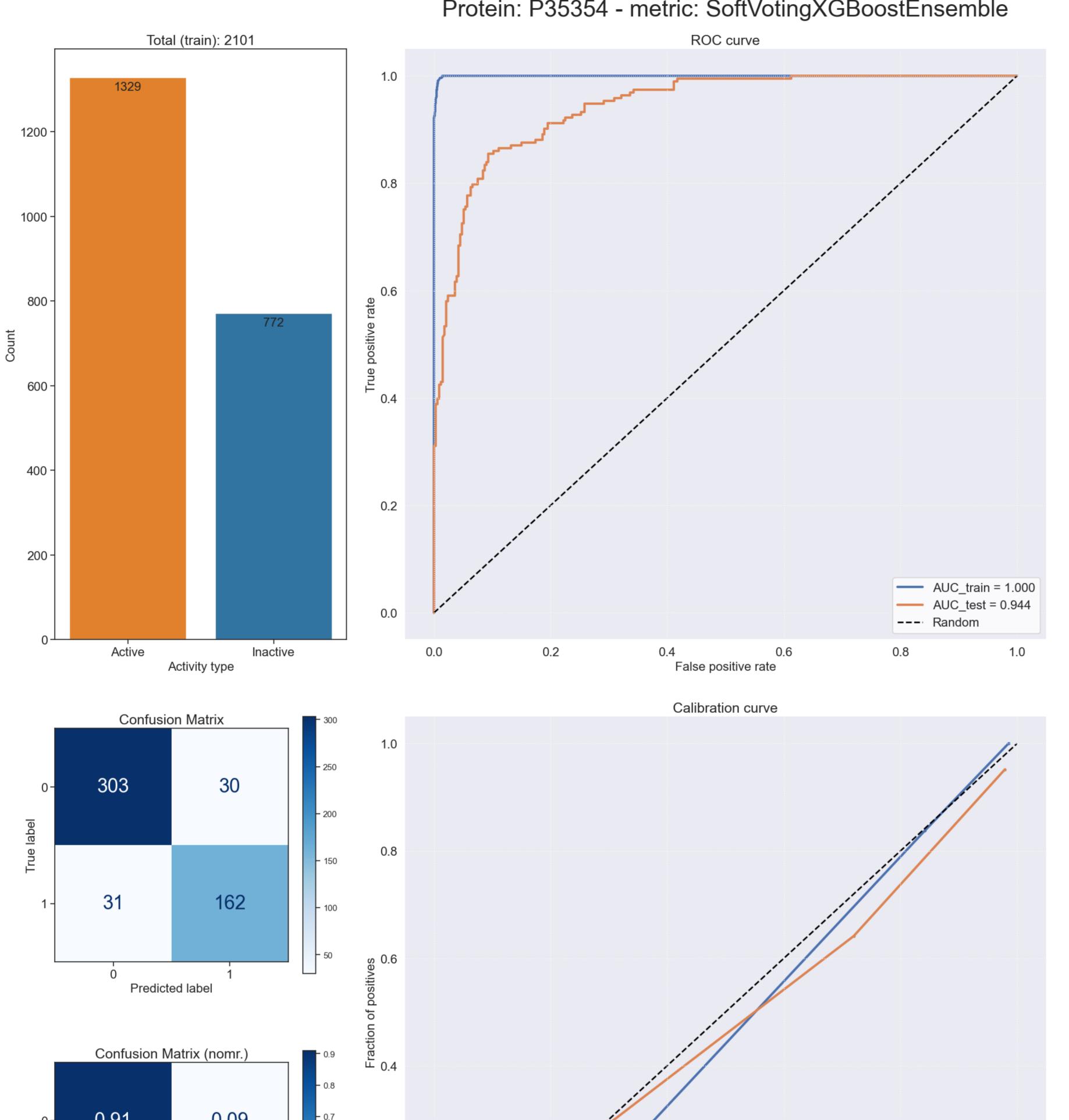
test: 0.0912

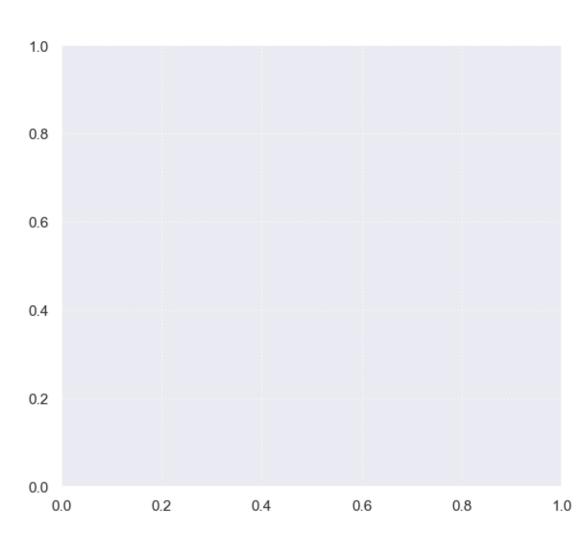
--- perfectly\_calibrated

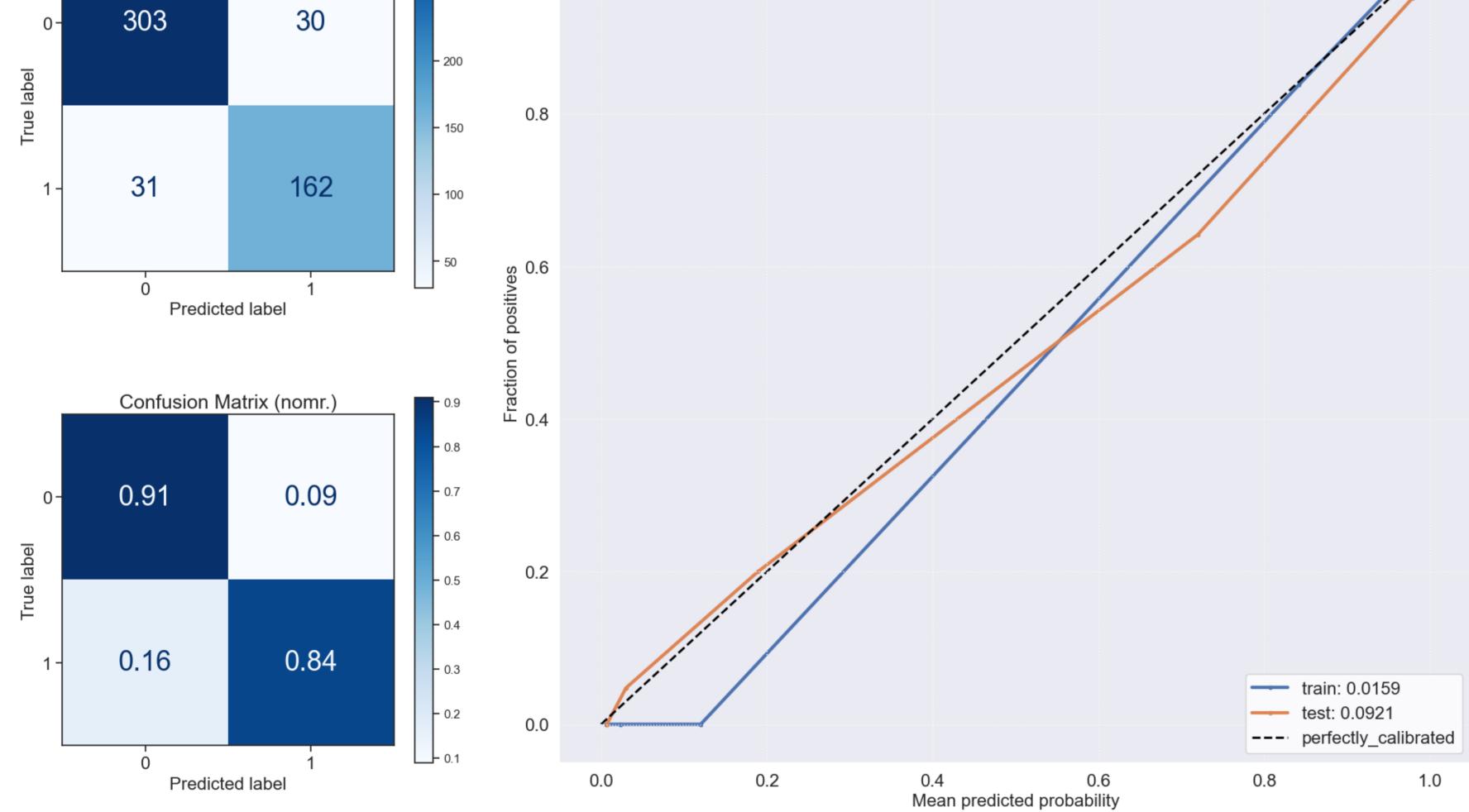
1.0

Active/mactive probability								
40 <del></del>		active zone					type active (193) inactive (333) decoy (524)	
20								
Percent 20								
15 <del></del>								
5								
0	0.0	0.2 0	.4 Probabi	0.6 lity	0.	8	1.0	

## Protein: P35354 - metric: SoftVotingXGBoostEnsemble







	Active/Inactive probability								
20	lna	ctive zone			Active zone	type active (193) inactive (333) decoy (524)			
15									
Percent 10									
5									
0									
(	0.0	.2	0.4 Probab	0.6 ility	0.8	1.0			

Metric	Train	Test
AUC	0.9997	0.9436
accuracy	0.9852	0.884
balanced accuracy	0.9883	0.8746
recall_(sens)	1.0	0.8394
specificity	0.9767	0.9099
precision	0.9614	0.8438
f1_score	0.9853	0.884