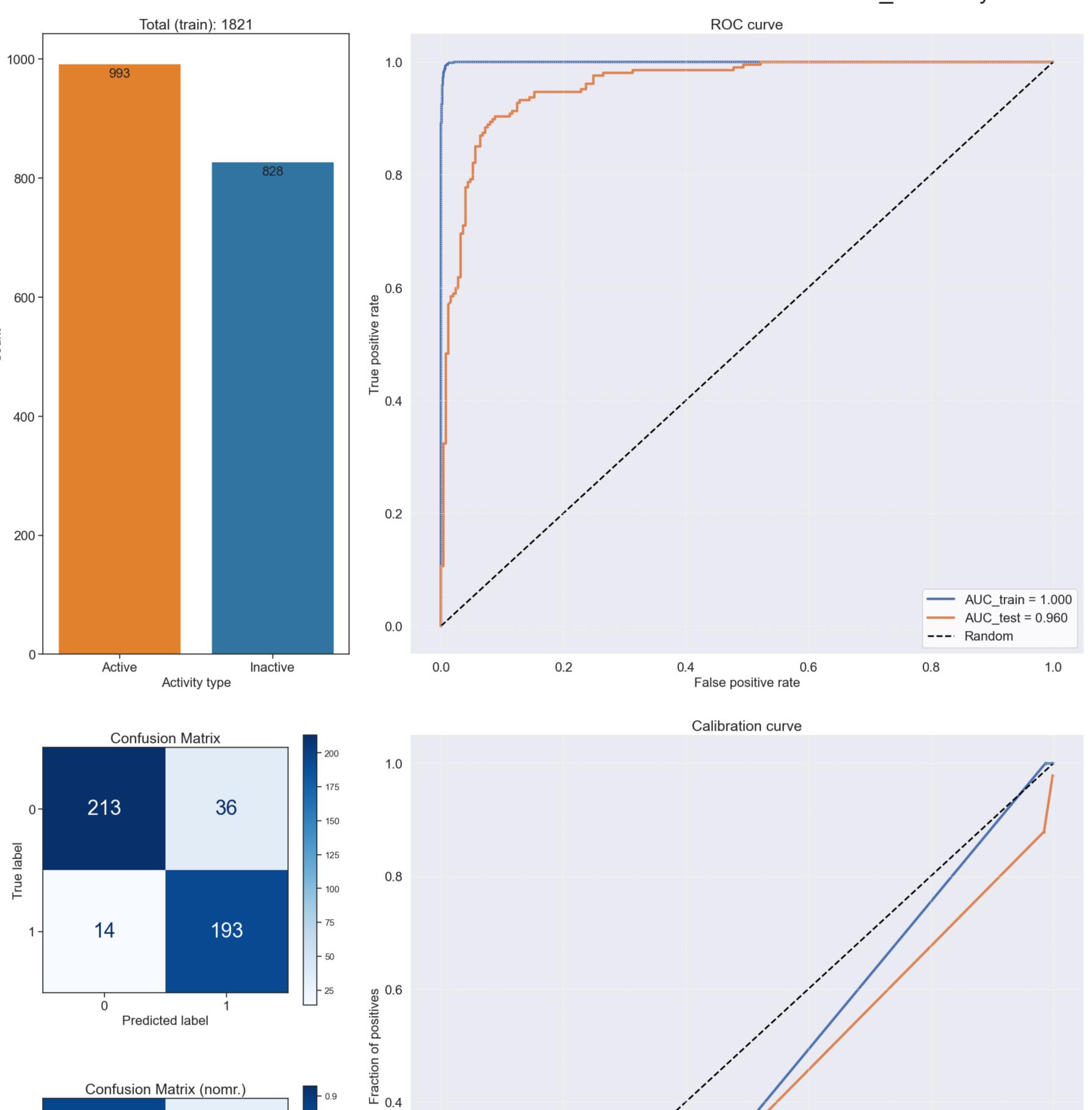
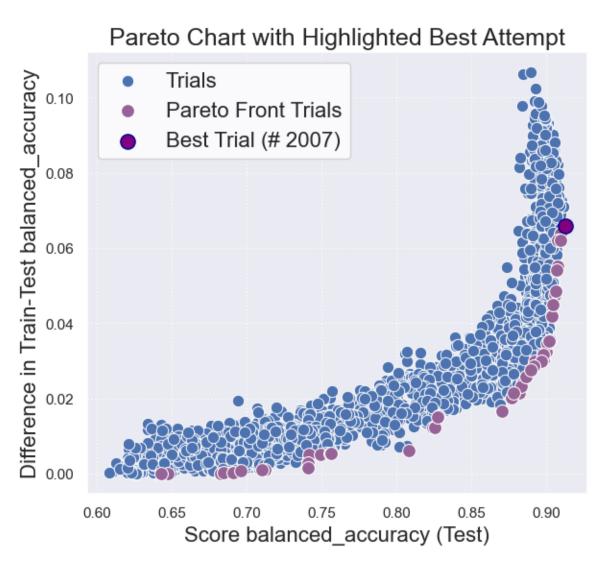
Protein: P28845 - metric: balanced_accuracy





| XGBoost Parameters | Value |
|--------------------|--------|
| alpha | 1.7009 |
| colsample_bytree | 0.6692 |
| gamma | 0.6129 |
| lambda | 2.2497 |
| learning_rate | 0.3801 |
| max_depth | 7.0 |
| min_child_weight | 5.0 |
| n_estimators | 768.0 |
| reg_alpha | 0.0299 |
| reg_lambda | 0.9424 |
| scale_pos_weight | 5.1608 |
| subsample | 0.9226 |
| | |

| | Active/Inactive probability | | | | | | |
|---------|--|------------|---|--------|--|--|--|
| 60 | Ina | ctive zone | | Active | type active (207) inactive (249) decoy (456) | | |
| 50 | | | | | | | |
| 40 | | | | | | | |
| Percent | | | | | | | |
| 20 | | | | | | | |
| 10 | | | | | | | |
| 0 | | 2 | 4 | 6 0 | 8 10 | | |
| | 0.0 0.2 0.4 0.6 0.8 1.0 Probability | | | | | | |

- 0.8

- 0.6

- 0.5

0.4

0.3

0.2

0.1

0.2

0.0

0.0

0.2

Confusion Matrix (nomr.)

Predicted label

0.14

0.93

0.86

0.068

0-

True label

| Metric | Train | Test |
|-------------------|--------|--------|
| AUC | 0.9997 | 0.9596 |
| accuracy | 0.9753 | 0.8904 |
| balanced accuracy | 0.9773 | 0.8939 |
| recall_(sens) | 1.0 | 0.9324 |
| specificity | 0.9547 | 0.8554 |
| precision | 0.9485 | 0.8428 |
| f1_score | 0.9753 | 0.8906 |

--- train: 0.0174

0.8

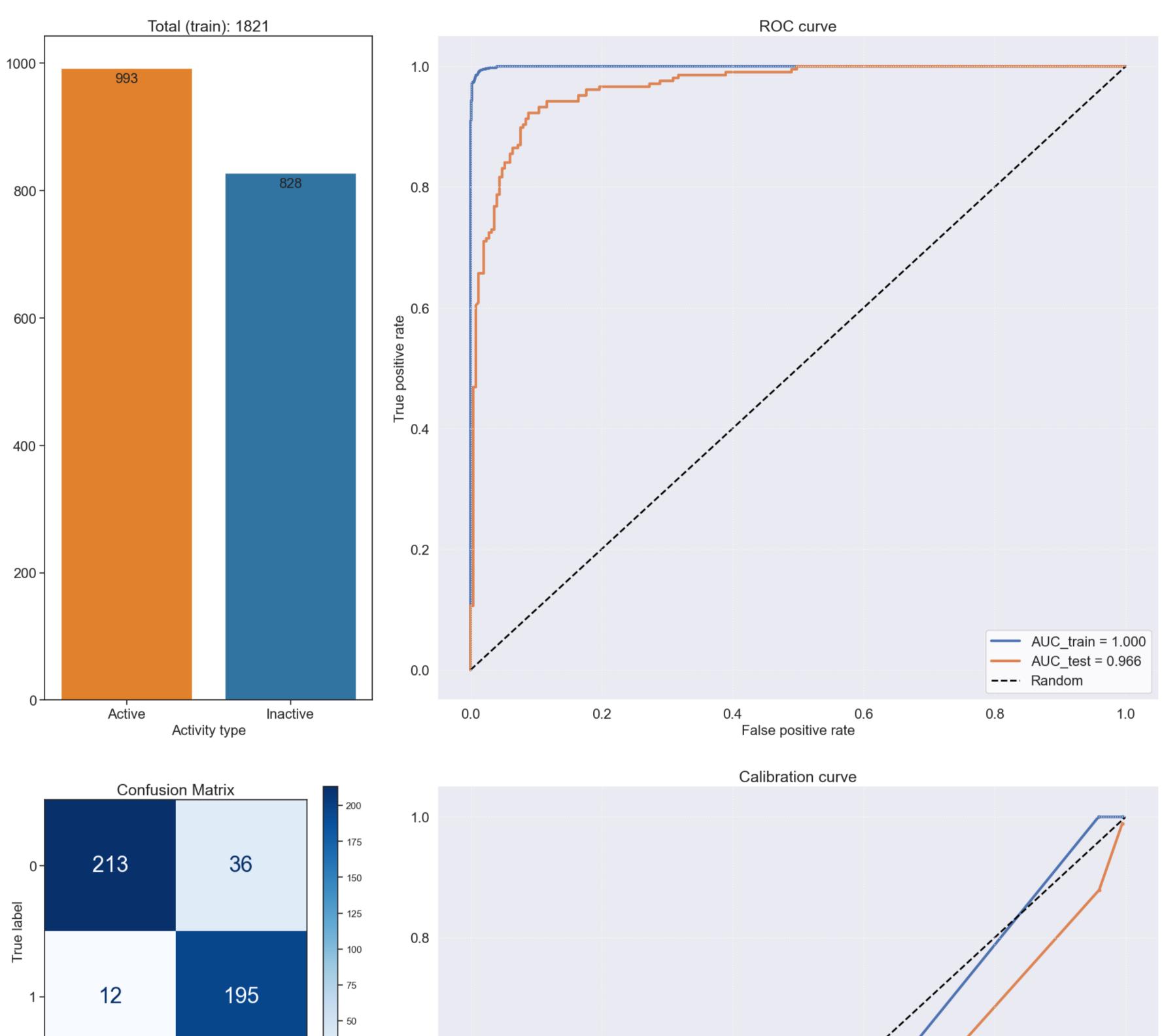
0.4 0.6 Mean predicted probability

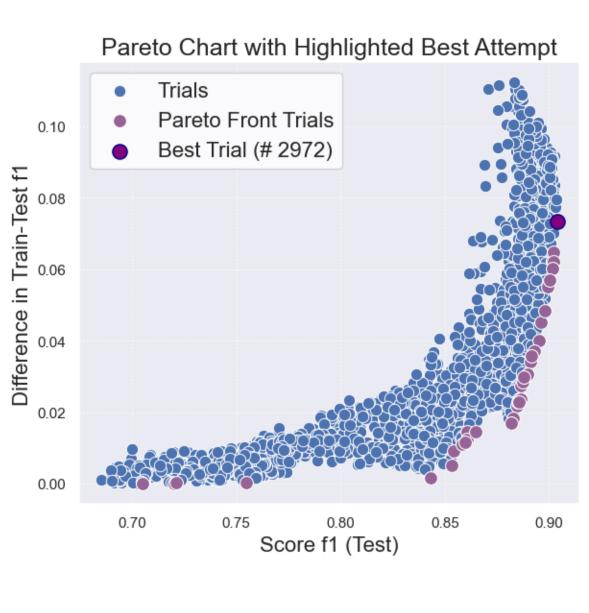
test: 0.0868

--- perfectly_calibrated

1.0

Protein: P28845 - metric: f1





| XGBoost Parameters | Value |
|--------------------|--------|
| alpha | 4.4425 |
| colsample_bytree | 0.554 |
| gamma | 0.0515 |
| lambda | 0.7395 |
| learning_rate | 0.0249 |
| max_depth | 9.0 |
| min_child_weight | 1.0 |
| n_estimators | 538.0 |
| reg_alpha | 0.0517 |
| reg_lambda | 0.8157 |
| scale_pos_weight | 2.3612 |
| subsample | 0.5654 |
| | |

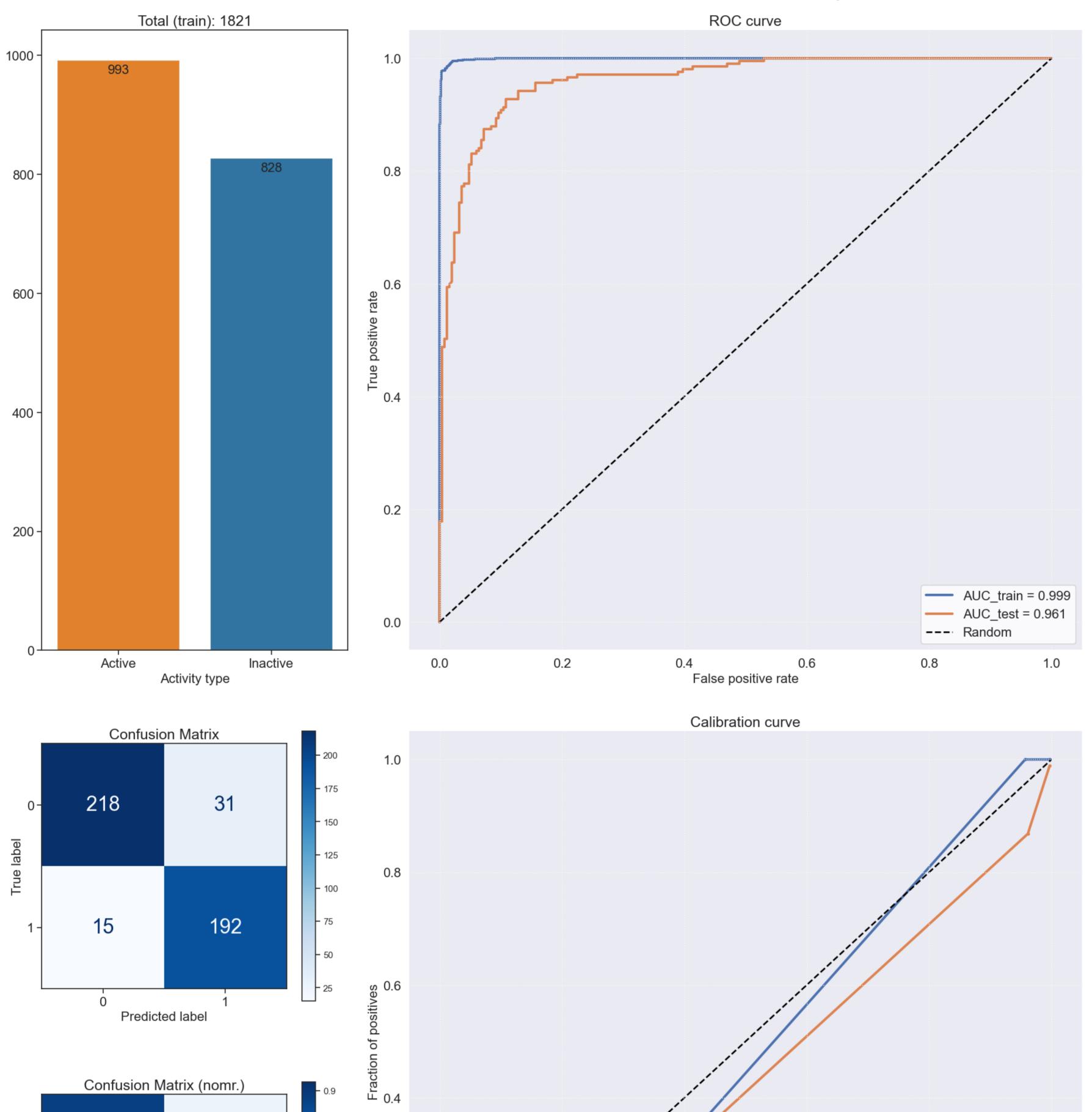
| abel O | 213 | 36 | - 175 - 150 - 125 | | | | | |
|------------|---------------|----------------|---|-----|---------|--|-------------------------|---|
| True label | 12 | 195 | - 100 - 75 - 50 | 0.8 | | | , and a second second | |
| | 0 Predicte | 1 ed label | 7 | | | | | |
| | Confusion M | latrix (nomr.) | - 0.9 L | 0.4 | | , para a la compania de la compania | | |
| label O | 0.86 | 0.14 | - 0.8 - 0.7 - 0.6 - 0.5 | 0.2 | | | | |
| True label | 0.058 | 0.94 | - 0.5 - 0.4 - 0.3 - 0.2 - 0.1 | 0.0 | | | | train: 0.0191 test: 0.0779 perfectly_calibrated |
| | 0 Predicte | 1 ed label | | | 0.0 0.2 | 0.4 Mean predi | 0.6 cted probability | 0.8 1.0 |

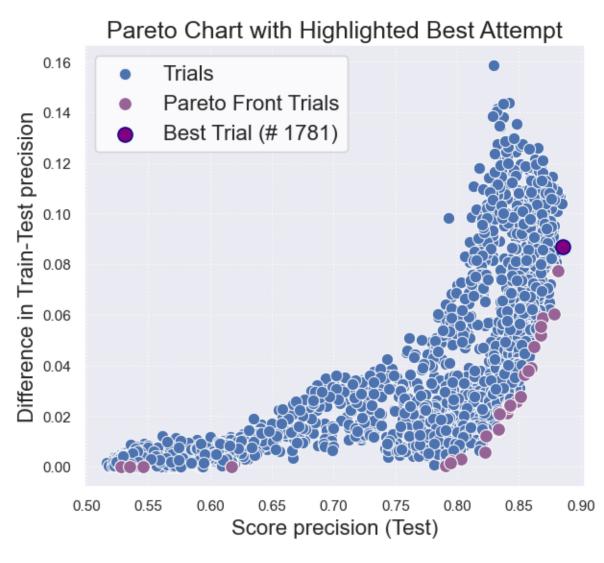
| 40 | ln | active zone | | | Active z | one typ | e (207) ive (249) |
|---------------|-----|-------------|---------------|--------------|----------|---------|----------------------|
| 30 | | | | | | deco | y (456) |
| Percent 00 | | | | | | | |
| | | | | | | | |
| 10 | | | | | | | |
| 0 | 0.0 | 0.2 | 0.4 Probab | 0.6 ility | 0.8 | 3 | 1.0 |

Active/Inactive probability

| Metric | Train | Test |
|-------------------|--------|--------|
| AUC | 0.9995 | 0.9657 |
| accuracy | 0.978 | 0.8947 |
| balanced accuracy | 0.9799 | 0.8987 |
| recall_(sens) | 1.0 | 0.942 |
| specificity | 0.9597 | 0.8554 |
| precision | 0.9539 | 0.8442 |
| f1_score | 0.9781 | 0.895 |

Protein: P28845 - metric: precision





| XGBoost Parameters | Value |
|--------------------|--------|
| alpha | 1.6475 |
| colsample_bytree | 0.5497 |
| gamma | 0.3794 |
| lambda | 1.2747 |
| learning_rate | 0.1833 |
| max_depth | 5.0 |
| min_child_weight | 3.0 |
| n_estimators | 476.0 |
| reg_alpha | 0.0735 |
| reg_lambda | 0.7563 |
| scale_pos_weight | 1.4966 |
| subsample | 0.9304 |

| | | | Active/Inactiv | e probability | | |
|----------------|-----|---------------|----------------|---------------|-------------|--|
| 50 | | Inactive zone | | | Active zone | type active (207) inactive (249) decoy (456) |
| 20 | | | | | | |
| Percent 00 | | | | | | |
| 10 | | | | | | |
| 0— | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
| | | | Proba | ibility | | |

- 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

0.12

0.93

0.88

0.072

0-

True label

| Metric | Train | Test |
|-------------------|--------|--------|
| AUC | 0.9993 | 0.9615 |
| accuracy | 0.9846 | 0.8991 |
| balanced accuracy | 0.9855 | 0.9015 |
| recall_(sens) | 0.9952 | 0.9275 |
| specificity | 0.9758 | 0.8755 |
| precision | 0.9717 | 0.861 |
| f1_score | 0.9846 | 0.8993 |

train: 0.016

8.0

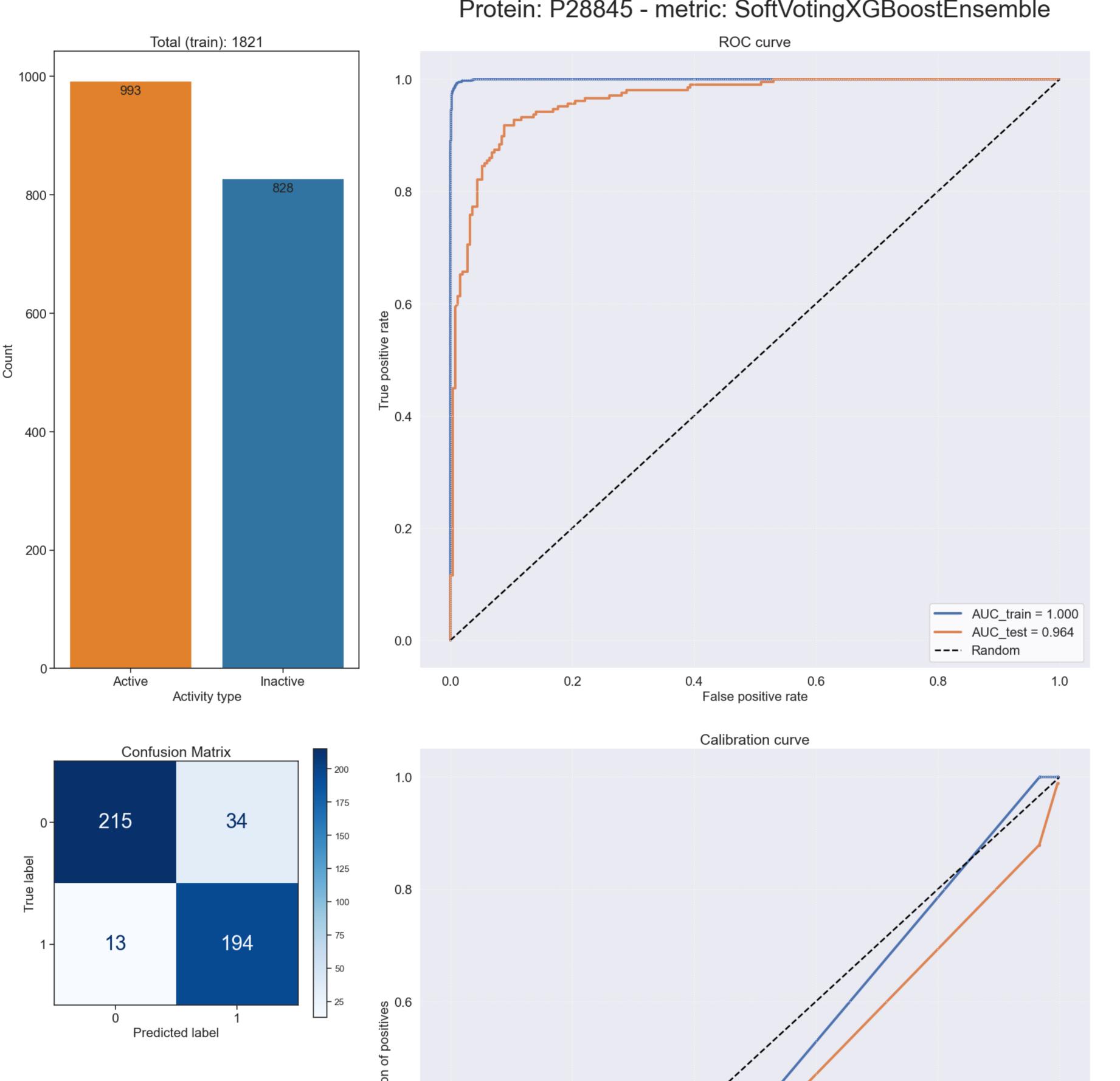
0.4 0.6 Mean predicted probability

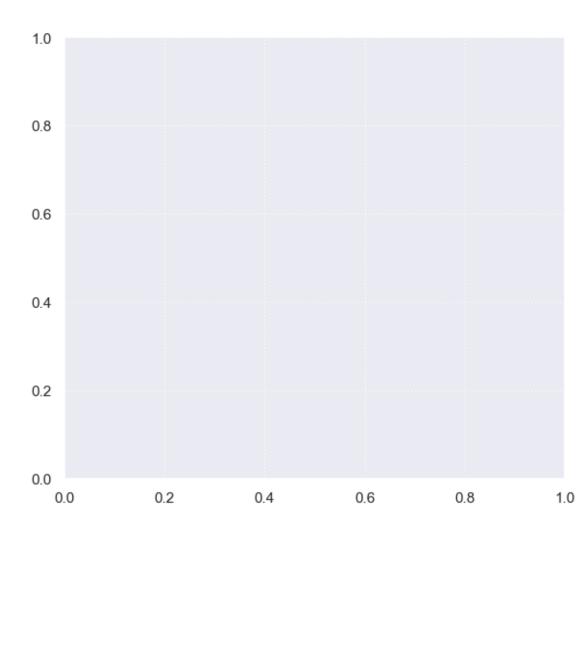
test: 0.0775

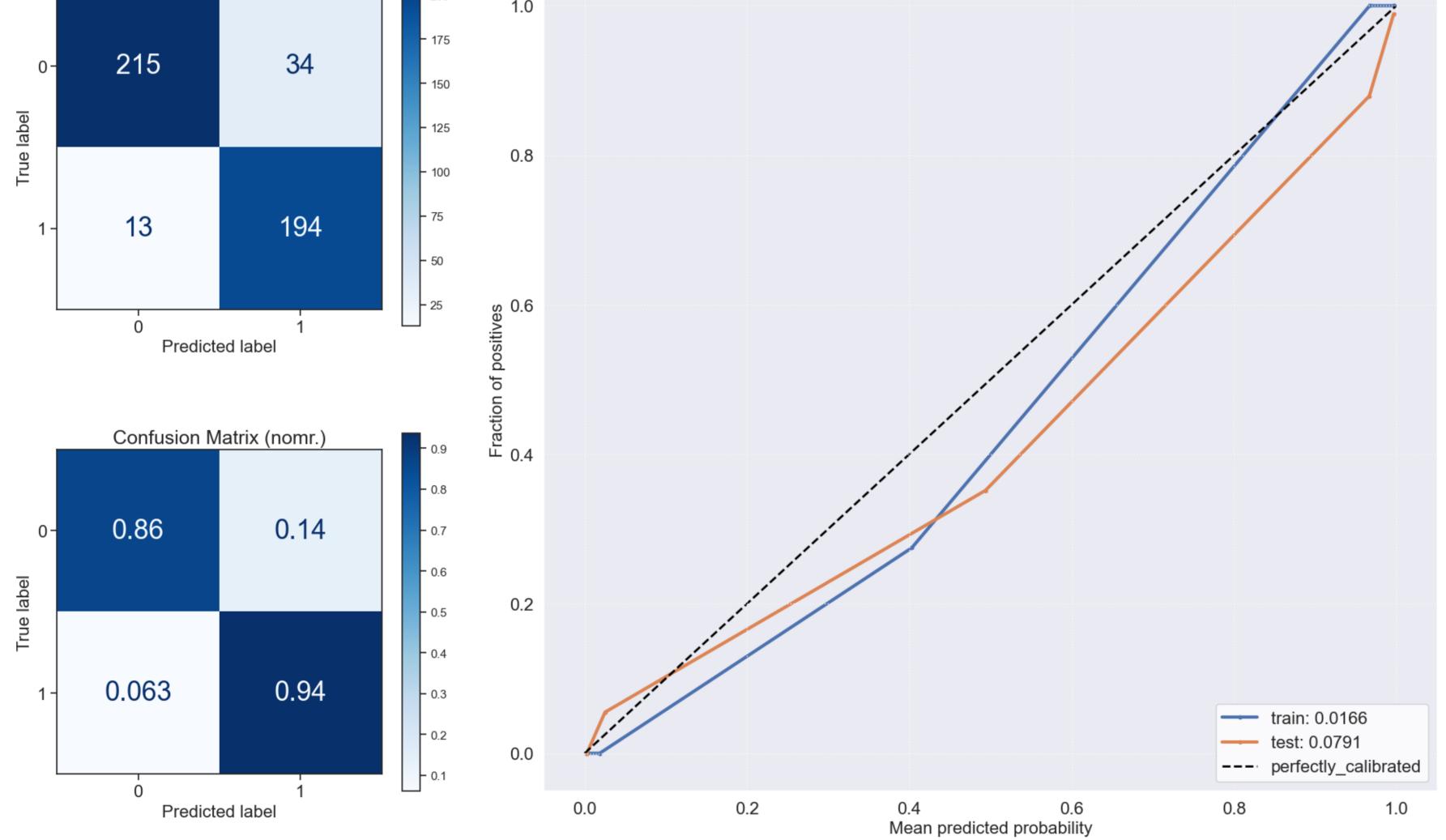
--- perfectly_calibrated

1.0

Protein: P28845 - metric: SoftVotingXGBoostEnsemble







| 50 | | | Active/Inactive | | | |
|-----------------|-----|---------------|-----------------|-------|-------------|--|
| 50 - | | Inactive zone | | | Active zone | type active (207) inactive (249) decoy (456) |
| 30 | | | | | | |
| Percent | | | | | | |
| 10 | | | | | | |
| 0 | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
| | | | Probab | ility | | |

| Metric | Train | Test |
|-------------------|--------|--------|
| AUC | 0.9996 | 0.9636 |
| accuracy | 0.9786 | 0.8969 |
| balanced accuracy | 0.9803 | 0.9003 |
| recall_(sens) | 0.9988 | 0.9372 |
| specificity | 0.9617 | 0.8635 |
| precision | 0.9561 | 0.8509 |
| f1_score | 0.9786 | 0.8971 |