Antideuteron search status

Charge confusion study

Marta Borchiellini (PhD student)
Kapteyn Astronomical Institute, Groningen

Antideuteron identification

Isotope identification

$$M = \frac{RZ\sqrt{1-\beta^2}}{\beta}$$

Antideuteron identification

Isotope identification

$$M = \frac{RZ\sqrt{1-\beta^2}}{\beta}$$

Background rejection

Wrong beta reconstruction

Antideuteron identification

Isotope identification

$$M = \frac{RZ\sqrt{1-\beta^2}}{\beta}$$

Background rejection

Wrong beta reconstruction

+

Wrong rigidity and charge sign reconstruction

Antideuteron identification

Isotope identification

$$M = \frac{RZ\sqrt{1-\beta^2}}{\beta}$$

Background rejection

Wrong beta reconstruction

+

Wrong rigidity and charge sign reconstruction

→

Search for Antideuterons

Antideuteron identification

Isotope identification

$$M = \frac{RZ\sqrt{1-\beta^2}}{\beta}$$

Background rejection

Wrong beta reconstruction

+

Wrong rigidity and charge sign reconstruction

Search for Antideuterons

Charge confusion study

Goal: build a charge confusion estimator (data-driven BDT) to reject background composed by events whose rigidity sign is wrongly reconstructed

Dataset

Data:

NAIA ISS Data v1.1.0/ISS.B1236/pass8

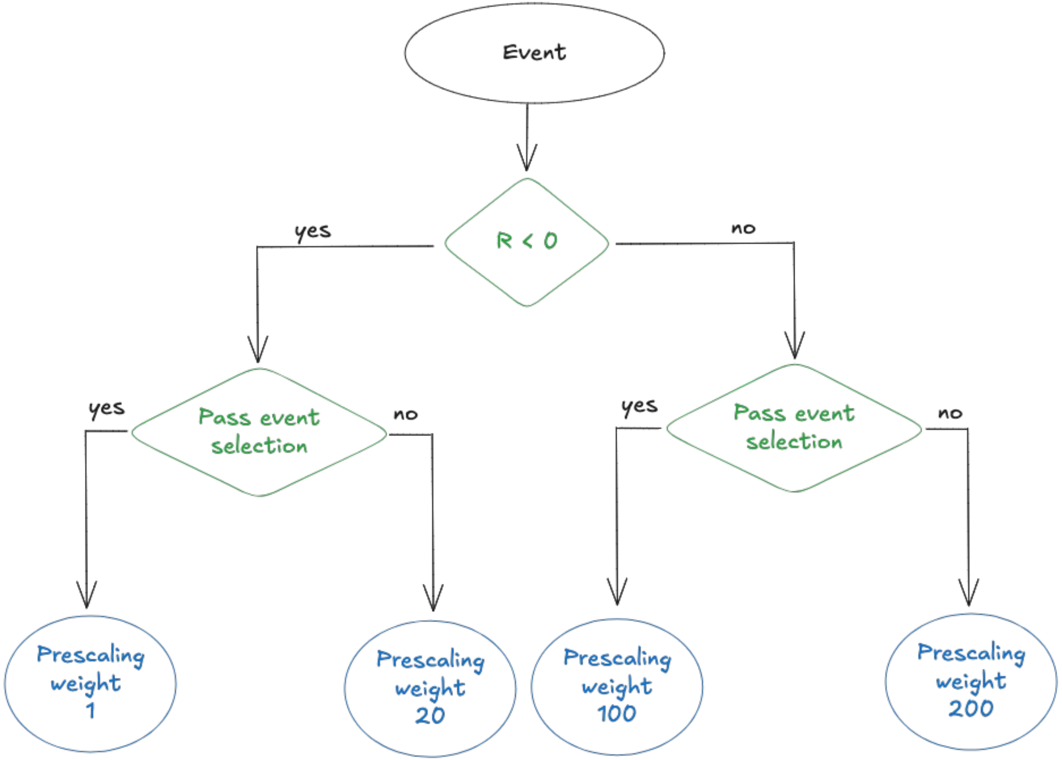
4 years data (2015-2018)

+ one year data (2023)

Dataset

Data:
NAIA ISS Data v1.1.0/ISS.B1236/pass8
4 years data (2015-2018)
+ one year data (2023)

Skimming:
original ntuples are reduced to decrease dataset size.



Skimming flowchart

Dataset

Data:

NAIA ISS Data v1.1.0/ISS.B1236/pass8

4 years data (2015-2018)

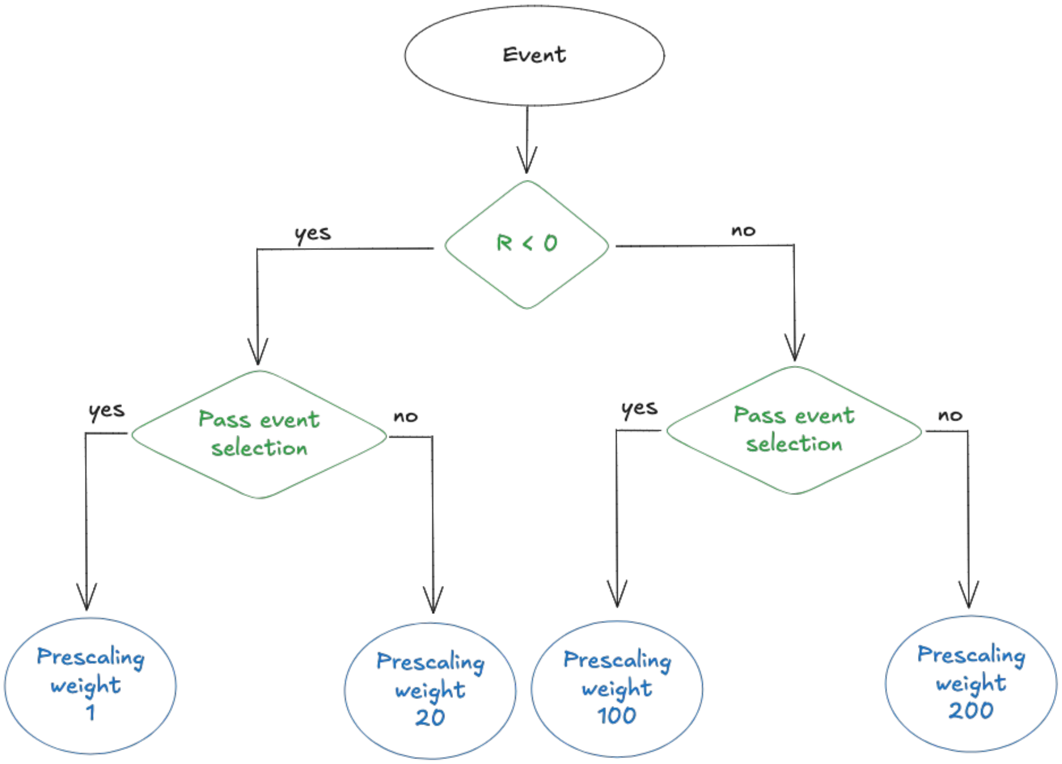
+ one year data (2023)

Skimming:

original ntuples are reduced to decrease dataset size.

Event selection:

Antiproton-like cut based selection



Skimming flowchart

Event selection

<table>
<tr><td>Z=1</td><td>Z=1 TOF</td><td>0.5<qtof<1.5 && qlowtof<2.0 && quptof<1.5</td></tr>
<tr><td></td><td>Z=1 Tracker</td><td>0.5<q_inntr<1.5.</td></tr>
<tr><td></td><td>God TOF Z</td><td>qup<1.5 && qdw<2.0</td></tr>
<tr><td>TOF</td><td>Good TOF NCluster</td><td>NBetaCluster == 4</td></tr>
<tr><td></td><td>Good TOF chisq</td><td>chisqtn < 10 && chisqcn < 10</td></tr>
<tr><td></td><td>Has Downgoing Track</td><td>Beta_tof>0.5</td></tr>
<tr><td>TRACKER</td><td>Good Inner tracker chisq</td><td>chisqInnerX_GBL< 10 && chisqnInnerY_GBL < 10</td></tr>
<tr><td></td><td>Single track</td><td>ntrtrack == 1</td></tr>
<tr><td></td><td>Tracker pattern</td><td>L2 && (L3 || L4) && (L5 || L6) && (L7 || L8)</td></tr>
<tr><td></td><td>XY Hits</td><td>At least 3 XY hits</td></tr>
<tr><td></td><td>Energy deposition</td><td>Less than 2.5 MeV deposited in Inner tracker (LayerEDep)</td></tr>
<tr><td>TRD</td><td>Enough TRD hits</td><td>NHitsOnTrack >10</td></tr>
<tr><td></td><td>Likelihood e/p</td><td>Likelihood e/p >0.8</td></tr>
<tr><td></td><td>Likelihood p/He</td><td>Likelihood p/He < 0.3</td></tr>
<tr><td></td><td>Physics Trigger</td><td>IsPhysicsTrigger() == True</td></tr>
<tr><td></td><td>Rigidity for isotope identification</td><td>|R_innner| < 20GV</td></tr>
</table>

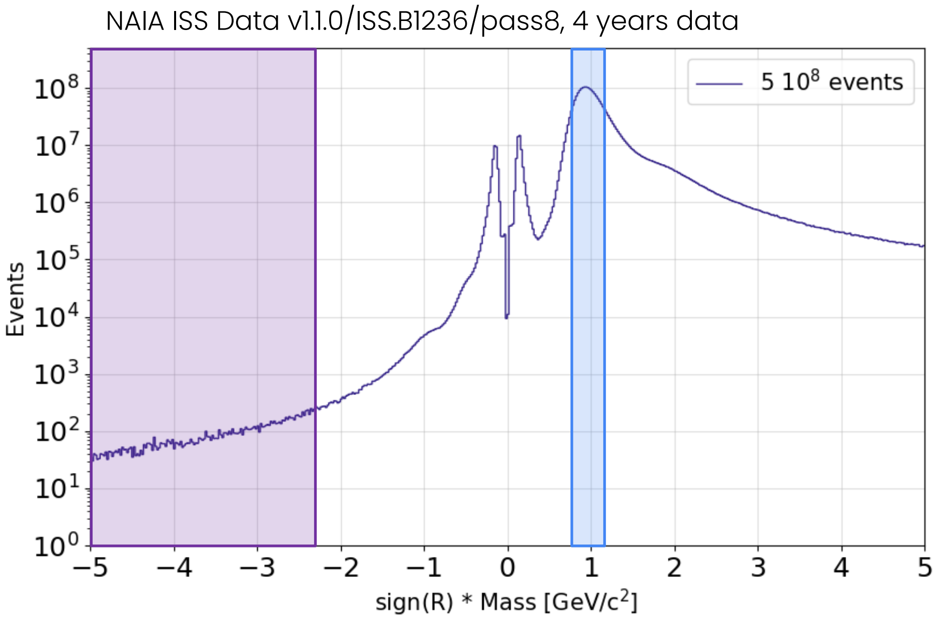
Event selection

+ requirement on Tof beta:
Beta_tof < 0.9

Z=1	**Z=1 TOF**	0.5<qtof<1.5 && qlowtof<2.0 && quptof<1.5
	Z=1 Tracker	0.5<q_inntr<1.5.
	God TOF Z	qup<1.5 && qdw<2.0
TOF	**Good TOF NCluster**	NBetaCluster == 4
	Good TOF chisq	chisqtn < 10 && chisqcn < 10
	Has Downgoing Track	Beta_tof>0.5
TRACKER	**Good Inner tracker chisq**	chisqInnerX_GBL< 10 && chisqnInnerY_GBL < 10
	Single track	ntrtrack == 1
	Tracker pattern	L2 && (L3 || L4) && (L5 || L6) && (L7 || L8)
	XY Hits	At least 3 XY hits
	Energy deposition	Less than 2.5 MeV deposited in Inner tracker (LayerEDep)
TRD	**Enough TRD hits**	NHitsOnTrack >10
	Likelihood e/p	Likelihood e/p >0.8
	Likelihood p/He	Likelihood p/He < 0.3

Physics Trigger	IsPhysicsTrigger() == True
Rigidity for isotope identification	|R_innner| < 20GV

Training samples



Signal Sample

events in mass proton range 0.75 < m < 1.25 Gev/c^2

Background sample

events with low negative rigidity (-20 <R< 0) in the high-mass tail (2.36 < m <5 Gev/c^2)

Feature selection and classification pipeline

Input dataset

Balanced dataset (5 years): ~32000 events equally divided in signal and background

'Complete' set of features (from ToF and Tracker) + *'Physics Driven'* sets of features

Feature selection and classification pipeline

Input dataset

Balanced dataset (5 years): ~32000 events equally divided in signal and background

‘Complete’ set of features (from ToF and Tracker) + *‘Physics Driven’* sets of features

4 years (2015-2018)

Feature Selection (6 months)

Classification (3.5 years)

Feature selection and classification pipeline

Input dataset

Balanced dataset (5 years): ~32000 events equally divided in signal and background
'Complete' set of features (from ToF and Tracker) + *'Physics Driven'* sets of features

4 years (2015-2018)

Feature Selection (6 months)	Classification (3.5 years)

Feature selection

- select the most discriminant features from complete set
- 4 different methods (Random Forest, kbest, linear regression, Pearson's correlation)

Training + Test

- 2 BDt for each set of features (XGB vs AdaBoost)
- Use cross-validation to evaluate performance
- Test models and produce evaluation plots

Feature selection and classification pipeline

Input dataset
Balanced dataset (5 years): ~32000 events equally divided in signal and background
'Complete' set of features (from ToF and Tracker) + *'Physics Driven'* sets of features

4 years (2015-2018)		1 year (2023)
Feature Selection (6 months)	Classification (3.5 years)	Validation

Feature selection
- select the most discriminant features from complete set
- 4 different methods (Random Forest, kbest, linear regression, Pearson's correlation)

Training + Test
- 2 BDt for each set of features (XGB vs AdaBoost)
- Use cross-validation to evaluate performance
- Test models and produce evaluation plots

Test
Cross-check the models with a completely distinct dataset

Feature selection and classification pipeline

Input dataset

Balanced dataset (5 years): ~32000 events equally divided in signal and background

'Complete' set of features (from ToF and Tracker) + *'Physics Driven'* set of features

4 years (2015-2018)		1 year (2023)
Feature Selection (6 months)	Classification (3.5 years)	Validation

Feature selection

- select the most discriminant features from complete set
- 4 different methods (Random Forest, kbest, linear regression, Pearson's correlation)

Training + Test

- 2 BDt for each set of features (XGB vs AdaBoost)
- Use cross-validation to evaluate performance
- Test models and produce evaluation plots

Test

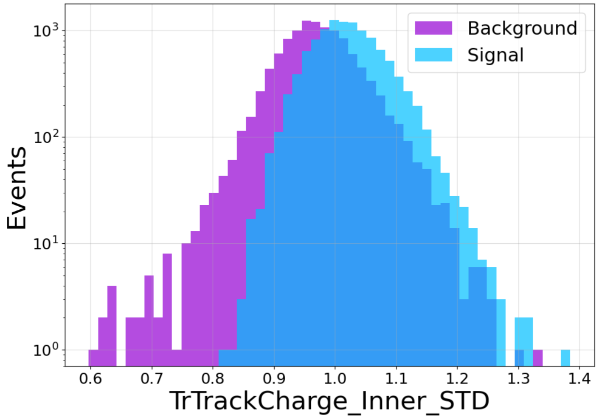
Cross-check the models with a completely distinct dataset

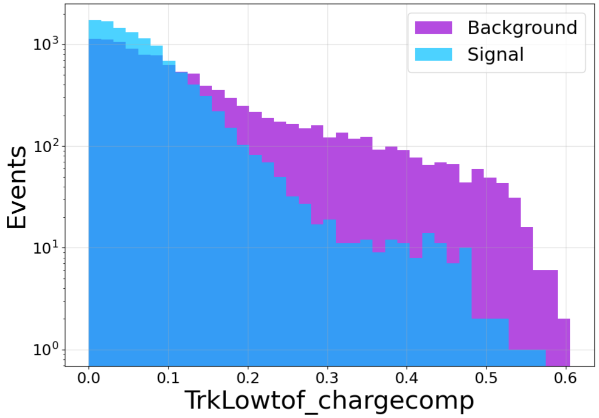
Physics Driven set

Chosen following a “physics-driven” approach based on the knowledge of the detectors and of the background to be rejected

How to choose the variables:

- Check that they do not introduce biases in the training phase (data leak)
- Check their discriminative power by looking at signal and background distributions





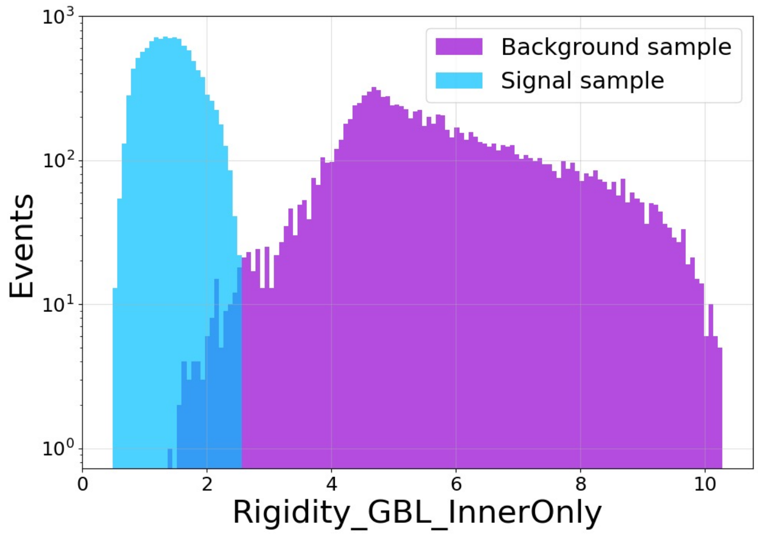
Data leak: dependence on rigidity

We choose the samples based on the event mass:

$$m = \frac{RZe}{\gamma\beta}$$

→ Features dependent on the rigidity introduce a bias in the training of the BDT

→ No true discriminative power, just given by our sample definition



Data leak: dependence on rigidity

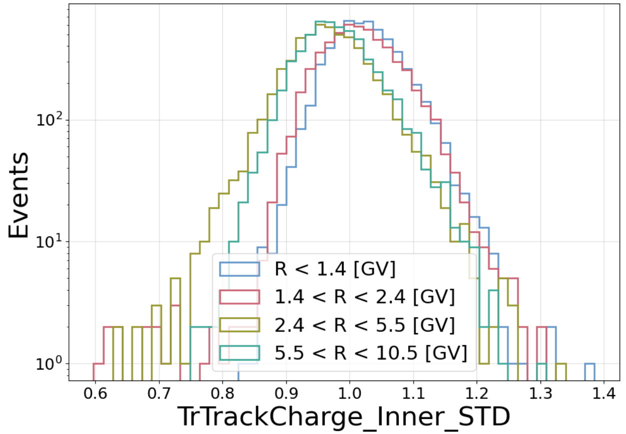
We choose the samples based on the event mass:

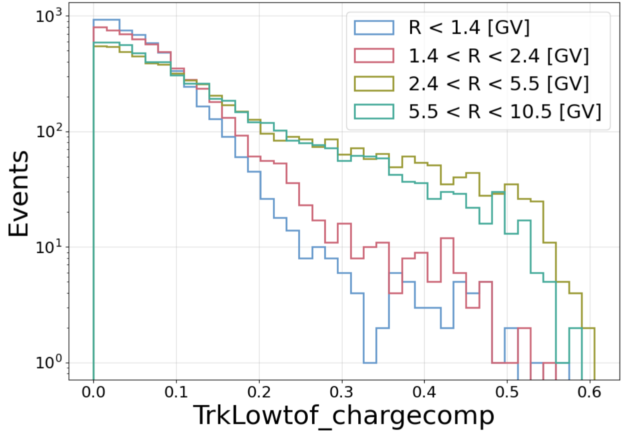
$$m = \frac{RZe}{\gamma\beta}$$

→ Features dependent on the rigidity introduce a bias in the training of the BDT

→ No true discriminative power, just given by our sample definition

→ Check dependence on the rigidty





Physics Driven set

- TrTrackChargeRMS_Inner_STD
- TrChiSq_GBLNoMS_InnerOnly_Y
- TrChiSq_GBNoMSL_InnerOnly_X
- LowerUpper_rigiditycomp

$$R_{LU} = \frac{|R_L - R_U|}{|R_L|}$$

- InnerPartial2_rigiditycomp

$$R_{IPl2} = \frac{|R_I - R_{Pl2}|}{|R_I|}$$

- TrkLowtof_chargecomp

$$Q_{IL} = \frac{|Q_{ITrack} - Q_{LToF}|}{|Q_{LToF}|}$$

- TrkUptof_chargecomp

$$Q_{IU} = \frac{|Q_{ITrack} - Q_{UToF}|}{|Q_{UToF}|}$$

- LowUptof_chargecomp

$$Q_{UL} = \frac{|Q_{UToF} - Q_{LToF}|}{|Q_{UToF}|}$$

Feature selection and classification pipeline

Input dataset

Balanced dataset (5 years): ~32000 events equally divided in signal and background

'Complete' set of features (from ToF and Tracker) + *'Physics Driven'* set of features

4 years (2015-2018)		1 year (2023)
Feature Selection (6 months)	Classification (3.5 years)	Validation

Feature selection

- select the most discriminant features from complete set
- 4 different methods (Random Forest, kbest, linear regression, Pearson's correlation)

Training + Test

- 2 BDt for each set of features (XGB vs AdaBoost)
- Use cross-validation to evaluate performance
- Test models and produce evaluation plots

Test

Cross-check the models with a completely distinct dataset

Rigidity dependence – statistical tests

We need a way to test the rigidity dependence of ~3000 features
Performed different statistical tests on the initial set of features :

- Spearman correction

test null hypothesis of no correlation

- Krukal-Wallis (KW)
- Kolmogorov Smirnov (KS)

tests if samples are drawn from the same distribution

We compute the p-value for each of the 3 statistical measures
We exclude features with p-value < 0.05.

→ We reject features for which there is less than 5% probability to measure the value of the statistic obtained with the chosen test given the null hypothesis

Rigidity dependence – statistical tests

Tested 3067 features from Tracker and Tof.

Number of features passing the tests:

- Spearman correlation: 710
- Krukal-Wallis: 670
- Kolmogorov Smirnov: 265

- Kolmogorov Smirnov (KS) is the more conservative
- The KS features seem to be independent of rigidity from visual inspection
- Features don’t seem to be very discriminative

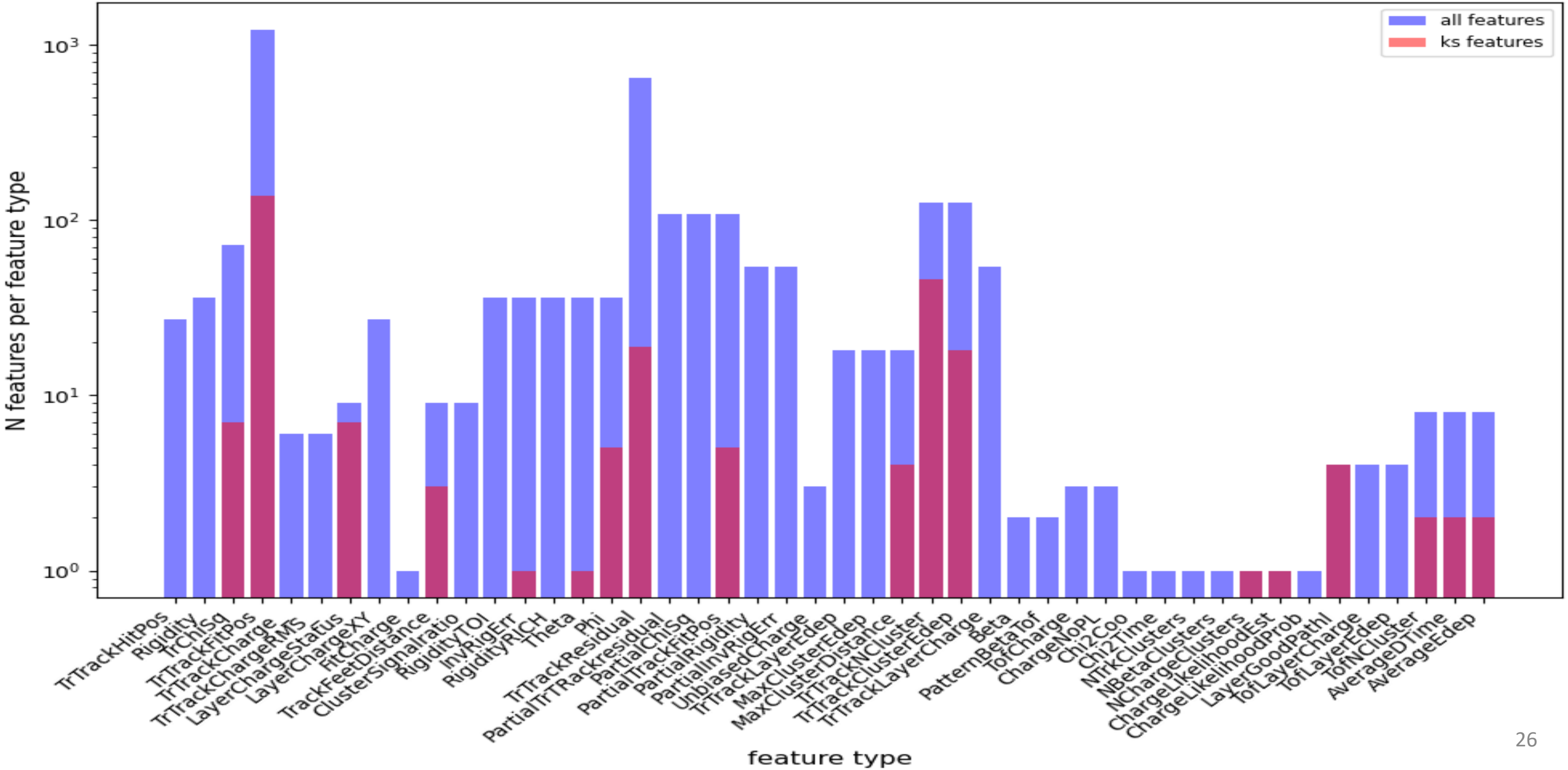
Rigidity dependence – statistical tests

Number of features passing the tests:

- Spearman correlation: 710
- Krukal-Wallis: 670
- Kolmogorov Smirnov test: 265 → Complete set

- Kolmogorov Smirnov (KS) is the more conservative
- The KS features seem to be independent of rigidity from visual inspection

Features per type – complete set



Feature selection and classification pipeline

Input dataset

Balanced dataset (5 years): ~32000 events equally divided in signal and background

'Complete' set of features (from ToF and Tracker) + *'Physics Driven'* sets of features

4 years (2015-2018)		1 year (2023)
Feature Selection (6 months)	Classification (3.5 years)	Validation

Feature selection

- select the most discriminant features from complete set
- 4 different methods (Random Forest, kbest, linear regression, Pearson's correlation)

Training + Test

- 2 BDt for each set of features (XGB vs AdaBoost)
- Use cross-validation to evaluate performance
- Test models and produce evaluation plots

Test

Cross-check the models with a completely distinct dataset

Feature selection

We input in the pipeline 265 features passing KS test

Technique	Number of features selected
kbest	103
Random Forest	53
Pearson’s correlation	160
Linear regression	1

Feature sets used for the training

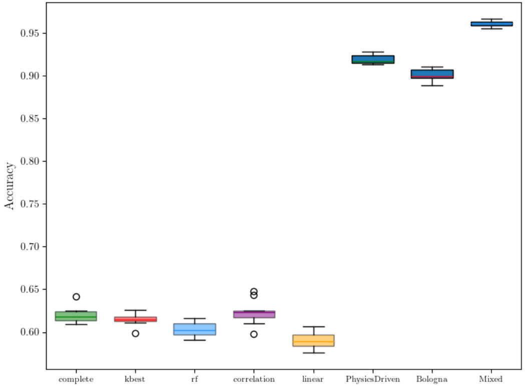
We input in the pipeline 265 features passing KS test

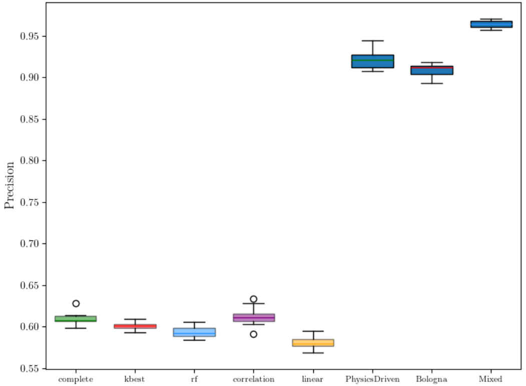
Technique	Number of features selected
kbest	103
Random Forest	53
Pearson’s correlation	160
Linear regression	1
Physics driven (Groningen)	8
Bologna	56
Mixed	63

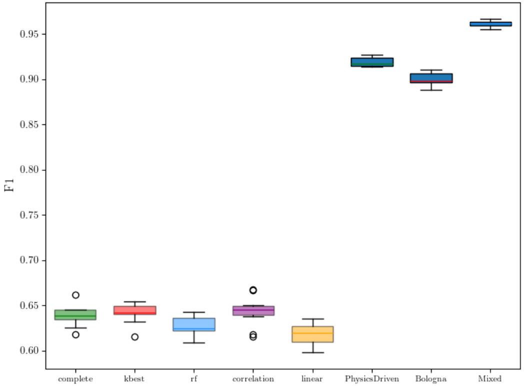
Mixed set = Physics Driven + Bologna

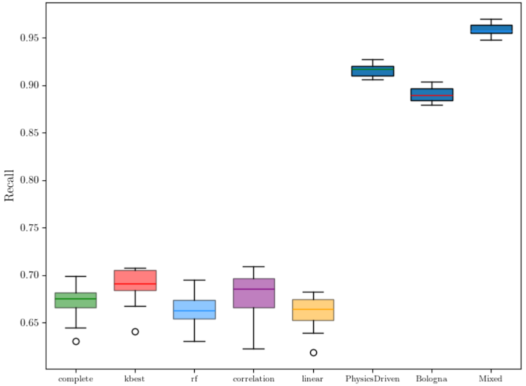
N.B. Physics Driven features and Bologna features do not pass the KS test with 0.05 threshold

Performance metrics - XGBoost









The BDTs trained with physics driven sets perform better than the others selected with feature selection algorithms

The BDT trained with Mixed set shows the best performances

Performance metrics – XGBoost vs AdaBoost

XGBoost

Metrics	complete	kbest	rf	correlation	linear	PhysicsDriven	Bologna	Mixed
Accuracy	0.6196 ± 0.0182	0.6146 ± 0.0134	0.6032 ± 0.0166	0.6232 ± 0.0274	0.5907 ± 0.020	0.9188 ± 0.0108	0.9002 ± 0.0146	0.9613 ± 0.0066
Precision	0.61 ± 0.015	0.6012 ± 0.0088	0.5936 ± 0.0134	0.6123 ± 0.023	0.581 ± 0.017	0.9218 ± 0.022	0.9087 ± 0.016	0.9643 ± 0.0086
Recall	0.6706 ± 0.0386	0.688 ± 0.04	0.663 ± 0.0346	0.679 ± 0.0524	0.6604 ± 0.038	0.916 ± 0.013	0.8905 ± 0.0164	0.9583 ± 0.0126
F1-Score	0.6388 ± 0.0228	0.6416 ± 0.021	0.6263 ± 0.0208	0.6437 ± 0.0324	0.6181 ± 0.024	0.9188 ± 0.01	0.8995 ± 0.0148	0.9613 ± 0.0068
ROC-AUC	0.6783 ± 0.0232	0.6671 ± 0.0238	0.6506 ± 0.023	0.6808 ± 0.028	0.6405 ± 0.019	0.9758 ± 0.0052	0.9665 ± 0.0076	0.9938 ± 0.0024

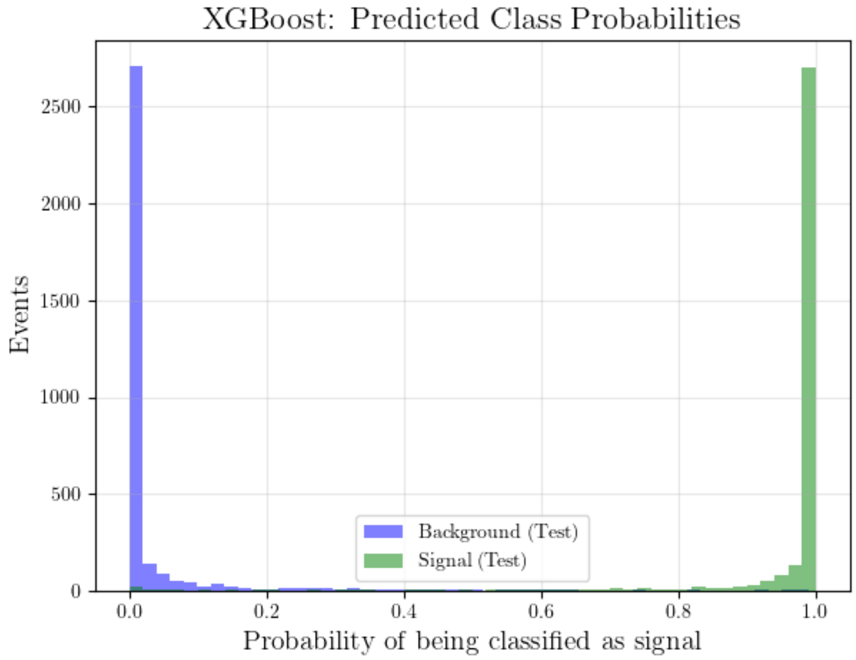
AdaBoost

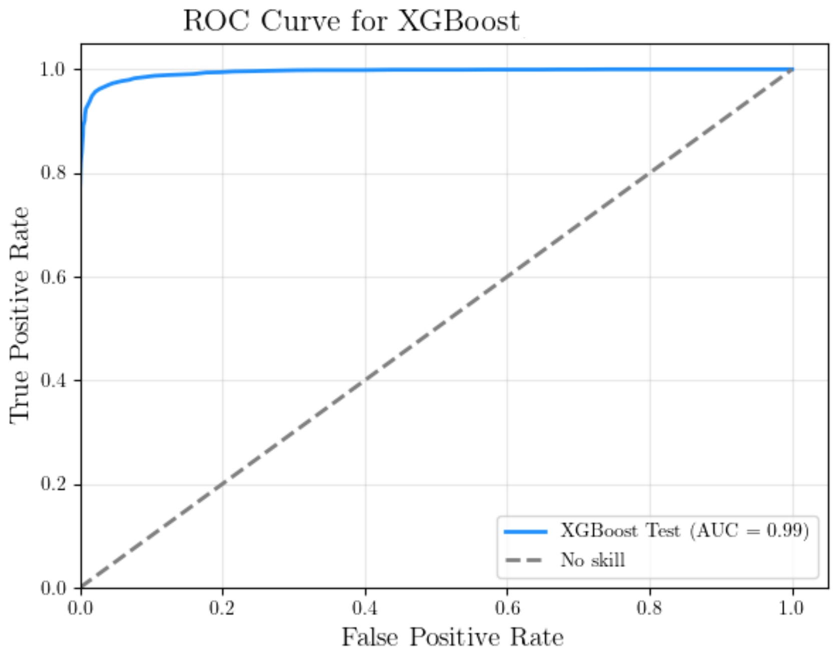
Metrics	complete	kbest	rf	correlation	linear	PhysicsDriven	Bologna	Mixed
Accuracy	0.6299 ± 0.0148	0.6293 ± 0.0206	0.6059 ± 0.0172	0.6284 ± 0.0222	0.5943 ± 0.0168	0.9093 ± 0.0134	0.809 ± 0.0168	0.9227 ± 0.0108
Precision	0.6048 ± 0.0126	0.6025 ± 0.016	0.5856 ± 0.013	0.6023 ± 0.015	0.5775 ± 0.015	0.9156 ± 0.0154	0.8331 ± 0.0184	0.9301 ± 0.017
Recall	0.7575 ± 0.033	0.7678 ± 0.0388	0.7336 ± 0.0354	0.763 ± 0.046	0.7133 ± 0.0428	0.9024 ± 0.0204	0.7746 ± 0.032	0.9149 ± 0.0144
F1-Score	0.6725 ± 0.016	0.6751 ± 0.0216	0.6512 ± 0.0198	0.6731 ± 0.026	0.6381 ± 0.0196	0.9089 ± 0.0138	0.8027 ± 0.0196	0.9224 ± 0.0106
ROC-AUC	0.6863 ± 0.0192	0.6849 ± 0.0208	0.6575 ± 0.0264	0.6862 ± 0.0222	0.6426 ± 0.0236	0.9691 ± 0.0062	0.8869 ± 0.0134	0.9763 ± 0.0046

In general, XGBoost models show slightly better metrics than AdaBoost ones

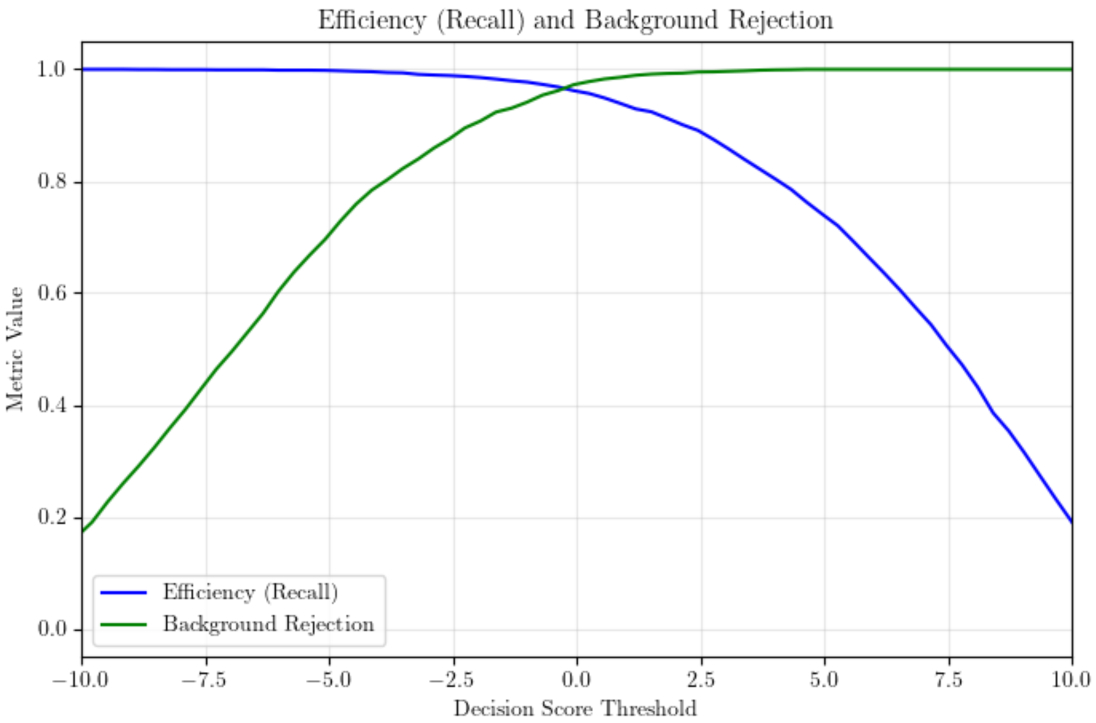
Performance Mixed set - XGBoost

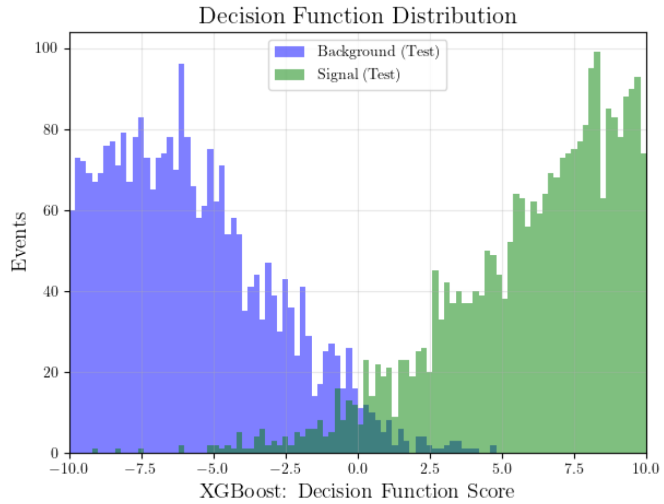
XGBoost model trained with the Mixed features set is the best performing method



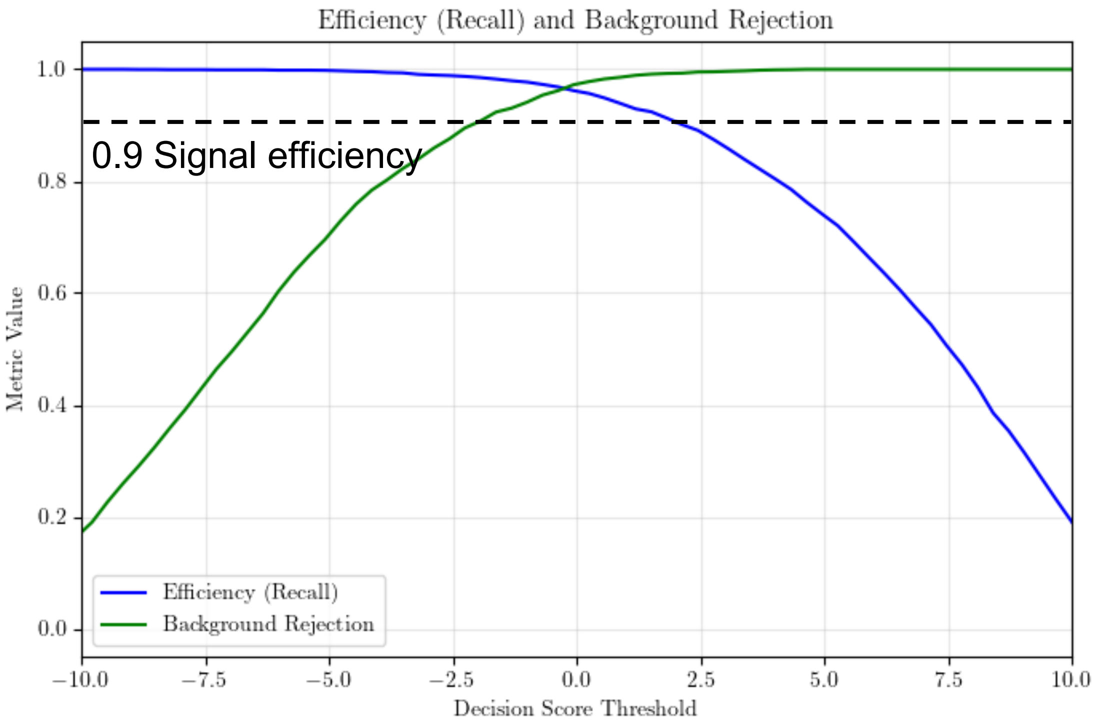


Performance Mixed set - XGBoost

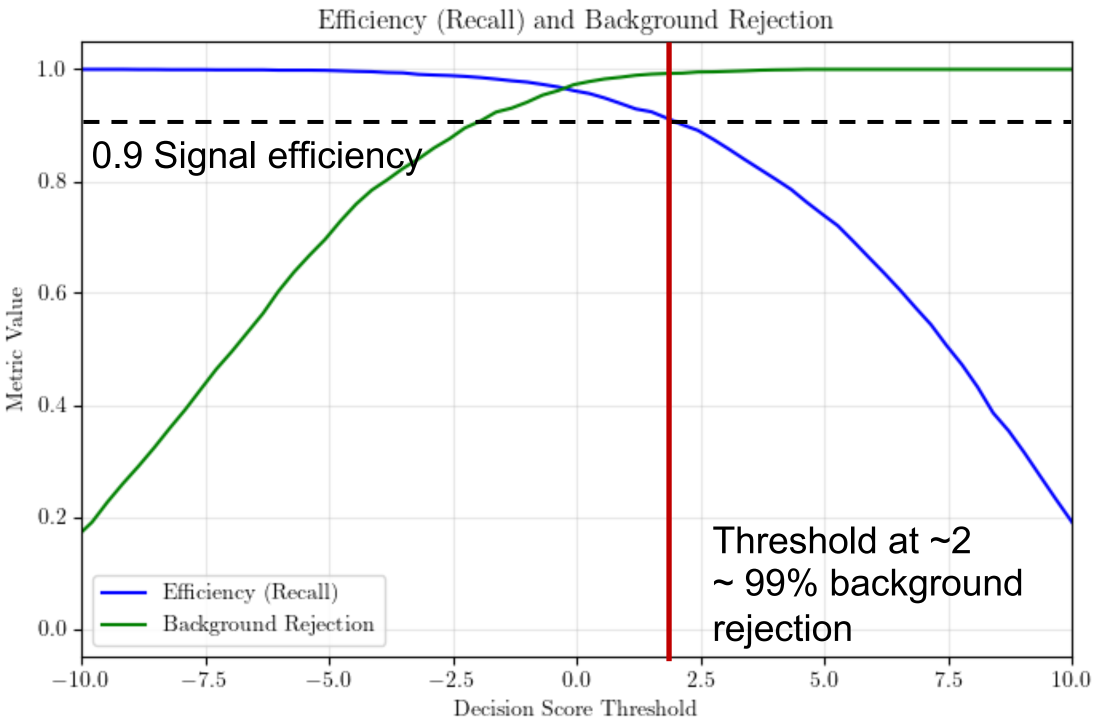




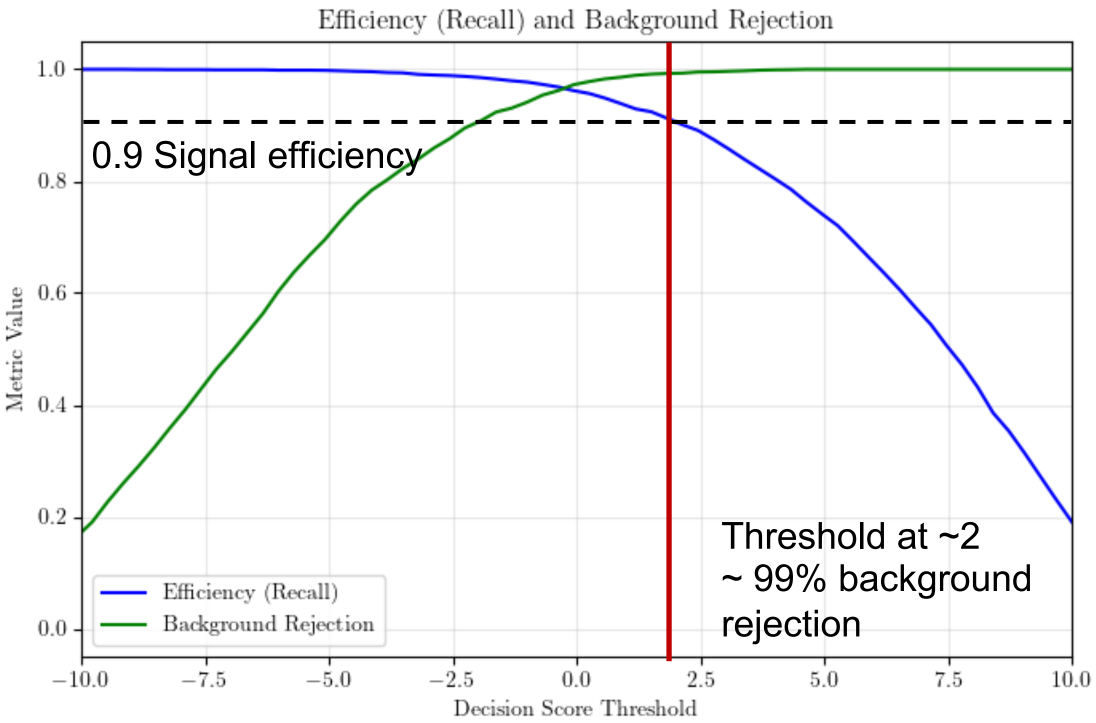
Performance Mixed set - XGBoost



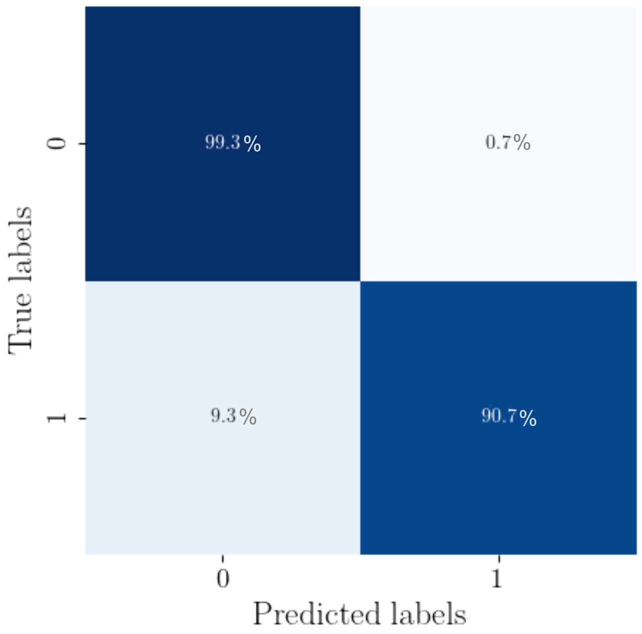
Performance Mixed set - XGBoost



Performance Mixed set - XGBoost



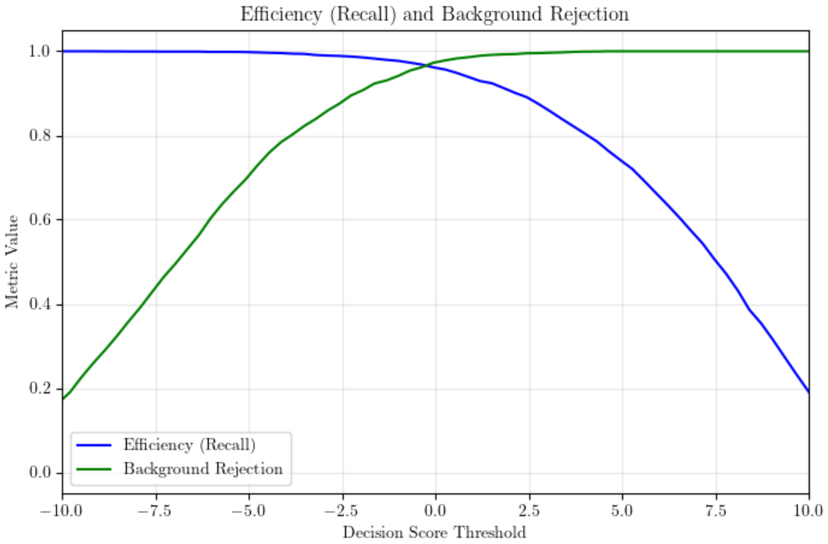
0 → background
1 → signal



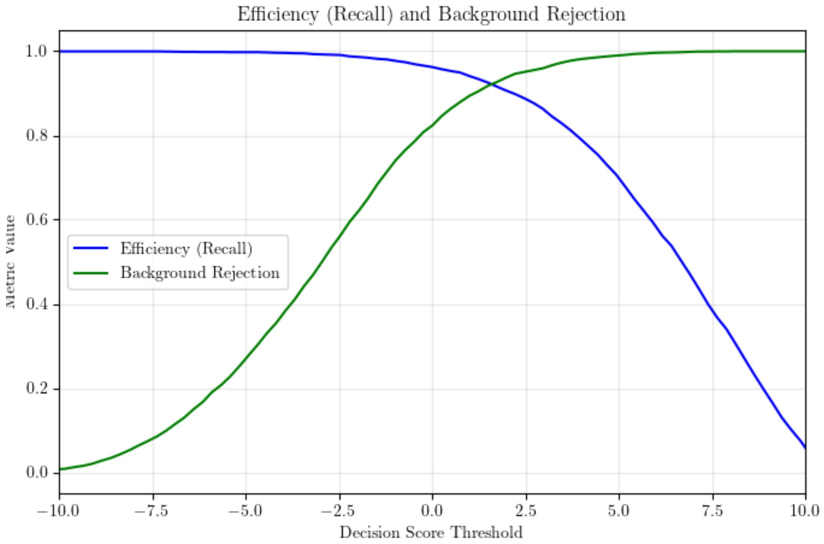
Validation on the 2023 dataset

We validate the models performance on a time-wise distinct dataset of 1 year (2023):

→ *All the models perform slightly worse than on the regular validation dataset*



Validation 2015-2018 dataset

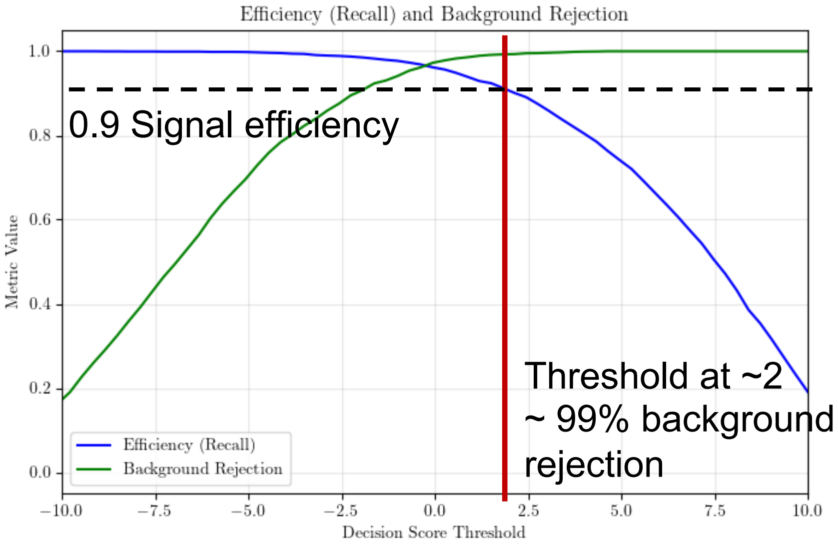


Validation 2023 dataset

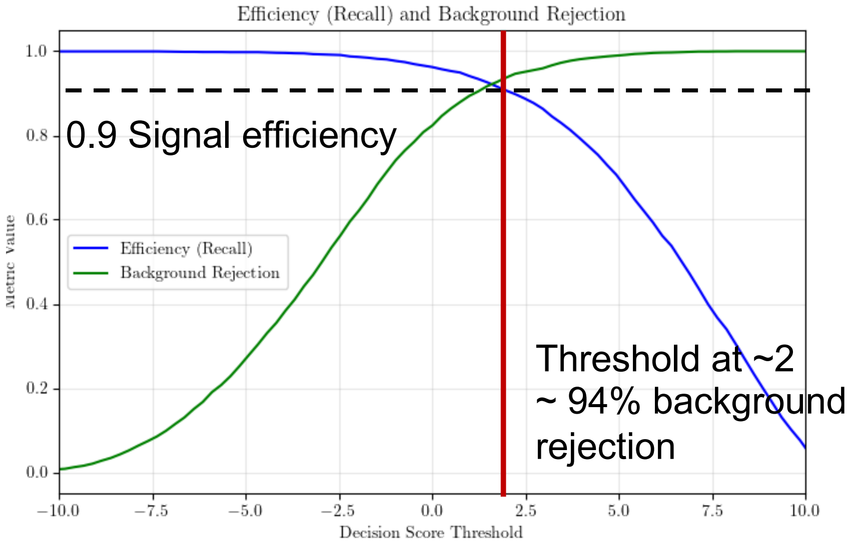
Validation on the 2023 dataset

We validate the models performance on a time-wise distinct dataset of 1 year (2023):

→ *All the models perform slightly worse than on the regular validation dataset*



Validation 2015-2018 dataset

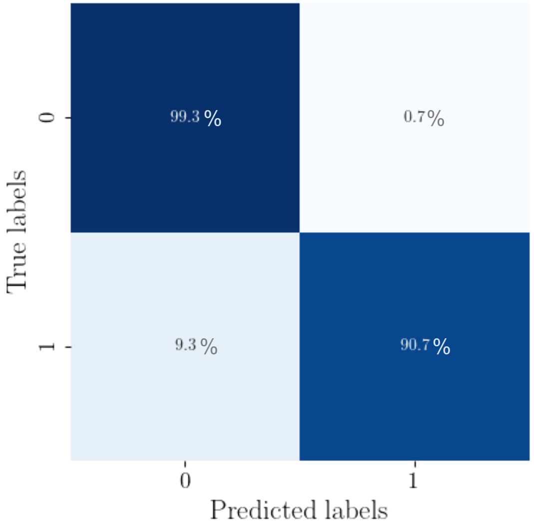


Validation 2023 dataset

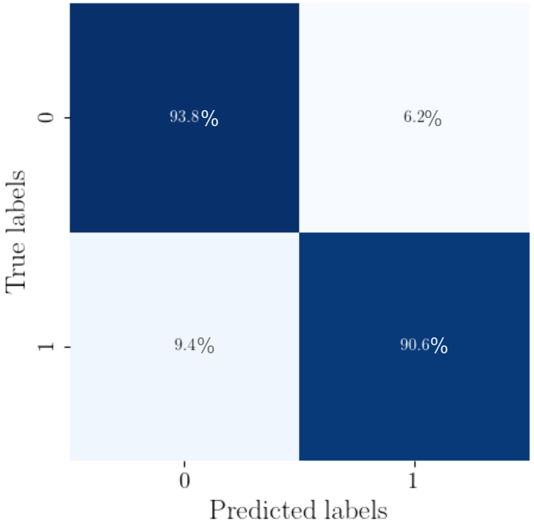
Validation on the 2023 dataset

We validate the models performance on a time-wise distinct dataset of 1 year (2023):

→ *All the models perform slightly worse than on the regular validation dataset*



Validation 2015-2018 dataset



Validation 2023 dataset

Conclusions

- ✓ We build a charge confusion estimator to reject the background events with wrongly reconstructed charge sign
- ✓ We use different methods to select features capable of discriminating between background and signal events
- ✓ We check for possible data leakages coming from rigidity dependent features using statistical test (e.g. KS) and visual inspection

- → Features passing Kolmogorov-Smirnov test lead to poor performances in the BDT models trained with them
- → The BDT trained using XGBoost with ‘Mixed’ features set leads to the best perfomances
- → All the models perform slightly worse in the validation on the 2023 dataset

Open questions

→ Features passing Kolmogorov-Smirnov test lead to poor performances in the BDT models trained with them

→ KS test is too much conservative?

- Check set of features passing the other two tests
- Enlarge dataset releasing requirement on beta ToF (beta_ToF < 0.9)
- Find another way of assessing the rigidity dependence

→ All the models perform slightly worse in the validation on the 2023 dataset

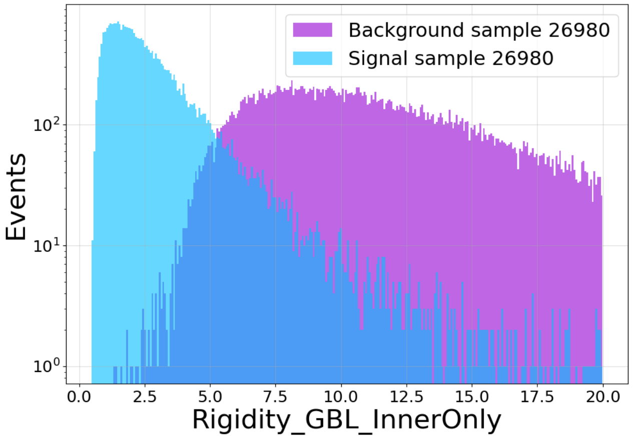
→ The model is not generalizing?

- Check generalization capabilities of the models on another dataset of one year
- Cross-check perfomances of the Bologna dataset with Francesco’s ntuples

BACK UP

Dataset without beta requirements

Rigidity distributions now overlap more.

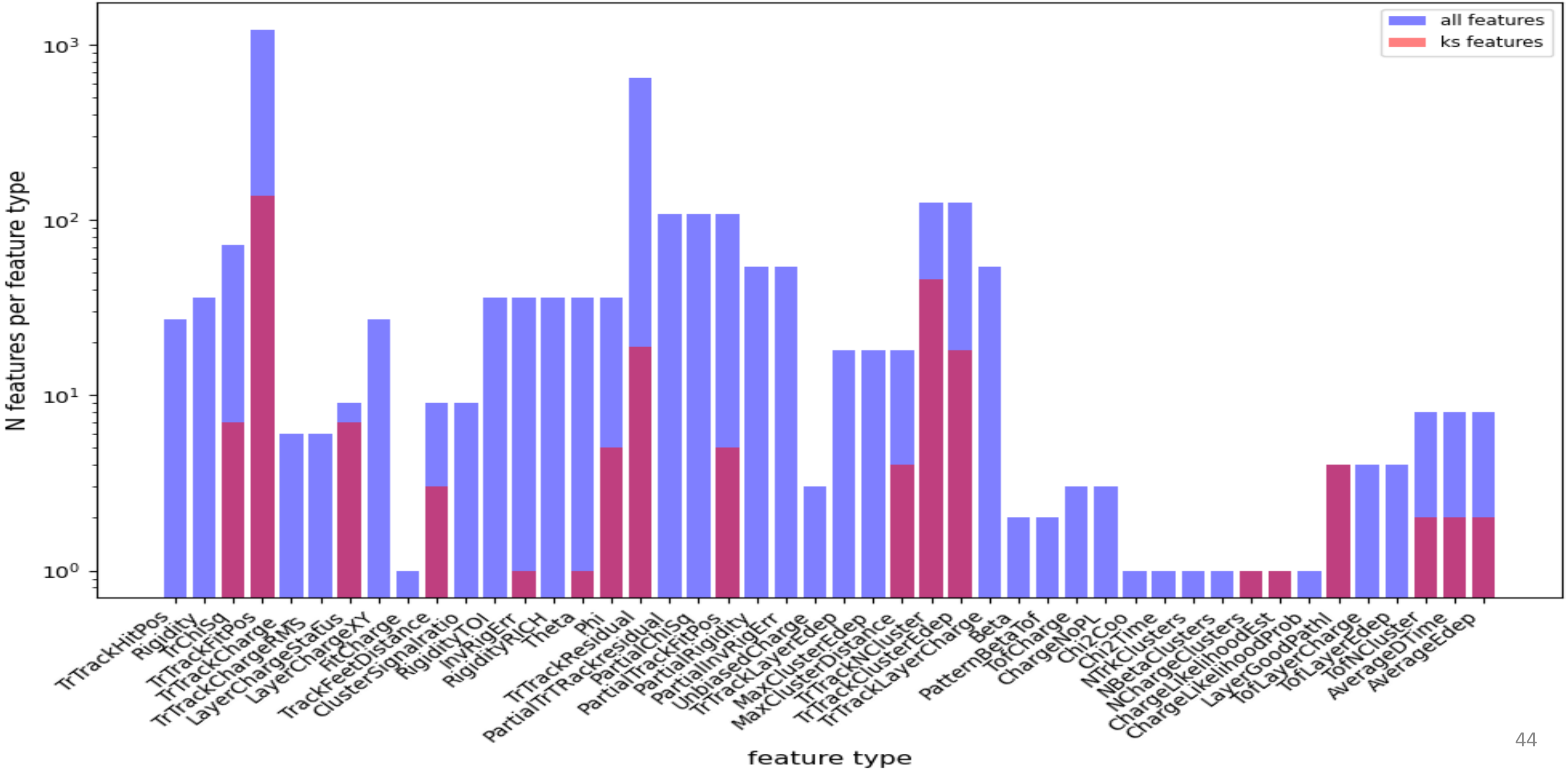


Number of features passing the tests:

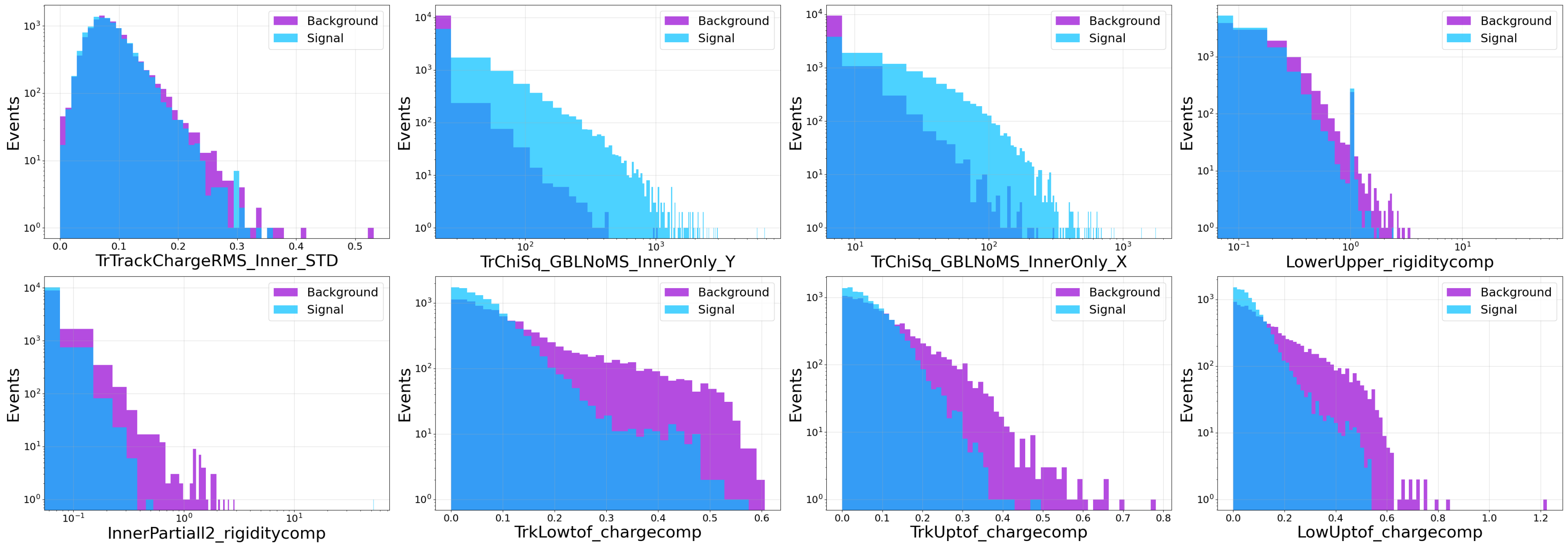
- Kolmogorov Smirnov test: 450
- Kruskal-Wallis: 715
- Spearman correlation: 932

- Kolmogorov Smirnov (KS) is the more conservative
- The KS features seem to be independent of rigidity from visual inspection
- More features pass the test but they still don't seem to be very discriminative

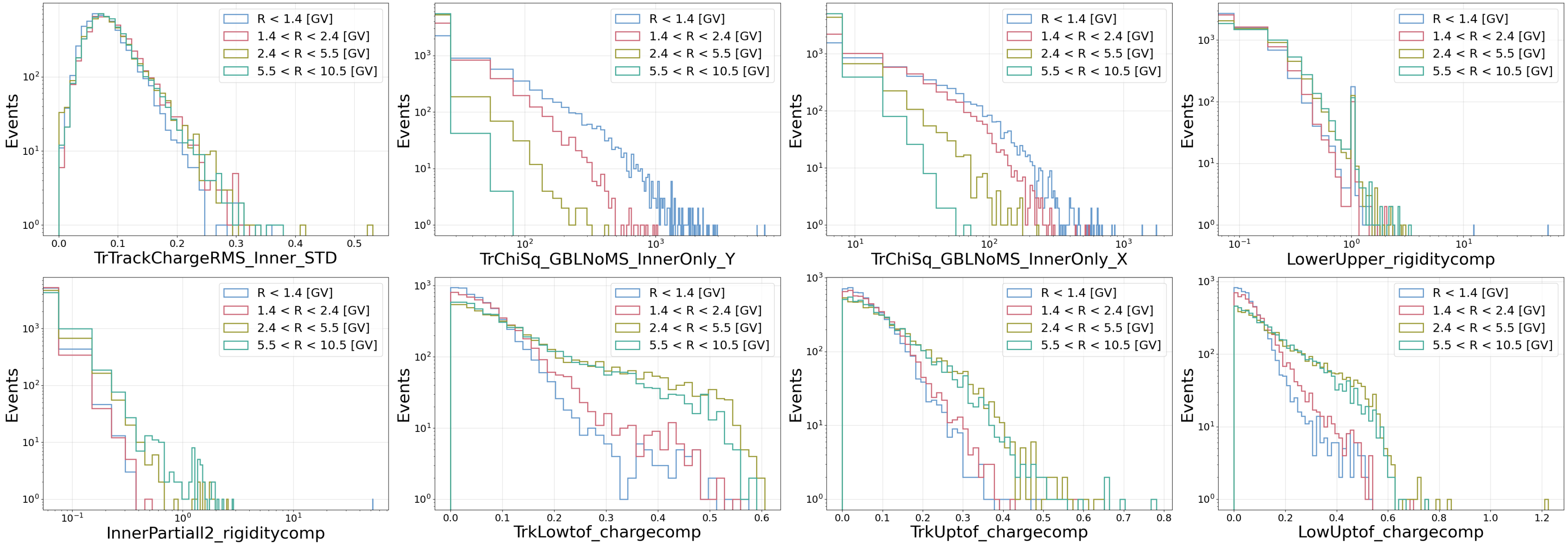
Features per type –KS set



Physics Driven



Physics Driven

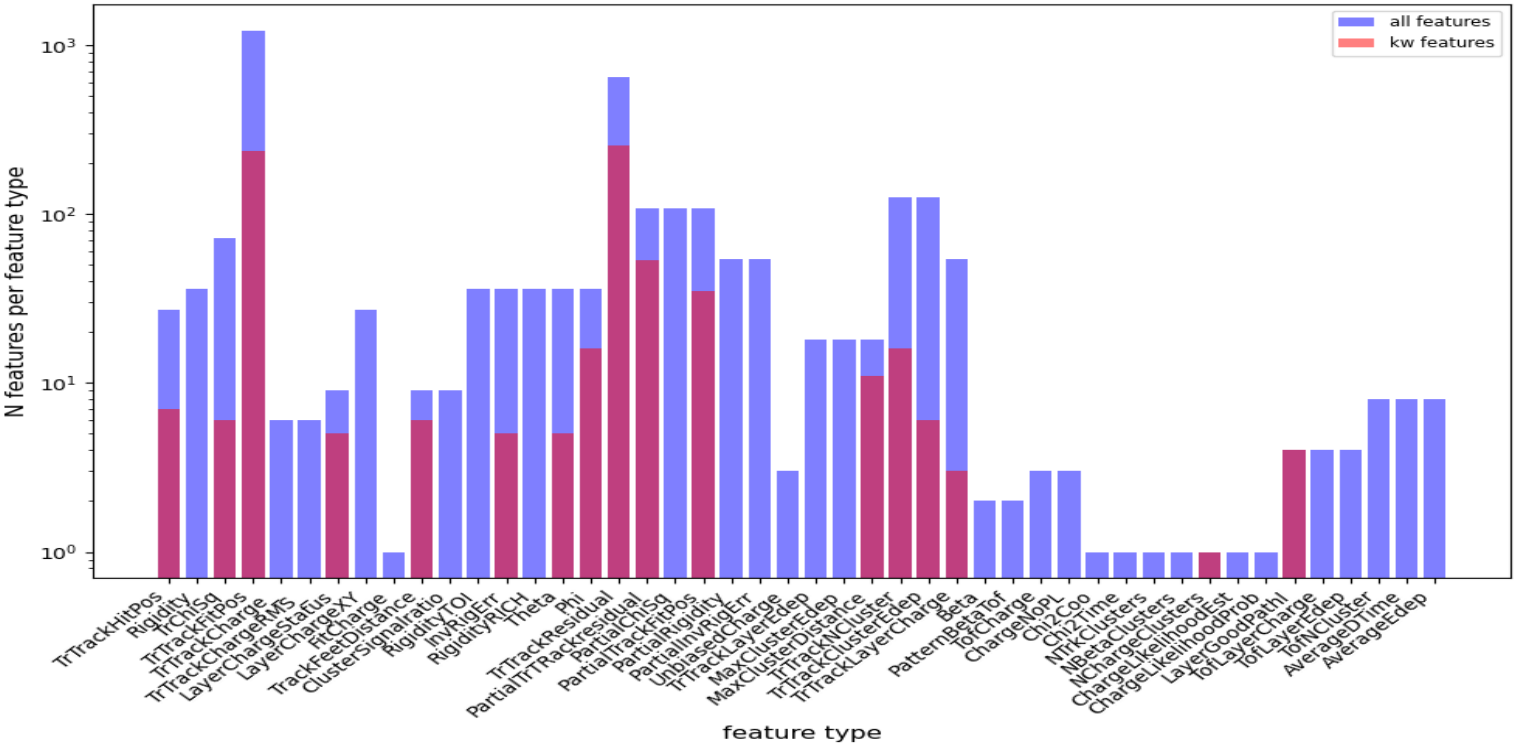


Bologna features

- N° 1 ChisquareX (IT)
- N°7 PartialR asymmetries (L2-L8) -> $\text{Partial R asymmetry}_{\text{i}} = \frac{\text{PartialR}_{\text{i}} - \text{R}_{\text{rec}}}{\text{R}_{\text{rec}}}$
- N°7 NCluster (IT), Side X, 1mm (L2-L8)
- N°7 NCluster (IT), Side X, 1cm from Track (L2-L8)
- N°7 NCluster (IT), Side X, 2cm from Track (L2-L8)
- N°7 NCluster (IT), Side Y, 1mm (L2-L8)
- N°7 NCluster (IT), Side Y, 1cm from Track (L2-L8)
- N°7 NCluster (IT), Side Y, 2cm from Track (L2-L8)
- NTotalCluster (IT), SideX, 1mm (sum on L2-L8)
- NTotalCluster (IT), SideX, 1cm (sum on L2-L8)
- NTotalCluster (IT), SideX, 2cm (sum on L2-L8)
- NTotalCluster (IT), SideY, 1mm (sum on L2-L8)
- NTotalCluster (IT), SideY, 1cm (sum on L2-L8)
- NTotalCluster (IT), SideY, 2cm (sum on L2-L8)

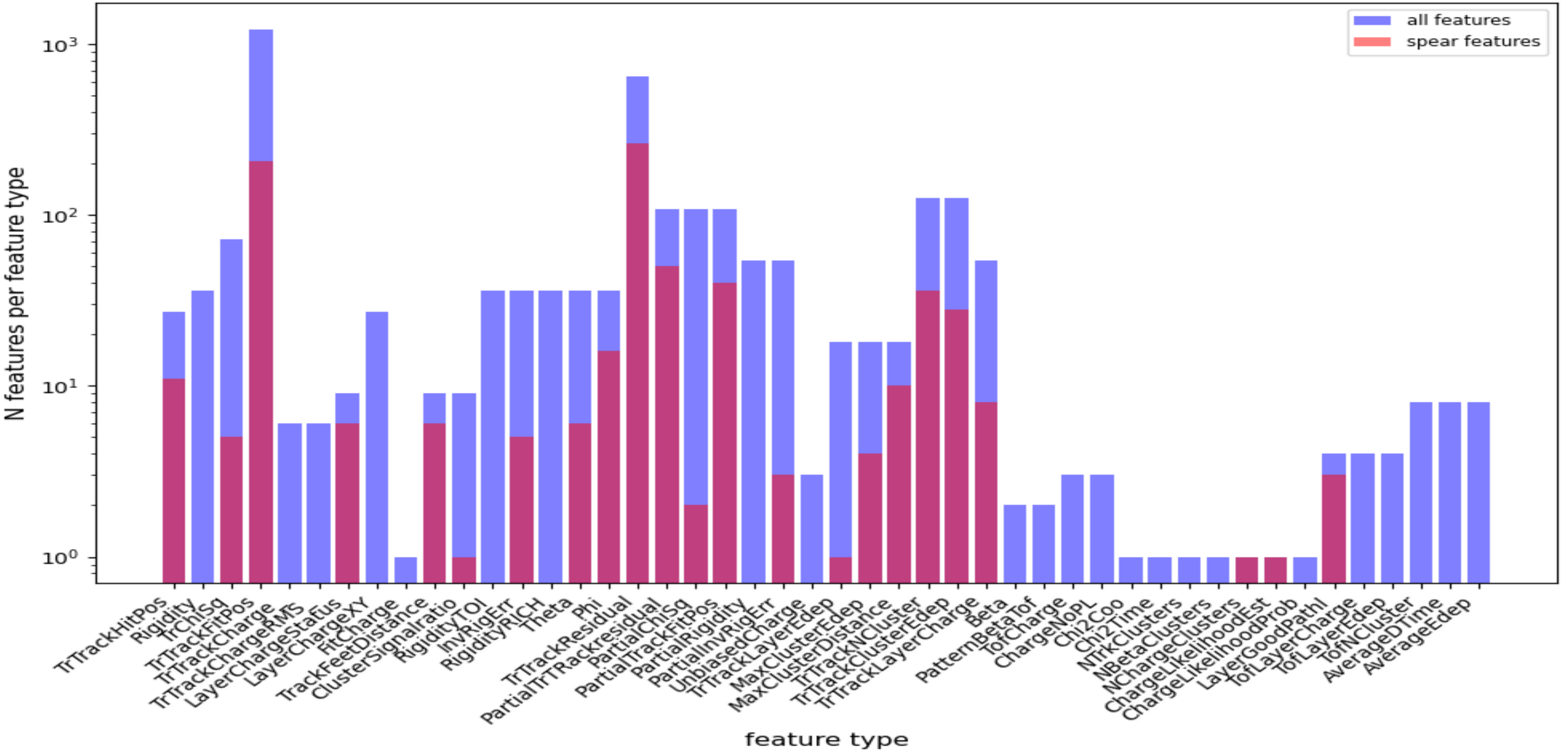
Features per type - kw set

Features considered: 3067 features from Tracker and ToF detectors



Features per type – spearman set

Features considered: 3067 features from Tracker and ToF detectors



Performance metrics - XGBoost

Metrics	complete	kbest	rf	correlation	linear	PhysicsDriven	Bologna	Mixed
Accuracy	0.6196 ± 0.0182	0.6146 ± 0.0134	0.6032 ± 0.0166	0.6232 ± 0.0274	0.5907 ± 0.0202	0.9188 ± 0.0108	0.9002 ± 0.0146	0.9613 ± 0.0066
Precision	0.61 ± 0.015	0.6012 ± 0.0088	0.5936 ± 0.0134	0.6123 ± 0.023	0.581 ± 0.017	0.9218 ± 0.022	0.9087 ± 0.016	0.9643 ± 0.0086
Recall	0.6706 ± 0.0386	0.688 ± 0.04	0.663 ± 0.0346	0.679 ± 0.0524	0.6604 ± 0.0384	0.916 ± 0.013	0.8905 ± 0.0164	0.9583 ± 0.0126
F1-Score	0.6388 ± 0.0228	0.6416 ± 0.021	0.6263 ± 0.0208	0.6437 ± 0.0324	0.6181 ± 0.024	0.9188 ± 0.01	0.8995 ± 0.0148	0.9613 ± 0.0068
ROC-AUC	0.6783 ± 0.0232	0.6671 ± 0.0238	0.6506 ± 0.023	0.6808 ± 0.028	0.6405 ± 0.0194	0.9758 ± 0.0052	0.9665 ± 0.0076	0.9938 ± 0.0024

Metrics	complete	kbest	rf	correlation	linear	PhysicsDriven	Bologna	Mixed
Accuracy	0,6145	0,6148	0,5948	0,6217	0,5925	0,917	0,8994	0,9673
Precision	0,6167	0,6178	0,597	0,6239	0,5949	0,917	0,8996	0,9674
Recall	0,6145	0,6148	0,5948	0,6217	0,5925	0,917	0,8994	0,9673
F1-Score	0,6132	0,6129	0,593	0,6204	0,5906	0,917	0,8994	0,9673
ROC-AUC	0,615	0,6154	0,5953	0,6222	0,5931	0,917	0,8994	0,9672

Metrics	complete	kbest	rf	correlation	linear	PhysicsDriven	Bologna	Mixed
Accuracy	0,5498	0,5419	0,5014	0,5142	0,5873	0,8837	0,8149	0,8934
Precision	0,5593	0,5901	0,5242	0,5856	0,5897	0,8851	0,8174	0,9012
Recall	0,5498	0,5419	0,5014	0,5142	0,5873	0,8837	0,8149	0,8934
F1-Score	0,5311	0,4713	0,3481	0,386	0,5845	0,8836	0,8145	0,8928
ROC-AUC	0,5498	0,5419	0,5014	0,5142	0,5873	0,8837	0,8149	0,8934

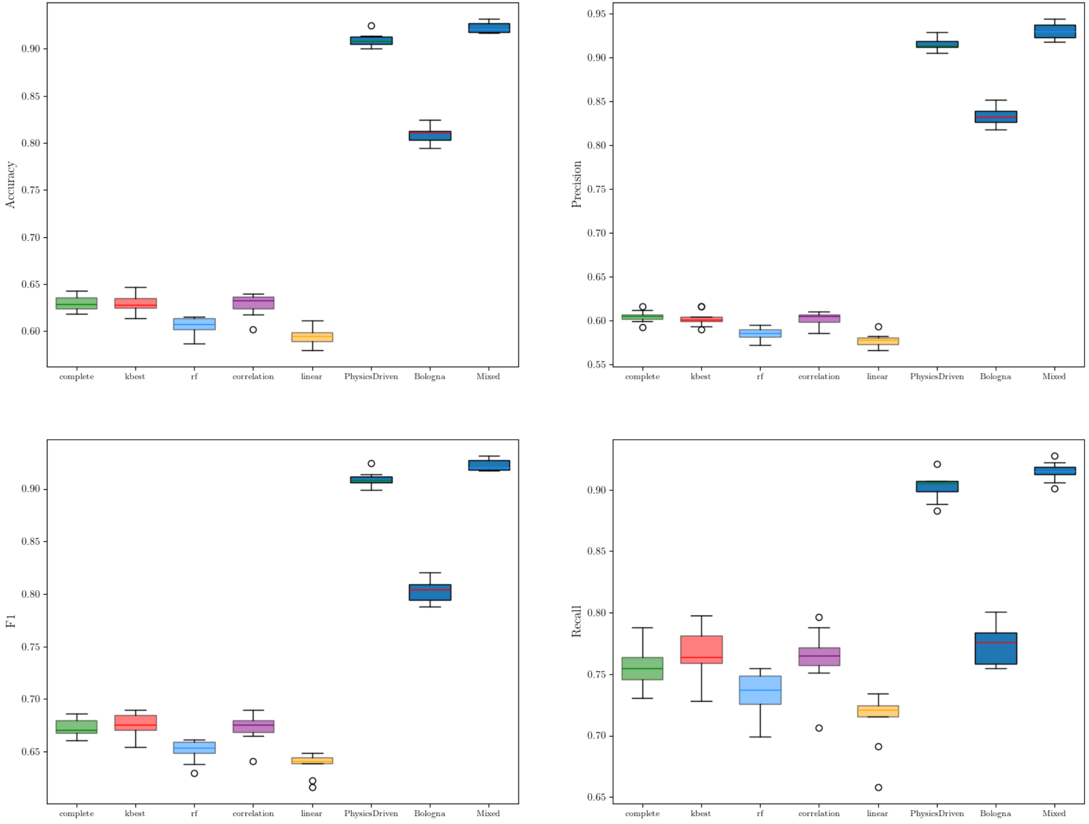
Performance metrics - AdaBoost

Metrics	complete	kbest	rf	correlation	linear	PhysicsDriven	Bologna	Mixed
Accuracy	0.6299 ± 0.0148	0.6293 ± 0.0206	0.6059 ± 0.0172	0.6284 ± 0.0222	0.5943 ± 0.0168	0.9093 ± 0.0134	0.809 ± 0.0168	0.9227 ± 0.0108
Precision	0.6048 ± 0.0126	0.6025 ± 0.016	0.5856 ± 0.013	0.6023 ± 0.015	0.5775 ± 0.015	0.9156 ± 0.0154	0.8331 ± 0.0184	0.9301 ± 0.017
Recall	0.7575 ± 0.033	0.7678 ± 0.0388	0.7336 ± 0.0354	0.763 ± 0.046	0.7133 ± 0.0428	0.9024 ± 0.0204	0.7746 ± 0.032	0.9149 ± 0.0144
F1-Score	0.6725 ± 0.016	0.6751 ± 0.0216	0.6512 ± 0.0198	0.6731 ± 0.026	0.6381 ± 0.0196	0.9089 ± 0.0138	0.8027 ± 0.0196	0.9224 ± 0.0106
ROC-AUC	0.6863 ± 0.0192	0.6849 ± 0.0208	0.6575 ± 0.0264	0.6862 ± 0.0222	0.6426 ± 0.0236	0.9691 ± 0.0062	0.8869 ± 0.0134	0.9763 ± 0.0046

Metrics	complete	kbest	rf	correlation	linear	PhysicsDriven	Bologna	Mixed
Accuracy	0,6311	0,6287	0,6122	0,6278	0,5952	0,9049	0,8063	0,9226
Precision	0,6419	0,6413	0,6214	0,6401	0,6017	0,9049	0,8077	0,9228
Recall	0,6311	0,6287	0,6122	0,6278	0,5952	0,9049	0,8063	0,9226
F1-Score	0,6246	0,621	0,6055	0,6202	0,5896	0,9049	0,806	0,9226
ROC-AUC	0,6321	0,6298	0,6132	0,6289	0,5962	0,9049	0,806	0,9226

Metrics	complete	kbest	rf	correlation	linear	PhysicsDriven	Bologna	Mixed
Accuracy	0,5185	0,524	0,5009	0,5172	0,5925	0,8756	0,7323	0,8758
Precision	0,605	0,6173	0,7502	0,6146	0,6001	0,8797	0,7375	0,8807
Recall	0,5185	0,524	0,5009	0,5172	0,5925	0,8756	0,7323	0,8758
F1-Score	0,3935	0,4059	0,3353	0,387	0,5846	0,8753	0,7308	0,8754
ROC-AUC	0,5185	0,524	0,5009	0,5172	0,5925	0,8756	0,7323	0,8758

Performance metrics - AdaBoost

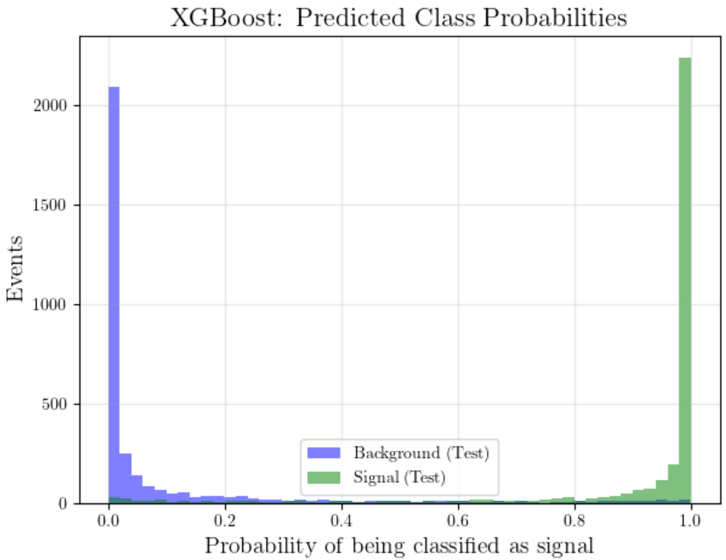


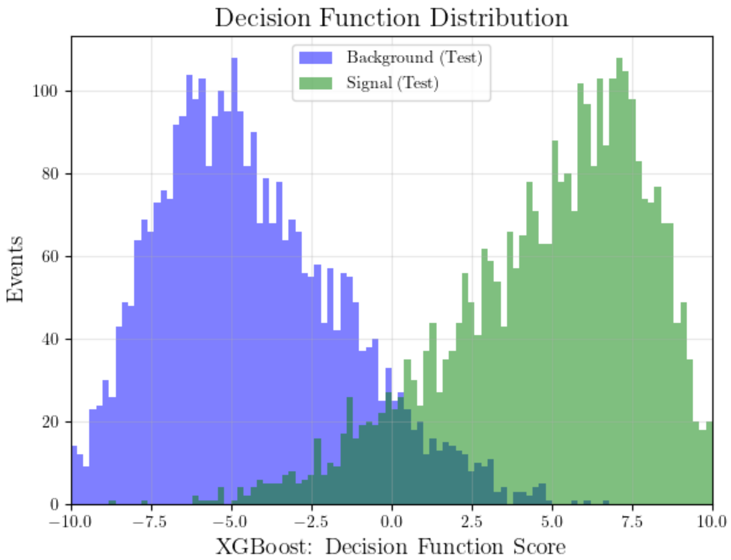
Decision score and probability distributions

Model XGBoost

Feature set: Groningen physics driven

Dataset: Validation



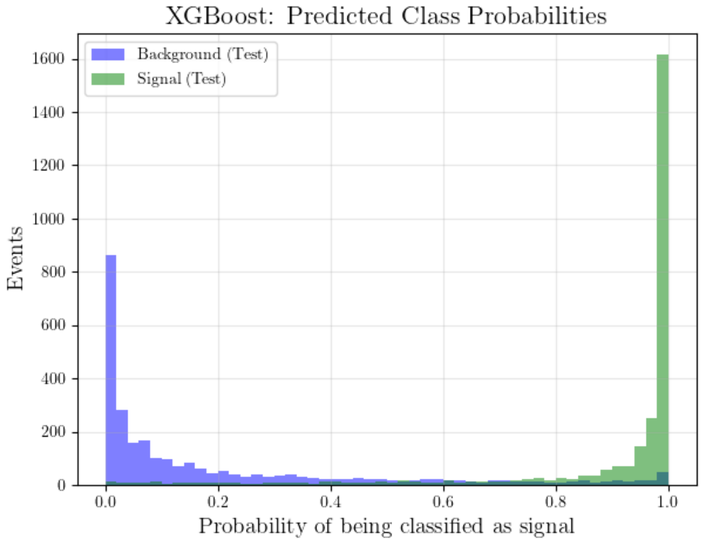


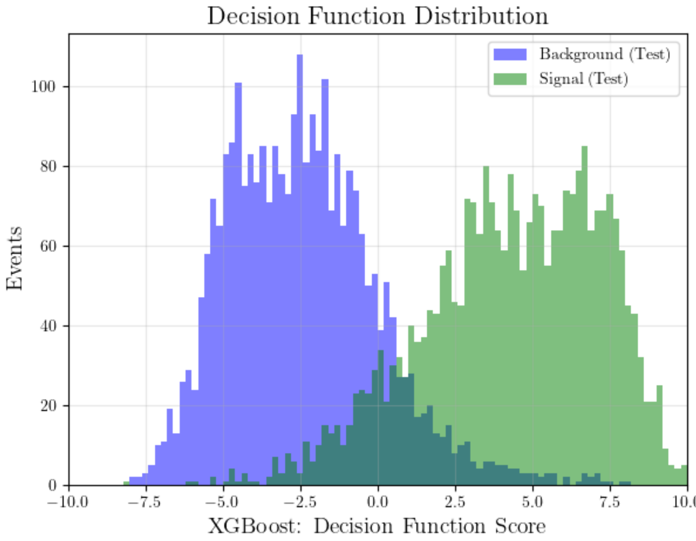
Decision score and probability distributions

Model XGBoost

Feature set: Groningen physics driven

Dataset : Validation 2023



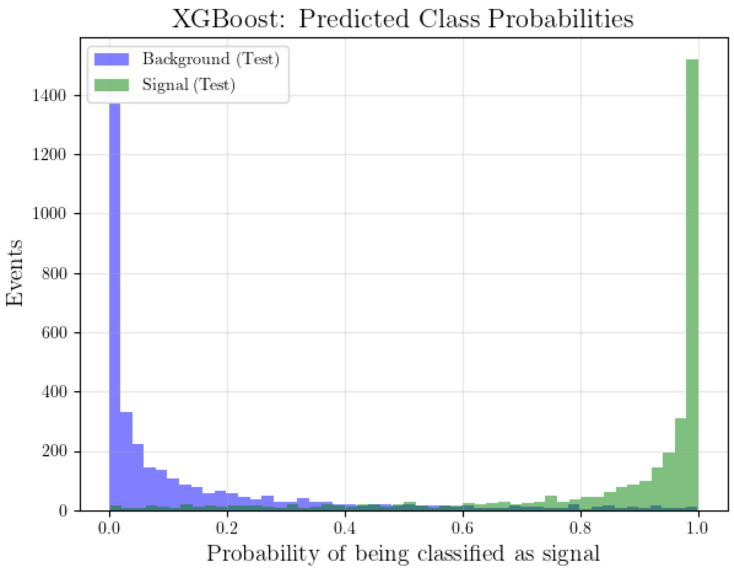


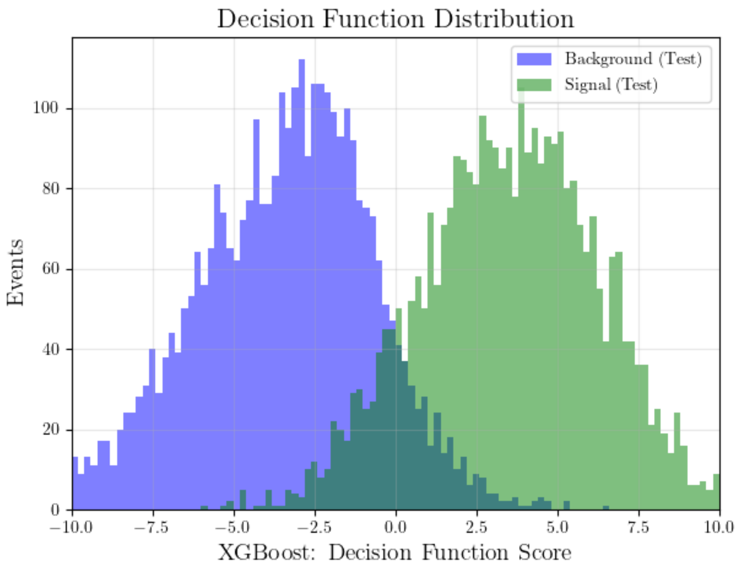
Decision score and probability distributions

Model XGBoost

Feature set: Bologna physics driven

Dataset: validation



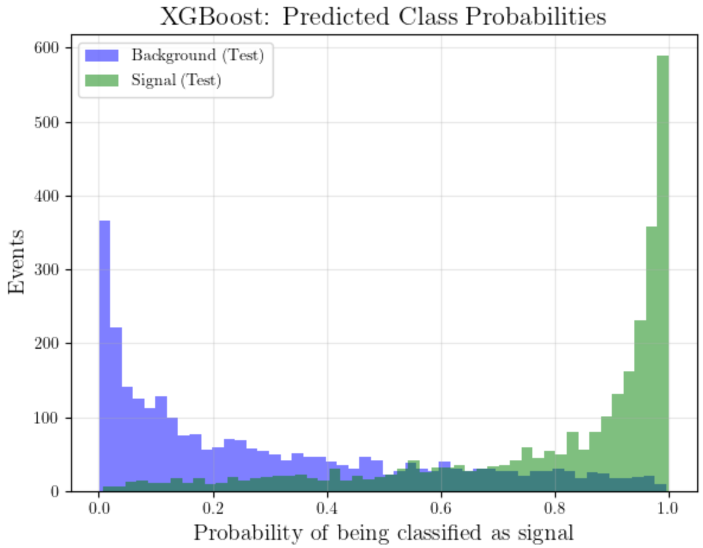


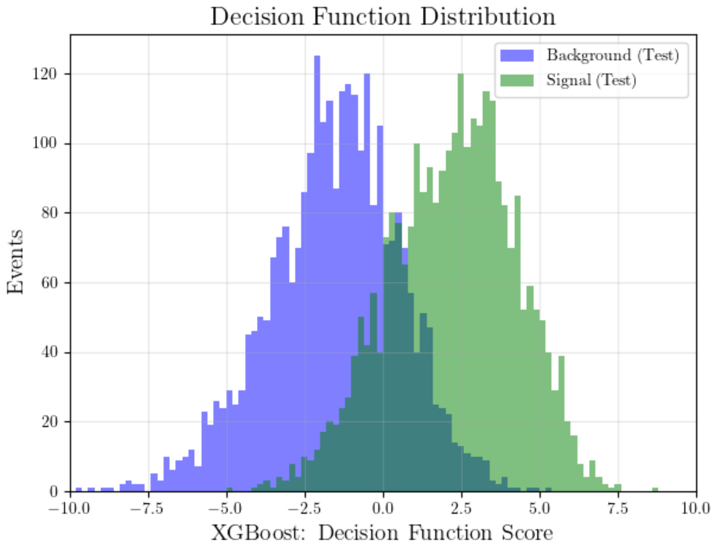
Decision score and probability distributions

Model XGBoost

Feature set: Bologna physics driven

Dataset: validation 2023



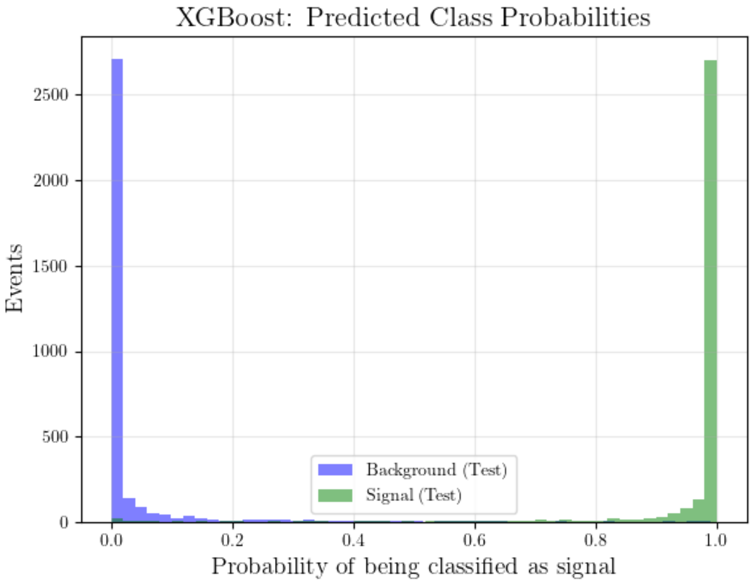


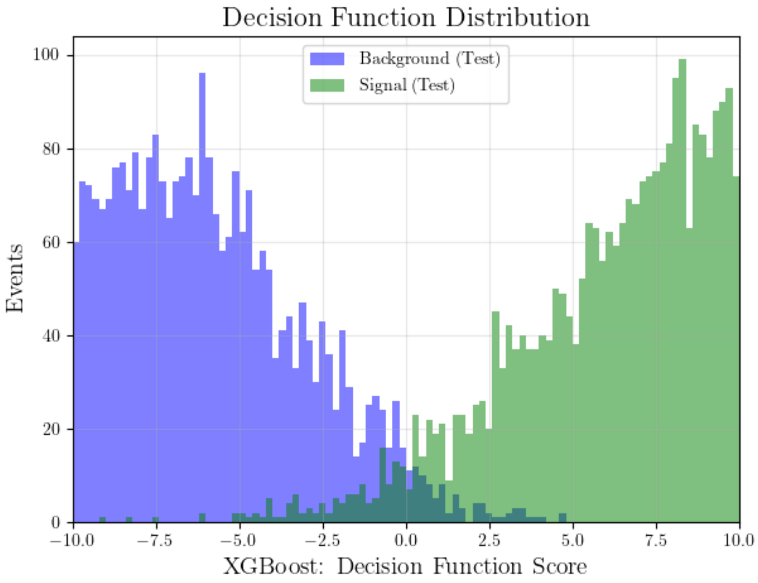
Decision score and probability distributions

Model XGBoost

Feature set: Mixed physics driven

Dataset: validation



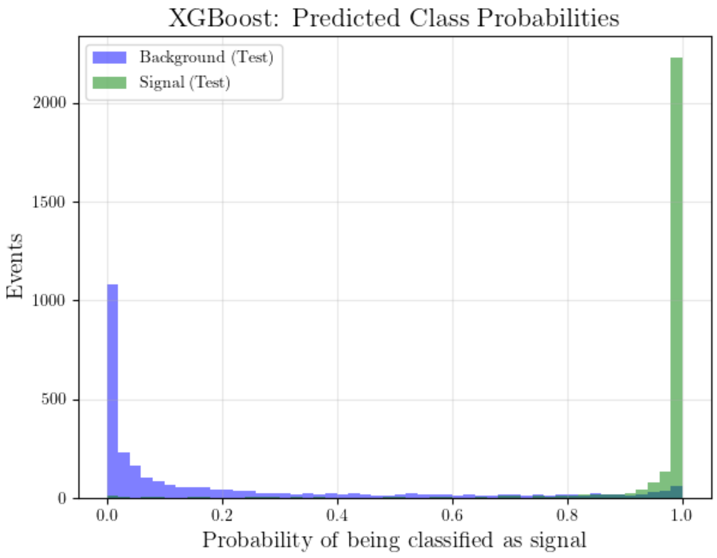


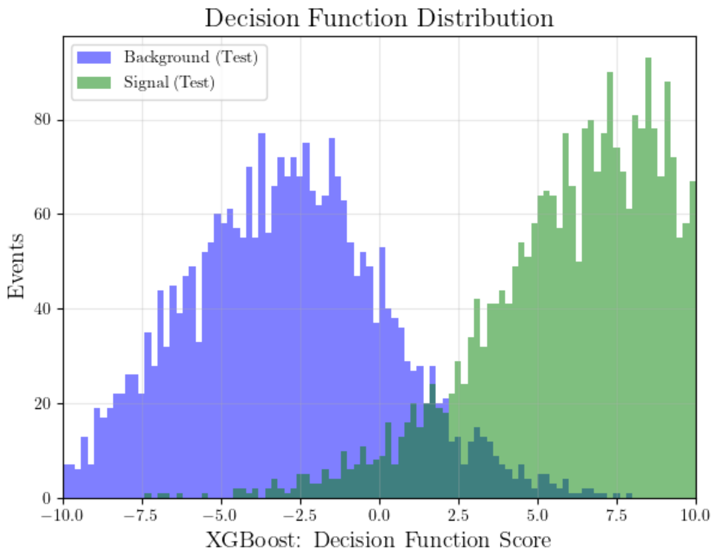
Decision score and probability distributions

Model XGBoost

Feature set: Mixed physics driven

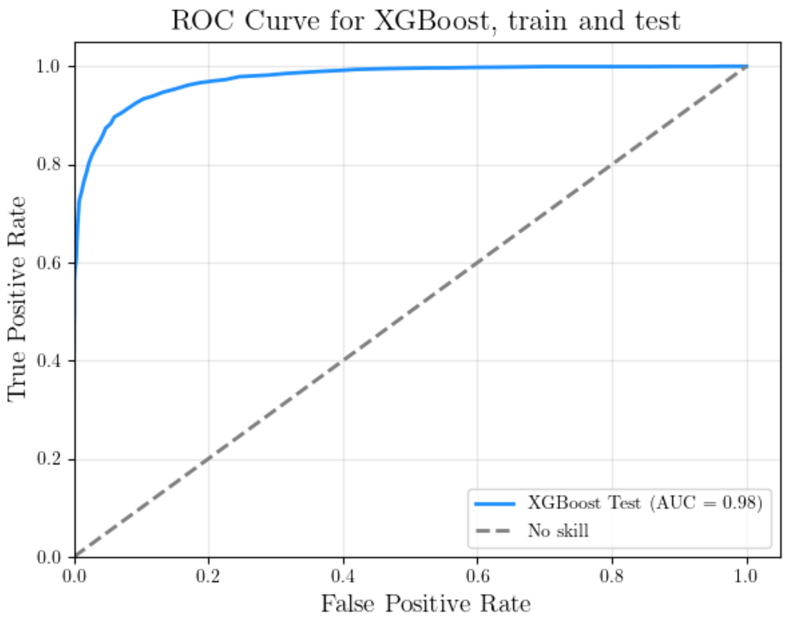
Dataset: validation 2023



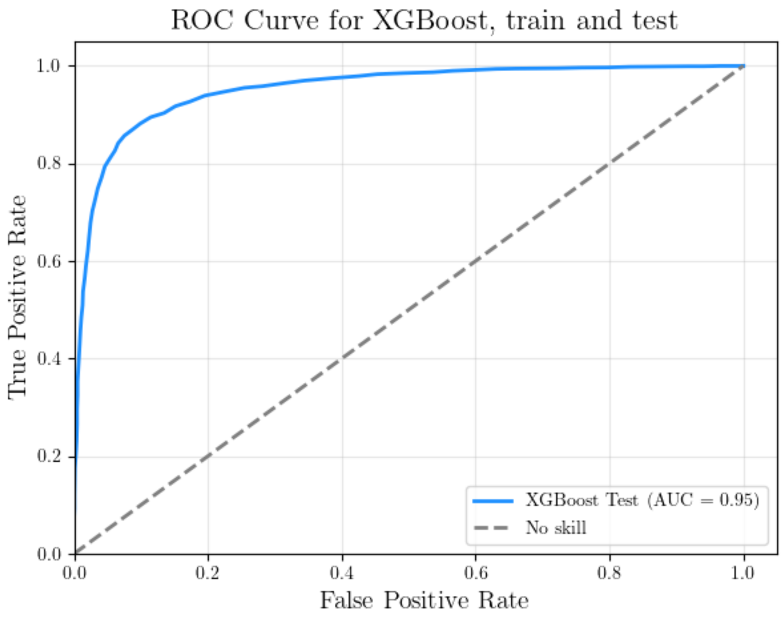


ROC curve – XGBoost, Groningen set

Validation 2015-2018 dataset

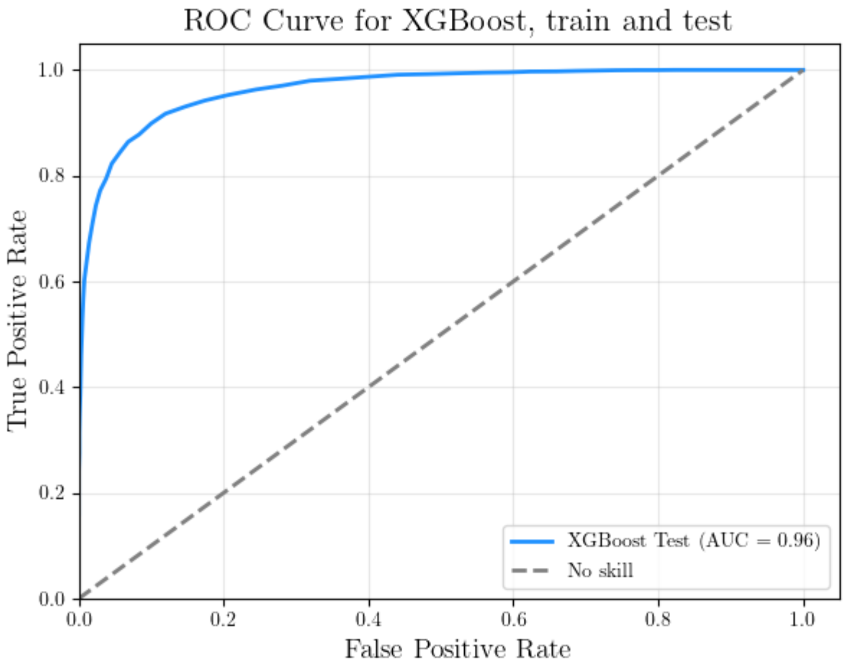


Validation 2023 dataset

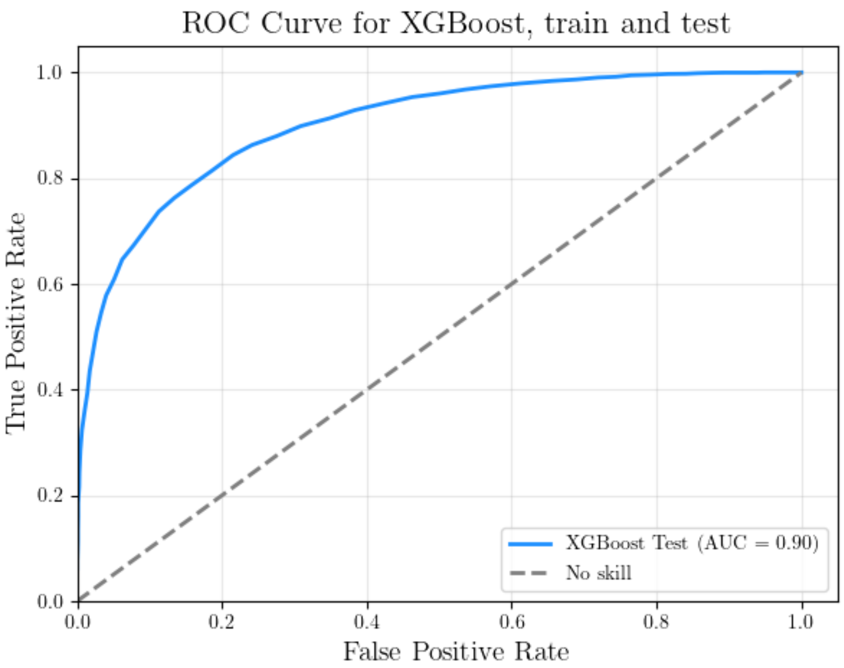


ROC curve – XGBoost, Bologna set

Validation 2015-2018 dataset

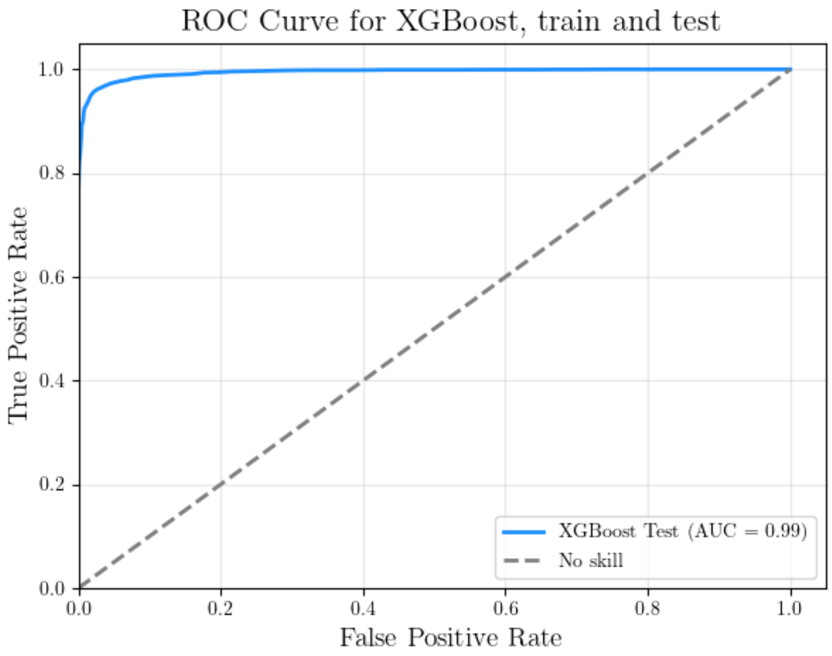


Validation 2023 dataset

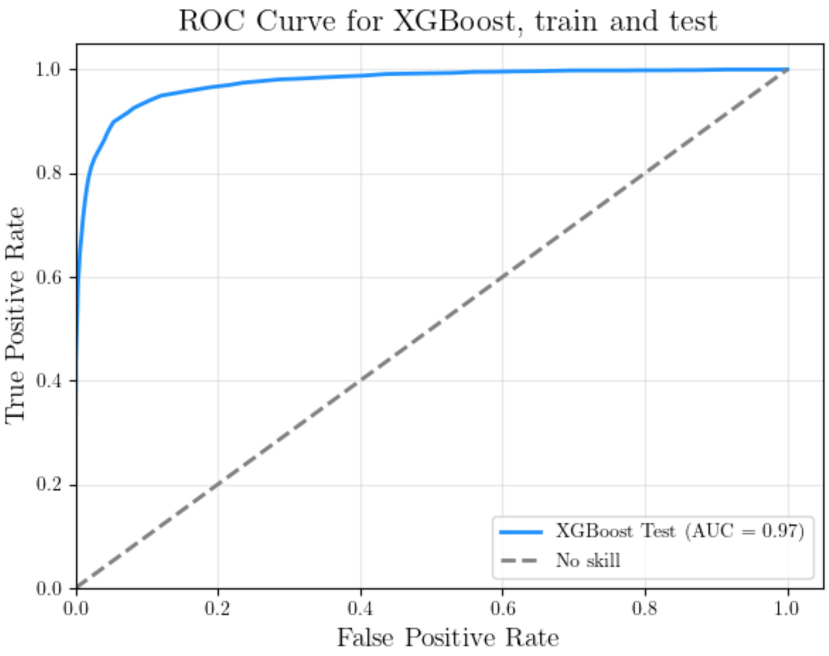


ROC curve – XGBoost, Mixed set

Validation 2015-2018 dataset



Validation 2023 dataset

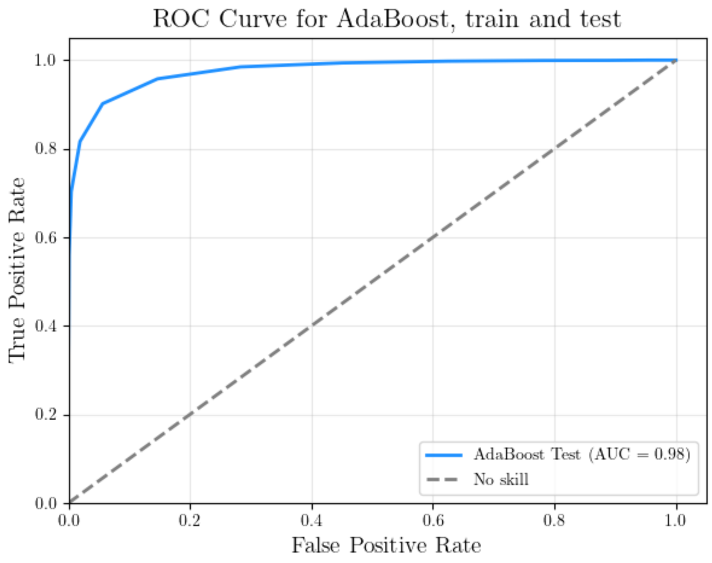


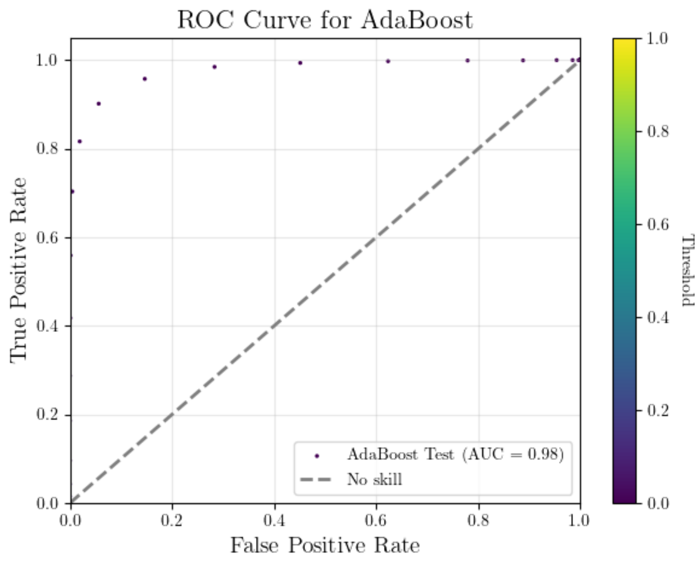
ROC curve

Model AdaBoost

Feature set: Mixed physics driven

Dataset: Validation





ROC curve

Model AdaBoost

Feature set: Mixed physics driven

Dataset: Validation 2023