National Research
Tomsk
State
University

Лаборатория
анализа данных
физики высоких энергий
Томского
государственного
университета

Measurement of differential cross-sections of a single top quark produced in association with a W boson with ATLAS at $\sqrt{s} = 13$ TeV

Progress Report

Neda Firoz

Goal: separate **tW (top+anti-top)** from **$t\bar{t}$** in the **1j1b dilepton** region

•**Inputs (9 variables used):**
bdt_centrality_ll_recalc_NOSYS,
bdt_delta_pT_ll_MET_recalc_NOSYS,
bdt_delta_pT_llb_MET_recalc_NOSYS,
bdt_eta_llMetB_recalc_NOSYS,
bdt_m_l1b_recalc_NOSYS,
bdt_m_l2b_recalc_NOSYS,
bdt_pT_llMetB_recalc_NOSYS,
bdt_pT_llb_recalc_NOSYS,
bdt_sum_ET_recalc_NOSYS.

•**Samples / tree:** all files' tree name is analysis

Signal: tW (top) + $\bar{t}$W (anti-top)
Background: $t\bar{t}$ (non-all-had)

•**Event weights:** auto-resolved to weight_mc_NOSYS * weight_pileup_NOSYS * globalTriggerEffSF_NOSYS.

•**Bad values:** any variable ≤ **−990** or non-finite is masked per event.

Article’s Report on BDT

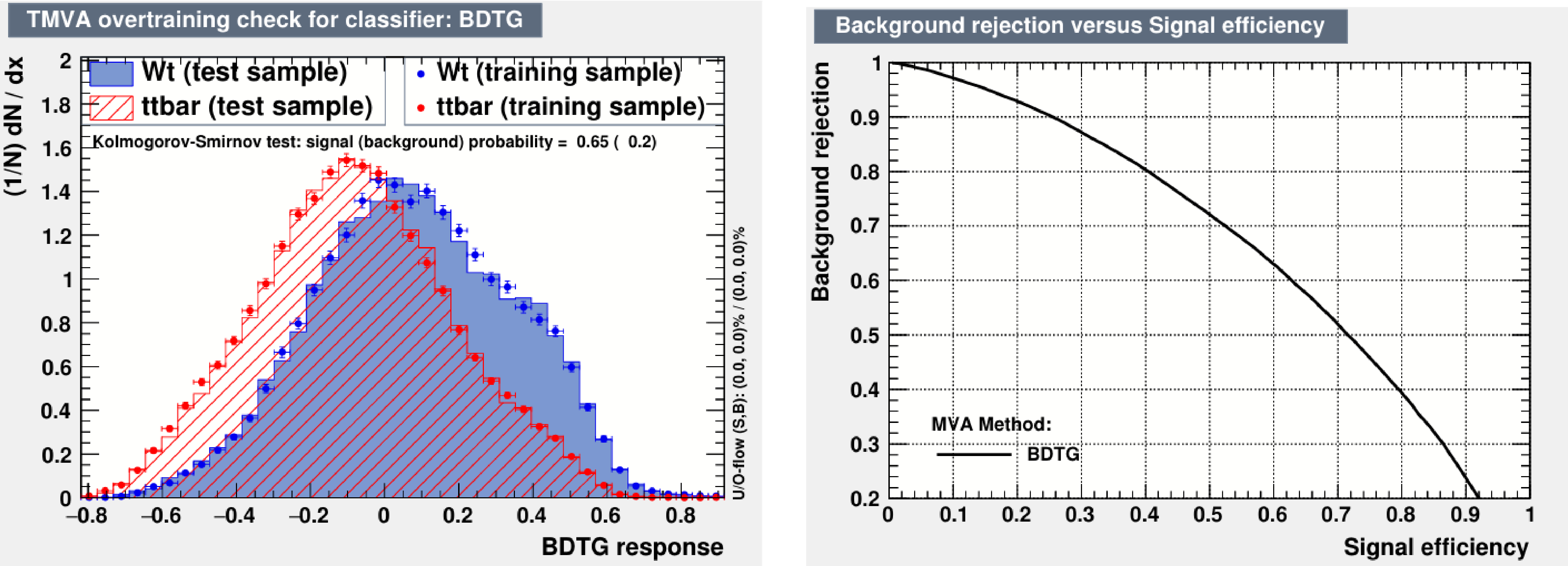
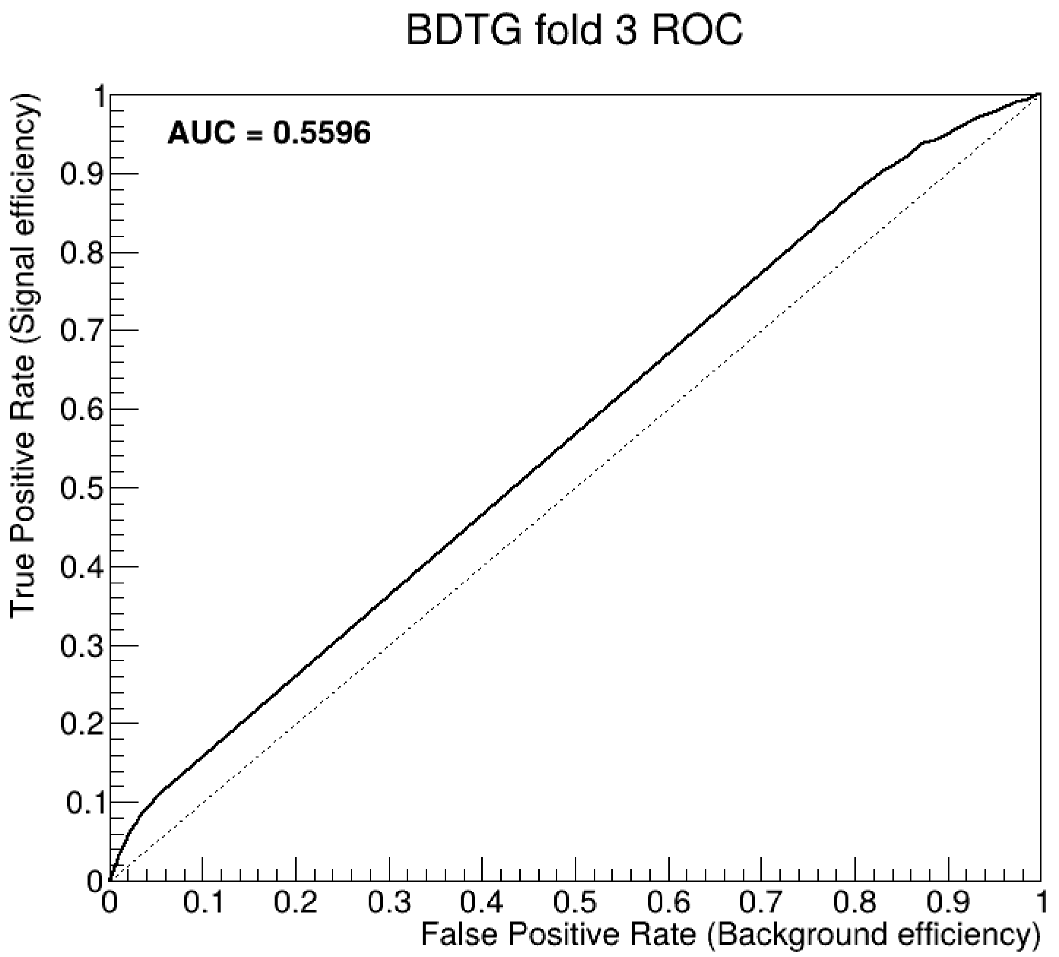


Figure 9: Comparison of test/training sample distributions and background rejection factor versus signal efficiency.

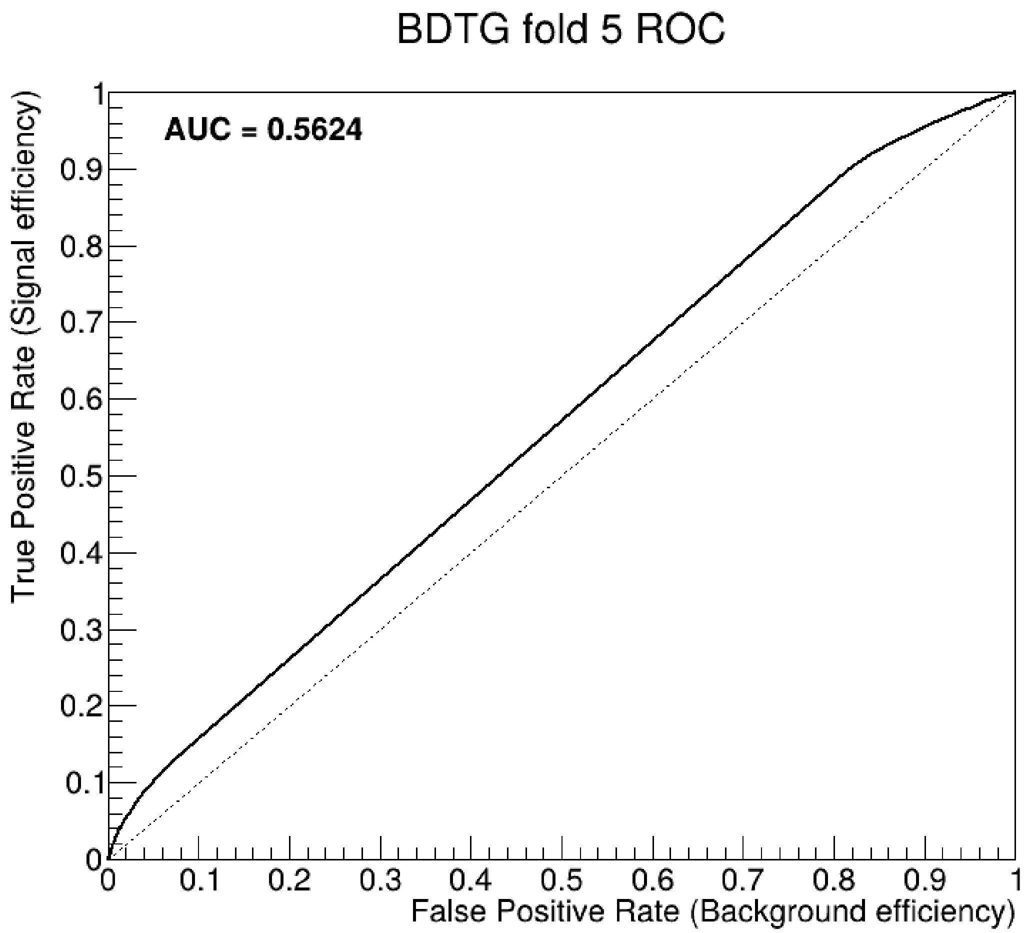
Results of Performance of ML Algorithms using K-fold on Python Trial 2
27th January 2026

Algorithms	ROC_AUC_MEAN	ROC_AUC_STD	n_folds_used
BDTB	0.560854	0.003277	5
BDTG	0.55978	0.003049	5
Fisher	0.550102	0.00274	5
KNN	0.549632	0.003916	5
LD	0.547795	0.001821	5
Likelihood	0.523526	0.004158	5
MLP	0.561797	0.002013	5

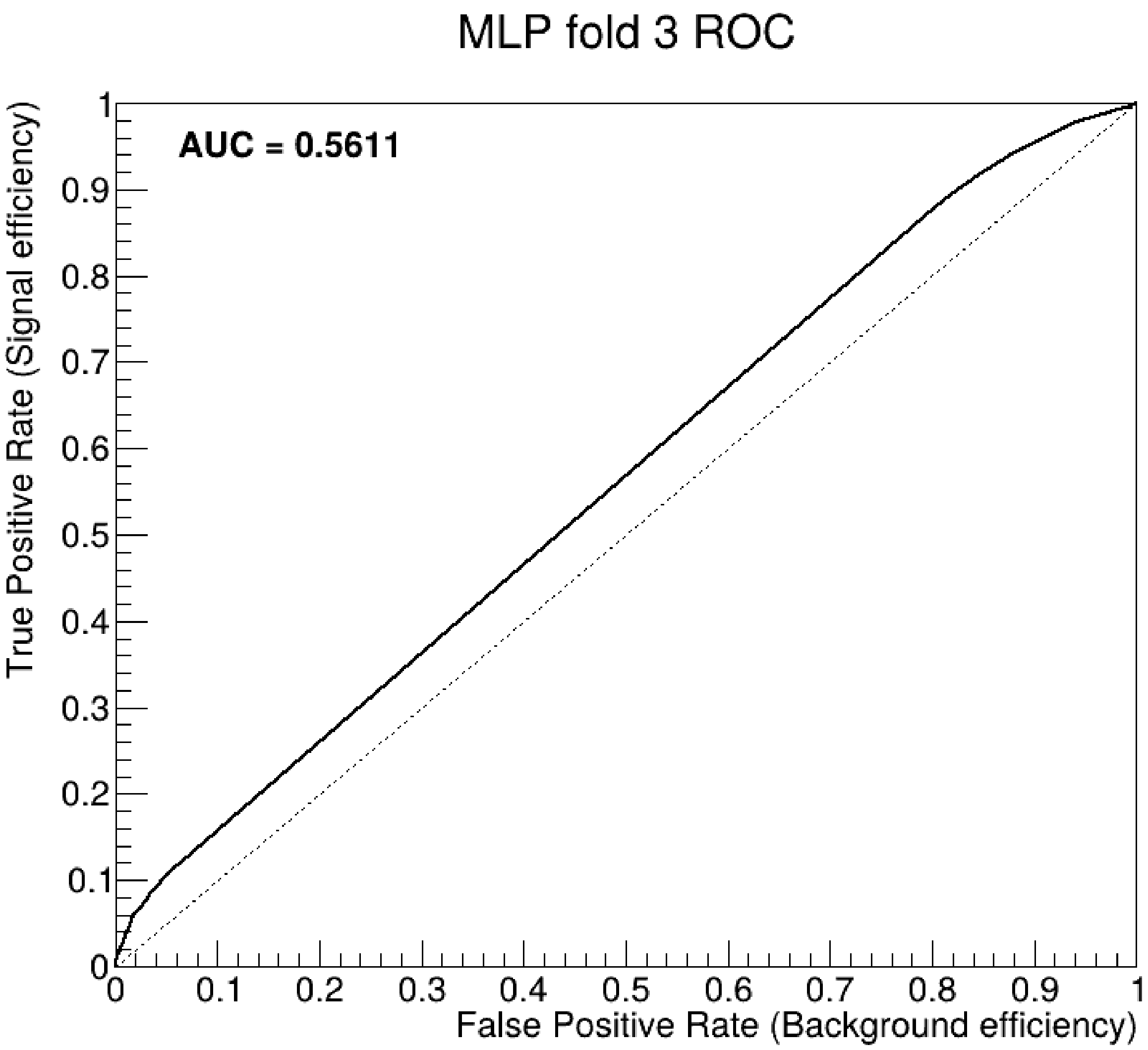
I was doing an ordinary train/test split before cross-validation (CV).
So, CV was effectively training on (roughly) half the data, then splitting that again into folds. That alone could have pushed AUC down noticeably and make results unstable.
So, I faced three times unstable results on tmva. I am still trying to prepare correct code with kfold on TMVA.



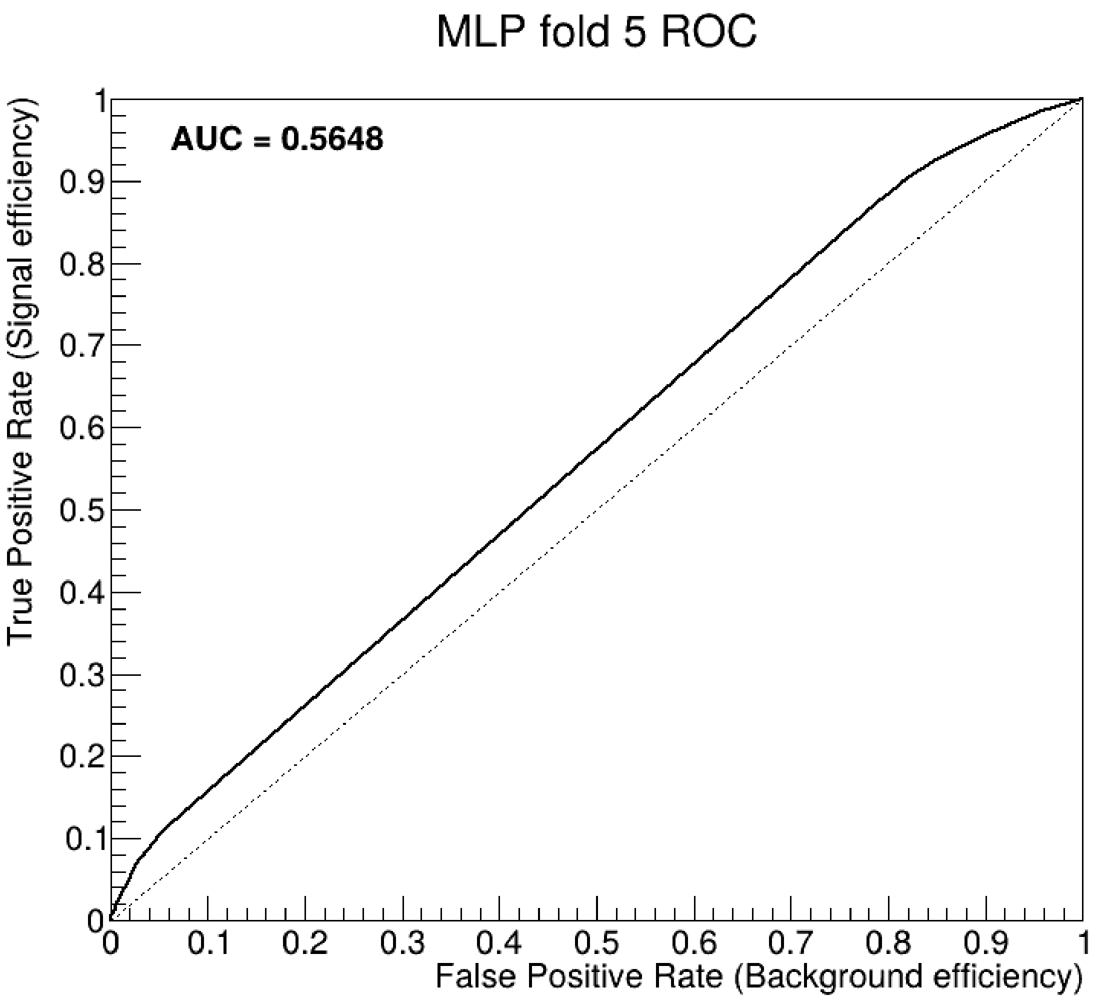
BDTG fold 3 ROC



BDTG fold 5 ROC



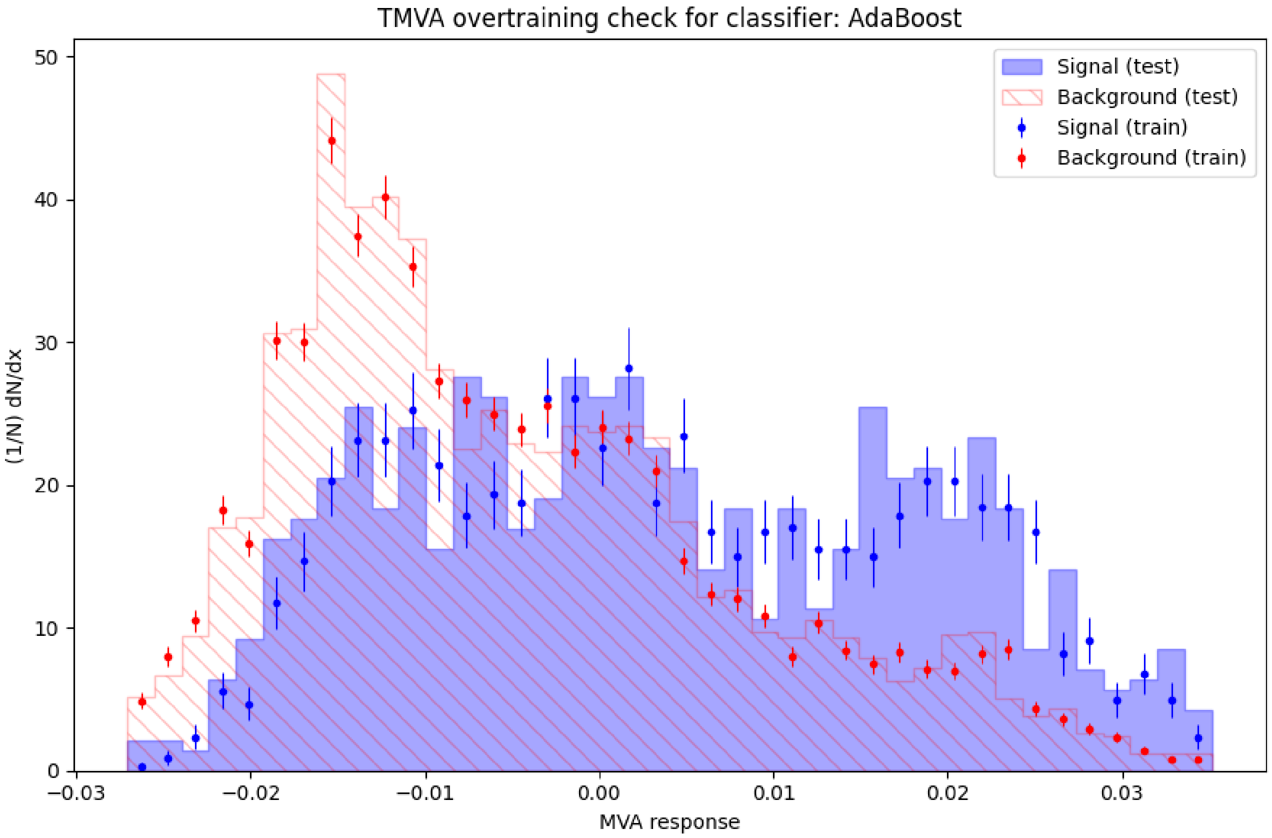
MLP fold 3 ROC
AUC = 0.5611
True Positive Rate (Signal efficiency)
False Positive Rate (Background efficiency)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

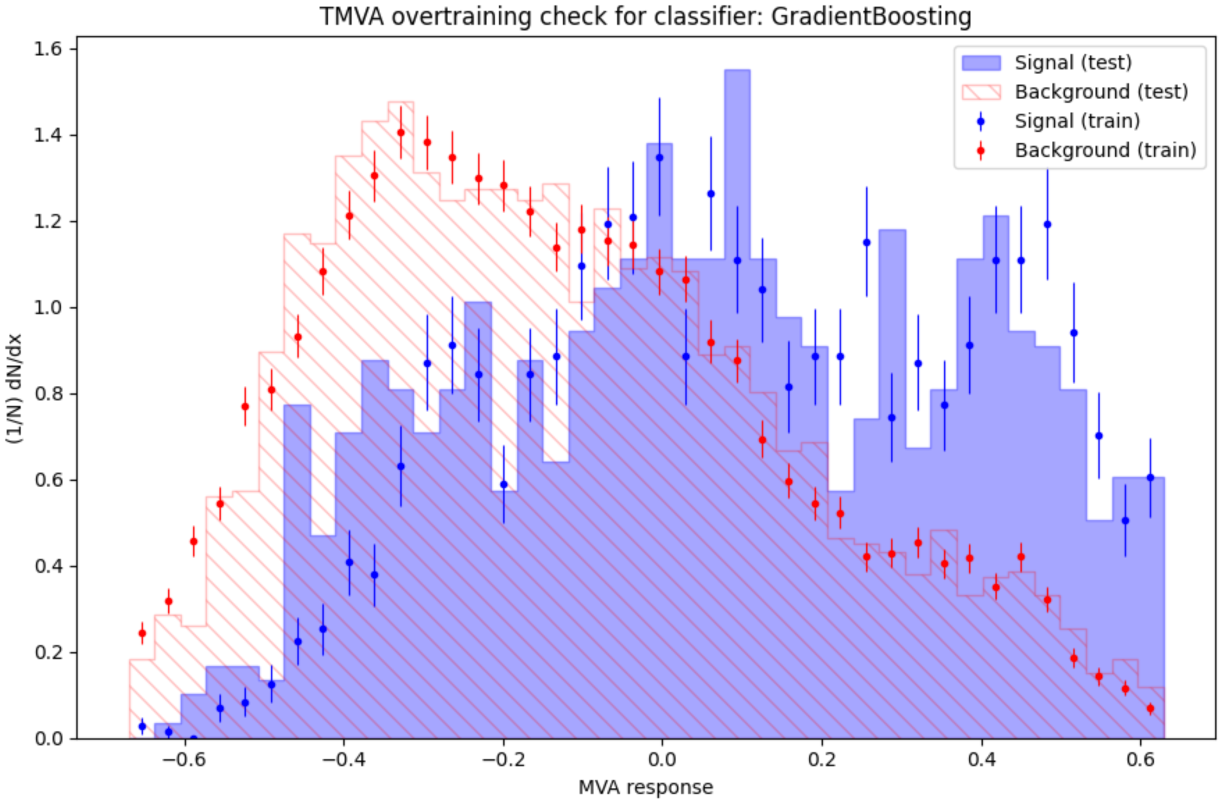


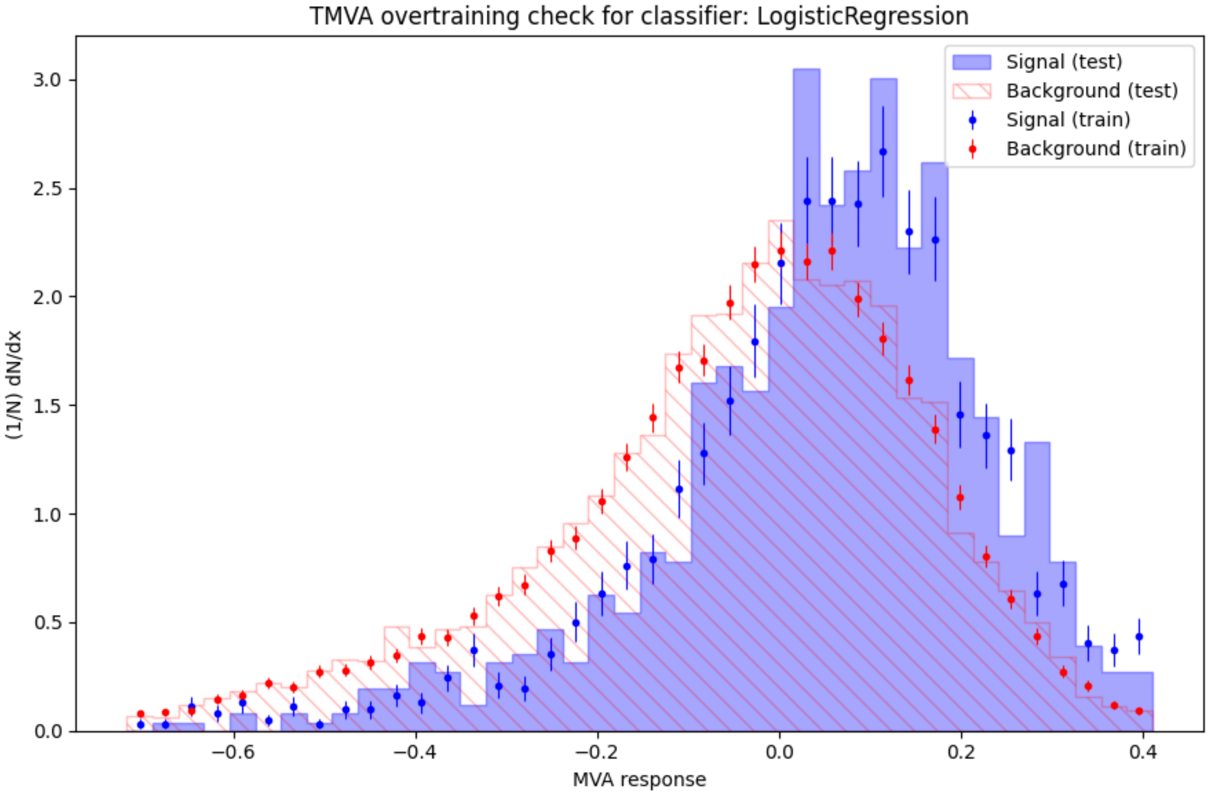
MLP fold 5 ROC
AUC = 0.5648
True Positive Rate (Signal efficiency)
False Positive Rate (Background efficiency)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

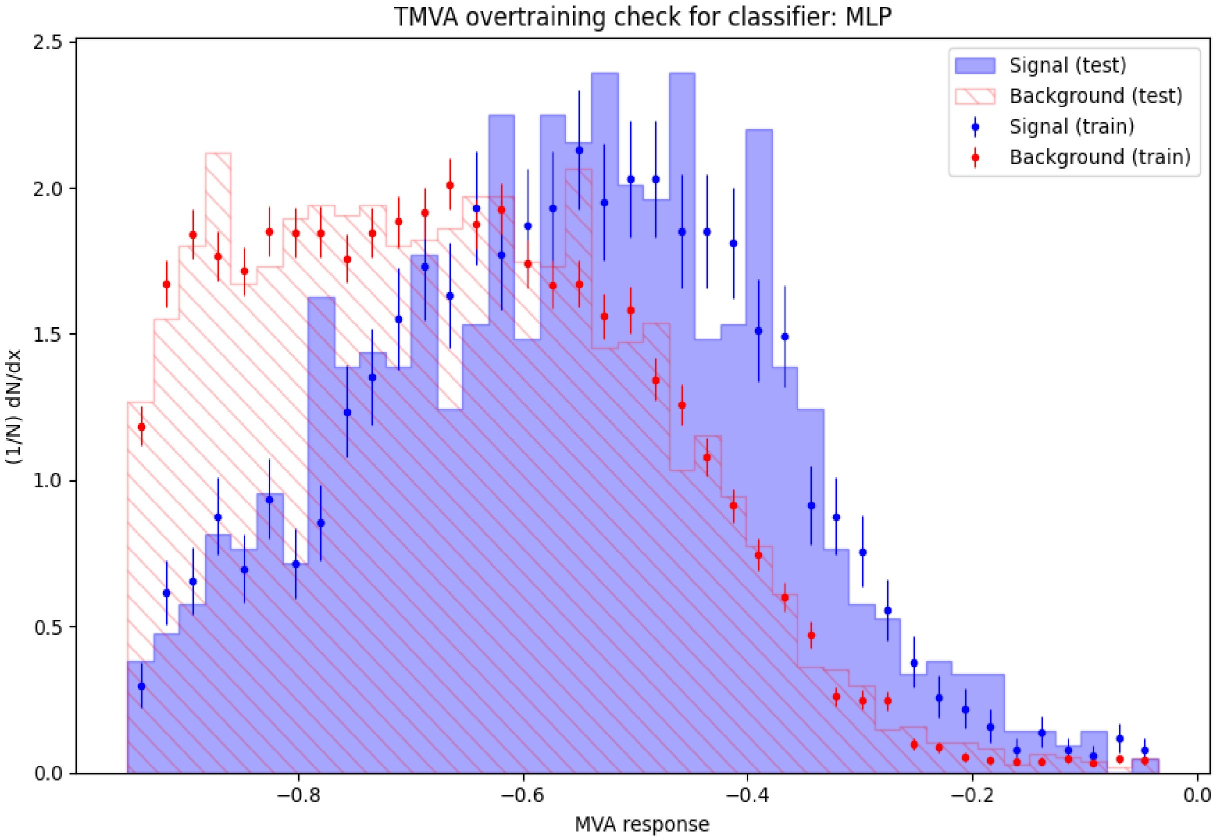
Results of Performance of ML Algorithms using K-fold on Python Trial 3
27th January 2026

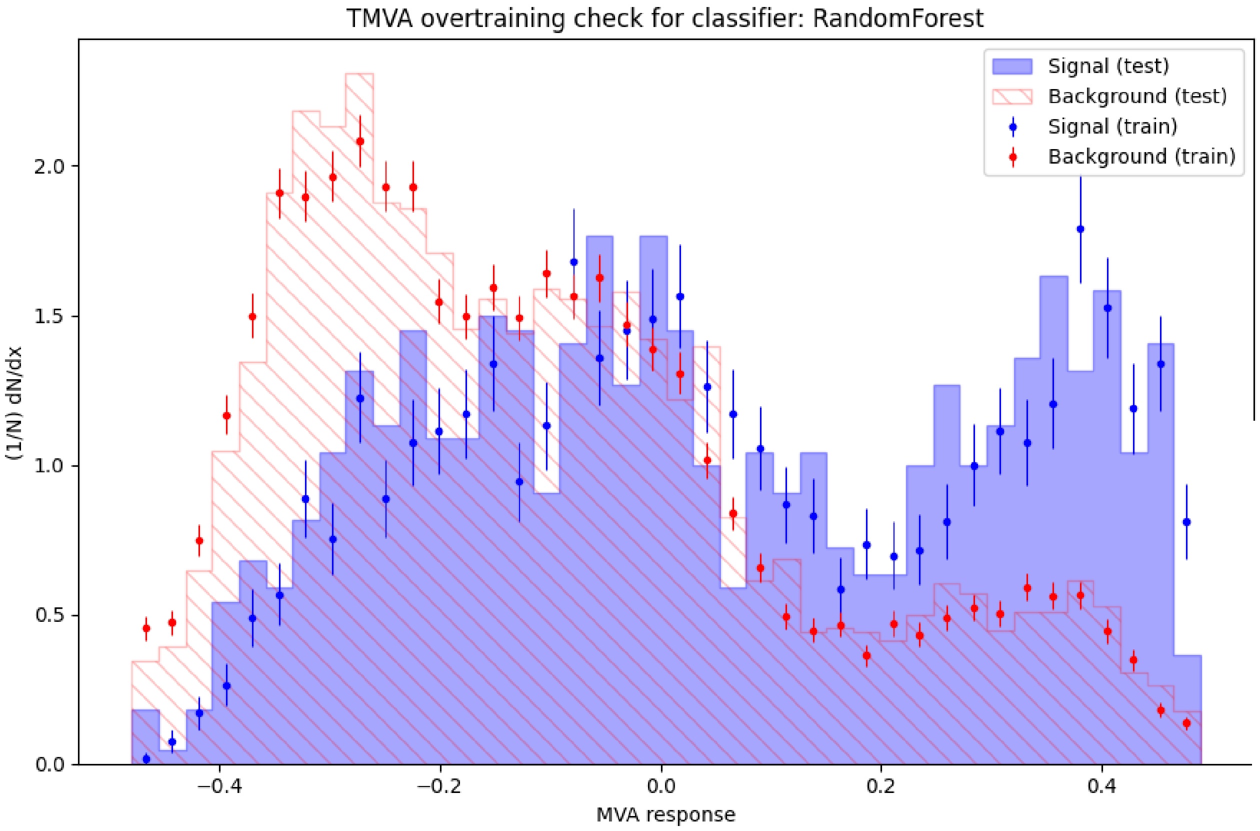
Algorithms	outer_k	inner_k	n_iters	mean_fold_auc	std_fold_auc	overall_oof_auc
AdaBoost	5	3	25	0.668024	0.008952	0.662409
Gaussian-NB	5	3	25	0.6407	0.006545	0.639867
Gradient Boosting	5	3	25	0.675134	0.01051	0.67482
KNN	5	3	25	0.643668	0.007173	0.643566
Logistic Regression	5	3	25	0.639506	0.011228	0.639507
MLP	5	3	25	0.665761	0.010691	0.664427
Random Forest	5	3	25	0.672187	0.010309	0.671608
SVC	5	3	10	0.676572	0.01301	0.676277

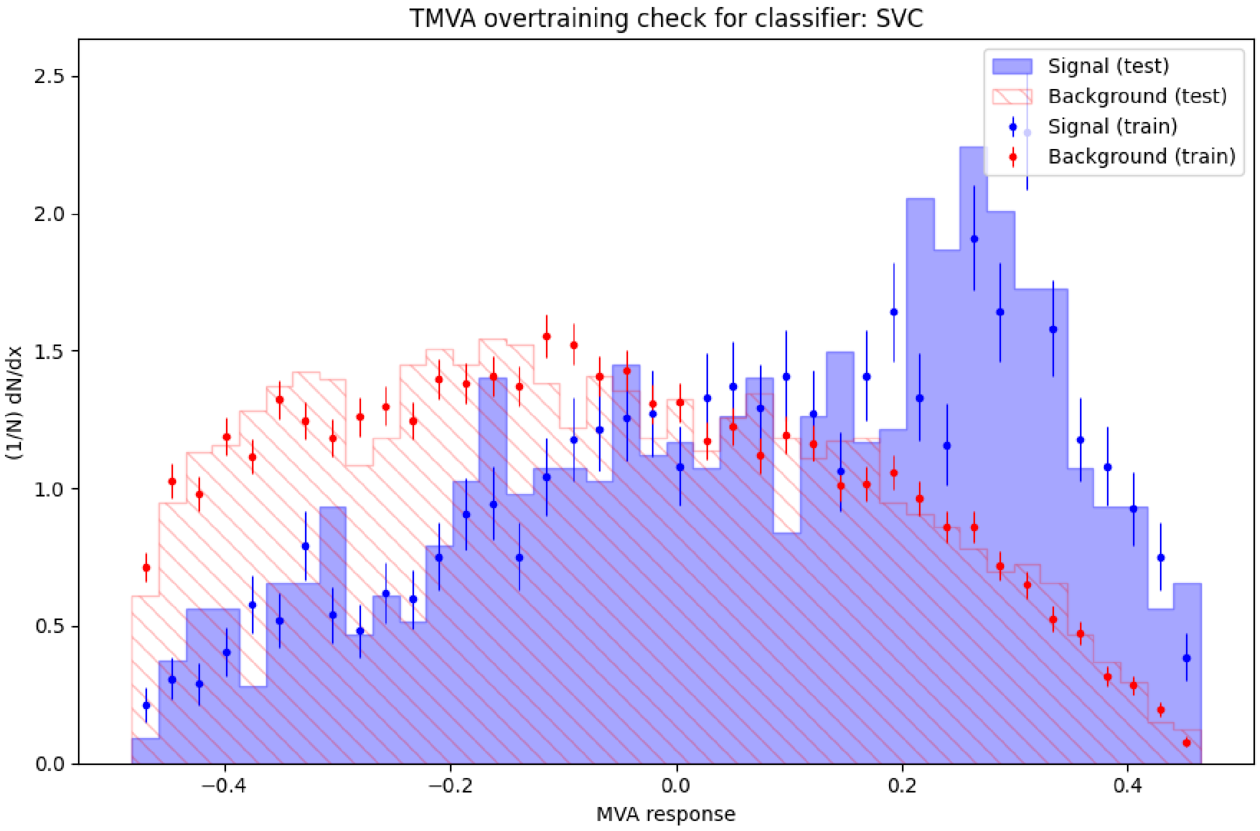

TMVA overtraining check for classifier: AdaBoost
Signal (test)
Background (test)
Signal (train)
Background (train)
(1/N) dN/dx
0
10
20
30
40
50
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03
MVA response

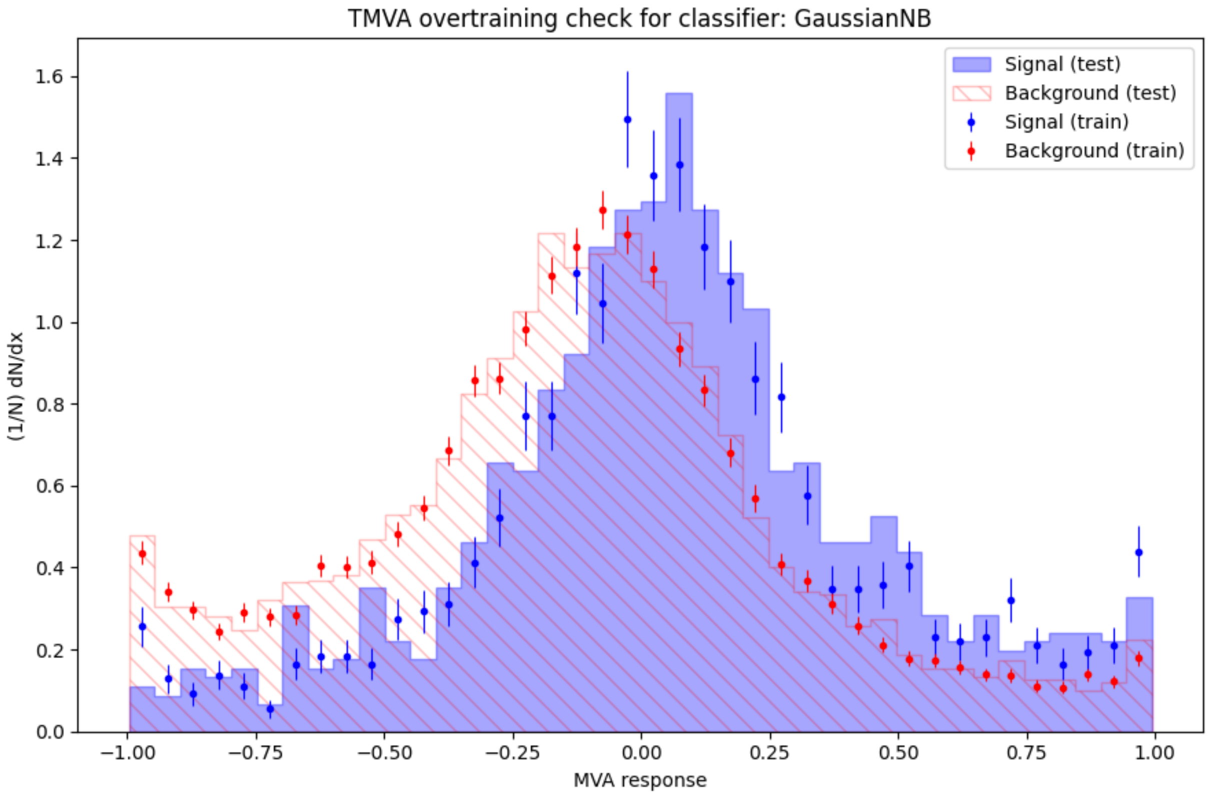

TMVA overtraining check for classifier: GradientBoosting
Signal (test)
Background (test)
Signal (train)
Background (train)
(1/N) dN/dx
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
MVA response

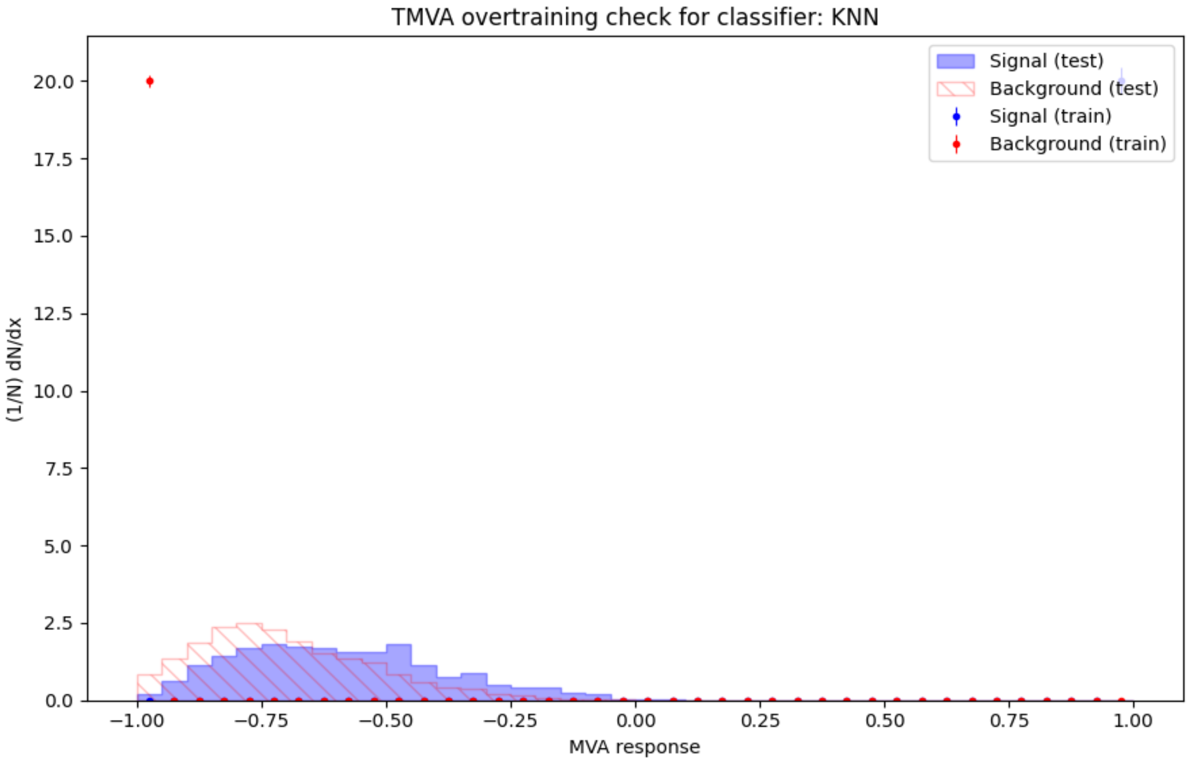

TMVA overtraining check for classifier: LogisticRegression
Signal (test)
Background (test)
Signal (train)
Background (train)
(1/N) dN/dx
0.0
0.5
1.0
1.5
2.0
2.5
3.0
−0.6
−0.4
−0.2
0.0
0.2
0.4
MVA response

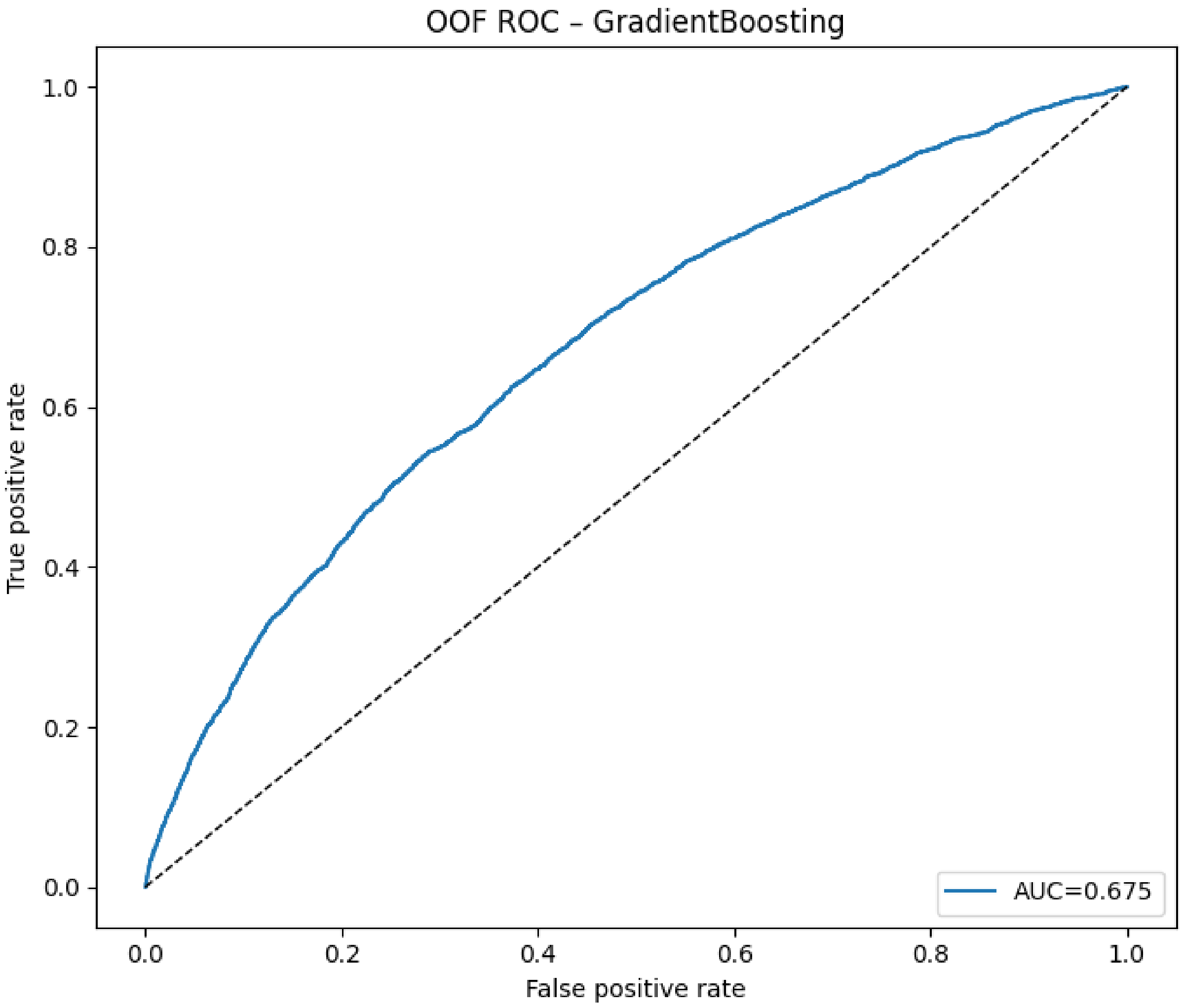

TMVA overtraining check for classifier: MLP
Signal (test)
Background (test)
Signal (train)
Background (train)
(1/N) dN/dx
0.0
0.5
1.0
1.5
2.0
2.5
−0.8
−0.6
−0.4
−0.2
0.0
MVA response

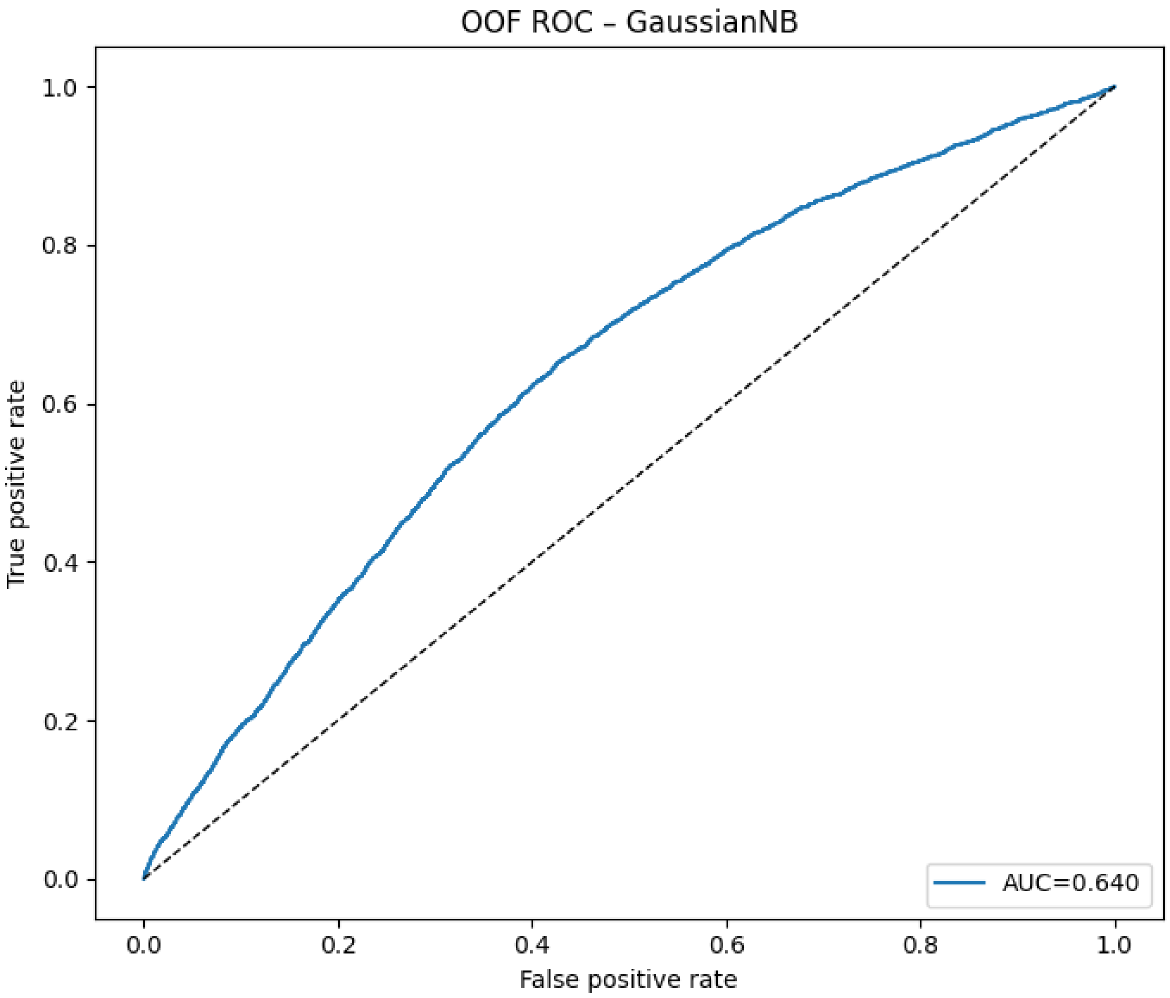

TMVA overtraining check for classifier: RandomForest
Signal (test)
Background (test)
Signal (train)
Background (train)
(1/N) dN/dx
0.0
0.5
1.0
1.5
2.0
−0.4
−0.2
0.0
0.2
0.4
MVA response

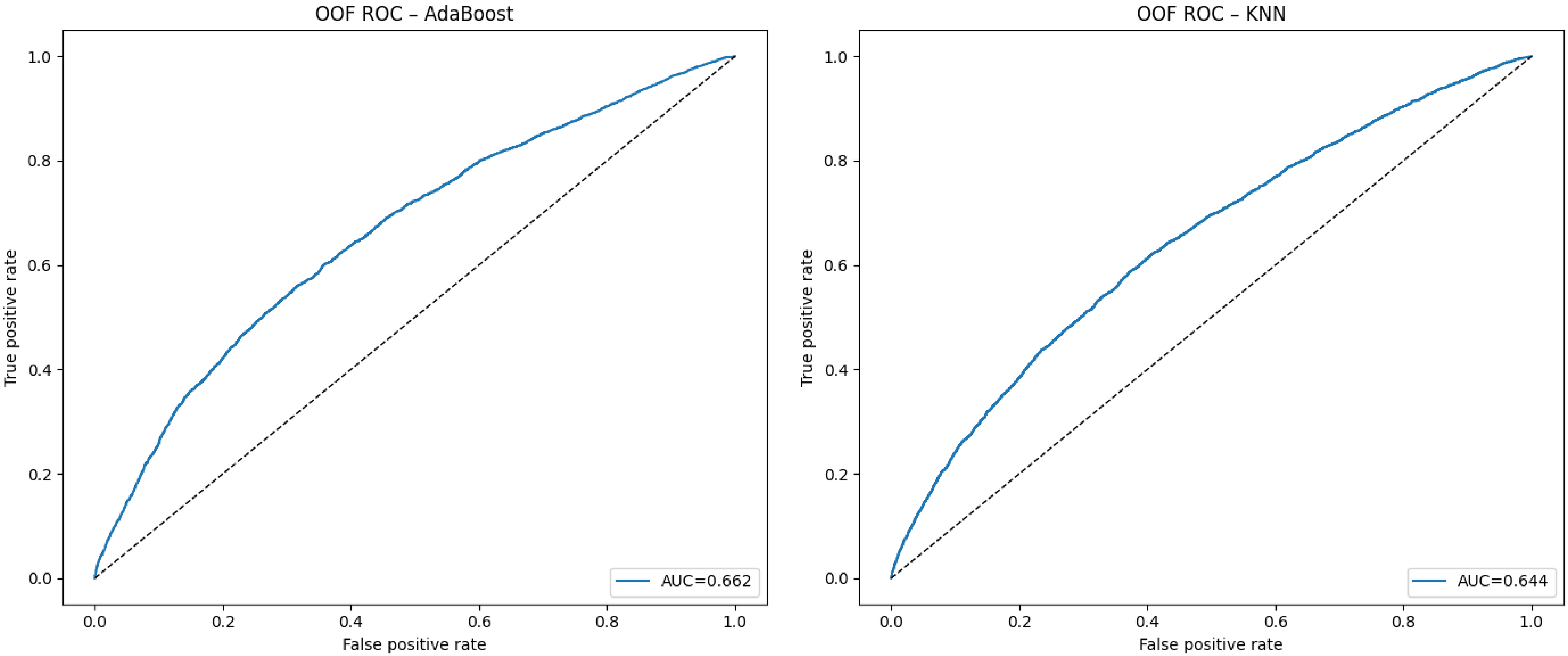

TMVA overtraining check for classifier: SVC
Signal (test)
Background (test)
Signal (train)
Background (train)
(1/N) dN/dx
0.0
0.5
1.0
1.5
2.0
2.5
−0.4
−0.2
0.0
0.2
0.4
MVA response

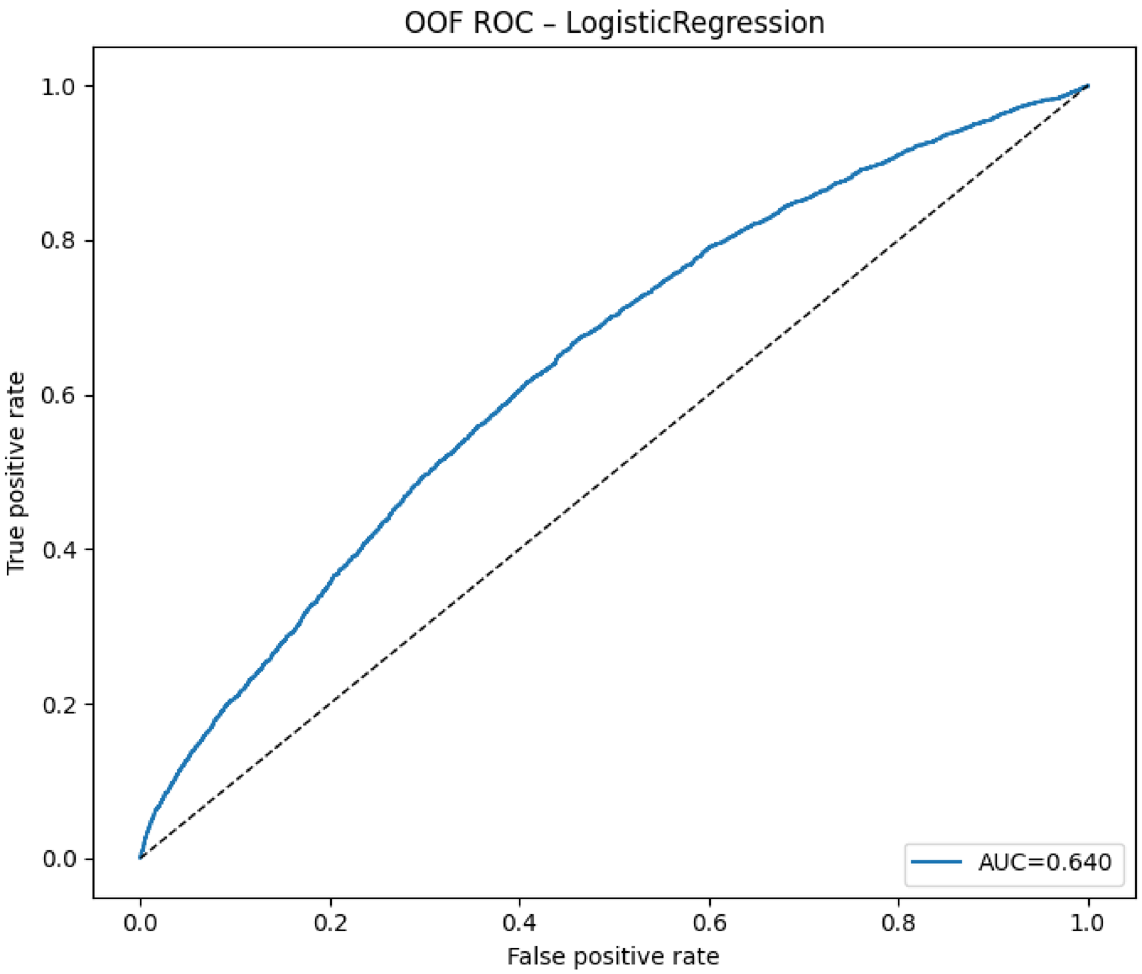

TMVA overtraining check for classifier: GaussianNB
Signal (test)
Background (test)
Signal (train)
Background (train)
(1/N) dN/dx
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
MVA response

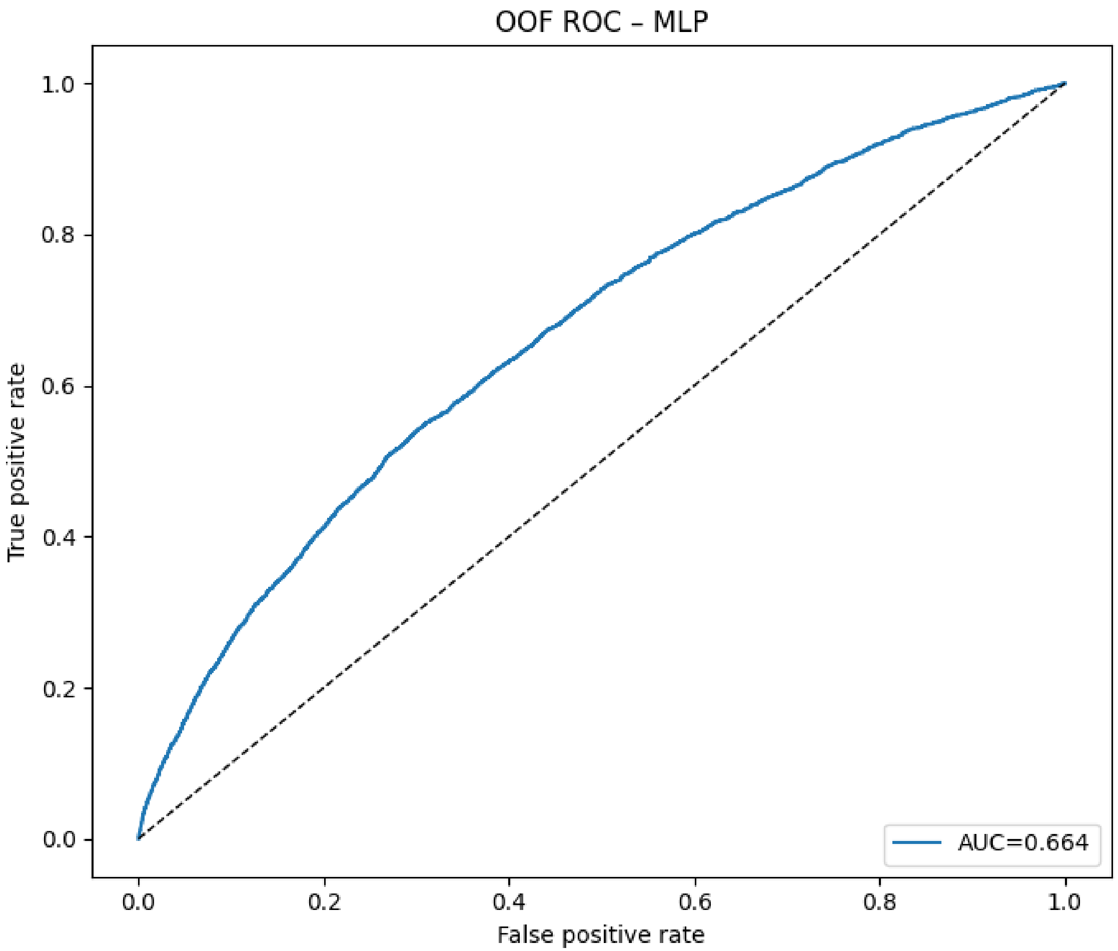

TMVA overtraining check for classifier: KNN
Signal (test)
Background (test)
Signal (train)
Background (train)
(1/N) dN/dx
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
MVA response

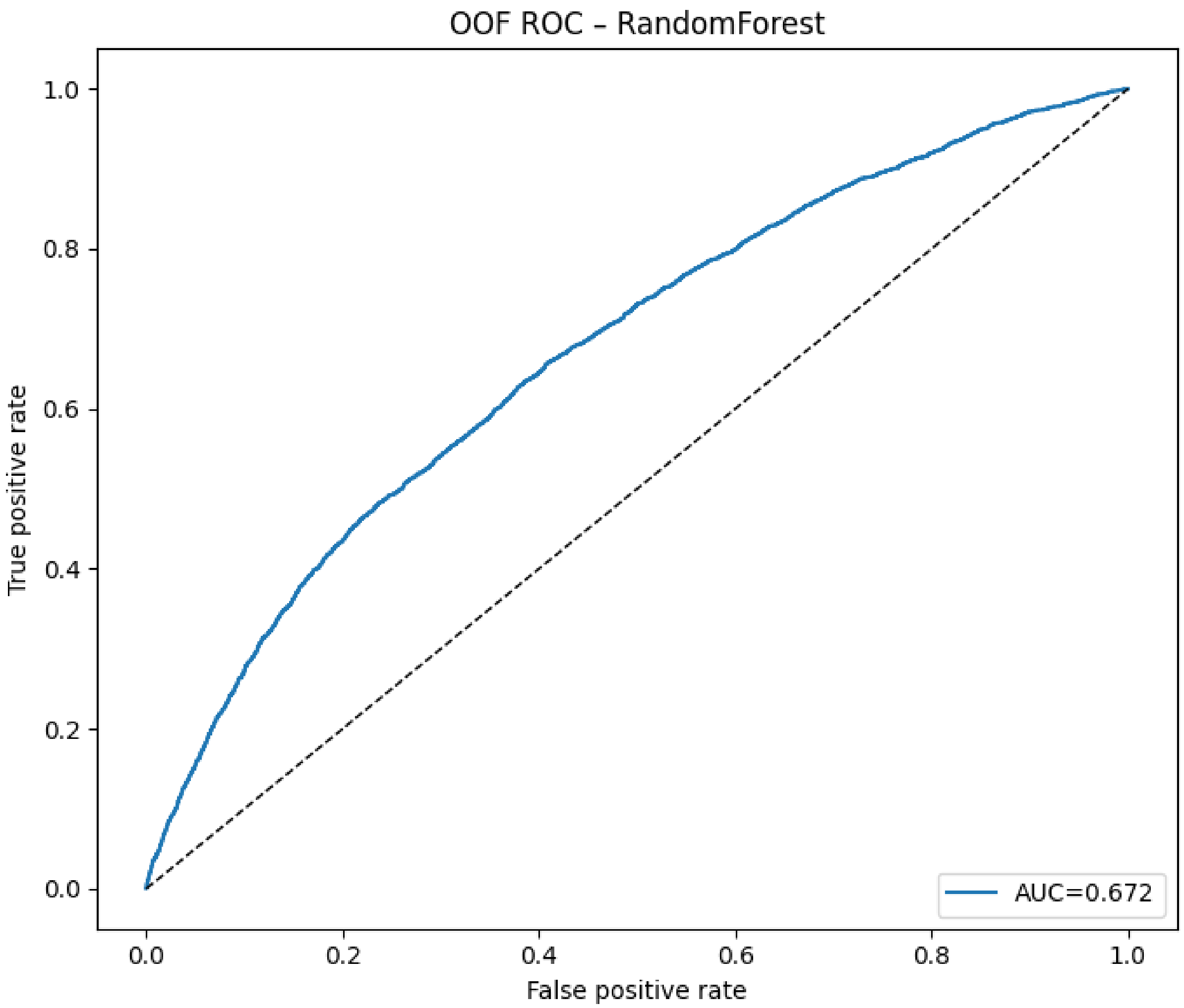

OOF ROC – GradientBoosting
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC=0.675

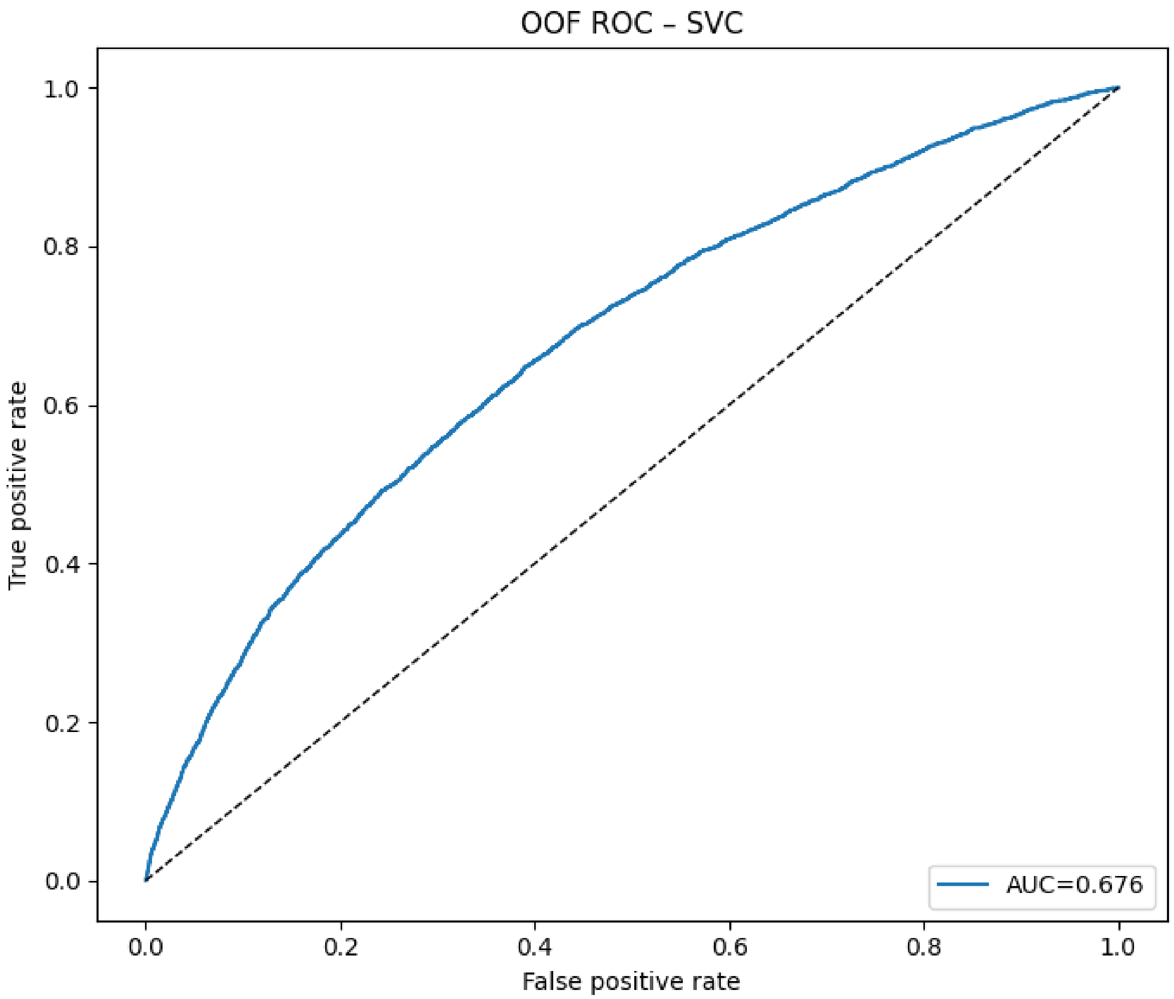

OOF ROC – GaussianNB
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC=0.640


OOF ROC – AdaBoost
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC=0.662
OOF ROC – KNN
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC=0.644


OOF ROC – LogisticRegression


OOF ROC – MLP


OOF ROC – RandomForest
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
AUC=0.672


OOF ROC – SVC
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
AUC=0.676

National Research
Tomsk
State
University

Лаборатория
анализа данных
физики высоких энергий

Томского
государственного
университета

Thank you for your attention!!!