Study of the VH($H \to b\bar{b}$) production by MVA methods

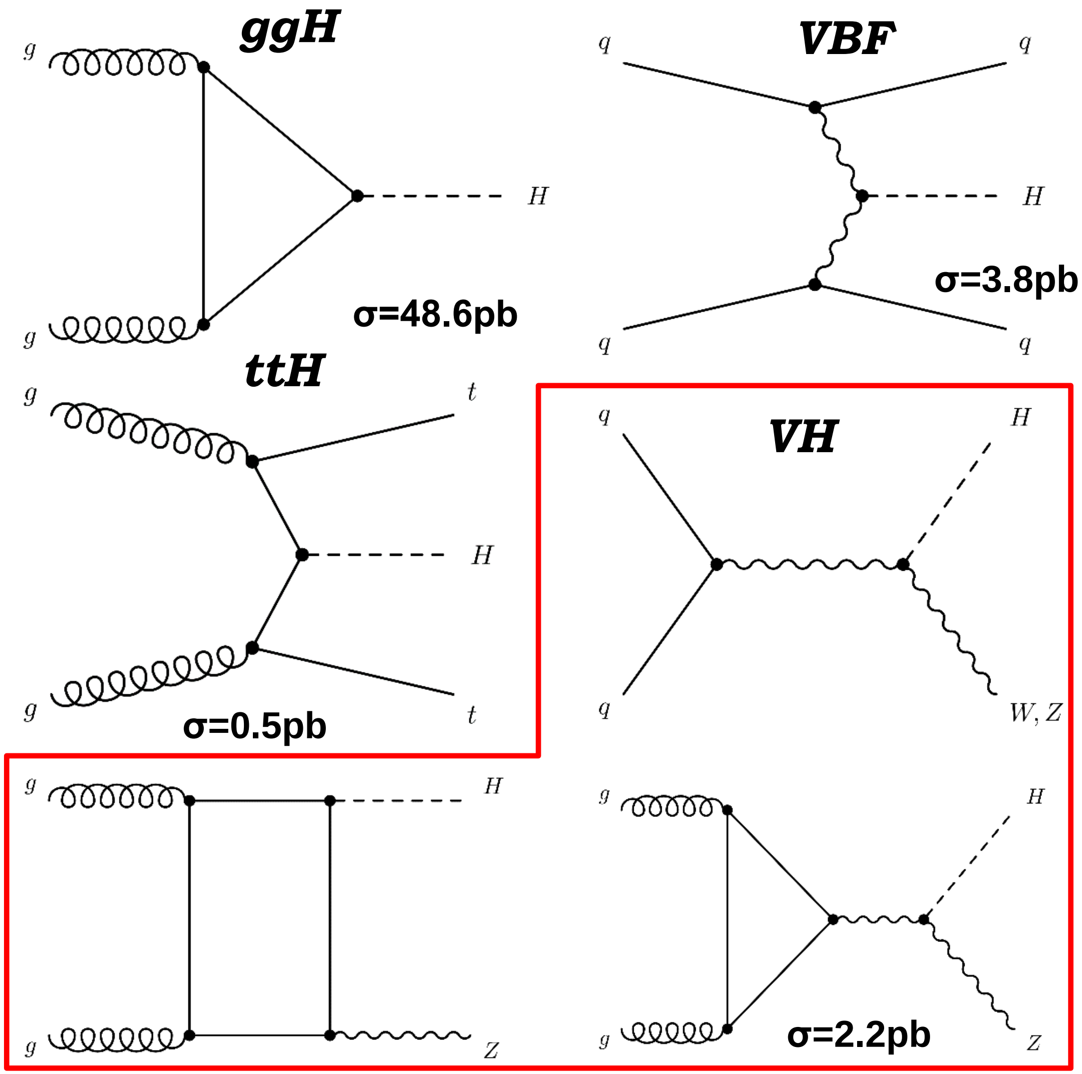
Faig Ahmadov
JINR & IP, ANAS

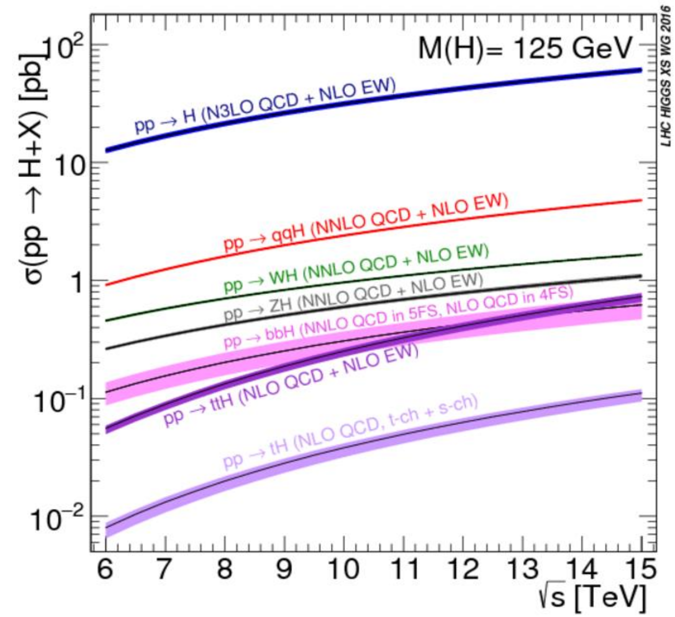
6th International Workshop DLCP-2022
06-08.07.2022

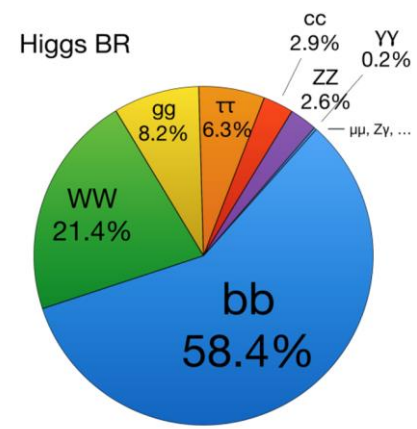
Introduction

- Theoretically, the existence of the Higgs boson was introduced in 1964 with the Higgs mechanism.
- The Higgs boson observed on 4 July 2012 (ATLAS & CMS) *in* $H \to \gamma\gamma$, $H \to ZZ$, *decay channels*;
- *Late observation of the* $H \to W^+W^-$, $H \to \tau^+\tau^-$ *decay and ggH, VBF production channels.*
- $H \to b\bar{b}$ decay channel plays an important role in studying the properties of the Higgs boson, since this is the dominant decay mode (**BR ≈ 58%**).
- 2018 - observation of $\boldsymbol{H \to b\bar{b}}$ decay and ***VH*** production channels.
- After observing the main production and decay channels, the study of the Higgs boson moves from discovery to the era of precise measurements.

Production & decay of the Higgs boson

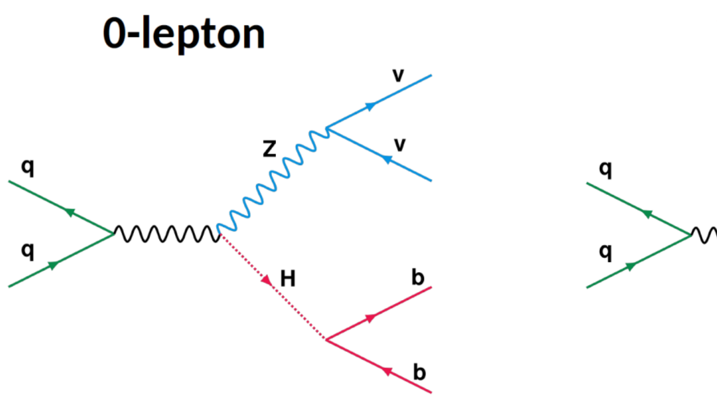






Signal & background

3 lepton channels: 0-lep. $\boldsymbol{ZH \to \nu\nu\, b\bar{b}}$, *1-lep.* $\boldsymbol{WH \to \ell\nu\, b\bar{b}}$, *2-lep.* $\boldsymbol{ZH \to \ell\ell\, b\bar{b}}$



Main backgrounds: ***Z+jets, W+jets, ttbar*** and ***single top***.

In 2-lep. channel *ttbar* estimated by data-driven method.

Backgrounds similar to signal: $\boldsymbol{ZZ \to \nu\nu\, b\bar{b}}$, $\boldsymbol{WZ \to \ell\nu\, b\bar{b}}$, $\boldsymbol{ZZ \to \ell\ell\, b\bar{b}}$

Multijet background: Estimated using data-driven method in 1-lep.

Negligible in 0 and 2-lep. channel.

Object reconstruction

- **Loose el.:** p_T > 7 GeV, $|\eta|$ < 2.47, low quality, track-based iso.
  Signal el.: loose el + high quality, calo-based iso. and p_T > 27 GeV.
- **Loose muons:** p_T > 7 GeV, $|\eta|$ < 2.7, low quality, loose track iso.
  Signal mu: loose mu + medium quality, tighter track iso.
  and p_T > 25 GeV.
- $\boldsymbol{E_T^{miss}}$**:** negative vector sum of the p_T of leptons, photons, had. τ, jets, and a '*soft-term*'.
- **Hadronically decaying τ :** p_T > 20 GeV, $|\eta|$ < 2.5 (exc. 1.37<$|\eta|$<1.52), 'medium' quality. Not used for ev. sel., only for E_T^{miss} calculation and to avoid double-counting τ as other objects.
- **Jets:** Topo-cluster of calo. cell using anti-k_t algorithm, R=0.4,
  p_T > 20 GeV, $|\eta|$<2.5 and p_T > 30 GeV if 2.5<$|\eta|$<4.5,
  Jet Vertex Tagger (JVT(to remove jets with p_T < 120 GeV, $|\eta|$<2.5 if not associated primary vertex.
- **B-jets** identified using a multivariate discriminant (MV2), with an average efficiency of 70%, which corresponds to light jet (u-, d-,s-quark and gluon) and c-jet misidentification efficiencies of 0.3% and 12.5% respectively.

Event selection

Selection	0-lepton	1-lepton: e sub-channel	1-lepton: μ sub-channel	2-lepton
Trigger	$E_{\mathrm{T}}^{\mathrm{miss}}$	Single lepton	$E_{\mathrm{T}}^{\mathrm{miss}}$	Single lepton
Leptons	0 *loose* leptons	Exactly 1 *tight* electron 0 additional *loose* leptons $p_{\mathrm{T}} > 27$ GeV	Exactly 1 *tight* muon 0 additional *loose* leptons $p_{\mathrm{T}} > 25$ GeV	Exactly 2 *loose* leptons $p_{\mathrm{T}} > 27$ GeV Same-flavour Opposite-sign charges ($\mu\mu$)
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 150 GeV	> 30 GeV	–	–
$m_{\ell\ell}$	–	–	–	81 GeV $< m_{\ell\ell} <$ 101 GeV
Jet p_{T}	> 20 GeV for $\lvert\eta\rvert < 2.5$ > 30 GeV for $2.5 < \lvert\eta\rvert < 4.5$			
b-jets	Exactly 2 b-tagged jets			
Leading b-tagged jet p_{T}	> 45 GeV			
Jet categories	Exactly 2 / Exactly 3 jets	Exactly 2 / Exactly 3 jets		Exactly 2 / ≥ 3 jets
H_{T}	> 120 GeV (2 jets), >150 GeV (3 jets)	–		–
$\min[\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}, \vec{\mathrm{jets}})]$	$> 20^{\circ}$ (2 jets), $> 30^{\circ}$ (3 jets)	–		–
$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}, \vec{bb})$	$> 120^{\circ}$	–		–
$\Delta\phi(\vec{b_1}, \vec{b_2})$	$< 140^{\circ}$	–		–
$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	$< 90^{\circ}$	–		–
p_{T}^{V} regions	– 150 GeV $< p_{\mathrm{T}}^{V} <$ 250 GeV $p_{\mathrm{T}}^{V} > 250$ GeV	– 150 GeV $< p_{\mathrm{T}}^{V} <$ 250 GeV $p_{\mathrm{T}}^{V} > 250$ GeV		75 GeV $< p_{\mathrm{T}}^{V} <$ 150 GeV 150 GeV $< p_{\mathrm{T}}^{V} <$ 250 GeV $p_{\mathrm{T}}^{V} > 250$ GeV
Signal regions	$\Delta R(\vec{b_1}, \vec{b_2})$ signal selection			
Control regions	High and low $\Delta R(\vec{b_1}, \vec{b_2})$ side-bands			

Analysis methods

Two multivariate analysis techniques are compared:

- ***B**oosted **D**ecision **T**ree (**BDT**)*
- ***N**eural **N**etwork (**NN**).*
- *In the LHC experiments:*
  - ✓ *ATLAS VHbb analysis uses BDT*
  - ✓ *CMS VHbb analysis uses Deep NN.*
  - ✓ *Previously, the cut-based method was used in both experiments.*

Cut-flow analysis

1. *Object selection - electron, muon, neutrino, b-jet selection*
2. *Final event selection*
3. Additional event selection cuts for cut-based analysis:

ZH→ννbb: *0 Lepton*

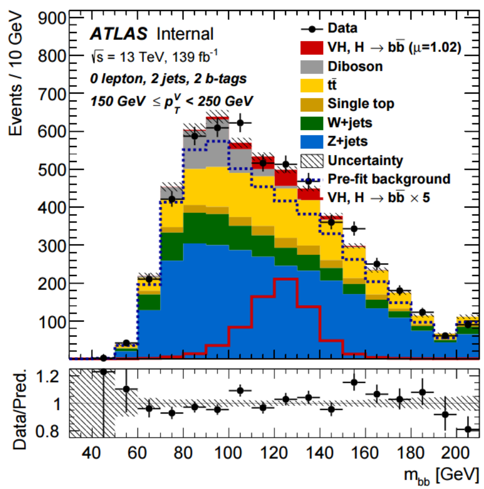
dR(b1,b2)<1.8 (reg. 1)
dR(b1,b2)<1.2 (reg. 2)

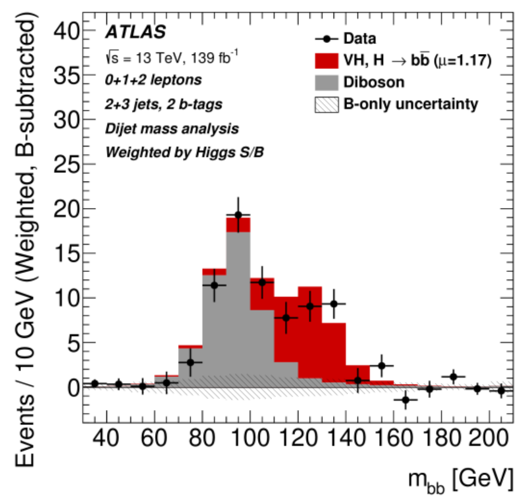
WH→lνbb: *1 Lepton*

dR(b1,b2)<1.8 (reg. 1)
dR(b1,b2)<1.2 (reg. 2)
mTW < 120 GeV

ZH→llbb: *2 Leptons*

dR(b1,b2)<1.8 (reg. 1)
dR(b1,b2)<1.2 (reg. 2)
81 < m(ll) < 101





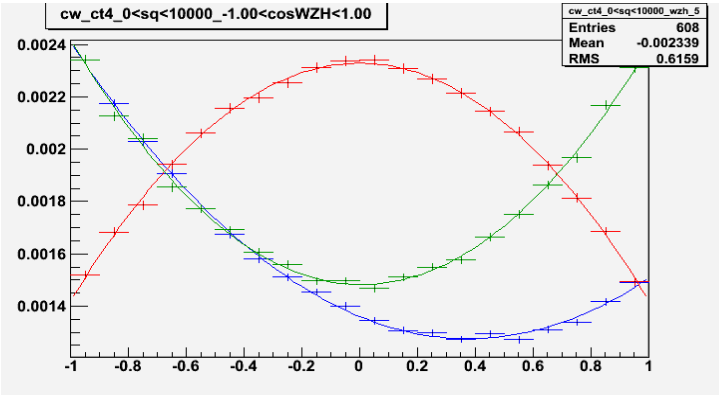
Multivariate Analysis

- The input variables are chosen in order to maximise the separation of the signal and background events.
- MVA is used to improve cut-flow analysis results.
- The selection criteria are looser in the MVA than in the cut-flow analysis in order to maximize the information available to the final discriminant.
- BDT & NN technique as an effective multivariate method, is used to properly account for correlations between variables.

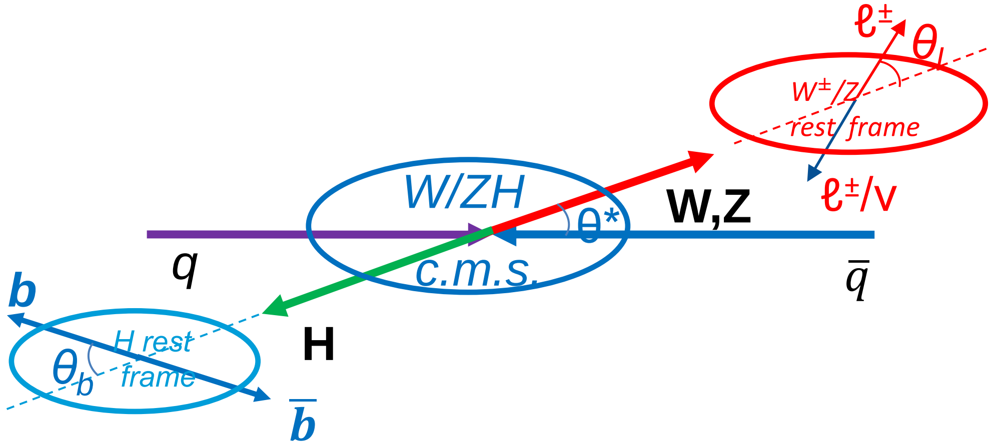
Input variables

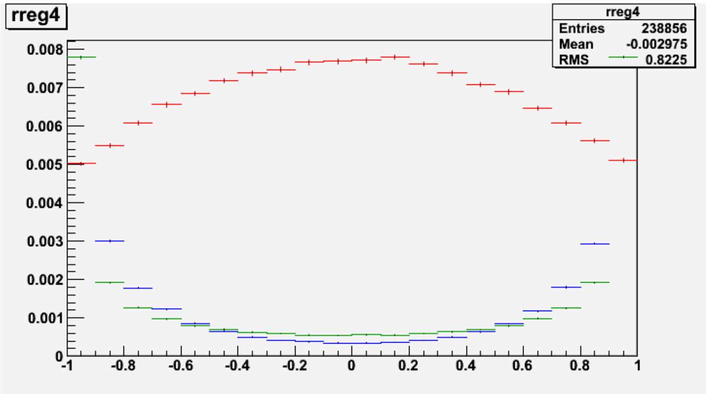
Variable	0-lepton	1-lepton	2-lepton
p_{T}^{V}		$\times$	$\times$
$E_{\mathrm{T}}^{\mathrm{miss}}$	$\times$	$\times$	$\times$
$p_{\mathrm{T}}^{b_1}$	$\times$	$\times$	$\times$
$p_{\mathrm{T}}^{b_2}$	$\times$	$\times$	$\times$
m_{bb}	$\times$	$\times$	$\times$
$\Delta R(b_1, b_2)$	$\times$	$\times$	$\times$
$\|\Delta\eta(b_1, b_2)\|$	$\times$		$\times$
$\Delta\phi(V, bb)$	$\times$	$\times$	$\times$
$\|\Delta\eta(V, bb)\|$			$\times$
H_{T}	$\times$		
$\min[\Delta\phi(\ell, b)]$		$\times$	
m_{T}^{W}		$\times$	
m_{ll}			$\times$
m_{Top}		$\times$	
$\|\Delta Y(V, H)\|$		$\times$	
	Only in 3-jet events		
$p_{\mathrm{T}}^{\mathrm{jet}_3}$	$\times$	$\times$	$\times$
m_{bbj}	$\times$	$\times$	$\times$

Some new variables

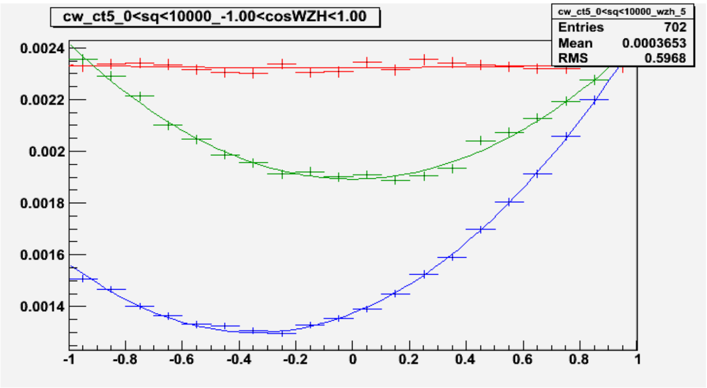


$\cos\theta_e^*$





$\cos\theta_W$



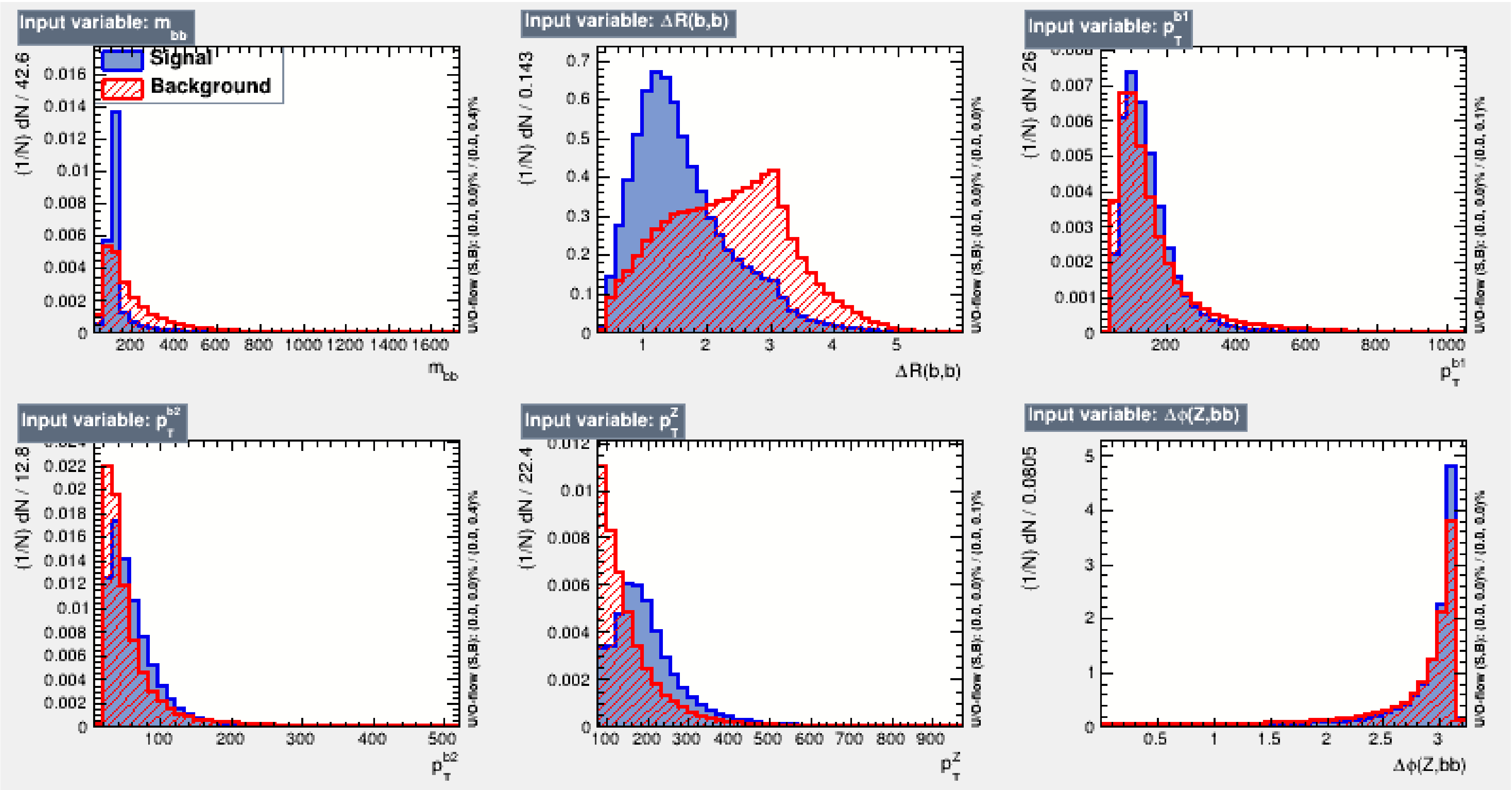
$\cos\theta_b^*$

$WHb\bar{b}$

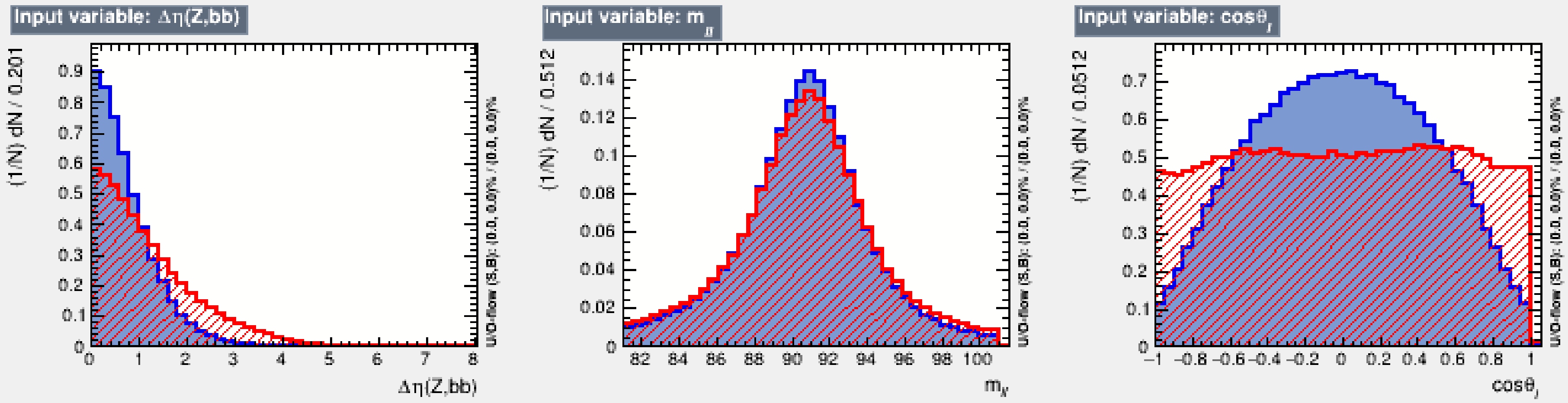
$WZb\bar{b}$

$Wb\bar{b}$

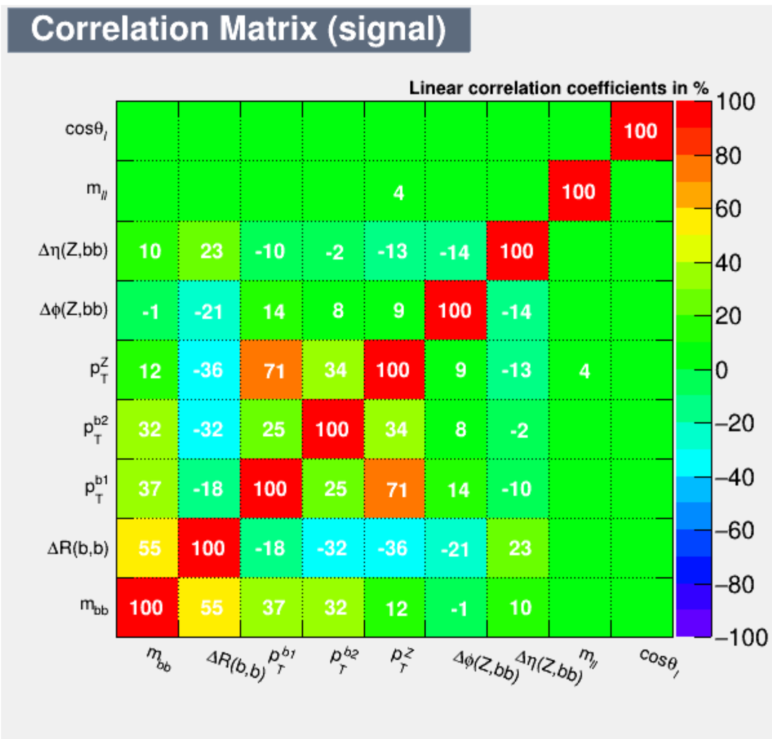
MVA input variables

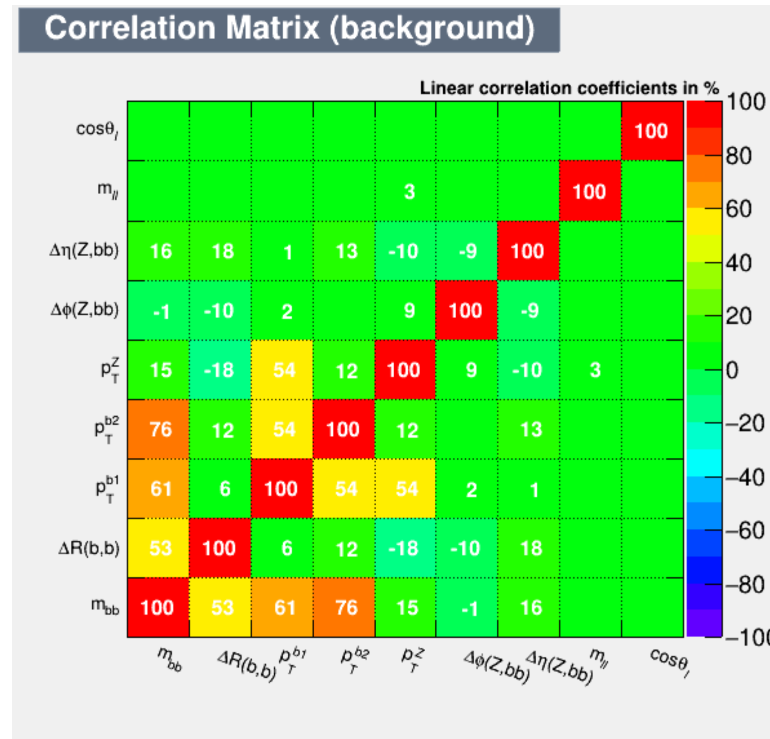


MVA input variables



Correlation Matrix





Signal **Background**

BDT training settings

Input variables:

2-Lepton channel:

"mBB:dRBB:dPhiVBB:dEtaVBB:pTV:pTB1:pTB2:mLL:cosThetaLep

+ mBBJ:pTJ3 for ≥ 3 jet events

hyper-parameters:

BoostType= ["Grad"]
Shrinkage = [0.5]
SeparationType = ["GiniIndex"]
PruneMethod = ["NoPruning"]

NTrees = [**100,200,300**]
MaxDepth = [**3,4,5**]
nCuts = [**50,100,150**]
MinNodeSize = [**4,5,6**]

81 hyper-parameters combinations

NN training settings

Input variables:

2-Lepton channel:

"mBB:dRBB:dPhiVBB:dEtaVBB:pTV:pTB1:pTB2:mLL:cosThetaLep

+ mBBJ:pTJ3 for ≥ 3 jet events

hyper-parameters:

Types=TMlpANN;

NCycles=200, 1000, 2000, 3000, 4000, 6000;

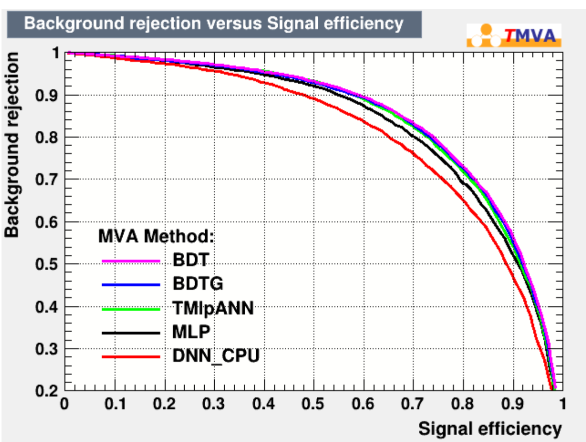
Hidden Layers=N,N-1; N,N; N,N+1;

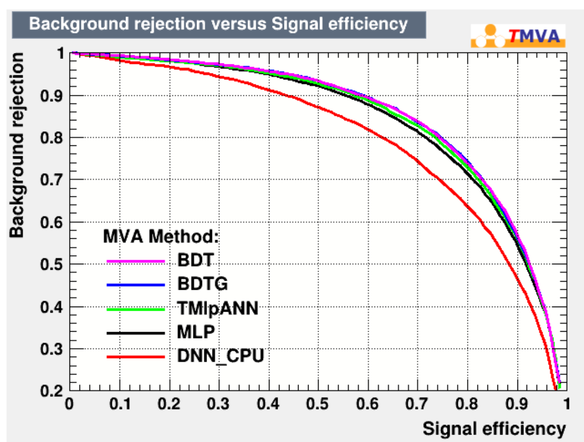
Learning Method=BFGS;

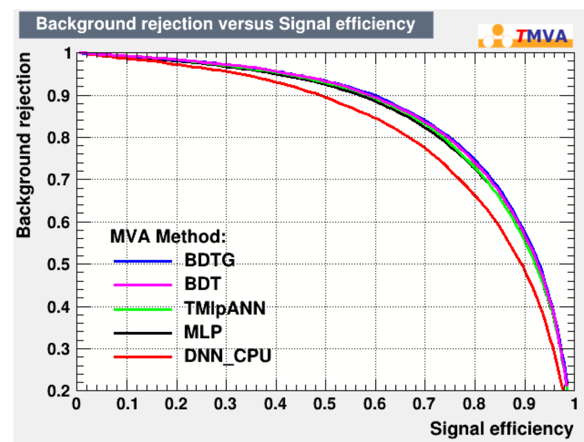
Validation Fraction=0.3, 0.5.

ROC from BDT & NN

- **BDT with optimised hyper-parameter setting**
  Types=BDTG,NTrees=200:MaxDepth=4:BoostType=Grad:Shrinkage=0.5:
  SeparationType=GiniIndex:nCuts=100:MinNodeSize=5:PruneMethod=NoPruning.
- **NN with default hyper-parameter setting**
  Types=TMlpANN, NCycles=200:HiddenLayers=N,N-1:LearningMethod=BFGS:
  ValidationFraction=0.3;
- **Number of signal (N_s) & background (N_b) events:** $N_s = N_b$;
- **Number of training (N_{tr}) & Number of testing (N_{ts}) events:** $N_{tr} = N_{ts}$;
- **Number of training signal events:** $N_{tr,s}$ = 10K, 20K, 50K, 100K, 200K, 400K.







N_{tr} = 10K **N_{tr} = 20K** **N_{tr} = 50K**

ROC from BDT & NN

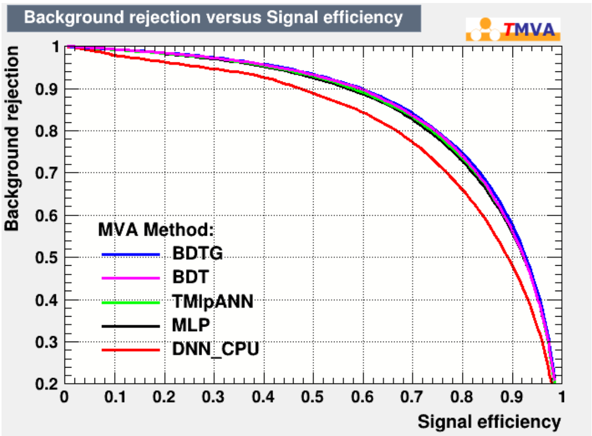
- **BDT with optimised hyper-parameter setting**

  Types=BDTG,NTrees=200:MaxDepth=4:BoostType=Grad:Shrinkage=0.5:
  SeparationType=GiniIndex:nCuts=100:MinNodeSize=5:PruneMethod=NoPruning.

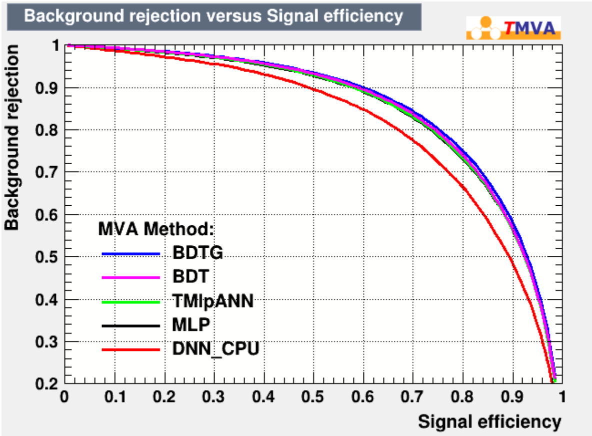
- **NN with default hyper-parameter setting**

  Types=TMlpANN, NCycles=200:HiddenLayers=N,N-1:LearningMethod=BFGS:
  ValidationFraction=0.3;

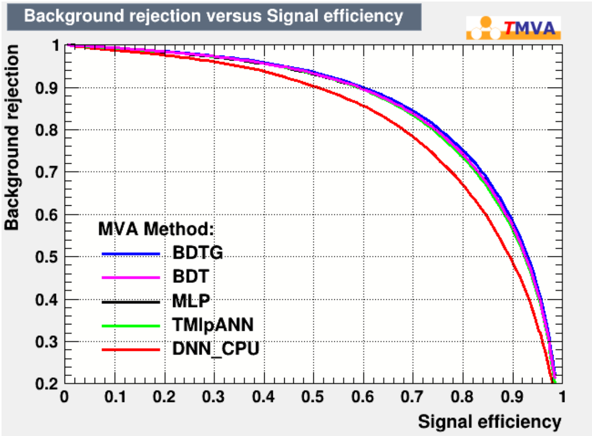
- **Number of signal (N_s) & background (N_b) events:** $N_s = N_b$;
- **Number of training (N_{tr}) & Number of testing (N_{ts}) events:** $N_{tr} = N_{ts}$;
- **Number of training signal events:** $N_{tr,s}$ = 10K, 20K, 50K, 100K, 200K, 400K.



N_{tr} = 100K



N_{tr} = 200K



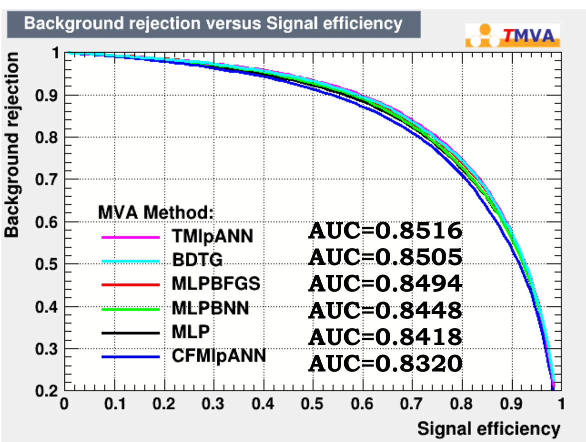
N_{tr} = 400K

ROC from BDT & NN

- **BDT with optimised hyper-parameter setting**

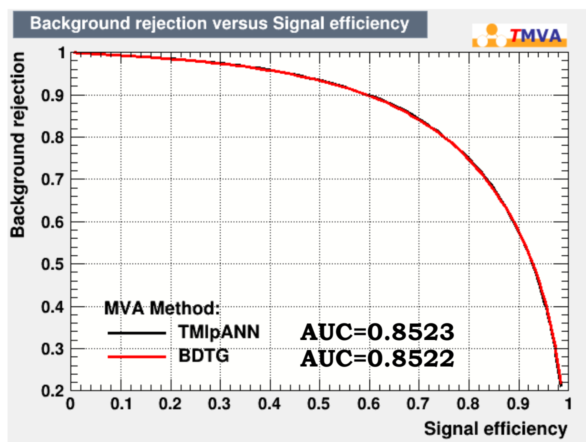
  Types=BDTG,NTrees=200:MaxDepth=4:BoostType=Grad:Shrinkage=0.5:
  SeparationType=GiniIndex:nCuts=100:MinNodeSize=5:PruneMethod=NoPruning.
- **NN with optimised hyper-parameter setting**
  Types=TMlpANN, NCycles=**2000/3000/4000**:HiddenLayers=N,N-1:
  LearningMethod=BFGS: ValidationFraction=**0.5**;
- **Number of signal (N_s) & background (N_b) events:** $N_s = N_b$;
- **Number of training (N_{tr}) & Number of testing (N_{ts}) events:** $N_{tr} = N_{ts}$;
- **Number of training signal events:** $N_{tr,s}$ = 10K, 20K, 50K, 100K, 200K, 400K.

NCycles=2000



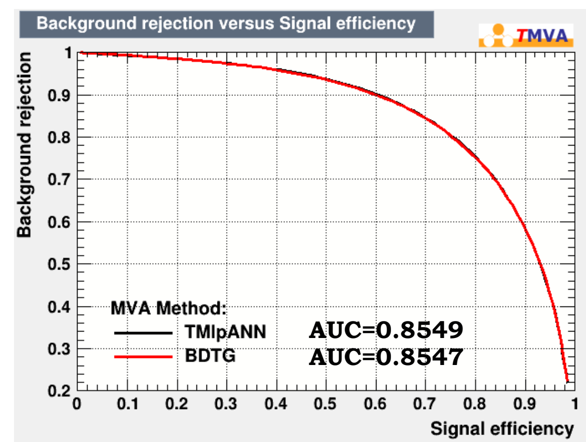
$N_{tr,s}$ = 50K

NCycles=4000



$N_{tr,s}$ = 100K

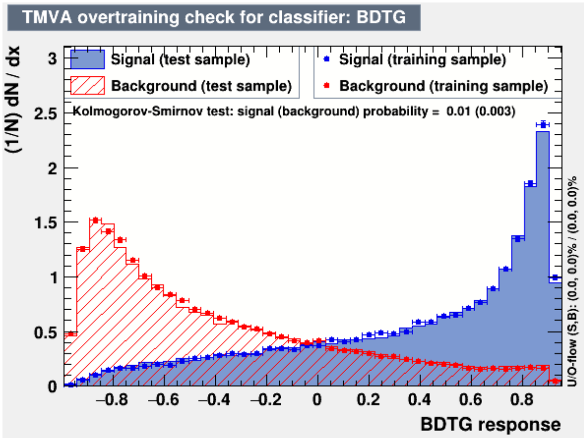
NCycles=3000

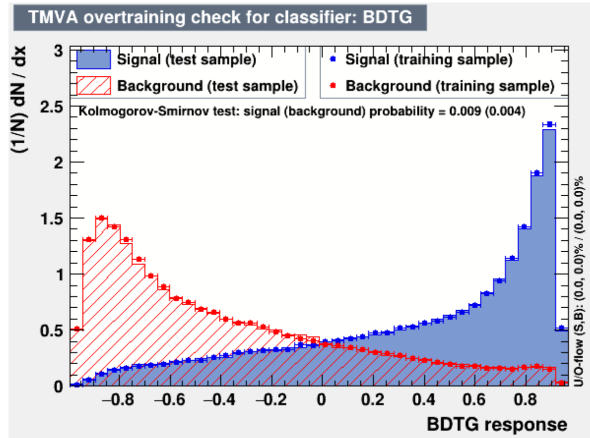


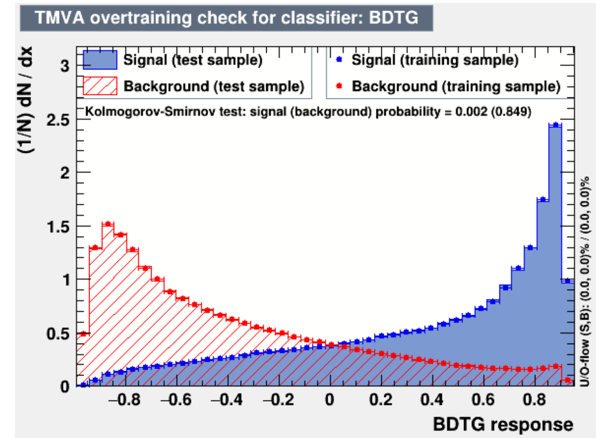
$N_{tr,s}$ = 400K

Overtraining check

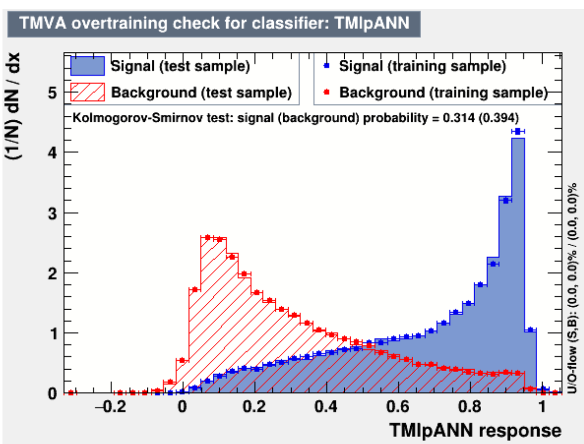
BDT

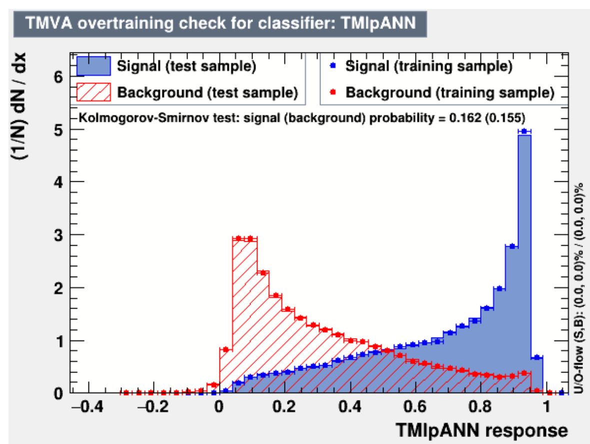


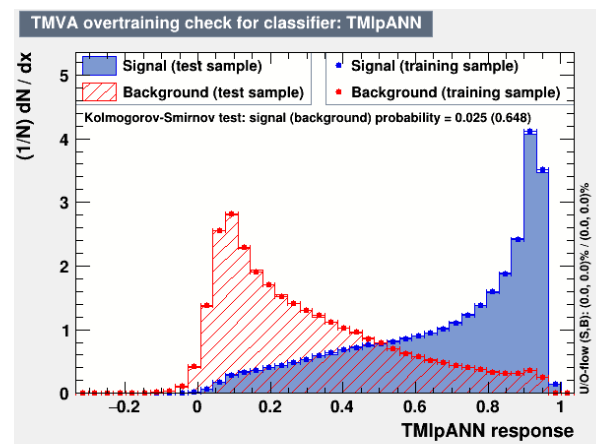




NN





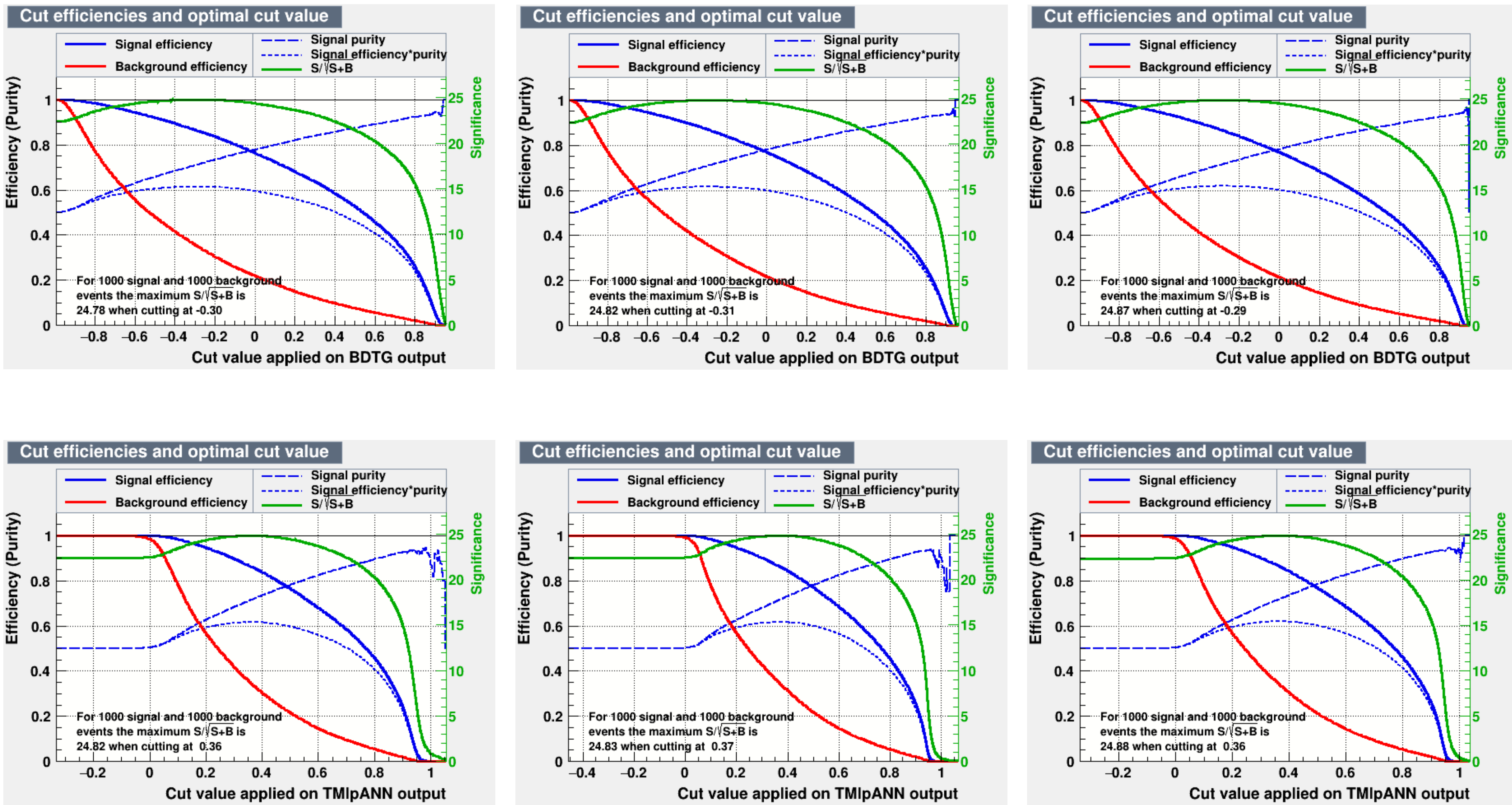


$N_{tr,s}$ = 50K | $N_{tr,s}$ = 100K | $N_{tr,s}$ = 400K

Efficiencies

BDT

NN



$N_{tr,s}$ = 50K $N_{tr,s}$ = 100K $N_{tr,s}$ = 400K

Conclusions

- The two MVA methods BDT and NN were compared to find the methods with the best performance for *VHbb* analysis.
- Up to 0.8 million signals and the same number of background events were used for training.
- The settings with the best performance were chosen to tune the BDT hyperparameters.
- A different number of events are trained and different settings for NN are obtained, providing performance that exceeds that of BDT.
- It turns out that for any number of training events, it is possible to find corresponding NN settings with better performance than BDT.
- The only problem with NN training is that, compared to BDT, it requires more computational resources.

Thank you for your attention!