Boosted Higgs boson in association with a vector boson

Laura Caviedes[2], Juan Manuel Moreno[2]

Reina Camacho Toro[1], Carlos Sandoval[2],Dimitris Varouchas[3]

UNIVERSIDAD NACIONAL DE COLOMBIA

7th COMHEP
01/12/2022

ATLAS EXPERIMENT

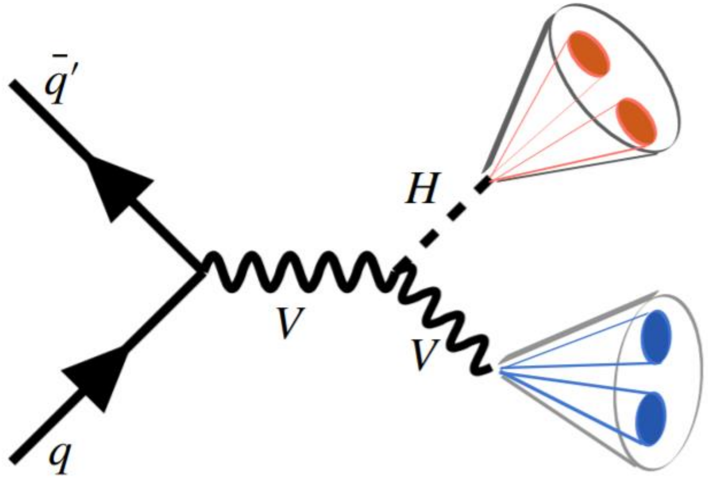
[1] LPNHE, [2] UNAL, [3] IJCLab

Outline

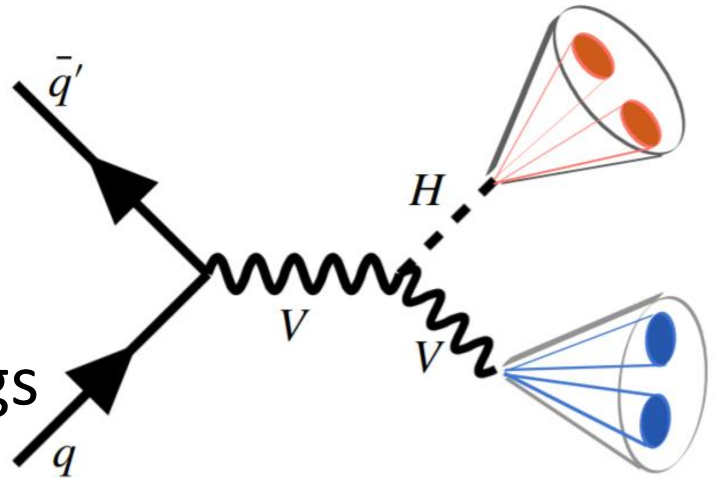
1. Objective
2. Summary
3. Variable distribution
4. Machine Learning Training
5. Conclusions

Objective

Improve the event selection for $VH \to q\bar{q}b\bar{b}$ production using machine learning techniques and calculate the signal significance.



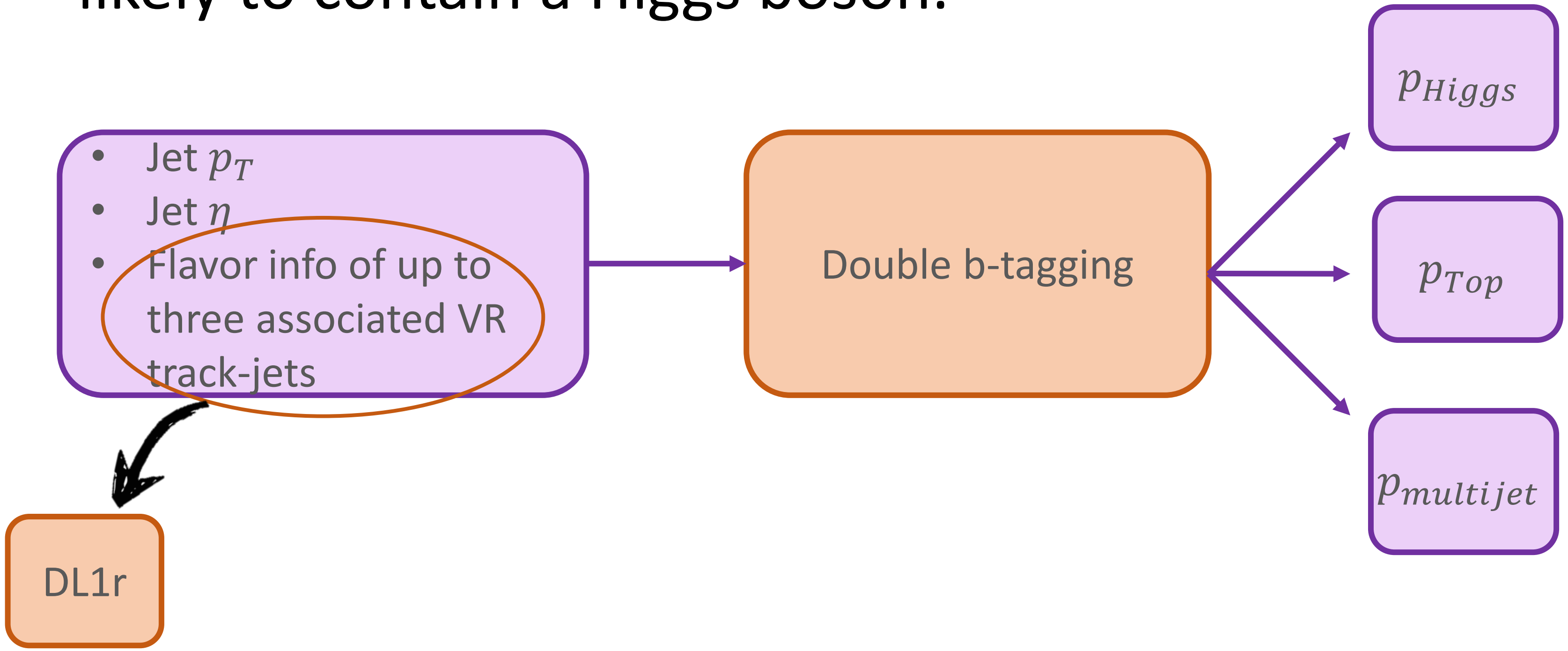
Motivation



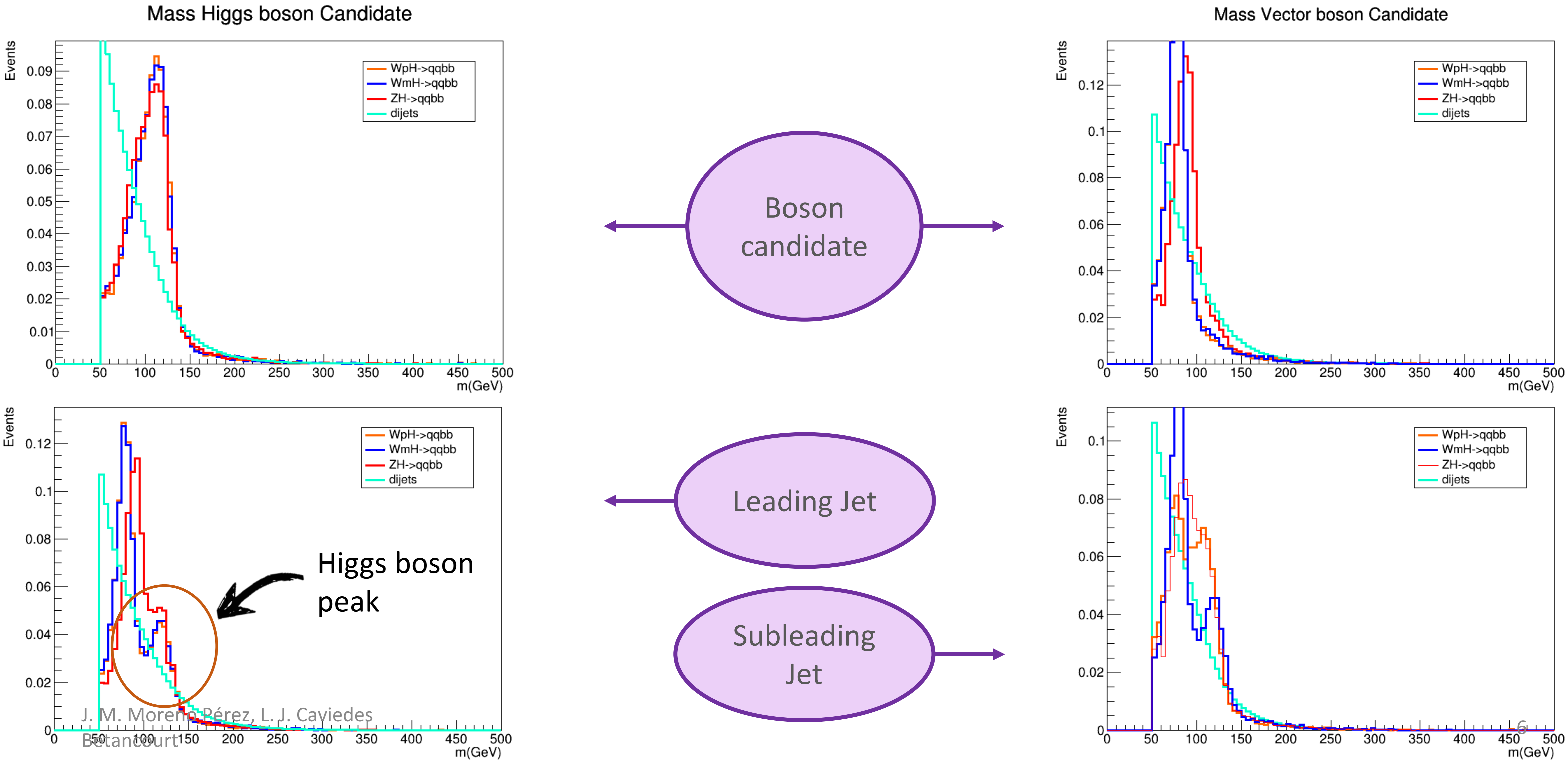
- The $H \rightarrow b\bar{b}$ decay allows us to access to the Higgs couplings for the second generation of fermions.
- This decay was seen for the first time by the ATLAS and CMS collaboration in 2018 using Higgs production associated with W/Z bosons.
- The identification of massive particles decaying into bottom-antibottom quark pairs at high transverse momenta in ATLAS is performed using a multivariate algorithm, denoted an $X \rightarrow b\bar{b}$ tagger.

Double b-tagging

Is a neural-network (NN) used to identifying large R jets originated from massive particles decaying into $b\bar{b}$ pair and determine if an event is likely to contain a Higgs boson.



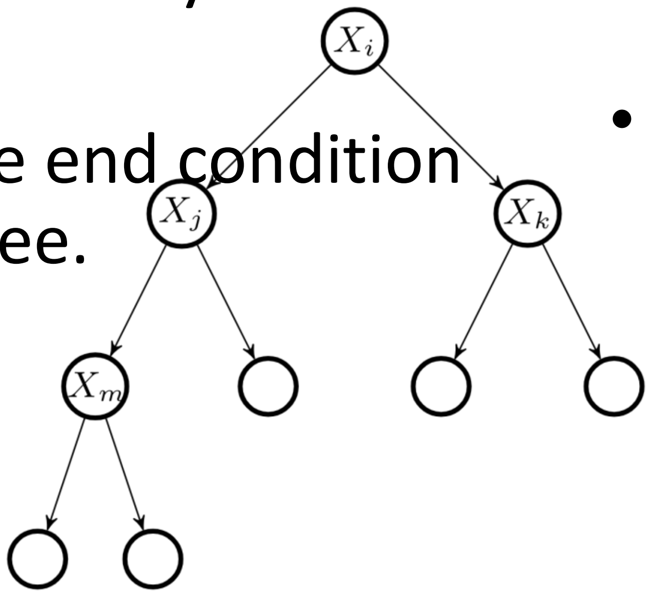
Boson tagging



MLP and BDT methods

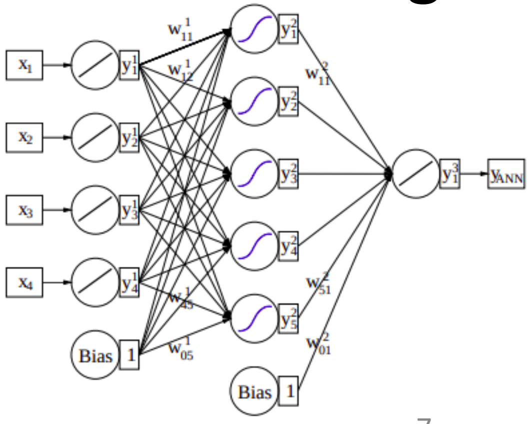
Boosted Decision Tree

- It takes a set of input features and splits input data recursively based on those features.
- Parameter define the end condition for building a new tree.

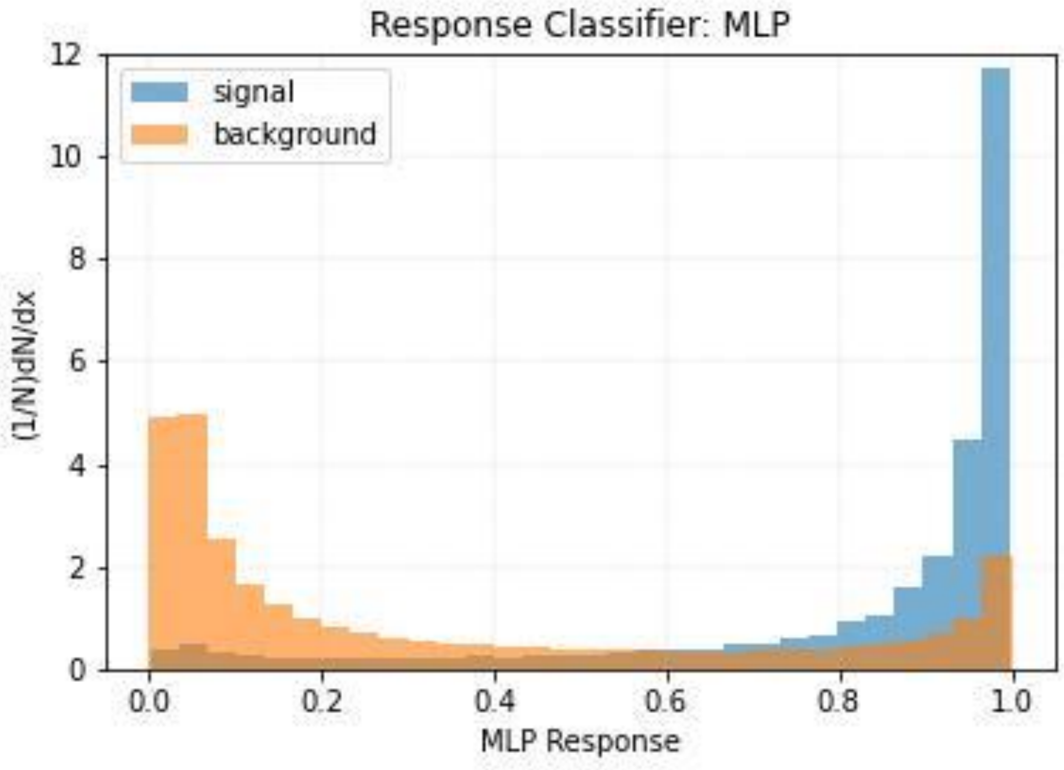


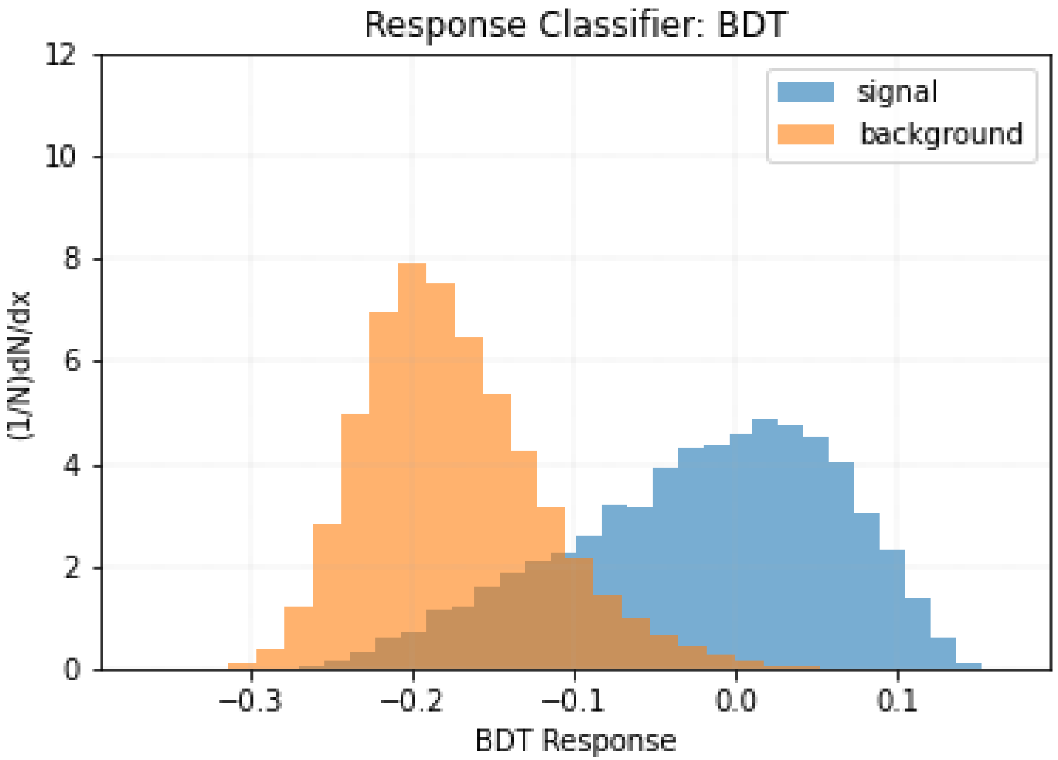
MLP

- A neural network is composed by neurons, in this case they are arranged by layers.
- Connections between neurons get weight assigned.

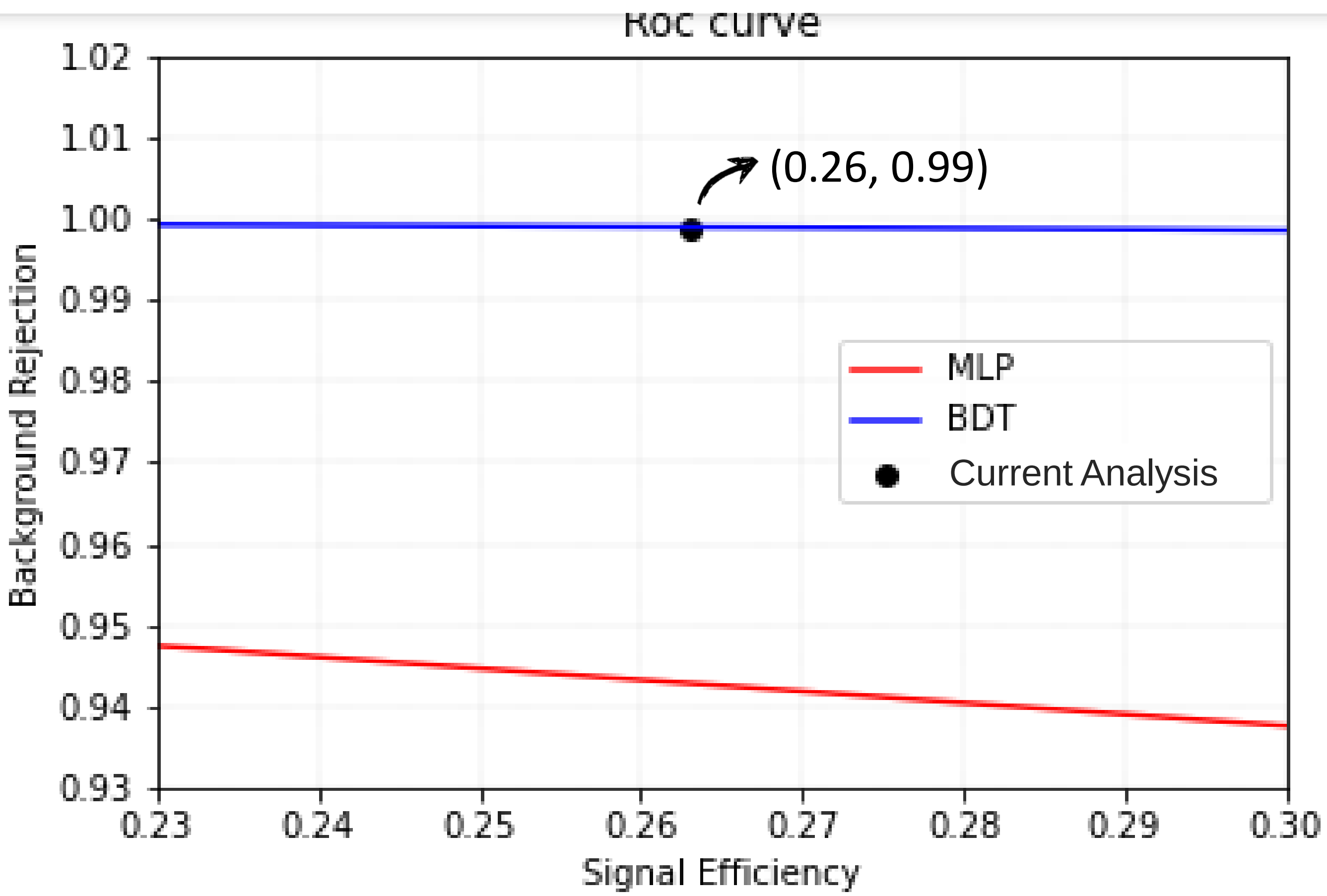


ML techniques response



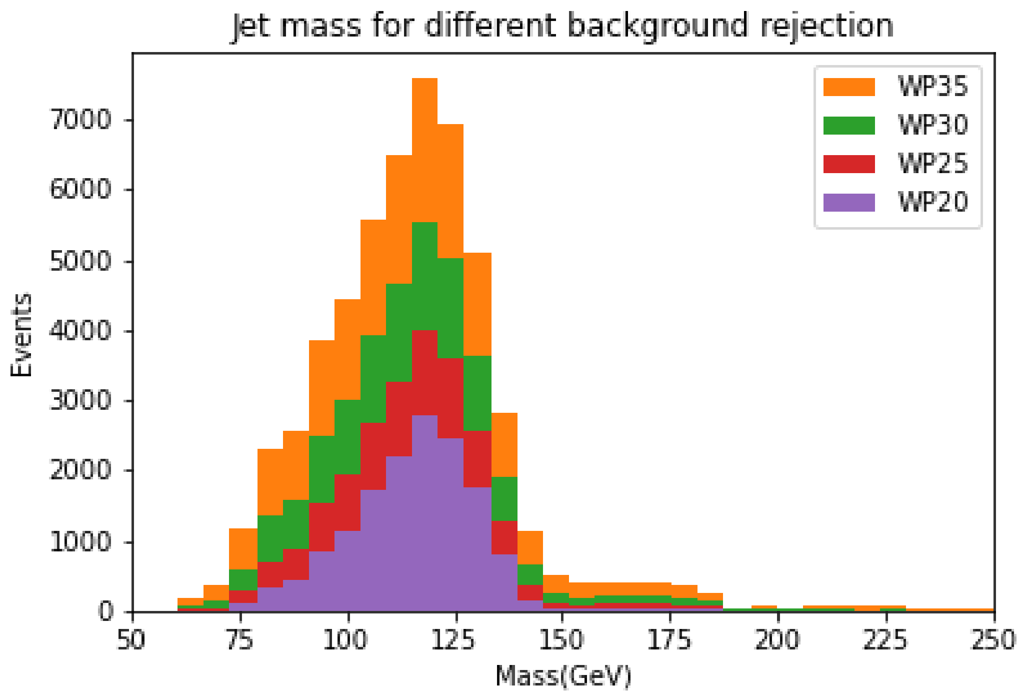


Working points signal efficiency



Although slight at large scale there is indeed an improvement on the efficiency after the BDT application.

Mass distribution of Higgs Candidate



Sample with signal ($VH \rightarrow qqb\bar{b}$) and background (dijets)

Conclusions

ML techniques such as BDT which was our chosen technique provide a wide range of possibilities for improvement in HEP analysis.

Further studies should include systematic uncertainties produced by the BDT to evaluate if the improvement is robust and study of different background rejection efficiency.

Thanks!!

Special Thanks to ERASMUS for the finantial support, the University Paris-Saclay and to our supervisors for the oportunity and support.

Backup

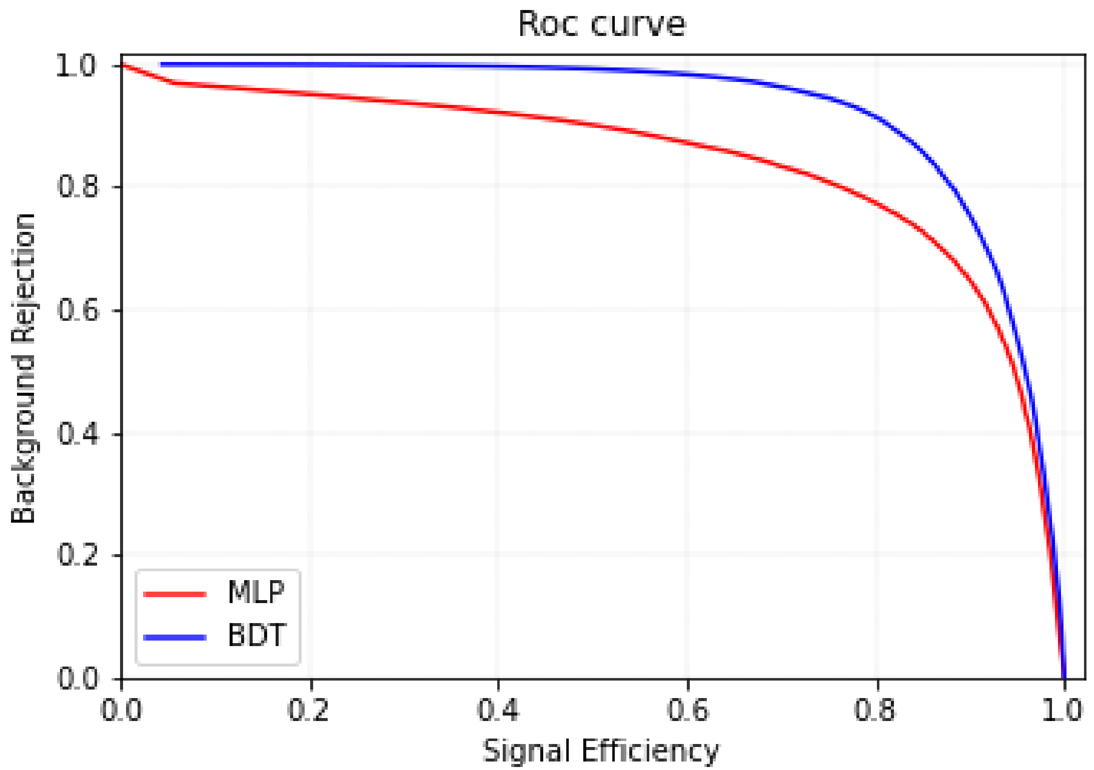
Variable	Mean	RMS	[Min	Max]
Jet1_M:	121.58	59.161	[60.000	2746.5]
Jet2_M:	122.84	55.513	[60.000	2733.4]
HCand_M:	127.98	56.501	[60.000	2697.8]
Jet1_pT:	592.58	152.68	[450.00	7519.2]
Jet2_pT:	497.74	161.32	[200.00	6156.9]
HCand_XbbScore:	0.024535	2.9759	[-7.7451	8.9881]
Jet1_D2:	1.4125	4.0357	[-999.00	1959.6]
Jet2_D2:	1.4207	3.7323	[-999.00	2403.8]
Jet1_nTrackJets:	2.8248	0.91322	[0.0000	13.000]
H_trackJetPtAsym:	-9.2556	97.902	[-999.00	1.0000]
V_trackJetPtAsym:	-14.903	122.82	[-999.00	1.0000]
dPhiVH:	3.0026	0.18703	[0.0015870	3.1416]
dEtaVH:	0.87970	0.69580	[8.9407e-08	3.9646]
dyVH:	0.86094	0.67944	[4.3183e-07	3.9330]
pTAsym:	0.095641	0.090335	[0.0000	0.92641]
dR:	3.3071	0.32939	[0.94583	6.6112]
VCand_Tau32:	0.63033	0.15472	[-1.0000	0.94114]
VCand_Tau21:	0.38290	0.16715	[-1.0000	0.91376]

Background rejection on different samples

Background Rejection	
Current Analysis	99.86 %
BDT on training background	99.88 %
BDT on Wqq	99.69 %
BDT on Zqq	99.77 %
BDT on ttbar_allhad	99.63 %

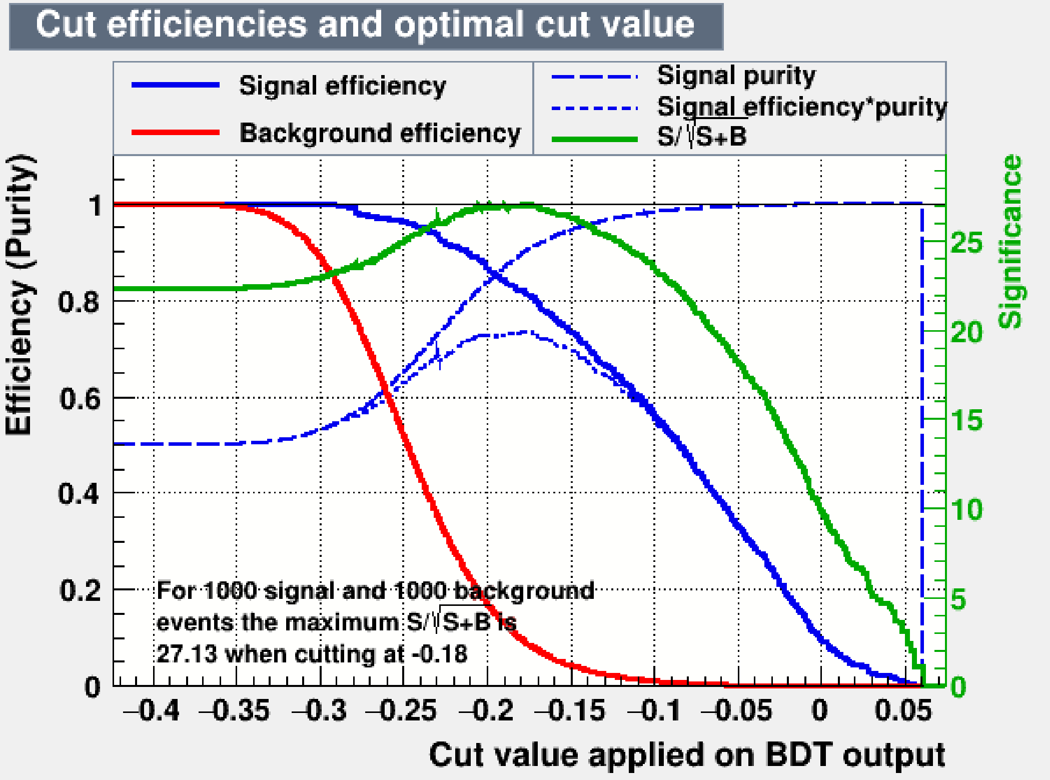
Machine Learning Training

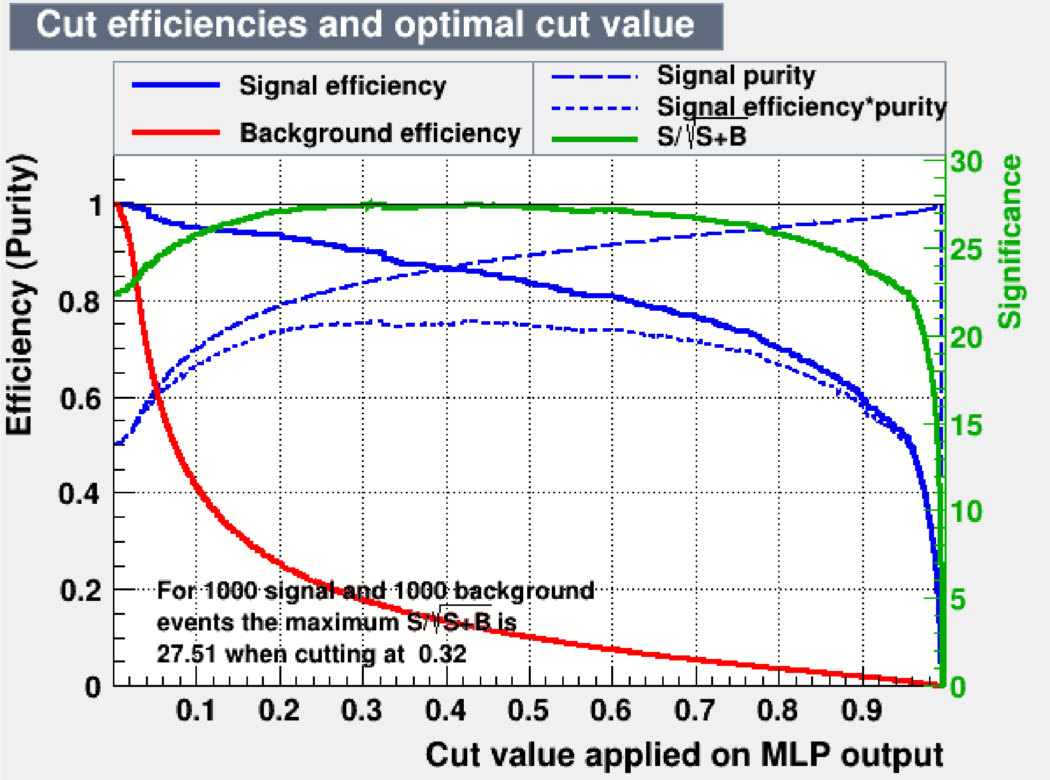
- Deployment of a BDT and MPL methods
- Analyze changes in significance and signal efficiency



Machine Learning Training

- MLP is more stable and provides safer cutting range.
- BDT results are better overall





Machine Learning Training

- Training time is about 4 times greater for the MLP but execution over testing sample is about 18 times faster so is better for classifying large samples.

```
                                : Elapsed time for training with 3305178 events: 2.3e+03 sec
BDT                             : [datasetW] : Evaluation of BDT on training sample (3305178 events)
                                : Elapsed time for evaluation of 3305178 events: 235 sec
```

```
                                : Elapsed time for training with 3305178 events: 7.93e+03 sec        671/0.5732]   39]
MLP                             : [datasetW] : Evaluation of MLP on training sample (3305178 events)
                                : Elapsed time for evaluation of 3305178 events: 8.82 sec
```

```
Factory                  : Test method: BDT for Classification performance
                         :
BDT                      : [datasetW] : Evaluation of BDT on testing sample (3305178 events)
                         : Elapsed time for evaluation of 3305178 events: 178 sec
Factory                  : Test method: MLP for Classification performance
                         :
MLP                      : [datasetW] : Evaluation of MLP on testing sample (3305178 events)
                         : Elapsed time for evaluation of 3305178 events: 9.59 sec
```