Machine learning methods to improve boosted Higgs boson tagging at ATLAS

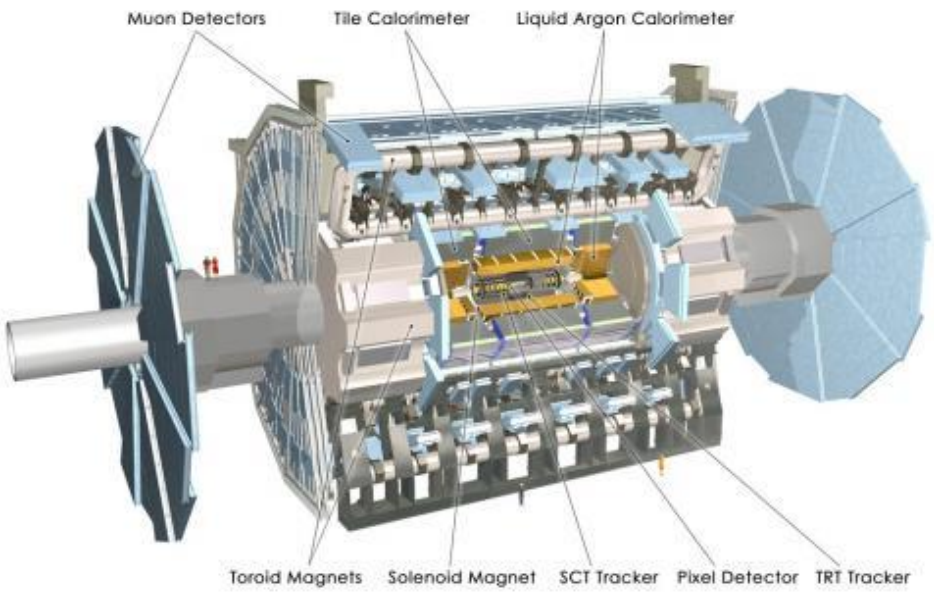
Speaker: Vladlen G.
Supervisor: Inês Ochoa

A brief intro...

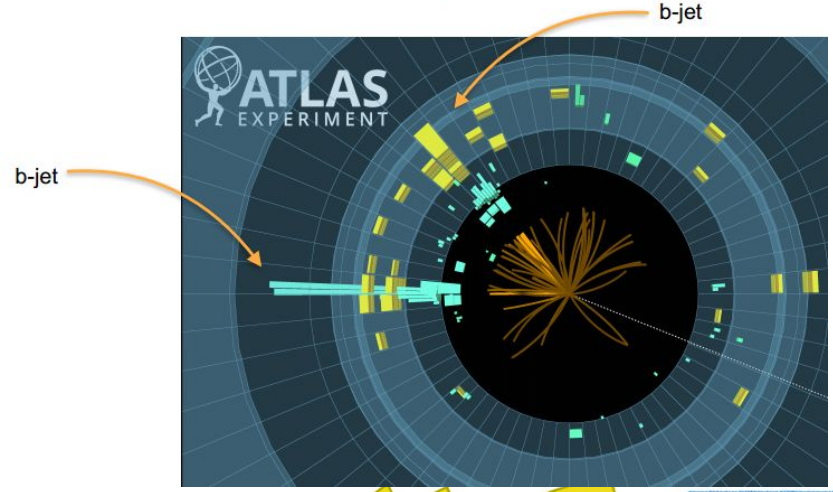
Why?
-To Study Beyond the Standard Model physics :

-There, heavy resonances may appear, which can decay into Higgs bosons with high pt (boosted).

-Reconstruction techniques for boosted higgs boson play an important role.



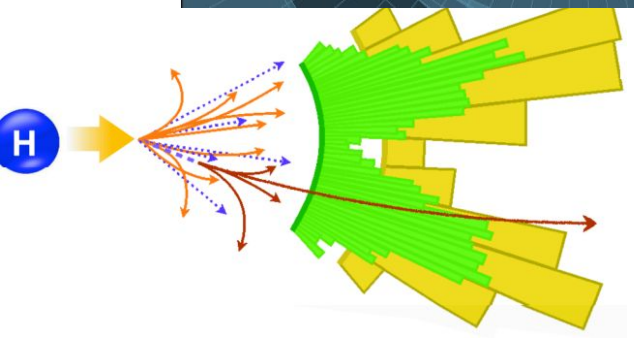
Boosted Higgs and difficulties

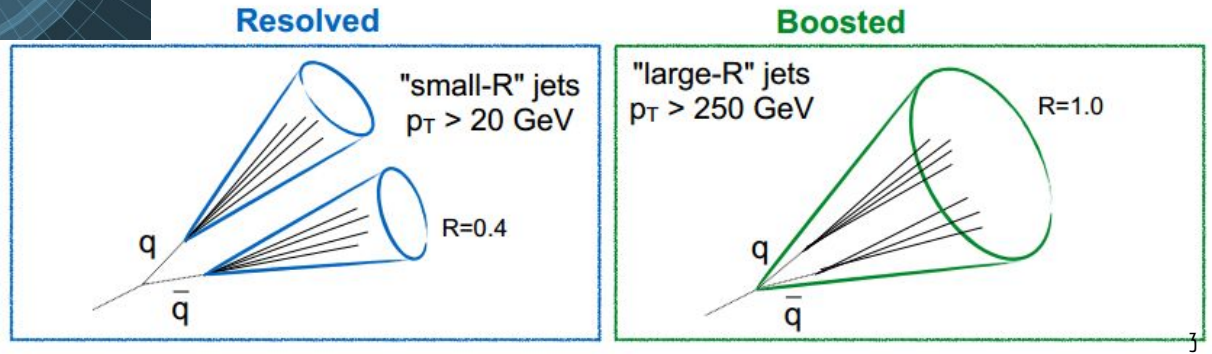


What are jets?

What's the difference between boosted and not boosted regimes?

Why is it so difficult to characterize them?



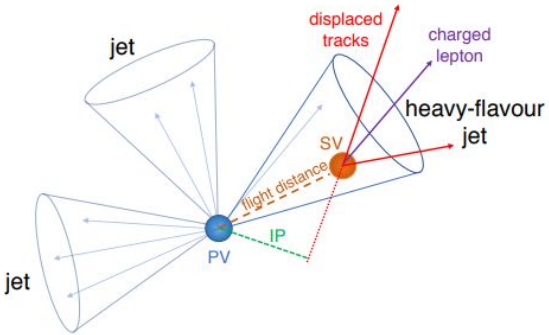


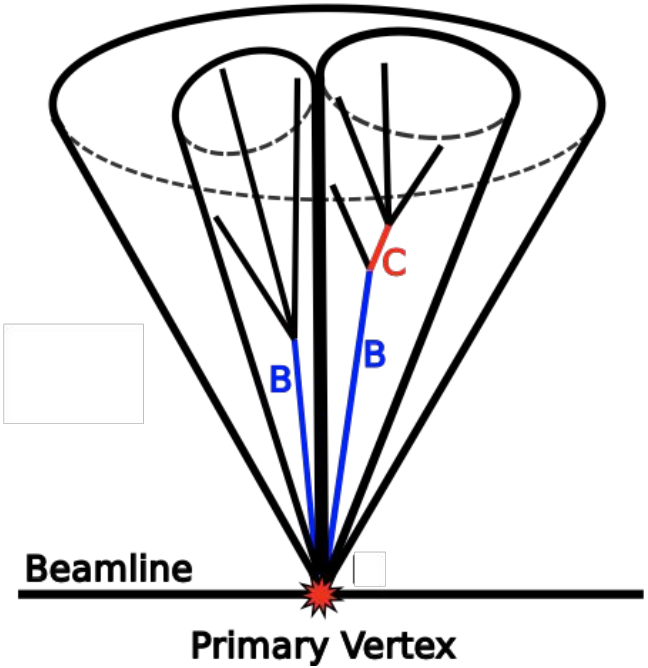
b-tagging method

What is b-tagging?

- b Quarks fragment into B-hadrons

- Displaced vertex (secondary vertex) from primary vertex due to its long life (~1.5ps)





A deep NN and how to feed it

Architecture

Layers: **5 hidden**
Nodes: **112, 96, 48, 24, 12, 6, 3**
Optimizer: **Adam**
Loss: **Categorical Crossentropy**
Learning rate: **0.01**
Batch size: **256**
Activation function: **Relu**

-Classes: **Higgs, QCD (dijets), Top**

Food: **Data (treated)**

With **sauce**: **Features**

-pt: Transverse momentum
-eta: Pseudorapidity
-DL1rT_pu
-DL1rT_pb
-DL1rT_pc
x3 Jets that we got
9 features

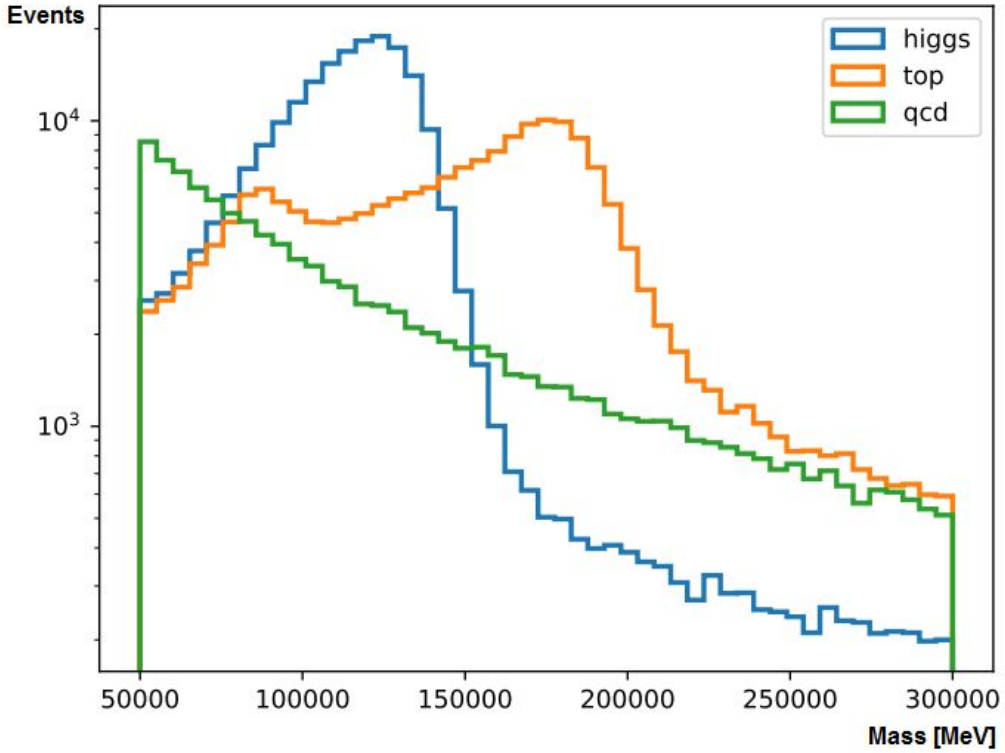
ex: DL1rT_pu_1: probability distribution of the jet 1 being created from an up quark

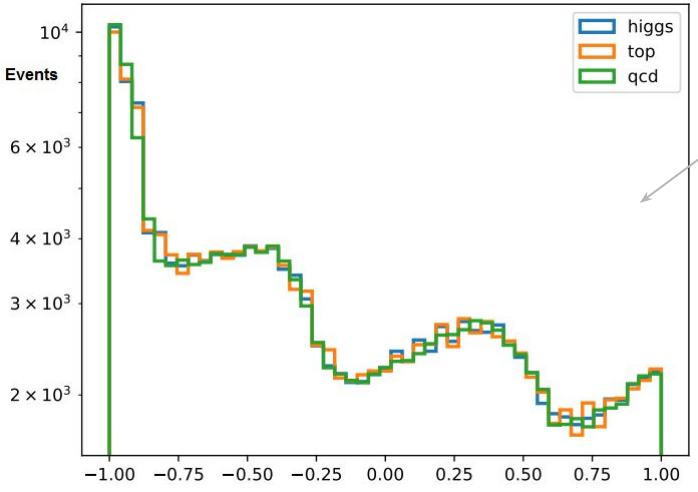
11 Features in Total!

Higgs,QCD and Top mass distribution

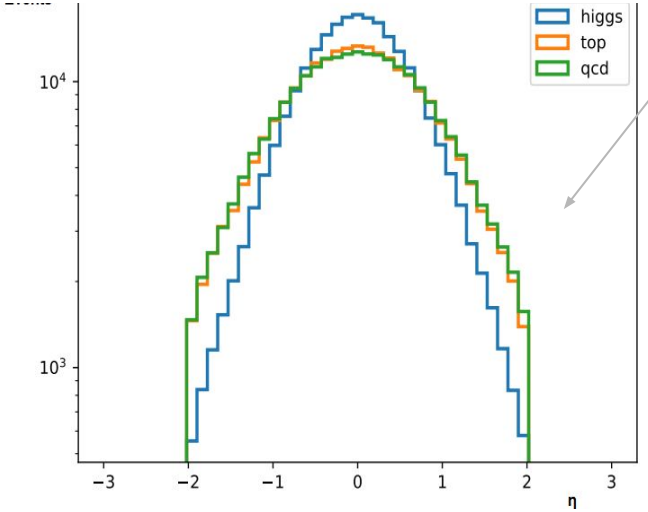
Using validation plots we get:

Mass does not enter the NN, cause it will be used to create our fits



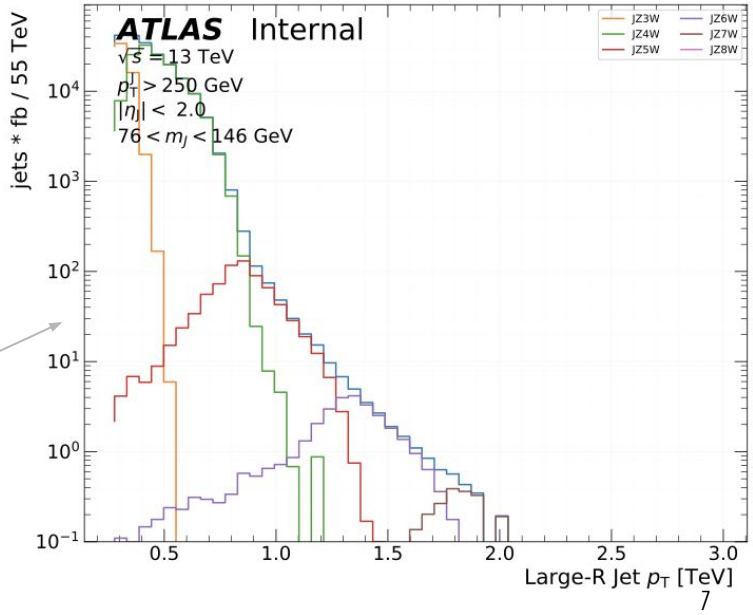


pt-Transverse momentum of large-radius jet



Pseudorapidity for large-radius jet,eta

We need weights to create a physical distribution of pts



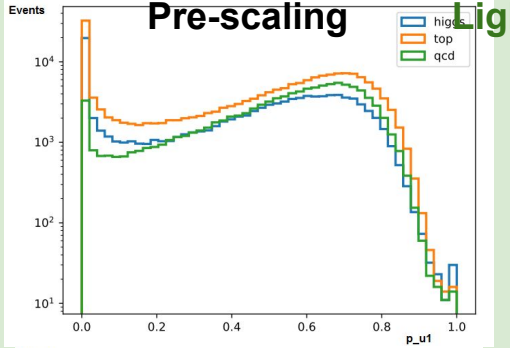
Validation, features:

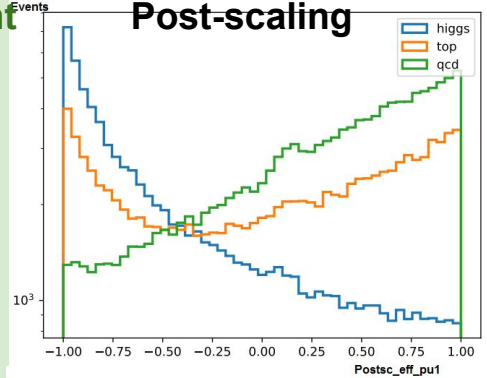
Nº of events vs pts/eta

Pre-scaling

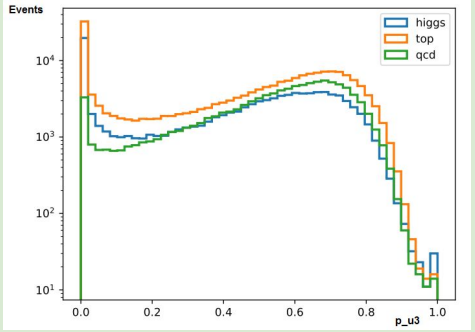
Light

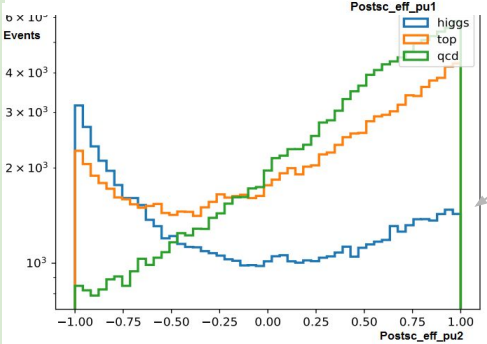
Post-scaling



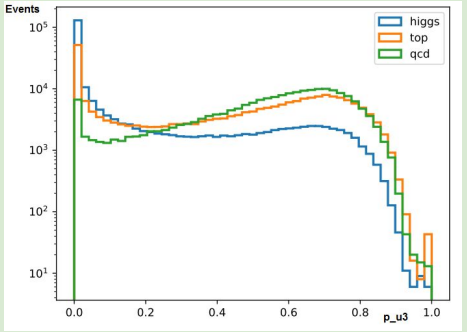


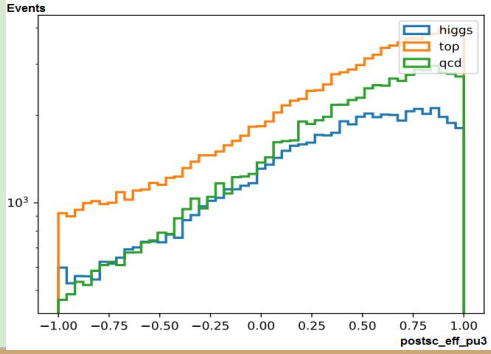
J1





J2





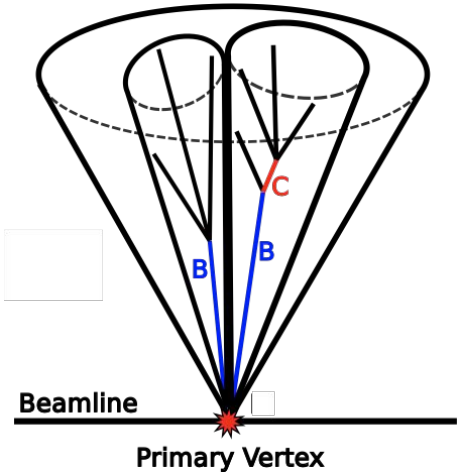
J3

LIP

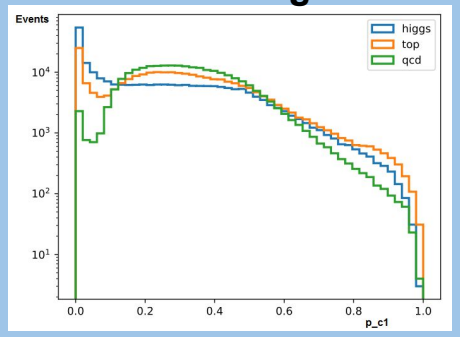
ATLAS EXPERIMENT

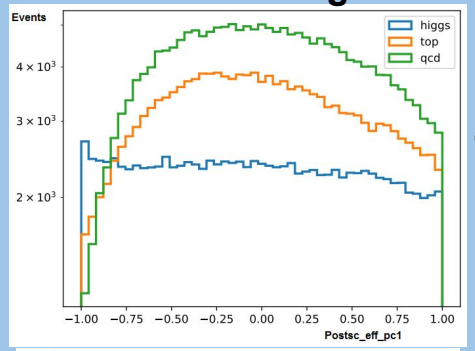
Nº of events vs DL1rT_pu_x

Probability of three VR track-jets to be light, for large-R jets in Higgs,Top and QCD

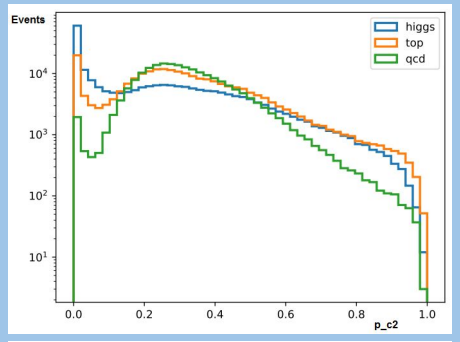


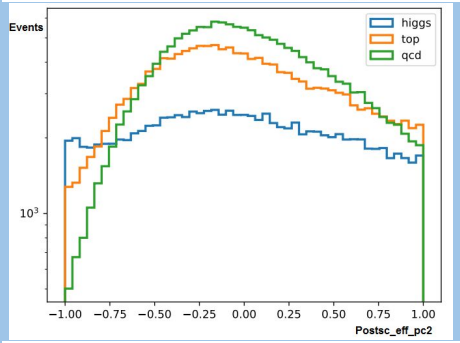
Pre-scaling CHARM Post-scaling



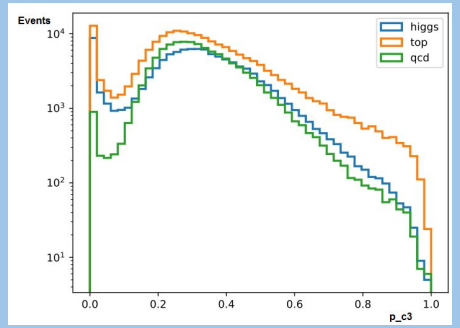


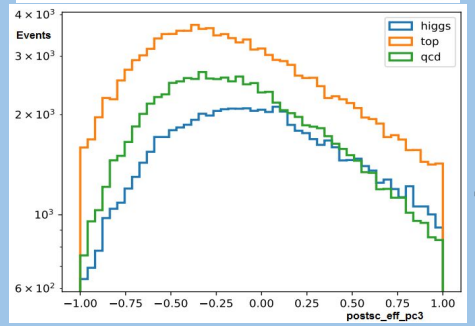
J1





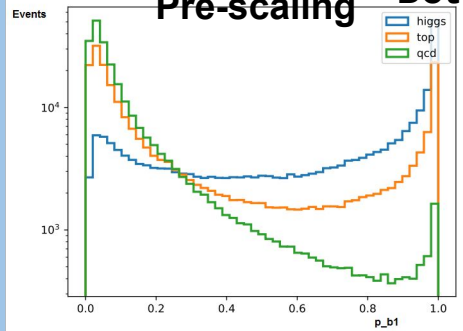
J2

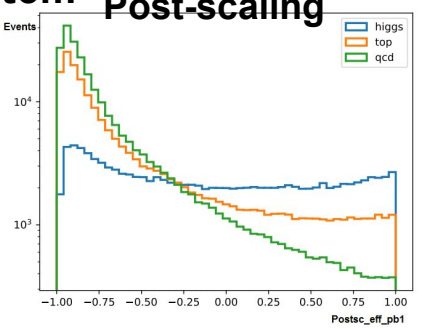


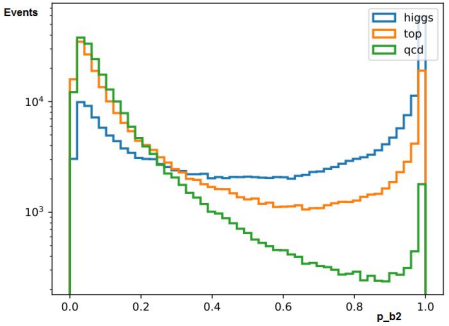


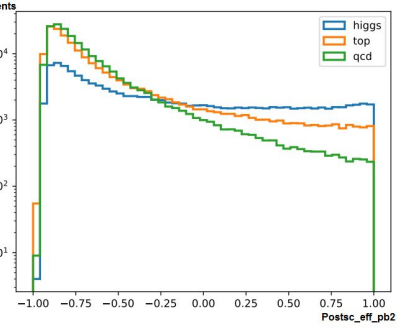
J3

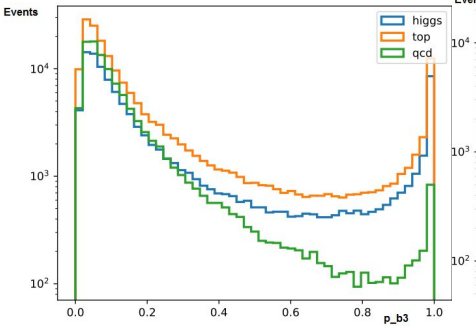
Bottom Pre-scaling Post-scaling

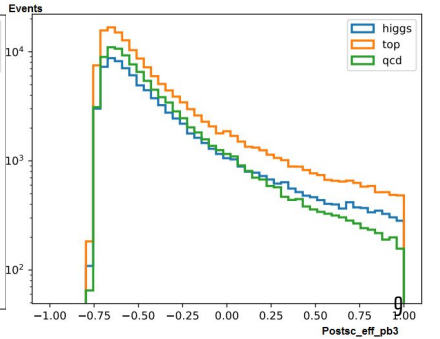








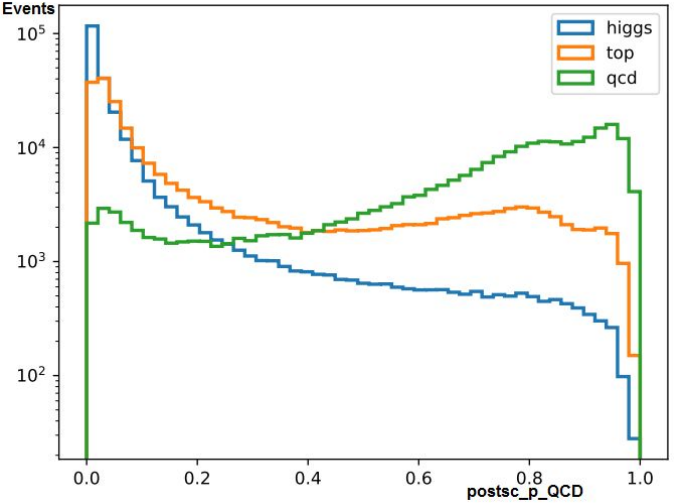




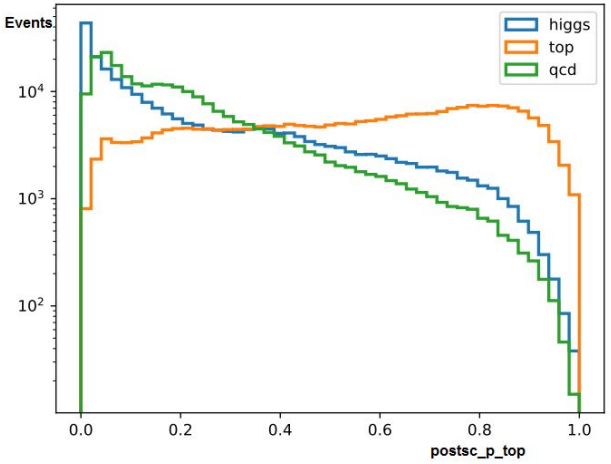
Results to reproduce...

https://cds.cern.ch/record/2724739/files/ATL-PHYS-PUB-2020-019.pdf

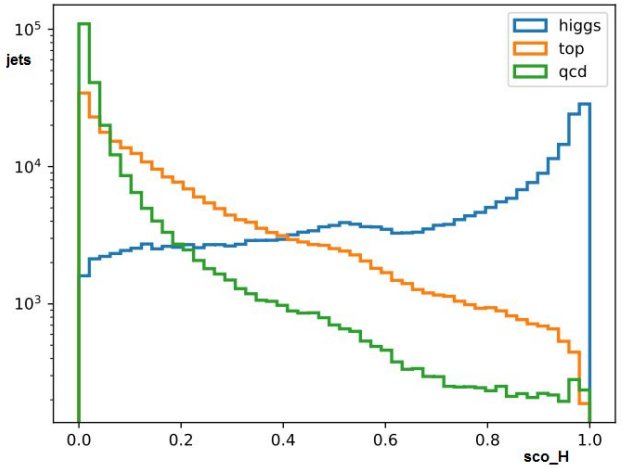
Probability of large-R jet to be QCD



Probability of large-R jet to be Top



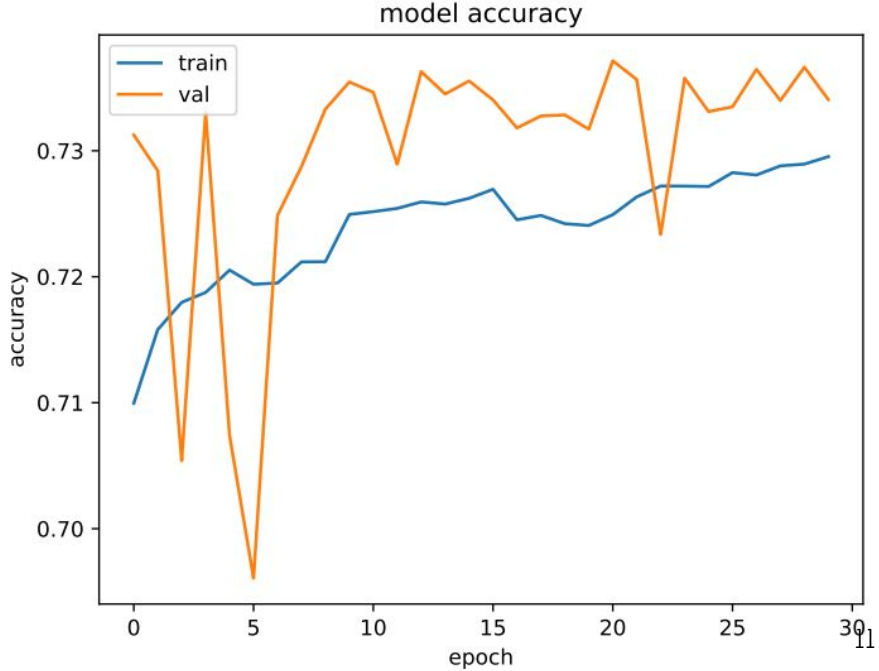
Probability of large-R jet to be Higgs



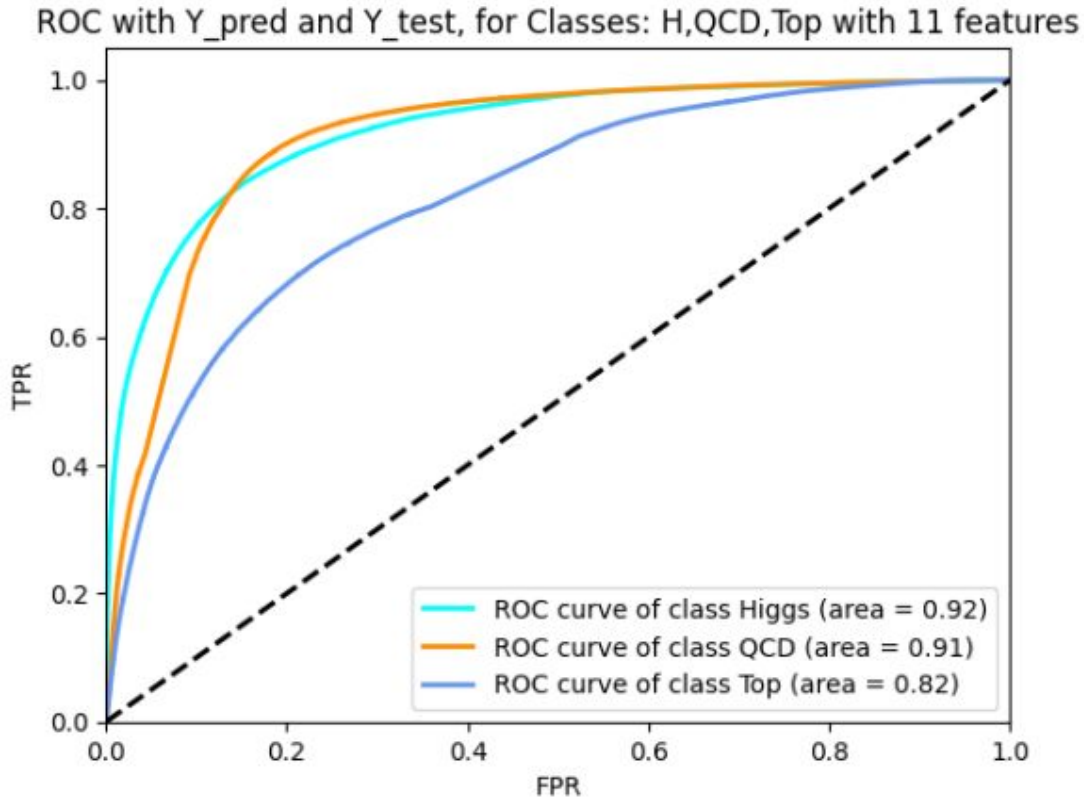
Proceed for NN

We finally feed our NN!

Considering 30 epochs and 11 features



ROC curves

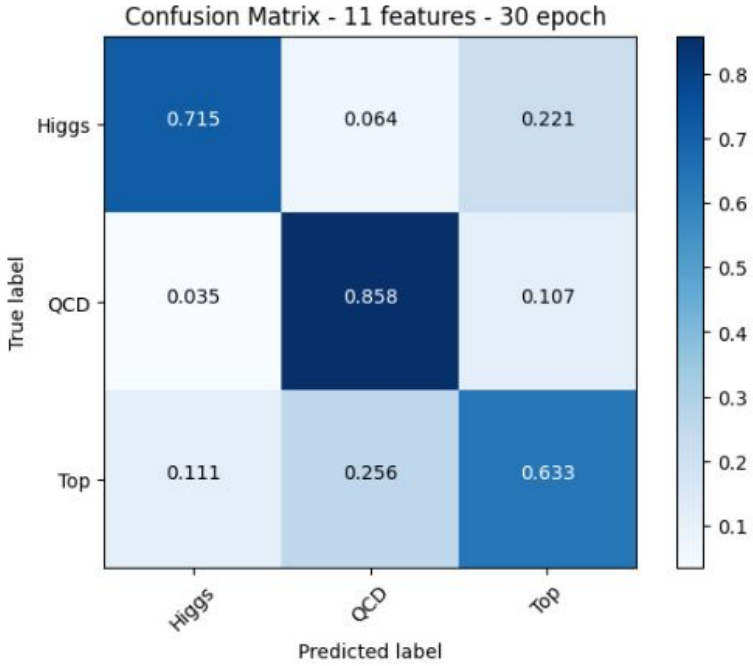


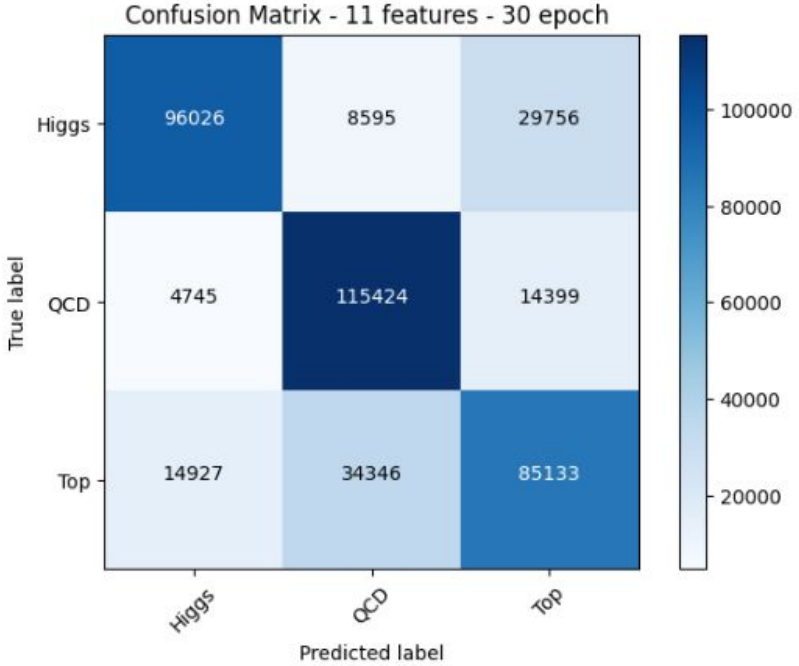
True Positive Rate vs False Positive Rate?

Ideal performance of the architecture?

Area is an indicator of the model's ability to distinguish between classes.

Confusion Matrices



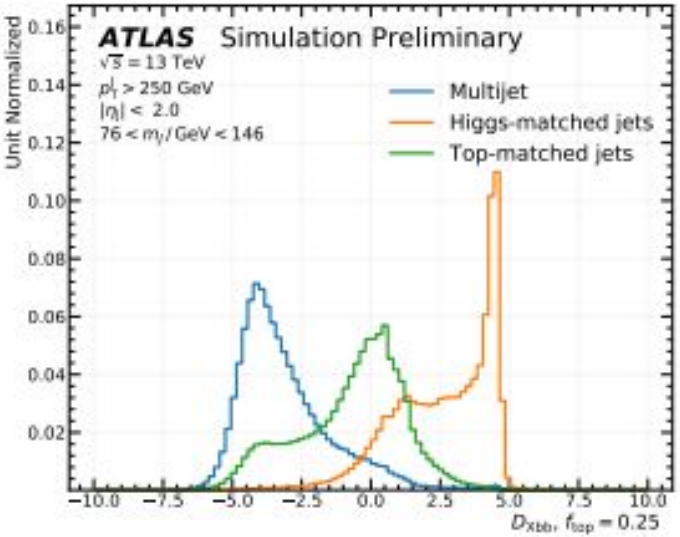


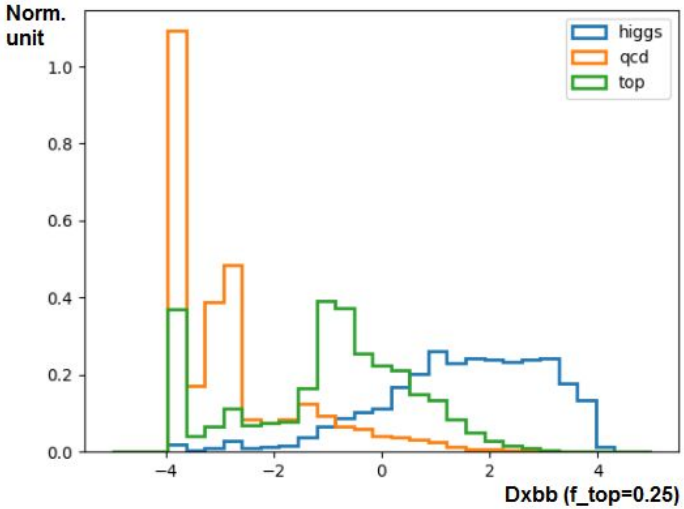
Allows the visualization of the performance of the algorithm, by comparing the jet class predicted by the NN and the real one.

Discrimination variable

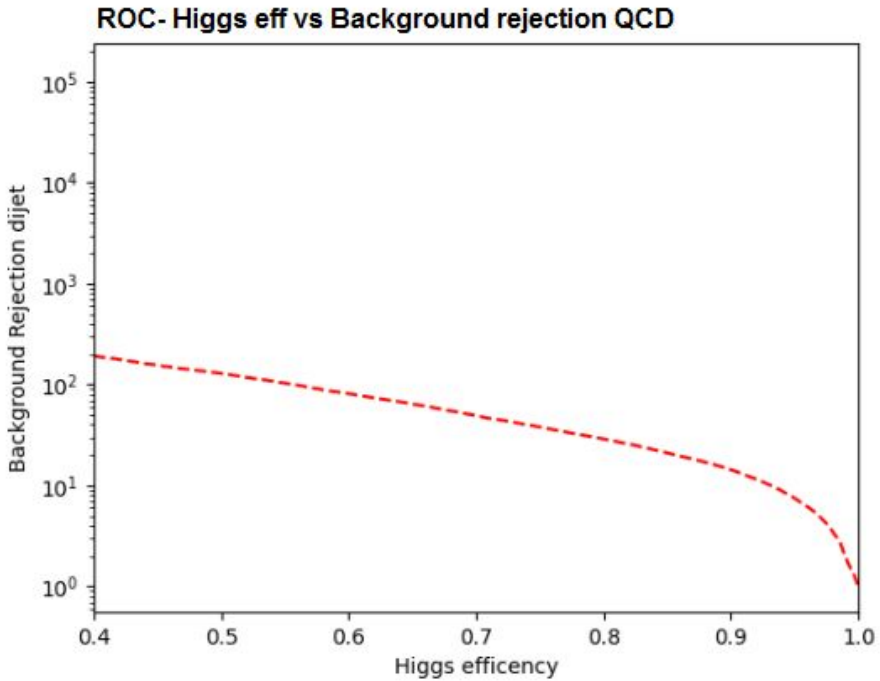
$$D_{\mathrm{Xbb}} = \ln \frac{p_{\mathrm{Higgs}}}{f_{\mathrm{top}} \cdot p_{\mathrm{top}} + (1 - f_{\mathrm{top}}) \cdot p_{\mathrm{multijet}}}$$

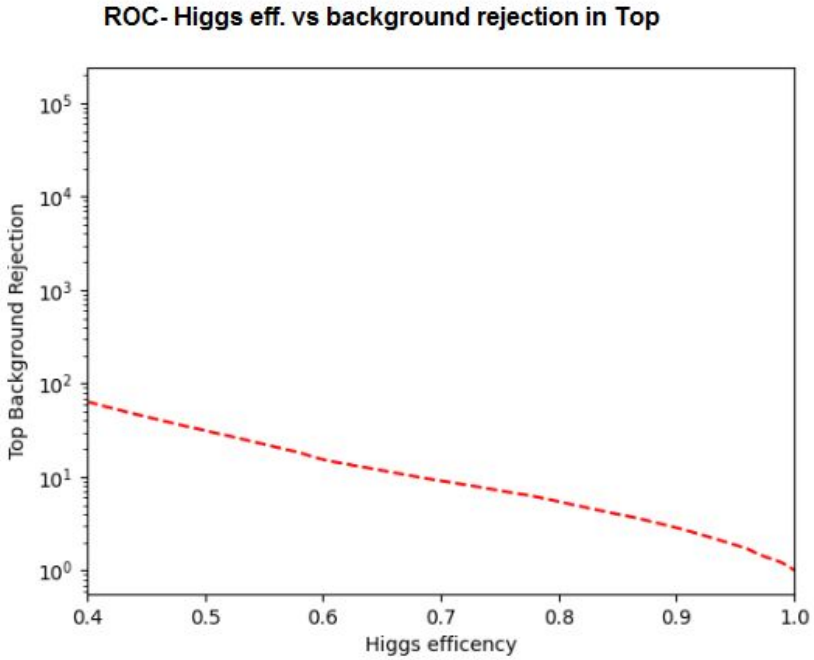
-Used to discriminate between classes
-f_{top}= 0.25
pHiggs, ptop,pmultijet predicted output (probability added to 1) from the NN, for the 3 classes.





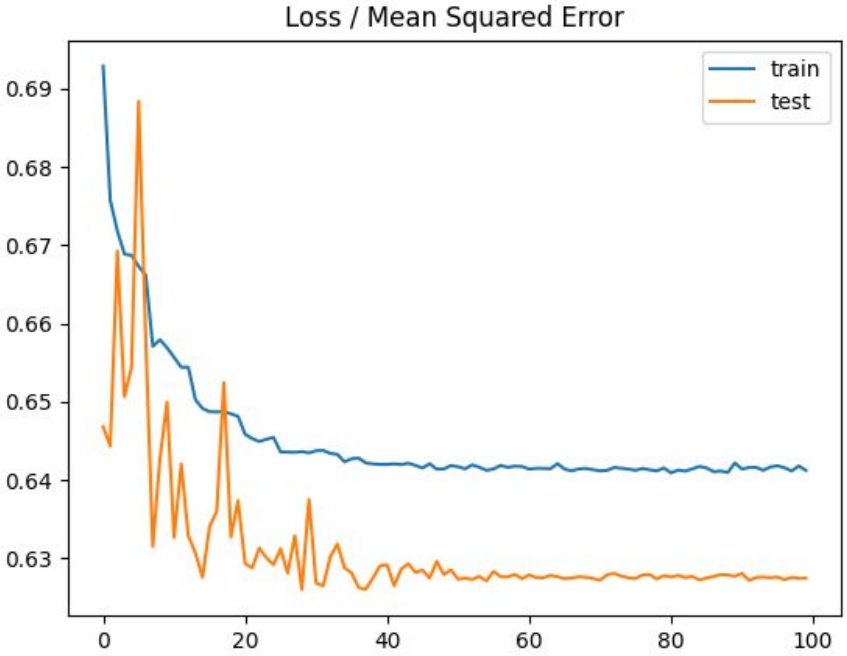
ROC but for Signal eff. and background

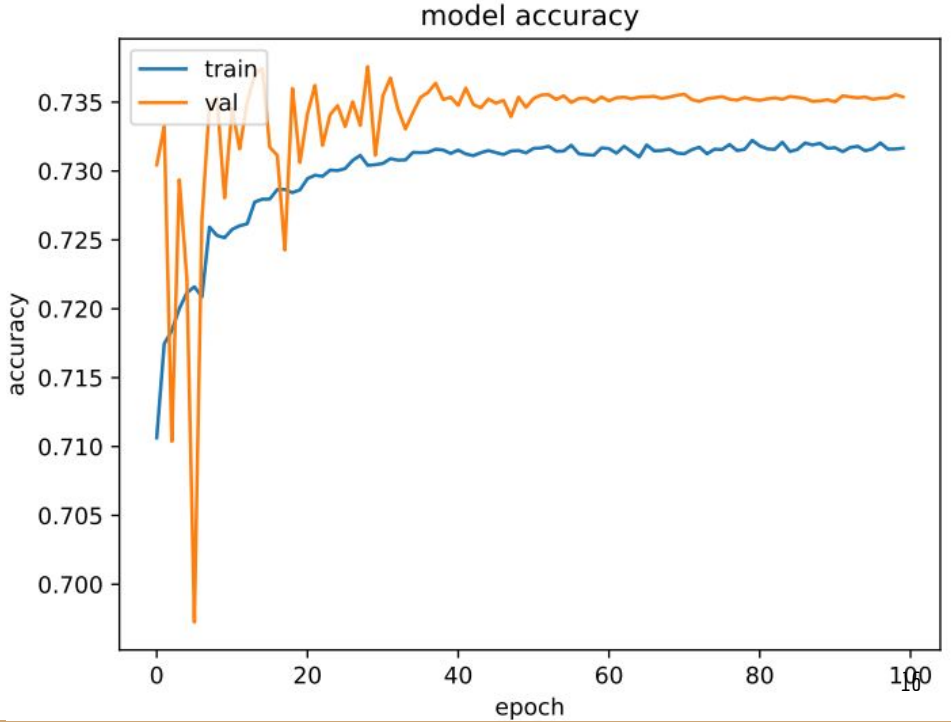




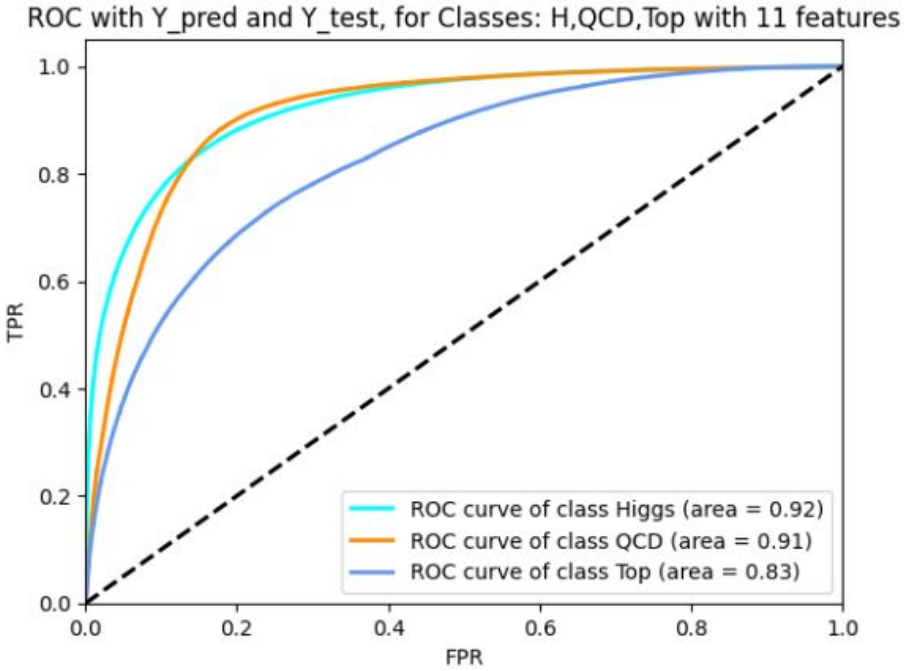
We do the same, but better...

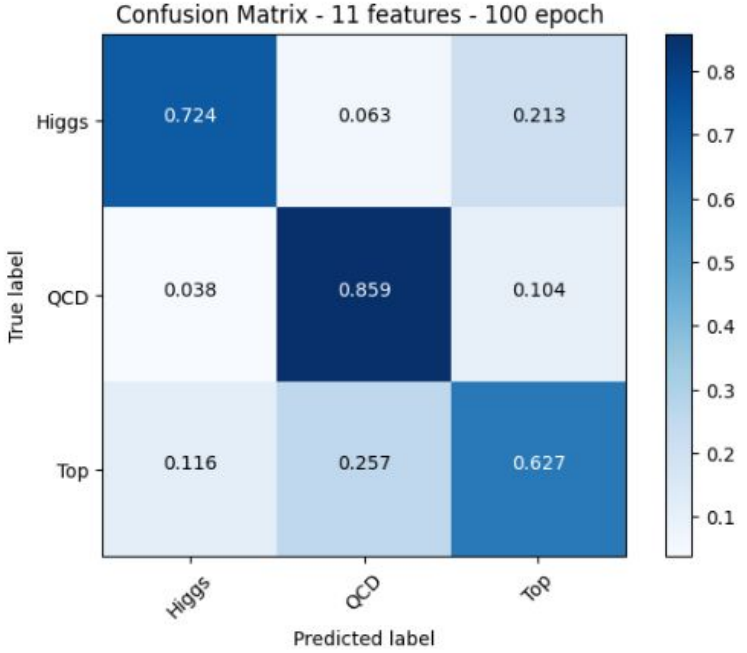
100 Epochs with 11 features:





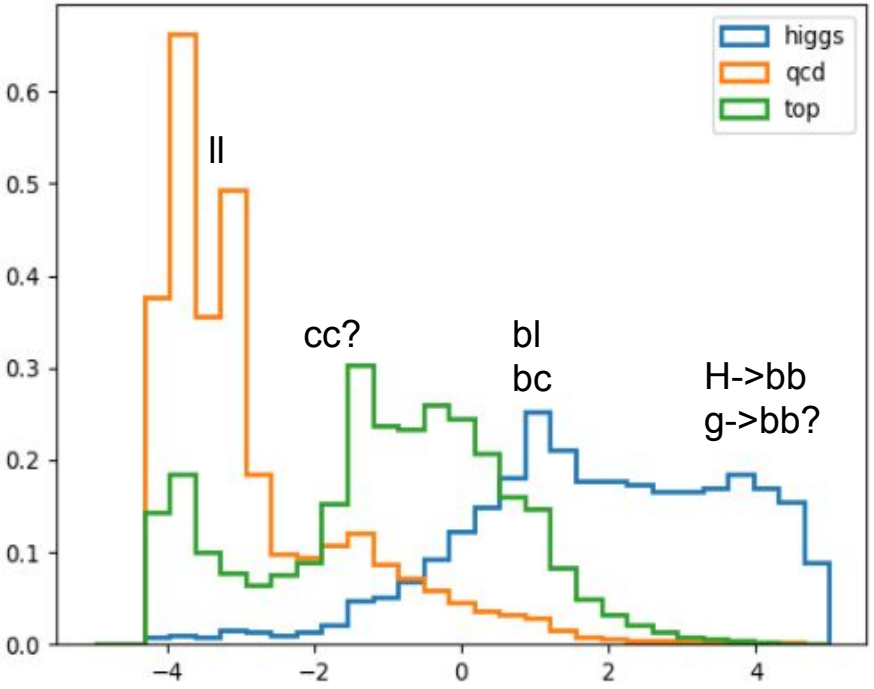
ROC (and roll) with some Confusion

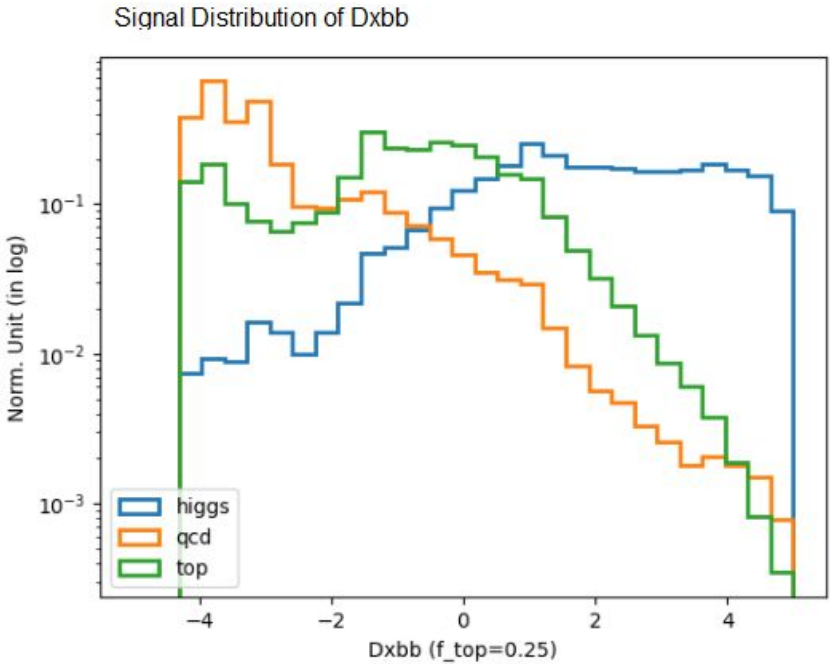




-Top the most interchangeable
-Most complex (it has W's) and it intersects the range of QCD and Higgs

DXbb Distribution for 100 epoch



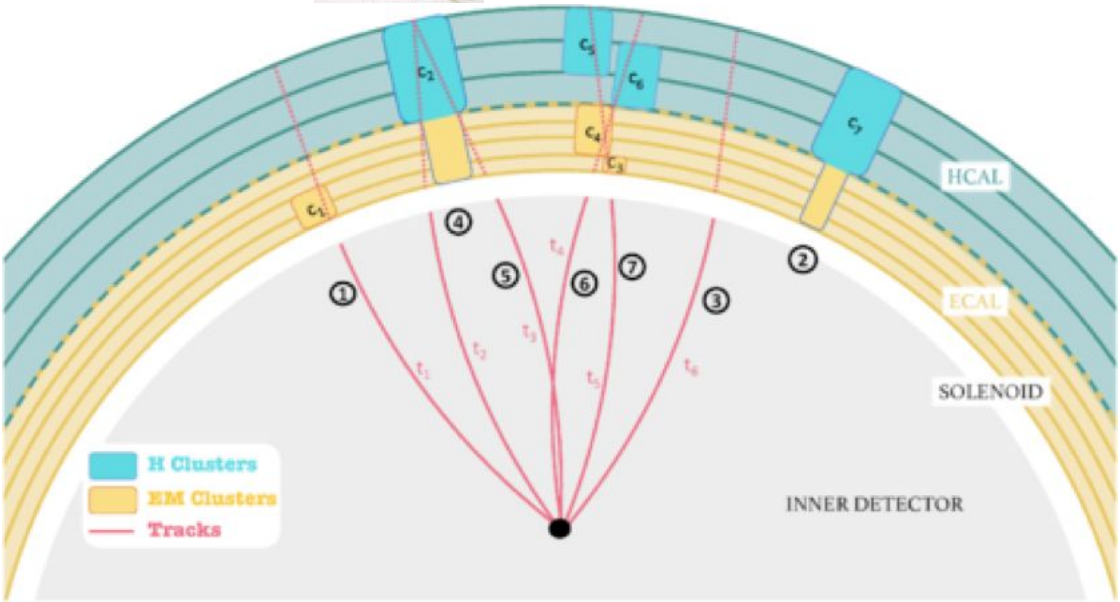


-The Intersection is visible here
-We could also study the performance vs flavour composition of the jets in the simulation.

Substructure Variables (adding )

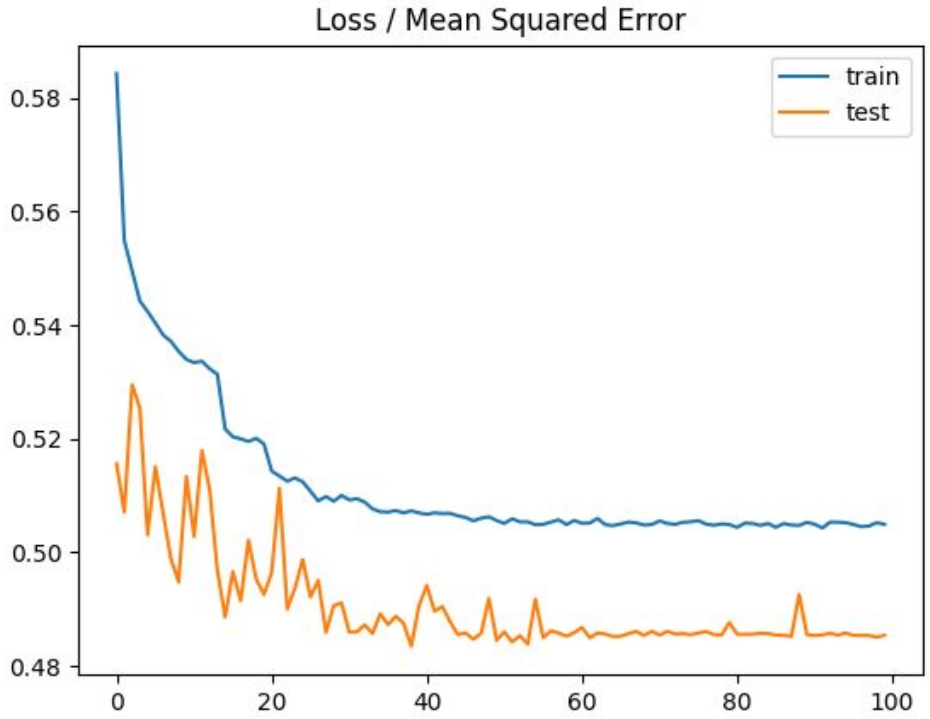
Adding to the information given from the b-tagging, we also could have additional knowledge of the structure of the jets via the calorimeter and the tracks.

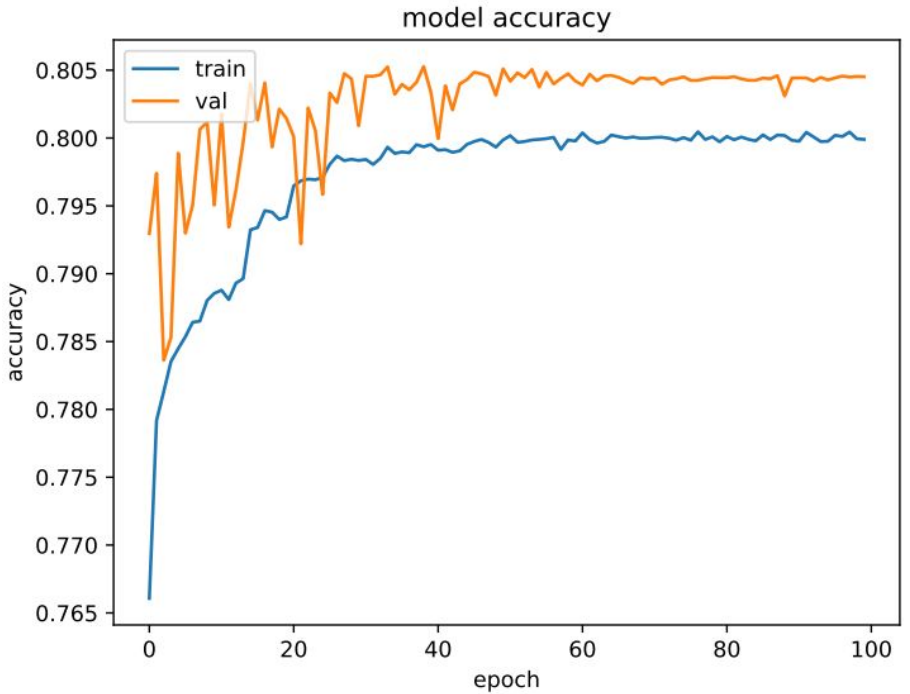
**This could be used to characterize better our system,
among Higgs, QCD and Top**



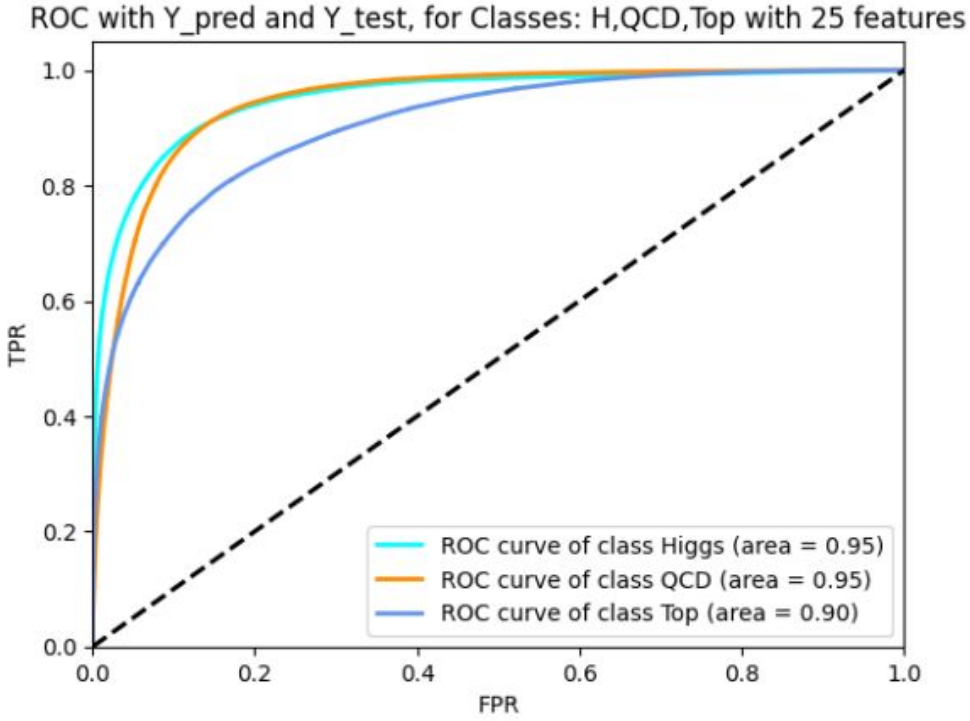
**Adding a total of 14 additional variables
for example “split12” that uses symmetry to distinguish asymmetric QCD splittings from heavy symmetric particle decays**

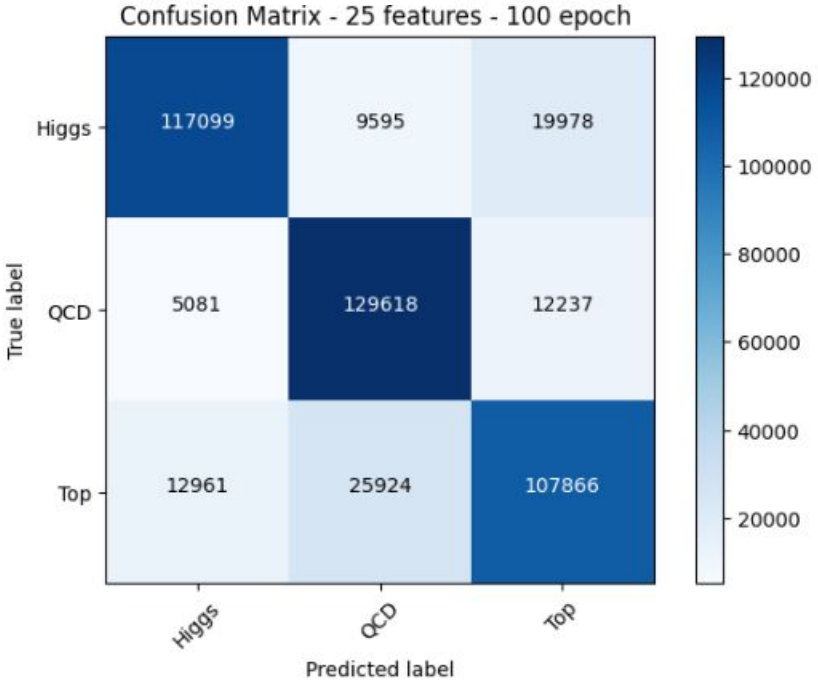
Now 100 epochs with 25 features



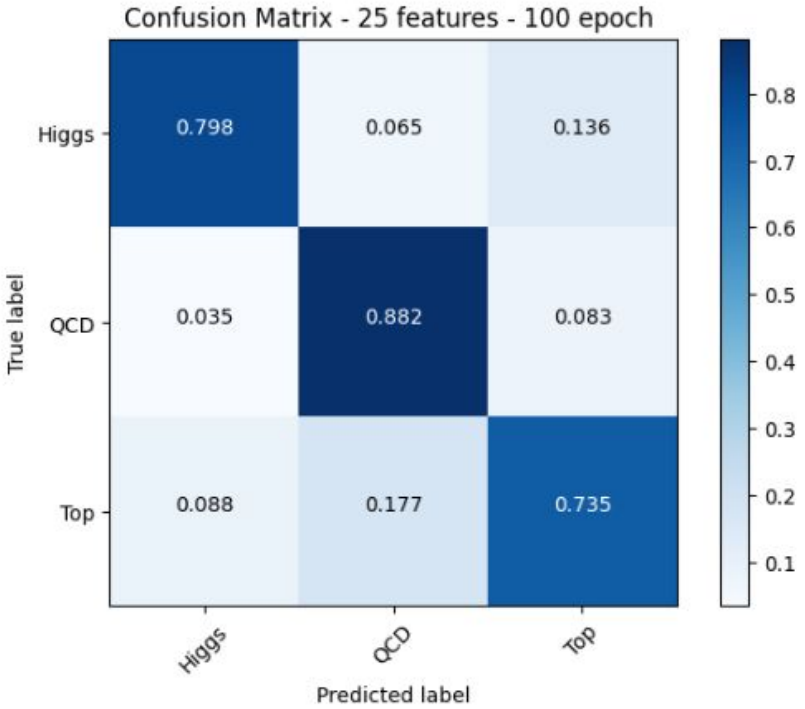


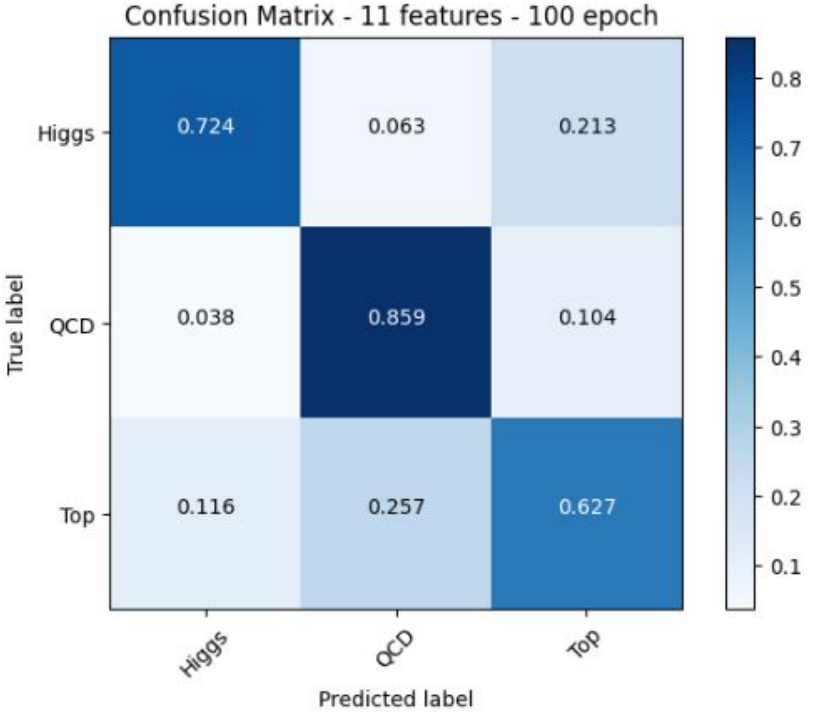
A new ROC and total CM



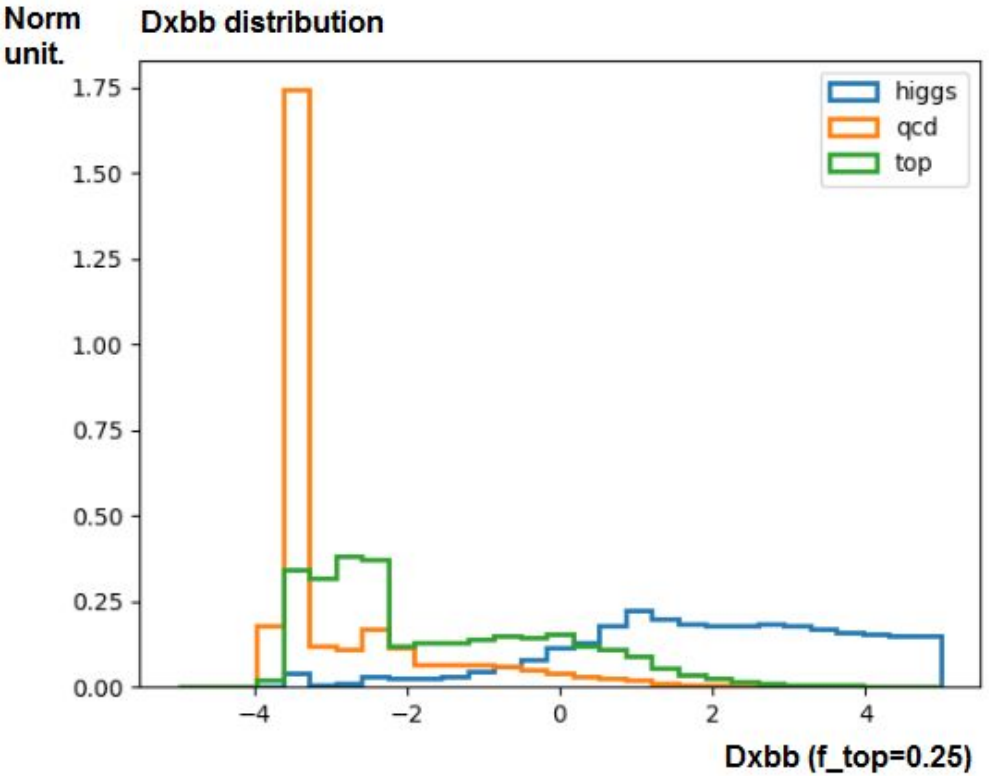


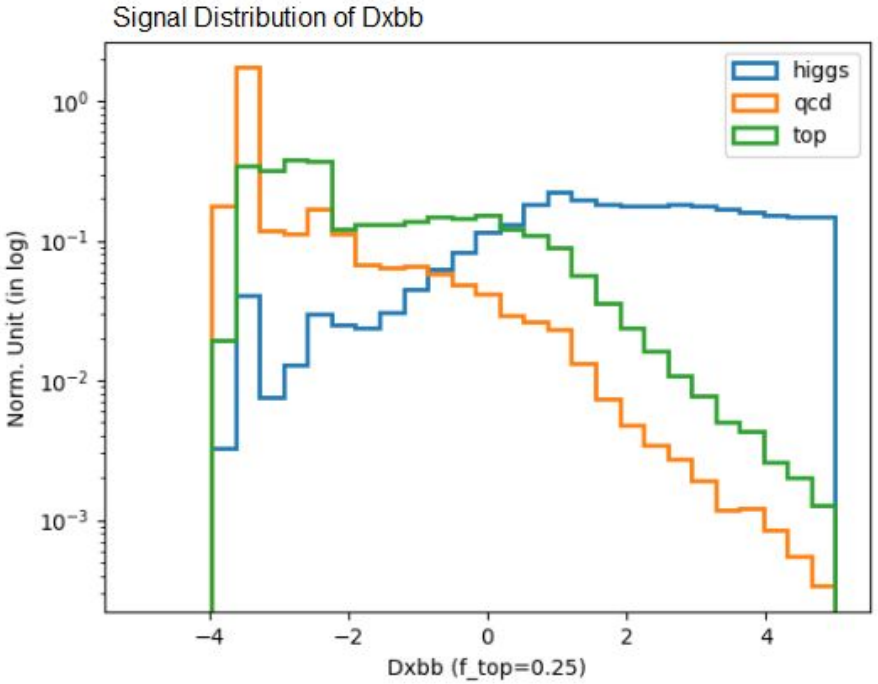
Comparing Confusion matrices



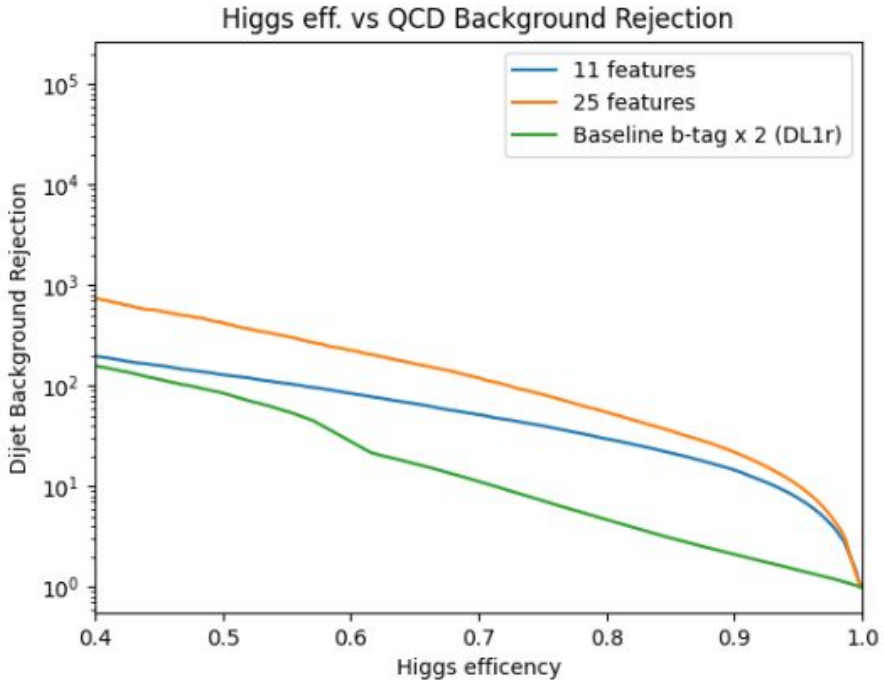


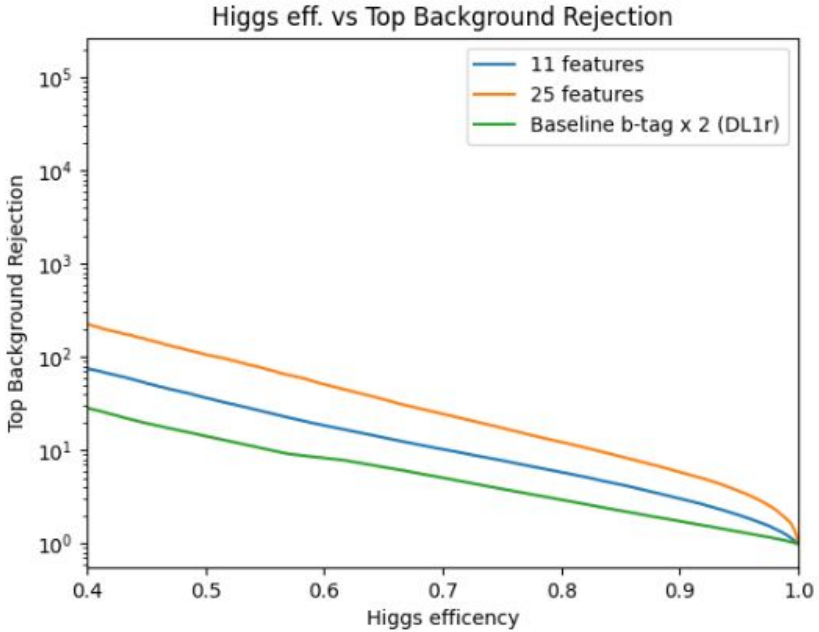
Dxbb distribution with 25 features





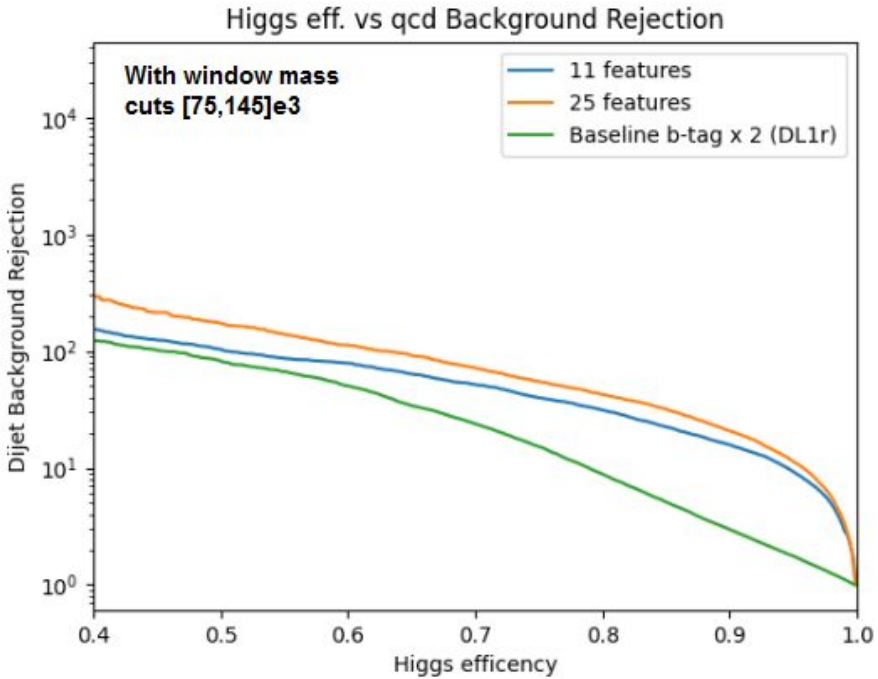
ROC with signal eff. and background

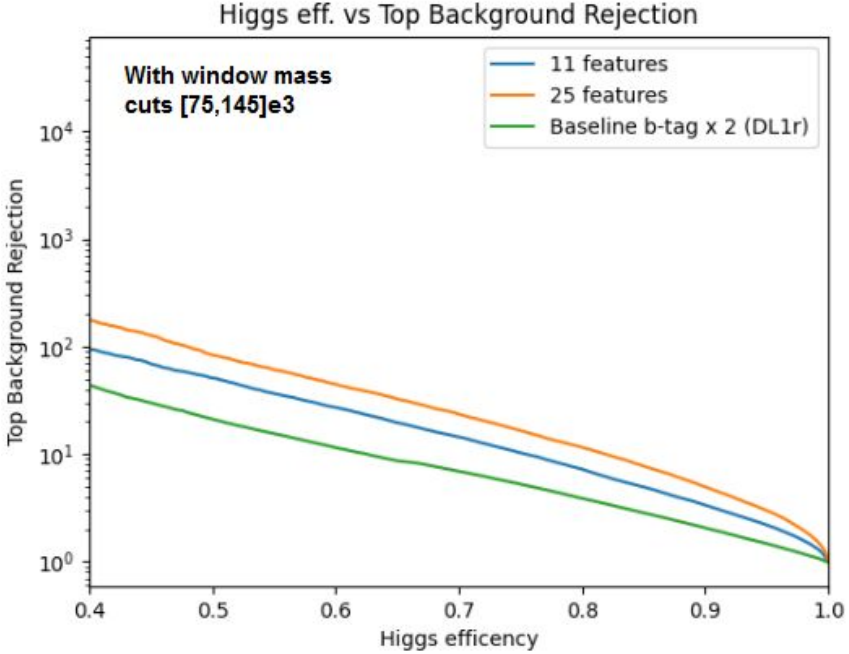




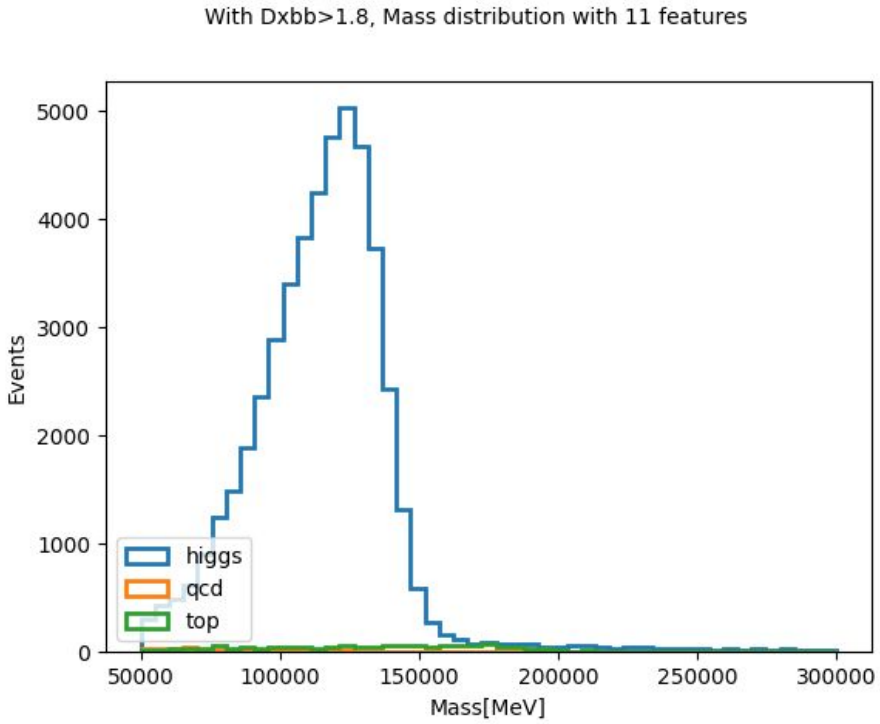
ROC with signal eff. and background

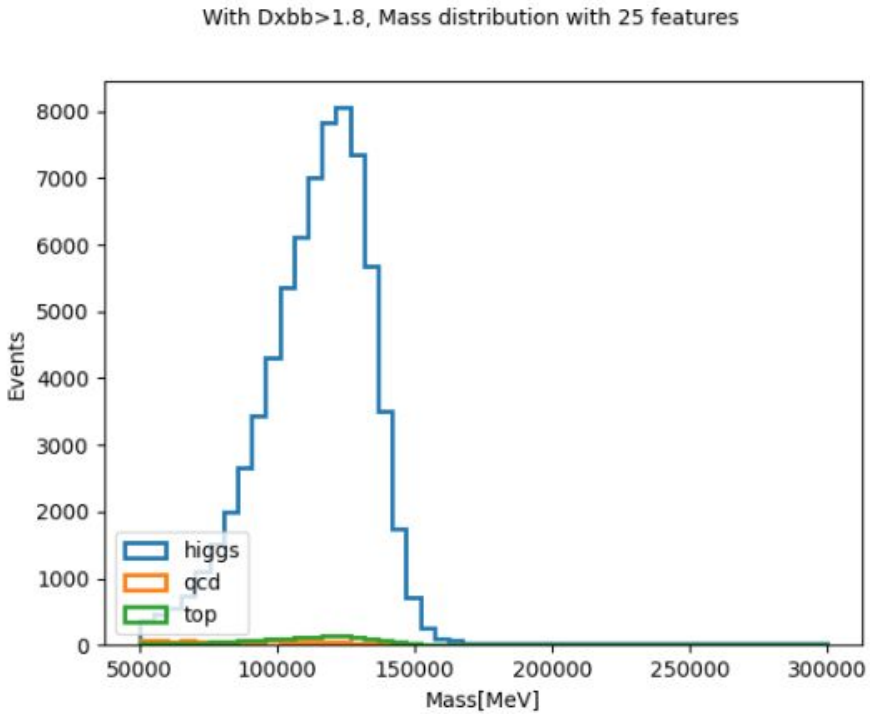
With mass cuts in the range of [75, 145] GeV to define Higgs boson mass region



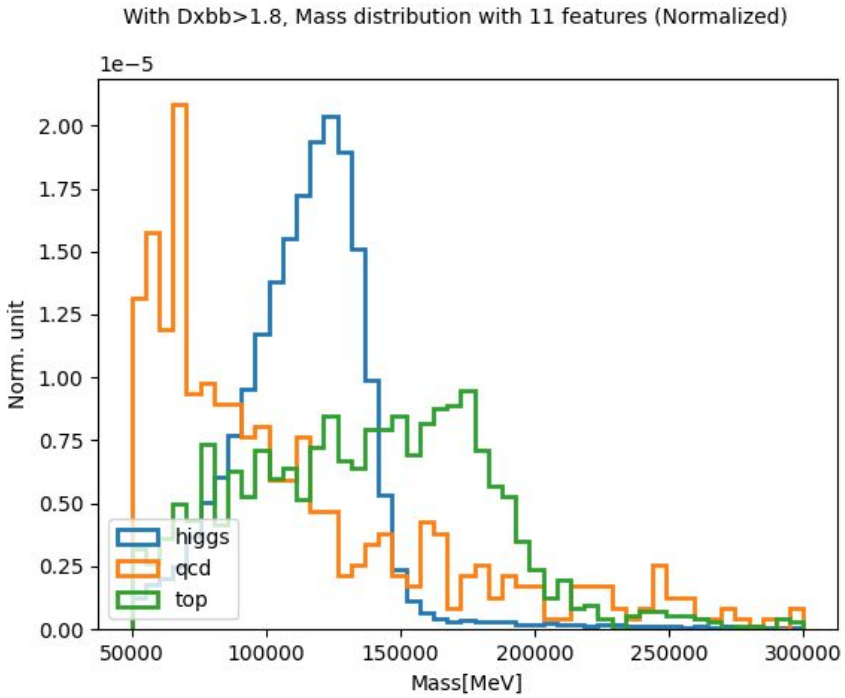


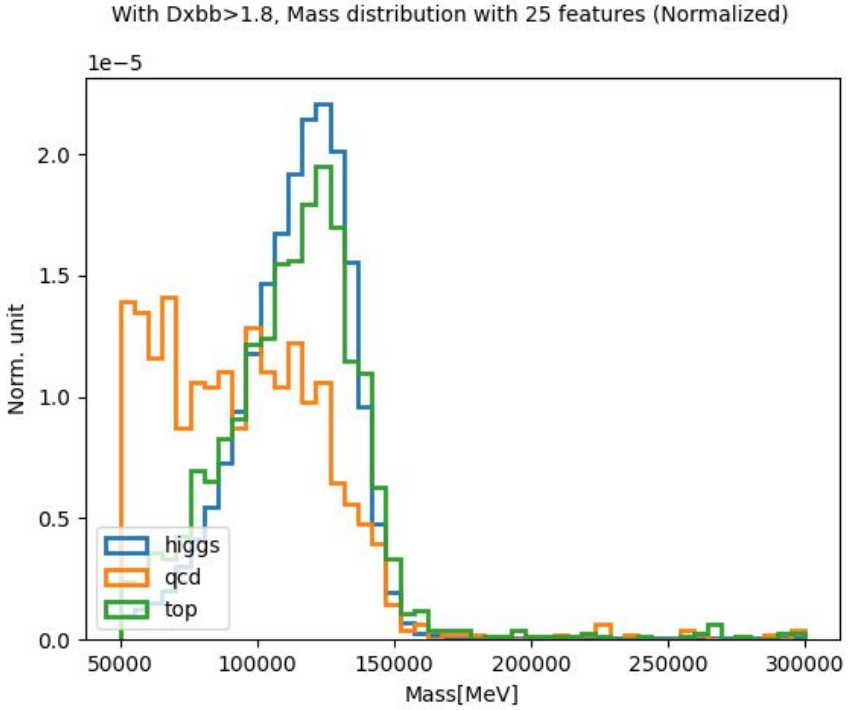
Mass distribution, after Dxbb›1.8





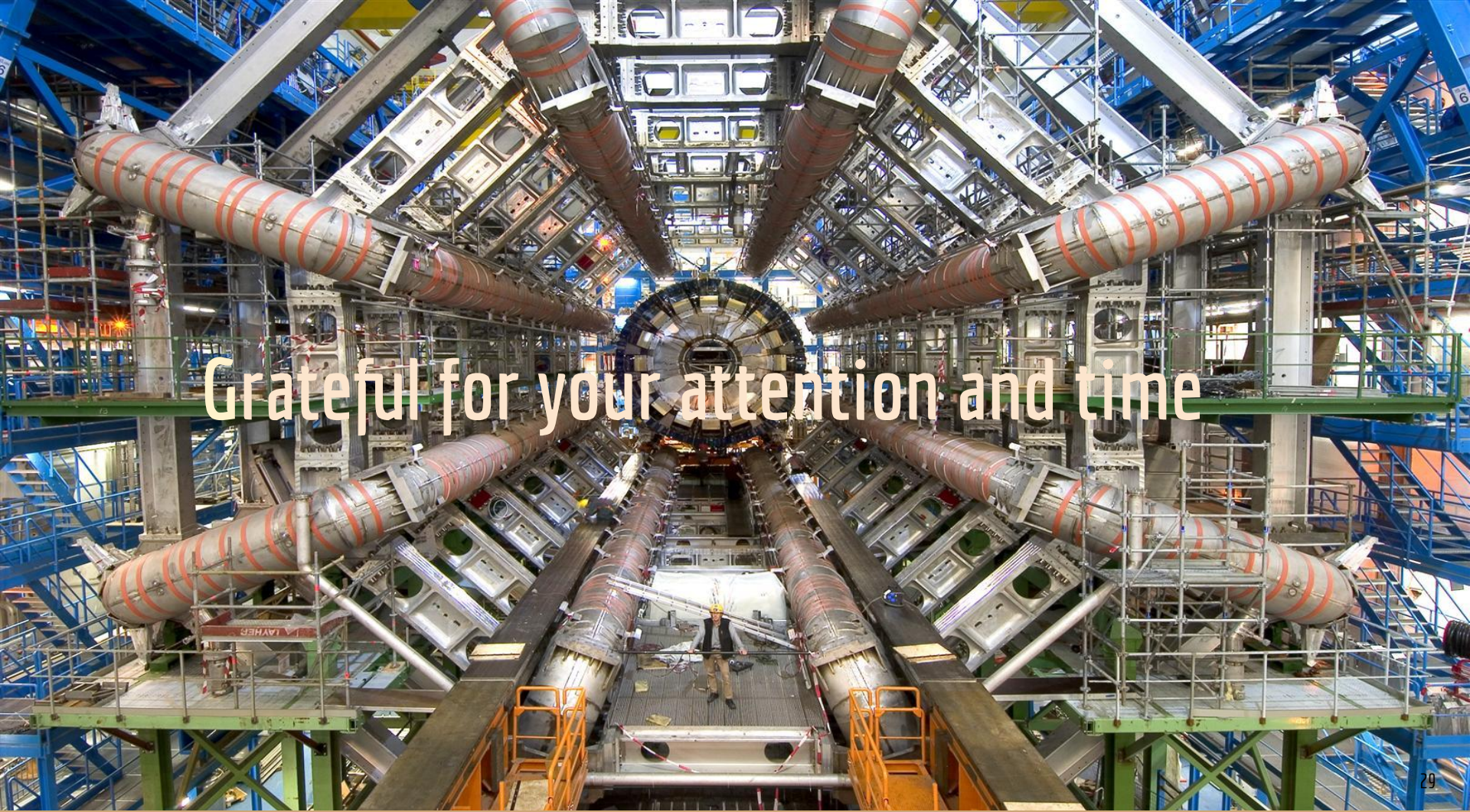
Something interesting appears...





Conclusions:

- Looking at the Confusion matrices, we identified that the **Top background was the most difficult to filter through** by being the most complex and by intersecting the ranges of Higgs and QCD.
- With the b-tagging variables alone, the **background does not peak in the Higgs mass region** after the cut in the DXbb distribution.
- The comparison between 11 features and 11+ substructure variables shows an increase in background rejection **up to factor of 2**.
- However with the substructure we can also **distort the background jet mass distribution**. As a next step this could also be corrected via a NN technique.


Grateful for your attention and time