LIP-STUDENTS-20-04

Machine learning methods to improve boosted Higgs boson tagging at ATLAS

Galetsky Vladlen[1,a]

[1] *Instituto Superior Técnico, Lisboa, Portugal*

Project supervisor: I. Ochoa

October 2020

Abstract. Identifying Higgs bosons decaying via the dominant H → bb⁻ mode is an essential ingredient to many searches for new physics happening at LHC. ATLAS reconstructs hadronic Higgs bosons by using a large-radius jet to which smaller variable-radius subjets are associated and assigned with flavor discriminants using standard flavor tagging (b-tagging). This work introduces an algorithm to select Higgs bosons in the high-momentum regime that combines flavor discriminants from up to three subjets using a feed-forward neural network. The algorithm performance is studied in terms of its accuracy to discriminate between jets originating from Higgs decays, top decays or QCD processes. The background rejection is also studied after adding substructure variables.

KEYWORDS: ATLAS, Neural Networks, b-tagging

1 Introduction

Many Beyond the Standard Model scenarios introduce new heavy resonances which can decay to Higgs bosons with high transverse momentum (p_T). Therefore precise reconstruction techniques for high-momentum (boosted) Higgs boson decays in which the two b-quarks are highly collimated are crucial for improving the sensitivity of those searches. The Higgs bosons that decay into pairs of bottom quarks have a high branching ratio of about 0.6. As the Higgs boson decays quickly we can only measure the respective products of the decay, where we use the reconstruction of the large radius (large-R) jets with the intention of capturing the entire process, while smaller variable-radius jets (VR subjets) are used to resolve and identify the individual b-hadrons. A neural network (NN) is introduced to solve the H → bb⁻ tagging problem, taking advantage of the strong tagging performance of the b-tagging algorithm in order to discriminate the large-R jets according to the processes originating them: Higgs to bb decays, QCD processes, or hadronic top-quark decays.

2 Methods

2.1 b-tagging method

The b-tagging algorithms take advantage of the fact that b-hadrons have a significant lifetime, they fragment and hadronize into B-hadrons providing a measureable displaced secondary vertex, as seen in Figure 1, where we show the H->bb decay capture by a large-R jet with individual b-jets resolved by variable-radius jets. Using this information allows us to identify jets containing b-hadrons.

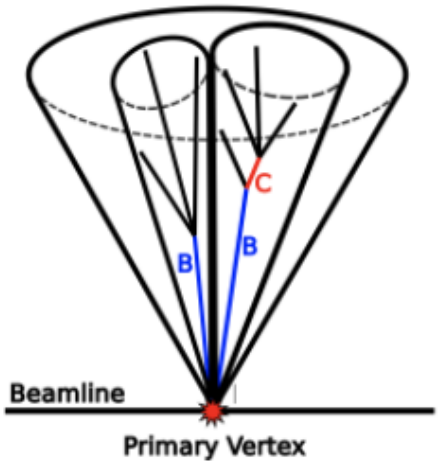

Figure 1. b-hadron long lifetime is present in the blue segments of the figure [3].

2.2 Neural Network

As described before for our tagging mechanism we will use a Deep Neural Network, the main characteristics that define it are described below:

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

2.2.1 Features

The data of our jets needs to be classified by our NN, in order to discriminate each jet and to categorize it towards the

[a] e-mail: vladlen.galetsky@tecnico.ulisboa.pt

available classes we introduce a subset of variables called features that can be used as good predictors for our model. In total we use 11 features for our input, we can identify them below:

- p_T: Transverse momentum of the large-R jets;
- η: Pseudo-rapidity of the large-R jets;
- p_u (3x): Probability distribution of a VR-subjet being light;
- p_b (3x): Probability distribution of a VR-subjet being bottom;
- p_c (3x) : Probability distribution of a VR-subjet being a charm;

Here p_T is the transverse momentum of the large-R jets. In our model we consider the boosted regime (high p_T), which implies a more difficult recognition of the subjets which become collimated and may overlap in the laboratory frame, this can be seen in the Figure below:

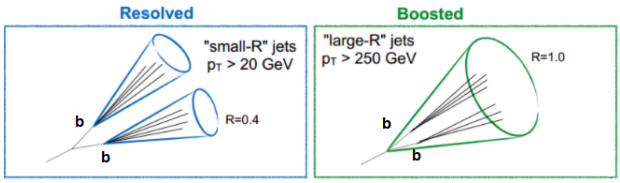

Figure 2. Resolved and Boosted regime of the subjets.

η is related to the angle that a particle has relative to the beam axis, p_u, p_b and p_c are the probability distributions of the VR-subjet originating from a light, bottom or charm quark. As we are dealing with up to 3 VR subjets in our model, we have in total 11 features as discriminant variables.

2.2.2 Classes

From features described before, our jets will be distributed towards the 3 available classes: QCD, Higgs and Top. Looking at the number of events in terms of the mass distribution we can understand better the physics behind them:

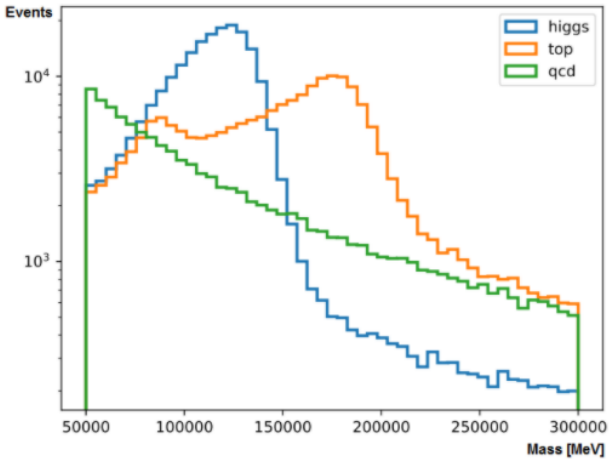

Figure 3. Large-R jet mass distribution of Top, Higgs and QCD jets.

We observe a smoothly falling distribution for QCD and a resonant peak for the Higgs (125 GeV) and two distinct peaks for Top, one corresponding to the W-boson (80.37 GeV) and the other to the top-quark (172 GeV), this due to the top decaying into b-hadron and a W-boson where consequently the W-boson decays into two quarks. For high enough transverse momentum the top-quarks may be combined into a single jet, which corresponds to the top-quark peak, if it is not the case the W-boson can be also combined into a single jet leaving out the b-hadron, which corresponds to the W-boson peak.

The invariant mass in our model will not be considered as a feature, not entering the NN. It will be used to characterize and create our fits for the model externally.

3 Results and Discussion

3.1 Validation Plots

Before training the NN, we first need to study and validate our data to know what are the respective orders of magnitude and the corresponding behaviour for our features, it also needs to be weighted in order for it to be physical, so we plot the corresponding number of events in terms of all features shown before:

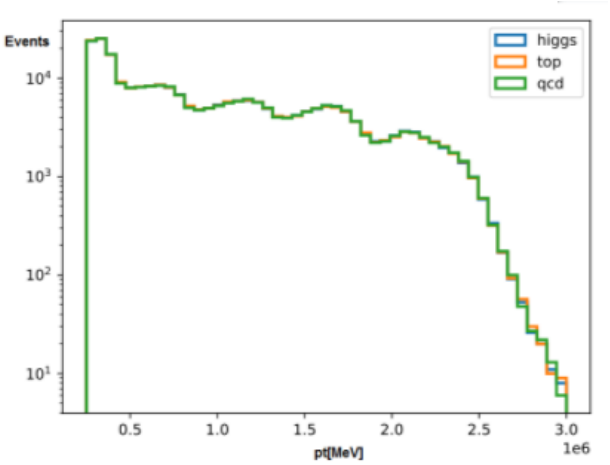

Figure 4. p_T-Transverse momentum of large-radius jets.

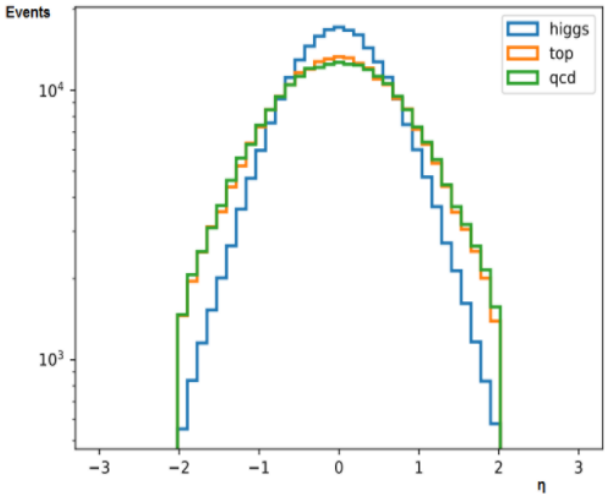

Figure 5. η- Pseudo-rapidity of large-radius jets.

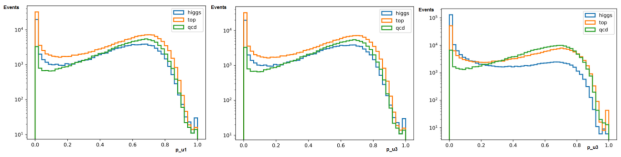

Figure 6. Probability of three VR track-jets to be light, for large-R jets in Higgs,Top and QCD.

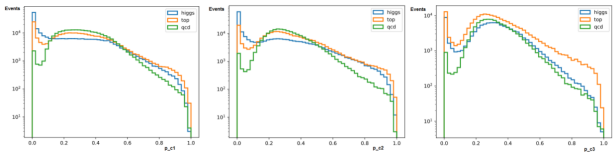

Figure 7. Probability of three VR track-jets to be Charm, for large-R jets in Higgs,Top and QCD.

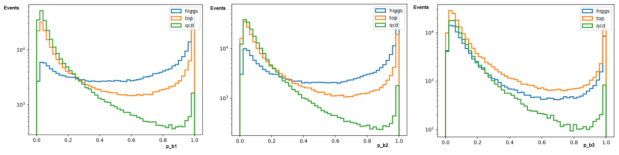

Figure 8. Probability of three VR track-jets to be Bottom, for large-R jets in Higgs,Top and QCD.

Additionally we compared this data with the one obtained from the figures 18-21 from [1] recreating and reproducing the probabilities of the VR-trackjets and large-R jet to be QCD, Higgs or Top.

3.2 NN results

With our data validated, we divide it into train and test samples, and we start training our NN for 100 Epochs with the corresponding features. We then check our model predictions that categorizes into the corresponding classes, giving us a normalized probability.

Below we show the loss function and the model accuracy for our NN along the 100 epochs for the train and test samples:

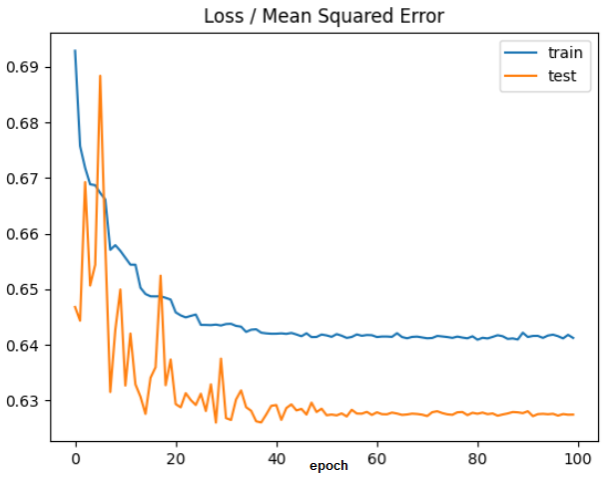

Figure 9. Loss function in terms of the mean squared error along 100 epochs.

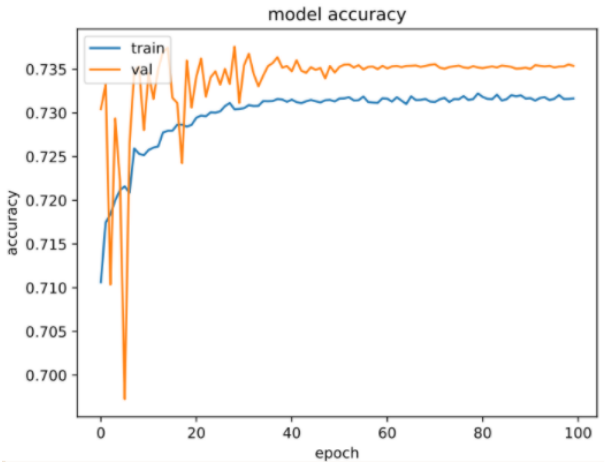

Figure 10. Accuracy of the model along 100 epochs.

These functions represent how well the model describes and categorizes the data, as the training progresses. As we can see, we could have practically the same results if the training/testing phase was done for 50 epochs.

We now plot the ROC curve that quantifies how well the model identifies the signal and rejects the background jets. This is studied by comparing the True Positive Rate and the False Positive Rate. The area below each curve gives an indication towards the performance of our model, ideally it should converge on a top-left corner of the plot if our model predicts correctly all the data.

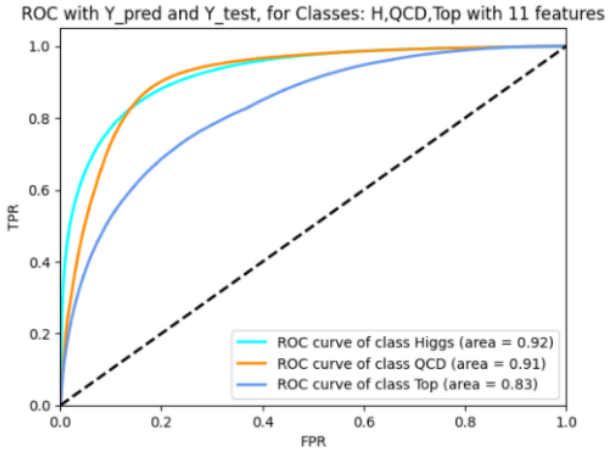

Figure 11. ROC curve for model with 11 features, for Higgs, QCD and Top.

To allow a better visualization of the performance of the algorithm, we can compare the jet class predicted by the NN and the true one by creating a normalized Confusion Matrix as shown down below. In an ideal case, we should have only the diagonal occupied.

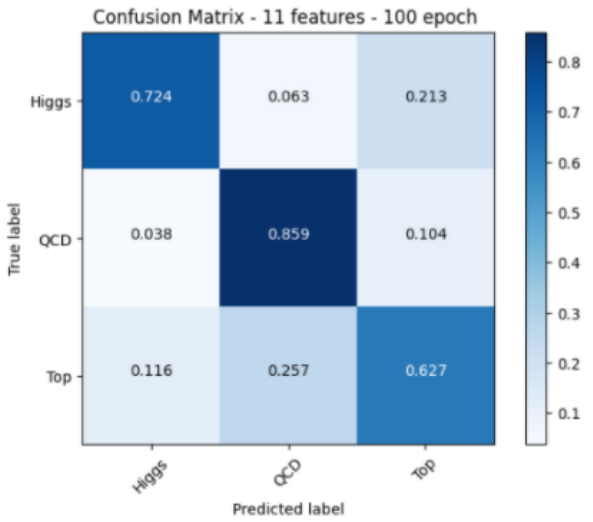

Figure 12. Normalized Confusion Matrix for model with 11 features, for Higgs, QCD and Top.

From here and from the ROC curve we can visualize the difficulty that the algorithm has to predict correctly the jet class of Top compared to the other 2 classes, this due the complexity of the processes occurring in that class as referred previously, intersecting both, the Higgs and QCD.

3.2.1 Discrimination Value

The three-class output can be combined into a single discriminant parameter described as:

$$D_{X_{bb}} = ln \frac{p_{Higgs}}{f_{top}.p_{top} + (1 - f_{top}).p_{multijet}} \quad (1)$$

Here p_{Higgs}, p_{top} and $p_{multijet}$ are respectively the probability output from our NN to classify the jets into Higgs, Top and QCD. Where the numerator is the signal and the denominator is the respective background, $f_{top} = 0.25$ is an arbitrary parameter in order to define the type of background for our system.

From this we can plot in normalized units the $D_{x_{bb}}$ distribution for our system and also the respective logarithmic scale for easier identification of the peaks:

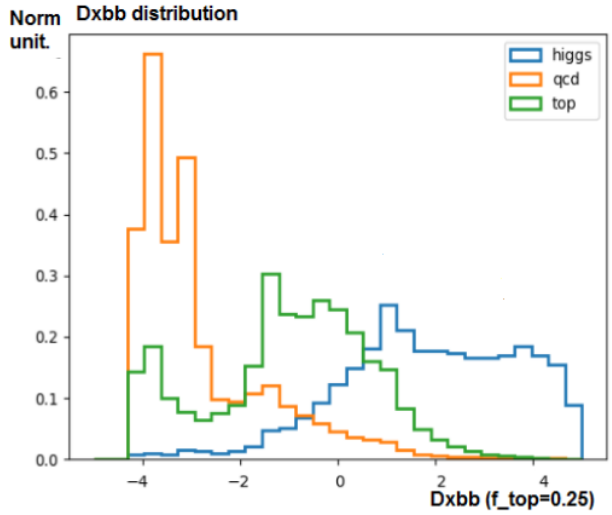

Figure 13. $D_{X_{bb}}$ distribution for 11 features for Higgs, QCD and Top.

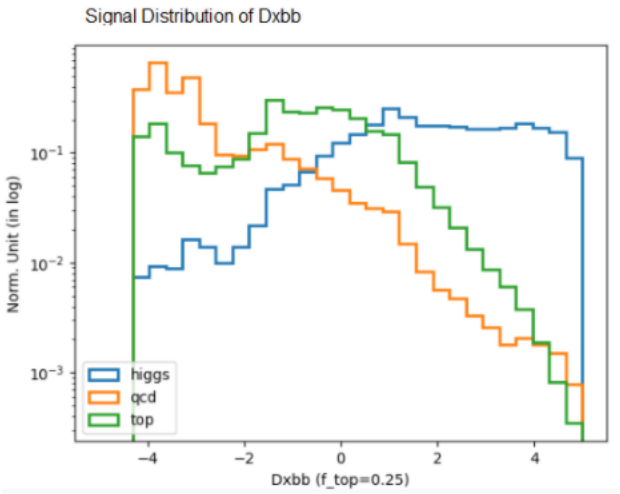

Figure 14. $D_{X_{bb}}$ distribution for 11 features for Higgs, QCD and Top in logarithmic scale.

Comparing our result with the figure 2 from [1] we identify a strong Top jet class intersection with the other two jet classes, as expected from the Confusion Matrix analysis. We also noticed a worse signal identification efficiency at high values of $D_{X_{bb}}$, identified by the missing peak structure at around 3.

Additionally we could also study the performance of the algorithm in terms of the flavour composition of the jets in the simulation, however in our case the data-sets had no information of that type and it is left for a future studies.

3.2.2 Substructure Variables

Adding to the information given from the b-tagging, we could also have additional knowledge of the structure of the jets via the calorimeter and the tracks. This could be used to characterize even better our system, among Higgs, QCD and Top. Adding a total of 14 additional variables to characterize the jets [2]:

- Split12, Split23, Zcut12: Splitting scale variables;
- Qw: Minimum pair invariant mass;
- Planar Flow, Angularity, Aplanarity: Jet shapes;
- KtDR: Kt-subjet Δ R;
- C2, D2: Energy Correlation ratios;
- e3: Energy Correlation function;
- Tau21, Tau32: N-Subjettiness;
- FoxWolfram20: Fox-Wolfram moment;

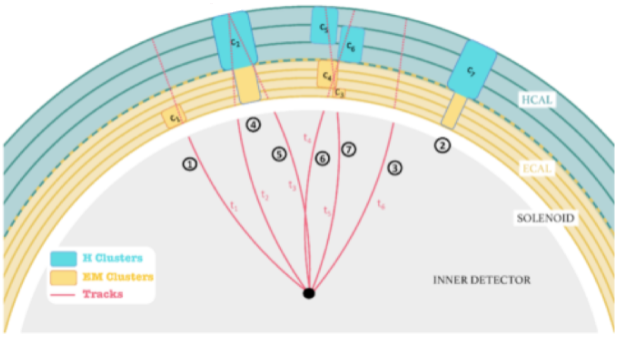

Figure 15. Tracks and Calorimeter visualization for the respective jets [4].

Meaning that we could repeat the procedure that was done with 11 features but now done with 25 in total. By training and testing our new model for 100 epochs we obtain the respective loss function and accuracy plot down below:

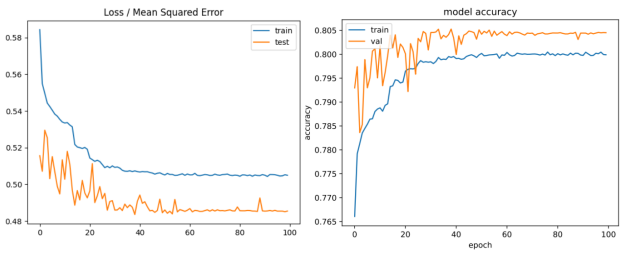

Figure 16. Loss in mean squared error and Accurracy model respectively along of 100 epochs.

Once again we define our ROC curve:

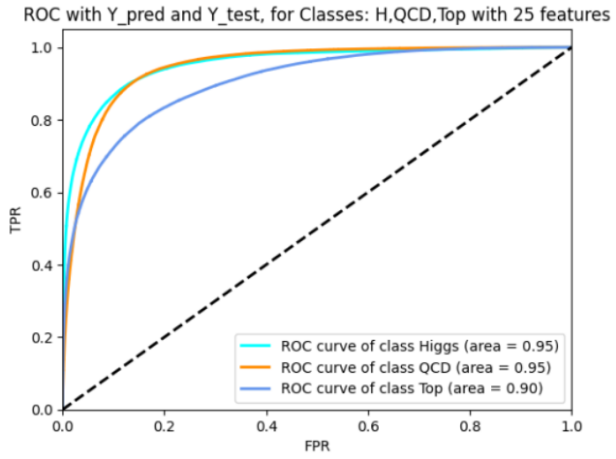

Figure 17. ROC curve with 25 features for Higgs, QCD and Top.

And now we compare the Confusion Matrices from 11 and 25 features:

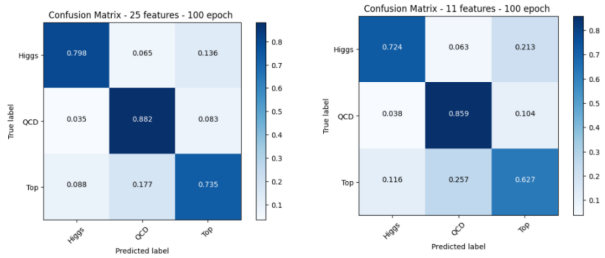

Figure 18. Confusion Matrix with 25 features for Higgs, QCD and Top.

We identify a better performance of our algorithm as expected, and this is more evident for the Top jet class identification increasing the overall probability of the correct data prediction from 0.63 to 0.74.

Also if the look at the $D_{X_{bb}}$ distribution we identify the same characteristics as for 11 features:

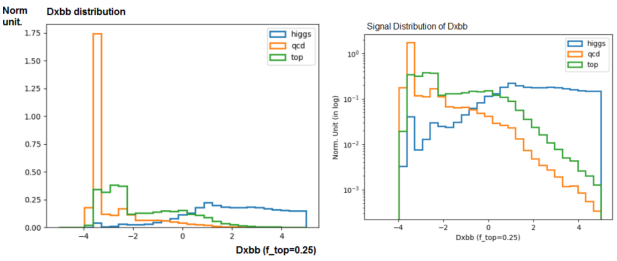

Figure 19. $D_{X_{bb}}$ distribution for 25 features, right image in logarithmic scale.

Here we are going to introduce the DL1r method (baseline) and then do a comparison between the performances of the 11 and 25 feature algorithms. The DL1r method considers the b-tagging information of the VR subjets individually, without taking into account their correlations for each large-R jet.

Here we compare the three performances using a ROC curve, where we study the Dijet or Top background rejection in terms of the Higgs efficiency of the signal. Comparing the 11 and 25 features we see an increase of the background rejection up to a factor of 2, however as the flavour information is already discriminatory, where for example the information about the number of VR track-jets is already partially included in the NN, this results into not having an even better performance of the algorithm when the 14 substructure features are added.

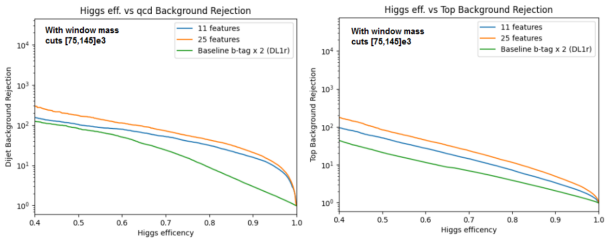

Figure 20. Background Rejection Top and Bijets in terms of the Higgs signal efficiency.

To study further the relation of adding the substructure variables we performed a mass distribution study with 11 and 25 features as shown below, having a cut in our discriminatory variable of $D_{X_{bb}} > 1.8$. This value is identified by requiring that 80% of the total signal passes through.

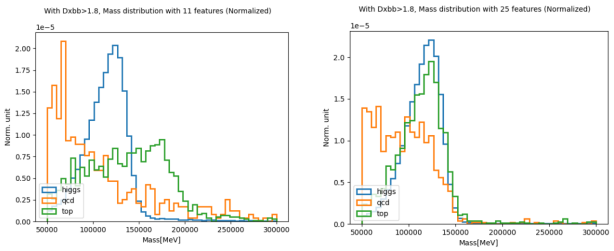

Figure 21. Mass distribution with 11 features (left) and 25 features (right).

From comparing the two distributions in Figure 21 it can clearly be seen that adding jet substructure information distorts the background jet mass distribution.

4 Conclusions

In terms of this work, by 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.

The comparison between 11 features and 25 features (where the substructure variables are added) 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 effect could be reduced for example by introducing another NN that could perform the decorrelation between the NN output and the jet mass.

5 Acknowledgements

I would like to thank my project supervisor, Inês Ochoa, whose guidance, commitment, and patience made this work possible. Also would like to thank to the Portuguese ATLAS team for the assistance given during the project.

References

[1] The ATLAS Collaboration, *"Identification of Boosted Higgs Bosons Decaying Into bb⁻ With Neural Networks and Variable Radius Subjets in ATLAS"*, https://cds.cern.ch/record/2724739/files/ATL-PHYS-PUB-2020-019.pdf

[2] The ATLAS Collaboration *"Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC"*, https://arxiv.org/abs/1808.07858

[3] The ATLAS collaboration, *"Variable Radius, Exclusive-kT, and Center-of-Mass Subjet Reconstruction for Higgs → bb⁻ Tagging in ATLAS"*, https://cds.cern.ch/record/2268678

[4] The ATLAS collaboration, *"Improving jet substructure performance in ATLAS using Track-CaloClusters"*, https://cds.cern.ch/record/2275636/files/ATL-PHYS-PUB-2017-015.pdf