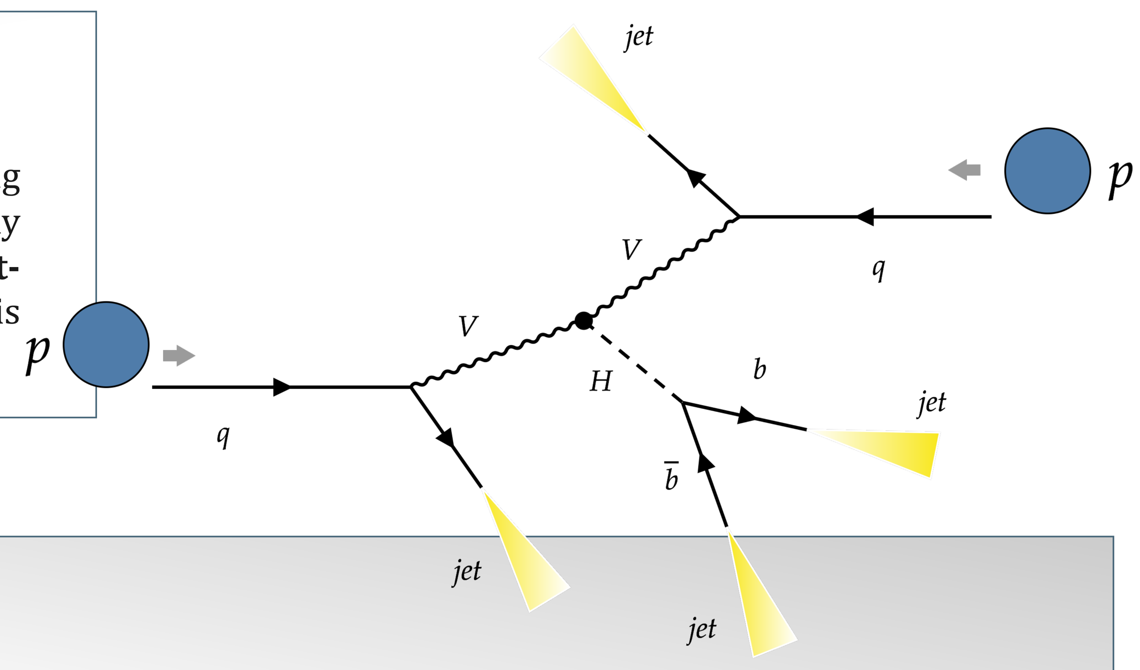
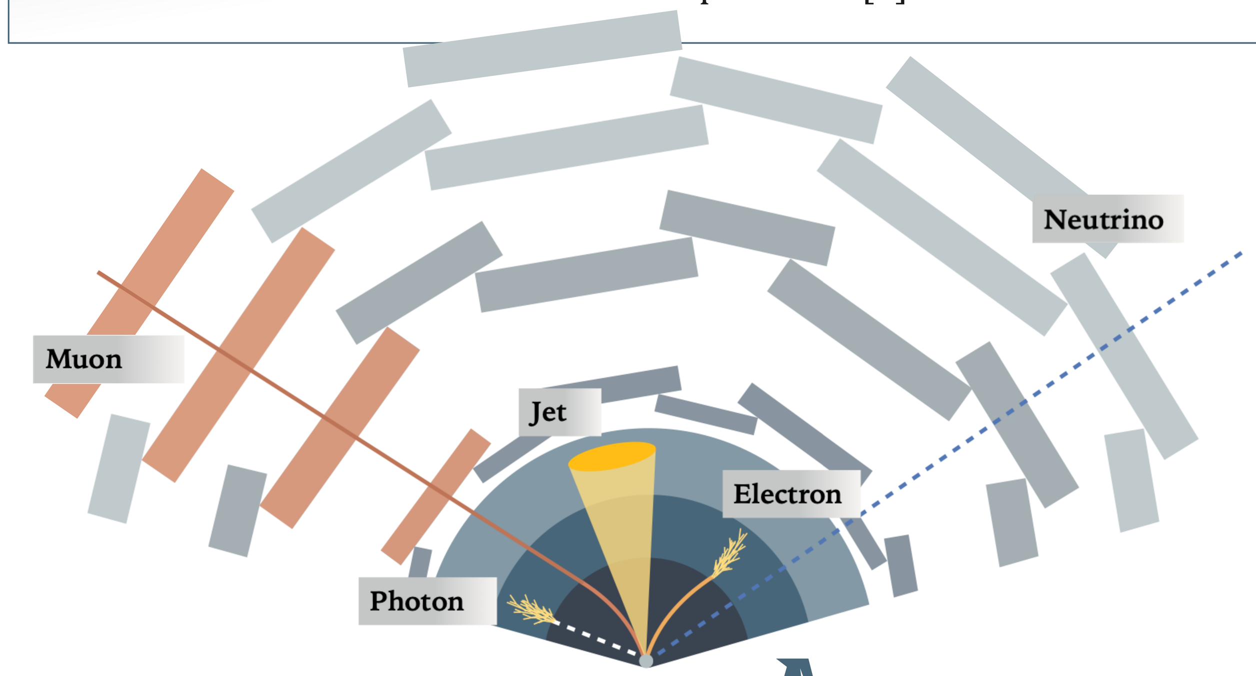
VBF production mode

The Higgs boson decaying into a quark pair is observed with most of the sensitivity associated with a vector boson V (V=W,Z) in the Higgs–strahlung production mode. Furthermore, the second most frequent Higgs boson production mechanism, the **Vector Boson Fusion (VBF)** can be exploited to study these channels. The VBF mechanism involves proton radiating weak vector bosons that fuse to form the Higgs boson. Its signature is represented by a **jet-dominated final state**: two quarks with a large rapidity gap and two b- or c-tagged jets coming from the Higgs boson decay. A VBF H → b$\bar{b}$ channel analysis for the Run 2 data with ATLAS has been provided. [1]





Higgs hadronic channel investigation

Detecting the decay of the Higgs boson into a quark pair (b$\bar{b}$ or c$\bar{c}$) is challenging due to the strong QCD background in proton-proton collision events at the Large Hadron Collider. In the dominant production mode, gluon-gluon fusion (ggF), these channels are **overwhelmed by non-resonant QCD** and **multi-jet background**. Studies of these channels are necessary to enhance the effect of **Beyond the Standard Model physics** on the Higgs coupling with quarks. **Neural networks** are used in these analyses to improve the sensitivity to signal events. Several studies on these channels are being conducted at the ATLAS and CMS experiments at the Large Hadron Collider of CERN.

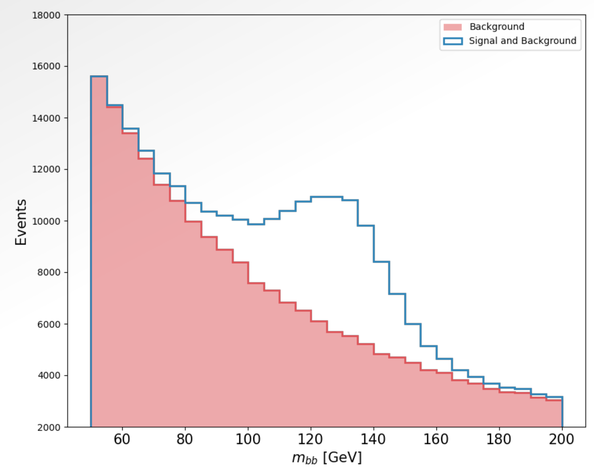
Greta Brianti[1,2]
1 University of Trento
2 TIFPA – INFN

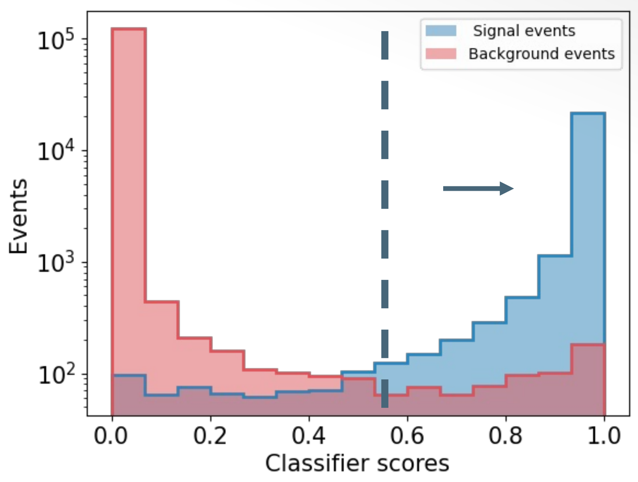
Advanced Decorrelating Techniques in DNN-Based Classification for VBF Higgs boson Analysis

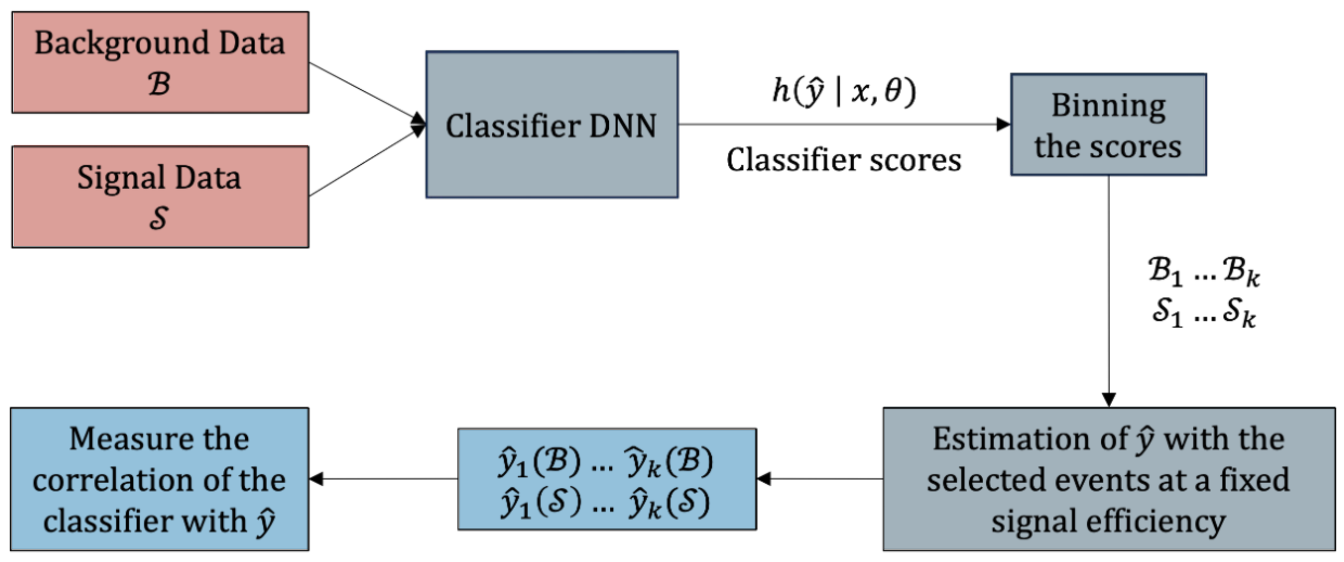
Dataset

- A dataset of VBF H → b$\bar{b}$ signal and QCD multijet background has been simulated integrating multiple frameworks: MadGraph, Pythia, and Delphes. The response of particle detectors to the final-state particles has been produced with the fast simulation of the ATLAS detector.

Input features (12)	Description
m_{jj}	Invariant mass of the VBF jet pair
$p_{T,jj}$	Transverse momentum of the VBF jet pair
$p_T^{balance}$	Ratio of the vectorial and scalar sums of the transverse momenta of b_1 , b_2, j_1 and j_2
$(p_T^{j_1} - p_T^{j_2})\ (p_T^{j_1} + p_T^{j_2})$	Asymmetry in the VBF jet transverse momenta
$\Delta\eta(\mathrm{bb}, jj)$	Separation in η between the b-tagged jet pair and the VBF jet pair
$\Delta\phi(\mathrm{bb}, jj)$	Separation in ϕ between the b-tagged jet pair and the VBF jet pair
$\tan^{-1}\left(\tan\left(\frac{\Delta\phi(bb)}{2}\right) / \tanh\left(\frac{\Delta\eta(bb)}{2}\right)\right)$	Measure of the relative angle of η and ϕ between the two b-tagged jets
n_{jets}	Number of jets with $p_T > 20$ GeV and $\lvert\eta\rvert < 4.5$
$\min \Delta R(j_{1(2)})$	Minimum separation in R between the (sub)leading VBF jet and any jet if is not a part of the b-tagged or VBF jet pair
$N_{trk}^{j_{1(2)}}$	Number of tracks matched to the (sub)leading VBF jet







Event selection from classifier scores

- In the context of a binary classifier used for signal vs background classification, selecting based on the score refers to choosing a threshold value on the classifier's output probability or score.
- Adjusting this threshold allows you to control the trade-off between true positives (correctly identified signals) and false positives (backgrounds incorrectly identified as signals), **optimizing the classifier's performance** based on the specific requirements of your application.
- In High Energy Physics analysis, Deep Neural Networks (DNN) are widely used to enhance the detection of **rare signal events.** However, the application of DNN in classifications for physical events often involves **the correlation of the latter with physical quantities.**
- This correlation could bring to the mis-identification of signal events and sculpting effects on the feature distribution.
- This sculpting effect occurs because of the **correlation between the DNN tagger and the Higgs boson mass.** As a result, QCD events with a mass similar to the Higgs boson are misclassified as signals.

Correlation metric

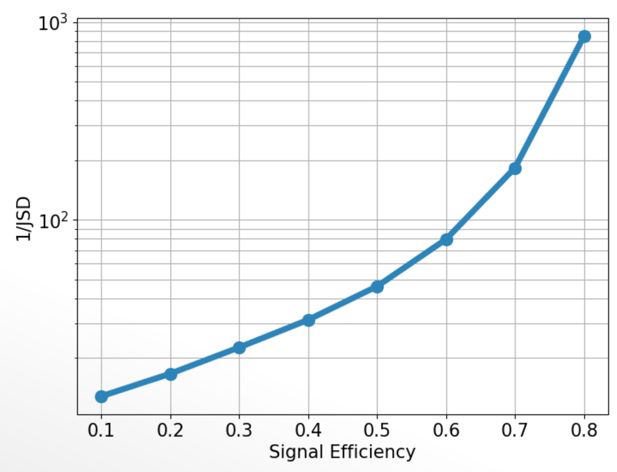
- The Jensen-Shannon divergence is defined as

$$JSD(P||Q) = \frac{1}{2}(KL(P||M) + KL(Q||M))$$

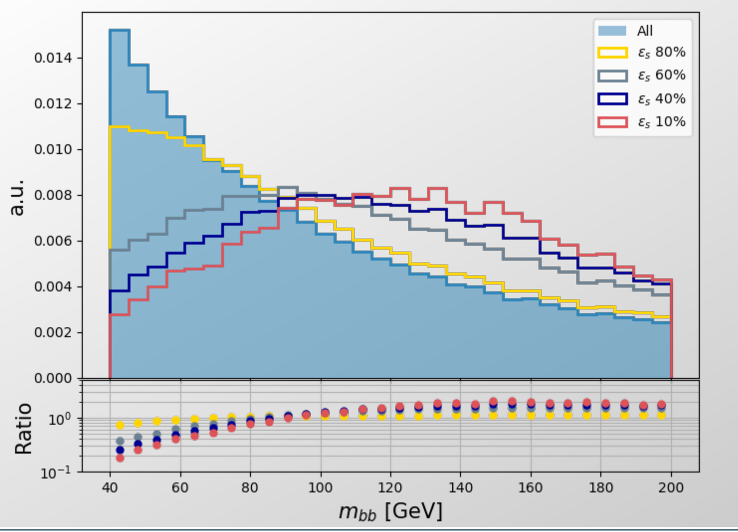
with $M = \frac{P+Q}{2}$ and KL the Kullback-Leibler divergence [2].

- In our case, the divergence is used to measure the **entropy between the normalised mass distributions of the background jets passing and total**, respectively, a given jet tagger cut:

$$JSD(P||Q) = JSD\left(\frac{N_{bkg}^{pass}(m)}{\sum_i N_{bkg}^{pass}} \Big|\Big| \frac{N_{bkg}^{tot}(m)}{\sum_i N_{bkg}^{tot}}\right)$$

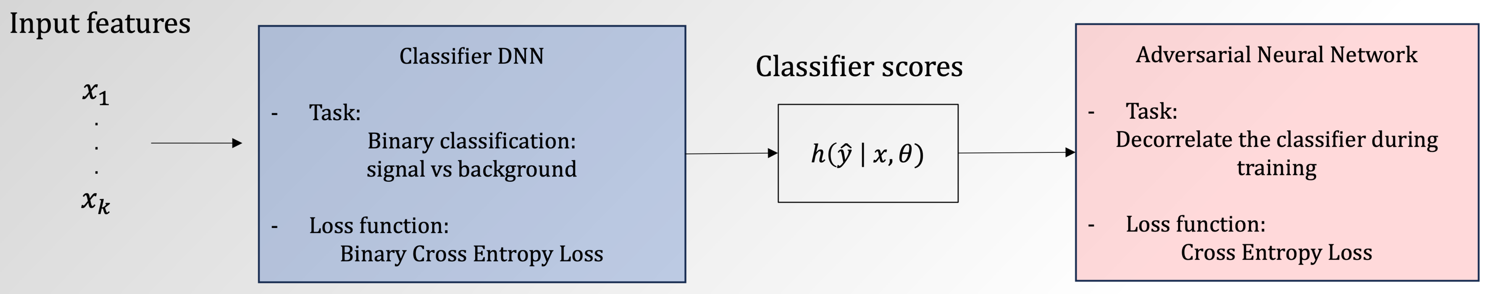


- When it comes to decorrelating the classifier into the workflow, two main strategies can be employed. These strategies involve different approaches that can be used depending on the system's specific requirements.
- The classifier can be decorrelated **during training.**
- The classifier can be decorrelated **after training, acting directly on the classifier scores.**



Adversarial Neural Network

The Adversarial Neural Network (ANN) acts on the classifier during training. **Its goal is to predict the di-b-jet invariant mass bin for each event using the classifier's output [3].**

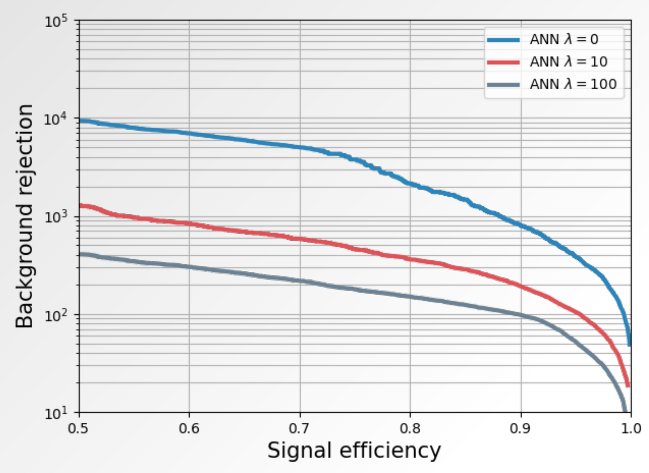


Modified loss function of the network: $L = L_{cl} - \lambda L_{ad}$

Where λ is an additional hyperparameter that weights the action of the ANN on the classifier.

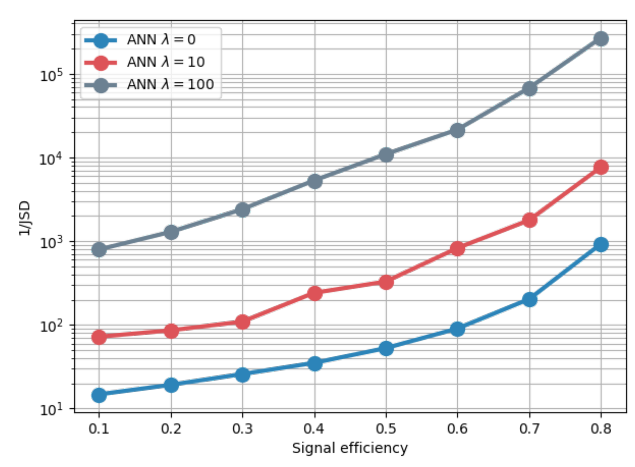
Two training steps:

1. Train the classifier on signal and background with loss L
2. Train the adversarial on background only with loss L_{ad}

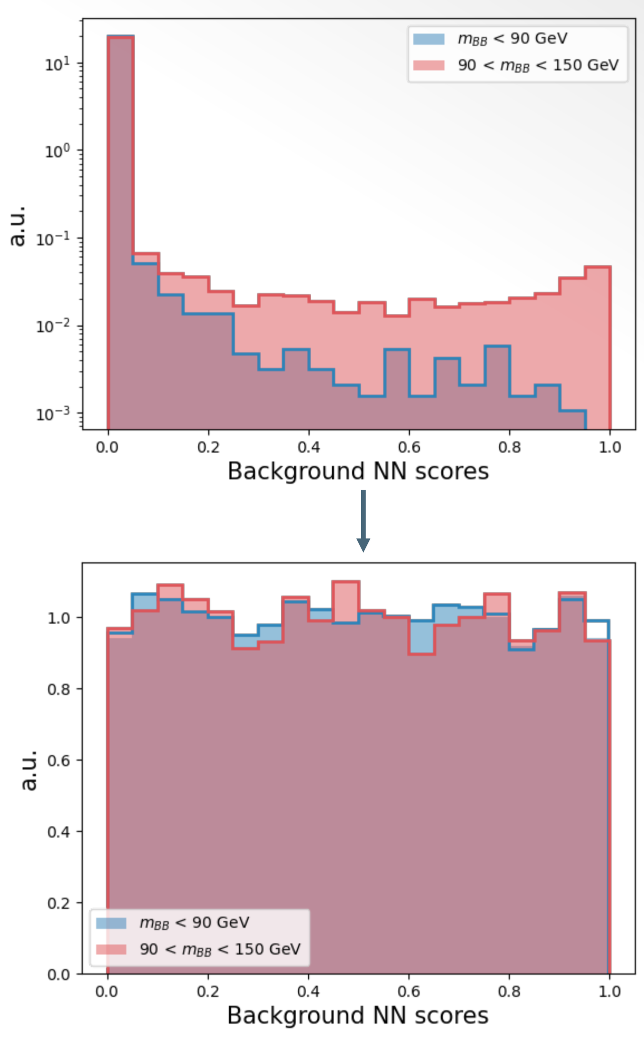


- Classifier: 4 hidden linear layers with 64 nodes
- Adversarial: 4 hidden linear layers with 128 nodes
- The ROC curves show the classifier performance at a fixed architecture trained for 100 epochs with λ = 0,10,100.
- The action of the ANN affects the classifier performance due to the decorrelation from the $m_{b\bar{b}}$.

- The correlation of the DNN with $m_{b\bar{b}}$ can be measured with JSD at different signal efficiencies.
- The larger 1/JSD the more the classifier is decorrelated.
- The correlation depends on the λ value.



Conditional normalizing flow



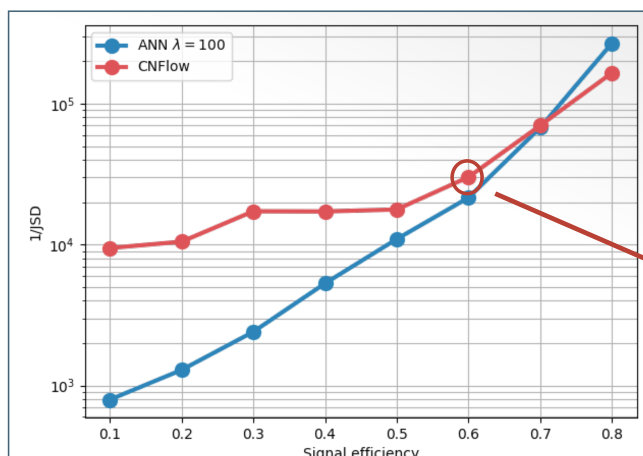
- A normalizing flow is an **invertible map between two distributions.** [4]
- A conditional normalizing flow (CNFlow) p_θ can approximate a data distribution $p_D(h(x)|m)$ by defining a neural network $f_\theta(h(x), m)$ that is invertible given m and a base distribution p that is independent of m.
- The model is fit to data by maximizing the log-likelihood under the change of variables formula:

$$\log p_\theta(h(x)|m) = \log p(f_\theta(h(x), m)) + \log|\det J_{f_\theta}(h(x), m)|$$

where J_{f_θ} is the Jacobian of f_θ .

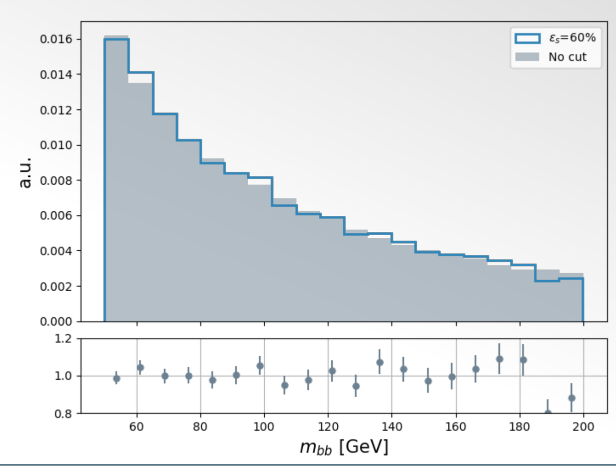
- **The CNFlow are a fast and simple method that can be applied directly on classifier scores.**
- The target distribution in this studies is a uniform distribution.

Final result



- The performance of CNFlow in decorrelation is superior to that of ANN with λ=100.
- The CNFlow can be applied to scores of different classifiers.

- Thanks to the decorrelation we can avoid the background sculpting and enhance the sensitivity to rare signals as the VBF H bb.



Reference

[1] https://link.springer.com/article/10.1140/epjc/s10052-021-09192-8
[2] https://www.sciencedirect.com/science/article/abs/pii/S0016003296000634
[3] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERSx/HIGG-2019-04/
[4] https://arxiv.org/abs/2211.02486

FONDAZIONE BRUNO KESSLER

UNIVERSITY OF TRENTO