PROCEEDINGS OF SCIENCE

Deep Learning applied to VBF Higgs Boson in the $b\bar{b}$ channel: a study of Neural Networks impact on High Energy Physics analysis

Greta Brianti[a,b,c,d,*]

[a] *University of Trento, Via Sommarive 14, Trento, Italy*
[b] *Fondazione Bruno Kessler, Via Sommarive 18, Trento, Italy*
[c] *INFN-TIFPA, Via Sommarive 14, Trento, Italy*
[d] *CERN, Espl. des Particules 1/1211, 23 Genève, Switzerland*

E-mail: greta.brianti@cern.ch

In this research, we investigated the influence of a Fully Connected Deep Neural Network (FCN) for signal-to-background classification on a sample of Vector Boson Fusion (VBF) Higgs bosons decaying into b-quark pairs. The FCN improves the identification of the signal events overwhelmed by the QCD background. However, the selection of the signal efficiency Working Point has a sculpting effect on the background distribution of the invariant mass of the tagged jets. This condition is generated by the algorithm correlation with the Higgs boson mass. In fact, due to the input features non-linear dependence on the tagged jets' invariant mass, the algorithm learns that the signal event mass of the b-jet pair is close to the Higgs boson mass. Therefore the background events that have similar b-jets mass are mis-identified as signal. In this paper, the correlation impact has been studied. Moreover, two different decorrelation approaches have been tested on Monte Carlo datasets self-produced.

42nd International Conference on High Energy Physics (ICHEP2024)
18-24 July 2024
Prague, Czech Republic

*Speaker

1. Introduction

The second most frequent Higgs boson production mechanism, the Vector Boson Fusion (VBF) can be exploited to study the hadronic decay channels of the Higgs Boson. 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 $H \to b\bar{b}$ channel analysis for the Run 2 data with ATLAS has been provided [1]. 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. Fully Connected Neural Networks (FCN) 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.

2. Dataset

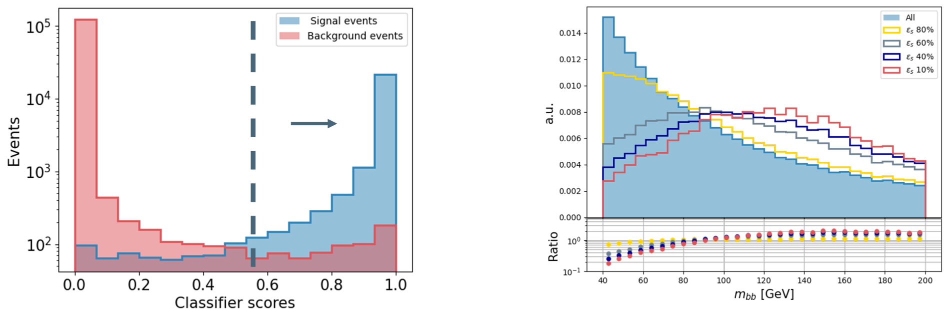
A dataset of VBF $H \to b\bar{b}$ signal and NLO QCD multijet background has been simulated integrating multiple frameworks: MadGraph, Pythia, and Delphes [2–4] at a centre-of-mass energy of $\sqrt{s} = 14$ TeV. The response of particle detectors to the final-state particles has been produced with the fast simulation of the ATLAS detector. From the simulated datasets, 12 features, described in Table 1, have been extracted to mimic the analysis on the Run 2 data in ATLAS.

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
$\Delta\eta(bb, jj)$	Separation in η between the b-tagged jet pair and the VBF jet pair
$\Delta\phi(bb, jj)$	Separation in ϕ between the b-tagged jet pair and the VBF jet pair
$\tan^{-1}[\tan(\frac{\Delta\phi(bb)}{2})/\tanh(\frac{\Delta\eta(bb)}{2})]$	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 it 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

Table 1: Input features of the FCN classifier for signal-to-background discrimination

3. Decorrelation methods and results

To enhance the rare signal events, the selection based on the score is applied. The selection consists of choosing a threshold value on the classifier's output probability or score, as shown in Figure 1 (a). Adjusting this threshold allows for control of the trade-off between true positives (correctly identified signals) and false positives (backgrounds incorrectly identified as signals). However, the selected background events in the Signal Region show sculpting on the invariant mass distribution of the b-tagged jets (see Figure 1 (b)). Therefore, the background events that pass the fixed cut are misidentified as signals due to the similarity of the b-jet invariant mass with that of the Higgs boson.

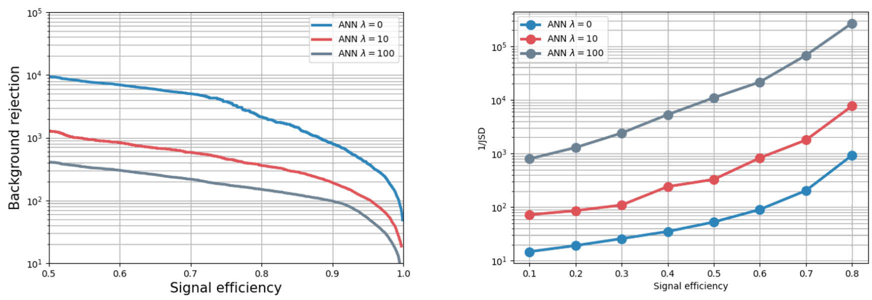


(a) **(b)**

Figure 1: (a) FCN scores with cut representation for the event selection. (b) Sculpting effect due to the FCN correlation with the invariant mass of the b-tagged jets at different Signal Efficiency selections.

To decorrelate the FCN from $m_{b\bar{b}}$, two main strategies can be employed. These strategies involve different approaches that can be used depending on the system's specific requirements. The FCN can undergo decorrelation either during the training process or post-training, directly influencing the classifier scores. This paper explores two distinct approaches for achieving this: the Adversarial Neural Network (ANN) [1] and the Conditional Normalizing Flow (CNFlow) [5]. The subsequent sections will delineate the application procedure and present the results.

3.1 Adversarial Neural Network



(a) **(b)**

Figure 2: Adversarial Neural Network performances at $\lambda = 0, 10, 100$. (a) ROC curve for the classification performance. (b) 1/JSD measurement for the decorrelation performance. As JSD quantifies the entropy between two distributions, 1/JSD serves to indicate the similarity between them. Consequently, a higher 1/JSD value signifies a more effective decorrelation process.

The ANN acts on the FCN during training. Its goal is to predict the di-b-jet invariant mass bin for each event using the classifier's output. The classifier is trained with a modified loss function,

$L_{comb} = L_{cl} - \lambda L_{ann}$, where L_{cl} is the loss function of the FCN, L_{ann} is the loss function of the ANN and λ is an additional hyperparameter that weights the action of the ANN on the classifier. The correlation of the FCN+ANN with $m_{b\bar{b}}$ can be measured with the Jenson-Shannon divergence (JSD) [6] at different signal efficiencies. As expected, the correlation and the FCN performance depend on the λ value. Figures 2 (a) and (b) show, respectively, the classifier performance and the decorrelation metrics trend at different Signal Efficiency selections on the FCN scores distribution for $\lambda = 0, 10, 100$.

3.2 Conditional Normalizing Flow

A normalizing flow is an invertible map between two distributions. A 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. In this study, $h(x)$ is the FCN scores depending on the set of input features x and m is the invariant mass of the b-tagged jets. 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 $J[f_\theta(h(x), m)]$ is the Jacobian of $f_\theta(h(x), m)$. The CNFlow is a fast and simple method that can be applied directly

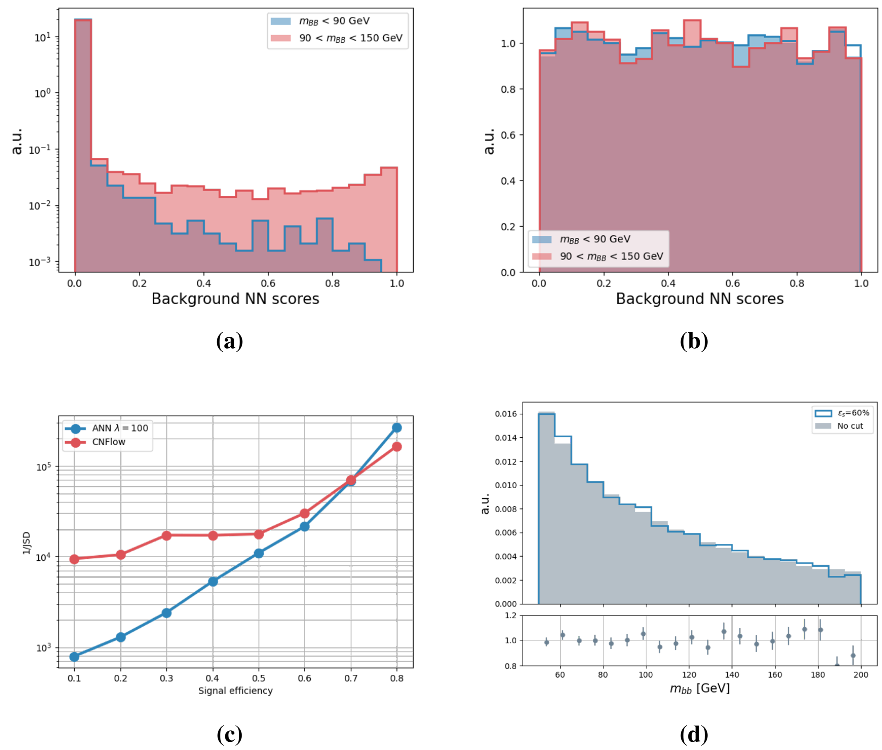


Figure 3: Action of the CNFlow algorithm on the FCN scores. (a) FCN scores in different invariant masses of the b-tagged jets. (b) FCN scores in different invariant masses of the b-tagged jets after the CNFlow application. (c) Comparison of 1/JSD between ANN $\lambda = 100$ and CNFlow. (d) Final distribution of the invariant mass of the b-tagged jets after the CNFlow application on FCN scores at 60% Signal Efficiency.

to classifier scores. The target distribution in this study is uniform as shown by Figure 3 (a), (b).

The performance of CNFlow in decorrelation is superior to that of ANN with $\lambda = 100$ for Signal Efficiency under 60% (see Figures 3 (c)). Thanks to the decorrelation we can avoid the background sculpting and enhance the sensitivity to rare signals as the VBF $H \rightarrow b\bar{b}$, as shown in Figure 3 (d).

4. Conclusion

In this paper, the application of different decorrelation methods on a VBF $H \rightarrow b\bar{b}$ dataset has been presented. The decorrelation method to be chosen is strictly related to the physical problem to deal with. It is possible to consider different approaches that involve diverse advanced machine-learning techniques.

References

[1] Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s} = 13$ TeV, ATLAS Collaboration, Eur. Phys. J. C 81, 537 (2021), https://doi.org/10.1140/epjc/s10052-021-09192-8

[2] The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, J. Alwall et al., Journal of High Energy Physics 2014.7, (2014), https://doi.org/10.1007%2Fjhep07%282014%29079

[3] A comprehensive guide to the physics and usage of PYTHIA 8.3.2022, Christian Bierlich et al., (2022), https://doi.org/10.48550/arXiv.2203.11601

[4] DELPHES 3: a modular framework for fast simulation of a generic collider experiment, J. de Favereau et al., Journal of High Energy Physics 2014.2, (2014), https://doi.org/10.1007%2Fjhep02%282014%29057

[5] Decorrelation with conditional normalizing flows, S. Klein, T. Golling, (2022), https://doi.org/10.48550/arXiv.2211.02486

[6] The Jensen-Shannon divergence, M.L. Menéndez, J.A. Pardo, L. Pardo, M.C. Pardo, Journal of the Franklin Institute Volume 334 Issue 2, (1997), https://doi.org/10.1016/S0016-0032(96)00063-4.

PoS(ICHEP2024)1050