Benedetta Camaiani PhD Research Project

Towards precision Higgs boson measurements with Adversarial Neural Networks

The current Higgs physics context Following the discovery of the Higgs boson by the ATLAS and CMS collaborations and with the increasing amount of data collected by the LHC, the properties of the newly observed particle have been measured with ever improving precision. Up to now, CMS has collected about 190 fb^{-1} of proton-proton collision data at energies of 7, 8 (Run 1) and 13 TeV (Run 2). In April 2022 the LHC is going to restart operations with an energy increase up to 14 TeV. During the Run 3 data acquisition, it is expected to double the integrated luminosity of Run 2, reaching ~ 350 fb^{-1} by the end of 2024.
During my PhD program, I plan to contribute to the development of the field in two directions, using data from the CMS experiment: one is the precision measurement of the Higgs boson differential production cross section, the second is the measurement of its total decay width, one of the most challenging properties to experimentally test.

Differential and fiducial measurement of the production cross section of the Higgs boson In this context, differential and fiducial measurements play a key role: they allow to determine the cross section of the process of interest as a function of one or more relevant kinematic variables whitin a restricted phase space, called fiducial volume, tailored to reproduce the experimental selection. The interest in this type of measurement is due to the fact that they provide results that can be directly compared to any physics model, with no need to simulate the detector's response. Thus, they can be easily reinterpreted in the future, representing the legacy of today's experiments.
During my master thesis I studied the feasibility of the new measurement of the production cross section of the Higgs boson, as a function of the signed azimuthal angle difference $\Delta\phi_{jj}$ between the two leading hadronic jets produced in association. In this study, the Vector Boson Fusion (VBF) production mode is considered as the signal process. The $\Delta\phi_{jj}$ observable is sensitive to the Higgs boson CP-properties and because these are predicted to differ from the SM case in many Beyond Standard Model (BSM) theories, it is particularly important that the measurement is not biased towards the SM expectation. For this reason I extracted the signal by developing an ad-hoc machine learning (ML) discriminator, made independent on the signal model of choice. I performed this preliminary analysis using proton-proton collision data collected by the CMS detector in 2018 at a center of mass energy of 13 TeV, and the results I have obtained show a sizable reduction of model dependency in the final result. Therefore, as the first step of my PhD studies, I will extend and further improve this analysis including all the data collected during Run 2 and part of those that will be acquired during Run 3. The analysis will benefit from the planned increases of integrated luminosity of the next data acquisitions, since the VBF channel is statistically limited.

Measurement of the Higgs boson off-shell production Among the properties of the Higgs boson, its total decay width represents a pivotal parameter in the SM and needs to be precisely measured. However, the direct measurement is limited by the instrumental resolution of the detector, which is of the order of 1 GeV. It is indeed three orders of magnitude larger than the expected natural width of the SM Higgs boson, which is $\Gamma_H \sim 4$ MeV. An alternative approach is to use events away from the on-peak region in the Higgs boson decay into two W bosons. These events are generally called *virtual* or *off-shell* processes, meaning that the invariant mass of the two W bosons is greater than the Higgs boson mass. This analysis represents a unique way to constrain the total decay width of the Higgs boson since the ratio of the off-shell and on-shell production and decay rates in the $H \to WW$ channel leads to an indirect measurement of Γ_H. In the latest CMS published analysis [1], a boosted decision tree discriminant has been used to distinguish between the off-shell Higgs boson signal and the background; the training procedure has been performed on a Monte Carlo sample generated by assuming the SM decay width as input parameter, potentially biasing the result towards the SM.

As far as the on-shell events measurement is concerned, it has been shown that the Gluon Fusion (GF) production channel of the Higgs boson is limited by the systematic uncertainties. For instance, in the Higgs boson properties measurements [2], the dominant impact is due to the theoretical uncertainties in modeling the signal and for this reason, not even the forthcoming increase of data statistics will bring a marked improvement. In this context, the crucial goal would be to make the analysis as independent as possible from these sources of uncertainty. I plan to reduce both the model dependency in the extraction of the off-shell signal, and the impacts of systematics in the extraction of the on-shell component using the *Domain Adaptation* (DA) [3].

Methodology DA is a set of techniques that aims at making a ML algorithm perform well when applied on a data set, or domain, that is different from the one it was trained on. I have already used this ML technique for my master thesis, in order to obtain a differential measurement of $\Delta\phi_{jj}$ which is agnostic with respect to several BSM scenarios. The implementation of DA I used is based on a two networks system, called *Adversarial Neural Network* (ANN), consisting of a classifier (C) and an adversary (A). The role of the classifier is to discriminate signal and background. C is trained on a data sample including events coming from different domains, each of which identified by an index. On the other hand, the adversary is designed to guess the data set to which the input event belongs, regressing the domain. The two networks are trained in a competitive way so that C classifies the events correctly while A fails to recognize the domains.
In the case of the differential analysis, the different domains correspond to several BSM scenarios and A tries to identify which of these a given signal event was drawn from. This is exactly the implementation that I set up in the context of $\Delta\phi_{jj}$ differential analysis to build a discriminator with good signal to background discrimination performances, whose output contains little residual information on the specific model of the signal events and is in this sense model independent. The same idea can be applied to the Γ_H measurement. Regarding the off-shell category definition, the different domains correspond to a continuum of possible Γ_H values. This will allow me to realize a discriminator which is agnostic with respect to the expected Higgs boson total decay width and thus will reduce the model dependence of the analysis. Besides, in the on-shell cross section measurement, which is systematically limited, I would train the ANN against an admixture of events with variations of systematic sources, to limit the discriminator's dependency on them. In this way the ANN output will be as independent as possible from these variations and the impact of the systematic uncertainties on the analysis will be reduced.
Mathematically, the adversarial training is performed minimizing a single loss function:

$$\mathcal{L}(C + A) = \mathcal{L}(C) - \alpha \cdot \mathcal{L}(A)$$

where α is an additional hyperparameter balancing the performances of the two networks. All the hyperparameters defining the ANN have to be optimized so that C is able to discriminate events while simultaneously preventing A from regressing the domain index. When an equilibrium between the two networks' performances is reached, the output of the classifier is independent of the domain index. This technique allowed me to obtain successful results in my master thesis studies and I believe that it will unlock great potential in the measurements that I plan to pursue.

References

[1] The CMS Collaboration, "Search for Higgs boson off-shell production in proton-proton collisions at 7 and 8 TeV and derivation of constraints on its total decay width", 2016

[2] The CMS Collaboration, "Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV", 2019.

[3] S. Ben-David, J. Blitzer, K. Crammer, A. Kalesza, F. Pereira, J.Wortman Vaughan, "A theory of learning from different domains", 2009.