Comparison of CP-sensitive observables to search for CP-violation in vector-boson-fusion production of the Higgs boson exploiting the $H \to \tau_{\text{had}}\tau_{\text{had}}$ decay at the ATLAS experiment

Bachelor's Thesis
presented for the degree of Bachelor of Science (B.Sc.) in Physics

Submitted by
Lea Kuttler

Supervised by
Prof. Dr. Markus Schumacher

Albert-Ludwigs-Universität Freiburg
Faculty for Mathematics and Physics

February 6, 2024

universität freiburg

Declaration

I hereby declare, that I am the sole author and composer of this Thesis and that no sources or aids, other than those listed, have been used. Furthermore, I declare that I have cited all literal or paraphrased content from other sources in accordance with recognized academic standards (lege artis).
The submitted bachelor's thesis has not been the subject of another examination procedure, either in its entirety or in significant parts.

Place, Date	Signature

Abstract

This thesis compares the sensitivities of various CP-odd observables to possible CP-violating contributions in the HVV vertex of the vector-boson-fusion (VBF) Higgs-boson production mode, considering the decay channel $H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$. Simulated event samples, corresponding to proton-proton collision data, measured by the ATLAS detector at the LHC during the full Run2, at a center of mass energy of $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 139 fb^{-1} are used. The analysis considers established observables, like the Optimal Observable ($\mathcal{OO}$) and the signed difference between the azimuthal angles ($\Delta\Phi_{\text{jj}}^{\text{sgn}}$) of the two leading VBF tagging jets and observables constructed using various machine learning methods, namely the observable $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{final}}$ constructed directly from the output of a multiclass neural network (NN), the observable $\mathcal{O}_{\text{NN}}^{\text{Reg}}$ obtained by targeting the $\mathcal{OO}$ using a regression NN and the observable $P_{j_0}^{T} P_{j_1}^{T} \sin\Delta\Phi_{\text{jj}}^{\text{sgn}}$, resulting from a symbolic regression method, when targeting the $\mathcal{OO}$. The CP-sensitivity is estimated via a maximum likelihood (ML) method that allows constraining the dimensionless parameter $\tilde{d}$, that defines the strength of possible new CP-violating couplings, through the construction of central confidence intervals for this parameter, considering only the shape of the CP-odd observable distributions. The expected CP-sensitivity of the $\mathcal{OO}$ is $\tilde{d}$ ∈[-0.0150(5),0.0149(5)] at 1σ confidence level and $\tilde{d}$ ∈[-0.0301(5),0.0300(5)] at 2 σ confidence level. This CP-sensitivity is not exceeded by the other considered CP-odd observables. The expected CP-sensitivity achieved by $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ is $\tilde{d}$ ∈[-0.0158(5),0.0157(5)] at 1σ confidence level and $\tilde{d}$ ∈[-0.0318(5),0.0316(5)] at 2σ confidence level. The observable $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{final}}$ constrains $\tilde{d}$ to [-0.0152(5),0.0152(5)] at 1σ confidence level and to [-0.0306(5),0.0306(5)] at 2σ confidence level. $\mathcal{O}_{\text{NN}}^{\text{Reg}}$, allows imposing the following constraints: $\tilde{d}$ ∈[-0.0149(5),0.0150(5)] at 1σ confidence level and [-0.0301(5),0.0301(5)] at 2σ confidence level. The observable $P_{j_0}^{T} P_{j_1}^{T} \sin\Delta\Phi_{\text{jj}}^{\text{sgn}}$ constrains $\tilde{d}$ to [-0.0150(5),0.0151(5)] at 1σ confidence level and to [-0.0303(5),0.0303(5)] at 2σ confidence level.

Zusammenfassung

Diese Arbeit vergleicht die Sensitivitäten verschiedener CP-ungerader Observablen gegenüber potenziellen CP-verletzenden Beiträgen im HVV-Vertex des Vektor-Boson-Fusions (VBF) Higgs-Boson-Produktionsmodus unter Berücksichtigung des Zerfallskanals $H \to \tau_{\text{had}}\tau_{\text{had}}$. In der vorgelegten Analyse werden simulierte Ereignisse verwendet, die dem gesamten Datensatz, der mit dem ATLAS Detektor am LHC während des Run2 bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV entsprechend einer integrierten Luminosität von 139 fb^{-1} aufgenommen wurde, entsprechen. Die Analyse umfasst etablierte CP-ungerade Observablen wie die Optimal Observable ($\mathcal{OO}$) und die vorzeichenbehaftete Differenz zwischen den Azimutwinkeln ($\Delta\Phi_{\text{jj}}^{\text{sgn}}$) der beiden führenden VBF-Tagging-Jets. Zusätzlich werden mittels verschiedener Methoden des maschinellen Lernens konstruierte Observablen untersucht, darunter $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{final}}$, $\mathcal{O}_{\text{NN}}^{\text{Reg}}$ und $P_{j_0}^T P_{j_1}^T \sin\Delta\Phi_{\text{jj}}^{\text{sgn}}$. $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{final}}$ resultiert direkt aus den Vorhersagen eines mehrklassigen Klassifizierungs-Neuronalen-Netzwerks, während $\mathcal{O}_{\text{NN}}^{\text{Reg}}$ die $\mathcal{OO}$-Verteilung numerisch approximiert und $P_{j_0}^T P_{j_1}^T \sin\Delta\Phi_{\text{jj}}^{\text{sgn}}$ eine analytische approximation der $\mathcal{OO}$, resultierend aus einer symbolischen Regressionsmethode, darstellt. Die CP-Sensitivität wird mithilfe eines Maximum-Likelihood-(ML)-Verfahrens abgeschätzt, das es ermöglicht, die Stärke $\tilde{d}$ der Beiträge möglicher neuer, CP-verletzender Kopplungen, durch die Konstruktion zentraler Konfidenzintervalle einzugrenzen. Die erwartete CP-Sensitivität der $\mathcal{OO}$ liegt bei $\tilde{d}$ ∈[-0.0150(5),0.0149(5)] auf dem 1σ-Konfidenzniveau und bei $\tilde{d} \in [-0.0301(5), 0.0300(5)]$ auf dem 2σ-Konfidenzniveau. Diese CP-Sensitivität wird von allen anderen betrachteten CP-ungeraden Observablen nicht übertroffen. Die für $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ erwartete CP-Sensitivität, liegt bei $\tilde{d}$ ∈[-0.0158(5),0.0157(5)] auf dem 1σ-Konfidenzniveau und bei $\tilde{d} \in [-0.0318(5), 0.0316(5)]$ auf dem 2σ-Konfidenzniveau. Die Observable $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{final}}$ beschränkt $\tilde{d}$ auf [-0.0152(5),0.0152(5)] auf dem 1σ-Konfidenzniveau und auf [-0.0306(5),0.0306(5)] auf dem 2σ-Konfidenzniveau. $\mathcal{O}_{\text{NN}}^{\text{Reg}}$ ermöglicht die folgenden Eingrenzungen: $\tilde{d}$ ∈[-0.0149(5),0.0150(5)] auf dem 1σ-Konfidenzniveau und [-0.0301(5),0.0301(5)] auf dem 2σ-Konfidenzniveau. Die Observable $P_{j_0}^T P_{j_1}^T \sin\Delta\Phi_{\text{jj}}^{\text{sgn}}$ beschränkt $\tilde{d}$ auf [-0.0150(5),0.0151(5)] auf dem 1σ-Konfidenzniveau und auf $[-0.0303(5), 0.0303(5)]$ auf dem 2σ-Konfidenzniveau.

Contents

Chapter 1

Introduction

The Standard Model (SM) of particle physics [1–3] is the current best theory to describe elementary particles - the fundamental constituents of the universe - and the forces that govern their interactions. A central principle of the SM is the concept of local gauge invariance, which forms the basis for its successful description of three out of the four fundamental forces: electromagnetism, the weak force and the strong force.

The SM demonstrated its predictive power on numerous occasions, anticipating the existence of several particles, for example the Z and the W bosons, which were later confirmed through experiments at the SPS Proton-Antiproton Collider [4, 5] at CERN or the top quark which was first observed at the Fermilab Tevatron [6]. The discovery of the Higgs boson in 2012 at the Large Hadron Collider (LHC) [7, 8], validated the Brout-Englert-Higgs-Guralnik-Hagen-Kibble mechanism [9–11], formulated in 1964, by which elementary particles acquire mass within the SM framework and simultaneously confirmed the existence of the last missing piece of the SM.

However, some observed phenomena can not be explained by the predictions of the SM. One such phenomenon is the imbalance of the amount of baryons and anti-baryons that is observed in our universe. This observation can only be explained if three conditions, formulated by A.D. Sakharov in 1967 [12], are fulfilled. One of these conditions requires the existence of processes that violate the invariance under the combination of charge conjugation and parity transformation (CP). In the SM, CP- symmetry is broken only in the electroweak sector. Here, CP-violation is incorporated through complex phases in the Cabibbo-Kobayashi-Maskawa (CKM) matrix [13]. However, the amount of CP-violation resulting from this SM source is not sufficient to explain the baryon asymmetry of our universe (BAU). This motivates the experimental search for additional sources of CP-violation beyond the standard model (BSM).

Given that the electroweak sector already encompasses SM sources of CP-violation, attempts have been made multiple times to search for BSM sources within this sector. Analyses, probing the CP-invariance of Higgs boson couplings to electroweak gauge bosons (HVV couplings) have been performed using data from the vector-boson-fusion (VBF) Higgs boson production through proton-proton (pp) collisions, exploiting the decay $H \to \tau\tau$, at the ATLAS detector [14] in 2016 at an integrated luminosity 20.3 fb^{-1} and at a center of mass energy $\sqrt{s} = 8$ TeV [15] and in 2020 with 36.1fb $^{-1}$ at $\sqrt{s} = 13T$ eV [16]. In both cases, no signs for CP-violation have been found. These analyses utilized the Optimal Observable ($\mathcal{OO}$) method, where the structure of the HVV couplings is probed using the CP-odd observable $\mathcal{OO}$, constructed from additional CP-violating contributions that appear with a strength $\tilde{d}$, when introducing CP-odd BSM interactions in the framework of an effective field theory (EFT). The strength parameter $\tilde{d}$ could be constrained to the interval [-0.11,0.05][15] and [-0.090,0.035][16] at 1σ confidence level, compatible with the SM prediction of $\tilde{d} = 0$. A further established CP-odd observable, besides the $\mathcal{OO}$, is given by the signed difference $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ between the azimuthal angles of the two leading tagging jets that occur in the VBF Higgs boson production [17]. Recently, new machine learning based CP-odd observables have been introduced. For example, in [18] a CP-odd observable is constructed directly from the output of classification neural networks (NNs). Furthermore, [19] employs a symbolic regression method to construct an analytical expression that approximates the $\mathcal{OO}$.

In this thesis, the sensitivity of CP-odd observables, constructed following the machine learning approaches proposed in [18, 19], to CP-violating contributions in the HVV vertex of the VBF Higgs boson production process is evaluated using a maximum likelihood (ML) method and compared to the CP-sensitivities of the well established observables $\mathcal{OO}$ and $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$.

The presented analysis particularly focuses on the decay channel $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$ of the VBF Higgs boson, using simulated event samples corresponding to data recorded by the ATLAS detector during the full Run2 with an integrated luminosity of 139 fb^{-1} at a center of mass energy of $\sqrt{s} = 13$ TeV.

The thesis is structured as follows: Chapter 2 describes the signal and background processes considered during the analysis. In chapter 3 the MC generated dataset and the corresponding event weights are discussed. Furthermore, this chapter specifies selection criteria imposed on the events in the dataset, aiming to enhance the signal to background ratio during the analysis.

The approach of utilizing CP-odd observables for performing CP-tests within the framework of effective field theories is elaborated in chapter 4. Moreover, the technique of estimating sensitives of CP-odd observables to CP-violating HVV couplings by utilizing negative log likelihood (NLL) scans, a maximum likelihood (ML) method that allows constraining $\tilde{d}$ through the construction of central confidence intervals, is discussed. This method requires predictions on the observable distributions in different CP-violating scenarios. The generation of these predictions, through a matrix elements based reweighting technique is also introduced in this chapter. Chapter 5 introduces well established observables that are commonly used in CP-tests, namely the Optimal Observable $\mathcal{OO}$ and the signed difference in azimuthal angles of the leading tagging jets $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ and provides a first demonstration of estimating CP-sensitivity via NLL scans. In chapter 6 various machine learning methods are employed for the construction of CP-odd observables which are then also evaluated in terms of their CP-sensitivities via NLL scans. When introducing additional CP-odd couplings in the EFT approach, their contribution appears in the corresponding matrix elements as a CP-odd interference term and an additional CP-even quadratic term. Classification NNs are used to construct CP-odd observables that comprise the information on the CP-structure of the HVV-vertex that is present in the interference term, following the approach proposed in [18]. Several binary and multiclass NNs are considered. Different studies are performed regarding the impact of specific conditions during the NN training, i.e impacts of the choice of the specific CP-violating $\tilde{d} \neq 0$ scenario assumed during training, the choice of input features and the inclusion of the matrix element based weights in the loss function. Moreover, regression neural networks are employed to resemble the $\mathcal{OO}$ distribution and finally the method of symbolic regression is applied, first following [19] to analytically approximate the $\mathcal{OO}$ distribution as it is predicted by the SM and in the next step, aiming to additionally approximate the $\mathcal{OO}$ distribution as predicted for a CP-violating scenario. Chapter 7 compares the well established CP-odd observables and the constructed machine learning based observables regarding their CP-sensitivites. Finally, chapter 8 provides a discussion of the results obtained during this analysis.

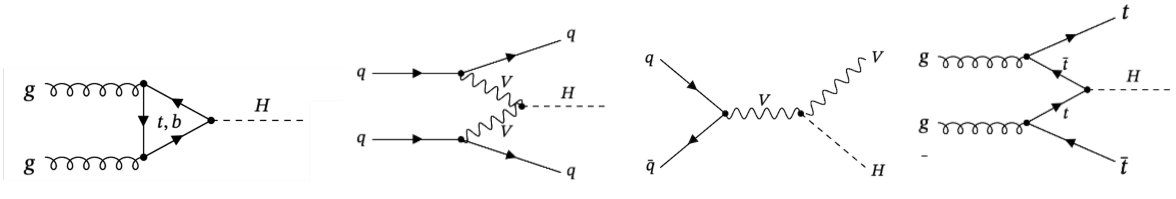
Chapter 2

Signal and Background Processes

The Large Hadron Collider (LHC) [20] is the world's most powerful and complex particle collider, located beneath the Franco-Swiss border near Geneva. Operated by CERN (Conseil Européen pour la Recherche Nucléaire), the LHC is a scientific instrument designed to explore the fundamental constituents of matter and their interactions. In the LHC, protons are collided at high energies.

One of the main motivations for building the LHC was the search for the Higgs boson, which is a short-lived, electrically neutral, fundamental particle with a spin of 0. After its prediction through the Brout-Englert-Higgs-Guralnik-Hagen-Kibble mechanism, formulated in 1964 [9–11], the Higgs boson was first experimentally observed in 2012 by the ATLAS and CMS experiments at CERN, using data from proton-proton (*pp*) collisions at center-of-mass-energies of 7 TeV and 8 TeV [7, 8]. According to the current ATLAS measurement [21], the mass of the Higgs boson is $M_{\text{Higgs}} = 125.11 \pm 0.11$ GeV.

Since protons are hadrons, their collision ultimately comes down to the interaction of partons, enabling the emergence of the Higgs boson through different production mechanisms. Figure 2.1 shows example Feynman diagrams for the four dominant Higgs-boson production modes at the LHC, namely the gluon-gluon-fusion (ggF) production, the vector-boson-fusion (VBF) production, the associated production with a vector boson (VH) and the associated production with two top-quarks (ttH) [22].



(a) (b) (c) (d)

Figure 2.1: Example Feynman diagrams for the dominant Higgs-boson production modes at the LHC: (a) Gluon-gluon-fusion (ggF) (b) vector-boson fusion (VBF) (c) associated production with vector bosons (VH) and (d) associated production with top-quark pairs (ttH).

2.1 VBF Higgs-Boson Signal

The presented analysis considers the production of Higgs bosons via VBF at the LHC at a center of mass energy of $\sqrt{s} = 13$ TeV, to investigate the CP-invariance of the couplings of electroweak gauge bosons $V \in \{W^+, W^-, Z\}$ to the Higgs-boson (HVV-couplings). This process is described by

$$pp \to Hqq. \tag{2.1}$$

The corresponding cross-section is $\sigma_{\text{VBF}} = (3.779 \pm 0.794)$ pb (for $M_{\text{Higgs}} = 125.09$ GeV) [23]. During the collision of the initial state protons p, two quarks q interact through the emission of electroweak gauge bosons, which then fuse, resulting in the production of the Higgs boson. Since quarks cannot exist in unbound states [22], each final state quark hadronises [22] and causes a cascade of secondary particles, which is referred to as tagging jet. These tagging jets are characterized by having a large invariant mass M_{jj} and a large separation in pseudorapidity (see section 3.2) $\Delta\eta_{jj}$.

Particularly, this analysis focuses on the decay of the VBF Higgs-boson into a pair of τ-leptons, for which the branching ratio is [23]

$$\text{BR}(H \to \tau\tau) = 6.27\% \ \pm 1.6\% \ . \tag{2.2}$$

Since τ-leptons have a mean lifetime of $\tau_\tau = 2.9 \cdot 10^{-13}$ s [24], they can only be detected through their decay products. Specifically, processes where the τ-leptons decay into a fully hadronic (visible) final state are considered:

$$H \to \tau_{\text{had}}\tau_{\text{had}} \to \nu + (n\pi^- m\pi^0) + \bar{\nu} + (n\pi^+ m\pi^0). \quad \text{with } n = 1, 3 \text{ and } m = 0, 1, 2.$$

An example Feynman diagram for this process is shown in figure 2.2

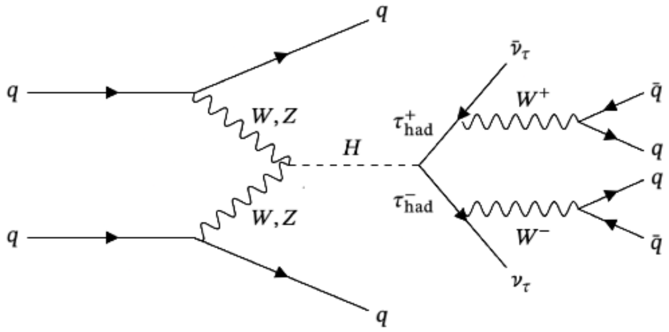


Figure 2.2: Example Feynman diagram for the VBF production of the Higgs-Boson decaying into a pair of hydronically decaying τ-leptons

The branching ratio for the decay of a τ-lepton in a hadron and a τ-neutrino is given as [24]

$$\text{BR}(\tau \to \text{had} + \nu_\tau) = (64.79 \pm 0.06)\% \tag{2.3}$$

The analyzed dataset corresponds to an integrated luminosity of $\int Ldt$ =139 fb^{-1}. The total number of produced signal events is then given as

$$N_{\text{prod}} = (\int Ldt) \cdot \sigma_{\text{VBF}} \cdot \text{BR}(H \to \tau\tau) \cdot \text{BR}(\tau \to \text{had} + \nu_\tau)^2 = 13825. \tag{2.4}$$

To reconstruct the four momentum of the Higgs boson, the kinematic information of all final state decay products is required. However, since the final state includes neutrinos (see figure 2.2), which do not interact with the detector, this information is not entirely accessible. Since the momentum in the transverse plane is a conserved quantity, assuming the transverse momenta of the incoming partons are zero allows to reconstruct a part of this information. Due to the missing neutrinos, calculating the negative sum of the transverse momenta of all detected final state particles will result in a non-zero value, which represents the missing transverse energy (E_T^{miss}). However, this method only allows restoring the transverse kinematics of the di-τ system and cannot fully reconstruct the corresponding four momenta. In this thesis, the Missing Mass Calculator (MMC) algorithm [25] is utilized to obtain the complete four momenta of the di-τ sytem.

2.2 Background Processes

Proton-proton collisions give rise to a range of processes, that can end up with the same final state as the signal process. These processes are referred to as irreducible background. Conversely, the detector may misidentify objects in certain processes, mistakenly categorizing them as having the same final state as the signal process, e.g. a jet being misidentified as a hadronically decaying τ-lepton. These are called reducible background processes. In the following, the most dominant background contributions in this analysis are discussed.

2.2.1 $Z \to \tau\tau$

Since the Z boson can be produced via the VBF process, the decay $Z \to \tau\tau$ can result in the same final state objects as $H \to \tau\tau$, namely two jets, a pair of τ-leptons and missing transverse mass due to undetectable neutrinos. This process provides the most dominant, irreducible background contribution. An example Feynman diagram of the VBF $Z \to \tau\tau$ process is shown in figure 2.3.

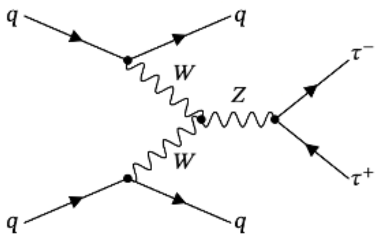


Figure 2.3: Example Feynman diagram of a Z boson that is produced via VBF and decays in a pair of τ-leptons.

Since the mass of the Z boson $M_{\mathrm{Z}} = 91.1876 \pm 0.0021$ GeV [24] differs from the mass of the Higgs boson, these background contributions can be reduced by applying requirements on the reconstructed invariant mass of the di-τ system.

2.2.2 $Z \to ll$

The Z boson can also decay in electrons e or muons μ. In these $Z \to ll$ decays, no missing transverse energy E_T^{miss} occurs. In principle, this can be utilized to suppress the contribution of these processes by requiring a lower limit for E_T^{miss}. However, E_T^{miss} can also be caused by detector deficiencies, such that this requirement only allows a partial reduction of this background. Furthermore, particularly considering the hadronic decay channel of the τ-leptons, automatically reduced the contribution of these background processes. In figure 2.4, an example Feynman diagram for $Z \to ll$ is shown.

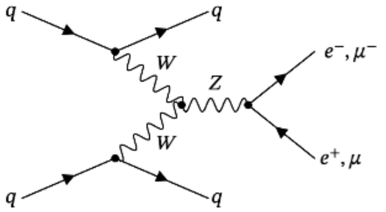
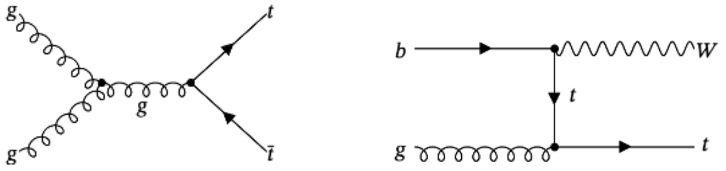


Figure 2.4: Example Feynman diagram of a Z boson that is produced via VBF and decays in a pair of electrons $e^{\pm}$ or a pair of muons $\mu^{\pm}$.

2.2.3 Top Quark Production

In figure 2.5, example Feynman diagrams for the production of a single top quark t or a top-quark pair $t\bar{t}$ are illustrated. Since the t quark can decay both, hadronically and leptonically, these processes can lead to a final state, similar to the VBF Higgs boson final state, that consists of two jets and a pair of τ-leptons.



(a) (b)

Figure 2.5: Example Feynman diagrams for (a) pair top quark production (b) single top quark production.

Since most of the time, t quarks decay in a bottom quark b and a W boson, the contribution of these background processes can be reduced by applying a veto on jets [26] originating from b quarks, utilizing that, due to the long mean lifetime of b-quarks, their corresponding hadrons are characterized by traveling a relatively long distance of a few mm before decaying.

2.2.4 Diboson Production

Since W and Z bosons can decay leptonically and hadronically, di-boson processes, which refer to the production of the boson pairs ZZ, WW or WZ, can result in a final state that includes two leptons and two jets. The most dominant background contribution results from the decay $WW \rightarrow \tau_{\text{had}}\nu_{\tau}\tau_{\text{had}}\nu_{\tau}$, since, like the signal process, they include missing transverse energy. Jets may arise, for instance, through emission of gluons. Example Feynman diagrams for these processes are shown in figure 2.6.

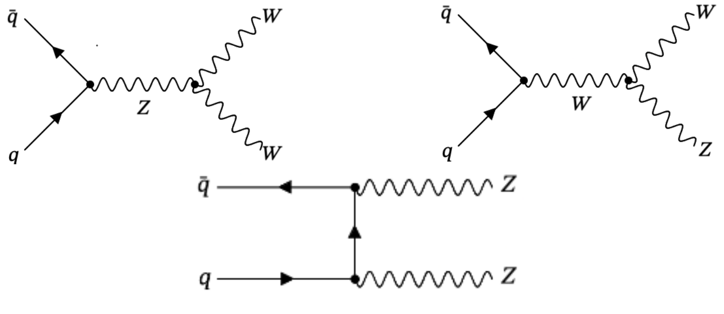


Figure 2.6: Example Feynman diagrams for diboson production.

2.2.5 W Boson Production

The production of W bosons with associated jets that are misidentified as hadronically decaying τ-leptons also contributes to the background. Example diagrams for such processes are shown in figure 2.7

(a) (b)

Figure 2.7: Example Feynman diagrams for W boson production with associated jets.

For the following discussions and figures, the processes that were introduced in this chapter are summarized into the four categories presented in table 2.1. The analysis uses simulated samples that are generated separately for each of these processes using the Monte Carlo (MC) method. The corresponding generators are also shown in table 2.1. For the VBF Higgs boson signal process, the non VBF Higgs processes and the t production processes, the matrix elements and the parton showers are calculated seperately using `POWHEG` [27] and `PYTHIA` [28], respectively. For the other processes, `SHERPA` [29] generates both, the matrix element and the parton shower. Additionally, this table contains the cross-section σ for each process.

Category	Processes	Generator	σ [pb]
VBF Higgs	$H \to \tau\tau$	`POWHEG+PYTHIA`	0.23
Non VBF Higgs	ggF, ttH, WH/ ZH	`POWHEG+PYTHIA`	12.84
$Z \to \tau\tau$	$Z \to \tau\tau$	`SHERPA 2.2.1`	7099.32
Others	$Z \to ll$	`SHERPA 2.2.1`	13457.54
	W boson prod.	`SHERPA 2.2.1`	184597.94
	Diboson prod.	`SHERPA 2.2.1`	125.97
	single t and $t\bar{t}$ prod.	`POWHEG+PYTHIA`	866.07

Table 2.1: Processes relevant for the presented analysis, grouped into four categories. For each process, the corresponding Monte Carlo generators are specified. Furthermore, the corresponding cross-sections σ, which are the same as in [30] are displayed.

Chapter 3

Simulated Event Samples and Event Selection

This section specifies the data that is used in the presented analysis. First, in section 3.1, the considered dataset, consisting of simulated event samples, is introduced. Next, in section 3.1.1, the applied event weights are discussed. Lastly, in section 3.2, the criteria considered in the event selection are reviewed.

3.1 Simulated Event Samples

The presented analysis aims to derive the sensitivities to constrain $\tilde{d}$ with the full data set collected in pp-collisions at a center of mass energy of $\sqrt{s} = 13$ TeV during Run-2 of the LHC corresponding to an integrated luminosity of 139 fb^{-1}.

The simulated samples used for this analysis contain both, events resulting from the VBF $H \to \tau_{\text{had}}\tau_{\text{had}}$ signal process, for which an example Feynman diagram is shown in figure 2.2 and contributions from the background processes introduced in section 2.2. They include parton level information about the four-momenta of the in- and outgoing partons and the Higgs boson, also referred to as truth-level information. Furthermore, they include information that is reconstructed after running the detector simulation [31], referred to as reconstruction-level information. This encompasses the four momenta of the two leading tagging jets with the highest transverse momenta, the four momenta of the τ-leptons, reconstructed only from their visible decay products and the four momentum of the di-τ-system (Higgs boson), fully reconstructed using the missing mass calculator (see section 2.1). If not stated differently, the presented results are obtained with reconstruction level information.

3.1.1 Event Weights

In order to attain a reasonable statistical accuracy, a large number N_{gen} of events needs to be generated in simulations. To adjust this number of events to match the expectation N_{exp} for a measurement at an integrated luminosity $\int Ldt = 139$ fb^{-1} and a given cross-section σ, the MC events need to be rescaled by applying the weights w_i:

$$N_{\text{exp}} = \sum_{i=1}^{N_{\text{gen}}} w_i. \tag{3.1}$$

These event weights are defined as

$$w_i = \cdot w_i^{xs} \cdot w_i^{\text{PU}} \cdot w_i^{\text{corr}} \cdot \left(\int L dt\right) \cdot \underbrace{\text{SF}}_{\text{reco only}} \quad . \tag{3.2}$$

Here, w_i^{xs} are weights that normalize the number of events to the correct cross-section (see table 2.1), where normalization is performed to an integrated luminosity of $\int L dt = 1\ \text{fb}^{-1}$. At the LHC, protons are collided in bunches, which results in multiple simultaneous collisions in every bunch crossing. This effect is called Pile up. Since MC samples are usually generated before data taking, the MC pile-up conditions need to be adjusted to actual pile-up conditions in retrospect. This is achieved by applying pile up weights w_i^{PU}. The weights w_i^{corr}, correct the phase space, for example by considering next-to-leading-order corrections. In case of reconstruction (reco) level events, the scale factor SF, that corrects differences between simulation and data, has to be applied. It is important to consider that ATLAS Run 2 involved multiple data-taking periods or ”campaigns” with different pile-up conditions and varying integrated luminosities. Table 3.1 presents the integrated luminosity for each campaign, that are used in the weight calculations.

Campaign	$\int L dt$ [fb^{-1}]
mc16a (2015/16)	36.207
mc16d (2017)	44.307
mc16e (2018)	58.450
⇒ Full Run 2: $\int L dt =$	139.000

Table 3.1: Integrated luminosities $\int L dt$ during the different campaigns of ATLAS Run 2 [32], that are used to calculate the event weights in eq. (3.2) which, in total, scale the MC events to an integrated luminosity of 139 fb^{-1}.

3.2 Event Selection

In Chapter 2, both, the signal processes and background contributions have been characterized. In this section, the identified characteristics are utilized to impose particular requirements on the analyzed data, aiming to reduce the contribution of background events to the analysis. Here, three different sets of requirements, also referred to as cuts, are considered: a preselection cut, the VBF cut [30] and the NN cut (Daniel Bahner, private communication). The preselection cut aims to reduce background contributions by exploiting kinematic differences between signal and background processes. It also ensures that only events, fully reconstructed with high efficiency by ATLAS are taken into account. The VBF topology cut further tightens these requirements and includes new criteria that allow to select events that show topological characteristics typical for VBF processes. Finally, the neural network cut relies on the predicted probability assigned by a neural network to classify an event as a signal event.

In the following most of the criteria required for each cut are given in terms of the components of the four momenta P^μ of final state constituents, expressed using ATLAS coordinates

$$P^\mu = (P^T, \eta, \phi, M). \tag{3.3}$$

Here, P^T denotes the transverse momentum, the spacial coordinates are given in terms of the pseudorapidity $\eta = -\log(\tan\theta/2)$, which is defined in terms of the polar angle θ and the azimuthal angle ϕ, illustrated in figure 3.1 and M denotes the mass. The subscripts $\tau_{0,1}$ refer to the hadronic decay products of the leading and subleading τ-leptons and $j_{0,1}$ denote the leading and subleading tagging jets.

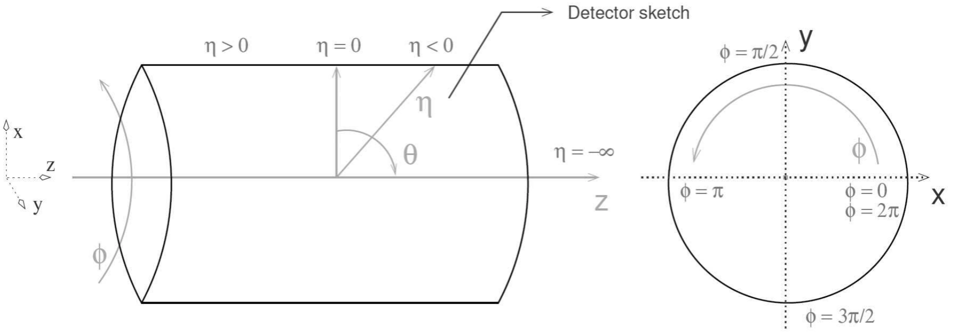


Figure 3.1: ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\log(\tan\theta/2)$ [33].

Preselection cut Candidate τ-leptons are identified as hadronic τ-leptons if they pass a set of criteria, referred to as working point (WP), based on a neural network score [34]. The WP "medium", which ensures an identification efficiency of 55% [34], is used.

Since the Higgs-boson is electrically neutral, τ-lepton candidates are only considered if they have opposite charges.

Furthermore, a pair of τ-leptons is only considered if their spacial separation does not exceed certain thresholds. Specifically, for the difference in pseudorapidities, $\Delta\eta_{\tau\tau} < 1.5$ is required. Their radial distance

$$\Delta R_{\tau\tau} = \sqrt{(\Delta\eta_{\tau\tau})^2 + (\Delta\Phi_{\tau\tau})^2}, \tag{3.4}$$

must fulfill $0.6 \leq \Delta R_{\tau\tau} \leq 2.5$.

Additionally, thresholds for the transverse momenta of the τ-leptons are defined: $P^T_{\tau_0} > 40$ GeV, $P^T_{\tau_1} > 30$ GeV. To only select events resulting from processes whose final state involves neutrinos, a missing transverse energy $E^{\text{miss}}_T > 20$ GeV is required. This allows reducing the contribution from $Z \to ll$ processes. Furthermore, the ratios $x_{0,1}$ of the visible τ-lepton energies and the true τ-lepton energies are restricted to $0.1 \leq x_{0,1} \leq 1.4$. The true τ-lepton energies are obtained through the reconstruction of the collinear mass [35]. This assures that the direction of the missing transverse energy fits the direction expected from the di-τ decay.

Events resulting from the production of top quarks, can be suppressed by applying a b-jet veto (see section 2.2) at the 70% WP [26]. Since the VBF final state involves the hadronisation of two partons, the presence of at least two tagging jets is required. These tagging jets are only considered in the analysis if $P^T_{j_0} > 40$ GeV and $P^T_{j_1} > 30$ GeV. Only the leading and subleading jets, i.e. the jets detected with the highest and second highest P^T are considered in the analysis. Additionally, the invariant mass M_{jj} of these leading and subleading tagging jets is required to exceed 350 GeV. To ensure that only jets which ATLAS can detect with a sufficient precision, are considered, the absolute value of their pseudorapidities must fulfill $|\eta_{j_0}| < 3.2$.

VBF Cut Considering the topology of the VBF process, the tagging jets are required to have a spacial separation $|\Delta\eta_{jj}| > 3$ and they must be detected in different detector hemispheres, i.e. $\eta_{j_0} \times \eta_{j1} < 0$. Additionally, $|\eta_{j_1}| < 3.2$ is required. Another requirement that considers the VBF topology is lepton centrality, which demands that the η coordinates of the visible decay products of the τ-leptons are detected between the two tagging jets with respect to the η coordinates. Furthermore, the threshold for the transverse momentum of the subleading jet is enhanced with respect to the treshhold required in the preselection cut, such that $P^T_{j_1} > 40$ GeV.

Figure 3.2 shows the mass of the di-τ system that was reconstructed using the MMC (MMC Mass) after applying the VBF cut.

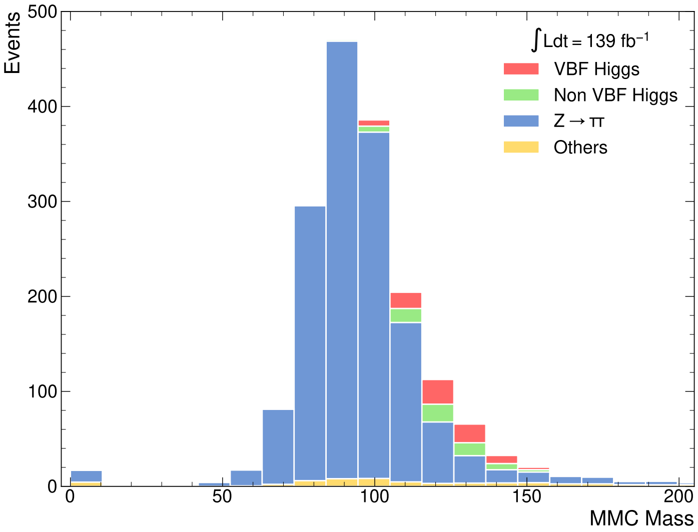


Figure 3.2: Distribution of the mass of the di-τ system that was reconstructed using the MMC (MMC mass), after the VBF cut for an integrated luminosity of $\int \text{Ldt} = 139\text{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2.

The VBF cut, which focuses on topological characteristics of the VBFH signal process, cannot sufficiently reduce $Z \to \tau\tau$ events, since as discussed in section 2.2.1 this process can result in the same final state objects as the signal process.

Neural Network Cut A neural network (NN) is applied to predict the probability or score for each event to be a signal event. The NN cut requires a score ≥ 0.92.

Figure 3.3 illustrates the distribution of this NN score after applying the VBF cut. Here, the NN cut is represented by the dashed line at 0.92.

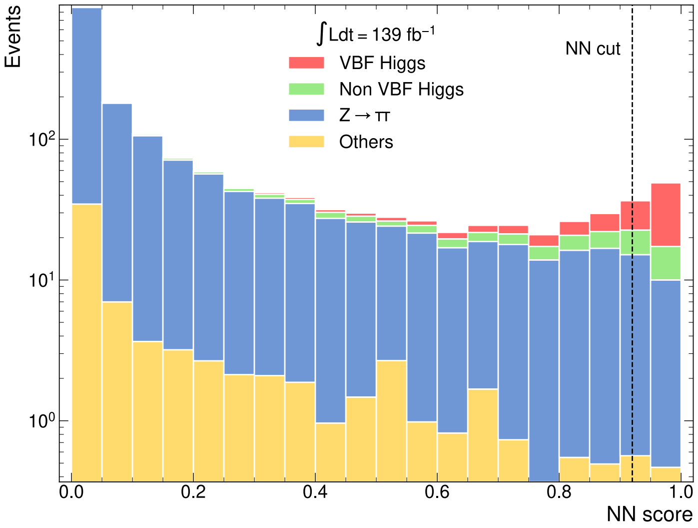


Figure 3.3: Neural network score after the VBF cut for an integrated luminosity of $\int \mathrm{Ldt} = 139\mathrm{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. The dashed line represents the NN cut at a score of 0.92.

To illustrate the effects of the NN cut, the distribution of the Optimal Observable ($\mathcal{OO}$) (see section 5.1) is shown in figure 3.4 using only events that fulfill the requirement of a score≥ 0.92. Especially, the contributions of $Z \to \tau\tau$ processes is reduced significantly by the NN cut. Based on the NN score illustrated in figure 3.3, this outcome is expected, given that the NN demonstrates strong performance in classifying these processes as background contributions.

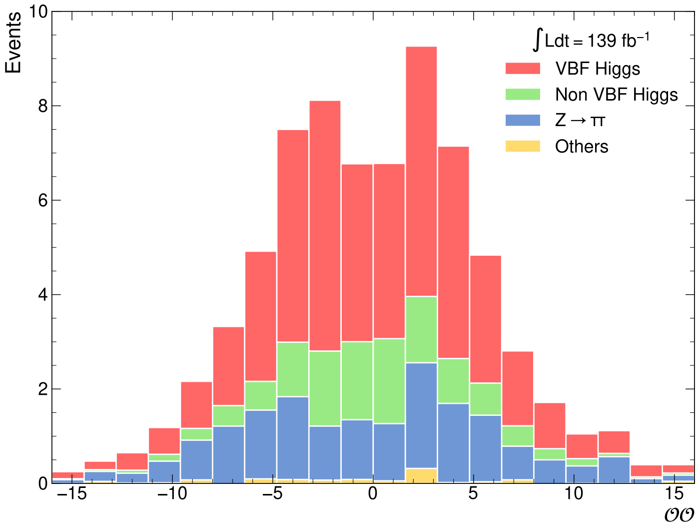


Figure 3.4: Distribution of the Optimal Observable $\mathcal{OO}$ (see section 5.1) after the NN cut for an integrated luminosity of $\int$ Ldt $= 139$ fb^{-1}. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2.

Additonal Cuts In addition to the cuts defined above, it is required that the di-τ mass, reconstructed with the missing mass calculator, also referred to as MMC mass, fulfills

$$\text{MMC Mass} > 1 \text{ GeV}. \tag{3.5}$$

This assures, that only events, where the Higgs boson mass could be properly reconstructed, are used. Furthermore, only events for which the Optimal Observable $\mathcal{OO}$, introduced in section 5.1, and calculated at reconstruction level, satisfies

$$|\mathcal{OO}| \leq 15. \tag{3.6}$$

are considered. This requirement allows avoiding outlier events.

Table 3.2 presents the remaining number of events after the application of each cut, for the VBF Higgs boson signal s, for the total background contribution b as well as for the individual contributions to b, grouped according to table 2.1. Furthermore, it contains the signal significance

$$\sigma = \frac{s}{\sqrt{s+b}}, \tag{3.7}$$

and the signal to background ratio s/b after each cut.

Cut	s	b	Non VBF Higgs	$Z \to \tau\tau$	Others	σ	s/b
Preselection cut	196	9624	246	8929	449	1.98	0.02
VBF cut	84	1692	68	1555	69	1.99	0.05
NN cut	41	32	12	19	1	4.80	1.28
MMC Mass > 1 GeV & $\lvert\mathcal{OO}\rvert \leq 15$	40	29	11	17	1	4.82	1.38

Table 3.2: Expected number s of VBF Higgs boson signal events and numbers b of total expected background events in 139 fb^{-1}, after each cut. The individual contributions of each dominant background processes to b are also shown, grouped according to table 2.1. Additionally, the corresponding signal significance σ and the signal to background ratio s/b are displayed.

The application of full event selection allows increasing the signal significance by a factor of 2.43 with respect to the significance obtained after the preselection. The signal to background ratio is improved by a factor of 69. The decay $Z \to \tau\tau$ remains to be the dominant background contribution after all cuts, however, the neural network cut allowed reducing its contribution by 98.9 %.

Chapter 4

Testing CP-Invariance with CP-Odd Observables

According to our current understanding, during the early moments of the universe, an equal amount of baryonic and anti-baryonic matter was created. However, since baryons and antibaryons annihilate, this scenario would have resulted in a universe where objects like stars, planets and ultimately all structures composed of baryonic matter that we observe today, could not have formed.

The observed baryon asymmetry of the universe (BAU) can be parameterized via the ratio η, which relates the excess of baryonic matter with the photon desity n_γ of our universe

$$\eta = \frac{n_B - n_{\bar{B}}}{n_\gamma} \sim 6 \cdot 10^{-10}. \tag{4.1}$$

Here, n_B and $n_{\bar{B}}$ denote the baryon and anti-baryon density, respectively. Notably, this value η, which was calculated using recent measurements of the average energy density of baryons made by the Planck satellite [36], represents a remarkably subtle imbalance, yet its implications are significant for the existence of the universe in its present form.

The favored production of baryons over anti-baryons can only be explained, if three conditions, formulated by A.D. Sakharov in 1967 [12] are met. One of these conditions requires the existence of processes that violate invariance under the combination of charge conjugation and parity transformation (CP). However, the Standard Model (SM) of particle physics, which is the current best theory to describe all fundamental particles and their interactions, excluding gravity, does not predict a sufficient amount of CP-violation to explain the observed baryon asymmetry. This motivates the search for new sources of CP-violation beyond the standard model (BSM).

The presented analysis focuses on the search for BSM sources of CP-violation in the Higgs sector, specifically in the process of VBF Higgs-boson production with $H \to \tau_\mathrm{had}\tau_\mathrm{had}$ that was introduced in section 2.1.

In this chapter, the approach of probing the CP-properties of the VBF Higgs-boson production process by utilizing CP-odd observables, within the framework of an effective field theory, is elaborated. First, in section 4.1 the concept of CP-symmetry is introduced. Next, in section 4.2 the CP-conserving couplings of electroweak gauge bosons to the Higgs-boson (HVV-couplings) predicted by the SM are augmented by introducing new, CP-violating interactions within the framework of an effective field theory. Section 4.3 describes how CP-odd observables can be used to test the CP-invariance of an interaction. One approach to evaluate the sensitivity of CP-odd observables to CP-violating contributions is the construction of central confidence intervals via the maximum likelihood (ML) method, which is introduced in section 4.4.

4.1 CP-Symmetry

In physics, a symmetry refers to a fundamental property of a physical system that remains unchanged under certain operations.

Parity ($\boldsymbol{P}$) - Symmetry Parity-symmetry refers tho the invariance of a physical theory under the inversion of the sign of all spacial coordinates. In the SM, the parity-symmetry is conserved for all interactions except the weak interaction. This was demonstrated in 1956 by C.S.Wu, by investigating the beta decay of ^{60}CO [37].

Charge Conjugation ($\boldsymbol{C}$) - Symmetry The charge conjugation operator $\boldsymbol{C}$ reverses all additive quantum numbers of a particle ψ, such that it is transformed into its anti-particle $\bar{\psi}$. In the SM, charge conjugation is conserved for all interactions except the weak interaction [38]. For example, the C operation would turn a left-handed neutrino into a left-handed anti-neutrino. However, in nature, only right-handed antineutrinos are observed [39].

CP-Symmetry For almost all processes via the weak interaction, where C or P symmetries are violated individually, the combination CP of the two transformations is conserved. For example, the CP operation transforms a left-handed neutrino into a right-handed antineutrino and thus resolves the symmetry breaking that is observed for neutrinos under pure C and P transformations. However, in 1964 the Croin and Fitch experiment demonstrated the violation of CP-symmetry in neutral Kaon decays [40]. This was followed by the discovery of CP-violation in various other meson-interactions [41, 42].

In the SM, CP-violation is explained through the complex phase of the Cabibbo-Kobayashi-Maskawa (CKM) matrix [13]. However, the amount of CP-violation resulting from this SM source is not sufficient to explain the observed baryon asymmetry in our universe [43]. This motivates the search for new sources of CP-violation beyond the SM.

4.2 Effective Field Theories

Effective field theories (EFTs) provide a framework that allows to parameterize new physics beyond the standard model (BSM). Current experiments are limited to observing a low-energy limit, where the particles corresponding to the new physics, which are expected to emerge at higher energies, remain unmeasurable. Thus, the only tangible effects of new physics are alterations in the observable interactions. EFTs allow including the corresponding new interaction terms into the theory.

New BSM coupling terms are introduced by operators with mass dimensions higher than the mass dimension $D = 4$ of the SM Lagrangian density $\mathcal{L}_{\mathrm{SM}}$. Only operators that are invariant under the local gauge transformations $SU(2)_{I,L} \times U(1)_Y$ are considered. Additionally, conservation of baryon and lepton numbers is required. This suppresses all operators with odd mass dimensions.

Under these assumptions, an effective Lagrangian density can be constructed as follows [17, 44]

$$\mathcal{L}_{\mathrm{eff}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{c_i^{(6)}}{\Lambda^2} \mathcal{O}_i^{(D=6)} + \sum_i \frac{c_i^{(8)}}{\Lambda^4} \mathcal{O}_i^{(D=8)} + ... \tag{4.2}$$

where $\mathcal{O}_i^{(D)}$ are the considered operators, the parameters $c_i^{(D)}$ denote the Wilson coefficients and Λ represents the energy scale of new physics [17, 22]. Since operators with mass dimension $D \geq 8$ are suppressed for large Λ, in the following, they will be neglected. Furthermore, only CP-odd operators are considered.

As suggested in [17, 44] after electroweak symmetry breaking, considering the assumptions made above, the effective Lagrangian that describes the coupling of electroweak gauge bosons to the Higgs-boson (HVV coupling) can be written in the mass basis of the Higgs-boson H, the photon A and the weak gauge bosons Z, W^+ and W^- as

$$\mathcal{L}_{\mathrm{eff}} = \mathcal{L}_{\mathrm{SM}} + \tilde{g}_{HAA} H \tilde{A}_{\mu\nu} A^{\mu\nu} + \tilde{g}_{HAZ} H \tilde{A}_{\mu\nu} Z^{\mu\nu} + \tilde{g}_{HZZ} H \tilde{Z}_{\mu\nu} Z^{\mu\nu} + \tilde{g}_{HWW} H \tilde{W}^{+}_{\mu\nu} W^{-\mu\nu}. \tag{4.3}$$

Here, $\hat{V}^{\mu\nu}$, $\tilde{\hat{V}}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} V_{\rho\sigma}$ represent the field-strength tensors/ dual field strength tensors of the gauge fields with $V \in \{A, Z, W^+, W^-\}$. Requiring $SU(2)_{I,L} \times U(1)_Y$ invariance allows expressing the HVV-coupling strengths $\tilde{g}_{\mathrm{HVV}}$ in eq. (4.3) in terms of two dimensionless parameters $\tilde{d}$ and $\tilde{d}_B$:

$$\begin{aligned}
\tilde{g}_{HAA} &= \frac{g}{2m_W}(\tilde{d}\sin^2\theta_W + \tilde{d}_B\cos^2\theta_W), \\
\tilde{g}_{HAZ} &= \frac{g}{2m_W}\sin 2\theta_W(\tilde{d} - \tilde{d}_B), \\
\tilde{g}_{HZZ} &= \frac{g}{2m_W}(\tilde{d}\cos^2\theta_W + \tilde{d}_B\sin^2\theta_W), \\
\tilde{g}_{HWW} &= \frac{g}{m_W}\tilde{d},
\end{aligned} \tag{4.4}$$

where g and g' are coupling strengths and the weak mixing angle θ_W is defined by $\tan\theta_W = \frac{g'}{g}$ [45].

Since experimentally, the contributions of W^+W^-, ZZ, AZ and AA fusion to VBF Higgs-production are indistinguishable [17], the arbitrary choice

$$\tilde{d} = \tilde{d}_B, \tag{4.5}$$

is made. This yields

$$\tilde{g}_{HAA} = \tilde{g}_{HZZ} = \frac{1}{2}\tilde{g}_{\mathrm{HWW}} = \frac{g}{2m_W}\tilde{d}, \text{ and } \tilde{g}_{HAZ} = 0. \tag{4.6}$$

Thus, the strength of CP-violation in the HVV coupling can be described by a single parameter $\tilde{d}$ [17, 22, 44]. The matrix element $\mathcal{M}$ of the VBF-Higgs-boson production process can then be decomposed in a CP-even part that results from the SM contributions and a CP-odd part, resulting from BSM couplings, whose contribution is governed by the magnitude of $\tilde{d}$

$$\mathcal{M} = \underbrace{\mathcal{M}_{\mathrm{SM}}}_{\mathrm{CP-even}} + \tilde{d} \cdot \underbrace{\mathcal{M}_{\mathrm{BSM}}}_{\mathrm{CP-odd}}. \tag{4.7}$$

Squaring this matrix element yields the following expression, which is proportional to the differential cross-section $d\sigma$

$$|\mathcal{M}|^2 = \underbrace{|\mathcal{M}_{\mathrm{SM}}|^2}_{\mathrm{CP-even}} + \underbrace{\tilde{d} \cdot 2\Re\{\mathcal{M}^*_{\mathrm{SM}}\mathcal{M}_{\mathrm{BSM}}\}}_{\text{CP-odd, source of CPV}} + \underbrace{\tilde{d}^2 \cdot |\mathcal{M}_{\mathrm{BSM}}|^2}_{\mathrm{CP-even}} \quad \sim \quad d\sigma. \tag{4.8}$$

The first and third terms in eq. (4.8) are CP-even and thus do not contribute to CP-violation. However, the interference term is CP-odd and therefore provides a source of CP-violation. In the presented analysis, the CP-odd contribution of this term will be utilized to define CP-odd observables for testing CP-invariance in the process of VBF Higgs-boson production.

4.3 CP-Odd Observables

A genuine CP-odd observable $\mathcal{O}$ is defined by having the following property under a CP-transformation

$$\mathbf{CP}\mathcal{O} = -\mathcal{O}. \tag{4.9}$$

In the effective field theory framework that was introduced in section 4.2, CP-violation yields two additional contributions in the matrix element $\mathcal{M}$ in eq. (4.8) compared to the pure SM prediction. The first new term is CP-odd and linear in the parameter $\tilde{d}$, whereas the second term is CP-even and quadratic in $\tilde{d}$. In this framework, the mean value of an observable $\mathcal{O}$, which is obtained by integrating over the cross-section $d\sigma \sim |\mathcal{M}|^2$, can be written as

$$\langle\mathcal{O}\rangle = \frac{\int \mathcal{O} d\sigma_{\mathrm{SM}} + \tilde{d} \cdot \int \mathcal{O} d\sigma_{\mathrm{CP-odd}} + \tilde{d}^2 \int \mathcal{O} d\sigma_{\mathrm{CP-even}}}{\int d\sigma_{\mathrm{SM}} + \tilde{d} \cdot \int d\sigma_{\mathrm{CP-odd}} + \tilde{d}^2 \int d\sigma_{\mathrm{CP-even}}}. \tag{4.10}$$

In case of a CP-odd observable, the integrals over the CP-even phase spaces $d\sigma_{\mathrm{SM}}$ and $d\sigma_{\mathrm{CP-even}}$ in the numerator of eq. (4.10), will vanish. In the denominator, the integral over $d\sigma_{\mathrm{CP-odd}}$ vanishes. The mean value of a CP-odd observable can then be written as

$$\langle \mathcal{O} \rangle = \frac{\tilde{d} \cdot \int \mathcal{O} d\sigma_{\mathrm{CP-odd}}}{\int d\sigma_{\mathrm{SM}} + \tilde{d}^2 \int d\sigma_{\mathrm{CP-even}}}. \tag{4.11}$$

This indicates that if CP-violation is present in nature, realized through a strength $\tilde{d} \neq 0$, asymmetries are induced in the distributions of CP-odd observables, leading to non-zero mean values. This behavior of CP-odd observables is demonstrated in figure 4.1 for two CP-violating scenarios, corresponding to $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$. For $\tilde{d} > 0$, a shift of the distribution towards positive values can be observed, resulting in a mean $\langle \mathcal{O} \rangle > 0$, whereas negative $\tilde{d}$ lead to a skew in negative direction and thus $\langle \mathcal{O} \rangle < 0$. Consequently, measuring a mean value $\langle \mathcal{O} \rangle \neq 0$ for a CP-odd observable would be a strong sign for CP-violation.

Furthermore, it is worth noting that in case of small $\tilde{d}$-values, the quadratic term in the denominator in eq. (4.11) can be neglected, resulting in a linear dependency between $\tilde{d}$ and $\langle \mathcal{O} \rangle$.

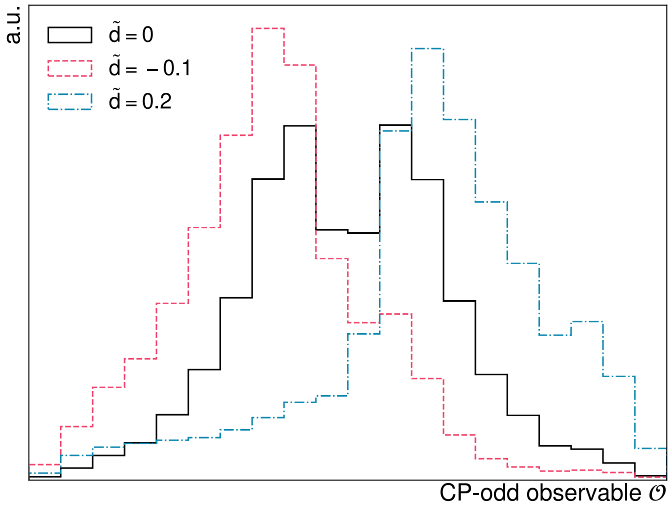


Figure 4.1: Illustration of the characteristic behavior of a CP-odd observable $\mathcal{O}$. In the SM prediction ($\tilde{d}$=0), represented by the black histogram, the distribution is symmetric and yields a mean value $\langle \mathcal{O} \rangle = 0$. Introducing CP-violation induces asymmetries in the distribution that are governed by the sign and magnitude of $\tilde{d}$. This is demonstrated for the scenario $\tilde{d} = -0.1$ (red histogram) which yields $\langle \mathcal{O} \rangle < 0$ and $\tilde{d} = 0.2$ (blue histogram), where $\langle \mathcal{O} \rangle > 0$.

4.4 Evaluating CP-Sensitivity via Maximum Likelihood Fits

In this thesis, the sensitivity of CP-odd observables to CP-violating BSM contributions in the HVV coupling (see section 4.2) is estimated using the maximum likelihood (ML) method for binned data [46], which allows determining the expected confidence intervals for the parameter $\tilde{d}$, using only information about the shape of the distributions of these observables.

Considering a histogram, consisting of N_{bins} bins, with the prediction $\nu_i(\tilde{d})$ for the contents of the i-th bin for a given BSM scenario, corresponding to a $\tilde{d} \neq 0$ and observations n_i, corresponding to the contents of this bin expected from the SM ($\tilde{d} = 0$), a likelihood function can be introduced in form of a multinomial distribution [46]

$$L(\tilde{d}) = N! \prod_{i}^{N_{\text{bins}}} \left(\frac{\nu_i(\tilde{d})}{N} \right)^{n_i} \frac{1}{n_i!}, \tag{4.12}$$

where N denotes the total number of observed events to which the predicted histogram is normalized:

$$\sum_{i=1}^{N_{\text{bins}}} n_i = N = \sum_{i=1}^{N_{\text{bins}}} \nu_i(\tilde{d}). \tag{4.13}$$

The goal is to find the best estimate for $\tilde{d}$. This translates into maximizing the likelihood function eq. (4.13) or, equivalently, minimizing the negative log likelihood, which is defined as

$$\text{NLL}(\tilde{d}) = -\log L(\tilde{d}) \quad = -\sum_{i=1}^{N} n_i \log \nu_i(\tilde{d}). \tag{4.14}$$

The additional terms that appear, when calculating the logarithm of eq. (4.13) do not depend on $\tilde{d}$. For this reason, they are neglected in eq. (4.14).

As described in [46], generally, a change of N' standard deviations σ in a single best-estimate parameter $\hat{\theta}_i$ yields an increase of $N'^2/2$ in the NLL, relative to its minimum

$$\Delta\text{NLL} = \text{NLL}(\hat{\theta}_i \pm N'\sigma) - \text{NLL}(\hat{\theta}_i) = \frac{N'^2}{2}. \tag{4.15}$$

The parameters $\theta^{\pm}_{1\sigma} = \hat{\theta} \pm (N' = 1)\sigma$, for which eq. (4.15) yields $\Delta\text{NLL} = 0.5$, define the borders of the 1σ confidence interval. Accordingly, the borders of the 2σ confidence interval are given by $\theta^{\pm}_{2\sigma} = \hat{\theta} \pm (N' = 1.96)\sigma$, for which eq. (4.15) in NLL becomes $\Delta\text{NLL} = 1.92$.

This is now utilized to construct the confidence intervals for the best estimate of the parameter $\tilde{d}$. For this, different $\tilde{d} \neq 0$ hypothesis are considered. The NLL is evaluated for each hypothesis according to eq. (4.14). This methodology is also refereed to as a NLL-scan. The values of the NNL-scan can be plotted as a function of $\tilde{d}$, resulting in a curve with its minimum at the best fit-value. Since in this case, the observation corresponds to the SM prediction, this minimum will occur at $\tilde{d} = 0$. Calculating the difference between this minimum and the results for all remaining NLL-values yields a curve, similar to the one portrayed in figure 4.2 serving as an arbitrary illustration, which allows to directly read off the 1σ and 2σ CIs according to eq. (4.15):

$$\Delta\text{NLL} = \text{NLL}(\tilde{d}) - \text{NLL}(\tilde{d} = 0) \overset{!}{=} \begin{cases} 0.5, & \text{for the } 1\sigma \text{ CI },\\ 1.92, & \text{for the } 2\sigma \text{ CI }. \end{cases} \tag{4.16}$$

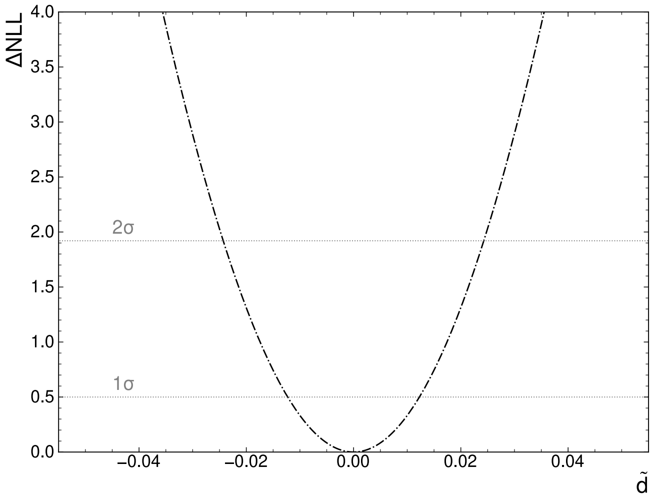


Figure 4.2: Illustration of a ΔNLL-curve that can be obtained by calculating the NLL for different $\tilde{d}$-hypotheses and allows to directly read off the 1σ and 2σ confidence intervals at ΔNLL = 0.5 and ΔNLL=1.92, respectively.

The 1σ (2σ) confidence intervals, by definition, cover the true value in 68.3% (95%) of all cases. Consequently, the length of these intervals, derived from a specific CP-odd observable distribution, acts as a metric for the sensitivity to CP-violating contributions of that observable.

4.5 Matrix Element based Reweighting

To generate the distributions of CP-odd observables for different CP-violating $\tilde{d} \neq 0$ hypotheses, a matrix element based reweighting technique is applied to the simulated samples introduced in chapter 3.

This approach utilizes the matrix element defined in eq. (4.8), where additional contributions due to the CP-violating HVV couplings, introduced in section 4.2, occur with a strength $\tilde{d}$. The information about these new couplings can be incorporated in the observable distributions by reweighing the VBF $H \to \tau_{\text{had}}\tau_{\text{had}}$ signal samples using

$$\hat{w}_{\tilde{d}} = 1 + \tilde{d} \cdot \frac{2\Re\{\mathcal{M}_{\text{SM}}^{*}\mathcal{M}_{\text{BSM}}\}}{|\mathcal{M}_{\text{SM}}|^2} + \tilde{d}^2 \cdot \frac{|\mathcal{M}_{\text{BSM}}|^2}{|\mathcal{M}_{\text{SM}}|^2}. \tag{4.17}$$

Since in this analysis the background processes are assumed to be CP-invariant, they are not reweighted but only considered in the SM prediction.

The matrix elements defining the weights $\hat{w}_{\tilde{d}}$ are calculated using HAWK [47, 48] and depend on the four momenta of incoming and outgoing partons, as well as the Higgs-boson four momentum

$$\mathcal{M} = \mathcal{M}\left(P^{\mu}_{\text{in}}, P^{\mu}_{\text{out}}, P^{\mu}_{\text{Higgs}}\right). \tag{4.18}$$

The complete weight of a given MC event e_i in some $\tilde{d}$ scenario is then given by

$$w_{\tilde{d}} = w_i \cdot \hat{w}_{\tilde{d}}, \tag{4.19}$$

where w_i are the nominal event weights defined in section 3.1.1, such that in case of the SM prediction, where $\tilde{d} = 0$ eq. (4.19) reduces to

$$w_{\text{SM}} := w_{\tilde{d}=0} = w_i. \tag{4.20}$$

Chapter 5

Established CP-Odd Observables

The Optimal Observable ($\mathcal{OO}$) and the signed difference in the azimuthal angles ($\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$) of the two tagging jets that emerge in the process of VBF Higgs-boson production, are commonly used observables for performing CP-Tests [49]. This chapter introduces these two CP-sensitive observables. Further, in this chapter a first demonstration of estimating the CP-sensitivity via Negative Log Likelihood (NLL) scans, as it was introduced in section 4.4, is provided for the examples of $\mathcal{OO}$ and $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$.

5.1 Optimal Observable $\mathcal{OO}$

The Optimal Observable [50, 51], is defined as the ratio of the interference term and the squared SM matrix element introduced in eq. (4.8), both calculated using HAWK [47, 48]

$$\mathcal{OO} = \frac{2\Re\{\mathcal{M}_{\mathrm{SM}}^{*}\mathcal{M}_{\mathrm{BSM}}\}}{|\mathcal{M}_{\mathrm{SM}}|^2}. \tag{5.1}$$

By construction, for small $\tilde{d}$ values, when the quadratic term in eq. (4.8) is negligible, this observable contains the full information on the seven dimensional phase space that describes the final state of the VBF Higgs-boson production mode, making it highly sensitive to the CP-structure of this interaction.

In this section, reconstruction level information is used to calculate the $\mathcal{OO}$. As explained in section 4.5, the matrix elements depend on the four momenta of incoming and outgoing partons $P_{\mathrm{in}}^{\mu}, P_{\mathrm{out}}^{\mu}$, as well as the Higgs-boson four momentum P_{in}^{μ}. Since, experimentally and thus in the case of simulated samples at reconstruction level, the full parton information is not directly accessible, for P_{out}^{μ}, the four momenta $P_{j_0}^{\mu}, P_{j_1}^{\mu}$ of the leading and subleading tagging jets, resulting from the hadronisation of the final state quarks, are used. The kinematic information about the incoming partons is derived from energy and three-momentum conservation. The corresponding four momenta are

$$P_{\mathrm{in}}^{\mu} = x_{0,1}\frac{\sqrt{s}}{2}(1, 0, 0, \pm 1), \tag{5.2}$$

where $x_{0,1}$ denote the fractions of momenta carried by the two incoming partons

$$x_{0,1} = \frac{M_{\mathrm{H_{final}}}}{\sqrt{s}} e^{\pm y_{H_{\mathrm{final}}}}, \tag{5.3}$$

with the invariant mass M_{final} and the rapidity $y_{H_{\mathrm{final}}}$ of the final state, consisting of the vectorial sum of the tagging jets and the Higgs boson. As discussed in section 2.1, the Higgs-boson information P^{μ}_{Higgs} is reconstructed using the Missing Mass Calculator [25].

Figure 5.1 shows the distribution of the Optimal Observable as predicted by the SM and for two additional cases with $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$. The predictions for these CP-violating scenarios are obtained by reweighting the SM signal samples according to the method introduced in section 4.5. As described in section 3.2, events are considered only if $|\mathcal{OO}_{\mathrm{org}}| \leq 15$. The distribution is normalized as follows

$$\mathcal{OO} = \frac{\mathcal{OO}_{\mathrm{org}}}{15}. \tag{5.4}$$

The SM prediction yields a symmetric distribution, whereas introducing CP-violation causes asymmetries, controlled in direction and strength by the sign and the magnitude of $\tilde{d}$. Positive $\tilde{d}$-values cause asymmetries towards positive values and vice versa.

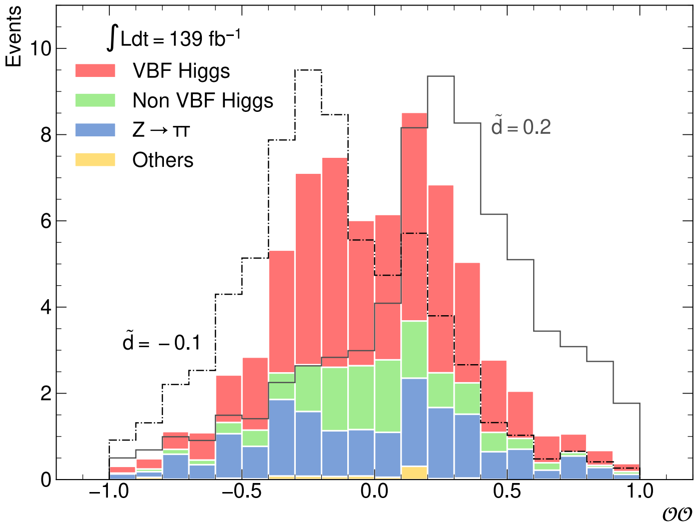


Figure 5.1: Distribution of the Optimal Observable $\mathcal{OO}$ as predicted by the SM for an integrated luminosity of $\int \mathrm{Ldt} = 139\ \mathrm{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. Additionally, distributions predicted for two CP-violating scenarios with the strengths $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$ are presented. These distributions are normalized to the total number $N_{\mathrm{SM}} = 69$ of expected SM events.

The asymmetry induced by CP-violation then results in non-vanishing mean values $\langle \mathcal{OO} \rangle$.

For small values of $\tilde{d}$, the quadratic term in eq. (4.11) can be neglected, and the mean value of a CP-odd observable is expected to depend linearly on $\tilde{d}$. As demonstrated in figure 5.2, this behavior can be observed for the $\mathcal{OO}$.

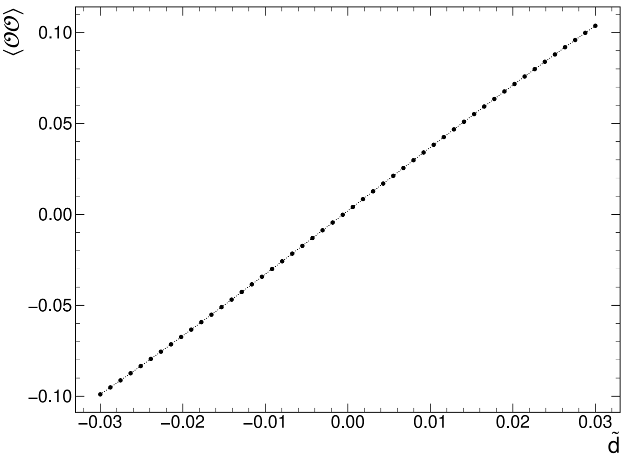


Figure 5.2: Linear dependency between the mean value of the Optimal Observable $\mathcal{OO}$ and $\tilde{d}$ in case of small $\tilde{d}$-values. Only signal events are considered.

To estimate the CP-sensitivity of the Optimal Observable, a Negative Log Likelihood (NLL) scan in $\tilde{d}$, as introduced in section 4.4 is performed. For this purpose, the NLL is calculated considering different CP-violating hypotheses corresponding to 200 $\tilde{d}$-values within the range [-0.05, 0.05]. The choice of this specific range is based on the constraints imposed on $\tilde{d}$ in recent ATLAS analyses [52, 53]. The scan is performed in evenly spaced steps of 0.0005, resulting in an uncertainty of

$$\Delta b = 0.0005, \tag{5.5}$$

on the interval borders b. When calculating interval lengths ℓ, this error propagates as follows

$$\Delta \ell = \sqrt{2}\Delta b = 0.0007. \tag{5.6}$$

The distribution of all considered CP-odd observables is divided into 20 bins, for evaluating the NLL eq. (4.14). The resulting ΔNLL curves for the Optimal Observable are shown in figure 5.3. To obtain the borders of the 1σ interval, linear fits between the two points closest to ΔNLL=0.5 are performed using the Python function `scipy.optimize.curve_fit`. This interpolation then allows calculating the two $\tilde{d}$-values, for which ΔNLL=0.5 is fulfilled. Analogously, the 2σ interval borders are obtained by finding the $\tilde{d}$-values, for which ΔNLL=1.92. The advantage of this procedure, compared to simply extracting the interval borders via parabolic fits, is that it provides sensitivity to possible asymmetries in the ΔNLL curves. Table 5.1 displays the resulting confidence intervals, as well as their lengths, as a measure of CP-sensitivity. The results are shown separately for the case where only signal events are considered and the case where also background events are included. As all considered background interactions are assumed to be CP-invariant and do not depend on $\tilde{d}$, their inclusion in the NLL calculations results in a decrease in CP-sensitivity. This explains the increase in confidence interval lengths by 37% at 1σ confidence level and by 39% at 2σ confidence level that is observed when including background events compared to only considering events from the VBFH signal.

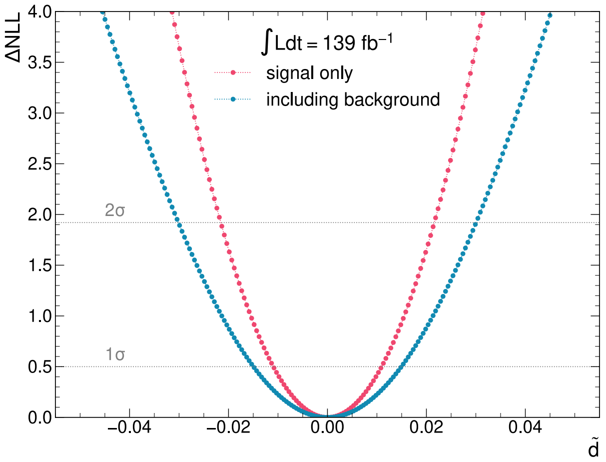


Figure 5.3: ΔNLL curves and resulting 1σ and 2σ confidence intervals, obtained with the Optimal Observable, shown for the case where only signal events are considered (red) as well as for the case where background events are included (blue).

	1σ CI	1σ CI length	2σ CI	2σ CI length
Signal Only	[-0.0109, 0.0109]	0.0218	[-0.0216, 0.0216]	0.0432
Including Background	[-0.0150, 0.0149]	0.0299	[-0.0301, 0.0300]	0.0601

Table 5.1: 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from ΔNLL scans on the Optimal Observable distribution. Results are shown separately for the case where only signal events are considered and for the case where background events are included. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths of $\Delta \ell = 0.0007$.

5.2 $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$

Another CP-sensitive observable, first proposed in [17], can be constructed as the signed difference in the azimuthal angles Φ of the two outgoing tagging jets, described in section 2.1

$$\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}} = \Phi_{j+} - \Phi_{j-}. \tag{5.7}$$

The sign in the subscript denotes, whether the jet is located in the positive (+) or negative (-) detector hemisphere. Thus, as already considered in the event selection defined in section 3.2, only events with jets detected in different hemispheres ($\eta_{j_0}\eta_{j_1} < 0$) can be taken into account when calculating $\Delta\Phi_{\text{jj}}^{\text{sgn}}$.

Figure 5.4 shows the distribution of $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ as predicted by the SM and additionally for two CP-violating scenarios with $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$. Physically, $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ is distributed within the range $[-\pi, \pi]$. Here, the distribution is normalized as follows

$$\Delta\Phi_{\text{jj}}^{\text{sgn}} = \frac{(\Delta\Phi_{\text{jj}}^{\text{sgn}})_{\text{org}}}{\pi}. \tag{5.8}$$

For negative $\tilde{d}$-values, CP-violation manifests in $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ as a preference of negatively signed azimuthal angles between the two jets, whereas for positive $\tilde{d}$ a preference towards positive values is observed.

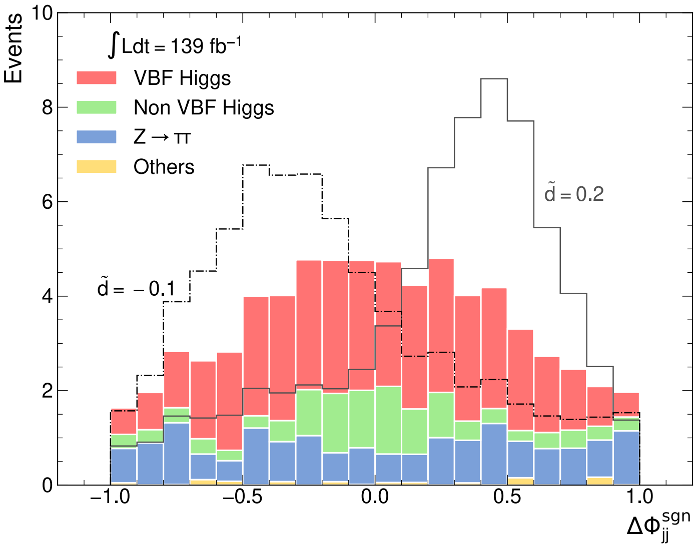


Figure 5.4: Distribution of $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ as predicted by the SM for an integrated luminosity of $\int \text{Ldt} = 139\ \text{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. Additionally, distributions predicted for two CP-violating scenarios with the strengths $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$ are presented. These distributions are normalized to the total number $N_{\text{SM}} = 69$ of expected SM events.

Figure 5.5 illustrates that in the case of small $\tilde{d}$-values, the mean values of $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ in different BSM scenarios show a linear dependency on $\tilde{d}$, as it is expected for CP-odd observables from eq. (4.11).

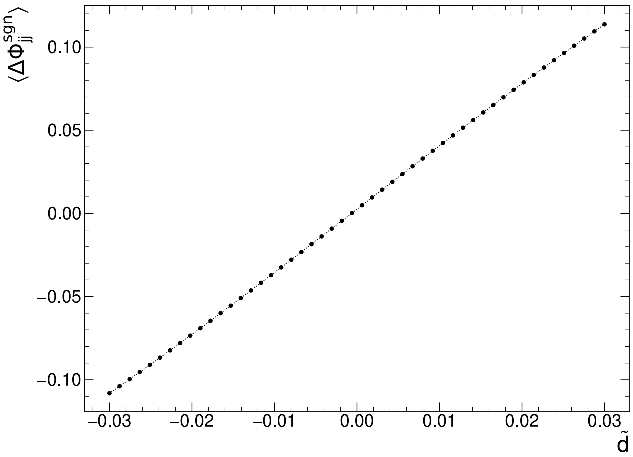


Figure 5.5: Linear dependency between the mean value of $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ and $\tilde{d}$ in case of small $\tilde{d}$-values. Only signal events are considered.

The CP-sensitivity of $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ is estimated with the same procedure as for the Optimal Observable in section 5.1. As before, hypotheses for 200 different $\tilde{d}$-scenarios, taken in equally spaced steps of 0.0005 from the range [-0.05, 0.05], are considered in the NLL scan, leading to the uncertainty $\Delta b = 0.0005$ for the confidence interval borders and $\Delta\ell = 0.0007$ for the confidence interval lengths. The resulting ΔNLL curves are shown in figure 5.6 for the case where only signal events are considered, as well as the case where background events are included in the calculation. In table 5.2, the extracted 1σ and 2σ confidence intervals and their lengths are displayed. Including background events results in an increase in confidence interval lengths by 28% at 1σ confidence level and by 30% at 2σ confidence level compared to only considering events from the VBFH signal.

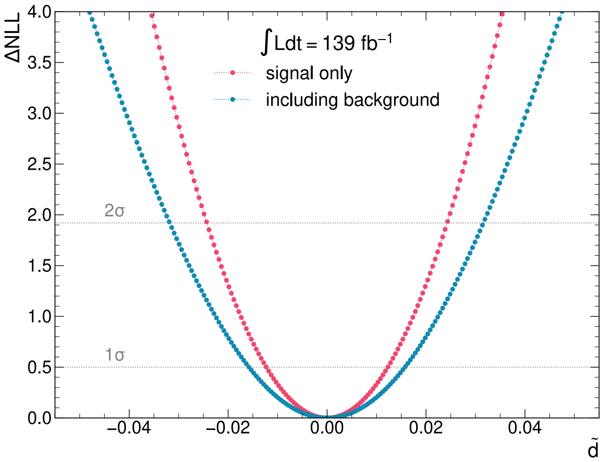


Figure 5.6: ΔNLL curves and resulting 1σ and 2σ confidence intervals, obtained with the $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ distribution, shown for the case where only signal events are considered (red) as well as for the case where background events are included (blue).

	1σ CI	1σ CI length	2σ CI	2σ CI length
Signal Only	[-0.0123, 0.0123]	0.0246	[-0.0243, 0.0243]	0.0486
Including Background	[-0.0158, 0.0157]	0.0315	[-0.0318, 0.0316]	0.0634

Table 5.2: 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from ΔNLL scans on the $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ distribution Results are shown separately for the case where only signal events are considered and for the case where background events are included. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths of $\Delta\ell = 0.0007$.

Comparing these results for $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ with the findings presented in table 5.1 for the $\mathcal{OO}$ shows that $\mathcal{OO}$ provides a higher sensitivity to CP-violating effects. Specifically, in the signal only case, the $\mathcal{OO}$ yields a decrease in interval lengths by 11% at both confidence levels, compared to the lengths obtained with $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$. When including background, a decrease by 5% is observed for both confidence levels. This difference in CP-sensitivity is expected, considering that the $\mathcal{OO}$ contains the full information about the seven dimensional phase space whereas $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ only carries the (signed) angular information of the two outgoing tagging jets.

Chapter 6

Machine Learning CP-Sensitive Observables

CP-odd observables are crucial for probing the CP-properties of interactions. Exploring new approaches to construct such observables is an important step towards advancing future analyses in their search for new sources of CP-violation.

As recently demonstrated in [18] and [19], Machnine Learning (ML) methods are promising candidates for such new approaches. The Neural Networks (NNs) used in these ML methods are based on recognizing kinematic correlations between given input features, enabling a deeper understanding of the structure and properties of the new sources of CP-violation. Furthermore, machine learning based CP-sensitive observables provide an easier accessibility than established observables like the $\mathcal{OO}$, whose construction requires the calculation of matrix elements demanding access to special tools like HAWK.

In the first section of this chapter, selected concepts regarding the machine learning methods that are utilized in this thesis, are introduced. Section 6.2 presents all input features that are considered in this chapter. Next, in section 6.3, a CP-sensitive observable is constructed directly from the output of a classification Neural Network, following the approach in [18]. Section 6.4 employs a Regression Neural Network, to resemble the distribution of the Optimal Observable. Finally, section 6.5 aims to derive analytical expressions describing the Optimal Observable as predicted by the SM as well as for CP-violating predictions, using the method of symbolic regression, following the approach in [19].

6.1 Machine Learning Methods

This section aims to introduce a selection of fundamental principles regarding the different types of supervised machine learning (ML) methods that are employed in this thesis. First, in section 6.1.1, an overview about the working principles of fully connected feed-forward neural networks (NNs) is provided. In this context, a distinction is made between regression and classification neural networks. Next, section 6.1.2 discusses the optimization of the hyperparameters that define the classification and regression NNs constructed in this thesis. In section 6.1.3, the permutation method, an approach for evaluating the importance of different input features during the training of NNs, is introduced. Section 6.1.4 discusses the general concept behind predicting data through analytical expressions via genetic programming based symbolic regression (SR).

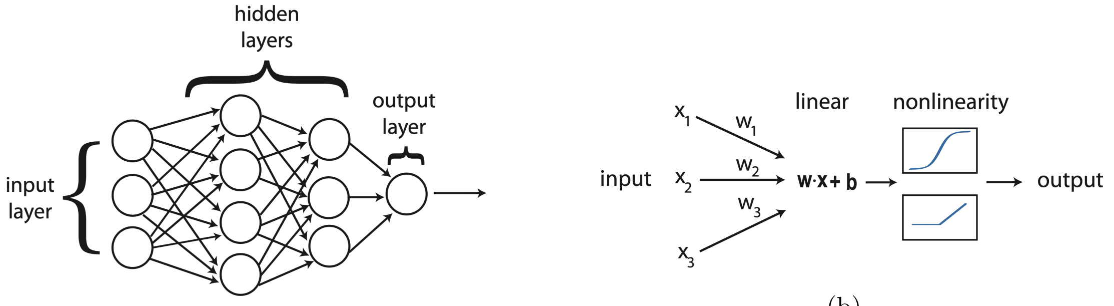
6.1.1 Fully Connected Neural Networks

The basic building blocks of neural networks are units called nodes. As illustrated in figure 6.1a, these nodes are arranged into layers, where, in the case of fully connected NNs, as considered in this section, each node in one layer is connected to every node in the subsequent layer. The first layer of a NN, where each node processes one entry of the initial input vector, is called input layer. The last layer, which returns the prediction of the NN, is called output layer. All intermediate layers are referred to as hidden layers.

Each node i in a given layer can take a vector $\vec{x} = (x_1, ...x_n)$, consisting of n input features, which it processes into one single, scalar output $o_i(\vec{x})$, that is passed as an input to nodes of the subsequent layer. Figure 6.1b illustrates how a given input vector $\vec{x}$ is processed inside one single node i. Generally, this processing can be decomposed in two steps. First, the input is transformed linearly by calculating the dot product with a weight vector $\vec{w}_i$ and adding a bias b_i. Next, non-linearity is introduced by applying non-linear activation functions A, such that the total operation performed by one node to produce the output $o_i(\vec{x})$ can be summarized as

$$o_i(\vec{x}) = A(\underbrace{\vec{w}_i \cdot \vec{x} + b_i}_{=:z_i}), \tag{6.1}$$

where z_i are referred to as the activations of a given node i.



(a)

(b)

Figure 6.1: Illustration of (a) the architecture of a NN, where the nodes (represented as circles) are arranged into layers, that can be categorized in three types: the input layer, the hidden layers and the output layer and (b) the processing of an input-vector $\vec{x} = (x_1, x_2, x_3)$ into a scalar output by first performing a linear transformation using the weights $\vec{w} = (w_1, w_2, w_3)$ and adding the bias b and then applying a non-linear activation function [54].

In this thesis, the rectified linear unit (ReLU) is used as an activation function in the hidden layers. It is defined as follows

$$\mathrm{ReLU}(z) = \max(0, z). \tag{6.2}$$

The training of neural networks is guided by the objective of minimizing loss functions, which measure the discrepancy between the NN prediction and the target data. Throughout this chapter, different loss functions are considered. They are defined in the respective sections. The minimization of these loss functions is performed using the stochastic gradient descent based optimizer `ADAM` [55].

Classification Neural Networks Classification neural networks are employed for tasks that involve categorizing input data into pre-defined classes. In the general case of multiclass neural networks, which are used to categorize data in $n > 2$ classes, the number of output nodes corresponds to the number of considered classes, such that each output node j returns a probability P_j for an given input set to belong to the respective class. To ensure that all outputs conform to the characteristics of actual probabilities, i.e., are normalized such that

$$\sum_{j=0}^{n-1} P_j = 1, \tag{6.3}$$

the output layer of multiclass NNs is activated using the softmax function

$$\mathrm{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=0}^{n-1} e^{z_j}}, \tag{6.4}$$

where z_i are the activations of the considered output node and j runs over the number of all n output nodes. For binary neural networks, employed in the special case of $n = 2$ possible classes, only one output node is necessary. It is activated by the sigmoid function $\sigma(z)$

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \tag{6.5}$$

which again ensures that the output probability P_1 for an input set to belong to class 1, is normalized, such that the probability to belong to class 0 can be obtained via

$$P_0 = 1 - P_1. \tag{6.6}$$

Regression Neural Networks If the target data is given in terms of continuous variable distributions, regression neural networks are employed. Regression neural networks yield numerical outputs that aim to approximate the values of the target variables. The activation function of the single output node is the identity function.

6.1.2 Hyperparameter Optimization

The choice of hyperparameters and the architecture of the neural networks constructed in this chapter is based on the results of studies performed using `OPTUNA`, a Bayesian optimization algorithm, utilizing the Tree-Structured Parzen estimator (TPE) [56]. An `OPTUNA` study is performed over several trials, where in each trial a NN, defined by a set of hyperparameters that is sampled from a pre-defined search space, is trained. The performance of this neural network is evaluated based on a customizable figure of merit and the search for hyperparameters is guided by optimizing this target metric. In this chapter, the figure of merit that defines a neural networks' performance is the CP-sensitivity of the observable constructed from its output. Specifically, the objective of the conducted optimization studies is to minimize the length of the 1σ confidence intervals extracted from NLL scans on the observable distributions. Besides the number of hidden layers and the number of nodes in each individual layer, the following hyperparameters are considered in these optimization studies:

L2 regularization strength To prevent overfitting, the method of L2-regularization [54] is applied. Since models that mimic the training data too closely, tend to have large weights w_i, they can be penalized in training, by considering an additional term inside the loss function that is quadratic in w

$$\mathrm{loss}_{\mathrm{L2}} = \mathrm{loss}(w_1, .., w_n, b_1, ..., b_n) + s_{\mathrm{L2}} \cdot \sum_{i=1}^{n} w_i^2. \tag{6.7}$$

The hyperparameter $s_{\mathrm{L2}} < 1$, allows controlling the strength of this penalization.

Learning rate decay The learning rate α is a hyperparameter that enables to control the step size used to approach the minimum of the loss function during training. To avoid becoming trapped in local minima of the loss function early in the training process, while still enabling convergence to the global minimum, it is advantageous to decay the learning rate during training. This learning rate decay is realized as follows

$$\alpha(\mathrm{step}) = \alpha_{\mathrm{in}} \cdot \lambda^{\frac{\mathrm{step}}{\mathrm{decay\ steps}}}, \tag{6.8}$$

where α_{in} is the initial learning rate, *step* refers to the specific optimizer step that this exponential decay is applied to and *decay steps* and the decay parameter λ are hyperparameters that allow to control the strength of the decay.

Batch size The hyperparameter batch size defines the number of input samples that are processed by the NN before updating the model parameters.

6.1.3 Feature Importance: The Permutation Method

The permutation method is a way to estimate the impact of a given input feature on the performance of a neural network. First, the performance p of the NN trained with the initial set of input features is evaluated. Then the NN is trained again with the entries of one specific feature being randomly permuted. If the resulting NN performance p_{perm} is worse than the performance p obtained with the initial feature set, the permuted feature is classified to be beneficial for the training process. The choice of the figure of merit p that quantifies the NN performance is problem specific. Since in this chapter, the NN performance is quantified by the CP-sensitivity of the observable constructed from its output, p and p_{perm} are defined as follows

$p \equiv 1\sigma$ interval-length obtained with initial set of input features

$p_{\text{perm}} \equiv 1\sigma$ interval-length obtained with a specific input feature randomly permuted

A score that classifies the importance of a given feature can then be constructed as

$$\text{feature importance score} = 1 - \frac{p}{p_{\text{perm}}}. \tag{6.9}$$

6.1.4 Genetic Programming Based Symbolic Regression

Similar to usual regression methods, symbolic regression (SR) aims to approximate continuous distributions. However, in this case, the model is constructed in terms of mathematical operators and variables, which are combined by the algorithm such that the output is given in terms of analytical expressions, rather than numerical values. In this thesis, the framework `PySR` [57] is used for SR. The fitness of a constructed analytical expression for describing the target data is evaluated by a function (score) that considers deviations between the analytical approximation and the target data (often evaluated via the mean squared error) and simultaneously favors expressions with low complexity. The search for analytical expressions is guided by optimizing this score-function. For this, `PySR` employs an algorithm that is based on genetic programming [58]. This algorithm operates inspired by natural selection. As a starting point, a population of randomly generated mathematical expressions, represented as trees consisting of nodes that contain operators or variables, is selected from a pre-defined search space. The expressions that best fit the target data are selected to produce the next generation. In the process of producing a new generation, modifications are included in the form of mutations and crossovers. As illustrated in figure 6.2, mutations refer to the process of randomly replacing nodes in existing trees and in crossover new trees are created by combining random subtrees of existing trees [59].

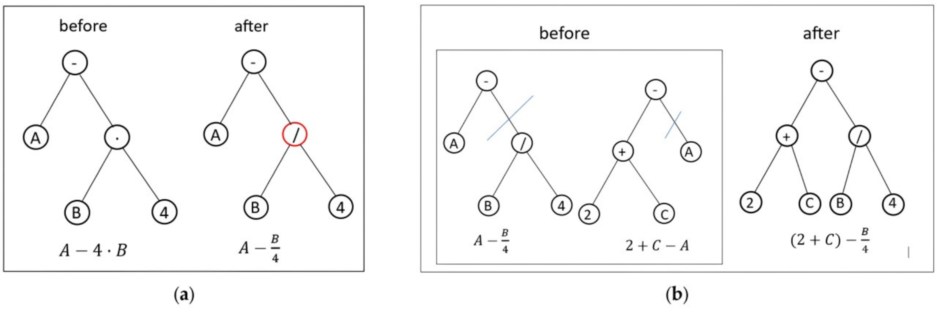


Figure 6.2: Illustration of modification of existing trees when breeding a new generation by (a) mutation, where nodes are randomly replaced and (b) crossover, where new trees are created by combining random subtrees of existing trees [59]

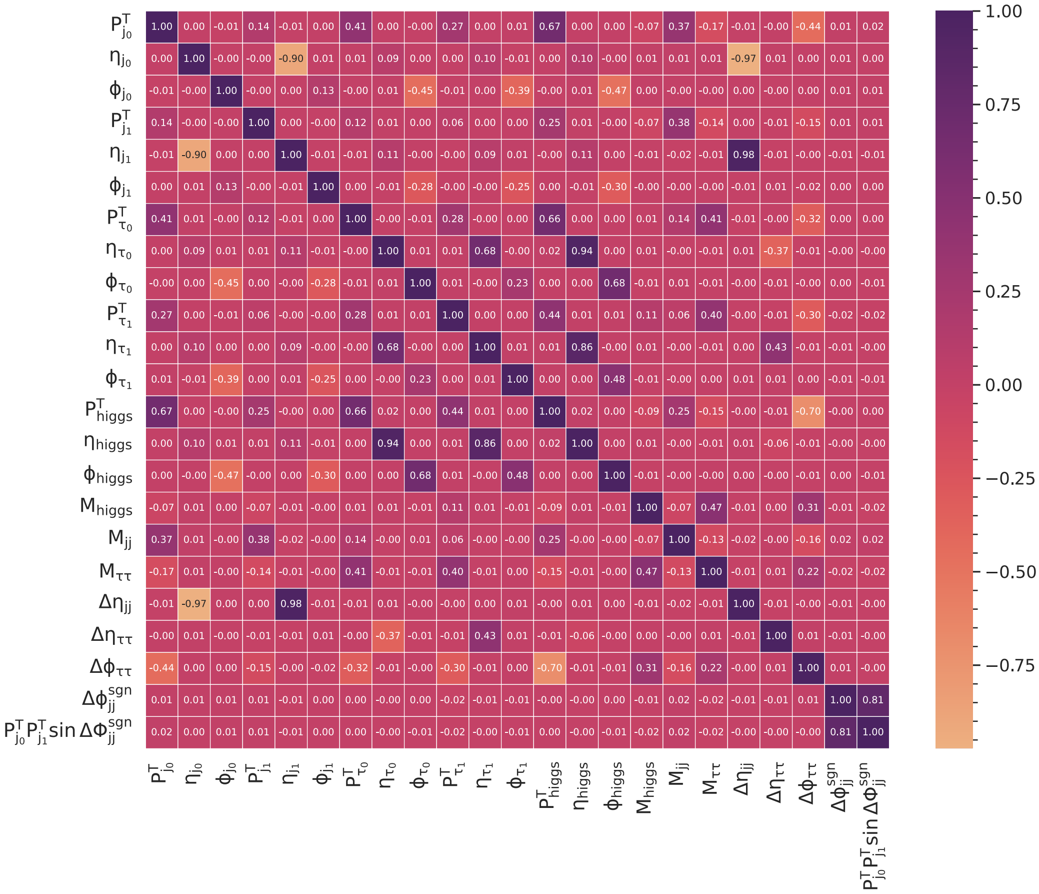
6.2 Input Features

In table 6.1, all input features that are considered in the training of the neural networks discussed in this chapter, are presented. These include both low and high-level features. The low-level features consist of simple kinematic information about all final state constituents in the analyzed VBF Higgs-boson production mode, namely the two tagging jets, the τ-leptons, whose kinematic information is reconstructed from only their visible decay products (see section 2.1 and section 3.1), and the Higgs-boson, reconstructed using the MMC. On the other hand, the high-level features are more complex quantities, build through combining different low level features. Examples for such high level features are the signed difference in azimuthal angles $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ introduced in section 5.2 and the variable $P_{j_0}^T P_{j_1}^T \sin\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ which is the result of a symbolic regression, found in [19] and verified in section 6.5.

Variable	Description
$P_{j_0}^T, \Phi_{j_0}, \eta_{j_0}$	Leading jet (j_0) three vector components
$P_{j_1}^T, \Phi_{j_1}, \eta_{j_1}$	Subleading jet (j_1) three vector components
$P_{\tau_0}^T, \Phi_{\tau_0}, \eta_{\tau_0}$	Leading tau-lepton (τ_0) three vector components
$P_{\tau_1}^T, \Phi_{\tau_1}, \eta_{\tau_1}$	Subleading tau-lepton (τ_1) three vector components
$P_{\mathrm{higgs}}^T, \Phi_{\mathrm{higgs}}, \eta_{\mathrm{higgs}}, M_{\mathrm{higgs}}$	Higgs-boson four vector components (obtained with the MMC)
M_{jj}	Invariant mass of the two tagging jets
$M_{\tau\tau}$	Invariant mass of the two tau-leptons
$\Delta\eta_{jj}$	Difference in pseudorapidity of the two tagging jets
$\Delta\eta_{\tau\tau}$	Difference in pseudorapidity of the two tau-leptons
$\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$	Signed difference in azimuthal angles of the two tagging jets [17]
$\Delta\Phi_{\tau\tau}$	Difference in azimuthal angles of the two tau-leptons
$P_{j_0}^T P_{j_1}^T \sin\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$	Result of symbolic regression [19]

Table 6.1: Collection of all considered input features.

For each pair of input features, the Pearson Correlation Coefficient (PCC) [46] is displayed in the correlation matrix shown in figure 6.3. As the PCC is a measure of the linear relationship between two variables, it is not able to register more complex feature dependencies. This explains the low Pearson correlation between the high level input features $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$, $P_{j0}^T P_{j1}^T \sin\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ and all low level input features. Figure 6.4 illustrates the correlation between these two high level features and selected low level features. The separated structure observed in the correlation plots that include $\eta_{j_{0,1}}$ reflects the discrimination between positive and negative detector hemispheres, included in $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ by considering the sign of $\eta_{j_{0,1}}$ when determining the angle difference of the two tagging jets (see section 5.2). On the other hand, this requirement removes the correlation between $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ and $\Phi_{j_{0,1}}$. As expected, a strong correlation between $P_{j0}^T P_{j1}^T \sin\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ and $P_{j_{0,1}}^T$ is observed.



	$P^T_{j_0}$	η_{j_0}	ϕ_{j_0}	$P^T_{j_1}$	η_{j_1}	ϕ_{j_1}	$P^T_{\tau_0}$	η_{τ_0}	ϕ_{τ_0}	$P^T_{\tau_1}$	η_{τ_1}	ϕ_{τ_1}	P^T_{higgs}	η_{higgs}	ϕ_{higgs}	M_{higgs}	M_{jj}	$M_{\tau\tau}$	$\Delta\eta_{jj}$	$\Delta\eta_{\tau\tau}$	$\Delta\phi_{\tau\tau}$	$\Delta\phi^{sgn}_{jj}$	$P^T_{j_0}P^T_{j_1}\sin\Delta\Phi^{sgn}_{jj}$
$P^T_{j_0}$	1.00	0.00	-0.01	0.14	-0.01	0.00	0.41	0.00	-0.00	0.27	0.00	0.01	0.67	0.00	0.00	-0.07	0.37	-0.17	-0.01	-0.00	-0.44	0.01	0.02
η_{j_0}	0.00	1.00	-0.00	-0.00	-0.90	0.01	0.01	0.09	0.00	0.00	0.10	-0.01	0.00	0.10	-0.00	0.01	0.01	0.01	-0.97	0.01	0.00	0.01	0.00
ϕ_{j_0}	-0.01	-0.00	1.00	-0.00	0.00	0.13	-0.00	0.01	-0.45	-0.01	0.00	-0.39	-0.00	0.01	-0.47	0.00	-0.00	-0.00	0.00	-0.00	0.00	0.01	0.01
$P^T_{j_1}$	0.14	-0.00	-0.00	1.00	0.00	-0.00	0.12	0.01	0.00	0.06	0.00	0.00	0.25	0.01	-0.00	-0.07	0.38	-0.14	0.00	-0.01	-0.15	0.01	0.01
η_{j_1}	-0.01	-0.90	0.00	0.00	1.00	-0.01	-0.01	0.11	-0.00	-0.00	0.09	0.01	-0.00	0.11	0.00	-0.01	-0.02	-0.01	0.98	-0.01	-0.00	-0.01	-0.01
ϕ_{j_1}	0.00	0.01	0.13	-0.00	-0.01	1.00	0.00	-0.01	-0.28	-0.00	-0.00	-0.25	0.00	-0.01	-0.30	-0.00	-0.00	-0.00	-0.01	0.01	-0.02	0.00	0.00
$P^T_{\tau_0}$	0.41	0.01	-0.00	0.12	-0.01	0.00	1.00	-0.00	-0.01	0.28	-0.00	0.00	0.66	0.00	0.00	0.01	0.14	0.41	-0.01	-0.00	-0.32	0.00	0.00
η_{τ_0}	0.00	0.09	0.01	0.01	0.11	-0.01	-0.00	1.00	0.01	0.01	0.68	-0.00	0.02	0.94	0.00	0.01	-0.00	-0.01	0.01	-0.37	-0.01	-0.00	-0.00
ϕ_{τ_0}	-0.00	0.00	-0.45	0.00	-0.00	-0.28	-0.01	0.01	1.00	0.01	-0.00	0.23	0.00	0.00	0.68	-0.01	0.01	-0.01	-0.00	-0.01	-0.00	0.00	-0.01
$P^T_{\tau_1}$	0.27	0.00	-0.01	0.06	-0.00	-0.00	0.28	0.01	0.01	1.00	0.00	0.00	0.44	0.01	0.01	0.11	0.06	0.40	-0.00	-0.01	-0.30	-0.02	-0.02
η_{τ_1}	0.00	0.10	0.00	0.00	0.09	-0.00	-0.00	0.68	-0.00	0.00	1.00	0.01	0.01	0.86	-0.00	0.01	-0.00	-0.01	0.00	0.43	-0.01	-0.01	-0.00
ϕ_{τ_1}	0.01	-0.01	-0.39	0.00	0.01	-0.25	0.00	-0.00	0.23	0.00	0.01	1.00	0.00	0.00	0.48	-0.01	-0.00	0.00	0.01	0.01	0.00	-0.01	-0.01
P^T_{higgs}	0.67	0.00	-0.00	0.25	-0.00	0.00	0.66	0.02	0.00	0.44	0.01	0.00	1.00	0.02	0.00	-0.09	0.25	-0.15	-0.00	-0.01	-0.70	-0.00	0.00
η_{higgs}	0.00	0.10	0.01	0.01	0.11	-0.01	0.00	0.94	0.00	0.01	0.86	0.00	0.02	1.00	0.00	0.01	-0.00	-0.01	0.01	-0.06	-0.01	-0.00	-0.00
ϕ_{higgs}	0.00	-0.00	-0.47	-0.00	0.00	-0.30	0.00	0.00	0.68	0.01	-0.00	0.48	0.00	0.00	1.00	-0.01	-0.00	-0.00	0.00	-0.00	-0.01	0.00	-0.01
M_{higgs}	-0.07	0.01	0.00	-0.07	-0.01	-0.00	0.01	0.01	-0.01	0.11	0.01	-0.01	-0.09	0.01	-0.01	1.00	-0.07	0.47	-0.01	0.00	0.31	-0.01	-0.02
M_{jj}	0.37	0.01	-0.00	0.38	-0.02	-0.00	0.14	-0.00	0.01	0.06	-0.00	-0.00	0.25	-0.00	-0.00	-0.07	1.00	-0.13	-0.02	-0.00	-0.16	0.02	0.02
$M_{\tau\tau}$	-0.17	0.01	-0.00	-0.14	-0.01	-0.00	0.41	-0.01	-0.01	0.40	-0.01	0.00	-0.15	-0.01	-0.00	0.47	-0.13	1.00	-0.01	0.01	0.22	-0.02	-0.02
$\Delta\eta_{jj}$	-0.01	-0.97	0.00	0.00	0.98	-0.01	-0.01	0.01	-0.00	-0.00	0.00	0.01	-0.00	0.01	0.00	-0.01	-0.02	-0.01	1.00	-0.01	-0.00	-0.01	-0.00
$\Delta\eta_{\tau\tau}$	-0.00	0.01	-0.00	-0.01	-0.01	0.01	-0.00	-0.37	-0.01	-0.01	0.43	0.01	-0.01	-0.06	-0.00	0.00	-0.00	0.01	-0.01	1.00	0.01	-0.01	-0.00
$\Delta\phi_{\tau\tau}$	-0.44	0.00	0.00	-0.15	-0.00	-0.02	-0.32	-0.01	-0.00	-0.30	-0.01	0.00	-0.70	-0.01	-0.01	0.31	-0.16	0.22	-0.00	0.01	1.00	0.01	-0.00
$\Delta\phi^{sgn}_{jj}$	0.01	0.01	0.01	0.01	-0.01	0.00	0.00	-0.00	0.00	-0.02	-0.01	-0.01	-0.00	-0.00	0.00	-0.01	0.02	-0.02	-0.01	-0.01	0.01	1.00	0.81
$P^T_{j_0}P^T_{j_1}\sin\Delta\Phi^{sgn}_{jj}$	0.02	0.00	0.01	0.01	-0.01	0.00	0.00	-0.00	-0.01	-0.02	-0.00	-0.01	0.00	-0.00	-0.01	-0.02	0.02	-0.02	-0.00	-0.00	-0.00	0.81	1.00

Figure 6.3: Correlation matrix, including the Pearson coefficients for each input feature pair.

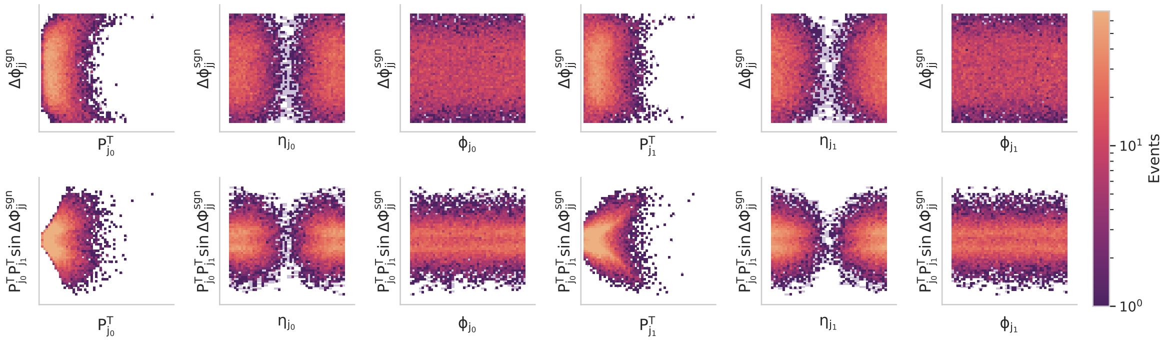


Figure 6.4: Correlation between the high level input features $\Delta\Phi_{\text{jj}}^{\text{sgn}}$, $P_{j_0}^T P_{j_1}^T \sin\Delta\Phi_{\text{jj}}^{\text{sgn}}$ and selected low level input features.

6.3 Classification Neural Network Observable $\mathcal{O}_{\text{NN}}^{\text{Class}}$

This section investigates the construction of CP-sensitive observables, directly from the output of classification neural networks, as it was first proposed in [18].

As described in chapter 4, the interference term in eq. (4.8) is responsible for introducing new sources of CP-violation. For a CP-even observable, this interference leaves no measurable traces in its distribution. However, in the distributions of CP-odd observables, asymmetries are induced by interference effects.

This can be exploited by training a neural network that learns to distinguish between events in which the interference term $\Re\{\mathcal{M}_{\text{SM}}^*\mathcal{M}_{\text{BSM}}\}$ is positive and negative, respectively. A simple CP-odd observable can then be build from the probabilities P_+, P_- that the NN assigns to a given event, for having a positive/ negative interference term

$$\mathcal{O}_{\text{NN}}^{\text{Class}} = P_+ - P_-. \tag{6.10}$$

Positive $\tilde{d}$-values realized in nature will create an asymmetry towards positive $\mathcal{O}_{\text{NN}}^{\text{Class}}$ values and vice versa. Thus, this observable qualifies to be used in CP-tests as they are described in chapter 4.

6.3.1 Binary Neural Network

First, a binary neural network is trained, aiming to discriminate only between events for which the interference term is positive and events for which the interference term is negative. Specifically, the two classes are assigned and targeted as folows:

$$\begin{array}{lll} \text{Class 0:} & \Re\{\mathcal{M}_{\text{SM}}^*\mathcal{M}_{\text{BSM}}\} > 0, & \text{Target-value: } 0 \\ \text{Class 1:} & \Re\{\mathcal{M}_{\text{SM}}^*\mathcal{M}_{\text{BSM}}\} < 0, & \text{Target-value: } 1 \end{array}$$

The NN is constructed such that it has only one output node which is activated by the sigmoid function eq. (6.5), such that it returns the probability P_- for an event to have a negative interference term (class 1). Respectively, the probability P_+ to have a positive interference term (class 0) can be obtained via

$$P_+ = 1 - P_-. \tag{6.11}$$

In contrast to the methodology presented in [18], where positive and negative interference samples are generated through MC simulations, this thesis follows a different approach. As explained in section 4.5, here, the information about interference effects that arise in a specific $\tilde{d} \neq 0$ scenario is incorporated by reweighting the signal events predicted by the SM using the event weights $w_{\tilde{d}}$ that are calculated via eq. (4.19) while using the respective $\tilde{d}$-value. These weights are included in the training of the NN by considering them when calculating the loss, which in this case is given by the binary cross entropy

$$\text{loss} = -\frac{\sum_i \hat{w}^i_{\tilde{d}} \cdot \left[y^i_{\text{true}} \cdot \log(P^i_-) + (1 - y^i_{\text{true}}) \cdot \log(1 - P^i_-) \right]}{\sum_i \hat{w}^i_{\tilde{d}}}. \tag{6.12}$$

Here, $\hat{w}^i_{\tilde{d}}$ are the event weights $w^i_{\tilde{d}}$ given in eq. (4.19) normalized by the factor

$$\text{NF}_\pm = \frac{N_{\text{SM}}}{\sum_i w^i_{\tilde{d}}}, \tag{6.13}$$

such that the total number of considered events adds up to the total number of signal events predicted by the SM, $N_{\text{SM}} = 40$. P^i_- denotes the probability for a given event to belong to class 1 (negative interference) and y^i_{true} denotes the target value for this event.

The architecture and hyperparameters of the NN are chosen based on the outcome of an optimization study that is conducted within the framework `OPTUNA`, over 100 trials. For this, the search space in the hyperparameters is defined as follows:

- Number of hidden layers $\in \{1, 2, 3, 4, 5\}$
- Number of nodes (individually for each layer) $\in \{5, 25, 45, 90, 100, 150, 200, 250, 300, 350, 400\}$
- Strength of L2 regularization $\in [10^{-5}, 10^{-1}]$, scanned logarithmically
- Initial learning rate $\in [10^{-5}, 10^{-1}]$, scanned logarithmically
- Learning rate decay steps $\in [1000, 10000]$, scanned linearly
- Learning rate decay parameter $\in [0, 0.7]$, scanend linearly
- Batch size $\in \{32, 64, 96, 128\}$

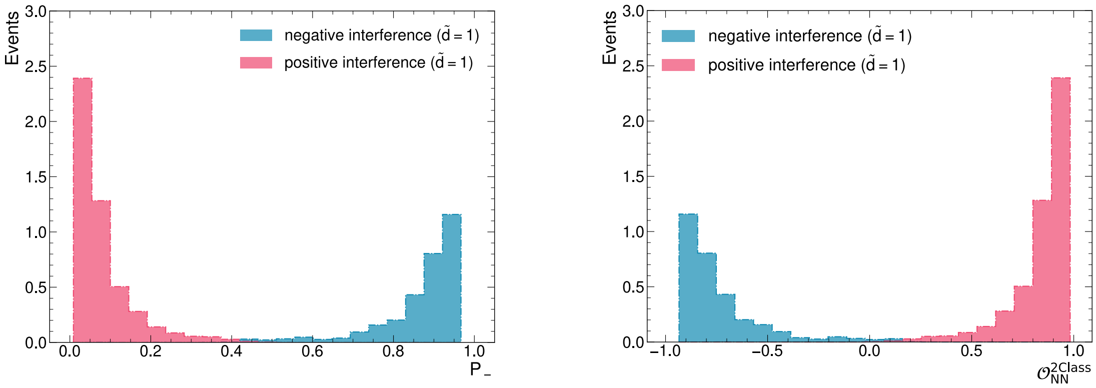
The outcome of this optimization study can be found in table 6.2.

Optimized hyperparameter	Study outcome
Number of hidden layers	4
Number of nodes in hidden layer 1	350
Number of nodes in hidden layer 2	400
Number of nodes in hidden layer 3	5
Number of nodes in hidden layer 4	150
L2 regularization strength	$2.9{\cdot}10^{-3}$
Initial learning rate	$1.8{\cdot}10^{-3}$
Learning rate decay steps	634.3
Learning rate decay rate	0.2
Batch size	128

Table 6.2: Architecture and hyperparameters of the binary neural network as found in a hyperparameter optimization study using `OPTUNA`.

This binary neural network is trained using the complete set of input features presented in section 6.2, table 6.1, taking only signal events into account. During training, 80% of the entire dataset are considered, while the remaining data is reserved for validation. The training is conducted assuming the CP-violating scenario corresponding to $\tilde{d} = 1$, i.e. using the weights $w_{\tilde{d}}$ from eq. (4.19) that correspond to this $\tilde{d}$-value, when calculating the loss eq. (6.13). Section 6.3.4 presents a study on the impacts of the specific $\tilde{d}$ scenario that is assumed during training.

Figure 6.5a shows the distribution of the predictions P_- made by the binary NN on the validation dataset in the $\tilde{d} = 1$-scenario. It demonstrates that the NN is able to effectively separate between positive and negative interference events, allowing the construction of a CP-sensitive observable $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$ according to eq. (6.10). The distribution of this observable, calculated using the predictions on the validation dataset, is shown in figure 6.5b for the $\tilde{d} = 1$-scenario. As expected for a CP-sensitive observable, the assumption of a non-zero $\tilde{d}$-value yields an asymmetric distribution. Specifically, since in this case the distribution is shown in the $\tilde{d} = 1$-scenario, an asymmetry towards positive values of $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$ can be observed.



(a) NN output P_{-} (b) CP-Sensitive observable $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$

Figure 6.5: Output P_{-} of the binary NN predicted for the validation dataset (a) and CP-sensitive Observable $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$ constructed directly from this output (b). Both distributions are shown as predicted for the CP-violating scenario, corresponding to $\tilde{d} = 1$. This validation dataset consists of 20% of the events in the full signal sample. This corresponds to an expected number of events $N_{\mathrm{val}} = 8$.

6.3.2 Multiclass Neural Network

The binary neural network described in the previous section is now expanded by an additional class, such that besides predicting the probabilities for positive and negative interference terms, it is also able to predict the probability P_{SM} that a given event is a SM event. This new multiclass neural network now has three output nodes that predict the probabilities P_{-}, P_{+} and P_{SM} and are activated by the softmax function eq. (6.4) such that

$$P_{+} + P_{-} + P_{\mathrm{SM}} = 1. \tag{6.14}$$

Specifically, the three classes are assigned and targeted as follows

$$\begin{array}{lll} \text{Class 0:} & \text{SM event} & \text{Target: } (1,\,0,\,0) \\ \text{Class 1:} & \Re\{\mathcal{M}_{\mathrm{SM}}^{*}\mathcal{M}_{\mathrm{BSM}}\} > 0 & \text{Target: } (0,\,1,\,0) \\ \text{Class 2:} & \Re\{\mathcal{M}_{\mathrm{SM}}^{*}\mathcal{M}_{\mathrm{BSM}}\} < 0 & \text{Target: } (0,\,0,\,1). \end{array}$$

In this case, the training set is created by first splitting the original data set, consisting of only SM signal events, in half, using the odd-even-splitting method, based on the event number assigned in the MC simulation. One half of the data set is now kept as SM sample. The remaining events are split in positive/ negative interference samples, based on the sign of their interference term. These interference samples are then normalized to the expected number of events $N_{\mathrm{SM}} = 20$ in the SM sample, using the normalization factors

$$\mathrm{NF}_{\pm} = \frac{N_{\mathrm{SM}}}{\sum_i w_{\tilde{d},i}^{\pm}}, \tag{6.15}$$

where $w_{\tilde{d}}^{\pm}$, denote the weights of positive (+) or negative (-) interference events, calculated according to eq. (4.19) assuming a value $\tilde{d} \neq 0$.

For multiclass NNs, the loss function is given by the categorical cross entropy

$$\text{loss} = -\frac{\sum_i^n \sum_{j=1}^3 w_{\tilde{d}}^{ij} \left[y_{\text{true}}^{ij} \cdot \log(P^{ij}) \right]}{\sum_i w_{\tilde{d}}^{ij}}. \tag{6.16}$$

Here, the first sum runs over all n data points and the second sum runs over the number of classes, such that y_{true}^{ij} denotes the j-th entry of the target vector for a given event i and P^{ij} denotes the according probability P_{SM}, P_+ or P_- for a given event i to belong to the respective class. The weight of an event i that belongs to class j is denoted by $w_{\tilde{d}}^{i,j}$.

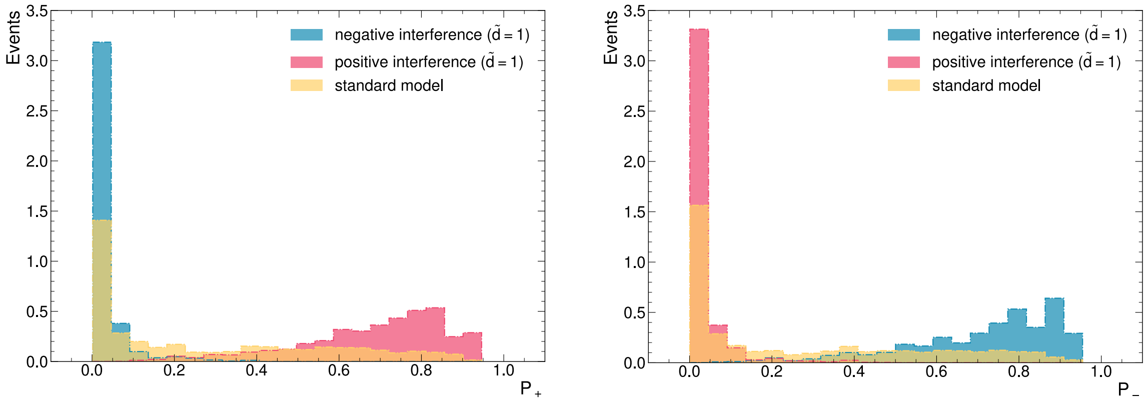
Again, the architecture and hyperparameters of the model are determined by the outcome of an optimization study conducted within the framework `OPTUNA` over 100 trials, using the search space defined in section 6.3.1. The outcome of this optimization study can be found in table 6.3.

Optimized hyperparameter	Study outcome
Number of hidden layers	2
Number of nodes in hidden layer 1	350
Number of nodes in hidden layer 2	400
L2 regularization strength	$1.7 \cdot 10^{-5}$
Initial learning rate	$8.2 \cdot 10^{-4}$
Learning rate decay steps	7849.9
Learning rate decay rate	0.3
Batch size	128

Table 6.3: Architecture and hyperparameters of the multiclass neural network as found in a hyperparameter optimization study using `OPTUNA`.

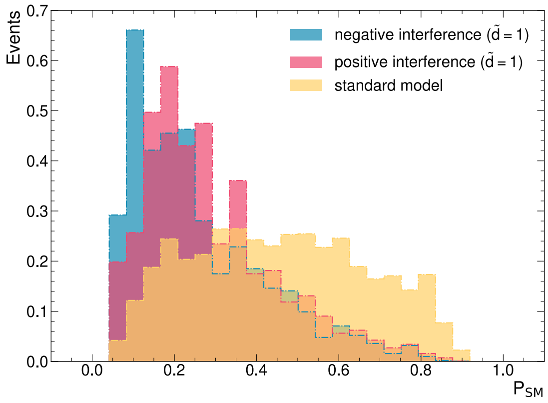
As in the previous section, the network is trained with the full input feature set presented in section 6.2, table 6.1. During training, 80% (corresponding to an expected number of events $N_{\text{train}} = 16$ per sample) of the dataset are considered, while the remaining 20% of the data (corresponding to a weighted number of events $N_{\text{val}} = 4$ per sample) is reserved for validation. The training is conducted assuming the CP-violating scenario corresponding to $\tilde{d} = 1$, i.e. using the weights $w_{\tilde{d}}$ according to eq. (4.19) that correspond to this $\tilde{d}$-value when calculating the loss in eq. (6.16).

Figure 6.6 shows the distribution of the probabilities predicted by the output nodes of the multiclass NN for the validation dataset. For positive and negative interference events, high probabilities P_+, P_- of belonging to their true classes are predicted. However, for SM events, the probabilities P_{SM} are more broadly distributed, indicating that the multiclass NN has difficulties recognizing SM specific characteristics.



(a) Positive interference class

(b) Negative interference class



(c) SM class

Figure 6.6: Outputs of the multiclass NN predicted for the validation dataset: Probability P_+ for positive interference (a), probability P_- for negative interference (b) and probability P_{SM} for SM prediction (c). Each sample (SM, positive interference, negative interference) contains an expected number of events $N_{\text{val}} = 4$.

Figure 6.7 shows the distribution of the CP-sensitive observable $\mathcal{O}_{\text{NN}}^{\text{3Class}}$ calculated from the predictions P_+, P_- made by the NN on the validation dataset. Standard model events are distributed symmetrically, whereas for $\tilde{d} = 1$ a distinct symmetry is observed.

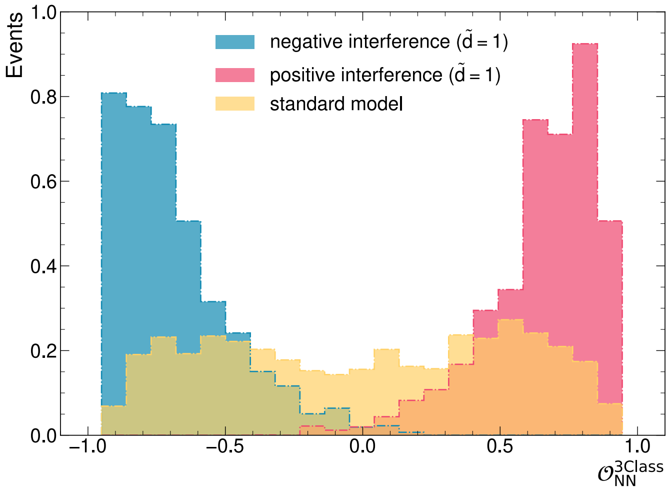
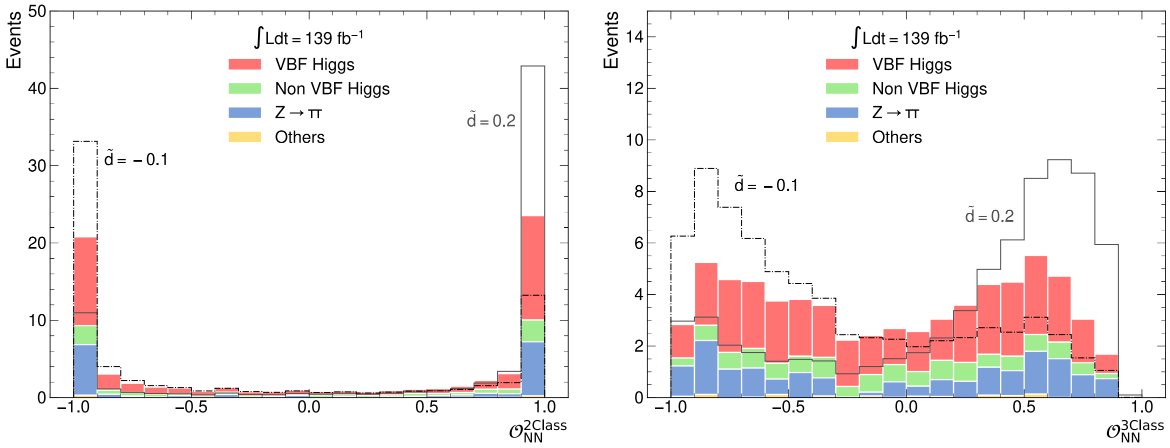


Figure 6.7: CP-sensitive Observable $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ constructed directly from the multiclass NN predictions P_+, P_- for the validation dataset. Each of the three histograms contains an expected number of events $N_{\mathrm{val}} = 4$.

6.3.3 Comparison of CP-sensitive observables constructed with Binary and Multiclass Neural Network

The binary neural network and the multiclass neural network constructed in the previous sections are now applied to the full dataset defined in section 3.2. Figure 6.8 shows the distributions of the CP-odd observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$, $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ that are constructed from the resulting NN outputs. The distributions are shown according to the SM prediction and additionally for two CP-vioalting cases with $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$. In both $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$ and $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$, CP-violating effects induce a shift of the distribution to the edge regions. For the binary NN observable distribution, this behavior can already be observed in the pure SM prediction, whereas the multiclass NN observable is distributed more broadly for the SM prediction and only gets shifted to the edge regions when CP-violating scenarios are considered.

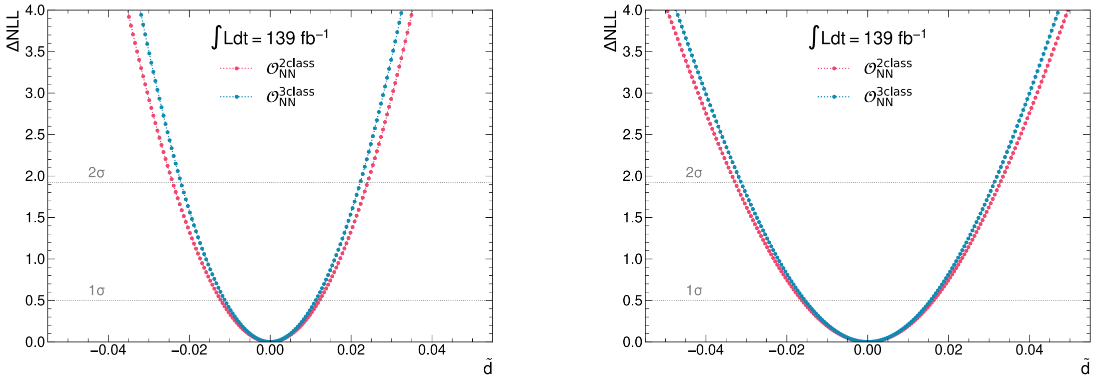


(a) Binary NN observable $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$

(b) Multiclass NN observable $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$

Figure 6.8: Distributions of the CP-sensitive observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$, $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$, constructed from the predictions of the binary/ multiclass NN on the full dataset, shown as as predicted by the SM for an integrated luminosity of $\int \mathrm{Ldt} = 139\ \mathrm{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. Additionally, distributions predicted for two CP-violating scenarios with the strengths $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$ are presented. These distributions are normalized to the total number $N_{\mathrm{SM}} = 69$ of expected SM events.

To estimate the CP-sensitivity of $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$ and $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$, NLL-scans according to section 4.4 are employed under the same conditions as for the Optimal Observable in section 5.1. Figure 6.9 shows the resulting ΔNLL-curves for the case where only signal events are considered and for the case where also background events are included. Table 6.4 and table 6.5 contain the corresponding confidence intervals, as well as their lengths. At both confidence levels, $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ is observed to have a higher CP-sensitivity than $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$. Specifically, $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ reduces the lengths of the 1σ and 2σ confidence intervals by 8% in the signal only case and by 5% when background is included.



(a) Signal only

(b) Including background

Figure 6.9: Comparison of the ΔNLL-curves obtained with the distributions of the observable $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}, \mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ constructed from the binary NN output and multiclass NN output when (a) considering only signal events and (b) also including background events.

	1σ CI	1σ CI length	2σ CI	2σ CI length
$\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$	[-0.0122, 0.0122]	0.0244	[-0.0241, 0.0241]	0.0482
$\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$	[-0.0112, 0.0112]	0.0224	[-0.0221, 0.0222]	0.0443

Table 6.4: **Signal only:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}, \mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ that are constructed from the binary and the multiclass neural network. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta \ell = 0.0007$.

	1σ CI	1σ CI length	2σ CI	2σ CI length
$\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$	[-0.0163, 0.0163]	0.0326	[-0.0328, 0.0328]	0.0656
$\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$	[-0.0155, 0.0155]	0.0310	[-0.0313, 0.0313]	0.0626

Table 6.5: **Including background:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}, \mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ that are constructed from the binary and the multiclass neural network. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta \ell = 0.0007$.

6.3.4 Impact of the Choice of $\tilde{d}$-scenario in Training on the Neural Network Performance

The neural networks introduced in the previous sections were trained considering interference events in the arbitrarily selected $\tilde{d} = 1$-scenario. This section investigates how the choice of the specific CP-violating scenario that is assumed in the calculation of the loss function during training, impacts the performance of both the binary and the multiclass neural network. For this, six different CP-violating scenarios corresponding to the following $\tilde{d}$-values are considered:

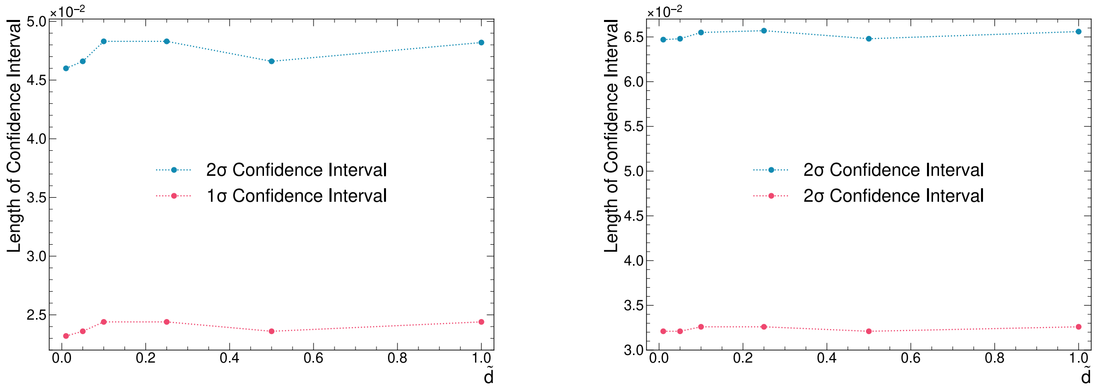
$$\tilde{d} \in \{0.01, 0.05, 0.1, 0.25, 0.5, 1\}.$$

For each of these $\tilde{d}$-values, the corresponding weights of the interference events are calculated according to eq. (4.19) and a new NN is trained and optimized while considering these specific weights when calculating the losses eq. (6.13) and eq. (6.16), respectively.

Each optimization study is executed in `OPTUNA` over 100 trials using the search space defined in section 6.3.1. The results of these optimization studies for all six $\tilde{d}$ scenarios can be found in table B.1 for the binary case and in table B.2 for the multiclass case.

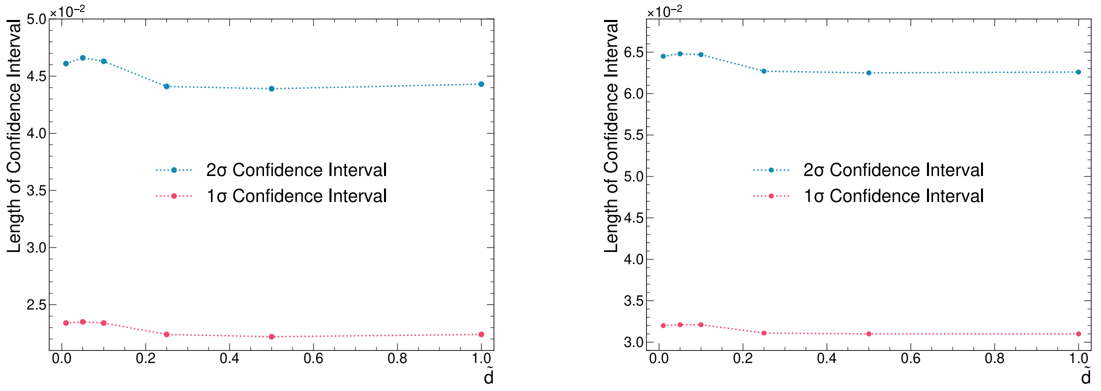
The predictions of each of these NNs on the full data set are then used for the construction of CP-sensitive observables $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{Class}})_{\tilde{d}}$ according to eq. (6.10). Figure A.1 shows the distributions of $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}})_{\tilde{d}}$ obtained from the binary NNs and figure A.2 shows the distributions of $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\tilde{d}}$ resulting from the multiclass NNs.

For each $\tilde{d}$-assumption, the 1σ and 2σ confidence intervals are extracted from the NLL curves obtained using the distributions of the corresponding observable $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{Class}})_{\tilde{d}}$ (see figure A.3). Figure 6.10 shows the dependency between the lengths of these confidence intervals and the respective choice of $\tilde{d}$ in training. The results are shown for the signal only cases and for the cases where background events are included. Table 6.6 and table 6.7 contain the corresponding numerical values obtained with the binary NN. Table 6.8 and table 6.9 contain the results obtained with the multiclass NN. In both cases, no clear dependency on $\tilde{d}$ can be observed. For small $\tilde{d}$-values, the confidence intervals lengths obtained from the binary and the multiclass neural networks are compatible within their uncertainty $\Delta\ell = 0.0007$ (see section 5.1, eq. (5.6)). However, in the multiclass case, choosing a $\tilde{d} > 0.1$ yields a significant decrease in confidence interval lengths, whereas in the binary case no such trend is observed. Further increasing $\tilde{d}$ does not yield notable changes in the multiclass results. These observations are true at both confidence levels for the signal only case and for the case where background events are included. Comparing the shortest interval lengths obtained using the multiclass NN with the shortest interval lengths obtained using the binary NN shows that with the multiclass NN an improvement by 4% at 1σ confidence level and by 5% at 2σ confidence level can be achieved in the signal only case. When including background, the interval lengths improve by 3% at both confidence levels. Based on these results, for further studies in this section, multiclass NNs will be utilized and the choice of considering the $\tilde{d} = 1$-scenario when training these NNs is retained.



(a) Binary NNs, signal only

(b) Binary NNs, including background



(c) Multiclass NNs, signal only

(d) Multiclass NNs, including background

Figure 6.10: Dependency between the lenghts of the 1σ and 2σ confidence intervals and the $\tilde{d}$-scenario that is assumed when training and optimizing the NNs. The first row shows the results obtained with the binary NNs when (a) only signal events are considered and (b) background events are included. The second row shows the results obtained with the multiclass NNs, again for the case when (c) only signal events are considered (d) when background events are included.

	1σ CI	1σ CI length	2σ CI	2σ CI length
$\tilde{d} = 0.01$	[-0.0116, 0.0116]	0.0232	[-0.0230, 0.0230]	0.0460
$\tilde{d} = 0.05$	[-0.0118, 0.0118]	0.0236	[-0.0233, 0.0233]	0.0466
$\tilde{d} = 0.1$	[-0.0122, 0.0122]	0.0244	[-0.0242, 0.0241]	0.0483
$\tilde{d} = 0.25$	[-0.0122, 0.0122]	0.0244	[-0.0242, 0.0241]	0.0483
$\tilde{d} = 0.5$	[-0.0118, 0.0118]	0.0236	[-0.0234, 0.0232]	0.0466
$\tilde{d} = 1$	[-0.0122, 0.0122]	0.0244	[-0.0241, 0.0241]	0.0482

Table 6.6: **Signal only:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$ constructed from the binary NN that was trained and optimized assuming different $\tilde{d}$-scenarios. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta \ell = 0.0007$.

	1σ CI	1σ CI length	2σ CI	2σ CI length
$\tilde{d} = 0.01$	[-0.0161, 0.0160]	0.0321	[-0.0324, 0.0323]	0.0647
$\tilde{d} = 0.05$	[-0.0161, 0.0160]	0.0321	[-0.0326, 0.0322]	0.0648
$\tilde{d} = 0.1$	[-0.0163, 0.0163]	0.0326	[-0.0328, 0.0327]	0.0655
$\tilde{d} = 0.25$	[-0.0163, 0.0163]	0.0326	[-0.0328, 0.0329]	0.0657
$\tilde{d} = 0.5$	[-0.0160, 0.0161]	0.0321	[-0.0323, 0.0325]	0.0648
$\tilde{d} = 1$	[-0.0163, 0.0163]	0.0326	[-0.0328, 0.0328]	0.0656

Table 6.7: **Including background:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the observables $\mathcal{O}_{\rm NN}^{\rm 2Class}$ constructed from the binary NN that was trained and optimized assuming different $\tilde{d}$-scenarios. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta \ell = 0.0007$.

	1σ CI	1σ CI length	2σ CI	2σ CI length
$\tilde{d} = 0.01$	[-0.0117, 0.0117]	0.0234	[-0.0230, 0.0231]	0.0461
$\tilde{d} = 0.05$	[-0.0117, 0.0118]	0.0235	[-0.0232, 0.0234]	0.0466
$\tilde{d} = 0.1$	[-0.0117, 0.0117]	0.0234	[-0.0231, 0.0232]	0.0463
$\tilde{d} = 0.25$	[-0.0112, 0.0112]	0.0224	[-0.0220, 0.0221]	0.0441
$\tilde{d} = 0.5$	[-0.0111, 0.0111]	0.0222	[-0.0219, 0.0220]	0.0439
$\tilde{d} = 1$	[-0.0112, 0.0112]	0.0224	[-0.0221, 0.0222]	0.0443

Table 6.8: **Signal only:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the observables $\mathcal{O}_{\rm NN}^{\rm 3Class}$ constructed from the multiclass NN that was trained and optimized assuming different $\tilde{d}$-scenarios. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta \ell = 0.0007$.

	1σ CI	1σ CI length	2σ CI	2σ CI length
$\tilde{d} = 0.01$	[-0.0160, 0.0160]	0.0320	[-0.0323, 0.0322]	0.0645
$\tilde{d} = 0.05$	[-0.0161, 0.0160]	0.0321	[-0.0326, 0.0322]	0.0648
$\tilde{d} = 0.1$	[-0.0161, 0.0160]	0.0321	[-0.0326, 0.0321]	0.0647
$\tilde{d} = 0.25$	[-0.0156, 0.0155]	0.0311	[-0.0314, 0.0313]	0.0627
$\tilde{d} = 0.5$	[-0.0155, 0.0155]	0.0310	[-0.0313, 0.0312]	0.0625
$\tilde{d} = 1$	[-0.0155, 0.0155]	0.0310	[-0.0313, 0.0313]	0.0626

Table 6.9: **Including background:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the observables $\mathcal{O}_{\rm NN}^{\rm 3Class}$ constructed from the multiclass NN that was trained and optimized assuming different $\tilde{d}$-scenarios. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta \ell = 0.0007$.

6.3.5 Optimization of Multiclass Neural Networks with different Sets of Input Variables

In this section, neural networks are trained and optimized using different subsets of the initial input feature set defined in section 6.2, table 6.1. Specifically, these subsets only include low level features. The effects of reducing the quantity and complexity of the input-information on the performance of the resulting NNs are then evaluated in terms of the CP-sensitivities reached with the observables constructed from their outputs.

Based on the results of section 6.3.3 and section 6.3.4 this study uses a multiclass neural network that is trained and optimized assuming the $\tilde{d} = 1$ scenario. The following input feature subsets, are considered

- Only tagging jet and tau-lepton information:

$$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}\}$$

- Information about all final state constituents (tagging jets, tau-lepton and Higgs-boson):

$$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}, \vec{P}_{\mathrm{Higgs}}\}$$

- Information about all final state constituents (tagging jets, tau-lepton and Higgs-boson), with all azimuthal angles Φ rotated such that $\Phi_{\mathrm{Higgs}} = 0$, to include topological information about the considered VBF Higgs-boson production process (see section 2.1):

$$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}, \vec{P}_{\mathrm{Higgs}}\}_{\Phi_{\mathrm{Higgs}} \equiv 0}$$

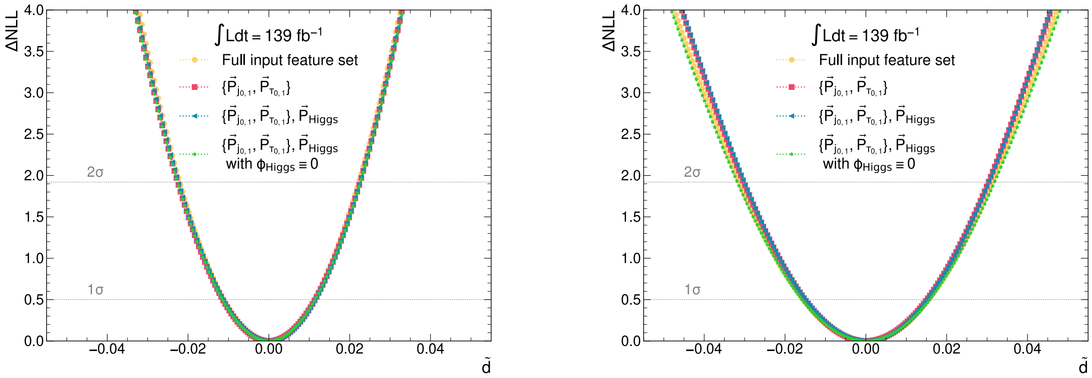
Here, $\vec{P}$ denote the three momenta of the respective final state constituents

$$\vec{P} = (P^T, \eta, \Phi).$$

With each of these subsets, a new NN is trained and optimized over 100 trials with `OPTUNA`, using the search space defined in section 6.3.1. Table B.3 contains the results of these studies.

A CP-sensitive observable $\mathcal{O}_{\mathrm{NN}}^{3\mathrm{Class}}$ is constructed from the predictions of each of these NNs. Figure A.4 shows the corresponding distributions and figure 6.11 displays the resulting ΔNLL-curves.

Table 6.10 contains the 1σ and 2σ confidence intervals and their lengths obtained from these curves when only considering singal events. The results that are obtained when including background events are presented in table 6.11. In the signal only case, rotating the Φ-components such that $\Phi_{\mathrm{Higgs}} = 0$ does not significantly influence the length of the confidence intervals within their uncertainty $\Delta\ell = 0.0007$. This is expected, since under these rotations, all physical information contained in the correlation of the Φ components is conserved. In the signal only case, including all input features does not lead to a significant improvement of the results compared to only including the kinematic information of the tagging jets and the tau-leptons at 1σ confidence level. At 2σ confidence level including the full set of input features improves the results by only 2%. Notably, the distributions obtained with the $\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}\}$ and $\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}, \vec{P}_{\mathrm{Higgs}}\}$ subsets are more resistant to background effects than the distributions obtained with full set of input features and the subset where the Φ-components are rotated. For $\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}\}$ and $\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}, \vec{P}_{\mathrm{Higgs}}\}$ the confidence interval lengths increase by 32% when including background compared to the length obtained when only considering signal events. The 2σ intervals increase by 34% after including background. When rotating the Φ components, including background yields an increase of 37% at the 1σ confidence level and 40% at the 2σ confidence level. For the full set of input features including background yields an increase of 38% at the 1σ confidence level and 41% at 2σ confidence level.



(a) Signal only (b) Including background

Figure 6.11: ΔNLL-curves obtained from the distributions of the CP-sensitive observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ constructed from the outputs of NNs that were trained with different sets of input features when (a) considering only signal events and (b) also including background events. Here, $\vec{P}$ denote the three momenta of the respective final state constituents.

	1σ CI	1σ CI length	2σ CI	2σ CI length
Full input feature set	[-0.0112, 0.0112]	0.0224	[-0.0221, 0.0222]	0.0443
$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}\}$	[-0.0114, 0.0114]	0.0228	[-0.0226, 0.0226]	0.0452
$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}, \vec{P}_{\mathrm{Higgs}}\}$	[-0.0114, 0.0114]	0.0228	[-0.0225, 0.0225]	0.0450
$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}, \vec{P}_{\mathrm{Higgs}}\}$ with $\Phi_{\mathrm{Higgs}} \equiv 0$	[-0.0115, 0.0115]	0.0230	[-0.0226, 0.0227]	0.0453

Table 6.10: **Signal only:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ obtained with the multiclass NN that was trained and optimized with different subsets of the initial input feature set defined in section 6.2. Here, $\vec{P}$ denote the three momenta of the respective final state constituents. For comparison, the results obtained with the multiclass neural network trained and optimized with the full input feature set in section 6.3.2 are also included.

	1σ CI	1σ CI length	2σ CI	2σ CI length
Full input feature set	[-0.0155, 0.0155]	0.0310	[-0.0313, 0.0313]	0.0626
$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}\}$	[-0.0150, 0.0151]	0.0301	[-0.0302, 0.0303]	0.0605
$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}, \vec{P}_{\mathrm{Higgs}}\}$	[-0.0150, 0.0150]	0.0300	[-0.0302, 0.0303]	0.0605
$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}, \vec{P}_{\mathrm{Higgs}}\}$ with $\Phi_{\mathrm{Higgs}} \equiv 0$	[-0.0157, 0.0157]	0.0314	[-0.0316, 0.0316]	0.0632

Table 6.11: **Including background:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ obtained with the multiclass NN that was trained and optimized with different subsets of the initial input feature set defined in section 6.2. Here, $\vec{P}$ denote the three momenta of the respective final state constituents. For comparison, the results obtained with the multiclass neural network trained and optimized with the full input feature set in section 6.3.2 are also included.

6.3.6 Linear Weights only in Training

Ideally, the neural network should only learn interference specific characteristics. However, the weights in eq. (4.19) that are used to incorporate the interference information during the training process do not purely consist of the desired interference term. They additionally contain a term that is quadratic in $\tilde{d}$. This term is CP-even and does not provide a source of CP-violation but only increases the overall production rate. This section investigates whether the NN is biased by the presence of this additional CP-even term during the training.

For this a NN is trained while only considering the part of the weights in eq. (4.19) that is linear in $\tilde{d}$ in the loss function eq. (6.16). This way, the network discriminates the interference events purely based on characteristics specific for positive/ negative interference.

Based on the results obtained previously in this chapter, this study uses a multiclass NN that is trained with the full set of input features defined in section 6.2 while assuming the $\tilde{d} = 1$-scenario. The NN architecture and hyperparameters are optimized with `OPTUNA`, over 100 trials using the search space defined in section 6.3.1. Table B.4 contains the results of this optimization.

Figure 6.12 shows the distribution of the observable $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\mathrm{lin}}$ constructed from the predictions of this neural network on the full dataset. The distribution is shown as predicted by the SM, as well as for two CP-violating scenarios with $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$.

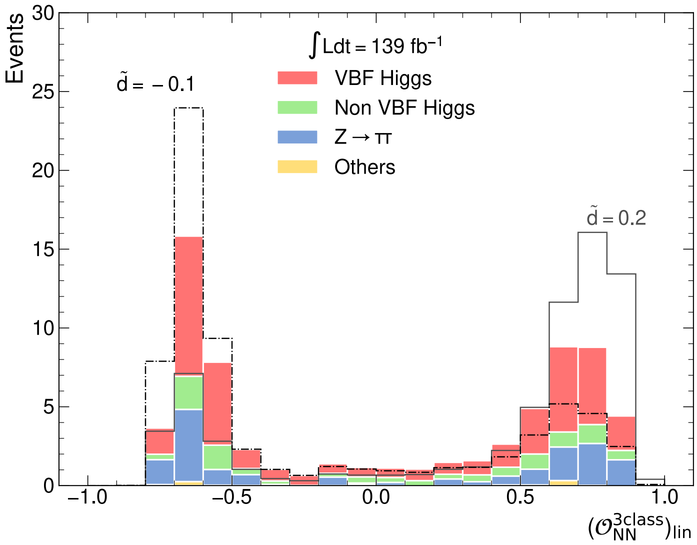
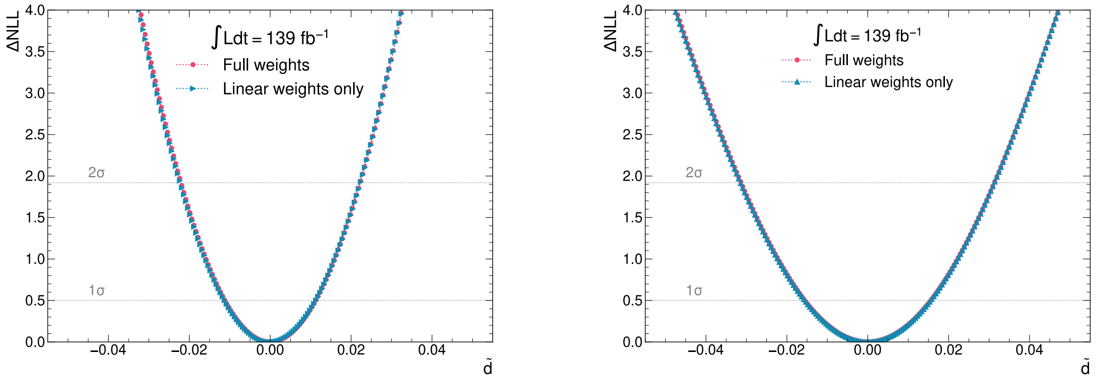


Figure 6.12: Distribution of the CP-sensitive observable $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\mathrm{lin}}$ constructed from the prediction by the multiclass NN that was trained and optimized while only using the linear term in the weights given in eq. (4.19) inside the loss function, as predicted by the SM for an integrated luminosity of $\int \mathrm{Ldt} = 139\ \mathrm{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. Additionally, distributions predicted for two CP-violating scenarios with the strengths $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$ are presented. These distributions are normalized to the total number $N_{\mathrm{SM}} = 69$ of expected SM events.

In Figure 6.13 the resulting ΔNLL curves are shown separately for the case where only signal events are considered and for the case where background events are included. For comparison, the curves corresponding to the distribution obtained with the multiclass model discussed in section 6.3.2 which was trained using the full weights in eq. (4.19) are included. Table 6.12 contains the confidence intervals extracted from these curves for the case where only signal events are considered. Table 6.13 shows the results obtained when including background events. Within the considered precision $\Delta\ell = 0.0007$, no changes in CP-sensitivity are observed.



(a) Signal only

(b) Including background

Figure 6.13: ΔNLL-curves obtained from the distribution of the CP-sensitive observable $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{lin}}$ constructed from the multiclass NN that was trained and optimized while only using the linear term in the weights given in eq. (4.19) inside the loss function when (a) considering only signal events and (b) also including background events. For comparison, the results obtained using the multiclass neural network trained and optimized while considering the full weights (see section 6.3.2) are also included.

	1σ CI	1σ CI length	2σ CI	2σ CI length
Full weights	[-0.0112, 0.0112]	0.0224	[-0.0221, 0.0222]	0.0443
Linear weights only	[-0.0113, 0.0113]	0.0226	[-0.0222, 0.0223]	0.0445

Table 6.12: **Signal only:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distribution of the observable $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{lin}}$ constructed from the multiclass NN that was trained and optimized while only using the linear term in the weights given in eq. (4.19) inside the loss function. For comparison, the results obtained with the multiclass neural network trained and optimized with interference samples created using the full weights (see section 6.3.2) are also included. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta \ell = 0.0007$.

	1σ CI	1σ CI length	2σ CI	2σ CI length
Full weights	[-0.0155, 0.0155]	0.0310	[-0.0313, 0.0313]	0.0626
Linear weights only	[-0.0156, 0.0156]	0.0312	[-0.0315, 0.0313]	0.0628

Table 6.13: **Including background:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distribution of the observable $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{lin}}$ constructed from the multiclass NN that was trained and optimized while only using the linear term in the weights given in eq. (4.19) inside the loss function. For comparison, the results obtained with the multiclass neural network trained and optimized with interference samples created using the full weights (see section 6.3.2) are also included. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta \ell = 0.0007$.

6.3.7 Final Classification Neural Network Observable

Based on the outcomes of the studies performed in the previous sections, a final classification neural network is constructed and trained such that it fulfills the following conditions:

- Multiclass NN with three classes: P_+, P_- and P_{SM}
- Creation of training samples for $\tilde{d} = 1$ scenario
- Full input feature set as defined in section 6.2 used in training
- Full weights in eq. (4.19), including the quadratic term used in training

In order to examine whether it is possible to reach a further increase of CP-sensitivity compared to what was obtained under these conditions with the NN constructed in section 6.3.2, another optimization study is performed. For this purpose, the ranges for the hyperparameter search are adapted, resulting in the following search space:

- Number of hidden layers $\in \{3, 4, 5, 6, 7, 8, 9\}$
- Number of nodes (individually for each layer) $\in [50, 400]$, scanned linearly
- Strength of L2 regularization $\in [10^{-6}, 10^{-3}]$, scanned logarithmically
- Initial learning rate $\in [10^{-4}, 10^{-1}]$, scanned logarithmically
- Learning rate decay steps $\in [1000, 10000]$, scanned linearly
- Learning rate decay parameter $\in [0, 0.4]$, scanned linearly
- Batch size $\in [30, 80]$, scanned linearly

The optimization study is performed over 1000 trials using `OPTUNA`. Table 6.14 shows the architecture and hyperparameters of the final classification neural network chosen based on the outcomes of this study.

Optimized hyperparameter	Study outcome
Number of hidden layers	9
Number of nodes in hidden layer 1	366
Number of nodes in hidden layer 2	125
Number of nodes in hidden layer 3	180
Number of nodes in hidden layer 4	92
Number of nodes in hidden layer 5	352
Number of nodes in hidden layer 6	199
Number of nodes in hidden layer 7	321
Number of nodes in hidden layer 8	373
Number of nodes in hidden layer 9	50
L2 regularization strength	$1.2 \cdot 10^{-6}$
Initial learning rate	$5.7 \cdot 10^{-3}$
Learning rate decay steps	4820.8
Learning rate decay rate	$9.0 \cdot 10^{-2}$
Batch size	32

Table 6.14: Architecture and hyperparameters of the final multiclass neural network as found in a hyperparameter optimization study using `OPTUNA`.

Figure 6.14 shows the distribution of the observable $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{final}}$ constructed from the predictions of this NN on the full dataset. Here, the distribution is shown as predicted by the SM and for two CP-violating scenarios with $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$.

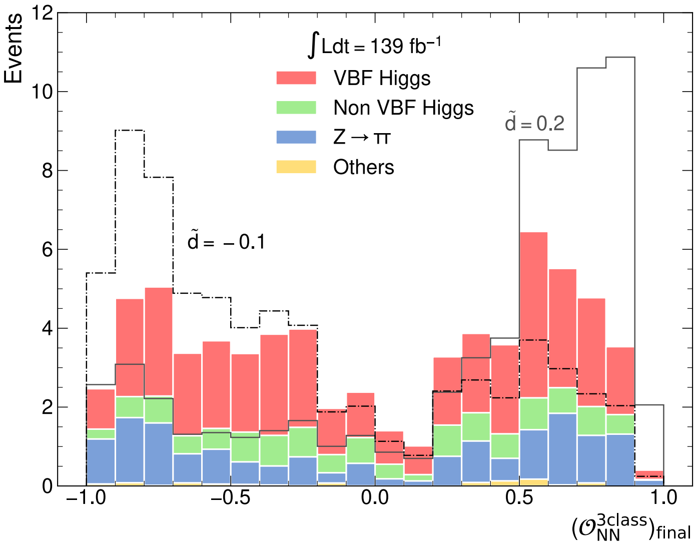


Figure 6.14: Distribution of the final classification NN obserbable $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{final}}$, as predicted by the SM for an integrated luminosity of $\int \text{Ldt} = 139\ \text{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. Additionally, distributions predicted for two CP-violating scenarios with the strengths $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$ are presented. These distributions are normalized to the total number $N_{\text{SM}} = 69$ of expected SM events.

Figure 6.15 shows the ΔNLL curves extracted from these distributions. The curves are shown for both, the signal only case and the case where background events are included. Table 6.15 contains the corresponding 1σ and 2σ confidence intervals. In the signal only case, the length of the 1σ confidence interval is consistent with the value reached with the less complex multiclass model constructed in section 6.3.2 (compare table 6.4). For the 2σ interval, a decrease in length by 2% is achieved. When including background events, the interval lengths also decrease by 2% compared to the results found in table 6.5 with the less complex multiclass model. This is true at both confidence levels.

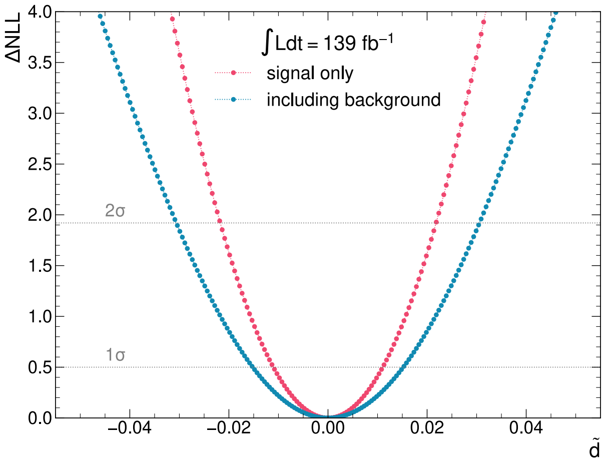


Figure 6.15: ΔNLL curves and resulting 1σ and 2σ confidence intervals, obtained with the $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\mathrm{final}}$ distribution, shown for the case where only signal events are considered (red) as well as for the case where background events are included (blue).

	1σ CI	1σ CI length	2σ CI	2σ CI length
Signal Only	[-0.0110, 0.0110]	0.0220	[-0.0217, 0.0218]	0.0435
Including Background	[-0.0152, 0.0152]	0.0304	[-0.0306, 0.0306]	0.0612

Table 6.15: 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the final classification NN observable $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\mathrm{final}}$. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta \ell = 0.0007$.

Figure 6.16 illustrates, that in the case of small $\tilde{d}$-values, the mean values of $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\mathrm{final}}$ in different BSM scenarios, show a linear dependency on $\tilde{d}$, as it is expected for CP-odd observables from eq. (4.11).

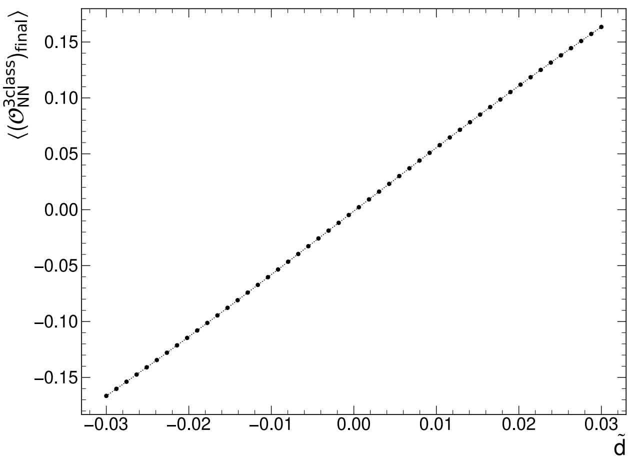


Figure 6.16: Linear dependency between the mean value of $(\mathcal{O}_{\mathrm{NN}}^{3\mathrm{Class}})_{\mathrm{final}}$ and $\tilde{d}$ in case of small $\tilde{d}$-values. Only signal events are considered.

6.3.8 Feature Importance

To investigate the impact of each input feature on the performance of the final classification NN, the feature importance score is calculated according to eq. (6.9). To reduce fluctuations caused by the specific choice of the initial weights used when training the NN, the feature importance scores are calculated 20 times, each time with a random choice of initial weights. The final feature importance scores, shown in figure 6.17, are given as the mean values of these 20 trials. The highest score is reached by $P^T_{j_0} P^T_{j_1} \sin \Delta\Phi^{\mathrm{sgn}}_{\mathrm{jj}}$. This is expected, since this high level feature is obtained from a symbolic regression which targets the highly CP-sensitive $\mathcal{OO}$ (see section 6.5). Since $\Delta\Phi^{\mathrm{sgn}}_{\mathrm{jj}}$ is strongly correlated with $P^T_{j_0} P^T_{j_1} \sin \Delta\Phi^{\mathrm{sgn}}_{\mathrm{jj}}$ it does not introduce new information to the neural network when $P^T_{j_0} P^T_{j_1} \sin \Delta\Phi^{\mathrm{sgn}}_{\mathrm{jj}}$ is already contained in the input set, explaining its low impact on the NN performance. Notably, the low level input feature η_{j_0} also yields significant contributions to the learning process. This could be attributed to the fact that it contains important spatial information, making it sensitive to the effects of CP-violation.

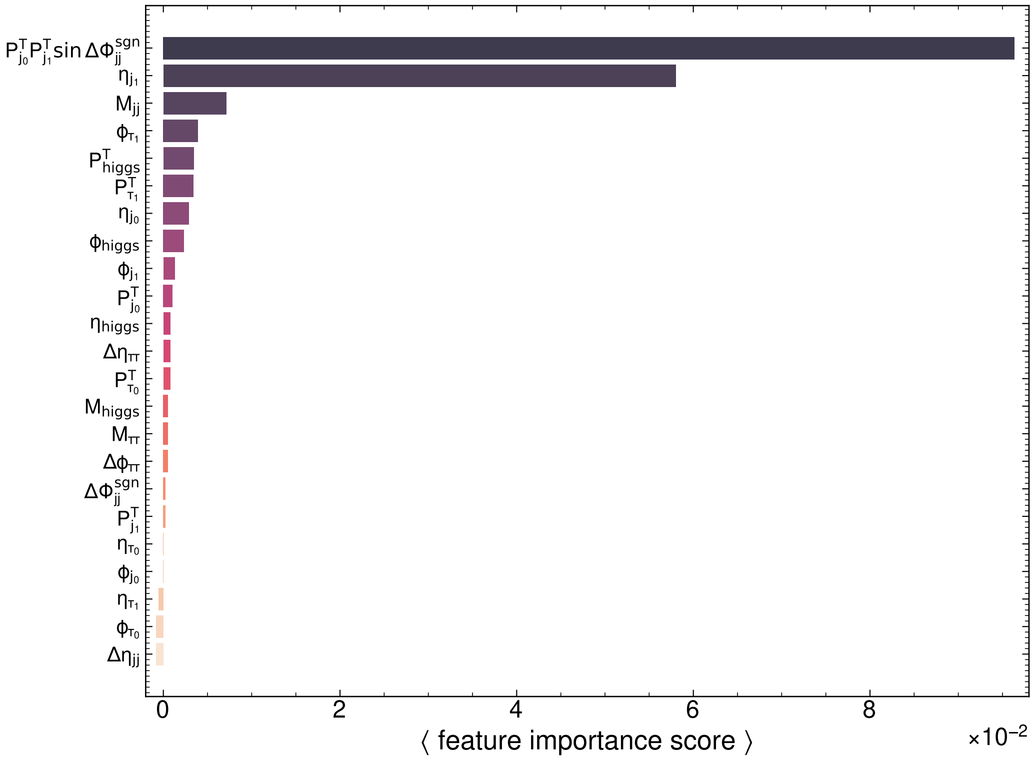


Figure 6.17: Feature importance score for all input features used in the training of the final classification neural network.

6.4 Regression Neural Network Observable $\mathcal{O}_{\text{NN}}^{\text{Reg}}$

Since the Optimal Observable contains the full information about the seven dimensional phase space in case of small $\tilde{d}$-values it is a very powerful tool for testing CP-invariance. This section aims to resemble the distribution of the Optimal Observable using the output of a regression neural network.

6.4.1 Optimization and Training of the Regression Neural Networks with different Sets of Input Variables

Two regression NNs are trained and optimized with different input feature sets. These NNs have one output node that returns predictions $\mathcal{O}_{\text{NN}}^{\text{Reg}}$, targeting the truth level $\mathcal{OO}$ distribution. The loss is then defined by the mean squared error between prediction and target

$$\text{MSE} = \frac{\sum_i w_{\text{SM},i}(\mathcal{O}_{\text{NN},i}^{\text{Reg}} - \mathcal{OO}_i)^2}{N_{\text{SM}}}, \tag{6.17}$$

where w_i are the nominal event weights given in eq. (3.2), discussed in section 3.1.1 and $N_{\text{SM}} = 40$ is the total number of signal events predicted by the SM.

The first NN is provided with only the three-momenta $\vec{P}_{j_{0,1}}, \vec{P}_{\tau_{0,1}}$ of the final state tagging jets and tau-leptons, whereas the second NN receives the full set of input features defined in table 6.1. In both cases, only SM signal events are considered in training. The data is split in training and validation sets in the ratio 80/20.

The architectures and hyperparameters of the regression neural networks are chosen based on the results of optimization studies that are conducted within the framework `OPTUNA` over 1000 trials in the following search space

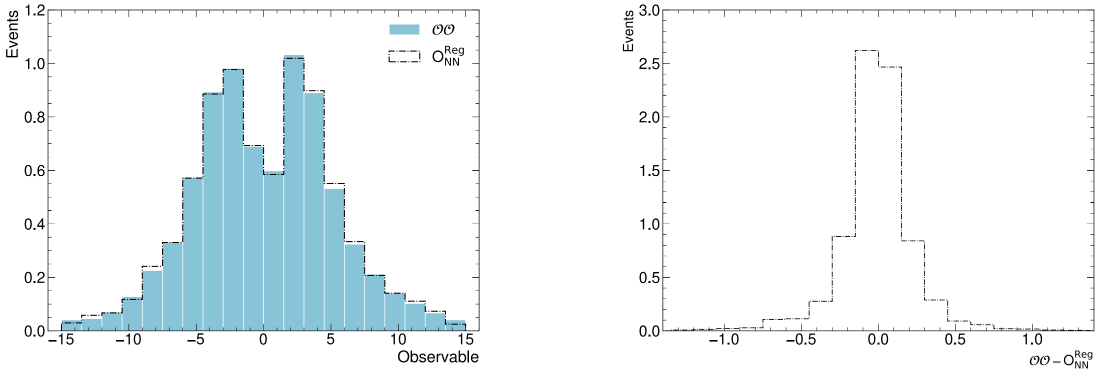
- Number of hidden layers $\in \{3, 4, 5, 6, 7\}$
- Number of nodes (individually for each layer) $\in [50, 400]$, scanned linearly
- Strength of L2 regularization $\in [10^{-7}, 10^{-1}]$, scanned logarithmically
- Initial learning rate $\in [10^{-5}, 10^{-1}]$, scanned logarithmically
- Learning rate decay steps $\in [1000, 5000]$, scanned linearly
- Learning rate decay parameter $\in [0, 0.7]$, scanned linearly
- Batch size $\in [50, 90]$, scanned linearly

Table 6.16 contains the results of these studies.

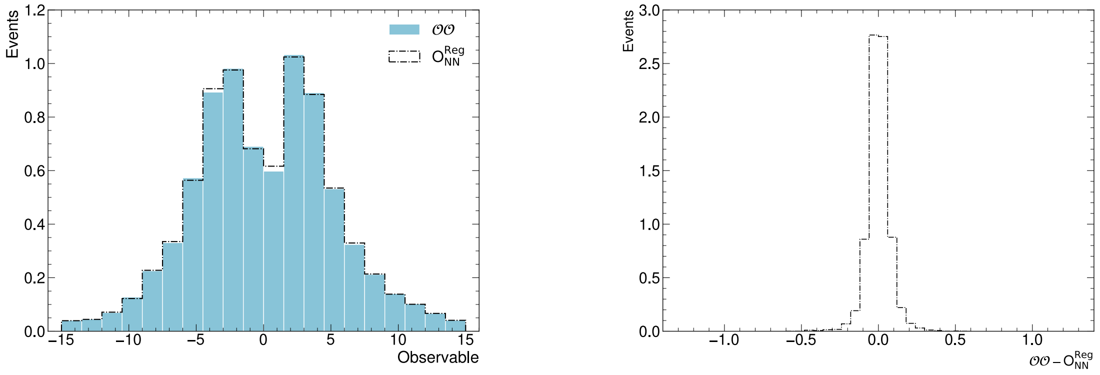
Optimized hyperparameter	$\vec{P}_{j_{0,1}}, \vec{P}_{\tau_{0,1}}$	Full input feature set
Number of hidden layers	4	6
Number of nodes in hidden layer 1	298	372
Number of nodes in hidden layer 2	235	272
Number of nodes in hidden layer 3	400	161
Number of nodes in hidden layer 4	384	143
Number of nodes in hidden layer 5	-	79
Number of nodes in hidden layer 6	-	119
L2 regularization strength	$2.1 \cdot 10^{-7}$	$5.2 \cdot 10^{-7}$
Initial learning rate	$6.5 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$
Learning rate decay steps	3734.3	1803.9
Learning rate decay rate	0.3	0.6
Batch size	87	85

Table 6.16: Architecture and hyperparameters of the regression NN trained and optimized with only the three momenta $\vec{P}_{j_{0,1}}, \vec{P}_{\tau_{0,1}}$ of the final state tagging jets and τ-leptons and the regression NN trained with the full set of input features.

Figure 6.18 compares the regression neural network observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{Reg}}$, that are predicted for the validation datasets to the targeted $\mathcal{OO}$. Providing additional and more complex information in the training process by including all input features allows the NN to resemble the targeted distribution more accurately and thus results in a significant decrease in event by event differences between $\mathcal{O}_{\mathrm{NN}}^{\mathrm{Reg}}$ and $\mathcal{OO}$.



(a) NN trained with only $P^{\mu}_{j_{0,1}}, P^{\mu}_{\tau_{0,1}}$ as input



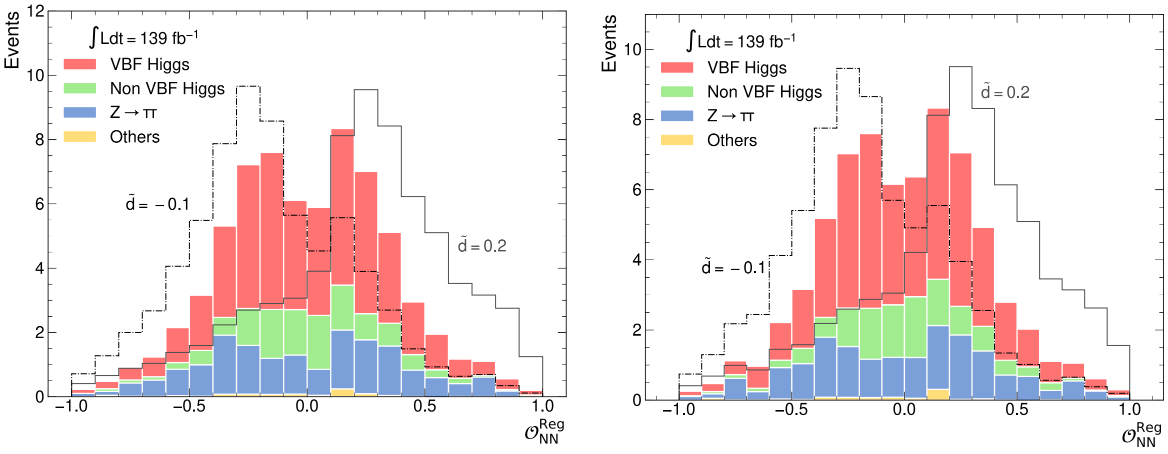
(b) NN trained with full input feature set

Figure 6.18: Left: Distributions of the regression NN predictions on the validation dataset $\mathcal{O}_{\text{NN}}^{\text{Reg}}$ compared to the targeted $\mathcal{OO}$. Right: Event by event differences between $\mathcal{OO}$ and $\mathcal{O}_{\text{NN}}^{\text{Reg}}$. The first row shows the results for the NN trained with only the three momenta $\vec{P}_{j_{0,1}}, \vec{P}_{\tau_{0,1}}$ of the final state tagging jets and τ-leptons and the second row shows the results for the NN trained with the full input feature set. This validation dataset contains 20% of the full SM signal sample. This corresponds to an expected number of events $N_{\text{val}} = 8$.

However, the primary goal is for the regression observable to capture the essential characteristics of the $\mathcal{OO}$ that contribute to its CP sensitivity. The regression NN is applied to the full dataset defined in chapter 3 to obtain the observable distributions of $\mathcal{O}_{\text{NN}}^{\text{Reg}}$ shown in figure 6.19. These distributions are shown as predicted by the SM and additionally for two CP-violating scenarios with $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$. Each distribution is normalized as follows

$$\mathcal{O}_{\text{NN}}^{\text{Reg}} = \frac{\mathcal{O}_{\text{NN,org}}^{\text{Reg}}}{\max(\mathcal{O}_{\text{NN,org}}^{\text{Reg}})}, \tag{6.18}$$

where in both cases $\max(\mathcal{O}_{\text{NN,org}}^{\text{Reg}}) = 15$.

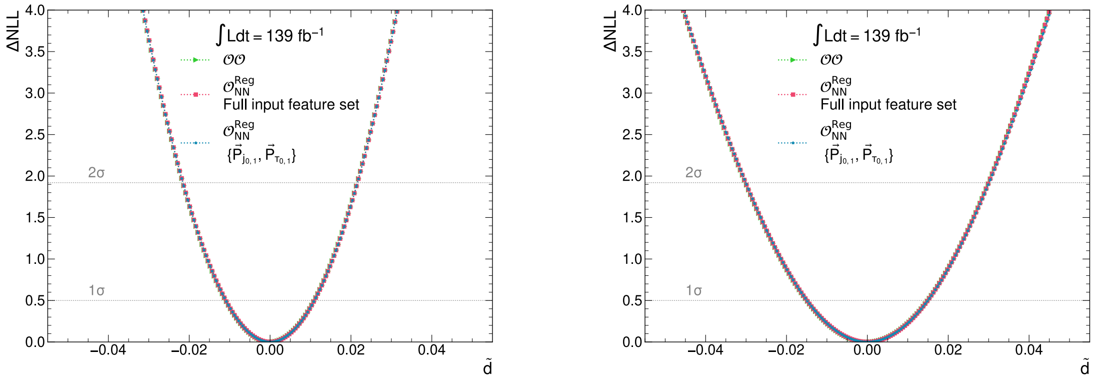


(a) Only $\{\vec{P}_{j_{0,1}}, \vec{P}_{\tau_{0,1}}\}$ as input

(b) Full input feature set

Figure 6.19: Distributions of the observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{Reg}}$ obtained from regression NNs that were trained with different input feature sets, as predicted by the SM for an integrated luminosity of $\int \mathrm{Ldt} = 139\ \mathrm{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. Additionally, distributions predicted for two CP-violating scenarios with the strengths $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$ are presented. These distributions are normalized to the total number $N_{\mathrm{SM}} = 69$ of expected SM events..

Figure 6.20 displays the ΔNLL curves obtained from these distributions. For comparison, the curve obtained with the reconstruction level $\mathcal{OO}$ is also shown. Table 6.17 contains the confidence intervals and corresponding lengths extracted from these curves when only considering signal events. Table 6.18 displays the results that are obtained when also including background events. The CP-sensitives of both regression NN observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{Reg}}$ are compatible with each other and with the CP-sensitivity achieved by the reconstruction level $\mathcal{OO}$ (see section 5.1) within the uncertainty on the length of the confidence intervals of $\Delta\ell = 0.0007$. This is true at both confidence levels, when considering only signal events and when also including background events.



(a) Signal only

(b) Including background

Figure 6.20: ΔNLL-curves obtained from the distributions of the observables $\mathcal{O}_{\text{NN}}^{\text{Reg}}$ resulting from regression NNs that were trained with different input feature sets. Additionally, the curve for the reconstruction level $\mathcal{OO}$ (see section 5.1) is shown. These results are shown separately for the case where (a) only signal events are considered and (b) background events are included. $\vec{P}$ denote the three momenta of the respective final state constituents.

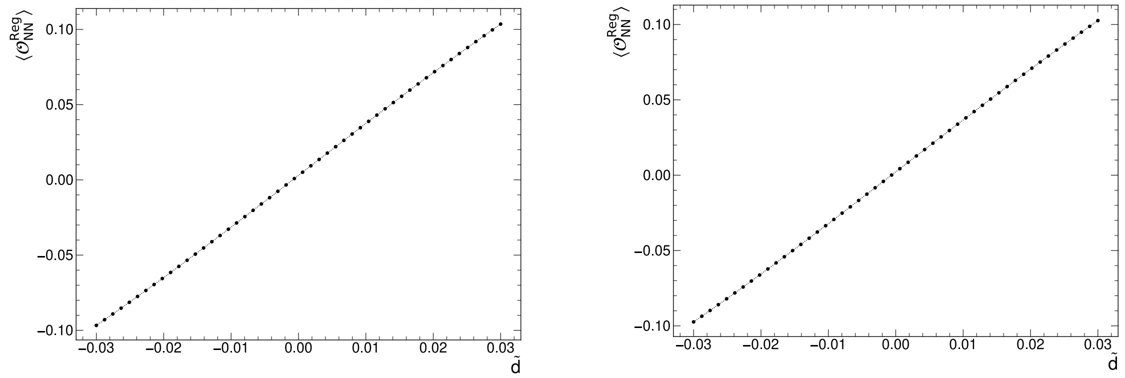
	1σ CI	1σ CI length	2σ CI	2σ CI length
$\mathcal{OO}$ (see section 5.1)	[-0.0109, 0.0109]	0.0218	[-0.0216, 0.0216]	0.0432
$\mathcal{O}_{\text{NN}}^{\text{Reg}}$ Full input feature set	[-0.0109, 0.0109]	0.0218	[-0.0216, 0.0216]	0.0432
$\mathcal{O}_{\text{NN}}^{\text{Reg}}$ $\{\vec{P}_{\text{j}0,1}, \vec{P}_{\tau0,1}\}$	[-0.0109, 0.0110]	0.0219	[-0.0216, 0.0216]	0.0432

Table 6.17: **Signal only:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the observables $\mathcal{O}_{\text{NN}}^{\text{Reg}}$ resulting from regression NNs that were trained and optimized with different sets of input features. $\vec{P}$ denote the three momenta of the respective final state constituents. For comparison, the results obtained with the reconstruction level $\mathcal{OO}$ are also shown. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta\ell = 0.0007$.

	1σ CI	1σ CI length	2σ CI	2σ CI length
$\mathcal{OO}$ (see section 5.1)	[-0.0150, 0.0149]	0.0299	[-0.0301, 0.0300]	0.0601
$\mathcal{O}_{\text{NN}}^{\text{Reg}}$ Full input feature set	[-0.0149, 0.0149]	0.0298	[-0.0301, 0.0300]	0.0601
$\mathcal{O}_{\text{NN}}^{\text{Reg}}$ $\{\vec{P}_{\text{j}0,1}, \vec{P}_{\tau0,1}\}$	[-0.0149, 0.0150]	0.0299	[-0.0301, 0.0301]	0.0602

Table 6.18: **Including background:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the observables $\mathcal{O}_{\text{NN}}^{\text{Reg}}$ resulting from regression NNs that were trained and optimized with different sets of input features. $\vec{P}$ denote the three momenta of the respective final state constituents. For comparison, the results obtained with the reconstruction level $\mathcal{OO}$ are also shown. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta\ell = 0.0007$.

Figure 6.21 illustrates that in the case of small $\tilde{d}$-values, the mean values of both regression observables $\mathcal{O}_{\text{NN}}^{\text{Reg}}$ show a linear dependency on $\tilde{d}$, as it is expected for CP-odd observables from eq. (4.11).



(a) Full set of input features

(b) Input: $\{\vec{P}_{\mathrm{j}_{0,1}}, \vec{P}_{\tau_{0,1}}\}$

Figure 6.21: Linear dependency between the mean values of the regression observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{Reg}}$ and $\tilde{d}$ in case of small $\tilde{d}$-values. Results are shown for the observables constructed with NNs that are trained and optimized using (a) the full input feature set defined in section 6.2 and (b) only the three momenta $\vec{P}$ of the tagging jets and the tau-leptons $\{\vec{P}_{\mathrm{j}_{0,1}}, \vec{P}_{\tau_{0,1}}\}$. Only signal events are considered.

6.5 Symbolic Regression Neural Network Observables

In contrast to the previous section, where the $\mathcal{OO}$ distribution was modeled, by utilizing the numerical output from a regression NN, this section aims to approximate the $\mathcal{OO}$ distribution through human-readable analytic expressions obtained via the method of symbolic regression (SR), following the approach presented in [19]. This study is performed at truth level. The resulting analytic expressions are then applied to reconstruction level data.

The symbolic regression neural networks considered in this thesis operate within the framework `PySR` [57]. For the construction of the desired analytic expressions, a tree structured search space, consisting of pre-defined variables and mathematical operators, represented as nodes, is defined. The search inside this space is performed using genetic programming (see section 6.1.4), following two objectives:

1. Minimizing a loss function, which in this thesis is given as the weighted mean squared error (MSE) between the target events $\mathcal{OO}_i$ and the respective analytical functions $f_i(\vec{x}_{\mathrm{in}})$ that are found to describe these events in terms of the input-features $\vec{x}_{\mathrm{in}}$

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n} w_{i,\tilde{d}} \cdot (f_i(\vec{x}_{\mathrm{in}}) - \mathcal{OO}_i)^2}{\sum_{i=1}^{n} w_{i,\tilde{d}}}, \tag{6.19}$$

where $w_{i,\tilde{d}}$ are the event weights defined eq. (4.19) for a given $\tilde{d}$-scenario.

2. Preferring expressions with low complexity, where

$$\text{complexity} \equiv \text{no. of nodes.} \tag{6.20}$$

`PySR` combines these two objectives into one single score

$$\text{score} = \text{NF} \cdot \text{MSE} + \text{SF} \cdot \text{complexity}, \tag{6.21}$$

where NF is a normalization factor, given as the inverse of the MSE between the data and the constant unit function (called `baseline` in `PySR`)

$$\text{NF} = \left(\frac{\sum_{i=1}^{n} w_{i,\tilde{d}} \cdot (1 - \mathcal{OO}_i)^2}{\sum_{i=1}^{n} w_{i,\tilde{d}}} \right)^{-1}, \tag{6.22}$$

and SF is a scaling factor, that allows to control how much the complexity contributes to the score. For this scaling factor (called `parsimony` in `PySR`) the default value SF$= 0.0032$ is used. By default, `PySR` uses the method of simulated annealing [60] to prevent getting trapped in a local minimum of the score function. This method defines a probability

$$P = e^{-\frac{\texttt{score}_{\text{new}} - \texttt{score}_{\text{old}}}{\texttt{alpha} \cdot T}} \tag{6.23}$$

for accepting mutations (see section 6.1.4). The parameter T, has an initial value of 1 and then linearly decreases with each mutation, such that in the last considered generation, it becomes 0. The number of generations to run can be controlled by the parameter `ncyclesperiteration`, whose default value 550 is used. The parameter `alpha` is also used at its default value, given as 0.1. If the score $\texttt{score}_{\text{new}}$ of a new function is lower than the score of a reference function $\texttt{score}_{\text{old}}$, this reflects in a positve sign of the exponent, which is used as a criterion to accept the new function. If the new score is larger than the reference score, the probability P for the new function to be accepted is exponentially suppressed, rather than strictly denied. In `PySR`, an iteration is defined by one full simulated annealing process. The number of considered iterations can be controlled by the parameter `niterations`, which in this study is set to 50. PySR provides a so-called hall of fame (HoF), where after each iteration, the algorithm saves the equation with the lowest score for a given complexity. New equations with higher complexities are only added to the HoF if their MSE is lower than the MSEs of previously saved equations. The final result of the symbolic regression is the equation with the best score and a MSE that is better than at least 1.5x the MSE of the most accurate expression. The migration of equations from different populations or the HoF is controlled by the parameters `fraction_replaced` $= 0.000364$ and `fraction_replaced_hof` $= 0.035$, both left at their default values.

In this thesis, the search space is defined by the following operators

- Binary operators: $+$, $-$, $\cdot$, $\div$
- Unary operators: cos, sin, exp, ln, sqrt

and by variables representing the following truth level information

$$\{\hat{P}^T_{p_0}, \phi_{p_0}, \eta_{p_0}, \hat{P}^T_{p_1}, \phi_{p_1}, \eta_{p_1}, \hat{P}^T_{\text{Higgs}}, \Phi_{\text{Higgs}}, \eta_{\text{Higgs}}, \Delta\Phi^{\text{sgn}}_{pp}, \Delta\eta_{\text{pp}}\}. \tag{6.24}$$

Here, the subscripts $p_{0,1}$ denote that the respective kinematic information belongs to the leading/ subleading outgoing partons. The transverse momenta, denoted as $\hat{P}^T_{\text{particle}}$, are rescaled to align their order of magnitude with the scale of the other input features,

$$\hat{P}^T_{\text{particle}} = \frac{P^T_{\text{particle}}}{M_{\text{higgs}}}. \tag{6.25}$$

This can help `PySR` with handling and processing the input features more efficiently.

6.5.1 Standard Model Target

First, symbolic regression is applied to find an analytic function f_{SM} that approximates the $\mathcal{OO}$ distribution as it is predicted by the SM. For this, the SM event weights w_{SM} discussed in chapter 3 are used when calculating the MSE via eq. (6.19). The result is

$$f_{\text{SM}} = A \cdot \hat{P}^T_{p_0} \hat{P}^T_{p_1} \sin \Delta\Phi^{\text{sgn}}_{pp}, \qquad \text{where } A = 5.9611454, \tag{6.26}$$

with the following MSE and complexity

$$\text{MSE} = 0.20$$
$$\text{complexity} = 8.$$

The full hall of fame is presented in table 6.19.

Complexity	Function	MSE
1	$\Delta\Phi_{pp}^{\mathrm{sgn}}$	7.89
3	$3.16961 \cdot \Delta\Phi_{pp}^{\mathrm{sgn}}$	4.77
4	$7.45406543983682 \cdot \sin \Delta\Phi_{pp}^{\mathrm{sgn}}$	3.05
5	$8.505658 \cdot \sin(\sin \Delta\Phi_{pp}^{\mathrm{sgn}})$	3.04
6	$8.505658 \cdot \hat{P}_{\mathrm{Higgs}}^{T} \sin \Delta\Phi_{pp}^{\mathrm{sgn}}$	1.89
8	$5.9611454 \cdot \hat{P}_{p_0}^{T} \hat{P}_{p_1}^{T} \sin \Delta\Phi_{pp}^{\mathrm{sgn}}$	0.20
10	$\hat{P}_{p_0}^{T}(5.6764827 \cdot \hat{P}_{p_1}^{T} + 0.3424119) \sin \Delta\Phi_{pp}^{\mathrm{sgn}}$	0.19
11	$\hat{P}_{p_0}^{T} \hat{P}_{p_1}^{T}(6.7836447 - \cos(\eta_{\mathrm{Higgs}})) \sin \Delta\Phi_{pp}^{\mathrm{sgn}}$	0.17
13	$\hat{P}_{p_0}^{T} \hat{P}_{p_1}^{T} \left(6.7836447 - \cos\left(\frac{\eta_{\mathrm{Higgs}}}{\Delta\Phi_{pp}^{\mathrm{sgn}}}\right)\right) \cdot \sin \Delta\Phi_{pp}^{\mathrm{sgn}}$	0.17
15	$\hat{P}_{p_0}^{T} \hat{P}_{p_1}^{T} \left(6.7836447 - \cos\left(\frac{\eta_{\mathrm{Higgs}} \hat{P}_{\mathrm{Higgs}}^{T}}{\eta_{p_1}}\right)\right) \cdot \sin \Delta\Phi_{pp}^{\mathrm{sgn}}$	0.15
17	$\hat{P}_{p_0}^{T} \hat{P}_{p_1}^{T} \left(6.7836447 - \cos\left(1.11295503027293 \cdot \frac{\eta_{\mathrm{Higgs}} \hat{P}_{\mathrm{Higgs}}^{T}}{\eta_{p_1}}\right)\right) \cdot \sin \Delta\Phi_{pp}^{\mathrm{sgn}}$	0.15
18	$\hat{P}_{p_0}^{T} \hat{P}_{p_1}^{T} \left(6.7836447 - \cos\left(\frac{\eta_{\mathrm{Higgs}} \hat{P}_{\mathrm{Higgs}}^{T}}{\eta_{p_1} \sqrt{\hat{P}_{\mathrm{p}_1}^{T}}}\right)\right) \cdot \sin \Delta\Phi_{pp}^{\mathrm{sgn}}$	0.15
19	$\hat{P}_{p_0}^{T} \hat{P}_{p_1}^{T} \left(6.7836447 - \cos\left(\frac{\eta_{\mathrm{Higgs}} \hat{P}_{\mathrm{Higgs}}^{T}}{\eta_{p_1} \sin \sqrt{\hat{P}_{\mathrm{p}_1}^{T}}}\right)\right) \cdot \sin \Delta\Phi_{pp}^{\mathrm{sgn}}$	0.14

Table 6.19: Hall of fame from symbolic regression targeting the $\mathcal{OO}$ distribution as it is predicted by the SM. These results are obtained at truth level.

The distribution obtained from eq. (6.26) when only considering (truth level) signal events is shown in figure 6.22. For comparison, the targeted truth level $\mathcal{OO}$ distribution is also displayed. Both distributions are shown as predicted by the SM ($\tilde{d} = 0$).

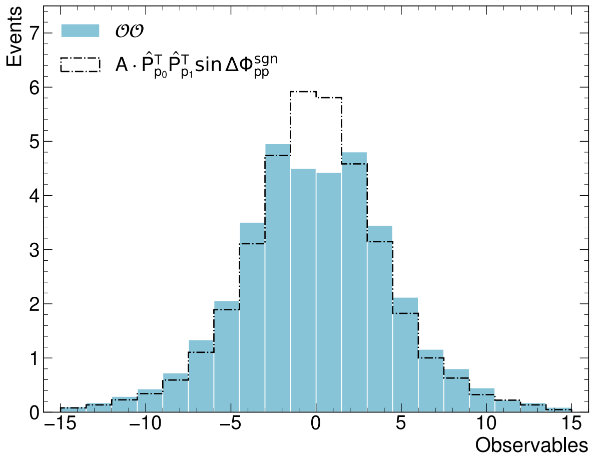


Figure 6.22: Approximation of truth level $\mathcal{OO}$ distribution by applying the analytic expression eq. (6.26) found through symbolic regression, targeting the truth level $\mathcal{OO}$ distribution as it is predicted by the SM. Both results are shown as predicted by the SM ($\tilde{d} = 0$) for signal events.

The result found in eq. (6.26) is now utilized to approximate the $\mathcal{OO}$ distribution at reconstruction level

$$\mathcal{OO} \sim P^T_{j0} P^T_{j1} \sin \Delta\Phi^{\mathrm{sgn}}_{\mathrm{jj}}. \tag{6.27}$$

Here, the parton information is replaced by the corresponding tagging jet information. The transverse momenta $P^T_{j0,1}$ are now given in their original magnitude. Since the primary focus lies in resembling the overall shape of the distribution, the scaling factor in eq. (6.26) is neglected.

The distribution obtained from eq. (6.27) when only considering signal events is shown in figure 6.23. For comparison, the reconstruction level $\mathcal{OO}$ distribution is also displayed. Both distributions are shown as predicted by the SM ($\tilde{d} = 0$). Here, the distributions are normalized as follows

$$\text{Observable} = \frac{\text{Observable}_{\text{org}}}{\max(\text{Observable}_{\text{org}})}, \tag{6.28}$$

where $\max([P^T_{j0} P^T_{j1} \sin \Delta\Phi^{\mathrm{sgn}}_{\mathrm{jj}}]_{\mathrm{org}}) = 45124$ and as discussed in section 5.1 $\max(\mathcal{OO}_{\mathrm{org}}) = 15$.

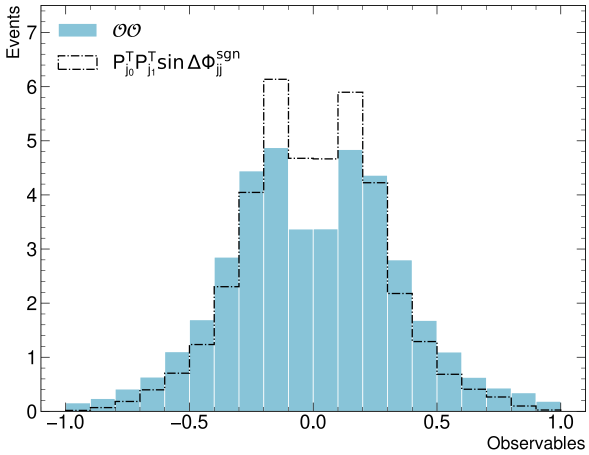


Figure 6.23: Approximation of reconstruction level $\mathcal{OO}$ distribution by applying the analytic expression eq. (6.27) found through symbolic regression, targeting the truth level $\mathcal{OO}$ distribution as it is predicted by the SM. Both results are shown as predicted by the SM ($\tilde{d} = 0$) for signal events and normalized according to eq. (6.28).

Remarkably, these straightforward and easily interpretable expressions, which solely rely on information about the final state tagging jets, allow resembling the overall shape of the $\mathcal{OO}$ distribution. This is true not only for the original target truth level information, but also when applied to the reconstruction level data.

6.5.2 Beyond the Standard Model Target

Now, the procedure introduced in the previous section is repeated, this time targeting the $\mathcal{OO}$ that is predicted in the CP-violating scenario $\tilde{d} = 0.5$. This is realized by considering the corresponding event weights $w_{i,\tilde{d}}$ according to eq. (4.19) when calculating the MSE eq. (6.19). Here, the number of BSM events is normalized to the number N_{SM} of predicted SM events, using the normalization factor

$$\mathrm{NF} = \frac{N_{\mathrm{SM}}}{\sum_i w_{i,\tilde{d}}}. \tag{6.29}$$

The symbolic regression yields the result

$$f_{\tilde{d}=0.5} = \hat{P}^T_{p0}\hat{P}^T_{p1}\left[\cos\Delta\Phi^{\mathrm{sgn}}_{pp} + B\right]\sin\Delta\Phi^{\mathrm{sgn}}_{pp}, \qquad \text{where B} = 6.281698, \tag{6.30}$$

with the following values for MSE and complexity

$$\begin{aligned}\mathrm{MSE} &= 0.72\\ \text{complexity} &= 11.\end{aligned}$$

The full hall of fame can be found in table 6.20.

Complexity	Function	MSE
1	$\Delta\Phi^{\mathrm{sgn}}_{pp}$	37.28
3	$6.1142375585507 \cdot \Delta\Phi^{\mathrm{sgn}}_{pp}$	16.68
4	$12.5012541883264 \cdot \sin\Delta\Phi^{\mathrm{sgn}}_{pp}$	15.60
5	$2.88670794254885 \cdot \hat{P}^T_{\mathrm{Higgs}}\Delta\Phi^{\mathrm{sgn}}_{pp}$	12.00
6	$8.373081 \cdot \hat{P}^T_{\mathrm{p}_1}\sin\Delta\Phi^{\mathrm{sgn}}_{pp}$	9.20
7	$6.02036774693938 \cdot \frac{\hat{P}^T_{\mathrm{Higgs}}\Delta\Phi^{\mathrm{sgn}}_{pp}}{\eta_{p_0}}$	8.85
8	$6.01319 \cdot \hat{P}^T_{p_0}\hat{P}^T_{p_1}\sin\Delta\Phi^{\mathrm{sgn}}_{pp}$	0.87
10	$5.6258187 \cdot \hat{P}^T_{p_0}(\hat{P}^T_{p_1} + 0.12879014) \cdot \sin\Delta\Phi^{\mathrm{sgn}}_{pp}$	0.86
11	$\hat{P}^T_{p_0}\hat{P}^T_{p_1}(\cos\Delta\Phi^{\mathrm{sgn}}_{pp} + 6.281698) \cdot \sin\Delta\Phi^{\mathrm{sgn}}_{pp}$	0.72
13	$\hat{P}^T_{p_0}\hat{P}^T_{p_1}\left(\hat{P}^T_{\mathrm{Higgs}} + 3.0322027\right)\sin\Delta\Phi^{\mathrm{sgn}}_{pp} + \sin\Delta\Phi^{\mathrm{sgn}}_{pp}$	0.71
15	$\hat{P}^T_{p_0}\hat{P}^T_{p_1}\left(\log(-\hat{P}^T_{p_1}) + e^{\sqrt{\hat{P}^T_{\mathrm{Higgs}}}} + 4.9699063\right) \cdot \sin\Delta\Phi^{\mathrm{sgn}}_{pp}$	0.69
16	$\hat{P}^T_{p_0}\hat{P}^T_{p_1}\left(1.8441851 \cdot \log(\hat{P}^T_{\mathrm{Higgs}} + 3.9669242))\right)) \cdot \sin\Delta\Phi^{\mathrm{sgn}}_{pp} + \sin\Delta\Phi^{\mathrm{sgn}}_{pp}$	0.66
19	$\sqrt{\hat{P}^T_{p_1}} \cdot \left(5.2899504 \cdot \hat{P}^T_{p_0} - 5.2899504 \cdot (-\sqrt{\hat{P}^T_{\mathrm{Higss}}} + \cos(\hat{P}^T_{p_1} - 0.3397412)) \cdot \eta^{-1}_{p_0}\right) \cdot \sin\Delta\Phi^{\mathrm{sgn}}_{pp}$	0.58
20	$\sqrt{\hat{P}^T_{p_1}} \cdot \left(5.2899504 \cdot \hat{P}^T_{p_0} - 5.2899504 \cdot (-\sqrt{\hat{P}^T_{\mathrm{Higss}}} + \cos(\hat{P}^T_{p_1} - 0.3397412)) \cdot \eta^{-1}_{p_0}\right) \cdot \sin\Delta\Phi^{\mathrm{sgn}}_{pp} - \sin\Delta\Phi^{\mathrm{sgn}}_{pp}$	0.57

Table 6.20: Hall of fame from symbolic regression targeting the $\mathcal{OO}$ distribution as it is predicted for the $\tilde{d} = 0.5$ scenario. These results are obtained at truth level.

The distribution obtained from eq. (6.30) when only considering (truth level) signal events is shown in figure 6.24. For comparison, the targeted truth level $\mathcal{OO}$ distribution is also displayed. Both distributions are shown in the $\tilde{d} = 0.5$ scenario.

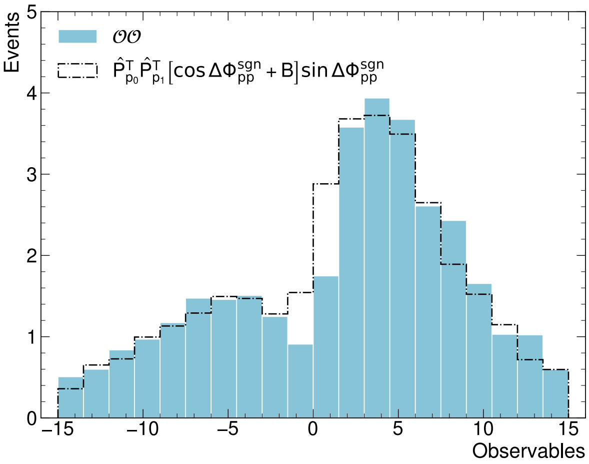


Figure 6.24: Approximation of truth level $\mathcal{O}\mathcal{O}$ distribution by applying the analytic expression eq. (6.26) found through symbolic regression, targeting the truth level $\mathcal{O}\mathcal{O}$ distribution as it is predicted in the $\tilde{d} = 0.5$ scenario. Both results are shown as predicted in the $\tilde{d} = 0.5$ scenario.

Again, the result found in eq. (6.30) is utilized to approximate the $\mathcal{O}\mathcal{O}$ distribution at reconstruction level

$$\mathcal{O}\mathcal{O} \sim P^T_{j_0} P^T_{j_1} \left[\cos\Delta\Phi^{\text{sgn}}_{jj} + B\right] \sin\Delta\Phi^{\text{sgn}}_{jj}, \qquad \text{where B} = 6.281698. \tag{6.31}$$

The distribution that results from eq. (6.31), when only considering signal events is shown in figure 6.25 as predicted for the $\tilde{d} = 0.5$ scenario. For comparison, the reconstruction level $\mathcal{O}\mathcal{O}$-distribution that is predicted for the $\tilde{d} = 0.5$ scenario is also displayed. Here, both distributions are normalized according to eq. (6.28), wit $\max([P^T_{j_0} P^T_{j_1} [\cos\Delta\Phi^{\text{sgn}}_{jj} + B] \sin\Delta\Phi^{\text{sgn}}_{jj}]_{\text{org}}) = 286586$ and as discussed in section 5.1 $\max(\mathcal{O}\mathcal{O}_{\text{org}}) = 15$.

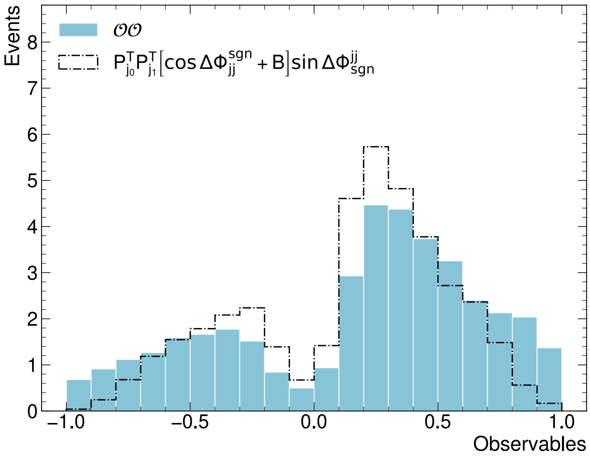


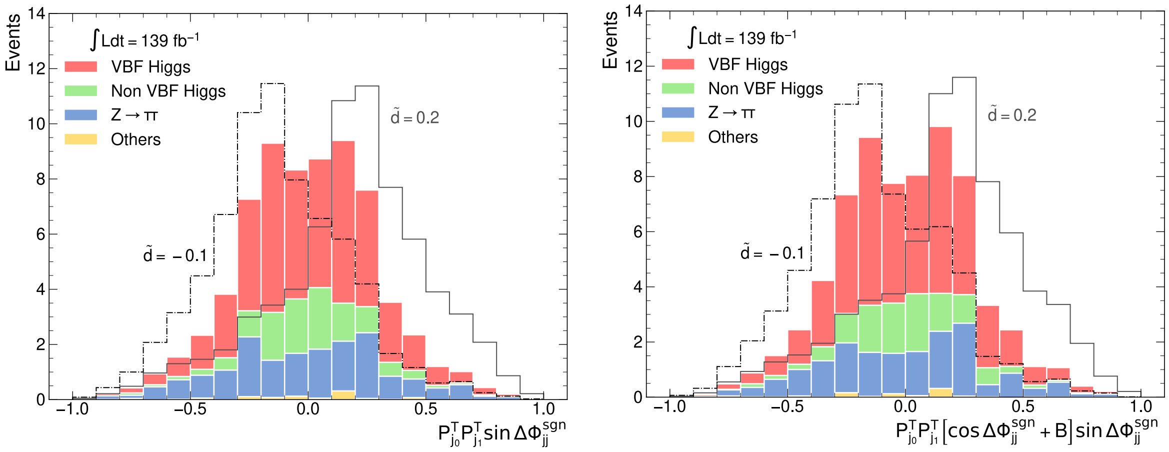
Figure 6.25: Reconstruction level $\mathcal{OO}$ distribution compared to distribution resulting from applying the analytical expression eq. (6.30), found through symbolic regression when targeting the truth level $\mathcal{OO}$ in the $\tilde{d} = 0.5$ scenario. Both distributions are shown as predicted for the $\tilde{d} = 0.5$ case. Only signal events are considered.

Describing the $\mathcal{OO}$ distribution in the BSM scenario with $\tilde{d} = 0.5$ requires more complexity compared to predicting the SM distribution. However, the result described in eq. (6.30) also relies solely on information of the final state tagging jets. It allows resembling of the overall shape of the $\mathcal{OO}$ distribution when applied to both the originally targeted truth level information and the reconstruction level data.

6.5.3 CP-Sensitivities

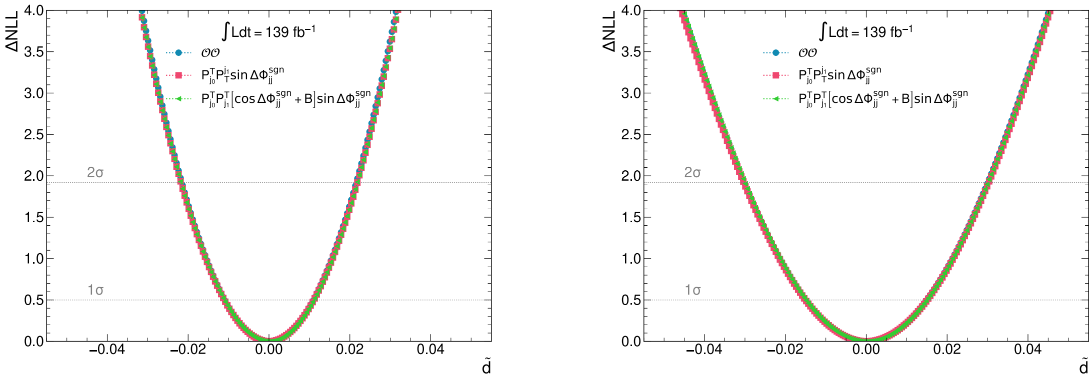
The distributions following from the analytical expressions eq. (6.27) and eq. (6.31) found in the previous sections, when considering the full dataset at reconstruction level, are shown in figure 6.26. Again, the distributions are normalized according to eq. (6.28), using the same maximum values as discussed in section 6.5.1 and section 6.5.2, since these were observed to remain the same when including background.

In Figure 6.27, the ΔNLL curves obtained from these distributions are displayed. The corresponding confidence-intervals as well as their lengths are presented in table 6.21 for the case where only signal events are considered and in 6.22 for the case where background events are included. Both analytical expressions yield observable distributions that allow to extract confidence interval lengths that are compatible with the results for the Optimal Observable within the uncertainty on the length of the confidence interval $\Delta\ell = 0.0007$. This is true at both confidence levels, when considering only signal events and when including background events in the calculations.



(a) Target: $\mathcal{OO}$ distribution as predicted by SM (b) Target: $\mathcal{OO}$ distribution predicted for $\tilde{d} = 0.5$

Figure 6.26: Distributions of the observables constructed from the analytical expressions found through symbolic regressions targeting the $\mathcal{OO}$ distribution that is predicted (a) by the SM and (b) in the $\tilde{d} = 0.5$ scenario, shown as predicted by the SM for an integrated luminosity of $\int$ Ldt = 139 fb^{-1}. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. Additionally, distributions predicted for two CP-violating scenarios with the strengths $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$ are presented. These distributions are normalized to the total number $N_{\text{SM}} = 69$ of expected SM events.



(a) Signal only (b) Including background

Figure 6.27: ΔNLL-curves obtained from the distributions of the observables constructed from the analytical expressions found through symbolic regressions targeting the $\mathcal{OO}$ distribution that is predicted by the SM ($P^T_{j0}P^T_{j1}\sin\Delta\Phi^{\text{sgn}}_{\text{jj}}$) and in the $\tilde{d} = 0.5$ scenario ($P^T_{j0}P^T_{j1}[\cos\Delta\Phi^{\text{sgn}}_{jj} + B]\sin\Delta\Phi^{\text{sgn}}_{jj}$). Additionally, the curve for the targeted $\mathcal{OO}$ is shown. These results are shown separately for the case where (a) only signal events are considered and (b) background events are included.

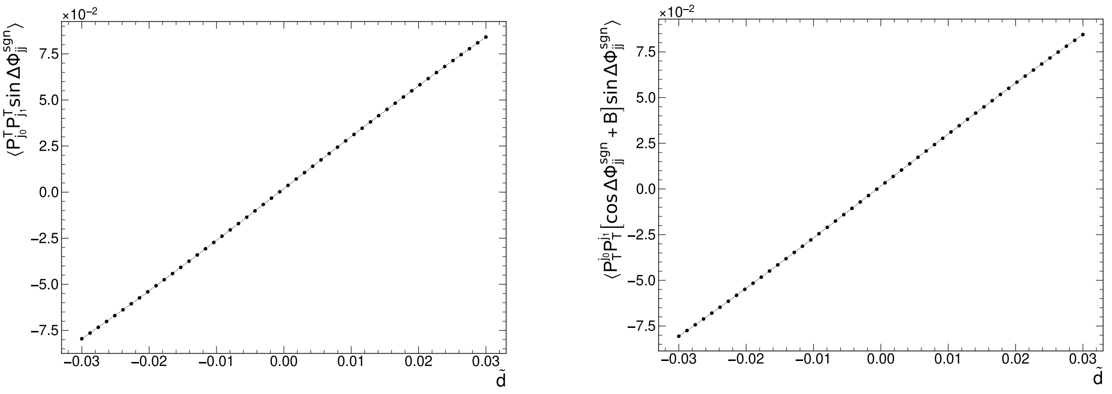
	1σ CI	1σ CI length	2σ CI	2σ CI length
$\mathcal{OO}$	[-0.0109, 0.0109]	0.0218	[-0.0216, 0.0216]	0.0432
$P^T_{j_0} P^{j_1}_T \sin\Delta\Phi^{\text{sgn}}_{jj}$	[-0.0110, 0.0111]	0.0221	[-0.0218, 0.0219]	0.0437
$P^T_{j_0} P^T_{j_1} \left[\cos\Delta\Phi^{\text{sgn}}_{\text{jj}} + B\right] \sin\Delta\Phi^{\text{sgn}}_{\text{jj}}$	[-0.0111, 0.0111]	0.0222	[-0.0218, 0.0218]	0.0436

Table 6.21: **Signal only:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the symbolic regression observables $P^T_{j_0} P^T_{j_1} \sin\Delta\Phi^{\text{sgn}}_{\text{jj}}$, found when targeting the $\mathcal{OO}$ distribution that is predicted by the SM and $P^T_{j_0} P^T_{j_1}[\cos\Delta\Phi^{\text{sgn}}_{jj} + B] \sin\Delta\Phi^{\text{sgn}}_{jj}$ found when targeting the $\mathcal{OO}$ distribution that is predicted for the $\tilde{d} = 0.5$ scenario. For comparison, the results obtained with the targeted $\mathcal{OO}$ are also shown. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta\ell = 0.0007$.

	1σ CI	1σ CI length	2σ CI	2σ CI length
$\mathcal{OO}$	[-0.0150, 0.0149]	0.0299	[-0.0301, 0.0300]	0.0601
$P^T_{j_0} P^{j_1}_T \sin\Delta\Phi^{\text{sgn}}_{jj}$	[-0.0150, 0.0151]	0.0301	[-0.0303, 0.0303]	0.0606
$P^T_{j_0} P^T_{j_1} \left[\cos\Delta\Phi^{\text{sgn}}_{\text{jj}} + B\right] \sin\Delta\Phi^{\text{sgn}}_{\text{jj}}$	[-0.0149, 0.0150]	0.0299	[-0.0300, 0.0301]	0.0601

Table 6.22: **Including background:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the symbolic regression observables $P^T_{j_0} P^T_{j_1} \sin\Delta\Phi^{\text{sgn}}_{\text{jj}}$, found when targeting the $\mathcal{OO}$ distribution that is predicted by the SM and $P^T_{j_0} P^T_{j_1}[\cos\Delta\Phi^{\text{sgn}}_{jj} + B] \sin\Delta\Phi^{\text{sgn}}_{jj}$ found when targeting the $\mathcal{OO}$ distribution that is predicted for the $\tilde{d} = 0.5$ scenario. For comparison, the results obtained with the targeted $\mathcal{OO}$ are also shown. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta\ell = 0.0007$.

Figure 6.28 illustrates that in the case of small $\tilde{d}$-values, the mean values of both symbolic regression observables show a linear dependency on $\tilde{d}$, as it is expected for CP-odd observables from eq. (4.11).



(a) Target: $\mathcal{OO}$ distribution as predicted by SM (b) Target: $\mathcal{OO}$ distribution predicted for $\tilde{d} = 0.5$

Figure 6.28: Linear dependency between the average values of the symbolic regression observables and $\tilde{d}$ in case of small $\tilde{d}$-values. Results are shown for (a) the observables $P^T_{j_0} P^T_{j_1} \sin \Delta\Phi^{\text{sgn}}_{\text{jj}}$, found when targeting the $\mathcal{OO}$ distribution that is predicted by the SM and (b) $P^T_{j_0} P^T_{j_1} [\cos \Delta\Phi^{\text{sgn}}_{jj} + B] \sin \Delta\Phi^{\text{sgn}}_{jj}$ found when targeting the $\mathcal{OO}$ distribution that is predicted for the $\tilde{d} = 0.5$ scenario. Only signal events are considered.

Chapter 7

Comparison of CP-sensitive Observables

In this chapter, the CP-sensitivities of the established CP-odd observables $\mathcal{OO}$ and $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$, that were introduced in chapter 5 and the machine learning CP-odd observables that were constructed in chapter 6 are compared. In chapter 6, several observables have been constructed with each machine learning method. This comparison considers only one result from each method.

Among the various classification NN observables obtained in section 6.3, the observable $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\mathrm{final}}$ is considered. This observable is constructed from a multiclass NN, that is trained and optimized using the full set of input features defined in section 6.2 and the full weights defined in eq. (4.19) assuming the $\tilde{d} = 1$ scenario.

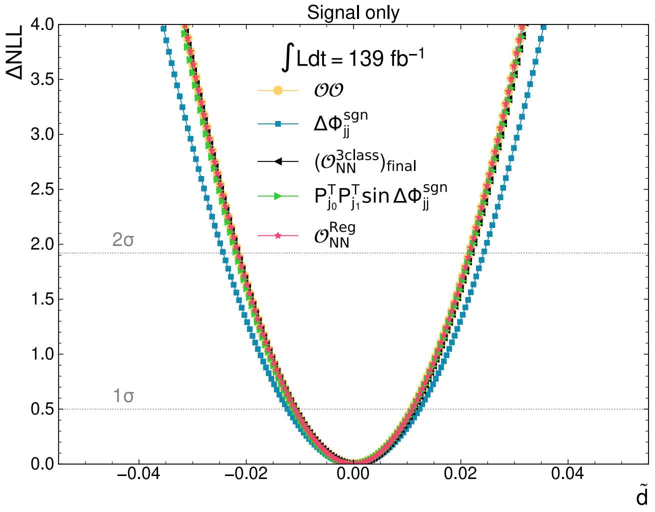
Two observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{Reg}}$ that resemble the $\mathcal{OO}$ distribution are obtained in section 6.4 from regression NNs that are trained and optimized with different sets of input features. Since within the considered uncertainties, the CP-sensitivities of these two observables are compatible with each other, in principle, it is arbitrary to choose one of them for the final comparison. However, for potential future applications a tool that requires less input information is favored. Thus, the result from the NN that only takes the three momenta $\vec{P}$ of the tagging jets and the tau-leptons as input features is chosen.

The symbolic regression observables found in section 6.5 also have compatible CP-sensitives. Here, $P_{j_0}^T P_{j_1}^T \sin\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ is preferred due to its lower complexity compared to $P_{j_0}^T P_{j_1}^T [\cos\Delta\Phi_{jj}^{\mathrm{sgn}} + B] \sin\Delta\Phi_{jj}^{\mathrm{sgn}}$.

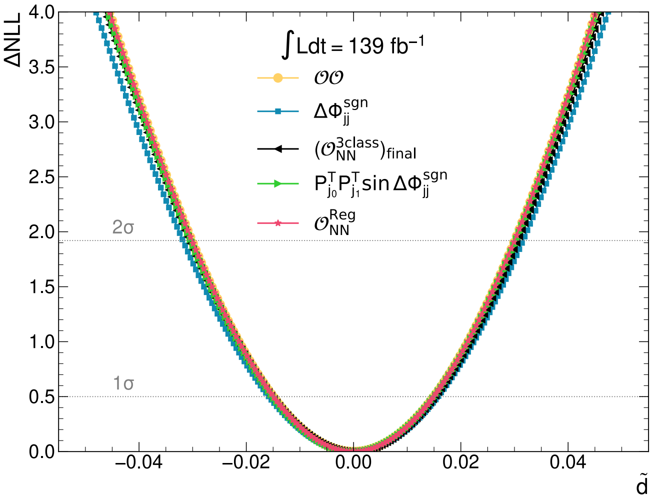
As in the previous chapters, the CP-sensitivities are estimated in terms of lengths of the confidence intervals extracted for $\tilde{d}$ from the ΔNLL curves that correspond to the respective distributions. The distributions of all considered CP-odd observables are divided into 20 evenly spaced bins for evaluating the NLL.

Figure 7.1 displays the ΔNLL curves obtained from all five considered observable distributions (individually displayed in figure 5.3, figure 5.6, figure 6.15, figure 6.20, figure 6.27). Table 7.1 contains the confidence intervals and corresponding lengths extracted from these curves when only considering signal events. Table 6.18 displays the results that are obtained when also including background events.

As previously mentioned, the $\mathcal{OO}$ contains the full information on the seven dimensional phase space for small $\tilde{d}$-values and is thus expected to yield the optimal sensitivity to new, CP-violating couplings. Strictly speaking, this optimality only applies at the truth level and solely to signal events. Therefore, it is noteworthy that even after considering event selection, detector resolution and background processes, its CP sensitivity still remains unbeaten. However, within their uncertainties $\Delta\ell = 0.0007$, the lengths of the confidence intervals that are obtained from all machine learning observables are compatible with the results for the $\mathcal{OO}$ when considering only signal events. This is true at both, the 1σ and the 2σ confidence level. When including background, this observation remains true for all ML observables, except for $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\mathrm{final}}$, whose 2σ confidence interval is now wider, compared to the result obtained with the $\mathcal{OO}$. Specifically, its 2σ confidence interval is increased by 2% compared to the 2σ interval obtained with the $\mathcal{OO}$. The observable $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ is outperformed by all ML observables. Remarkably, as demonstrated by $P_{j_0}^T P_{j_1}^T \sin\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$, only a minimal increase in complexity and the inclusion of one piece of additional kinematic information per tagging jet is necessary to significantly enhance the CP-sensitivity. Specifically, in the signal only case, the additional information stored in $P_{j_0}^T P_{j_1}^T \sin\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ yields a decrease in interval length by 10% compared to $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ at both confidence levels. When including background, a decrease by 4% is observed at both confidence levels.



(a)



(b)

Figure 7.1: ΔNLL-curves obtained from the distributions of the established CP-odd observables $\mathcal{O}\mathcal{O}$ and $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ that are introduced in chapter 5 and from the distributions of the machine learning observables $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\mathrm{final}}$ constructed from the output of a multiclass NN in section 6.3, $\mathcal{O}_{\mathrm{NN}}^{\mathrm{Reg}}$ constructed using a regression NN in section 6.4 and $P_{j_0}^T P_{j_1}^T \sin\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ found through symbolic regression in section 6.5. These results are shown separately for the case where (a) only signal events are considered and (b) background events are included.

	1σ CI	1σ CI length	2σ CI	2σ CI length
$\mathcal{OO}$	[-0.0109, 0.0109]	0.0218	[-0.0216, 0.0216]	0.0432
$\Delta\Phi^{\text{sgn}}_{jj}$	[-0.0123, 0.0123]	0.0246	[-0.0243, 0.0243]	0.0486
$(\mathcal{O}^{\text{3class}}_{\text{NN}})_{\text{final}}$	[-0.0110, 0.0110]	0.0220	[-0.0217, 0.0218]	0.0435
$P^T_{\text{j}_0} P^T_{\text{j}_1} \sin\Delta\Phi^{\text{sgn}}_{jj}$	[-0.0110, 0.0111]	0.0221	[-0.0218, 0.0219]	0.0437
$\mathcal{O}^{\text{Reg}}_{\text{NN}}$	[-0.0109, 0.0109]	0.0218	[-0.0216, 0.0216]	0.0432

Table 7.1: **Signal only:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the established CP-odd observables $\mathcal{OO}$ and $\Delta\Phi^{\text{sgn}}_{\text{jj}}$ that are introduced in chapter 5 and from the distributions of the machine learning observables $(\mathcal{O}^{\text{3Class}}_{\text{NN}})_{\text{final}}$ constructed from the output of a multiclass NN in section 6.3, $\mathcal{O}^{\text{Reg}}_{\text{NN}}$ constructed using a regression NN in section 6.4 and $P^T_{j_0} P^T_{j_1} \sin\Delta\Phi^{\text{sgn}}_{\text{jj}}$ found through symbolic regression in section 6.5. The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta\ell = 0.0007$.

	1σ CI	1σ CI length	2σ CI	2σ CI length
$\mathcal{OO}$	[-0.0150, 0.0149]	0.0299	[-0.0301, 0.0300]	0.0601
$\Delta\Phi^{\text{sgn}}_{jj}$	[-0.0158, 0.0157]	0.0315	[-0.0318, 0.0316]	0.0634
$(\mathcal{O}^{\text{3class}}_{\text{NN}})_{\text{final}}$	[-0.0152, 0.0152]	0.0304	[-0.0306, 0.0306]	0.0612
$P^T_{\text{j}_0} P^T_{\text{j}_1} \sin\Delta\Phi^{\text{sgn}}_{jj}$	[-0.0150, 0.0151]	0.0301	[-0.0303, 0.0303]	0.0606
$\mathcal{O}^{\text{Reg}}_{\text{NN}}$	[-0.0149, 0.0149]	0.0298	[-0.0301, 0.0300]	0.0601

Table 7.2: **Including background:** 1σ and 2σ confidence intervals (CIs), as well as their lengths, extracted from the distributions of the established CP-odd observables $\mathcal{OO}$ and $\Delta\Phi^{\text{sgn}}_{\text{jj}}$ that are introduced in chapter 5 and from the distributions of the machine learning observables $(\mathcal{O}^{\text{3Class}}_{\text{NN}})_{\text{final}}$ constructed from the output of a multiclass NN in section 6.3, $\mathcal{O}^{\text{Reg}}_{\text{NN}}$ constructed using a regression NN in section 6.4 and $P^T_{j_0} P^T_{j_1} \sin\Delta\Phi^{\text{sgn}}_{\text{jj}}$ found through symbolic regression in section 6.5.The uncertainties on the interval borders are $\Delta b = 0.0005$, resulting in uncertainties on the interval lengths $\Delta\ell = 0.0007$.

Chapter 8

Conclusion and Outlook

This thesis presents a comparison of the sensitivities of both established CP-odd observables and newly constructed machine learning based CP-odd observables, to constrain CP-violating couplings in the HVV-vertex of the Higgs-boson production via VBF in the $H \to \tau_{\text{had}}\tau_{\text{had}}$ decay channel. For this, simulated event samples, corresponding to the data recorded during the full Run2 of the ATLAS experiment with an integrated luminosity of 139 fb^{-1} at a center of mass energy of $\sqrt{s} = 13$ TeV are used.

Firstly, an event selection process was conducted following a three staged approach, which involved applying the following sets of requirements: The preselection cut, which specifically exploits kinematic differences between the signal and background processes and considers detector limitations, the VBF cut, which allows selecting events that show kinematic and topological features that are specific for the VBF process, and finally the NN cut which chooses events based on the probability a neural network assigns to them for resulting from the signal process. The additional requirements imposed through the VBF and NN cuts allowed to enhance the signal significance by a factor of 2.43 compared to only requiring the preselection criteria. The signal to background ratio was enhanced by a factor of 69.

Next, NLL scans were employed to estimate the sensitivity of the established CP-odd observables $\mathcal{OO}$ and $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ to CP-violating contributions in the HVV vertex, by constraining the parameter $\tilde{d}$ through the construction of 1σ and 2σ confidence intervals. The expected sensitivity of the $\mathcal{OO}$, using the full Run2 dataset, is $\tilde{d}$ ∈[-0.0150(5), 0.0149(5)] at 1σ confidence level and $\tilde{d}$ ∈[-0.0301(5),0.0300(5)] at 2σ confidence level. With $\Delta\Phi_{\text{jj}}^{\text{sgn}}$, $\tilde{d}$ could be constrained to the interval [-0.0158(5), 0.0157(5)] at 1σ confidence level and to [-0.0318(5), 0.0316(5)] at 2σ confidence level. The $\mathcal{OO}$ provides an increase in CP-sensitivity by 5% at both confidence levels, compared to $\Delta\Phi_{\text{jj}}^{\text{sgn}}$. This difference in CP-sensitivity is expected, considering that $\Delta\Phi_{\text{jj}}^{\text{sgn}}$ only carries the (signed) angular information of the two outgoing tagging jets, whereas the $\mathcal{OO}$ contains the full information about the seven dimensional phase space.

Various machine learning based CP-odd observables have been constructed. When introducing additional CP-odd couplings in the EFT approach, their contribution appears in the corresponding matrix elements as a CP-odd interference term. A binary NN was employed to trained to learn kinematic differences between events that have a positive and negative interference term, respectively. Its output was then used to construct a CP-odd observable $\mathcal{O}_{\text{NN}}^{\text{2Class}}$, whose distribution allowed to constrain $\tilde{d}$ to [-0.0163(5), 0.0163(5)] at 1σ confidence level and to [-0.0328(5), 0.0328(5)] at 2σ confidence level, when considering the full Run2 dataset.

The next step was to expand this NN by an additional class, allowing the NN to also learn features that are specific for SM events. The central confidence intervals constructed from the resulting observable $\mathcal{O}_{\text{NN}}^{\text{3Class}}$ are [-0.0155(5), 0.0155(5)] at 1σ confidence level and [-0.0313(5), 0.0313(5)] at 2σ confidence level, implying an increase in CP-sensitivity by 5% compared to $\mathcal{O}_{\text{NN}}^{\text{2Class}}$, when considering the full Run2 dataset.

Studies on the impact of the specific $\tilde{d}$ value used for calculating the weights $w_{\tilde{d}}$ that are considered in the loss function during training showed no clear dependency between NN performance and the considered $\tilde{d}$-scenario. This is true for both, the binary and the multiclass case. However, for $\tilde{d} > 0.1$, the multiclass NN outperformed the binary NN, which is why the analysis proceeded with a multiclass NN, trained in the $\tilde{d} = 1$ scenario.

Furthermore, increasing the quantity and complexity of the input features was found to yield no further improvement in the NN performance compared to only considering low level kinematic features of the final state constituents. Presenting topological information on the VBF process to the NN during training, by rotating all Φ-components, such that $\Phi_{\text{Higgs}} = 0$, also yields no improvement in the NN performance.

A further study, where the multiclass NN was trained while only considering the interference information during training by removing the CP-even terms in the weights $w_{\tilde{d}}$ used in the loss function, yields an observable $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{lin}}$, that constrains $\tilde{d}$ to [-0.0156(5), 0.0156(5)] at 1σ confidence level and to [-0.0315(5), 0.0313(5)] at 2σ confidence level, compatible with the results obtained when using the full weights $w_{\tilde{d}}$. This implies, that the additional CP-even contribution, does not bias the NN.

To investigate, whether further enhancement in CP-sensitivity could be achieved, a final multiclass NN was trained in the $\tilde{d} = 1$ scenario, using the full set of input features and the full weights $w_{\tilde{d}}$ and optimized over more trials using a search space, that that was expanded compared to the search space used in the optimization of the previous NNs. The resulting CP-odd observable $(\mathcal{O}_{\text{NN}}^{\text{3Class}})_{\text{final}}$ allowed constraining $\tilde{d}$ to [-0.0152(5), 0.0152(5)] at 1σ confidence level and to [-0.0306(5), 0.0306(5)] at 2σ confidence level, providing an improvement of 2%, compared to previous results. A study on the importance of different input features on the NN performance revealed that the high level input feature $P_{j_0}^T P_{j_1}^T \sin \Delta\Phi_{\text{jj}}^{\text{sgn}}$ plays the most crucial role in the learning process.

A regression neural network was employed to resemble the distribution of the $\mathcal{OO}$, as predicted by the SM. Studies on the impact of the set of input features chosen in training revealed that only the simple kinematic information contained in the three momenta of the final state tagging jets and τ-leptons are necessary to obtain an observable $\mathcal{O}_{\text{NN}}^{\text{Reg}}$, whose CP-sensitivity is compatible with the sensitivity of the $\mathcal{OO}$. Including additional information and high level features does not yield a significant improvement. The observable $\mathcal{O}_{\mathrm{NN}}^{\mathrm{Reg}}$, resulting from a regression NN trained with only this simple kinematic information, constrains $\tilde{d}$ to [-0.0149(5),0.0150(5)] at 1σ confidence level and to [-0.0301(5),0.0301(5)] at 2σ confidence level, whereas the observable that is obtained when considering the full set of input features defined in table 6.1 yields the compatible intervals [-0.0149(5),0.0149(5)] at 1σ confidence level and [-0.0301(5),0.0300(5)] at 2σ confidence level.

Furthermore, the $\mathcal{OO}$ distribution was approximated through analytical expressions obtained via the method of symbolic regression. When targeting the $\mathcal{OO}$ distribution predicted by the SM, the expression $P_{j_0}^T P_{j_1}^T \sin \Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ was found, verifying the result obtained in [19]. The distribution that results from applying this expression allows constraining $\tilde{d}$ to [-0.0150(5), 0.0151(5)] at 1σ confidence level and to [-0.0303(5), 0.0303(5)] at 2σ confidence level. When targeting the $\mathcal{OO}$ distribution in the $\tilde{d} = 0.5$ scenario, the more complex expression $P_T^{j_0} P_T^{j_1} [\cos \Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}} + B] \sin \Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ was found. Applying this expression yields a distribution that allows to constrain $\tilde{d}$ to [-0.0149(5), 0.0150(5)] at 1σ confidence level and to [-0.0300(5), 0.0301(5)] at 2σ confidence level.

Finally, the CP-sensitivities of $\mathcal{OO}, \Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}, (\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\mathrm{final}}, \mathcal{O}_{\mathrm{NN}}^{\mathrm{Reg}}$ and $P_{j_0}^T P_{j_1}^T \sin \Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ were compared. It is worth noting that the CP-sensitivity of the $\mathcal{OO}$ has not been exceeded. As previously mentioned, this is expected, since the $\mathcal{OO}$ contains the full information on the seven dimensional phase space for small $\tilde{d}$-values. The observable $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ was outperformed by all other observables. The CP-sensitives of all machine learning based observables were found to be consistent with the sensitivity of the $\mathcal{OO}$ at 1σ confidence level,. At 2σ confidence level, $(\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}})_{\mathrm{final}}$ is outperformed by the other ML observables.

To conclude, machine learning observables were found to be a promising alternative to established observables such as $\Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ and the $\mathcal{OO}$ for future analyses concerning the CP-properties of the HVV-vertex in VBF. While they may not surpass the performance of the current best observable - the $\mathcal{OO}$, they offer the advantage of not relying on specialized tools like HAWK, thus enhancing accessibility for potential applications. ML approaches represent a powerful tool capable of learning CP-properties of the HVV vertex solely from information about the final state constituents. $\mathcal{O}_{\mathrm{NN}}^{\mathrm{Reg}}$ and $P_{j_0}^T P_{j_1}^T \sin \Delta\Phi_{\mathrm{jj}}^{\mathrm{sgn}}$ can even achieve sufficient results without the need for reconstructing the Higgs boson or calculating high-level features for training - they rely solely on the visible τ-leptons information and the kinematics of the tagging jets.

Acknowledgments

Firstly, I would like to thank Prof. Dr. Markus Schumacher for his excellent supervision. I highly appreciate all the time and effort he dedicated to supporting me throughout the past three months. The numerous meetings we had for discussing the thesis (and physics in general) were very valuable to me.

I would also like to thank Dr. Lorenzo Rossini for his great supervision and for everything he patiently taught me during the past three months. I particularly appreciate that he was always immediately there to help whenever I had any sort of question or problem, regardless of how busy he was himself.

Thanks to Oğul Öncel for his support and advice, especially towards the end of the writing process. I also appreciate Naman Kumar Bhalla's support with all kinds of technical matters and the many helpful tips he has provided me.

Finally, I would like to thank all members of the research group. I am very grateful that I was allowed to write my thesis in such a friendly and supportive environment.

Bibliography

[1] A. Salam. "Weak and electromagnetic interactions". In: *Physical Review Letters* (1959), pp. 244–254. DOI: `10.1142/9789812795915_0034`.

[2] S. L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. DOI: `https://doi.org/10.1016/0029-5582(61)90469-2`.

[3] S. Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: `10.1103/PhysRevLett.19.1264`.

[4] UA2 Collaboration. "Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN pp collider". In: *Physics Letters B* 129.1 (1983), pp. 130–140. DOI: `https://doi.org/10.1016/0370-2693(83)90744-X`.

[5] UA2 Collaboration. "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV". In: *Physics Letters B* 122.1 (1983), pp. 103–116. URL: `https://www.sciencedirect.com/science/article/pii/0370269383911772`.

[6] D0 Collaboration. "Observation of the Top Quark". In: *Physical Review Letters* 74.14 (1995). DOI: `10.1103/physrevlett.74.2632`.

[7] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012). DOI: `10.1016/j.physletb.2012.08.021`.

[8] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. DOI: `10.1016/j.physletb.2012.08.020`.

[9] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: `10.1103/PhysRevLett.13.508`.

[10] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: `10.1103/PhysRevLett.13.321`.

[11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 1964), pp. 585–587. DOI: `10.1103/PhysRevLett.13.585`.

[12] A. D. Sakharov. “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Soviet Physics Uspekhi* 34.5 (1991), p. 392. DOI: `10.1070/PU1991v034n05ABEH002497`.

[13] M. Kobayashi and T. Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: `10.1143/PTP.49.652`.

[14] “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003. DOI: `10.1088/1748-0221/3/08/S08003`.

[15] ATLAS Collaboration. “Test of CP invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector”. In: *The European Physical Journal C* 76.12 (2016). DOI: `10.1140/epjc/s10052-016-4499-5`.

[16] ATLAS Collaboration. “Test of CP invariance in vector-boson fusion production of the Higgs boson in the $H \to \tau\tau$ channel in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Physics Letters B* 805 (2020). DOI: `10.1016/j.physletb.2020.135426`.

[17] V. Hankele et al. “Anomalous Higgs boson couplings in vector boson fusion at the CERN LHC”. In: *Physical Review D* 74.9 (2006). DOI: `10.1103/physrevd.74.095001`.

[18] A. Bhardwaj et al. “Machine-enhanced CP-asymmetries in the Higgs sector”. In: *Physics Letters B* 832 (2022), p. 137246. DOI: `10.1016/j.physletb.2022.137246`.

[19] A. Butter et al. “Back to the formula - LHC edition”. In: *SciPost Physics* 16.1 (2024). DOI: `10.21468/scipostphys.16.1.037`.

[20] L. Evans and P. Bryant. “LHC Machine”. In: 3.08 (Aug. 2008), S08001. DOI: `10.1088/1748-0221/3/08/S08001`.

[21] ATLAS Collaboration. “Combined Measurement of the Higgs Boson Mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ Decay Channels with the ATLAS Detector Using $\sqrt{s} = 7$, 8, and 13 TeV *pp* Collision Data”. In: *Phys. Rev. Lett.* 131 (25 2023), p. 251802. DOI: `10.1103/PhysRevLett.131.251802`.

[22] A. Lösle. “Measurement of the Higgs-Boson Production Cross-Section and Test of CP Invariance in Vector-Boson Fusion Production with $H \to \tau^+\tau^-$ Decays in the Fully-Leptonic Final State with the ATLAS Detector at the LHC”. In: *PhD thesis* (2020). URL: `https://freidok.uni-freiburg.de/data/167201`.

[23] LHC Higgs Cross Section Working Group. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. In: 2/2017 (Oct. 2016). DOI: `10.23731/CYRM-2017-002`. arXiv: `1610.07922 [hep-ph]`.

[24] Particle Data Group. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 2022), p. 030001. DOI: `10.1103/PhysRevD.98.030001`.

[25] A. Elagin et al. "A new mass reconstruction technique for resonances decaying to". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (2010). DOI: `10.1016/j.nima.2011.07.009`.

[26] Georges Aad et al. "ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset". In: *Eur. Phys. J. C* 83.7 (2023), p. 681. DOI: `10.1140/epjc/s10052-023-11699-1`. arXiv: `2211.16345 [physics.data-an]`.

[27] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: `10.1007/JHEP06(2010)043`. arXiv: `1002.2581 [hep-ph]`.

[28] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: `10.1016/j.cpc.2015.01.024`. arXiv: `1410.3012 [hep-ph]`.

[29] T. Gleisberg et al. "Event generation with SHERPA 1.1". In: *JHEP* 02 (2009), p. 007. DOI: `10.1088/1126-6708/2009/02/007`. arXiv: `0811.4622 [hep-ph]`.

[30] ATLAS Collaboration. "Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 08 (2022), p. 175. DOI: `10.1007/JHEP08(2022)175`. arXiv: `2201.08269 [hep-ex]`.

[31] ATLAS Collaboration. "The ATLAS Simulation Infrastructure". In: *The European Physical Journal C* 70.3 (2010). DOI: `10.1140/epjc/s10052-010-1429-9`.

[32] ATLAS Collaboration. "Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC". In: (2019). URL: `https://cds.cern.ch/record/2677054`.

[33] J. Seixas. "Self-organized mapping of calorimetry information for high efficient online electron/jet identification in ATLAS". In: (2009), p. 055. DOI: `10.22323/1.050.0055`.

[34] ATLAS Collaboration. "Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV". In: (2017). URL: `https://cds.cern.ch/record/2261772`.

[35] R.K. Ellis et al. "Higgs decay to + −A possible signature of intermediate mass Higgs bosons at high energy hadron colliders". In: *Nuclear Physics B* 297.2 (1988), pp. 221–243. DOI: `https://doi.org/10.1016/0550-3213(88)90019-3`.

[36] Planck Collaboration. "Planck 2015 results - XIII. Cosmological parameters". In: *A&A* 594 (2016), A13. DOI: `10.1051/0004-6361/201525830`.

[37] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105 (4 1957), pp. 1413–1415. DOI: `10.1103/PhysRev.105.1413`.

[38] W. Demtröder. *Experimentalphysik 4*. ISBN: 978-3-642-21476-9. Springer Spektrum Berlin, Heidelberg, 2013.

[39] M. Goldhaber, L. Grodzins, and A. W. Sunyar. "Helicity of Neutrinos". In: *Phys. Rev.* 109 (3 1958), pp. 1015–1017. DOI: `10.1103/PhysRev.109.1015`.

[40] J. H. Christenson et al. "Evidence for the 2π Decay of the K_2^0 Meson". In: *Phys. Rev. Lett.* 13 (4 1964), pp. 138–140. DOI: `10.1103/PhysRevLett.13.138`.

[41] M. Gronau. *CP Violation in the B Meson System.* 1997. arXiv: `hep-ph/9705262 [hep-ph]`.

[42] R. Aaij et al. "Observation of CP Violation in Charm Decays". In: *Phys. Rev. Lett.* 122.21 (2019), p. 211803. DOI: `10.1103/PhysRevLett.122.211803`. arXiv: `1903.08726 [hep-ex]`.

[43] P. Huet. *Electroweak Baryogenesis and the Standard Model.* 1994. arXiv: `hep-ph/9406301 [hep-ph]`.

[44] ATLAS Collaboration. "Test of CP invariance in vector-boson fusion production of the Higgs boson in the $H \to \tau\tau$ channel in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Physics Letters B* 805 (2020). DOI: `10.1016/j.physletb.2020.135426`.

[45] S. Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967). DOI: `10.1103/PhysRevLett.19.1264`.

[46] G. Cowan. *Statistical Data Analysis.* ISBN: 978-0-19-850156-5. Clarendon, Oxford: Oxford Science Publications, 1998.

[47] M. Ciccolini, A. Denner, and S. Dittmaier. "Strong and electroweak corrections to the production of Higgs+2jets via weak interactions at the LHC". In: *Physical Review Letters* 99.16 (2007). DOI: `10.1103/physrevlett.99.161803`.

[48] M. Ciccolini, Denner A., and Dittmaier S. "Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC". In: *Physical Review D* 77.1 (2008). DOI: `10.1103/physrevd.77.013002`.

[49] ATLAS Colaboration. "Test of CP invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector". In: *The European Physical Journal C* 76.12 (2016), p. 658. DOI: `10.1140/epjc/s10052-016-4499-5`.

[50] D. Atwood and A. Soni. "Analysis for magnetic moment and electric dipole moment form factors of the top quark via e^+e^-→tt$^-$". In: *Phys. Rev. D* 45 (7 1992), pp. 2405–2413. DOI: `10.1103/PhysRevD.45.2405`.

[51] M. Diehl and O. Nachtmann. "Optimal observables for the measurement of three gauge boson couplings ine+e−→W+W−". In: *Zeitschrift für Physik C Particles and Fields* 62.3 (1994). DOI: `10.1007/BF01555899`. URL: `https://doi.org/10.1007/BF01555899`.

[52] ATLAS Collaboration. *Test of CP-invariance of the Higgs boson in vector-boson fusion production and its decay into four leptons.* 2023. arXiv: `2304.09612 [hep-ex]`.

[53] ATLAS Colaboration. "Test of CP Invariance in Higgs Boson Vector-Boson-Fusion Production Using the $H \to \gamma\gamma$ Channel with the ATLAS Detector". In: *Phys. Rev. Lett.* 131 (6 2023), p. 061802. DOI: `10.1103/PhysRevLett.131.061802`.

[54] P. Mehta et al. "A high-bias, low-variance introduction to Machine Learning for physicists". In: *Physics Reports* 810 (2019). DOI: `10.1016/j.physrep.2019.03.001`.

[55] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization.* 2017. arXiv: `1412.6980 [cs.LG]`.

[56] Takuya A. et al. *Optuna: A Next-generation Hyperparameter Optimization Framework.* 2019. arXiv: `1907.10902 [cs.LG]`.

[57] M. Cranmer. *PySr: Fast & parallelized symbolic regression in python/julia.* URL: `https://zenodo.org/records/4041459`. (accessed: 21.01.2023).

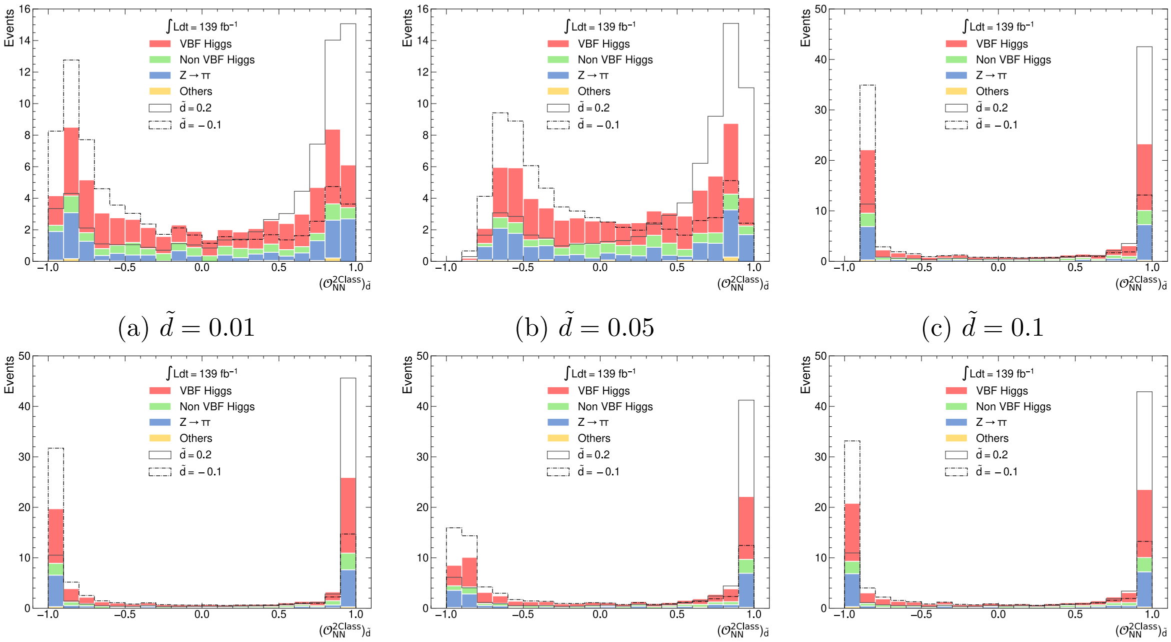
[58] J. R. Koza. "Genetic programming as a means for programming computers by natural selection". In: *Statistics and computing* 4 (1994), pp. 87–112.

[59] D. Brkić et al. "Symbolic Regression Approaches for the Direct Calculation of Pipe Diameter". In: *Axioms* 12.9 (2023). DOI: `10.3390/axioms12090850`.

[60] S. Kirkpatrick, Jr. C. D. Gelatt, and M. P. Vecchi. "Optimization by Simulated Annealing". In: *Spin Glass Theory and Beyond.* 1983, pp. 339–348. DOI: `10.1142/9789812799371_0035`.

Appendix A

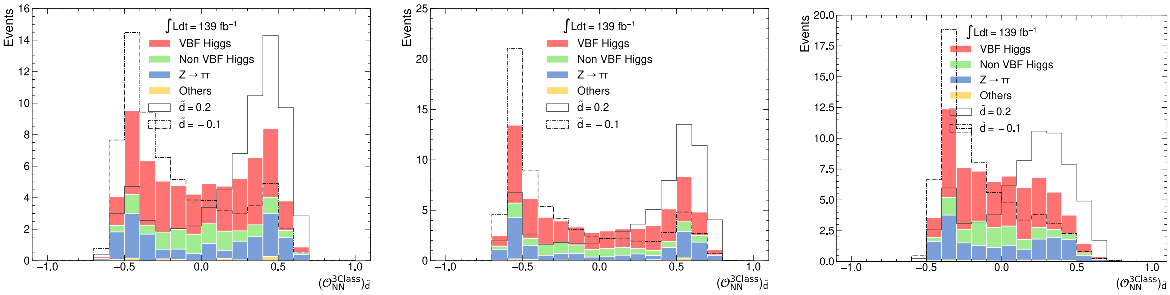
Figures



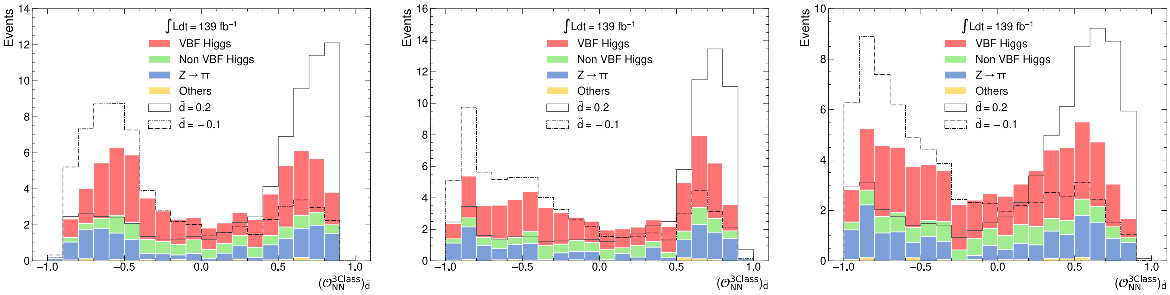
(a) $\tilde{d} = 0.01$ (b) $\tilde{d} = 0.05$ (c) $\tilde{d} = 0.1$

(d) $\tilde{d} = 0.25$ (e) $\tilde{d} = 0.5$ (f) $\tilde{d} = 1$

Figure A.1: Distribution of of the CP-sensitive observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{2Class}}$ constructed from the prediction on the full dataset by binary neural networks that were trained and optimized while assuming different $\tilde{d}$-scenarios, as predicted by the SM for an integrated luminosity of $\int \mathrm{Ldt} = 139\ \mathrm{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. Additionally, distributions predicted for two CP-violating scenarios with the strengths $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$ are presented. These distributions are normalized to the total number $N_{\mathrm{SM}} = 69$ of expected SM events.

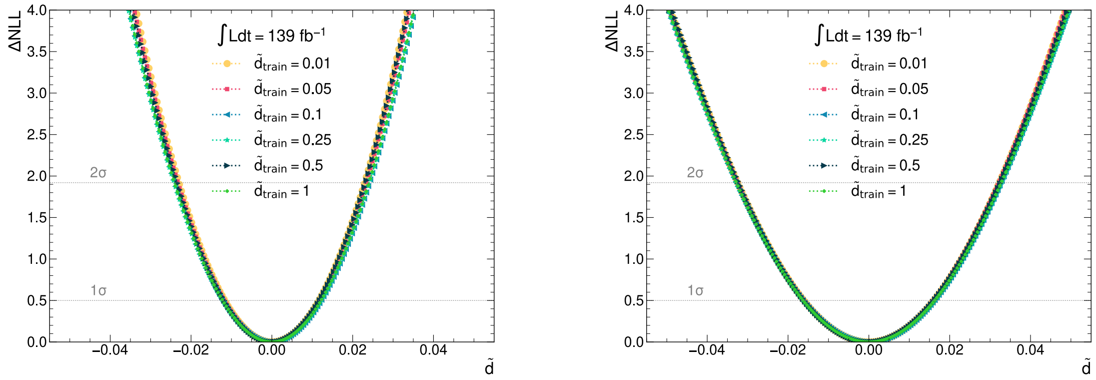


(a) $\tilde{d} = 0.01$ (b) $\tilde{d} = 0.05$ (c) $\tilde{d} = 0.1$



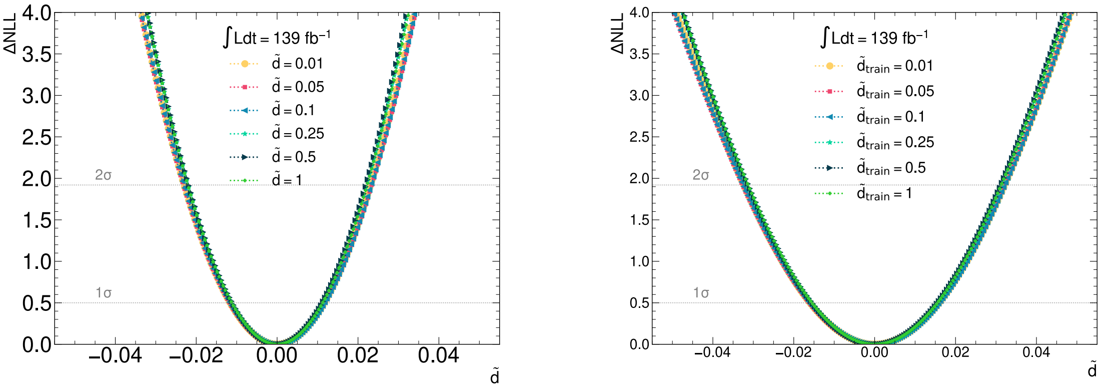
(d) $\tilde{d} = 0.25$ (e) $\tilde{d} = 0.5$ (f) $\tilde{d} = 1$

Figure A.2: Distributions the CP-sensitive observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ constructed from the prediction on the full dataset by multiclass neural networks that were trained and optimized while assuming different $\tilde{d}$-scenarios, as predicted by the SM for an integrated luminosity of $\int \mathrm{Ldt} = 139\ \mathrm{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. Additionally, distributions predicted for two CP-violating scenarios with the strengths $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$ are presented. These distributions are normalized to the total number $N_{\mathrm{SM}} = 69$ of expected SM events.



(a) $(\mathcal{O}^{\text{2Class}}_{\text{NN}})_{\tilde{d}}$, signal only

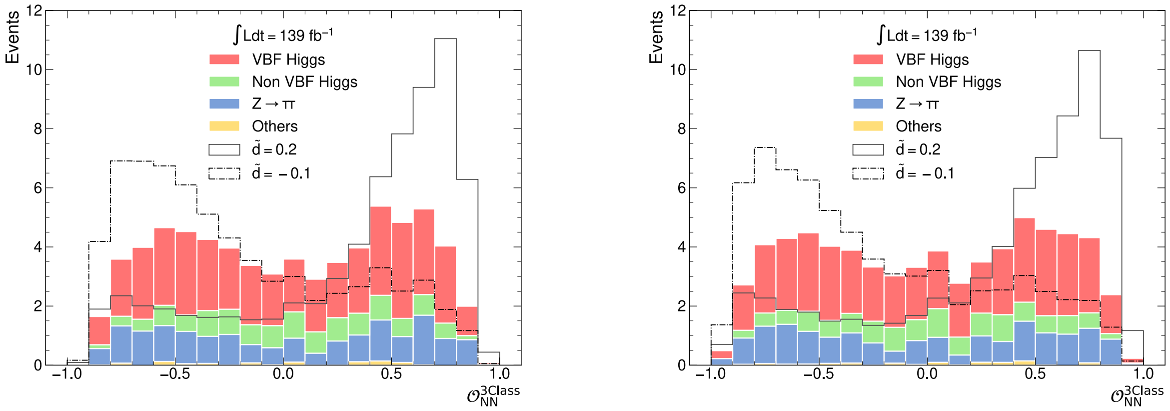
(b) $(\mathcal{O}^{\text{2Class}}_{\text{NN}})_{\tilde{d}}$, including background



(c) $(\mathcal{O}^{\text{3Class}}_{\text{NN}})_{\tilde{d}}$, signal only

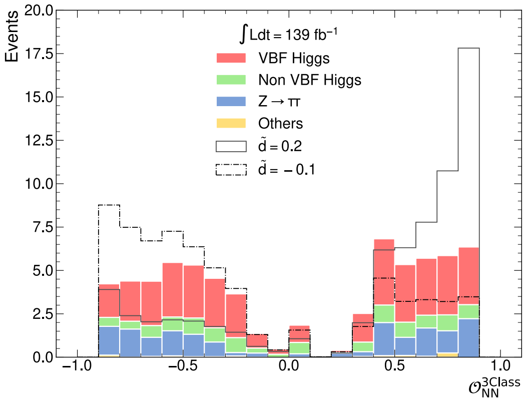
(d) $(\mathcal{O}^{\text{3Class}}_{\text{NN}})_{\tilde{d}}$, including background

Figure A.3: ΔNLL curves, obtained from the CP-sensitive observables constructed from the prediction on the full dataset by neural networks that were trained and optimized while assuming different $\tilde{d}$-scenarios. Results are shown for both the binary NN observables $(\mathcal{O}^{\text{2Class}}_{\text{NN}})_{\tilde{d}}$ and the multiclass NN observables $(\mathcal{O}^{\text{3Class}}_{\text{NN}})_{\tilde{d}}$, separately with and without considering background events.



(a) Input:$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}\}$

(b) Input:$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}, \vec{P}_{\mathrm{Higgs}}\}$



(c) Input:$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau 0,1}, \vec{P}_{\mathrm{Higgs}}\}_{\Phi_{\mathrm{Higgs}} \equiv 0}$

Figure A.4: Distributions of the CP-sensitive observables $\mathcal{O}_{\mathrm{NN}}^{\mathrm{3Class}}$ constructed from the prediction on the full dataset by multiclass neural networks that were trained and optimized using different subsets of the initial input dataset defined in section 6.2, as predicted by the SM for an integrated luminosity of $\int \mathrm{Ldt} = 139\ \mathrm{fb}^{-1}$. The VBF Higgs-boson signal is presented in red. Background processes are grouped in processes where a Higgs boson emerges from non VBF production modes (green) and processes where the detected τ-leptons emerge from the decay of a Z-boson (blue). The category *others* (yellow) contains all remaining background contributions, described in section 2.2. Additionally, distributions predicted for two CP-violating scenarios with the strengths $\tilde{d} = -0.1$ and $\tilde{d} = 0.2$ are presented. These distributions are normalized to the total number $N_{\mathrm{SM}} = 69$ of expected SM events.. $\vec{P}$ denote the three momenta of the respective final state constituents that are used as NN inputs.

Appendix B

Tables

Optimized hyperparameter	$\tilde{d} = 0.01$	$\tilde{d} = 0.05$	$\tilde{d} = 0.1$	$\tilde{d} = 0.25$	$\tilde{d} = 0.5$	$\tilde{d} = 1$
Number of hidden layers	2	1	5	2	5	4
Number of nodes in hidden layer 1	100	350	25	300	150	350
Number of nodes in hidden layer 2	150	-	300	200	200	400
Number of nodes in hidden layer 3	-	-	90	-	350	5
Number of nodes in hidden layer 4	-	-	5	-	150	150
Number of nodes in hidden layer 5	-	-	400	-	200	-
L2 regularization strength	$1.6{\cdot}10^{-3}$	$1.2{\cdot}10^{-5}$	$1.6{\cdot}10^{-4}$	$3.7{\cdot}10^{-5}$	$6.1{\cdot}10^{-4}$	$2.9{\cdot}10^{-3}$
Initial learning rate	$1.0{\cdot}10^{-3}$	$1.6{\cdot}10^{-5}$	$2.0{\cdot}10^{-4}$	$4.6{\cdot}10^{-5}$	$2.8{\cdot}10^{-5}$	$1.8{\cdot}10^{-3}$
Learning rate decay steps	7371.4	5997.6	3157.3	9940.9	7727.3	634.3
Learning rate decay rate	$1.9{\cdot}10^{-2}$	$3.0{\cdot}10^{-3}$	$9.2{\cdot}10^{-2}$	0.1	0.6	0.2
Batch size	128	96	64	128	64	128

Table B.1: Architecture and hyperparameters of the binary neural networks trained and optimized while assuming different $\tilde{d}$-scenarios. These results are the outcome of hyperparameter optimization studies performed using optimization framework `OPTUNA` [56].

Optimized hyperparameter	$\tilde{d} = 0.01$	$\tilde{d} = 0.05$	$\tilde{d} = 0.1$	$\tilde{d} = 0.25$	$\tilde{d} = 0.5$	$\tilde{d} = 1$
Number of hidden layers	1	2	1	2	2	2
Number of node in hidden layer 1	200	150	100	90	25	350
Number of nodes in hidden layer 2	-	150	-	100	5	400
L2 regularization strength	$1.9{\cdot}10^{-3}$	$2.5{\cdot}10^{-4}$	$6.6{\cdot}10^{-3}$	$1.7{\cdot}10^{-4}$	$1.7{\cdot}10^{-5}$	$1.7{\cdot}10^{-5}$
Initial learning rate	$5.5{\cdot}10^{-5}$	$1.1{\cdot}10^{-4}$	$1.3{\cdot}10^{-3}$	$9.6{\cdot}10^{-4}$	$1.9{\cdot}10^{-3}$	$8.2{\cdot}10^{-4}$
Learning rate decay steps	3860.3	9584.2	7351.6	4111.6	4804.5	7849.9
Learning rate decay rate	0.2	0.1	0.6	0.7	0.7	0.3
Batch size	64	128	96	32	64	128

Table B.2: Architecture and hyperparameters of the multiclass neural networks trained and optimized while assuming different $\tilde{d}$-scenarios. These results are the outcome of hyperparameter optimization studies performed using the optimization framework `OPTUNA` [56].

Optimized hyperparameter	$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau_{0,1}}\}$	$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau_{0,1}}, \vec{P}_{\mathrm{Higgs}}\}$	$\{\vec{P}_{\mathrm{j}0,1}, \vec{P}_{\tau_{0,1}}, \vec{P}_{\mathrm{Higgs}}\}_{\Phi_{\mathrm{Higgs}}\equiv 0}$
Number of hidden layers	2	2	4
Number of nodes in hidden layer 1	400	300	150
Number of nodes in hidden layer 2	400	300	250
Number of nodes in hidden layer 3	-	-	90
Number of nodes in hidden layer 4	-	-	150
L2 regularization strength	$2.1{\cdot}10^{-5}$	$1.2{\cdot}10^{-5}$	$7.4{\cdot}10^{-7}$
Initial learning rate	$3.5{\cdot}10^{-4}$	$1.7{\cdot}10^{-4}$	$9.0{\cdot}10^{-5}$
Learning rate decay steps	1669.1	9096.8	10000
Learning rate decay rate	0.2	$3.6{\cdot}10^{-2}$	0.6
Batch size	128	32	64

Table B.3: Architecture and hyperparameters of the multiclass neural networks trained and optimized with different subsets of input features. These results are the outcome of hyperparameter optimization studies performed using the optimization framework `OPTUNA` [56].

Optimized hyperparameter	Study outcome
Number of hidden layers	3
Number of nodes in hidden layer 1	400
Number of nodes in hidden layer 2	250
Number of nodes in hidden layer 3	300
L2 regularization strength	$1.0{\cdot}10^{-5}$
Initial learning rate	$4.6{\cdot}10^{-3}$
Learning rate decay steps	5833.4
Learning rate decay rate	$2.4{\cdot}10^{-3}$
Batch size	96

Table B.4: Architecture and hyperparameters of the multiclass neural network that was trained and optimized using only the term in the event weights in eq. (4.19) that is linear in $\tilde{d}$. These results are the outcome of a hyperparameter optimization study performed using the optimization framework `OPTUNA` [56].