Probing CP Violation in the Higgs sector at the Future Circular Collider

MPhys Project Report (June 2023)

by Matylda Hoffman 10453302
Project in collaboration with Lily Dickinson 10308668
Department of Physics and Astronomy, University of Manchester

May 15, 2023

Abstract

The mysterious matter-antimatter asymmetry in the universe is one of the most famous problems of modern-day particle physics, inspiring novel extensions for the Standard Model in hopes of a formalised explanation. The following project investigates the proposed Standard Model Effective Field Theory framework, and its implications, on the potential CP violation arising from the Higgs sector. This report aims to quantitatively constrain the strength of the interference of SMEFT CP-odd operators with the Standard Model through the simulation of electron-positron collisions at FCC-ee energies. These are analysed through the construction of CP-odd observables, firstly manually and later using a machine learning algorithm. The two are then compared revealing the latter to be more sensitive to new physics. The limiting values on the CP-violating operators' Wilson coefficients are found to be $[-0.14, +0.14]$, $[-0.023, +0.023]$, and $[-0.21, +0.21]$ for the $\mathcal{O}_{\mathcal{H}\widetilde{B}}$, $\mathcal{O}_{\mathcal{H}\widetilde{W}B}$, and $\mathcal{O}_{\mathcal{H}\widetilde{W}}$ operators respectively. Eventually, a comparison to FCC-hh simulated data is drawn where is it revealed that its high centre-of-mass energies and luminosity make it a more suitable candidate for CPV investigations in the Higgs sector.

Contents

1 Introduction

The Standard Model (SM) of particle physics refers to the well-established theoretical framework describing the fundamental particles making up the universe, and their interactions. Whilst it provides a coherent explanation for many ideas of modern particle physics, it is not a perfect theory as it fails to account for some evident phenomena. For instance, gravity, matter-antimatter asymmetry, and the hierarchy problem cannot be explained within the simplest version of this model, pointing to its incompleteness [1]. Consequently, extensions are posed which would unify it with other established theories, such as that of general relativity, which describes gravity. This paper will focus on the observed cosmological excess of matter over antimatter which deviates from the theoretical predictions of the SM. It will analyse events in the weakly interacting Higgs sector emergent from leptonic and hadronic collisions simulated at Future Circular Collider (FCC) centre-of-mass energies, aiming to estimate limits on the strength of the coupling of a new model to the SM. The partner report [2] will, in turn, investigate proton-proton interactions under High-Luminosity Large Hadron Collider (HL-LHC) and FCC-hh conditions. Where FCC-hh refers to the operational stage of the FCC which uses two proton beams.

The Sakharov conditions summarise the three requirements for this emergent matter-antimatter imbalance including CP symmetry violation (CPV), baryon number violation, and departure from thermal equilibrium [3]. However, known SM explanations are still not sufficient to justify the amount of observed baryon asymmetry by several orders of magnitude, motivating various Beyond Standard Model (BSM) theories. This project will explore possible CPV arising from the Higgs sector, assuming the Effective Field Theory extension of the Standard Model (SMEFT) and by looking at the Higgs interactions with weak gauge bosons. This new framework extends the SM Lagrangian by including dimension six operators

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda^2} \widetilde{\mathcal{O}}_i \tag{1}$$

with Λ representing the scale of new physics, $\mathcal{L}$ the Lagrangian, and $\widetilde{\mathcal{O}}_i$ denoting the operator. This sets all particle masses to be less than its value. $\frac{c_i}{\Lambda^2}$ are complex Wilson coefficients that signify the strength of interactions between SM particles and those arising from the additional operators in the Lagrangian. Their values depend on the energy at which interactions are probed. Considering the new Lagrangian the matrix element is updated to include an extra term, resulting in a binomial expansion when computing amplitude squared. This describes the production and decay of the Higgs boson including the relevant operators

$$\begin{aligned} |\mathcal{M}|^2 &= |\mathcal{M}_{SM} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{M}_i|^2 \\ &= |\mathcal{M}|^2 + \frac{c_i}{\Lambda^2} 2\mathcal{R}[\mathcal{M}_{SM}\mathcal{M}^*_{d6,i}] + \frac{c_i c_j}{\Lambda^4} \mathcal{M}_{d6,i}\mathcal{M}^*_{d6,j} \end{aligned}$$

where the first term is purely SM, the second term is the interference between the two leading amplitudes, and the last squared term is purely BSM suppressed by a factor of $1/\Lambda^4$ and hence are not included in the simulations. The interference term will be generated with each SMEFT CP-odd operator turned on individually to investigate their impact on the model separately. Those which are of interest describe electroweak interactions of the Higgs sector and are

$$\begin{aligned} \mathcal{O}_{\mathcal{H}\widetilde{B}} &= \mathcal{H}^\dagger \mathcal{H} B^{\mu\nu} \widetilde{B}_{\mu\nu} \\ \mathcal{O}_{\mathcal{H}\tilde{W}} &= \mathcal{H}^\dagger \mathcal{H} W^{i\mu\nu} \widetilde{W}^i_{\mu\nu} \\ \mathcal{O}_{\mathcal{H}\tilde{W}B} &= \mathcal{H}^\dagger \sigma^i \widetilde{B}^{i\mu\nu} B_{\mu\nu} \end{aligned}$$

where $\mathcal{H}$ is the Higgs field and W^μ and B^μ describe the weak and electromagnetic fields respectively derived from the tensor product of SU(2) and U(1) gauge-fields [4]. $\widetilde{X}^{\mu\nu}$ are dual field strength tensors whose complex phases are potential sources of CPV in Higgs interactions with fermions. The time component of the U(1) dual gauge field gives the electric field, while the magnetic field is related to its spatial components. In the case of the SU(2) gauge field, the dual field tensor is related to the weak force and is responsible for the breaking of the symmetry between the W and Z bosons. CPV occurs when these coefficients are non-vanishing [5].

1.1 Report Structure

The following report will first describe the concept of CP violation and CP-odd operators and observables. It will also provide an overview of the SMEFT framework and how it may lead to CP violation. Later, it will briefly describe the FCC and the software used to simulate collisions and process data. The two following sections will describe different methods of constructing CP-odd observables, one by making an equation from known angular variables and another using a machine learning algorithm. Their sensitivities to new physics are tested using likelihood tests in Section 8, and are later compared to one another. Section 9 provides a critical analysis of the data and the investigation and Section 10 concludes and comments on the sensibility of the results.

2 Theory

2.1 Lepton induced HZ Higgsstrahlung investigation channel

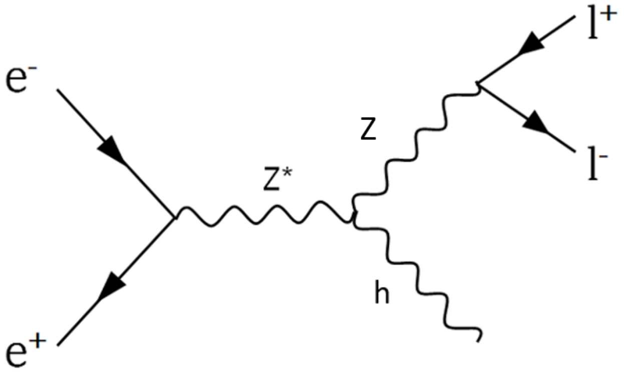

Figure 1: HZ Higgsstrahlung Feynman diagram for an electron-positron collision. The symbols Z, h, e, l represent the Z boson, the Higgs boson, the electron/positron, and a charged lepton respectively; and Z* indicates that the Z boson is virtual. Each vertex conceals additional particles and processes introduced in the SMEFT framework. The two leptons in the final state and their kinematic properties are investigated in further analysis. MadGraph generation command: $e + e- > l + l - h$.

The Higgs boson is the only scalar boson of the standard model, with spin-0 and invariant mass of 125GeV [6]. Its properties and interactions can be investigated in lepton colliders through stimulated electron-positron collisions in which the two oppositely-charged lepton annihilate to produce a virtual Z boson as shown in Figure . This particle then interacts with the Higgs field producing a real Higgs boson and a Z that recoil off each other. The Z may then undergo a dileptonic decay and the Higgs may be studied via their recoil kinematics. This process is especially useful as it allows for precision measurements of the coupling strengths between the Higgs and other particles whilst being relatively model-independent.

2.2 CP Violation

CPT symmetry is a set of three operations under the combination of which the laws of physics remain invariant. Charge conjugation (C) changes the charge of particles so that they are transformed into their corresponding antiparticles, parity transformation (P) causes a mirror reflection of all three spatial coordinates, and time reversal (T) reverses the time [7]. These can be represented in the form of eigenequations

$$\begin{aligned} C|\vec{p}, s, q\rangle &= \zeta_C |\vec{p}, s, -q\rangle \\ P|\vec{p}, s, q\rangle &= \zeta_P | -\vec{p}, s, q\rangle \\ T|\vec{p}, s, q\rangle &= \zeta_T | -\vec{p}, -s, q\rangle \end{aligned}$$

where C, P, T are operators for the three symmetries with respective eigenvalues, $\zeta_{C,P,T}$, and $|\vec{p}, s, q\rangle$ describes a particle state with defined momentum (p), spin (s), and charge (q). This invariance, however, does not always hold for the combination of C and P operators and suggests the laws of physics to not act analogously on particles and their antiparticles [8]. This phenomenon has been first observed in long-lived neutral Kaon decays that produced two charged pions ($K_L^0 \to \pi^+\pi^-$), which is a channel forbidden under CP symmetry [9].

A small contribution of the overall CPV arises from the CKM matrix, a unitary matrix that describes the mixing of the three generations of quarks in the weak interaction

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1-\frac{\lambda^2}{2} & \lambda & \lambda^3 A(\rho - i\eta) \\ -\lambda & 1-\frac{\lambda^2}{2} & \lambda^2 A \\ \lambda^3 A(1-\rho-i\eta) & -\lambda^2 A & 1 \end{pmatrix} + \mathcal{O}(\lambda^4)$$

Where $\lambda = L\sin\theta \approx 0.23$ with θ corresponding to the Cabbibo angle, $\eta \approx 0.4$, and $A \approx 0.8$. The off-diagonal, complex phase contributing components are what leads to interference between different decay pathways, causing differences in the rates of particle and antiparticle decays, resulting in CP violation [10].

2.3 CP operators

CP operators can be either odd or even. Even operators, as well as CP-even variables, do not show any asymmetries between their positive and negative values under this compound operation. CP-odd operators, on the other hand, are mathematical functions that change signs as a result of a charge and parity reflection which is then manifested in the data as a negatively weighted event. Consequently, this investigation uses exclusively CP-odd operators.

2.4 The Standard Model Effective Field Theory (SMEFT)

The SM can be interpreted as a low-energy approximation of some unknown fundamental theory. Effective Field Theory (EFT) hypothesises that at vertices of Feynman diagrams, there occur more intricate interactions involving more particles. The energy at which we induce the extra particles is much lower than their masses or interaction thresholds, so we will not observe them at current operational energies. Therefore, new technologies like the FCC will be incredibly useful in testing BSM theories.

This project will investigate CPV by making alternations to the SM Lagrangian by including sixth dimension operators ($\widetilde{\mathcal{O}}_i$). Higher-dimensional space Lagrangian can possess different symmetries to what we observe in our four-dimensional world. Thus, the inclusion of those operators can lead to additional symmetry violations that can manifest as CPV. The operators are built from SM fields and their derivatives, and the new theory obeys the same symmetries as the SM (Gauge, Lorentz, etc.). Dimension five, seven operators, etc. are not included as they violate lepton and baryon number conservation laws which we assume to hold [5]. Dimension eight operators are suppressed by a factor of $1/\Lambda^4$ which makes it negligibly small and in the context of this project is therefore omitted. The $1/\Lambda^2$ suppression of dimension-six is not sufficient for the effects to be dismissed at FCC's operating energies.

3 The FCC

	FCC-ee (ZH)	**FCC-hh**
$\sqrt{s}$ (GeV)	240	$100,000$
$\int \mathcal{L}$ (ab^{-1})	5	30
Number of expected Higgs events	10^6	10^{10}

Table 1: FCC-ee and FCC-hh collider parameters. $\sqrt{s}$ is the centre-of-mass energy, and $\int \mathcal{L}$ is the integrated luminosity over the course of a collider's operation.

With the LHC nearing its last operational stage, the HL-LHC, a suitable successor needs to be chosen which would allow for investigations at much larger energy scales. A potential candidate is the FCC, which would be situated a 100km ring and generate centre-of-mass energies of 250-350GeV for its leptonic phase of operation and up to 1000GeV for the hadronic. These will make the FCC a Higgs factory, with each type of collider producing 10^6 and 10^{10} Higgs events respectively over the course of their running. The FCC-ee, the electron-positron collider that this project focuses on, will be aimed at precision studies, whereas the incredibly high energies generated at the FCC-hh will allow for exploration of the limitations of the SM, with a special focus on the Higgs sector which can potentially provide explanations for some of the infamous problems of the SM. Moreover, the contribution of dimension-six operators scales as $(E/\Lambda)^2$, consequently growing with the centre-of-mass energy making the FCC the perfect facility for conducting SMEFT investigations due to its great centre-of-mass energies. This paper will use data simulated in MadGraph 5 with FCC parameters inputted in the run card to investigate the extent of its potential usefulness in investigations looking for CPV in the Higgs sector.

4 Methodology

4.1 Introduction

4.2 Events Generation

Events for analysis were generated using MadGraph 5 [11], with Pythia8 [12] imported for incorporation of partonic showers and hadronization. Then the data were first processed using a specialised package called Rivet which allows you to perform analysis at the generator level. Special consideration needs to be paid to commands used inside the terminal to ensure that the desired information from the HZZ vertex is carried over to the final state. When $e^+e^- \to Zh$ is used, as in last semester's report [13], and then two leptons are selected inside Rivet, the decay of the Z is done by Pythia8 which does not preserve the correlation between vertices. Consequently, the forward lepton no longer carries the information related to the Higgs vertex; resulting in the data not manifesting the CP asymmetry. Subsequently, $e^+e^- \to l^+l^-h$ is a better command to use as it already ensures that the final state is directly related to the Higgs vertex, and it is the Z boson that is implied through branching ratios. Whereas in my previous report, the emphasis was on maximising the number of generated signal events instead which is why a different command was used.

To produce interference terms the SMEFTsim package needs to be imported into the MadGraph environment [14]. This contains a complete set of dimension-six operators allowed by the symmetries of the SMEFT Lagrangian, along with the corresponding Wilson coefficients that can be edited to parameterize the strength of the new physics interactions. Then the CPV massless extension may be imported for only massive particles to be considered which significantly shortens the computational time, especially for FCC-hh simulations which take a while due to the multitude of possible interactions. The interference only data is generated by adding $NP^2 == 1$ at the end of the command line. Events are then selected in Rivet which requires them to have two electrons or muons in the final state and applies loose kinematic cuts shown in Table 2.

Variable	**Cut Applied**
m_{ll}[GeV]	$70 - 110$
ΔR	0.1
η_l	2.5

Table 2: Cuts applied to the kinematic variables during MadGraph 5 generation and Rivet analysis. The invariant mass of two leptons, m_{ll}, has been set to a 40GeV range around the Z resonance peak. The maximum angular separation between leptons in the final state, ΔR, has been set to 0.1 to ensure a full kinematic reconstruction by 'dressing' them in photons within that radius. Finally, the pseudorapidity of the charged leptons, η_l, is used to restrict the signal range as it is expected to be concentrated with respect to the beam axis.

The MadGraph algorithm uses the Monte Carlo pseudo-random method to generate events and assigns each of them a weight. When calculating interference effects of the SMEFT and SM frameworks these weights can be negative as a consequence of BSM terms interfering destructively with the SM contributions. This then leads to cancellations and results in a reduced integrated cross-section, within statistical fluctuations from zero.

4.3 Looking for CPV

CPV can be identified via asymmetric spreads of data when looking at CP-odd observables which are sensitive to the interference between the CP-conserving and CP-violating amplitudes of a process. Destructive interference contributions are indicated by negatively-weighted events from the MadGraph Monte Carlo events generation, hence this investigation aims to find variables that will successfully separate the data into positively and negatively weighted when looking at the dileptonic Z decay channel. Firstly, this is done by visual inspection of histogram plots of different kinematic variables after a CP-odd transformation and inspecting them visually for asymmetric spreads. Later, a new variable is constructed with the help of a machine learning algorithm which can display greater sensitivity to the SMEFT CP-odd operators' interference.

5 Primary Analysis

5.1 CP sensitive observables: Delta Phi

In this part of the analysis, an angular variable is constructed so that it effectively separates events into positively- and negatively-weighted ones. This is assessed by visual inspection of resultant histograms in search of an anti-symmetric spread. Angular variables are chosen as they are sensitive to the CP-even and CP-odd components of the decay amplitude which can cause an asymmetry in the distribution of the decay products. The $\Delta\phi$ compound variable is the difference between the azimuthal angles, ϕ, of the two leptons in the final state formulated as

$$\Delta\phi = \phi_1 - \phi_2 \tag{2}$$

where ϕ_1 corresponds to the leptons with higher relative rapidity.

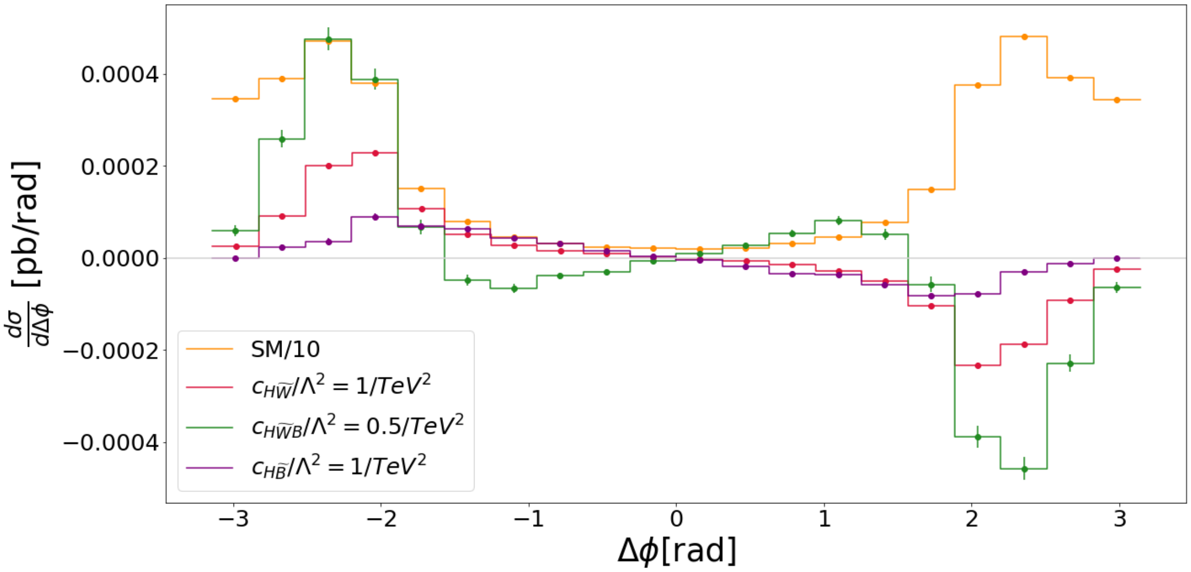

Figure 2: Plot contrasting $\Delta\phi$ for SM events against $d\sigma/d\Delta\phi$ data simulated with SMEFT CP-odd operators. Each BSM plot was generated with only one dimension-six operator at a time. SM and $\mathcal{O}_{\mathcal{H}\tilde{W}B}$ data have been scaled down by factors of 10 and 2 respectively for comparison. All SMEFT distributions are anti-symmetric, including errors, illustrating the constructive and destructive interference ranges of each operator with the SM.

The interference of CP-odd SMEFT operators is shown in Figure 2 where the negatively weighted events represent destructive interference and vice versa. The histogram values have been found by calculating the cross section of each bin and dividing by the width. However, the proportionality between the sum of weights, w, and the total cross section, $\sigma = A\sum_i w_i$, where i is the bin number, does not hold for the SMEFT data due to the negative contributions. Therefore the constant of proportionality, A, needed to be fixed manually to be $1/N_{generated}$ for partial cross section calculations. The expected number of events in each bin can then be found using

$$N = \epsilon\sigma \int \mathcal{L}, \tag{3}$$

where ϵ indicates the detector efficiency, $\mathcal{L}$ is the luminosity, and σ is the cross section found by multiplying the absolute value of the event weights by A. The predictive efficiency of 75% has only been considered when setting the limits [15].

6 Incorporating Machine Learning

6.1 Introduction

The above technique has limited accuracy and scope due to only taking in a single variable at a time unless one is constructed previously from multiple others manually. The latter approach, however, leaves room for human error and can be inefficient as many combinations would need to be tested. Therefore, a neural network was programmed to construct a compound observable from a selected range of inputs which would allow for events classification with improved accuracy. This was achieved using two independent models, the SciKit Learn Multi-layer Perceptron Classifier (MLPClassifier) model, and the TesorFlow Sequential model, which differ primarily by their API. They are both examples of Classification networks that are used for segregating data into classes based on set features. In this investigation a binary classification is conducted, meaning there are two final categories containing positively- and negatively-weighted events.

Those two models are then contrasted, and the MLPClassifier neural network is used to construct a new CP-odd observable, O_{NN}. This variable is then compared against $\Delta\phi$ with respect to its capacity to correctly separate the events. Hyperparameters are tuned for each CP-odd operator independently, giving three MLPClassifier and three Sequential models which are all used on SM data for comparison.

6.2 Principles of Neural Networks

Neural networks (NN) are a popular machine learning approach that is designed to imitate the structure and function of the human brain. Their primary objective is to identify and analyse the relationships between different features in a given data set to predict the output or similar data. Their fundamental components are *layers* of *synapses* connected by *neurons* (Figure 3) [16].

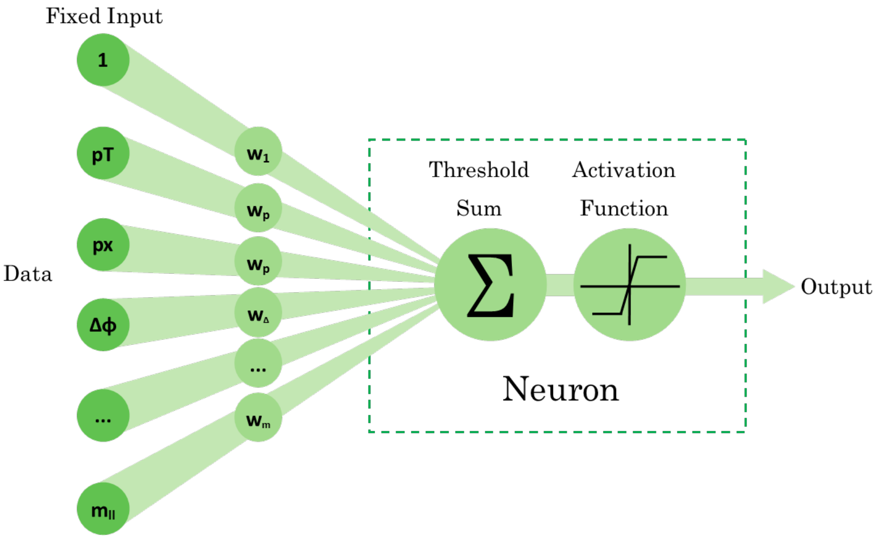

Figure 3: Schematic of a neuron in a neural network where w_i are the respective weights of each neuron that the newtowrk later updates.

The neural network begins by receiving input events that are passed into the input layer, which is usually sized according to the number of features in the data set. Next, the information is processed in the hidden layers to generate predictions, and the final output is produced by the output layer. The synapses that connect the neurons have different weights that are based on their respective contributions to the analysis. Additionally, each synapse has a bias which is then added to that of the connecting neuron, and their total is then added to the threshold sum. This sum is then passed through an activation function, which decides whether the neuron will be activated or not. If the neuron is activated, it sends a signal to each synapse in the next layer. The output of the neural network is determined by the neuron with the highest value in the output layer. ‘To improve the accuracy of the neural network, the weights, and biases of the model are updated through forward and backpropagation. Backpropagation calculates the *loss* at each *epoch*, which is another method of evaluating the effectiveness of a model by measuring the deviation of the prediction from the true value. The number of epochs refers to the number of times the neural network is being trained; it is usually set so that the accuracy and loss reach saturation. Model parameters are adjusted during every back-propagation using *optimizers*, which are algorithms looking to minimize the loss function.

The first step when building a classification model is to define two data sets, one containing the discriminant variable(s) you are using to determine the class of events, here these are the event weights and one with all the other variables which the model uses during learning. Initially, the entire large data set is split into *train* and *test* subsets consisting of 80% and 20% of all events respectively. The model is then trained on the former set by learning patterns in individual features and their combinations in order to predict the class of the event, its success is determined by comparing against the set of discriminant variables. A learning curve can be produced to display the improving accuracy of predictions as the forward and back propagation enhance the model's performance. The accuracy of a classification neural network is determined by how well the trained model can predict the class of an event from new data by looking at a set of features; this is calculated by applying the model to the test subset and dividing the number of events classified correctly by the total. A trained model can then be used for predictions.

The accuracy of a neural network is maximised by tuning the *hyperparameters*. This project will focus on the number of epochs, hidden and dropout layers, their dropout rates, the numbers of nodes in each layer, as well as types of activation functions and optimizers. Dropout layers are a regularization technique to prevent overfitting [17], which occurs when a model learns the training data in too much detail. The algorithm memorises it instead of spotting more general patterns which then leads to inaccurate predictions for unknown data sets. Dropout layers randomly remove some of the neurons in the network, making it smaller and hence, forcing it to learn more robust features. The dropout rate is another hyperparameter that can be tuned; it tells the proportion of neutrons that will be randomly deactivated. When predicting data from a model this feature is usually switched off and the entire neural network is used [18]. Moreover, both models use *early stopping*, a technique designed to prevent overfitting and improve efficiency. It involves setting a limit on the number of epochs the model can continue training for if its performance does not improve.

6.3 Model I: Tensorflow Sequential model

6.3.1 Introduction

The Sequential classification model is configured using the KERAS API which runs on top of the TENSORFLOW package, providing a more accessible interface [19]. It uses the Binary Cross-entropy loss function expressed as

$$L(y, p) = -(y \log(p) + (1 - y) \log(1 - p)) \tag{4}$$

where y and p are probability distributions from the training data and those predicted by the model respectively.

The main advantage of the Sequential classification model from TESORFLOW over SCIKIT LEARN's MLPClassifier model is its flexibility in building complex neural network architectures. It lets the user customise every layer's hyperparameters independently, whereas the latter is more restrictive as it applies some of them universally to all layers, for example, the activation function. However, this model was not used for further prediction or constructing the new CP-odd observable as its function used for predicting probabilities of events, which is necessary for constructing O_{NN}, did not give a sensible output. The function gave unreasonable values of either all ones or zeros, which is impossible, especially considering the accuracy this model converged on. Nevertheless, this model gives similar results to the MLPClassifier network, and successfully reconstructing $\Delta\phi$ is significant as it indicates robustness and reliability of prediction and reduces the likeliness of bis towards a particular neural network architecture. Moreover, the accuracy and loss curves were useful for indicating overfitting, which allowed for the selection of a suitable combination of data features for the later construction of O_{NN}.

This phase of the study solely employs one data set for analysis, following which it is determined that the model may not be used for further investigation. The training process, hyperparameter tuning, and feature importance analysis are carried out utilizing the data containing the CP-odd operator $\mathcal{O}_{\mathcal{H}\widetilde{B}}$. The two remaining pertinent operators will be examined using the MLPClassifier model as expounded in Section 6.4.

6.3.2 Reconstructing $\Delta\phi$ and feature selection for O_{NN}

Thanks to the primary analysis, one CP-odd observable which separates the data successfully based on the type interference with the SM has already been identified (Figure 2). This provides a solid base, however, ideally, a new one can be constructed which would allow for improved sensitivity and an even narrower constraint on the strength of the SMEFT interference with the Standard Model. A classification machine learning algorithm can help find one from a combination of suitable features which, when combined in a particular way, can help to identify negatively and positively interfering events with improved accuracy. However, not all variables should be fed into the model at once as it might confuse the network, resulting in inaccurate predictions or overfitting.

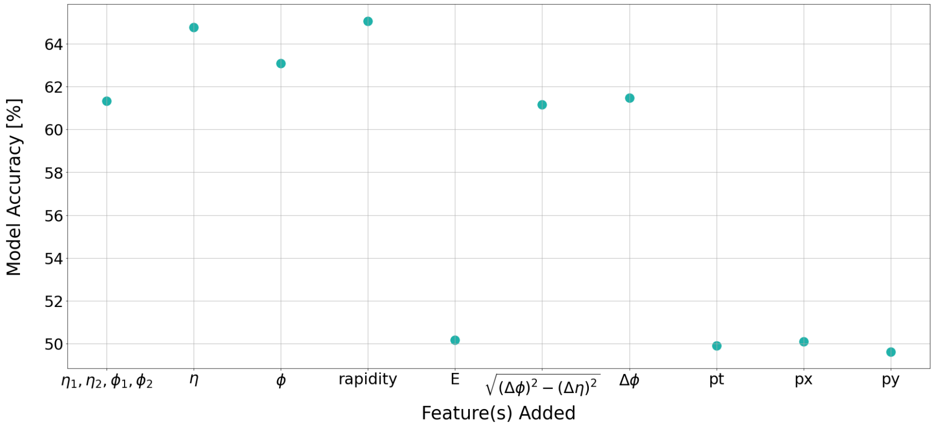

Figure 4: Plot showing the effects of different features on the accuracy of the model when added in addition to $\eta_1, \eta_2, \phi_1, \phi_2$. These four variables compose $\Delta\phi$ and their accuracy is plotted for reference as the first data point giving 61.32%. The rest are labeled with feature names that were added onto this base combination to test their effect on the model accuracy. The plotted rapidity was later deemed optimal and denoted as $rapidity_{ll}$.

The following analysis uses the variables for $\Delta\phi$ constructions from the primary analysis as guidance and then introduces new features independently into the combination one after another. FIG. 4 shows the effect of these on the accuracy of the model in comparison to $\Delta\phi$ only ($\eta_1, \eta_2, \phi_1, \phi_2$). From this, it can be inferred that the overall rapidity, or η, of the two-lepton final state, boosts the accuracy the most. Consequently, it is the preferred candidate for O_{NN} construction. Testing the addition of more than one variable at a time results in accuracy fluctuating around 50%, however, the existence of a more impactful combination, which was not tested, and therefore cannot be excluded from the realm of possibilities.

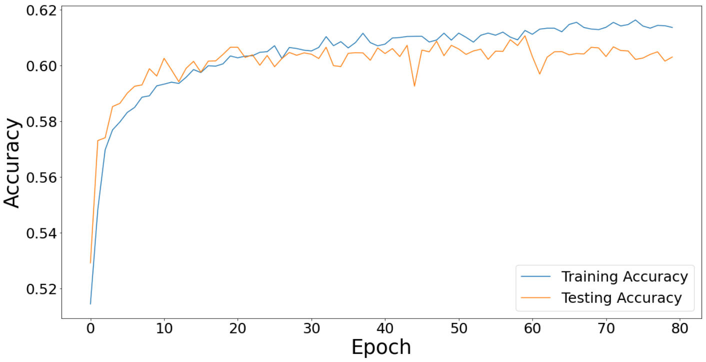

Figure 5: Accuracy curves for the Sequential model trained on: $\eta_1, \eta_2, \phi_1, \phi_2$

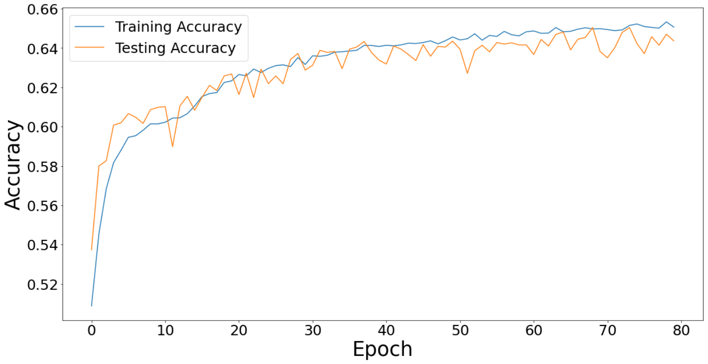

Figure 6: Accuracy curves for the Sequential model trained on: $\eta_1, \eta_2, \phi_1, \phi_2, rapidity_{ll}$

Although scaling the feature values to be of a similar order of magnitude is a fundamental step in data preparation for modeling, it was not performed in this project. The purpose of scaling is to ensure that each variable contributes equally to the final prediction and that the larger features do not dominate the smaller ones. However, in this particular project, this step led to a divergence between the testing and training accuracies, indicating overfitting, and therefore was not performed. The variables η and ϕ have ranges of values within the same order of magnitude and therefore do not require scaling with respect to each other. The weights values, which are three orders of magnitude smaller, have been constrained to be zero and one for negatively and positively weighted events respectively. This was especially important when using the $ReLU$ activation function during hyperparameter tuning, which transposes all negative contributions to zero anyway.

6.3.3 Feature Importance

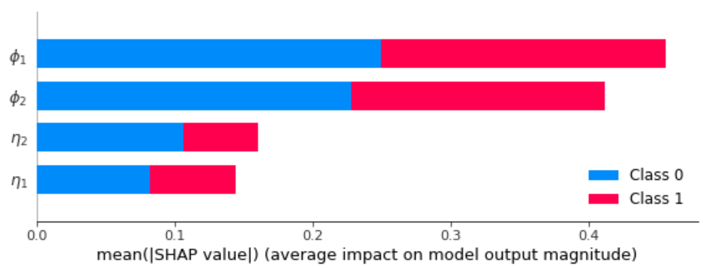

Figure 7: SHAP plot for the Sequential model trained on: $\eta_1, \eta_2, \phi_1, \phi_2$

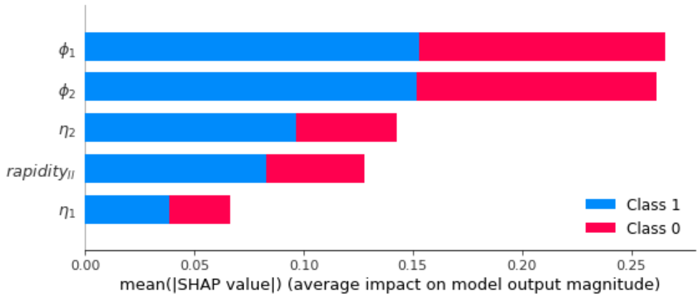

Figure 8: SHAP plot for the Sequential model trained on: $\eta_1, \eta_2, \phi_1, \phi_2$

The next important step in the analysis, after useful features have been selected, is finding out how exactly those variables affect the training process individually. This can help to expose more significant contributions which have a bigger impact on data classification. In addition, it can confirm if $\Delta\phi$ is, in fact, being reconstructed, if not then the model would utilise some features significantly less or not at all, suggesting that a different observable is being constructed instead. This investigation will use the SHapley Additive exPlanations, SHAP, package in KERAS to investigate the feature importance.

The SHAP algorithm facilitates the identification and ranking of features that significantly influence compound classification and activity prediction, utilizing any Machine Learning model. They use Shapley values, originally designed to be used in coalition game theory [20], to calculate the contribution of each feature to the difference between a model's prediction and a baseline, which typically is the average of all predictions in the data set. Computation of the importance of a feature x^i in model f involves calculating a weighted sum of this feature's impact on the model's output fx_i across all of its possible combinations. This weighted sum, $\sigma^i(f)$, is found using

$$\sigma^i(f) = \sum_{S \subseteq \{x^1,\ldots,x^n\}/\{x^i\}} \frac{|S|!(n-|S|-1)!}{n!}(f(S \cup \{x^i\}) - f(S)) \tag{5}$$

where S is a subset of features, and n is the number of features in the model [21]. However, in practice, as well as in this project, $f(S)$ is achieved by replacing the variables not currently tested with a selection drawn from a randomly generated background data set. Conclusively, making each SHAP value the sum of $f(x_i) - f(z)$ contributions of variable i, where z is from the background set.

SHAP plots visualise these weighted sums showing how much each feature contributes to categorising events to each class and ordering them from most to least impactful. From FIGURE 7 it can be inferred that it is in fact $\Delta\phi$ that is being reconstructed as all values are being used. Furthermore, η's contribute less which is consistent with them being used for ordering and not in direct calculation of the observable. FIGURE 8 shows the effect of adding the rapidity of the two-leptonic final state which improves the accuracy of the model. It becomes a more important feature than the pseudorapidity of one of the leptons, η_1. However, this does not indicate the quality of the leptons' which determines how this disparity in the significance of their pseudorapidities arises. MADGRAPH generates events according to the kinematics of the process being simulated, and the order of leptons in the event is determined by their order of production or decay, rather than their rapidity or other kinematics. Therefore, it cannot be stated why it is η_1 that is suppressed instead of η_2 and further analysis of how a NN operates is not possible. The model was further tested using only η_1, ϕ_2, ϕ_1 and $rapidity_{ll}$ which gave an accuracy of 56.22% proving that η_2 is indispensable for improved classification accuracy and so should not be omitted.

6.3.4 Hyperparameter Tuning

	Layer				
Hyperparameter	Input	I	II	III	Output
Number of Neurons	128	64	64	64	2
Drop-out rate	0	0	0.1	0	-
Activation Function	ReLu	tanh	sigmoid	ReLu	sigmoid

Table 3: Optimum hyperparameters of the Sequential model. I, II, and III refer to the three hidden layers from which each one was tested for the number of neurons, a dropout layer, and its rate and activation function. The last column was fixed and was not an outcome of the Keras Tuner analysis.

In order to optimise the model's accuracy its hyperparameters must be suitably adjusted, including the number of hidden dense and dropout layers, as well as their drop-out rates, the learning rate for the fixed *Adam* optimizer, the number of nodes in each layer and their activation functions. The KERAS Hyperband Tuner package was used to achieve this which works by running a large number of models with a small number of epochs and discards the poorly performing models. The best-performing models are then trained for more epochs, and the process is repeated until the remaining models converge or reach the maximum number of epochs. The *max_epochs* parameter sets an upper bound on the number of epochs that can be used for any model. However, the Hyperband algorithm may start with fewer epochs and progressively increase it for the top-performing models. By setting *max_epochs* to a number you can fix the number of epochs used for all models and the tuner can focus on finding the optimal values for other hyperparameters. The EarlyStopping callback is also used to stop the training early if the validation loss does not improve for a specified number of epochs.

Fixed hyperparameters included the input shape, which was set to the length of the training data set, and those of the output layer. The latter needed to consist of two final neurons representing the two classes, as well as use the *sigmoid* activation function, $sigmoid(x) = 1/(1 + exp(-x))$, which is a common choice for binary classification problems. Its output is then used to compute the binary cross-entropy loss. The activation functions *tanh* and *ReLu* are employed in several layers, with the latter activating only for positive inputs and returning zero for any negative or zero values. A different set of hyperparameters resulted in a lower overall accuracy. This analysis was done twice, once before the feature analysis which highlighted rapidity as an important variable, and another time after the final set of variables was decided (TABLE 3). The learning rate was found to be optimal at the default value of 0.001.

6.3.5 Conclusion

Overall, the Sequential model is highly effective and flexible, allowing for very precise hyperparameter tuning. These qualities were exploited for recognising significant features from the data, which then could be used in further models and for O_{NN} construction. It also provided a successful feature importance analysis which can be later compared against that of different models. However, the *predict* function was not able to produce a suitable array of probabilities that could later be used to construct the CP-odd observable O_{NN}, and therefore the extent of use of this model was limited. Instead, it predicted the probability of each event being in one class with absolute certainty, which is impossible, especially considering the model's restricted accuracy. For later construction of O_{NN} a range of values between negative and positive one would be optimal, hence a new model had to be tested.

6.4 Model II: SciKit Learn MLPClassifier model

6.4.1 Introduction

Another model was built as a consequence of the original model not predicting adequate values. The continuation of this project, therefore, is done using SCIKIT LEARN's MLPClassifier [22]. It is another type of classification NN that exploits a supervised learning algorithm to learn non-linear relationships between input and output data. It can also be configured with various hyperparameters to optimise the model's performance on the given data set, however with more restrictions than TESORFLOW's Sequential model. On the other hand, it is able to predict a sensible range of probability values for the construction of O_{NN}, and hence it has been chosen over the Sequential algorithm.

Firstly, the robustness of the selected features across models needs to be assessed in order to test if the same variables as selected in the previous investigation may be used. This is done by testing whether it provides a similar accuracy for the tested data sets as the Sequential model. To achieve this, firstly the hyperparameters will be tuned on the same set of variables, and if the outcome is comparable the investigation can be taken further.

6.4.2 Hyperparameter Tuning

This time hyperparameter tuning is done before feature importance, as the combination of variables necessary for improved accuracy is already known. This model, however, has restricted potential for personalisation as each layer cannot be tuned separately for some hyperparameters, meaning it can achieve slightly worse accuracy than the Sequential model. Only the number of neurons can be adjusted independently, whereas properties such as the activation function or optimiser type are forced to be uniform across all layers. This model was also not trained for drop-out as it proved to not have a significant impact on the accuracy.

As Scikit Learn does not have a built-in tuner tool specifically for the MLPClassifier model, instead, the Keras Tuner library can be imported or the hyperparameters can be varied using a set of embedded *for* loops. Considering the limited flexibility of this model the latter approach is chosen for efficiency.

	CP-odd operator		
Hyperparameter	$\mathcal{O}_{\mathcal{H}\widetilde{B}}$	$\mathcal{O}_{\mathcal{H}\widetilde{W}B}$	$\mathcal{O}_{\mathcal{H}\widetilde{W}}$
Number of neurons in hidden layer I	128	64	128
Number of neurons in hidden layer II	32	128	64
Number of neurons in output layer	2		
Activation Function	tanh		
Optimizer	Adam		
Accuracy	63%	65%	67%

Table 4: Optimum hyperparameters of the MLPClassifier model trained on $\eta_1, \eta_2, \phi_1, \phi_2, rapidity_{ll}$.The tuning was done using embedded for-loops which iterated through all combinations of these parameters. Resultant accuracies for each data set can also be seen, with the model trained on $\mathcal{O}_{\mathcal{H}\widetilde{W}}$ events performing the best, and $\mathcal{O}_{\mathcal{H}\widetilde{B}}$ the poorest.

This NN improved on the previous by also varying the kind of optimiser instead of only tuning the learning rate. The learning rate is important when aiming to avoid overfitting. The learning rate controls the step size at which the optimiser updates the weights and biases of the network. A high learning rate may cause the optimiser to overshoot the minimum of the loss function and fail to converge, while a low learning rate may result in slow convergence or the optimiser getting stuck in a local minimum. Its value for this model was propagated from the Keras Tuner analysis done previously (0.001), whereas the optimiser investigation has shown that the *Adam* algorithm is best for this task. It is a variation of the stochastic gradient descent method which computes the adaptive learning rates of all weight parameters individually. It computes the first and second moments of the gradients at each iteration and uses them to update each weight. This system based on using a history of gradient data makes it a very efficient optimiser in terms of running time, memory requirements, and required tuning [23].

The disparity in the number of neurons shown in Table 4 between the different data sets can be attributed to their complexities, with sets displaying higher degrees of non-linearity requiring more neurons. Additionally, the networks with more neurons in each layer perform better in the case of the three operators.

Overall, this model is less sophisticated and complex than the Sequential, however, it achieves comparable accuracy with less manipulation of the hyperparameters. It does so with only two hidden layers instead of three, and a universally applied *tanh* activation function, which also appears in one of the hidden layers of the previous model. Conclusively, this model has now been optimised for the construction of the new CP-odd observable O_{NN}.

6.4.3 Feature Importance

	Relative importance		
Feature	$\mathcal{O}_{\mathcal{H}\widetilde{B}}$	$\mathcal{O}_{\mathcal{H}\widetilde{W}B}$	$\mathcal{O}_{\mathcal{H}\widetilde{W}}$
ϕ_1	0.26	0.25	0.31
ϕ_2	0.26	0.25	0.32
η_1	0.17	0.24	0.22
η_2	0.17	0.10	0.10
$rapidity_{ll}$	0.13	0.16	0.05

Table 5: Table of relative importance of different features in the model trained on $\eta_1, \eta_2, \phi_1, \phi_2$ and $rapidity_{ll}$

In this model, the importance of each feature is assessed using the *permutation_importance* function from Scikit Learn [24]. It measures the importance of each feature in the model by randomly shuffling the values of that variable in the data set and computing the decrease in the model's performance compared to the original test set. The more significant impact it has on the model, the bigger this difference will be. The values produced are relative to each other and, here, they have been normalised to add up to unity.

Both ϕ variables always score the highest, meaning they provide the most indication about whether an event is negatively or positively interfering with the SM. Other contributions are less significant which agrees with the results obtained from SHAP plots. Interestingly, the $\mathcal{O}_{\mathcal{H}\widetilde{B}}$ data set uses the individual lepton pseudorapidities more than the joint final state rapidity, which is not the case for the other two operators. $\mathcal{O}_{\mathcal{H}\widetilde{W}B}$ data utilizes η_1 significantly more than η_2 and the joint rapidity, meanwhile events generated with the $\mathcal{O}_{\mathcal{H}\widetilde{W}}$ operator rely almost entirely on the angular variables and η_1. These differences arise from the varied non-linear complexities and are affected by the individual choices of hyperparameters.

7 ML-Constructed CP-odd Observable

The primary objective of this project is to develop an effective model for distinguishing between events by predicting how they interfere with the SM. To achieve this, a suitable CP-odd observable is needed which would best highlight these effects and allow for the analysis of the strength of this interference. Despite $\Delta\phi$ providing an adequate separation, a new, NN-constructed observable may prove more sensitive. Consequently, providing tighter constraints on the range of possible Wilson coefficients, and hence more rigorous guidance for experimentalists investigating the effects of these EFT operators in the future [25].

To construct O_{NN} in this study the latter model, the MLPClassifier binary algorithm, is employed for training the NN. The *predict_proba* function is used to generate an array of probabilities for all events in the data set. These indicate the likelihood of each event being in each of the two classes, with the first value corresponding to the probability that the event is positively weighted, P_+, and vice versa for the second entry, P_-. The two add up to unity and O_{NN} is defined as

$$O_{NN} = P_+ - P_- \tag{6}$$

[4].

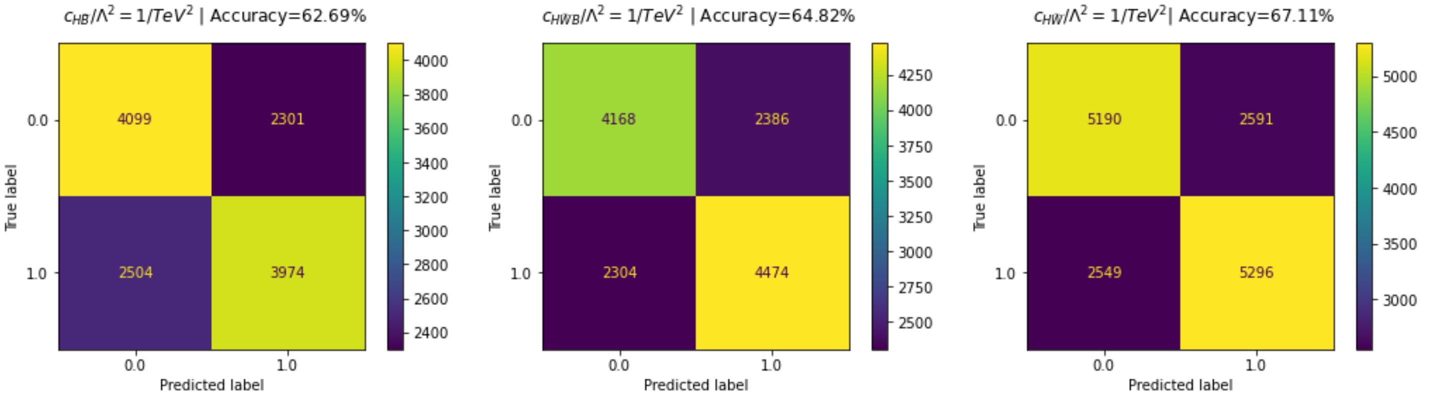

Figure 9: MLPClassifier model trained on: $\eta_1, \eta_2, \phi_1, \phi_2, rapidity_{ll}$. Confusion matrices show how many data points were classified correctly and incorrectly. Accuracy can be calculated by dividing the number of correctly classified events by the total.

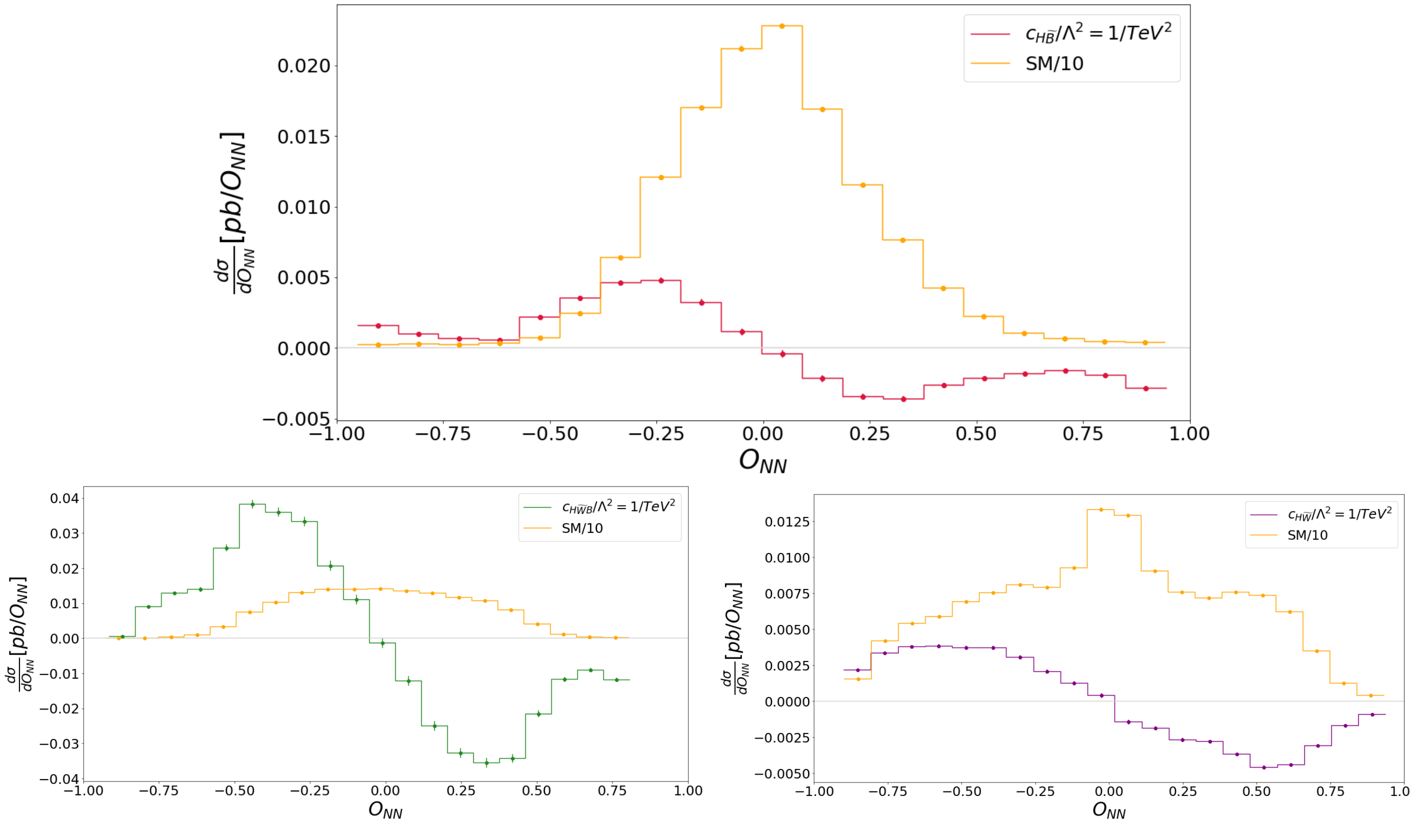

Figure 10: Cross-section plots as a functions of the CP-odd observable O_{NN} constructed using the MLP-Classifier model in Scikit Learn with hyperparameters tuned independently for each operator, $\mathcal{O}_{\mathcal{H}\widetilde{B}}$ (top), $\mathcal{O}_{\mathcal{H}\widetilde{W}B}$ (bottom left), $\mathcal{O}_{\mathcal{H}\widetilde{W}}$ (bottom right). The SM has been generated using the same model as the SMEFT variable and has been scaled by a factor of 10 for comparison.

This was then plotted (Figure 10) and each spread was inspected visually for anti-symmetry and appropriate range (-1 to 1). To improve the model's stability the *random_state* hyperparameter was fixed to 1 for all operators after a range of values has been tested. As the weights of the model are initialised randomly, it may converge on different solutions with every run. This parameter is used to set the seed for the random number generator used during training and ensures that the results are consistent and reproducible. A further improvement could be made by testing a larger variety of integers to be confident that the optimal is used.

The SM distribution is symmetrical, or close to, around zero, which contrasts the anti-symmetrical spreads of the interference data containing CP-odd SMEFT operators. Despite greater accuracies of prediction of their respectful models, the distributions of $\mathcal{O}_{\mathcal{H}\widetilde{W}B}$ and $\mathcal{O}_{\mathcal{H}\widetilde{W}}$ are not as even as that of $\mathcal{O}_{\mathcal{H}\widetilde{B}}$.

8 Likelihood Testing and Limit setting

Finally, the two observables, $\Delta\phi$ and O_{NN} , may be compared on their sensitivities to the SMEFT interference. This is done through a likelihood test, a statistical technique used to determine how well a model fits and explains a set of data. It involves calculating the likelihood values, L, and comparing them. The null hypothesis here is the SM, and the alternative hypothesis assumes a SMEFT extension. The test evaluates whether the new model significantly improves the fit to the data in comparison to the old one [26].

In counting experiments, such as this Monte Carlo simulation-based investigation, data follow a Poisson distribution

$$P(n) = \frac{\lambda^n e^{-\lambda}}{n!}, \tag{7}$$

where n represents the observed number of events, and λ represents the expected. The likelihood distribution is then obtained by binning the data and calculating the likelihood as

$$L = \Pi_i P(n_i), \tag{8}$$

where i represents the bin indexes. The sensitivities of these hypotheses can be tested by looking at their likelihood ratios over a range of Wilson coefficients range of Wilson coefficients they allow, and finding its maximum. Equivalently, $-2ln(L)$ can be minimised. According to the Neyman-Pearson lemma [27], he most powerful tool for testing the contrast between the two hypotheses the ratio of their likelihood,

$$q^{NP} = -2ln\frac{L(H_{SM})}{L(H_{SMEFT})}, \tag{9}$$

A large likelihood ratio indicates that the data is more consistent with the null hypothesis than the BSM model, conversely, a small ratio represents evidence in favour of the new hypothesis. This is used to set limits on the parameter the hypothesis depends on by comparing the likelihood of data obtained under a range of its values. This investigation will vary the Wilson coefficients aiming to determine the range of parameter values that are contained within a 95% confidence level (CL) of statistical significance. The limiting values indicate the region of parameter space where the hypothesis is supported by the data.

The test statistic, q^{NP}, for a Poisson p.d.f. can be obtained by evaluating

$$q^{NP} = 2\sum_i(-b_i ln(1+\frac{cs_i}{b_i}) + cs_i), \tag{10}$$

where i is the bin index, b equals the number of SM events, s is the number of interference events, and c is the Wilson coefficient to be varied. Both b and s are obtained using (Equation 3) and are additionally multiplied by the value of the detector efficiency of single lepton reconstruction found in the FCC-ee design report [15], which was estimated to total 75%. Furthermore, a theorem established by S. Wilks [28] states that as the sample size tends to infinity, the distribution of this ratio converges to the χ^2 distribution for the null hypothesis

$$\chi^2 = \sum_i \frac{(n_i - \lambda_i)^2}{n_i}, \tag{11}$$

where n is the number of SM events, and λ is the number of SMEFT events. Making the numerator equivalent to the number of interference events squared.

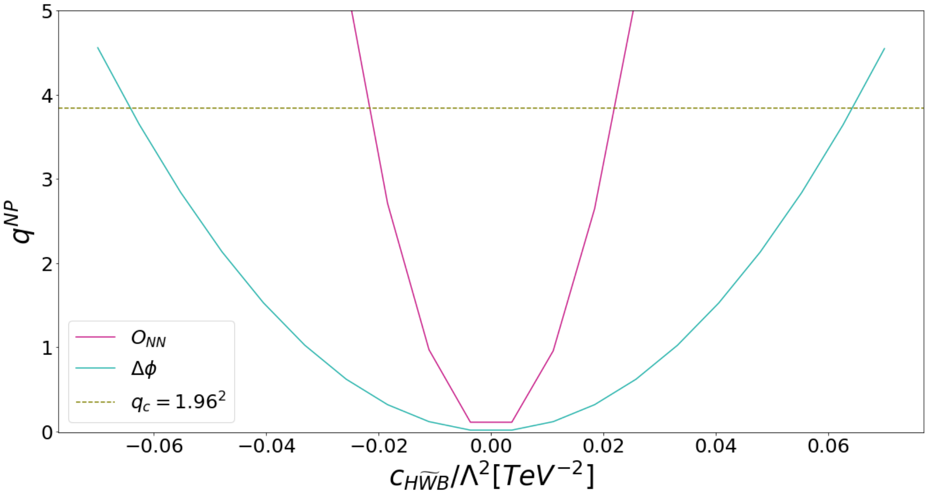

Figure 11: q^{NP} plot for the $\Delta\phi$ and O_{NN} observables. The narrower parabola, and consequently a larger gradient value, corresponding to O_{NN} implies it has greater sensitivity to the SMEFT interference. This is true for all CP-odd operator data. The biggest improvement can be observed in data generated with the $\mathcal{O}_{\mathcal{H}\widetilde{W}B}$ operator as displayed in this figure. The 95%CL is indicated with the dashed green line and the critical values are interpreted to be where the plots intersect it.

The biggest improvement between $\Delta\phi$ and O_{NN} can be seen in the CP-odd operator $\mathcal{O}_{\mathcal{H}\widetilde{W}B}$. The model used for its construction also gave the highest accuracy score out of all the interference data sets, however, the least improvement was seen in the $\mathcal{O}_{\mathcal{H}\widetilde{W}}$ operator data which had the second highest accuracy score, therefore it is not the accuracy of the model that determines this improvement.

9 Results and Discussion

This section will focus on the final results of this project. Firstly, it will give a summary of a similar investigation done for FCC-hh and HL-LHC conditions. Then, it will analyse the two observables from this report, $\Delta\phi$ and O_{NN}, and the extent to which they provide reliable results for the limit setting. It will then compare the limiting values for the strength of SMEFT CP-odd operators' interference from the lepton collider simulation, to those taken from a report investigating hadronic collisions. Finally, it will outline some criticisms of the investigations and provide guidance for potential future continuation.

9.1 HL-LHC and FCC-hh investigation overview

A twin investigation was performed for hadron colliders, FCC-hh, and HL-LHC, which used the same simulation method. It looked at the process of Higgs production via gluon-gluon fusion, subsequently decaying to two Z bosons, which then produce a leptonic pair each. The final state was assumed to contain an electron-positron pair, and a muon-antimuon pair for easier distinction during analysis. In future investigations, ideally, all combinations of two dileptonic decays should be considered. The observable used to perform interference-type based classification is taken from a 2D histogram function generated by plotting the invariant mass of the primary leptons, m_{12}, against Φ_{4l}, defined as

$$\Phi_{4l} = \frac{\mathbf{q_1} \cdot (\hat{\mathbf{n_1}} \times \hat{\mathbf{n_2}})}{\mid \mathbf{q_1} \cdot (\hat{\mathbf{n_1}} \times \hat{\mathbf{n_2}}) \mid} \times \cos^{-1}(\hat{\mathbf{n_1}} \cdot \hat{\mathbf{n_2}}), \tag{12}$$

with the normal vectors defined as

$$\hat{\mathbf{n_1}} = \frac{\mathbf{q_{11}} \times \mathbf{q_{12}}}{\mid \mathbf{q_{11}} \times \mathbf{q_{12}} \mid} \qquad and \qquad \hat{\mathbf{n_2}} = \frac{\mathbf{q_{21}} \times \mathbf{q_{22}}}{\mid \mathbf{q_{21}} \times \mathbf{q_{22}} \mid}, \tag{13}$$

where $\mathbf{q}_{\alpha\beta}$ corresponds to the three-momentum of the (anti)lepton β from the Z boson decay $Z_\alpha \to l\bar{l}$, and the three-momentum of Z_α is found from $\mathbf{q}_\alpha = \mathbf{q}_{\alpha\mathbf{1}} + \mathbf{q}_{\alpha\mathbf{2}}$. All values are in the Higgs boson centre-of-mass frame.

CP-odd Observable	$c_{\mathcal{H}\tilde{B}}/\Lambda^2[TeV^{-2}]$	$c_{\mathcal{H}\tilde{W}B}/\Lambda^2[TeV^{-2}]$	$c_{\mathcal{H}\tilde{W}}/\Lambda^2[TeV^{-2}]$
FCC-ee			
$\Delta\phi$	$[-0.34, +0.34]$	$[-0.064, +0.064]$	$[-0.24, +0.24]$
O_{NN}	$[-0.14, +0.14]$	$[-0.023, +0.023]$	$[-0.21, +0.21]$
FCC-hh			
Φ_{4l}, m_{12}	$[-0.014, +0.014]$	$[-0.026, +0.027]$	$[-0.019, +0.020]$
HL-LHC			
Φ_{4l}, m_{12}	$[-0.14, +0.14]$	$[-0.48, +0.48]$	$[-0.87, +0.86]$

Table 6: Limits on the coupling strength constant of different CP-odd SMEFT operators, c, to the SM. O_{NN}'s much narrower limits suggest greater sensitivity to new physics. This table summarises results from three different colliders operating at various centre-of-mass energies.

9.2 Limits Evaluation

Table 6 displays a summary of the constraints on the Wilson coefficients of different CP-odd SMEFT operators. Starting with the FCC-ee data, the most tightly constrained constant is $\mathcal{O}_{\mathcal{H}\widetilde{W}B}$, with its interference strength limits predicted to be over an order of magnitude smaller than those of $\mathcal{O}_{\mathcal{H}\widetilde{W}}$, which has the widest limiting coefficients. These indicate that, for lepton-induced HZ Higgsstrahlung, the operator describing the interference of the Higgs field with the electroweak field may contribute more to the relevant physical process, having a more pronounced effect on it at the FCC-ee energy scale. Consequently, the larger Wilson coefficient would suggest more CPV in the interactions of those fields in the vertex as they signify the size of the operators' contribution to the resultant amplitude, and CPV arises from the interference between their different complex phases.

When comparing these results to those obtained from the same investigation done in simulated FCC-hh and HL-LHC conditions (Table 6), it is clear that the FCC-hh will far outperform HL-LHC and FCC-ee in its potential for exploration of CPV in the Higgs sector, assuming SMEFT. It is comparative to FCC-ee, however it uses a harsher efficiency estimate and accounts for background, which the lepton collider simulation does not do. The FCC-ee will perform better than the HL-LHC, giving the same result for the constraints on $\mathcal{O}_{\mathcal{H}\widetilde{B}}$, and with improvement seen in $\mathcal{O}_{\mathcal{H}\widetilde{W}B}$ and $\mathcal{O}_{\mathcal{H}\widetilde{W}}$ when using O_{NN}. The FCC-hh investigation gives relatively reasonable results, using the Φ_{4l}, m_{12} CP-odd observable. All in all, these values suggest that looking in newly accessible regions of phase space at FCC-hh is not as important as increased luminosity for experiments looking at CPV in the Higgs sector.

9.3 Comparing Observables

This section will compare the limits obtained from the observables listed in Table 6, with particular focus on the variable constructed using machine learning. Final results from the limit setting analysis for the FCC-ee show the biggest improvements in Wilson coefficient limits for the $\mathcal{O}_{\mathcal{H}\widetilde{W}B}$ operator, as an observable is constructed using a neural network. The limiting value for this operator is reduced by a factor of 2.8 when O_{NN} separates the data by interference, meanwhile the $\mathcal{O}_{\mathcal{H}\widetilde{B}}$ and $\mathcal{O}_{\mathcal{H}\widetilde{W}}$ limits are reduced by factors of 2.6 and 1.1 respectively. For all data sets O_{NN} proved relatively more successful in separating events by their interference than $\Delta\phi$, however, this particular operator's $\Delta\phi$ distribution could be fit with a fifth-order polynomial in the $[-\pi, +\pi]$ interval, with negatively- and positively-weighted events in both halves of the histogram. Whereas, the other data sets displayed spreads which could be modelled with a third order polynomial in the same region, making the improvement between the two observables not as prominent.

Overall, both observables, $\Delta\phi$, as well as O_{NN}, are able to successfully separate the data into regions of positive and negative interference. Improvement across all SMEFT interference data sets can be seen when using O_{NN}, despite the limited accuracies of the ML models. In future investigations, a multi-class classification model can be trained instead of a simpler, binary one. A multi-class architecture can capture more of the variation in the data, and is more sensitive to slight differences between events, therefore, it can make more accurate predictions. It would allow for increased model accuracy and lead to a construction of a more sensitive observable.

9.4 Critical Analysis

Overall, the FCC-hh proves the most promising candidate for estimating the strength of BSM interference from SMEFT CP-odd operators on the standard model, and hence for investigating CPV in the Higgs sector. Its outperformance of both, the FCC-ee and the HL-LHC, can be attributed to its higher centre-of-mass energy and luminosity values. With a larger number of Higgs events a more precise analysis can be done. It also leaves potential for further constraint of the results if combined with other, complementary experiments involving other Higgs vertices. Jointly these could provide a more comprehensive understanding of the nature of CPV arising from the Higgs sector. A potential next step for the FCC-hh investigation is to incorporate machine learning algorithms, as from the FCC-ee investigation one might assume they are likely to outperform simpler, manually-constructed observables on the grounds of sensitivity to new physics.

A large limitation to all above results lies in the rough estimation of detector effects. The accuracy of experimental lepton-objects reconstruction may significantly impact the results. In the hadron collider investigations the efficiency of events reconstruction in the detector is taken to be 60% for the limit setting. It has been implemented by multiplying each expected number of events (SM and interference) by its value. For more accurate results the efficiency should be modelled using a full detector simulation instead. Additionally, the FCC-hh and HL-LHC simulations account for the background using correction factors taken from last semester's investigation. This estimate is based on the ratio of the expected number of signal events to the background, which were also calculated from data generated in MadGraph. This correction was propagated into the final set of results by multiplying each SM prediction, b, in Equation 10 by its value. However, this technique is of limited reliability and a more accurate analysis of the background effects would be required in the future.

On the other hand, in the above FCC-ee analysis the efficiency estimate of 75% was taken from the design report [15]. This is a much more reliable assumption, therefore all expected numbers of events in Equation 10 have been scaled by the square of this value to account for both of the leptons being detected. Nevertheless, the FCC-ee limits on the Wilson coefficients are likely to be underestimates due to the negligence of background effects. In accordance with last semester's report [13], including an implemented drastic cut on the lepton recoil mass of 100GeV around the Higgs resonance, the background accounts for around 2% of the data and hence, may be neglected. However, this investigation uses this assumption meanwhile considering largely looser cuts on the kinematic variables and none on the recoil mass. Therefore, a future investigation would need to reprocess the MadGraph files in Rivet with the same cuts applied, and only then estimate the efficiency. Alternatively, a new model could be tuned and trained on data sets which include a cut on the recoil mass of the leptons. Therefore, the FCC-ee results need moderation and are not to be considered above FCC-hh.

10 Conclusion

Overall, this project has achieved a set of reasonable constraints on the values of the potential SMEFT interference Wilson coefficients. The FCC-hh collider shows the most promise for CPV investigations in the Higgs sector, whilst FCC-ee needs to be rerun with appropriate cuts on the recoil mass and background scaling factors for improved confidence. Nevertheless, the limits on the SMEFT Wilson coefficients for the Higgs interactios with electroweak bosons provide grounds for further investigation into this BSM theory in search for answers to the problem of matter-antimatter asymmetry in the universe. The FCC will be the ideal facility for analyses of the Higgs sector due to its high centre-of-mass energies and luminosities for both, hadronic and leptonic detectors, predicting very large yields.

Future investigations could especially benefit from incorporating machine learning in FCC-hh simulations, preferably using a more advanced classification algorithm, such as a multiclass model. Furthermore, a more thorough background reconstruction should be performed to achieve an improved scaling factor for both FCC colliders. The detector efficiency should be reconstructed fully using an accuracy simulator, especially for the FCC-hh, in order to produce a set of more realistic results. The FCC-ee simulation's reliability would be further improved by implementing strict cuts on the invariant and recoil masses around the Z boson and Higgs resonances respectively, and, only then, applying last semester's background scaling factor.

References

[1] E. R. Paudel. "Problems of Standard Model, Review". In: *BMC Journal of Scientific Research* 4.1 (2021), pp. 65–73.

[2] L. Dickinson. "Probing Higgs Boson Interactions at the Future Circular Collider". 2023.

[3] F. U. Bernlochner, C. Englert, C. Hays, K. Lohwasser, H. Mildner, A. Pilkington, D. D. Price, M. Spannowsky. "Angles on CP violation in Higgs boson interactions". In: *Physics Letters B* 790 (2019), pp. 372–379.

[4] A. Bhardwaj, C. Englert, R. Hankache, A. D. Pilkington. "Machine-enhanced CP-asymmetries in the Higgs sector". In: *Physics Letters B* 832 (Sept. 2022), p. 137246.

[5] J. Touchèque C. Degrande. "A reduced basis for CP violation in SMEFT at colliders and its application to diboson production". In: *Journal of High Energy Physics* 2022.4 (2022), pp. 1–45.

[6] D. S. Hwang J. Ellis. "Does the 'Higgs' have spin zero?" In: *Journal of High Energy Physics* 2012.9 (2012), pp. 1–22.

[7] R. Lehnert. "CPT Symmetry and Its Violation". In: *Symmetry* 8.11 (2016).

[8] T. Agatonovic Jovin, I. Bozovic Jelisavcic, I. Smiljanic, G. Kacarevic, N. Vukasinovic, G. Milutinovic Dumbelovic, J. Stevanovic, M. Radulovic, D. Jeans. *CP violation in the Higgs sector at ILC*. 2021. eprint: 2110.12830.

[9] J. W. Cronin. “CP Symmetry Violation: The Search for its Origin”. In: *Science* 212.4500 (1981), pp. 1221–1228. (Visited on 04/04/2023).

[10] A. Höcker, Z. Ligeti. “CP Violation and the CKM Matrix”. In: *Annual Review of Nuclear and Particle Science* 56.1 (2006), pp. 501–567.

[11] MadTeam. *MadGraph5*. Version 3.5.x.

[12] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten, L. Lönnblad, S. Mrenna, S. Prestel, C.T. Preuss, others. “A comprehensive guide to the physics and usage of PYTHIA 8.3”. In: *SciPost Physics Codebases* (2022), p. 008.

[13] M. Hoffman. “Probing Higgs Boson Interactions at the Future Circular Collider”. 2023.

[14] I. Brivio, Y. Jiang, M. Trott. “The SMEFTsim package, theory and tools”. In: *Journal of High Energy Physics* 2017.12 (2017), pp. 1–57.

[15] M. Benedikt, A. Blondel, O. Brunner, M. Capeans Garrido, F. Cerutti, J. Gutleber, P. Janot, J. M. Jimenez, V. Mertens, A. Milanese, K. Oide, J. A. Osborne, T. Otto, Y. Papaphilippou, J. Poole, L. J. Tavian, F. Zimmermann. *FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2. Future Circular Collider*. Tech. rep. 2. Geneva: CERN, 2019, p. 377.

[16] S. Kohli, S. Miglani, R. Rapariya. “Basics of artificial neural network”. In: *International Journal of Computer Science and Mobile Computing* 3.9 (2014), pp. 745–751.

[17] X .Ying. “An overview of overfitting and its solutions”. In: *Journal of physics: Conference series*. Vol. 1168. IOP Publishing. 2019, p. 022022.

[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov. “Dropout: a simple way to prevent neural networks from overfitting”. In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[19] *The Sequential model*. Jan. 2022.

[20] R. Rodrıguez-Pérez, J. Bajorath. “Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions”. In: *Journal of computer-aided molecular design* 34 (2020), pp. 1013–1026.

[21] L. Ungar D. Bowen. “Generalized SHAP: Generating multiple types of explanations in machine learning”. In: *arXiv preprint arXiv:2006.07155* (2020).

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[23] A. Gupta. “A Comprehensive Guide on Optimizers in Deep Learning”. In: (2023).

[24] Yang, Jian-Bo and Shen, Kai-Quan and Ong, Chong-Jin and Li, Xiao-Ping. “Feature selection for MLP neural network: the use of random permutation of probabilistic outputs”. In: *IEEE Transactions on Neural Networks* 20.12 (2009), pp. 1911–1922.

[25] C. Cornella, D.A. Faroughy, J. Fuentes-Martın, Javier, G. Isidori, M. Neubert. “Reading the footprints of the B-meson flavor anomalies”. In: *Journal of High Energy Physics* 2021.8 (2021), pp. 1–53.

[26] R. Hankache A. Pilkington. “Introduction to Limits Setting”. 2021.

[27] “Neyman-Pearson Lemma”. 2021.

[28] G. Cowan. “Goodness of fit and Wilks’ theorem”. In: (2013).