

DNN Regression: Neural Network Application in Higgs Boson CP State Measurement

Qintong Li

Student ID: 02158205

Project ID: MLBD_01

Supervised by: Prof. David Colling

Assessed by: Dr. Daniel Winterbottom

Word Count: 8211

Submitted: September 9th 2022

Layperson's summary - Exploring the Invisible Land in an Apple

I still remember the first time I read the book called *The Map of the Invisible*. The author asked a question at the beginning of the book: imagine you have an apple, you cut it in half, then cut it in half again, and again, and keep doing that. What will you end up with? I guess it is pretty normal to have such a stage in life, being curious about the most basic but eccentric questions. As time passed, most people forgot about these 'stupid' thoughts and started to focus on more 'important' things in life. Fortunately, you are not one of them. You are still wondering whether there is an answer to that question. Congratulations, this is the first step before entering the gate of particle physics.

Let's go back to the question about the apple. After billions of cutting, we end up with a set of minor constituents called atoms. If the same logic can be applied to this element, what will be inside it? Will there be anything inside? The answer is yes. The model of atoms was discovered and finalized a century ago by Thompson, Rutherford and Schrödinger. Atoms were proved to consist of sub-atomic particles. However, this is still not the end of this path.

"...Discoveries in this field since the 1930s have resulted in a remarkable insight into the fundamental structure of matter: Everything in the universe is found to be made from a few basic building blocks called fundamental particles. Our best understanding of how these particles are encapsulated in the Standard Model(SM) of particle physics." Quoted from the website of CERN introducing the establishment of SM.

Nevertheless, by then, there was one last missing puzzle in SM: the Higgs boson. It was predicted in the 1960s and finally discovered in 2012 at CERN's Large Hadron Collider. Analysing the properties of this particle became one of the principal interests in this field. This project is also designed for the same purpose.

This goal can be achieved by reconstructions of observables. The decay products of the Higgs boson can be an excellent point to start with for the reconstructions. The obstacle in this process is to predict the kinematics of neutrinos, as it is the so-called invisible particle and is not detectable. This project uses ML techniques, especially Deep Neural Networks, to predict neutrinos without building any explicit physics model. This model used Monte Carlo truth as the learning target. This way, further analysis of the observables can be carried out based on the prediction.

Apart from fundamental observables such as mass, spin and parity, the CP state is another essential property to explore. An angle between two decaying planes of Higgs decaying products is defined as acoplanarity angle. It was cumbersome to evaluate data for too many distinct acoplanarity angle sensitivities. ML techniques can then be implemented again to simplify the process. This study improved the previous separation of this angle for one of the decay channel by approximately 26%.

We can see that the improvement of equipment and technical tools sizably accelerated the development of science in the history of particle physics. From the discovery of the Higgs boson to the breakthrough in its further analysis, each step is significant and cheerful. No matter how little knowledge we have at each stage, extracting as much information as possible from the limited condition. We grabbed all the chances like beasts. This will always be the spirit on the way to this endless road.

Hope the thoughts of wondering about the result of cutting apples never fade in your mind.

Abstract

Machine learning(ML) techniques are finding increasing applications in high-energy physics data analysis. Extracting as much information as possible from low-level variables and implementing ML techniques with this information has become a principal interest in this field. This allows an algorithm to learn the results without explicit physics models. This paper discusses how much improvement can be introduced by ML techniques, especially by implementing DNN, to the measurement of CP-sensitive observable ϕ_{CP} of the Higgs boson. By implementing a multi-output DNN regressor to predict the momentum of neutrinos decayed from τ leptons, the discrimination between the CP-even(scalar) and CP-odd(pseudo-scalar) improved vastly when both τ leptons undergo π decays. More improvements can be expected for other decay modes once the reconstruction of ϕ_{CP} is correct.

Contents

1	Introduction	4
1.1	Measurement of CP states	4
1.2	History applications of ML techniques	5
1.2.1	Mixing Angle α_{cp}	5
1.2.2	The first successful measurement of Higgs CP state	6
1.3	Focus of this study and parametrisation of CP states	7
2	Methods	8
2.1	Source of Data	8
2.2	Separating $H \rightarrow \tau\tau$ signal from Drell-Yan background by τ -polarisation	9
2.3	Methods for CP sensitive observables reconstruction	9
2.3.1	Impact Parameter method	9
2.3.2	Neutral-pion method	9
2.3.3	Polarimetric vector method	9
2.4	Acoplanarity angle ϕ_{CP}	10
2.5	Main structure of the multi-output NN Regressor	11
2.5.1	Pre-selection of data	11
2.5.2	Input variables and targets	11
2.5.3	Define customised loss function	12
2.5.4	The structure of the multi-output DNN regressor	13
3	Results	14
3.1	Distributions of energy fraction for Higgs boson and Z boson	14
3.1.1	Higgs with 125 GeV mass	15
3.1.2	Higgs with 95 GeV mass	16
3.2	Linear regression for Higgs momentum	18
3.2.1	Loss curves for training and testing data	18
3.2.2	Linear regression for Higgs momentum and its components in 2D distribution	18
3.3	Linear regression for neutrino momentum using multi-output DNN regressor	20
3.3.1	Evaluation of predicted neutrino momentum components	20
3.3.2	Comparison between multi-output DNN regressor and basic ML models	21
3.4	Reconstruction of acoplanarity angle ϕ_{CP}	22
4	Conclusion	24

4.1	Possible application of GNN architecture	24
4.2	Main achievements	24
5	Acknowledgements	26
	Appendices	30
A	Important code	30
B	Repeated graphs	31
B.0.1	Higgs M95	31

1 Introduction

Machine learning(ML) techniques have been rapidly finding a place in high-energy physics data analysis such as data selection, data classification and further predictions [1]. Various applications of ML techniques allow improvements in reconstructing high-level physics quantities without specific physics models. It shows the probability of having a general model and improves the efficiency of data analysis. This study addressed the application of ML techniques, especially Deep Neural Network(DNN), in measuring the Higgs boson CP state.

1.1 Measurement of CP states

After the discovery of the Higgs boson by LHC experiments in 2012 [2, 3], a large and growing amount of literature has investigated other basic quantum numbers such as mass [4], spin [5, 6] and parity [6] which is crucial for studying the feature of Higgs boson. Its CP state is one of the most significant properties of Higgs boson that help us achieve a better understanding of this particle[7].

Extensive research has been conducted to measure the CP state of the Higgs boson[7]. For example, the Standard Model predicts that the Higgs boson is purely CP even. Therefore, any CP odd measurements will imply physics beyond Standard Model and worth carrying analysis.

As was stated in 1993 by Krämer [8], one way to achieve that can be accomplished by measuring the transverse spin correlations. One of the most powerful methods to do that is through its decay to τ leptons [9]. This is based on the sensitivity of the transverse spin to the parity of the Higgs. This sensitivity can be indicated by angular correlations of secondary decay products, particularly in the distribution of the products from τ decays.

However, the main limitation of this method is the validity of detecting the loss of neutrinos and evaluating the detector precision [7]. Therefore, based on Krämer's work, Bower improved the method of measuring the parity of the Higgs in 2002 [10]. Even though it does not need to adapt the measurements from LHC and the reconstruction of τ rest-frames, the method only applied to the dominant τ decay channels [7]. This is because it is relatively easier to define the acoplanarity angle in this decay mode. In light of this study, extending to other decay modes has become particularly important for the next stage.

In addition, the secondary decay products π^0 , π^\pm and most importantly, the neutrinos. The neutrinos are the so-called Invisibles and cannot be detected directly from the collision [11]. One crucial step before analyzing the Higgs boson's property is reconstructing the neutrinos with valuable features such as the directions and angular-related information, which will be mentioned later in the study.

In most of the following publications [9, 12, 13, 14, 15], a well-established idea is that the promising potential improvement in achieving sensitivity to Higgs CP states can be starting to use information in τ decay vertex. For example, Ralph Józefowicz and his colleagues have concentrated only on the sensitivity to CP states during Higgs decaying through analyzing acoplanarity angles [7]. Instead of exploring the effect of the decay vertex, they studied more essential details in the cascade τ decay. Based on Bower’s work, they improved the definition of acoplanarity angle and split the events into more groups so that it can be used to discriminate more decay modes [7].

1.2 History applications of ML techniques

When the analysis of decaying modes was extended from $\pi \pm \pi^0$ to 3π , it is found that part of the sensitivity could have been lost in the decays due to the more complicated nature of decay with the 3π product. The Deep Learning Neural Network classification was then introduced to quantify this effect. More details and formal discussion can be found in [16]. This new inclusion increases the available fraction of $H \rightarrow \tau\tau$ decay rate from 6.5 % to 11.9 %, for the analysis discussed here only considered either τ decay to 3 charged π or τ decay to $\pi \pm \pi^0$. The improvement is promising as they almost doubled available statistics [7].

The accomplishment of machine learning techniques in this field was approved by extensive literature. As it was supported by Barberio in 2017 [17], although it is currently challenging to differentiate the pseudo-scalar Higgs boson state from the scalar ones, which are favoured in measurements of CP states, the Deep Learning techniques have been applied in this study to extract the sensitivity as much as possible. Lasocha also pointed out in 2020 that machine learning offered a significant breakthrough in distinguishing scalar and pseudo-scalar Higgs states by their complex multidimensional signatures [18].

1.2.1 Mixing Angle α_{cp}

However, a closer look at the literature on the condition when applying machine learning techniques to this field reveals several gaps and shortcomings. For example, it did not discuss the possible contamination of the background signals or mis-reconstructed signals while training the Deep Learning Neural Network formally [7, 18]. Lasocha’s work also commented that the machine learning techniques applied in this study have limitations concerning the measurement of the hypothetical mixing angle of Higgs boson parity states as part of the study. This significant quantity was also defined in Józefowicz’s research and widely used in the following studies [7].

Lasocha investigated the same topic and extended it to the potential of determining the preferred mixing angle of the Higgs boson decays by DNN [18]. Based on Józefowicz’s work, Lasocha’s research mainly focused on the discussion of the distinction between two different machine learning techniques in Deep Neural Networks: classification and regression, when both of them are used to distinguish the preferred mixing angle of Higgs boson events.

The preferred mixing angle was defined for the clarity of the discussion on the DNN results: $\alpha_{CP} = 2\phi_{CP}$. The $\alpha_{CP} = 0, 2\pi$ indicate scalar state and the $\alpha_{CP} = \pi$ to pseudo-scalar one [18]. Therefore, it is obvious that some important information about the Higgs boson CP state is encoded in the angles between the decay products and the angle ϕ_{CP} between the two τ lepton decay planes in the H rest frame. An illustration of ϕ_{CP} of a single π decay is shown in Figure 1 [19].

To be compatible with the previous studies [7], all the data sources were consistent with Józefowicz’s work. The extended part of the research and the physics quantities such as spin

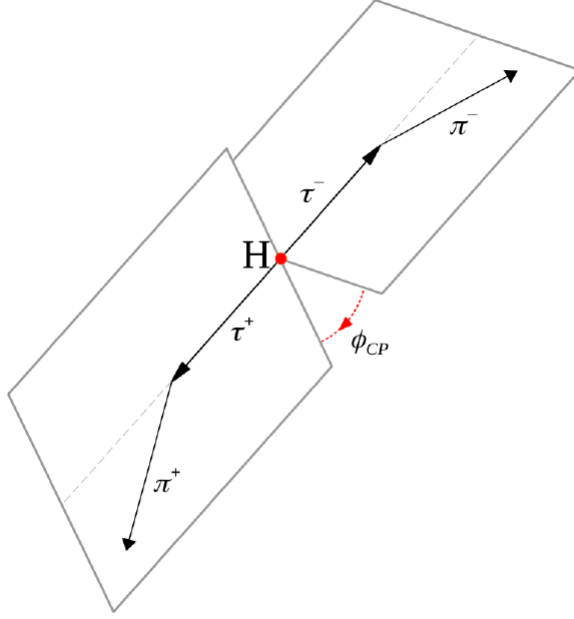


Figure 1: The decay planes of two τ leptons decaying to a single π^\pm . The angle ϕ_{CP} is the angle between the decay planes in the H rest frame.

weight of different CP mixing angles α_{CP} were also calculated and stored.

1.2.2 The first successful measurement of Higgs CP state

The latest breakthrough in this field is the first successfully measured CP structure of the Yukawa coupling between the Higgs boson and τ leptons by CMS collaboration in 2021 [19]. It showed the consistency of experimental measurements and the Standard Model prediction. Machine learning techniques DNN were also significantly applied in the analysis.

The data used for this study is collected at pp collision when centre-of-mass energy \sqrt{s} is equal to 13 TeV by the CMS detector at the LHC, corresponding to an integrated luminosity of 137 fb^{-1} . This is consistent with the research done by Józefowicz and Lasocha. A similar method of using mixing angles was applied in this study.

Based on previous studies, CMS collaboration developed a more concise and efficient method and defined an effective mixing angle $\alpha^{H\tau\tau}$ between CP-even and CP-odd τ Yukawa couplings. The effective mixing angle $\alpha^{H\tau\tau}$ is defined as [19]:

$$\tan \alpha^{H\tau\tau} = \frac{\tilde{\kappa}_\tau}{\kappa_\tau} \quad (1)$$

, where $\tilde{\kappa}_\tau$ and κ_τ are coupling strength modifiers.

The most noticeable development of this study was its deployment of machine learning techniques. The first application was similar to previous studies, the Deep Neural Network (DNN) called ‘DEEPTAU’ for τ lepton reconstruction.

Another improvement of this research was the experimental techniques developed for this analysis. The multivariate analysis discriminant (MVA) was deployed to identify τ decay modes. After establishing this technique, the other application of the machine learning technique is a

boosted decision tree (BDT). It was implemented to optimize the performance in discriminating the τ decay modes. This decision tree was trained by XGBoost framework [19].

Unlike the previous study, most research has been carried out only with binary classification because they focused on separating scalar and pseudo-scalar Higgs states [7, 18]. The machine learning techniques improved in this study could identify multiple features of events at once. Both multi-class neural networks and multi-class BDT algorithm combined with the XGBoost package were used to optimize the performance of categorization [19].

The effective mixing angle was found to be $-1 \pm 19^\circ$ as $0 \pm 21^\circ$ at the 68.3% confidence level. The results also disfavoured the pure CP-odd scenario at 3.0 standard deviations. The results showed compatibility with predictions from the standard model Higgs boson. More importantly, as the leading uncertainties were statistical, the precision would increase as the data accumulated. The limitation of this literature has not yet been evaluated academically for its advance. However, this research offered limited explanations in specific analysis with machine learning techniques compared to previous literature.

1.3 Focus of this study and parametrisation of CP states

Experimental data consists typically of events to the magnitude of 10^6 . Each event represents a particle in a multidimensional coordinate space. This is usually presented by four-momenta and other observable quantities of each detected particle in this event. Studying distributions of these quantities and the ones constructed from these quantities is one of the main goals of this research. Based on the analysis of these distributions, looking for physically meaningful interpretations of the results is also crucial for making a reasonable conclusion. In this case, ML techniques can be implemented to reduce the complexity of the problem.

However, using only low-level variables as input for the ML algorithm is inadequate to capture all information that can reconstruct the wanted quantity. Mixing the low-level inputs with the high-level ones becomes the seemingly best strategy. Similar methods were presented in these studies [20, 21].

Nevertheless, this previous literature not only provided opportunities for developing the implementations of ML techniques but also the prerequisites of future studies in related studies [1]. With the help of results from previous studies, we reconstructed the core variable: polarimetric vectors of two τ and their decayed products by using the predicted information made by the DNN model. The polarimetric vector can be considered as an estimation of the most likely direction of the spin vector of the τ lepton in the τ lepton rest frame [22]. We start by reconstructing the τ lepton momenta required to compute the polarimetric vectors.

The reconstruction of neutrinos is an inevitable step before starting reconstructing the τ lepton momenta. Therefore, reconstructing partial neutrino information, especially the four-momenta (6 variables in total), becomes the first step to obtaining the polarimetric vector. As Lasocha mentioned in their latest studies, only limited input information (hadronic visible decay products) was provided in their previous study. This limitation might be overcome by adding more high-level variables and hence being able to reconstruct the neutrinos (Invisibles) momenta approximately and other event kinematics (the momenta of visible τ decay products, overall missing energy met) [1]. The case in this study can be divided into the following steps:

1. Reconstruct the four-momenta of the decay products from each τ leptons, i.e., π^0 and π^\pm .
2. Compute the τ lepton mass from Monte Carlo truth and the one from reconstructed level information and missing transverse energy met for each event.

3. Build data set for inputs and output target.
4. Build DNN model with a customized loss function
5. Train model to predict the neutrino four-momenta(6 variables).
6. Compute polarimetric vectors for each decay and obtain the acoplanarity angle.

After step 1 and step 2, all the baseline information constructed from the Monte Carlo truth is ready to be compared to the result obtained by DNN models. The process can be considered an approach to achieving high-level variables from essential low-level variables without explicit physics models [1]. A similar strategy is applied by the following literature [23, 24]. By accomplishing these steps, the framework is ready for DNN implementation. Therefore, steps 3 and 4 start to build the DNN model customized to this case.

Once the DNN regressor makes prediction the neutrino four-momenta, it can be used to obtain the acoplanarity angle ϕ_{CP} with the help of the polarimetric vector method. The performance of the DNN model is also compared to traditional methods. The improvement and limit will be discussed in the conclusion.

This paper will be organised as follows. The Source of data and preprocessing of the H signals are outlined in the first part of Methods, followed by the reconstruction methods for CP-sensitive observables. After that, an essential variable Acoplanarity angle is presented in the next part of the Methods. With all the essential background fully outlined, the main structure of the DNN model is presented in the last bit of the Methods. In the Result part, the separation of Higgs signals from Drell-Yan background signals, linear regression for neutrino kinematics and the reconstruction of Acoplanarity angles are listed respectively. A conclusion and the improvement of this study are discussed in the end.

2 Methods

2.1 Source of Data

Some of the valuable databases in numerical studies were also listed here as background information. The signal and relevant background processes are modelled with Monte Carlo simulated events samples. The signal samples with an H produced through gluon-gluon fusion (ggH), vector boson fusion (VBF), or in association with a W or Z vector boson (denoted as WH or ZH, or VH when combined) are generated at next-to-leading order (NLO) in perturbative quantum chromodynamics (QCD) with the POWHEG 2.0 [25]-[26] event generator [19].

The decay of the H does not depend on its production. These samples are simulated without accounting for the τ spin correlations. After the samples have been generated, the TAUSPINER package [27] is used to calculate event weights that can be applied to the simulated signal samples to model τ polarisation effects for a boson with CP-mixing angles of 0, 45, and 90°. There is no normalisation effect from the re-weighting procedure, i.e. the integrated $H \rightarrow \tau\tau$ cross-section of the signal samples is invariant under rotations in $\alpha^{H\tau\tau}$.

Monte Carlo generated events are processed through a simulation of the CMS detector based on GEANT4 [28] and reconstructed with the same algorithms as the ones used for data. This is also referred to as generator-level information for later.

2.2 Separating $H \rightarrow \tau\tau$ signal from Drell-Yan background by τ -polarisation

Before applying any ML technique, an analysis to separate the Higgs signal from the background, the background signal is generated from the Drell-Yan process. The energy fraction of the decay products from each τ leptons is plotted in both 1D and 2D distributions for different decay modes. This is based on the theory presented by Moretti in 2002 [29]. It was shown in this study that the correlation between the τ -polarisation can serve as a distinctive way to separate the $H \rightarrow \tau\tau$ signal from the Drell-Yan background. The τ -polarisation is defined as the two decay modes we are interested in at the generator and reconstructed level:

$$R_{i,gen} = \frac{E_\pi}{E_\tau} = \frac{E_\pi}{E_\pi + E_\nu} \quad (2)$$

$$R_{i,gen/reco} = \frac{E_\rho}{E_{\tau_{vis}}} = \frac{E_{\pi^\pm}}{E_{\pi^\pm} + E_{\pi^0}} \quad (3)$$

, where the numbering i will be addressed later in the Results section.

By showing the 2D distribution of the selected events, the different distribution patterns can be used to distinguish the $H \rightarrow \tau\tau$ signal from the Drell-Yan background.

2.3 Methods for CP sensitive observables reconstruction

2.3.1 Impact Parameter method

The transverse impact parameter here is defined as the minimal distance between the primary vertex and the point on the track [30]. For each τ lepton, the impact parameter and its charged-particle momentum vector can be used to define a plane. This plane is constructed in the laboratory frame and only represents the plane of a single charged pion and neutrino instead of the genuine τ lepton frame[19].

2.3.2 Neutral-pion method

This method can be applied to hadronic decay channels where at least one τ lepton decays involving more than one outgoing hadron. We describe the method applied to the intermediate ρ meson decay. For the ρ meson decays, the reconstructed four-momentum vector in laboratory frame for the decay products vector λ is replaced by the four-momentum vector of the π^0 , which means we use the planes spanned by the ρ decay products to define the ϕ_{CP} observable [19].

2.3.3 Polarimetric vector method

This method can be applied to all decay channels. Only two decay modes of the τ leptons are considered in this case: $\tau^\pm \rightarrow \pi^\pm \nu$ and $\tau^\pm \rightarrow \rho^\pm \nu$, followed by $\rho^\pm \rightarrow \pi^\pm \pi^0$. Decay mode $\tau^\pm \rightarrow a_1^\pm \nu$ is also considered but not completed due to time limitations, which will be specified more in the conclusion part. The first two decay modes form four possible hadronic final-state configurations, each accompanied by a corresponding neutrino pair. This variable is defined as Eq.(4) and (5) for ρ decay mode as [31]:

$$h^i = \mathcal{N}(2(q \cdot N)q^i - q^2 N^i) \quad (4)$$

, where \mathcal{N} is a normalization function, q is the difference of the π^\pm and π^0 four-momenta and N is the four-momentum of the τ neutrino (all defined in the τ rest frame). The term $q \cdot N$ is defined as:

$$q \cdot N = (E_{\pi^\pm} - E_{\pi^0})m_\tau \quad (5)$$

For π decay mode, h^i is simply defined as the four-momenta vector of the single charged pion boosted into the τ frames as a three-vector and then boosted back to the laboratory frame.

2.4 Acoplanarity angle ϕ_{CP}

To compute ϕ_{CP} , the process is shown in the following equations(6-10) as the methods of calculating the polarimetric vector $\vec{h}_{1,2}$ (they can also be the corresponding four-vectors computed by the other methods) are outlined in the last section:

$$\vec{k}_{1,2} = \frac{\vec{h}_{1,2} \times \vec{h}_{1,2}}{|\vec{h}_{1,2} \times \vec{h}_{1,2}|} \quad (6)$$

, where $\vec{k}_{1,2}$ can now be used to reconstruct ϕ_{CP} . Nevertheless, before that, ϕ^* is calculated without applying the phase shift.

$$\phi^* = \arccos(\vec{k}_1 \cdot \vec{k}_2) \quad (7)$$

$$O^* = -(\vec{h}_1 \times \vec{h}_2) \cdot \vec{h}_1 \quad (8)$$

, where ϕ^* is ϕ_{CP} before phase shifting and O^* is the PV sign that is used to apply the corresponding phase shift. ϕ_{CP} can be shifted in the methods presented in equation (9):

$$\phi_{CP} = \begin{cases} \phi^* & \text{if } O^* \geq 0 \\ 2\pi - \phi^* & \text{if } O^* \leq 0 \end{cases} \quad (9)$$

Another valid angular variable to approximate the energy fraction is defined similarly in this process. This will be discussed more in Results.

$$C_{1,2} = \vec{h}_{1,2} \cdot \vec{n}_{1,2} \quad (10)$$

, where $\vec{n}_{1,2}$ are the unit vectors for the decayed products from the τ leptons in the laboratory frame.

After fully computing the acoplanarity angle, its distribution can be used to identify the CP properties of Higgs by comparing the distribution applied with scalar weight and pseudo-scalar.

Figure 2 shows the normalised distribution of ϕ_{CP} at the generator level. ϕ_{CP} here is calculated in the rest frame of the H, for the scalar, pseudo-scalar and the ϕ_{CP} distribution from Drell–Yan processes. These distributions are for one of the scenarios where both τ leptons decay to a charged pion and a neutrino [19].

It is important to note that the distribution of ϕ_{CP} for the Drell–Yan background is constant. The acoplanarity angle ϕ_{CP} was initially introduced in the context of e^+e^- collisions [8, 32],

where ϕ_{CP} can be calculated in the H rest frame. In this study, the methods for estimating ϕ_{CP} have been extended and optimised for hadronic collisions [33]. Throughout this study, the angle between the τ decay planes is denoted as ϕ_{CP} , irrespective of the frame in which it is calculated.

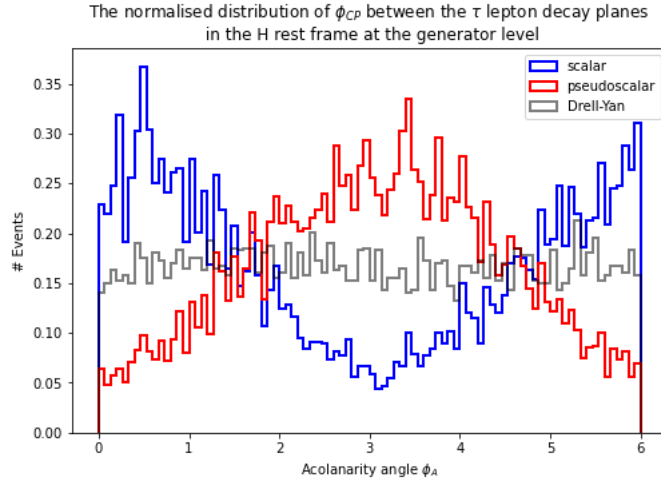


Figure 2: The normalised distribution of ϕ_{CP} between the τ lepton decay planes in the H rest frame at the generator level.

2.5 Main structure of the multi-output NN Regressor

In this study, the ML algorithm is applied to analyze the CP state in the decay channel $H \rightarrow \tau\tau$.

2.5.1 Pre-selection of data

Before analyzing the DNN model, pre-selection data is a fundamental prerequisite to optimizing the model's efficiency. The data set includes the information of decay mode flags: decay mode 1(dm_1) stands for leading τ decay mode and decay mode 2(dm_2) stand for sub-leading τ decay mode. As was stated in the previous section, only hadronic decay is discussed in this study. Therefore, $dm = 11$ is excluded in the data set, while $dm = 0, 1, 10$ represent single charged pion decay (π mode), neutral pi decay (ρ mode) and 3-prong decay (a1 mode).

2.5.2 Input variables and targets

The input with a dimension of 72 for the DNN model is categorized as:

1. Single charged π and neutral π (four-momenta) kinematics decays from each τ .
2. Missing energy met in x and y direction(Cartesian coordinate).
3. Impact parameters for both taus in Cartesian coordinate.
4. Secondary vertex information.
5. Reconstructed tau kinematics.

The 6 output targets are also listed here for each tau, 12 in total:

1. Neutrino momentum in Cartesian coordinate(3 variables for each tau). Energy is the magnitude of the 3-vector.

2. deltaRapidityPhi: $\Delta R = \Delta\phi^2 + \Delta\eta^2$, where ϕ is the spherical polar angle (from 0 to π , inclusive) and η is the rapidity which is a kind of spherical polar angle [34]. This quantity is defined to specify the difference between the neutrino vector and the visible decay product vectors.
3. Scalar product between the visible decay product and neutrino for each tau, which can also be treated as the projection of neutrino vector on visible decay product direction.
4. Cross product magnitude between the visible decay product and neutrino for each tau can also be treated as the distance between the neutrino and visible product vectors.

After the ϕ_{CP} is computed, the weight for the scalar and pseudo-scalar event included in the same data set will be applied for separation.

2.5.3 Define customised loss function

A loss function is a method of evaluating how well the algorithm models the data set. During training, the loss calculated by the loss function in each epoch tells how well the model is trained. If the predictions are off, the loss function should output a higher number and vice versa. Loss functions are also related to model accuracy, another critical component in evaluating the performance of the ML models.

A normal built-in loss function such as MAE is usually given in the structure for a baseline model. For each prediction made by the model, the loss function will measure the absolute difference between the predictions and the given target values. For example, MAE in mathematical notation can be defined as:

$$MAE = \frac{1}{N} \sum_{j=1}^n |y_{pred} - y_{true}| \quad (11)$$

Therefore, checking whether the loss function is well designed for the specific case is also essential. Due to the complexity of the case, giving a simple MSE or MASE loss function for the DNN regressor model is not capable of achieving the goal of regressing multiple variables. A more complicated loss function is defined to customize the study.

In this case, we have 12 outputs instead of only one. Due to the nature of each predicted physics quantity, the impact of each quantity on the final reconstruction is different. Therefore, to penalize each quantity equally or add bias to the preferred target, applying individual weight to each target value is important. This is addressed in more detail in Appendix A. Similarly to the MAE defined above, the loss function defined in this case can be written as:

$$MAE_{multi} = \frac{1}{N} \sum_{j=1, i=1}^n w_i \cdot |y_{i,pred} - y_{i,true}| \quad (12)$$

, where w_i is the weight for each output.

In this study, the information on neutrino kinematics is what we want to partially reconstruct at the end. However, if the information of four-momenta is the only target, it will be strongly biased towards the tau four momenta because the neutrino direction only deviates subtly from the tau direction. Therefore, adding more angular information related to the distance between the neutrino and the visible decayed products can separate these two vectors. The deltaRapidityPhi is added to address the vector difference in spherical coordinates. Similarly, the scalar and cross product are used to give information on the correlation between these two vectors.

2.5.4 The structure of the multi-output DNN regressor

The library used in this study to build the architecture is TensorFlow. Figure 3 is a flowchart showing the structure of the multi-output DNN regressor.

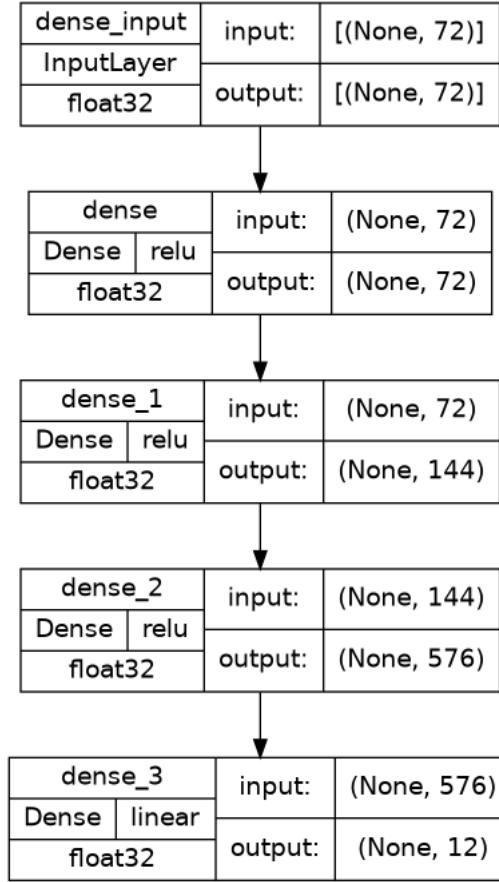
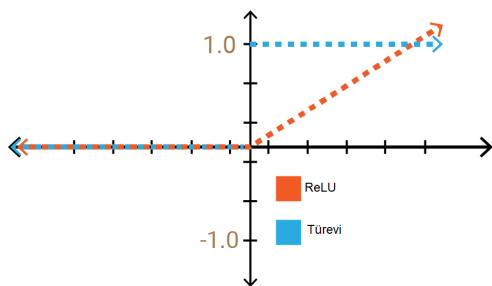


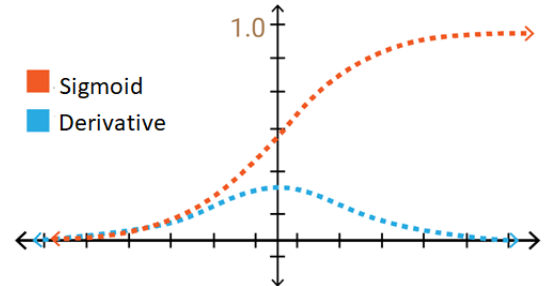
Figure 3: A visualisation of DNN model architecture

Activation functions

Here, an input layer with an input dimension of 72 is followed by three Dense layers with the ReLU function as the activation function. The last layer is the output layer with a linear activation function. The activation function ReLU function in the DNN regressor and a Sigmoid function is shown in Figure 4 [35]:



(a) A plot of ReLU activation function



(b) A plot of Sigmoid activation function

Figure 4: Activation function plots in this DNN regressor

Sigmoid activates almost all neurons in the same way. This means that the activation is very intensive [35]. Figure 4 shows the advantage of using a ReLU function as the activation function

compared to a Sigmoid function is evident when the NN is getting larger. The ReLU function with a value of 0 on the negative axis means that the network will run faster and makes ReLU a better choice when we got multi-layers of structure, which improves the efficiency sizably.

Nevertheless, the ReLU function is still not the best activation function as it might cause incapability of learning in the 0 activation region. The leaky-ReLU function gives a leaky value of 0.01, whereas the ReLU function will give a 0, which can avoid the problem encountered when using a ReLU function as an activation function. The plot of a leaky-ReLU function is shown in Figure 5.

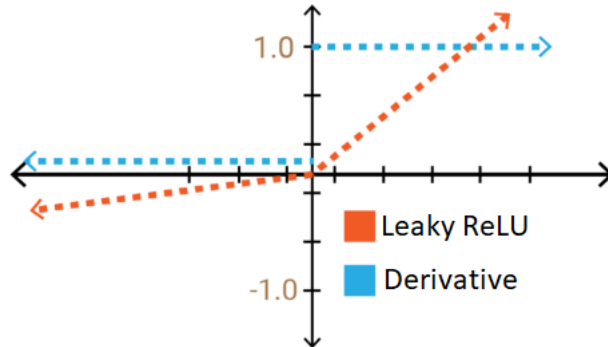


Figure 5: Plot of a Leaky-ReLU function

Weight initializer

The weight initializer is a critical consideration in the design of a neural network model. Weight initialization aims to prevent layer activation outputs from exploding or vanishing during the course of training. If either occurs in the process, loss gradients will either be infinite or approach 0 to be meaningful, and the network will take longer to converge or even not be able to do so [36].

In this model, optimizer was chosen as RandomNormal. This means the initializer will generate a normal distribution of weights every time.

3 Results

The results in this section follow the same logic as the Methods section. After the pre-processing of data, the Higgs boson signal is separated from the Drell-Yan background using the energy fraction. Then the polarimetric vectors and the corresponding ϕ_{CP} are reconstructed with the neutrino information predicted by the multi-output DNN regressor. All the necessary neutrino information at the reconstructed level in this part is the prediction from the DNN regressor. Therefore, the distributions in this section are plotted at the generator level. A series of comparisons will be made for the final possible decay configurations reconstructed by different methods listed in the Methods section such as impact parameter, neutral-pion method and polarimetric vector method.

3.1 Distributions of energy fraction for Higgs boson and Z boson

This section discusses how to use τ polarisation to distinguish the Higgs boson signal from the Z boson signal(Drell-Yan) in this case. The method is applied for both cases when Higgs boson mass is 125 GeV and 95 GeV.

3.1.1 Higgs with 125 GeV mass

The final decay configuration discussed in this section will be:

1. When both τ leptons decay into a single charged pion($\pi\pi$ decay).
2. When both τ leptons decay into one charged pion and a neutral pion($\rho\rho$ decay).
3. One τ lepton decays into a single charged pion, while the other decays into one charged pion and a neutral pion($\pi\rho$ decay).

π decay for both τ leptons

Figure 6 shows the distributions for Higgs and Z bosons when both τ leptons decayed from they undergo a single charged pion decay. Here, we redefine the energy fractions as R_1 and R_2 , where R_1 is the energy fraction of leading τ lepton and R_2 is the energy fraction for sub-leading τ lepton. The definition was stated earlier in the Methods section in equations (2) and (3). For Higgs, there is one tau energy larger than the other (the leading tau), while for the Z boson both τ leptons decayed from it should have similar energy.

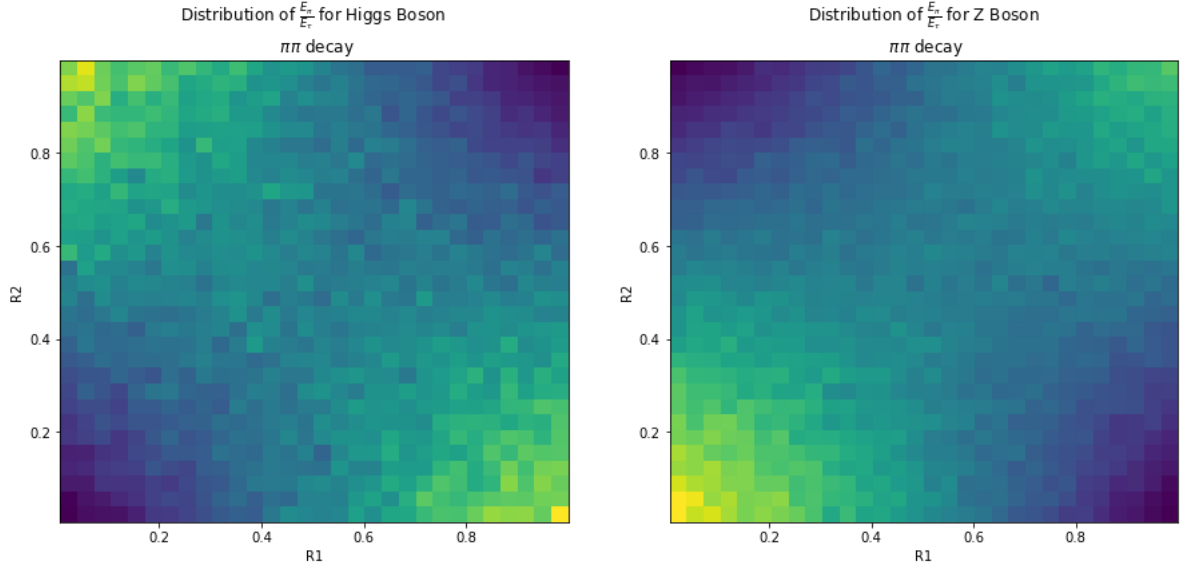


Figure 6: Distribution energy fraction of Higgs boson(left) and Z(right) boson

ρ decay for both τ leptons

Figure 7 is the 2D distribution for Higgs and Z boson $\rho\rho$ decay modes. As shown in the figure, Higgs boson distribution is biased towards the edges while Z boson biases towards the corners. The mass difference between them might cause this. Different weights applied to the Higgs boson distribution can also cause a difference in the distribution. The weight applied in this study is simply the weight for scalar products. The energy fraction for this final decay mode configuration is computed only from the visible decay products so it can also be applied to the reconstructed level.

ρ decay for leading τ lepton π decay for sub-leading τ

Figure 8 shows the distribution of energy fraction R_1 for leading τ and R_2 for sub-leading τ . In this final decay configuration, a 1D distribution plot is enough to determine the difference between a Higgs signal and a Drell-Yan background.

We also applied Pt cuts on the data set to achieve better separation. The constraint is set as: transverse momentum $P_T > 40$ GeV and rapidity $\eta < 2.3$ for both *tau* lepton decay products.

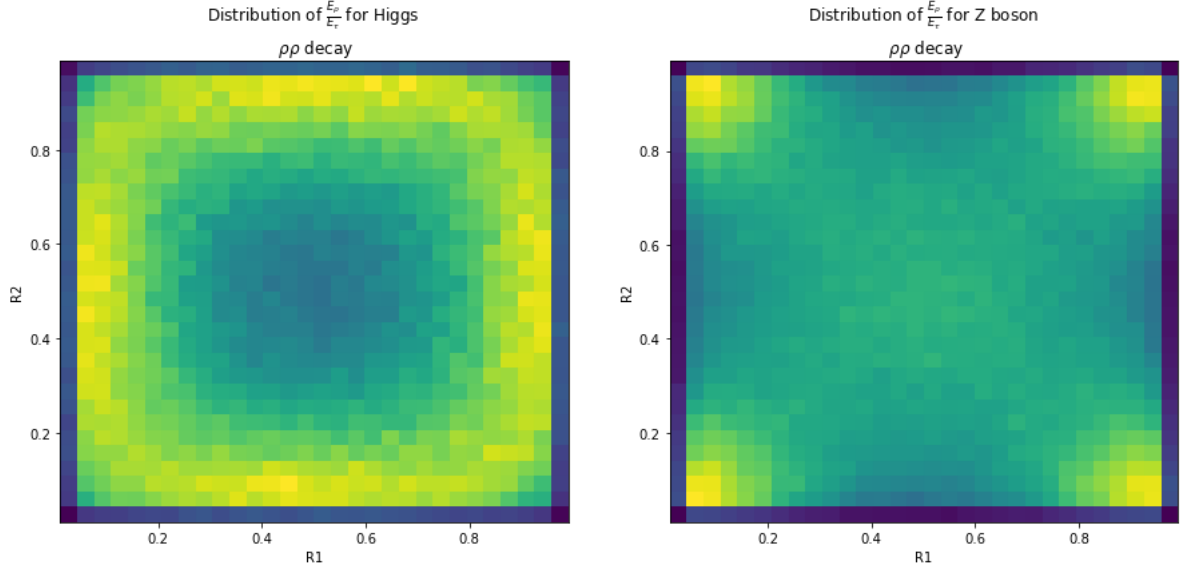


Figure 7: Distribution energy fraction of Higgs boson and Z boson

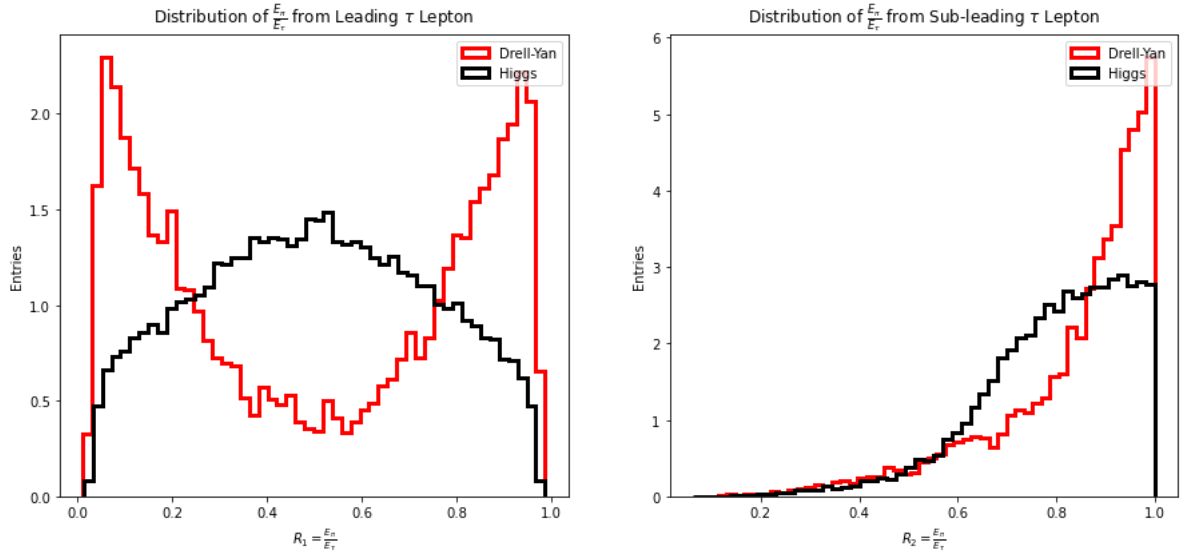


Figure 8: Distribution energy fraction of Higgs boson and Z boson

For decay mode a1, it might be better to use the polarimetric vector. Energy fraction can tell some information about the τ spin, but if the polarimetric vector can be constructed entirely, it will imply maximal information for the decay. However, the formula to reconstruct it might be very complicated in some cases. This will be discussed later in the subsection for DNN regressor.

3.1.2 Higgs with 95 GeV mass

This analysis repeats the process of the previous subsection when Higgs mass is 125 GeV. The results for final decay configuration $\pi\pi$ decay and $\rho\rho$ decay are similar to the previous case. For clarity of the study, distributions for these two decay configurations are presented in Appendix B with other repeated contents.

The distributions of energy fraction when the leading τ lepton decays into ρ and the sub-leading τ lepton undergoes a π decay is slightly different from the previous case. This is presented in Figure 9 below. The cut applied in this case is the same as the one in the previous

case.

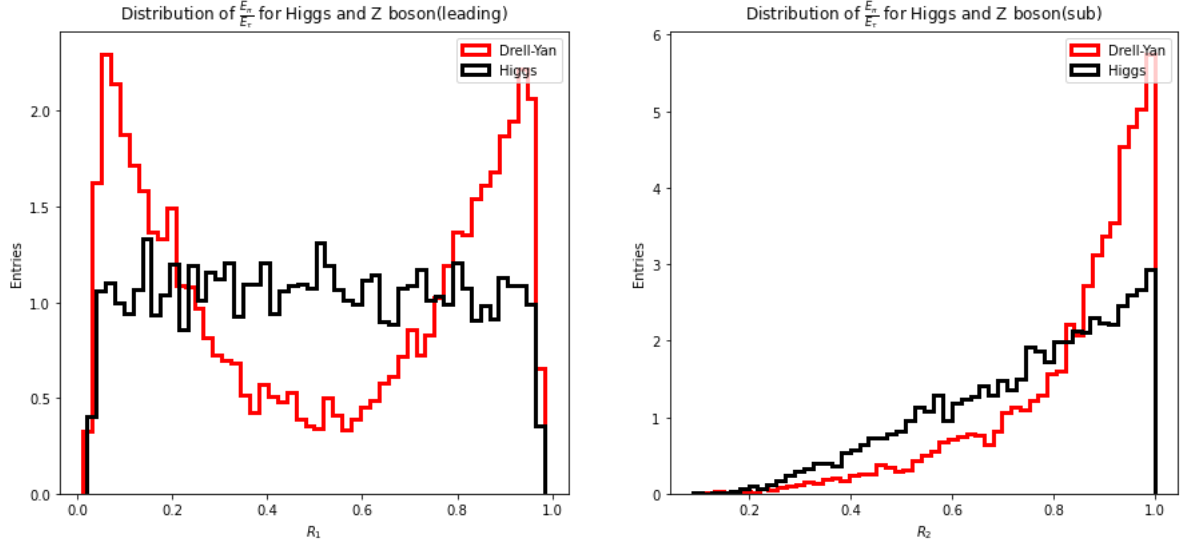


Figure 9: Distribution energy fraction of Higgs boson and Z boson

Figure 9 clearly shows that the distribution for the Higgs boson in this case is flatter than the one for Higgs with 125 GeV mass. This is because the change in energy fraction will be less evident for smaller masses.

Some other simple variables are also helpful: for τ lepton decay products, the angular information with how two taus are distributed (defined as $C_{1,2}$ in this case), the definition can be found in equation (10). It is similar to something we boosted into the Z rest frame, then took the angles between one of the τ lepton and the other. This will also create separation without using the information of τ lepton decay products. The distribution for this simple variable is shown in Figure 10:

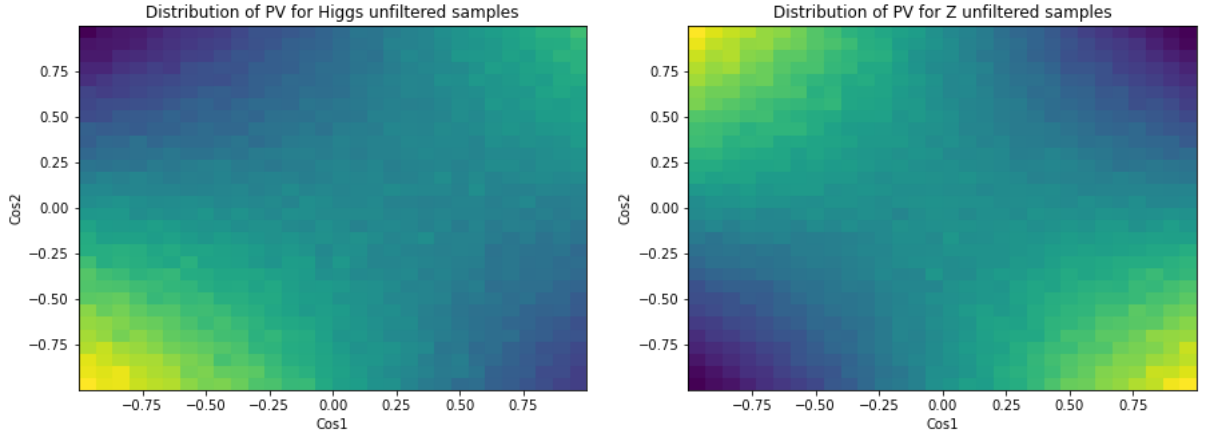


Figure 10: Distribution of $C_{1,2}$ for Higgs boson and Z boson

However, even if we use these variables to separate the Higgs from Z or non-Higgs, the sample separated out, or at least the properties we used for Higgs might not be identical as the case for the pure Higgs state. This means that the separation using this method introduces bias to the analysis. Therefore, more work should be done around ϕ_{CP} to check if the distribution of the ϕ_{CP} is the same in the selected sample as in the pure sample.

3.2 Linear regression for Higgs momentum

The quantities above are more accessible to reconstruct than angular information, which is more difficult to define, let alone the neutrino information. Regressing the Higgs boson momentum and its components (in Cartesian coordinate) can be the first step. This step can check whether the DNN setup can learn helpful information from the lowest level of variables. The inputs for this model are similar to the final DNN regressor while the outputs are only the momentum for Higgs and its components at the generator level. The baseline model architecture is set up as shown in Appendix B because it has a similar structure as the final DNN regressor but with a less customized setup.

3.2.1 Loss curves for training and testing data

Plotting the loss for each epoch during the training process and comparing it to a testing data set is a practical and basic way to evaluate the model's performance. The shape of curves can represent the goodness of choice for learning rate. A plot of different loss curves is shown in Figure 11 to address this property [37]:

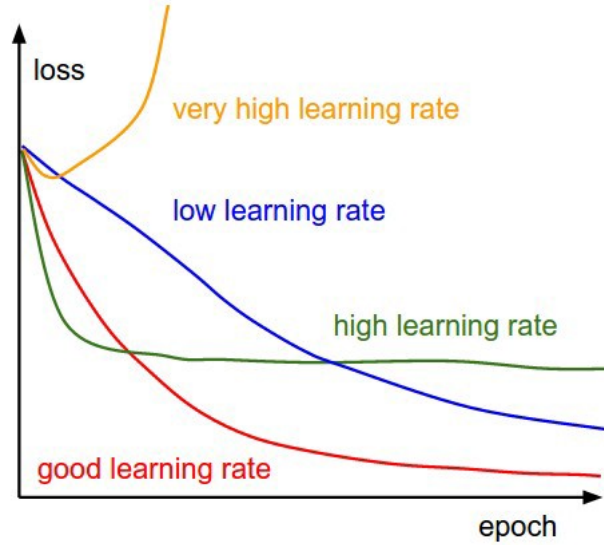


Figure 11: Loss curves for models with different learning rates

Apart from this property, comparing the curves for the training and testing sets is also informative. For example, the testing curve starts to go up after a certain point while the training curve continues to go down. This might imply the model is undergoing over-training [38].

With this background information, Figure 12 shows the learning curves for the baseline model used to regress the Higgs momentum and its components.

As it is clearly shown in the plot, the testing curve is closely in line with the training curve. This means the model makes predictions under the excellent condition without over-fitting or under-fitting problems.

3.2.2 Linear regression for Higgs momentum and its components in 2D distribution

After evaluating the model's performance, the next step is presenting the result with true values as this is the ultimate goal of making the predictions. The prediction made by this baseline model is presented in Figure 13:

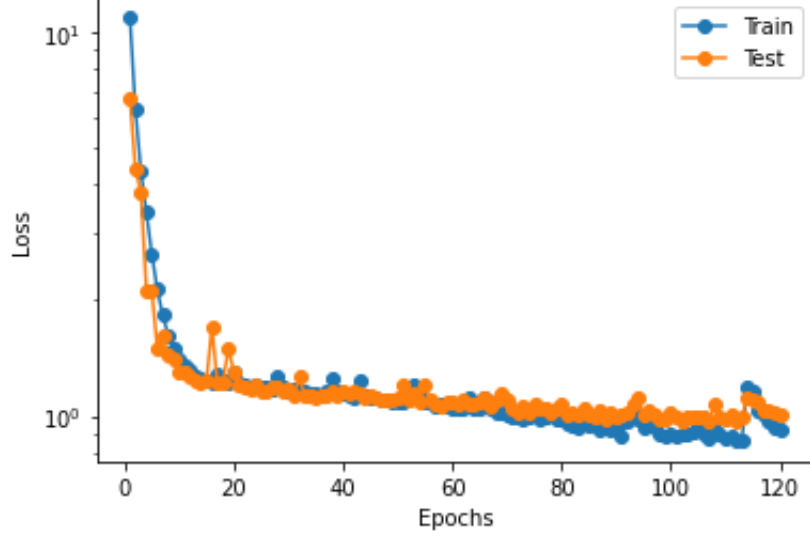


Figure 12: Loss curves for baseline model for training and testing data

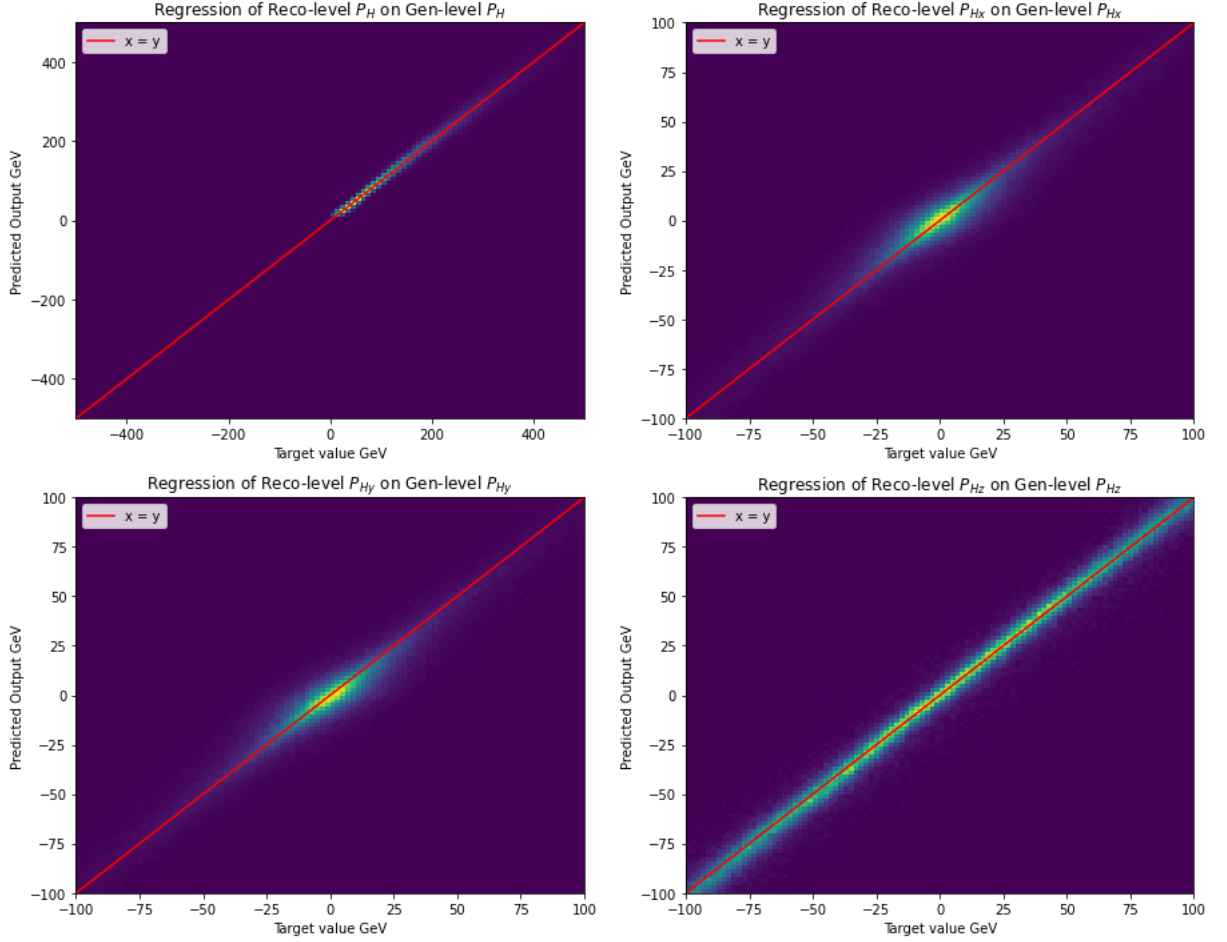


Figure 13: Linear regression plots for Higgs momentum and its three components in Cartesian coordinate. The 2D distribution is plotted as the predicted value against the true value at the generator level.

As shown in the plots, both the total momentum and the momentum components in the x,y and z directions demonstrate the linear regression on the generator level values. Here, the total momentum is the magnitude of the components so it only has value along the positive axis.

Compared with the components in the x and y direction, the momentum component in the z direction has the best prediction, while the components in the x and y direction are biased towards the origin.

3.3 Linear regression for neutrino momentum using multi-output DNN regressor

After testing the baseline model with Higgs momentum regression, regression for neutrino momentum components can be achieved in similar methods. Apart from the analysis from the previous subsection, the predicted values are also compared to SV reconstructed values (referred to as SV-fit later) as part of the evaluation. SV-fit values here are calculated with the reconstructed level visible information and τ kinematics reconstructed by Secondary Vertex [19]. The baseline model is also adjusted to the DNN regressor with more helpful input information. More details for this model are listed in the Methods section and appendix.

3.3.1 Evaluation of predicted neutrino momentum components

The first step to evaluating the prediction is still plotting the 2D distributions for predicted values against true values. The distributions for the x momentum component is presented as followed in Figure 14: a similar set of 2D distribution plots are made for neutrino momentum components in Cartesian coordinate for both leading τ lepton and sub-leading τ lepton. Only the results of the momentum component in the x direction are presented in this subsection because the other two components demonstrate a similar pattern. Therefore, the rest of the results will be included in Appendix B. are follows

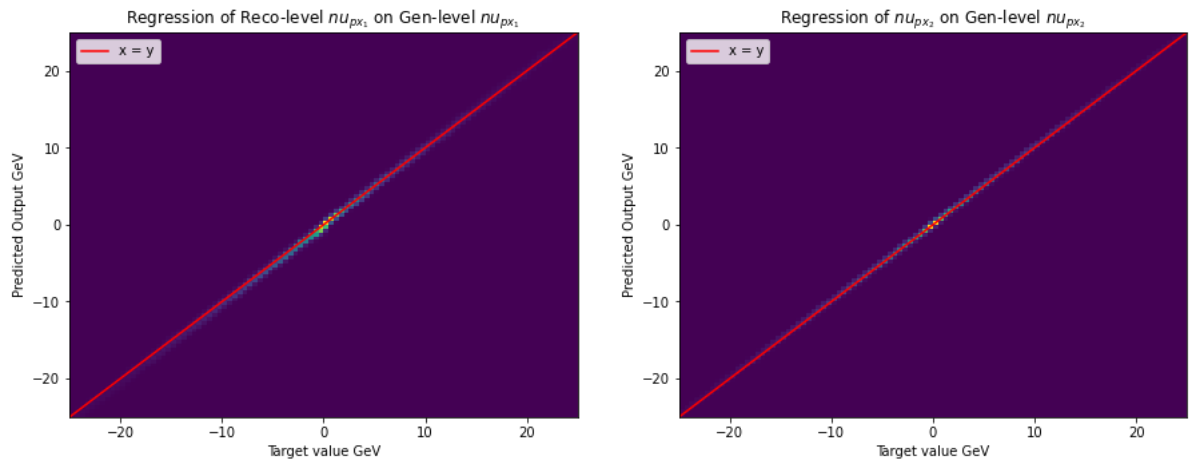


Figure 14: Linear regression plots for neutrino momentum x-component in Cartesian coordinate. The 2D distribution is plotted as the predicted value against the true value at the generator level.

As it is shown in Figure 14, the 2D distributions for neutrino momentum components are regressed better in all directions than in the baseline model. Therefore, the next evaluation step can be done using SV-fit values as a comparison. The computation of neutrino momentum components using SV-fit information can be explained by the equation below:

$$SV_{neutrino} = SV_{tau} - reco_{visible} \quad (13)$$

Figure 15 shows the difference distribution between predicted and true values. The true values for both differences are generator-level information. The comparison is made for the prediction

obtained by the DNN regressor and the prediction made by SV-fit using equation(13). The rest of the distributions are included in Appendix B for similar reasons to the 2D distribution plots.

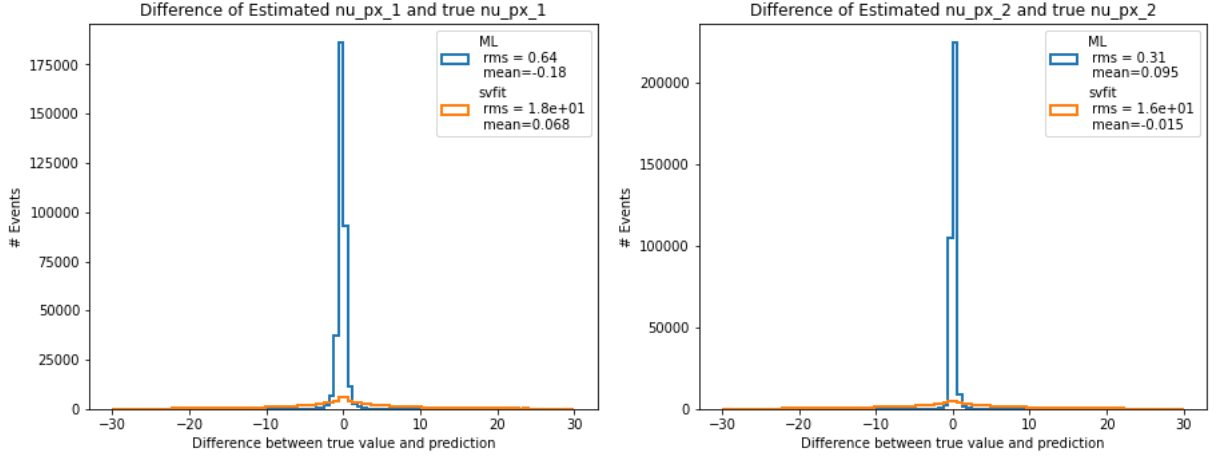
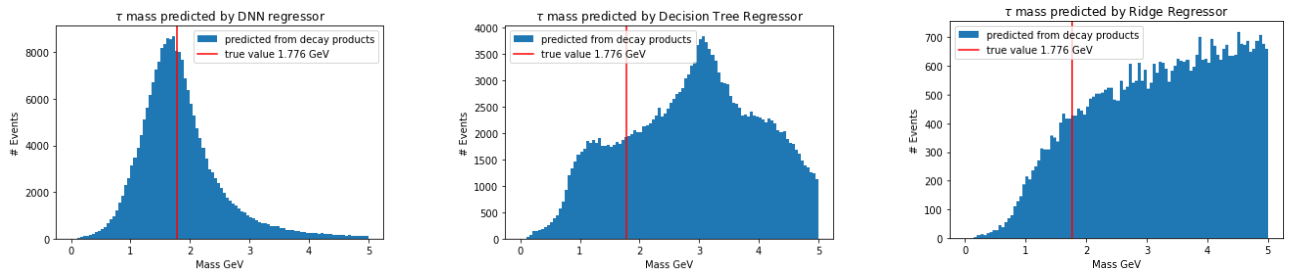


Figure 15: 1D distribution of the difference between true values and predicted values for neutrino momentum x-component in Cartesian coordinate. The true values are imported from the generator level and the predictions are made from the DNN model and calculated by SV-fit values.

It is clearly shown that the distribution peaks at 0. The peaks for the distributions for DNN regressor predictions are sharper than the SV-fit results. The root means square errors are also used to represent the width of the distributions. The distribution widths for DNN regressor results are smaller than the ones for SV-fit results by 2 orders of magnitude.

3.3.2 Comparison between multi-output DNN regressor and basic ML models

Comparing the reconstructed variables from the predictions is closer to the ultimate goal to analyze the predictions further. One simple variable that can be used to evaluate the performance is the mass of τ lepton. τ lepton mass is known to be around 1.776 GeV [39]. Therefore, comparing the peak of each prediction distribution can be an effective way to compare the performance of each model.



(a) A plot of τ mass computed from the predictions made by multi-output DNN regressor

(b) A plot of τ mass computed from the predictions made by Decision Tree regressor

(c) A plot of τ mass computed from the predictions made by Ridge regressor

Figure 16: Comparison of distributions of reconstructed τ mass computed from different models

The result distribution for the DNN regressor is the sharpest and peaks at the true value. This means that the DNN regressor gives the best predictions so that the reconstructed τ lepton mass peaks at 1.776 GeV, while the predictions made by the other two basic ML models failed in reconstructing this variable.

3.4 Reconstruction of acoplanarity angle ϕ_{CP}

All the evaluation above is to confirm the validity of the predictions before reconstructing the CP-sensitive variable ϕ_{CP} . The distribution of ϕ_{CP} is plotted with generator information and compared with the ϕ_{CP} computed from predictions made by the DNN regressor.

With the neutrino kinematics predicted in the previous section, the reconstruction of ϕ_{CP} and the polarimetric vectors for the decay products can be done following the process in the Methods section. The final possible decay configurations of interests are:

1. $\pi\pi$ decay: Both τ leptons decay into single charged pion π^\pm .
2. $\rho\rho$ decay: Both τ leptons decay into a charged pion and a neutral pion π^0 .
3. $\pi\rho$ decay: Leading τ lepton decays into a single charged pion π^\pm while sub-leading τ lepton decays into a charged pion and a neutral pion π^0 .
4. $\rho\pi$ decay: Leading τ lepton decays into a charged pion and a neutral pion π^0 while sub-leading τ lepton decays into a single charged pion π^\pm .

Figure 17 shows the ϕ_{CP} distribution when both τ leptons undergo π decays. The distributions are computed by impact parameter methods, and one is reconstructed by the polarimetric vector method with DNN regressor predictions.

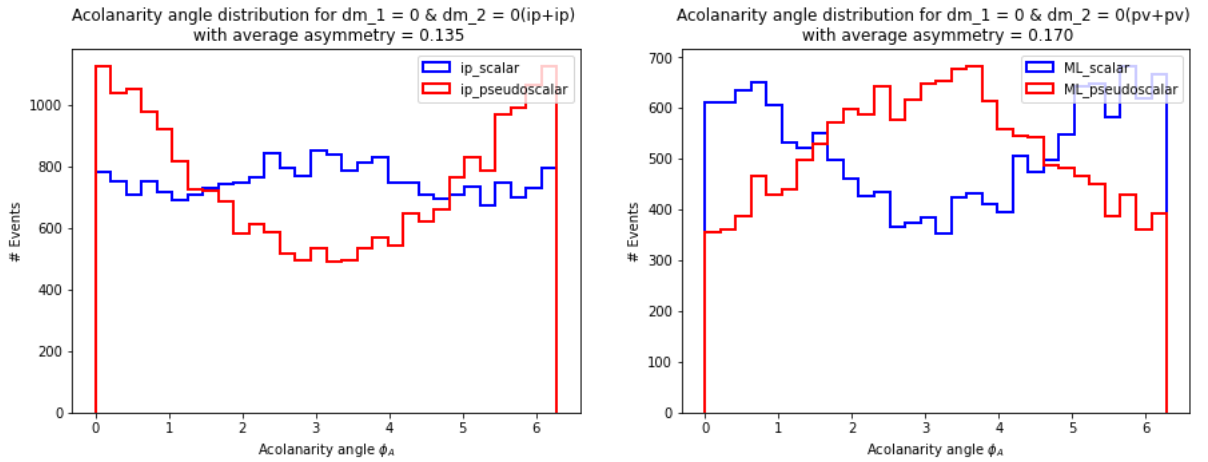


Figure 17: ϕ_{CP} distribution for scalar and pseudo-scalar reconstructed by impact parameter methods and polarimetric vector method with the prediction from DNN regressor.

In Figure 17, data of $\pi\pi$ is displayed. Events are included from all MVA score bins in these signal categories. The average asymmetry between the scalar and pseudo-scalar distributions. The definition of the average asymmetry per bin is $\frac{|CP_{even} - CP_{odd}|}{|CP_{even} + CP_{odd}|}$ and is normalised to the total number of bins. In this equation, CP_{even} and CP_{odd} are the scalar and pseudo-scalar contributions per bin [19].

However, the polarimetric vector methods for the rest of the decay modes configuration are not as promising as the original method. For ρ decays, the simple method to reconstruct ϕ_{CP} is the neutral-pion method. The distribution in this scenario is displayed in Figure 18. Similar results for reconstructed ϕ_{CP} distribution using the polarimetric vector method for $\pi\rho$ and $\rho\pi$ decays are included in Appendix B.

The average asymmetry of these two distributions from the polarimetric method is not as good as the case when both ϕ_{CP} are reconstructed by the neutral-pion method. This implies

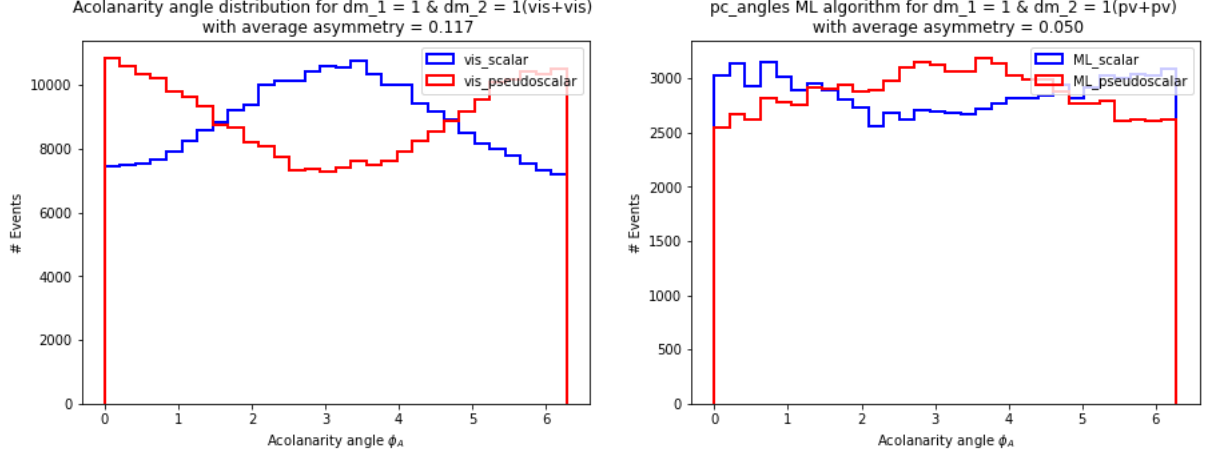


Figure 18: ϕ_{CP} distribution for scalar and pseudo-scalar reconstructed by neutral-pion methods and polarimetric vector method with the prediction from DNN regressor.

that a mixed method using neutral-pion methods for ρ decay and polarimetric vector methods for π decays can be applied to the other two configurations.

However, the ϕ_{CP} reconstructed by repeating the neutral-pion method using the reconstructed level visible decay products kinematics are not giving valid results. This might be caused by mistakes in the code. It can be corrected by plotting the same set of distributions at the generator level. This verifying process is displayed in Figure 19:

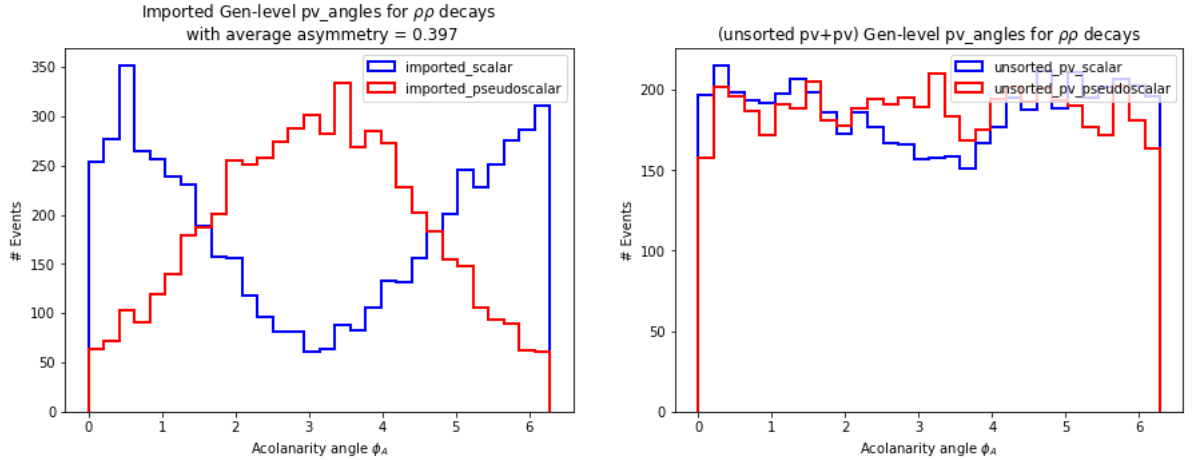


Figure 19: ϕ_{CP} distribution for scalar and pseudo-scalar reconstructed by neutral-pion methods using the information at generator level.

In Figure 19, the distribution of ϕ_{CP} on the left is plotted by sorting the value of ϕ_{CP} and applying weights sorted by the value of the acoplanarity angle computed by the neutral-pion method using generator level information. The one on the right is the distribution of ϕ_{CP} without this process. This is to check whether the algorithm used to compute the neutral-pion vectors will change the ϕ_{CP} order so that each ϕ_{CP} does not belong to the actual event it comes from. By sorting both the ϕ_{CP} values and weights in ascending order, it will reorganise the ϕ_{CP} so that each value corresponds to the event it comes from. Comparison between the sorted distributions and the unsorted one implies that having the wrong ϕ_{CP} for each event is one of the problems causing the failure of separation in this case.

4 Conclusion

According to the first part of the study, separating Higgs signals from Drell-Yan background signals with energy fractions or the $C_{1,2}$ is possible at the generator level. A similar conclusion can be drawn at the reconstructed level since we have successfully predicted the kinematics information of neutrinos. Once the separation is made by classification, it is also essential to check whether the ϕ_{CP} reconstructed from the separated Higgs signal is the same as the one reconstructed from the pure Higgs state.

The ML techniques were first applied to this field because it was cumbersome to evaluate data for too many distinct acoplanarity angle sensitivities, then evolved as the backbone in identifying decay modes. It started from the application of binary classification and was developed into a multi-class neural network tool. Implementing ML techniques not only simplified the analysis but also reduced the uncertainties of the results.

The multi-output DNN regressor implementation in this study shows promising potential for improving the reconstruction of CP-sensitive observables, especially the acoplanarity angles ϕ_{CP} . By customizing the loss function, the performance of the regressor was improved so that the average asymmetry of the CP_{even} (scalar) and CP_{odd} (pseudo-scalar) distributions for $\pi\pi$ decays improved from the 0.135 using the impact parameter method to 0.170 using the polarimetric method with the prediction made by the regressor.

Nevertheless, this only works well for $\pi\pi$ decays. The results of other possible configurations are still far from satisfactory. Improvements can be made by checking the algorithm computing the polarimetric vectors and ϕ_{CP} because of the problem stated in the Results section.

4.1 Possible application of GNN architecture

Apart from the potential mistake in the code, the DNN regressor can also be improved from a simple basic neural network structure to a more complicated model with more layers and appropriate structure.

Here, implementation of Graph Neural Network structure might be a possible way to go. A GNN is an optimizable transformation on all attributes of the graph (nodes, edges, global-context) that preserves graph symmetries (permutation invariance) [40]. The benefit of implementing a GNN is that it does not change the potential linkage between the inputs. The progressive transformation will be acted on the embeddings such as the information of nodes, edges and global context without changing the connectivity of the input graph.

Possible way of implementation is listed below:

1. Node feature: decay products kinematics(four-momenta, missing energy), impact parameter and secondary vertex information.
2. Edge features a one-hot encoded array showing the type of decay mode.
3. Label for regression: same output targets stated in this study earlier in the Methods section.

4.2 Main achievements

So far, most of the previous studies are predominantly focused on basic applications of machine learning and are far from optimization [7, 18, 19]. Therefore, the research gap in this field deserves to be discussed to optimize the applications and the measurements. This study critically

evaluates the development of the machine learning method applied in measuring the CP state of the Higgs boson at least in a bid to address the gap.

Overall, previous studies highlight the need to develop machine learning techniques in the analysis of the CP structure of the Higgs boson. Considering the evidence mentioned in the study, it was evident that developing corresponding machine learning techniques can bring significant improvements in measurements and extraction of sensitivities with more amendment if time allows.

However, due to the time limitation, some of the work in this study remains uncompleted. A more systematic and sufficient analysis is required for the training of the algorithm so that it can extract more information contained in the existing data. This is one of the toughest challenges for all researchers in this domain. Implementing a new model such as GNN might therefore be helpful to improve the performance of the results in the analysis of the CP-sensitive observables.

5 Acknowledgements

I would like to thank my supervisors, Prof. David Colling and Daniel Winterbottom for guiding me through this MRes project. I also appreciate the time I spent with the HEP team. The meetings and conversations were vital in inspiring me to think outside the box from multiple perspectives to form a comprehensive and objective critique. I have been feeling supported the whole year and enjoy working with all my colleagues.

References

- [1] K. Lasocha, E. Richter-Was, D. Tracz, Z. Was, and P. Winkowska. Machine learning classification: Case of higgs boson cp state in $h \rightarrow \tau\tau$ decay at the lhc. *Phys. Rev. D*, 100:113001, Dec 2019.
- [2] S. Chatrchyan, V. Khachatryan, A.M. Sirunyan, and etc. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, 2012.
- [3] S. Chatrchyan, V. Khachatryan, A.M. Sirunyan, and etc. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, sep 2012.
- [4] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, R. Aben, and etc. ”combined measurement of the higgs boson mass in pp collisions ats=7 and 8 tev with the atlas and cms experiments”. *Physical Review Letters*, 114(19), 5 2015.
- [5] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, and etc. Constraints on the spin-parity and anomalous hvv couplings of the higgs boson in proton collisions at 7 and 8 tev. *Phys. Rev. D*, 92:012004, 7 2015.
- [6] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, and etc. Evidence for the spin-0 nature of the higgs boson using atlas data. *Physics Letters B*, 726(1):120–144, 2013.
- [7] R. Józefowicz, E. Richter-Was, and Z. Was. Potential for optimizing the Higgs boson C P measurement in $H \rightarrow \tau\tau$ decays at the LHC including machine learning techniques. , 94(9):093001, November 2016.
- [8] M. Krämer, J. Kühn, M. L. Stong, and P. M. Zerwas. Prospects of measuring the parity of higgs particles. *Zeitschrift für Physik C Particles and Fields*, 64(1):21–30, 1994.
- [9] Stefan Berge and Werner Bernreuther. Determining the cp parity of higgs bosons at the lhc in the τ to 1-prong decay channels. *Physics Letters B*, 671(4):470–476, 2009.
- [10] G.R. Bower, T. Pierzchala, Z. Was, and M. Worek. Measuring the higgs boson’s parity using $\tau \rightarrow \rho\nu$. *Physics Letters B*, 543(3):227–234, 2002.
- [11] Osamu Seto. Large invisible decay of a higgs boson to neutrinos. *Phys. Rev. D*, 92:073005, Oct 2015.
- [12] S. Berge, W. Bernreuther, B. Niepelt, and H. Spiesberger. How to pin down the cp quantum numbers of a higgs boson in its τ decays at the lhc. *Phys. Rev. D*, 84:116003, Dec 2011.
- [13] Stefan Berge, Werner Bernreuther, and Hubert Spiesberger. Higgs cp properties using the τ decay modes at the ilc. *Physics Letters B*, 727(4):488–495, 2013.
- [14] Stefan Berge, Werner Bernreuther, and Sebastian Kirchner. Determination of the higgs cp -mixing angle in the tau decay channels at the lhc including the drell–yan background. *The European Physical Journal C*, 74(11), 2014.
- [15] Stefan Berge, Werner Bernreuther, and Sebastian Kirchner. Prospects of constraining the higgs boson’s cp nature in the tau decay channel at the lhc. *Phys. Rev. D*, 92:096012, Nov 2015.
- [16] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

- [17] E. Barberio, B. Le, E. Richter-Was, Z. Was, and D. Zanzi. Optimizing higgs boson cp measurement in $h \rightarrow \tau\tau$ decay with ml techniques. *Acta Physica Polonica B*, 48(6):1049, 2017.
- [18] K. Lasocha, E. Richter-Was, M. Sadowski, and Z. Was. Deep neural network application: Higgs boson CP state mixing angle in $H \rightarrow \tau\tau$ decay and at the LHC. , 103(3):036003, February 2021.
- [19] A. Tumasyan, W. Adam, J. W. Andrejkovic, and etc. Analysis of the CP structure of the yukawa coupling between the higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2022(6), 6 2022.
- [20] P. Baldi, P. Sadowski, and D. Whiteson. Searching for exotic particles in high-energy physics with deep learning. *Nature Communications*, 5(1), 2014.
- [21] P. Baldi, P. Sadowski, and D. Whiteson. Enhanced higgs boson to $\tau^+\tau^-$ search with deep learning. *Phys. Rev. Lett.*, 114:111801, Mar 2015.
- [22] Vladimir Cherepanov, Elzbieta Richter-Was, and Zbigniew Was. Monte Carlo, fitting and Machine Learning for Tau leptons. *SciPost Phys. Proc.*, page 018, 2019.
- [23] Stefan Berge and Werner Bernreuther. Determining the cp parity of higgs bosons at the lhc in the τ to 1-prong decay channels. *Physics Letters B*, 671(4):470–476, 2009.
- [24] André Rougé. Cp violation in a light higgs boson decay from τ -spin correlations at a linear collider. *Physics Letters B*, 619(1):43–49, 2005.
- [25] Paolo Nason. A new method for combining NLO QCD with shower monte carlo algorithms. *Journal of High Energy Physics*, 2004(11):040–040, nov 2004.
- [26] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *Journal of High Energy Physics*, 2007(11):070–070, nov 2007.
- [27] T. Przedziński, Richter-Was E., and Was Zbigniew. Tauspinner: A tool for simulating cp effects in $h \rightarrow \tau\tau$ $h \rightarrow \tau\tau$ decays at lhc. *The European Physical Journal C*, 74(11), 2014.
- [28] S. Agostinelli, J. Allison, K. Amako, and etc. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.
- [29] S Moretti and D.P Roy. The τ -polarisation test for the $h \rightarrow \tau\tau$ signal from the drell-yan background signal at the lhc. *Physics Letters B*, 545(3-4):329–336, 2002.
- [30] CMS Collaboration. Track impact parameter resolution for the full pseudo rapidity coverage in the 2017 dataset with the CMS Phase-1 Pixel detector. *CERN*, 2020.
- [31] Malgorzata Worek. Higgs cp from h/a -gt; tau tau decay. *arXiv*, 2003.
- [32] Joseph R. Dell Aquila and Charles A. Nelson. Distinguishing a spin-0 technipion and an elementary higgs boson. *Phys. Rev. D*, 33:93–100, Jan 1986.
- [33] Stefan Berge, Werner Bernreuther, and Sebastian Kirchner. Determination of the higgs cp-mixing angle in the tau decay channels at the lhc including the drell-yan background. *The European Physical Journal C*, 74(11), 2014.
- [34] Scikit-Hep. Scikit-hep/vector: Vector classes and utilities.

- [35] Ph.D. Ayyüce Kızrak. Comparison of activation functions for deep neural networks, Jan 2020.
- [36] James Dellinger. Weight initialization in neural networks: A journey from the basics to kaiming, Apr 2019.
- [37] Andrej Karpathy. convolutional neural networks for visual recognition, 2020.
- [38] A Baeldung. Learning curves in machine learning, Jul 2020.
- [39] R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.
- [40] Benjamin Sanchez-Lengeling, Emily Reif, Adam Pearce, and Alexander B. Wiltschko. A gentle introduction to graph neural networks, Sep 2021.

Appendices

A Important code

All the code will be uploaded to a GitHub repository and the link will be put at the end of the appendix.

```
# define new loss function
import keras.backend as K

def custom_loss(y_true, y_pred):
    """
    This function is defined for the Multi-output regressor.
    Inputs:
        y_true: Target value in training test/testing set
        y_pred: Prediction made by the DNN regressor
    Outputs:
        loss: loss calculated by math inside this function
        for each group of y_pred and y_true.
    """

    loss = K.square(y_pred - y_true)

    # multiplying the values with weights along batch dimension
    loss = loss * [0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.9, 0.9, 0.3, 0.3, 0.5, 0.5]

    # summing both loss values along batch dimension
    loss = K.sum(loss, axis=1)

    # calculate loss, using y_pred

    return loss
```

Figure 20: Code for customized loss function

the code is available at: https://github.com/ql1821/MLBD_01

B Repeated graphs

B.0.1 Higgs M95

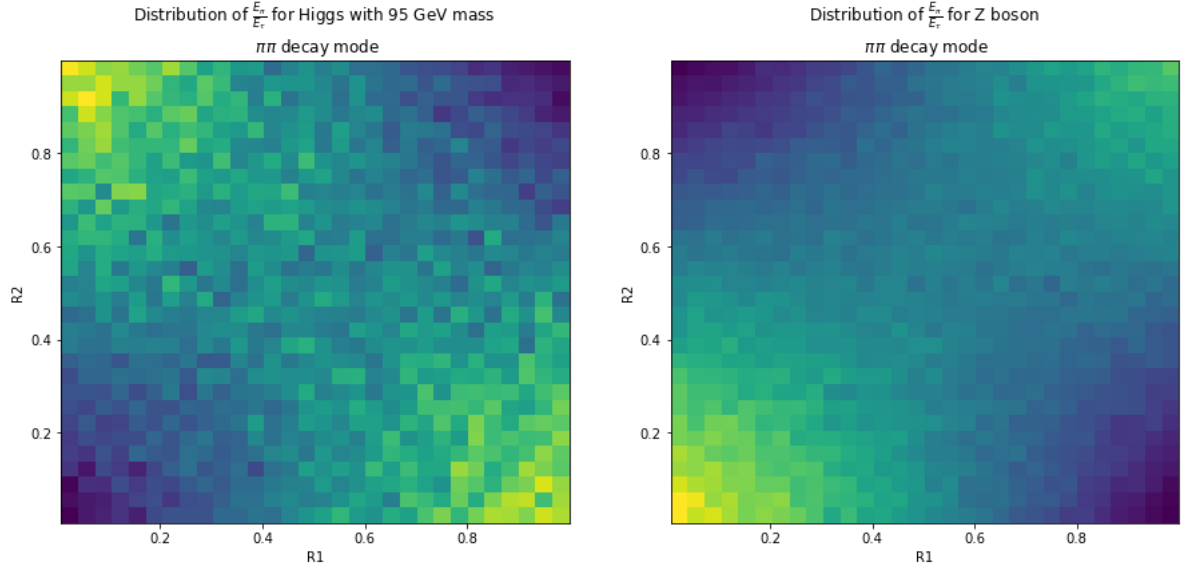


Figure 21: Distribution energy fraction of Higgs boson and Z boson

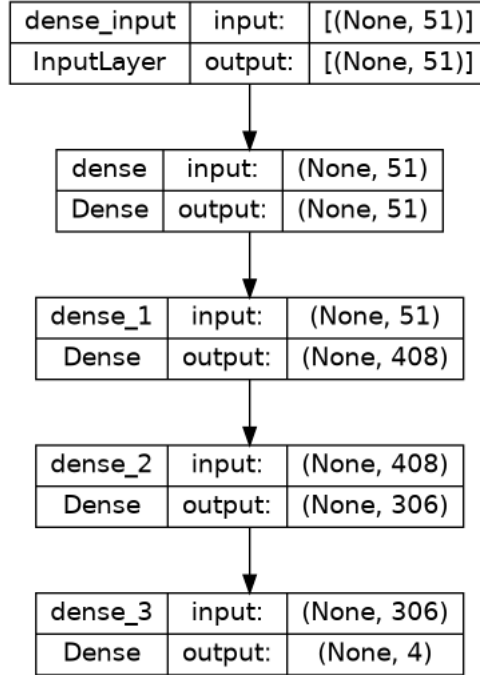


Figure 22: Flowchart of the baseline model used to predict Higgs momentum and its components

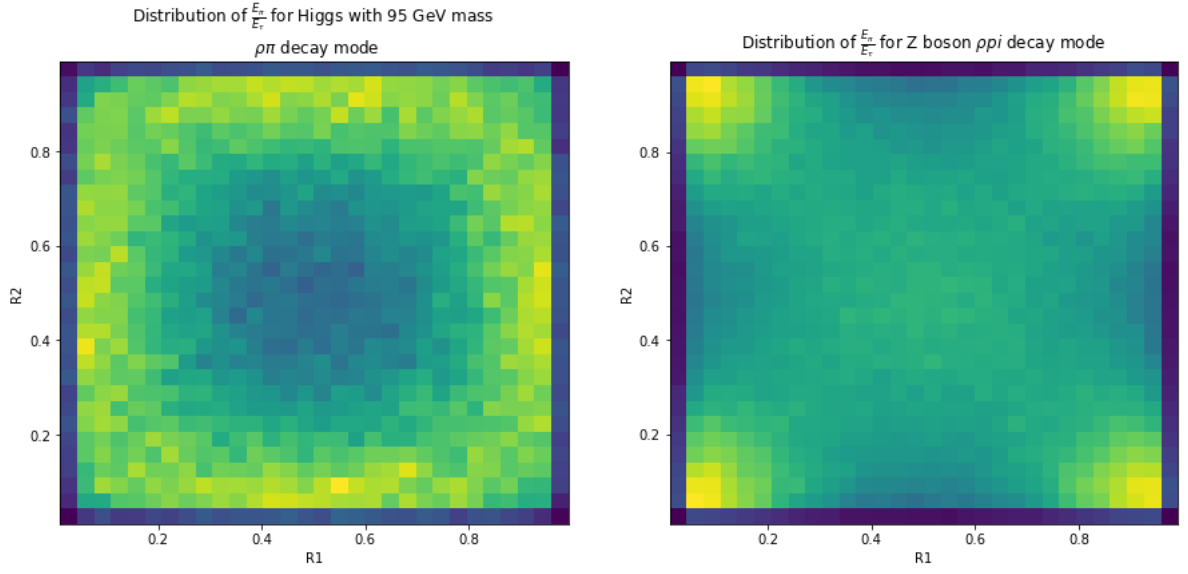


Figure 23: Distribution energy fraction of Higgs boson and Z boson

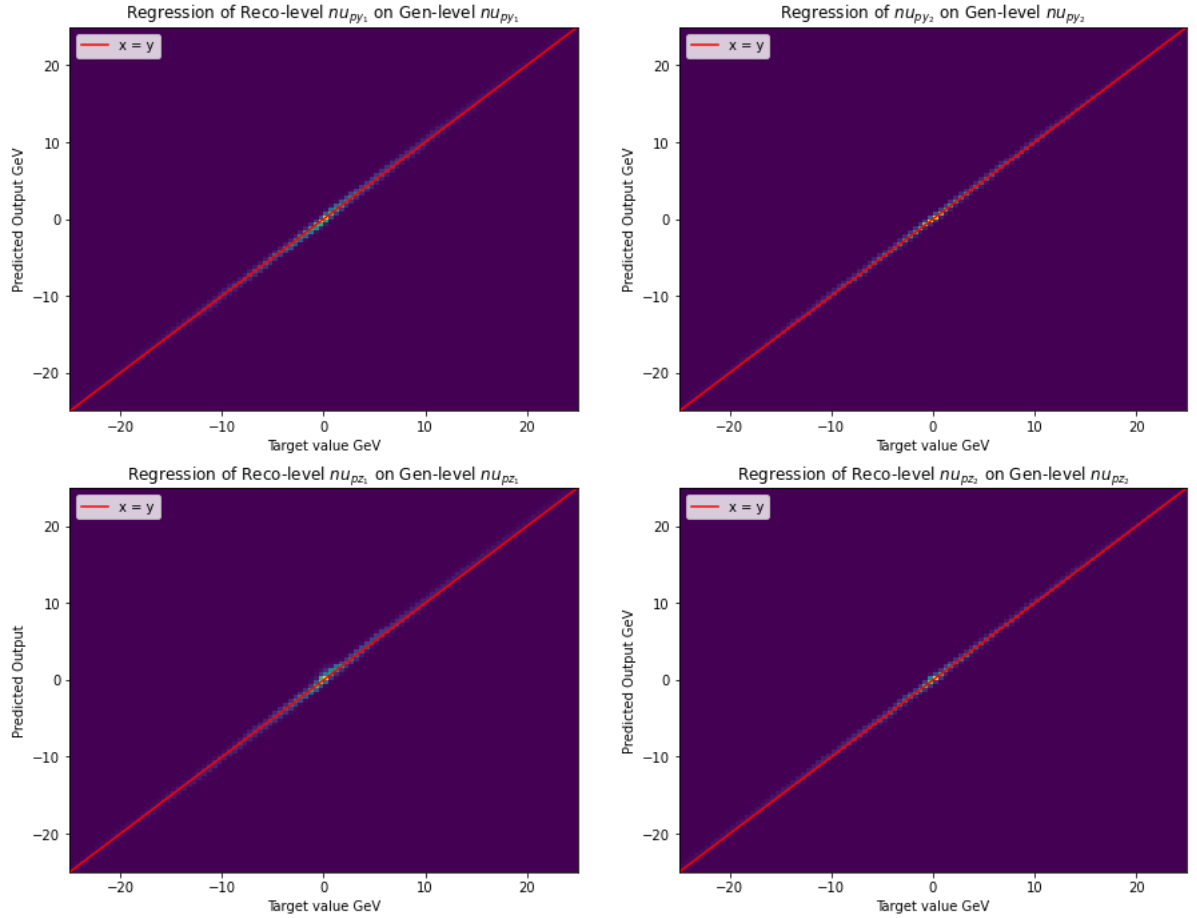


Figure 24: Linear regression plots for neutrino momentum y,z-component in Cartesian coordinate. The 2D distribution is plotted as the predicted value against the true value at the generator level.

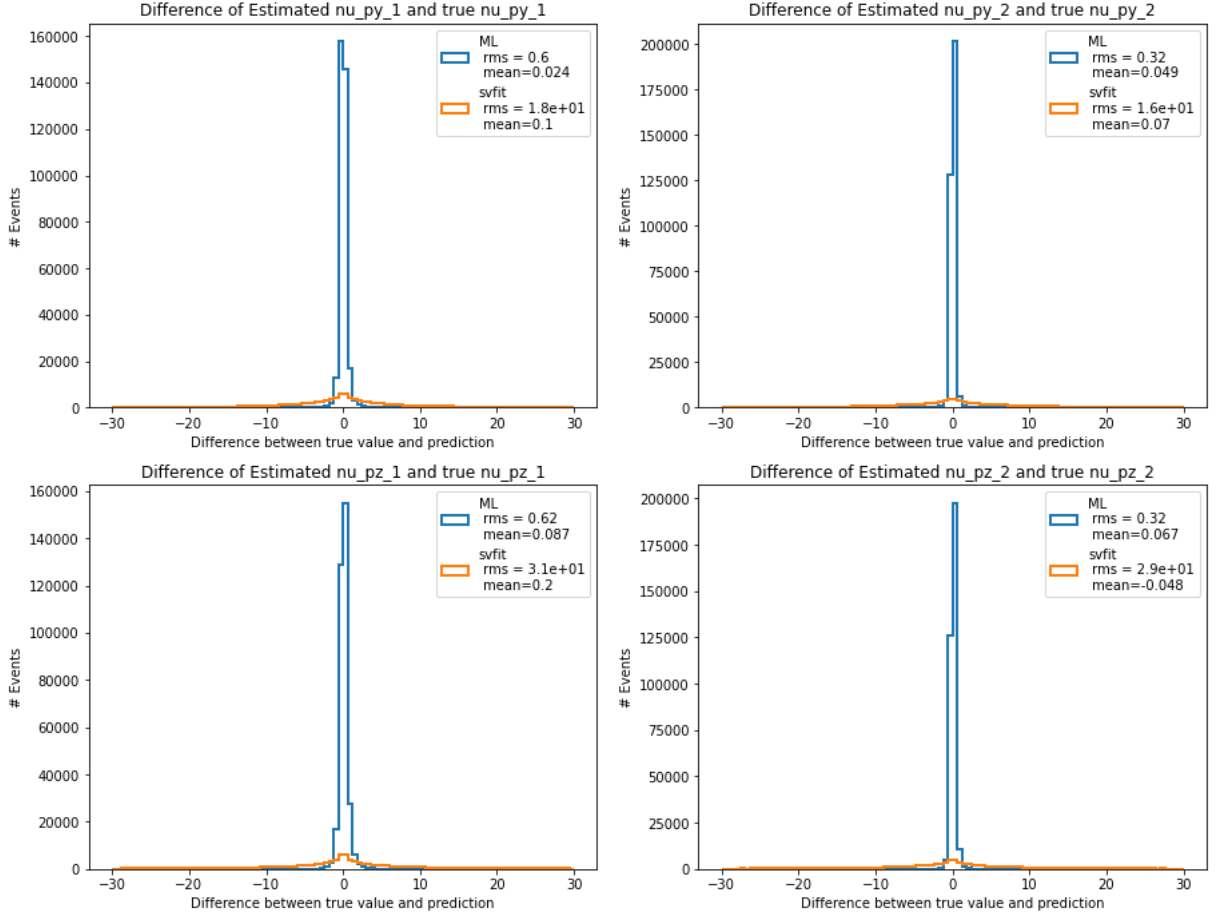


Figure 25: 1D distribution of the difference between true values and predicted values for neutrino momentum y,z-component in Cartesian coordinate. The true values are imported from the generator level and the predictions are made from the DNN model and calculated by SV-fit values.

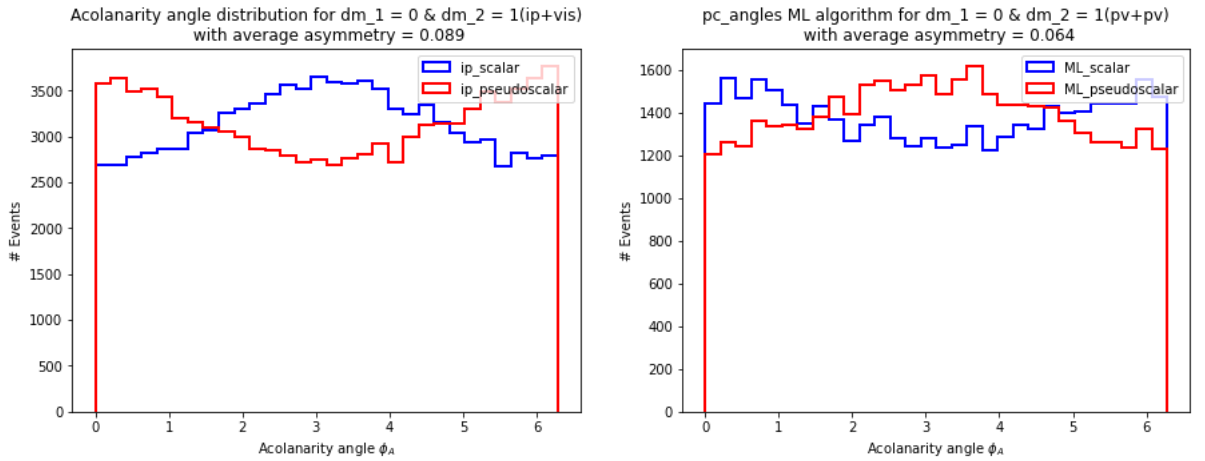


Figure 26: ϕ_{CP} distribution for scalar and pseudo-scalar reconstructed by impact parameter methods and polarimetric vector method with the prediction from DNN regressor.

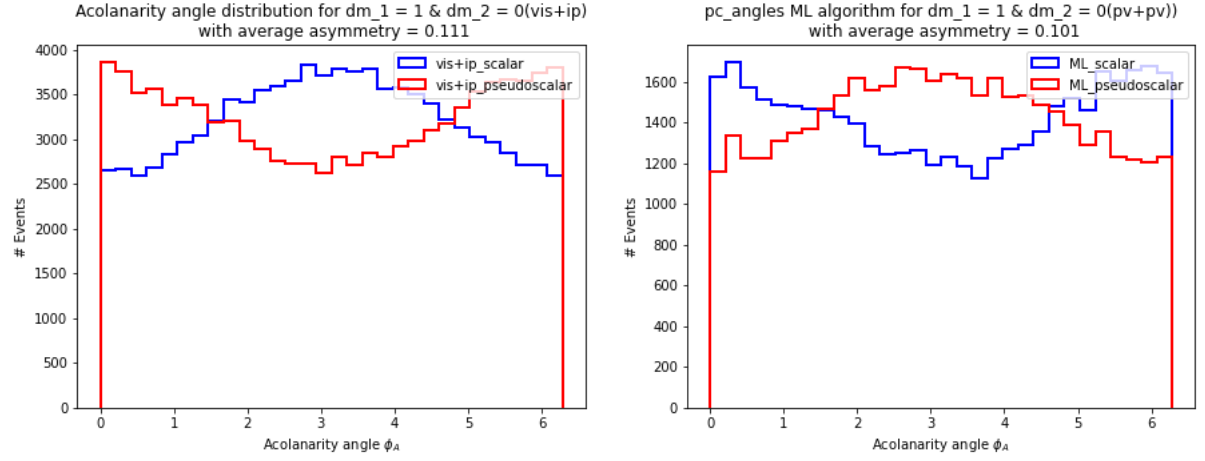


Figure 27: ϕ_{CP} distribution for scalar and pseudo-scalar reconstructed by impact parameter methods and polarimetric vector method with the prediction from DNN regressor.