

# Exploring the standard model EFT in $VH$ production with machine learning

Felipe F. Freitas,<sup>1,†</sup> Charanjit K. Khosa,<sup>2,\*</sup> and Verónica Sanz<sup>2,‡</sup>

<sup>1</sup>*CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics,  
Chinese Academy of Sciences, Beijing 100190, China*

<sup>2</sup>*Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, United Kingdom*



(Received 26 February 2019; published 30 August 2019)

In this paper we study the use of machine learning techniques to exploit kinematic information in  $VH$ , the production of a Higgs in association with a massive vector boson. We parametrize the effect of new physics in terms of the standard model effective field theory (SMEFT) framework. We find that the use of a shallow neural network allows us to dramatically increase the sensitivity to deviations in  $VH$  respect to previous estimates. We also discuss the relation between the usual measures of performance in machine learning, such as area under the curve or accuracy, with the more adept measure of Asimov significance. This relation is particularly relevant when parametrizing systematic uncertainties. Our results show the potential of incorporating machine learning techniques to the SMEFT studies using the current datasets.

DOI: [10.1103/PhysRevD.100.035040](https://doi.org/10.1103/PhysRevD.100.035040)

## I. INTRODUCTION

The particle physics community holds high hopes of discoveries in the Large Hadron Collider (LHC), the machine colliding protons at the highest energies in an Earth laboratory. Yet, after years of an intense effort searching for new phenomena, no clear evidence of new physics has been found.

To continue the search for new phenomena and improve the exploitation of the LHC data, we are shifting our focus from the low-hanging fruit, e.g., resonance searches, into more subtle (indirect) effects of new physics. A well-defined approach to develop the interpretation of data in terms of indirect probes is the framework of effective field theories [1], and in particular in the context of the standard model effective field theory (SMEFT) [2].

In a nutshell, the SMEFT is a consistent way of exploring new theories as deformations from the SM structures, with a large number of possible SM deviations taken into account.

As an example, in the SMEFT approach the Higgs couplings to vector bosons  $V = W, Z$  would be modified in the following way

$$\eta_{\mu\nu} g m_V \Rightarrow \eta_{\mu\nu} g m_V - \frac{2g c_{\text{HW}}}{m_W} p_\mu^V p_\nu^V + \dots \quad (1)$$

which in terms of Lagrangian terms would be equivalent to adding to the SM Lagrangian new terms suppressed by a scale of new physics

$$\mathcal{L}_{\text{SM}} \Rightarrow \mathcal{L}_{\text{SM}} + \frac{2ig c_{\text{HW}}}{m_W^2} [D^\mu H^\dagger T_{2k} D^\mu H] W_{\mu\nu}^k + \dots \quad (2)$$

where  $H$  is the Higgs  $SU(2)$  doublet and  $W^k$  is the electroweak gauge boson triplet.

One could trace the ultimate origin of these deformations to many different types of new physics, just too heavy to be discovered directly at the LHC. For example, the deformation (aka Wilson coefficient)  $c_{\text{HW}}$  could be the manifestation of a new set of scalar particles, such as in 2HDMs, too heavy or too complex to be seen in direct production, but still felt via virtual effects such as the one-loop contribution shown in Fig. 1 [3].

These new theories would then manifest themselves in the LHC environment as subtle deviations in physical observables, often in kinematic regions where the theoretical and experimental understanding is particularly poor. In contrast with a resonance search in a final state, SMEFT analyses forces us to deal with the LHC's inherently complex environment, where the understanding of extreme kinematic regions is required.

In the context of the SMEFT effects in the Higgs sector, the LHC analyses have moved from the basic use of total cross-sections ( $\kappa$  formalism [4]) to understand that pushing the boundaries of the SMEFT means using kinematic information [5,6]. Even that frontier is becoming a

\*charanjit.kaur@sussex.ac.uk

†fp4303@itp.ac.cn

‡V.Sanz@sussex.ac.uk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.

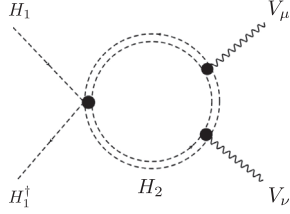


FIG. 1. An example of how operators like  $c_{\text{HW}}$  could arise from new theories.  $H_{1,2}$  denotes the light (heavy) Higgs.

well-trodden path with the Run 2 finished, and searches for new physics in SMEFT effects now moving towards identifying even subtler effects by looking at multidimensional information [7] and combining as many channels as possible [8].

This state of affairs, the need to quickly identify subtle effects in multidimensional distributions of information, clearly calls for artificial intelligence methods. Particularly the use of data mining techniques in machine learning [9]. The amount of information one single channel can provide is limited, though. Even in a complex final state such as vector boson fusion (VBF) and all the multidimensional correlations one can think of in this channel, the amount of information quickly saturates [7], just a manifestation that the kinematics of the final state particles (input information) satisfies a number of constraints (energy-momentum conservation, behaviour of parton distribution functions, experimental selection cuts and resolution), limiting the usefulness of single channels. In VBF, we showed the inherent limitations in a Bayesian context [7] and recently in Refs. [10] the authors pioneered the use of machine learning to identify SMEFT effects in VBF, including data augmentation. See also Ref. [11] for a detailed comparison of the information available in VBF,  $VH$ , and on  $H \rightarrow 4\ell$ .

Despite its importance to understand the electroweak sector, the measurement of Higgs production in VBF is not a reality yet, hence studies are based on future prospects. On the other hand, the production of the Higgs in association with a massive vector boson, or  $VH$ , is already firmly established [12,13]. As quality kinematic information in  $WH$  and  $ZH$ , better statistics and experimental understanding, will occur before VBF production is understood, we believe the approach of this paper would be the first step to push the boundaries of our understanding of SMEFT effects on Higgs LHC data, complemented later on with VBF information.

In this paper, we will illustrate the use of machine learning techniques in  $VH$  by switching on a single SMEFT effect on  $WH$  and  $ZH$ . In the past few months, we have witnessed an explosion works by the HEP community on the use of Machine Learning techniques, e.g., Refs. [14], and the analyses have quickly become more and more sophisticated. Although in this paper we use state-of-the-art techniques, we expect our results in  $VH$  will be surpassed by other works in the near future.

This paper is organized as follows. In Sec. II, we describe the current status of the SMEFT analyses and the experimental understanding of the  $VH$  channel. In Sec. III, we then move to describe the sort of kinematic information one could use in  $VH$ . The machine learning analysis, in particular the use of a shallow neural network is described in Sec. IV and Appendix A, where we provide a simple glossary of terms used in this paper. We present our results in Sec. V, and discuss possible new directions in Sec. VI.

## II. CURRENT STATUS: LIMITS ON THE SMEFT AND THE $VH$ AT THE LHC

We are going to illustrate the techniques using a particular deformation, the operator in Eq. (2) with Wilson coefficient  $c_{\text{HW}}$ . It is currently constrained to values in the range [8] (individual constraint)

$$c_{\text{HW}} = 0.002 \pm 0.014. \quad (3)$$

In this paper we will often illustrate points using a benchmark within the  $2\sigma$  region:

$$c_{\text{HW}} = 0.03. \quad (4)$$

The limits on SMEFT operators were obtained by performing a global fit including kinematic information on  $VH$  [15] and electroweak  $WW$  production at LEP2 and LHC [16] but only  $40 \text{ fb}^{-1}$  of data, half of the total Run2 dataset. A more recent global analysis was done by the groups in Refs. [17,18], but their analysis did not substantially change the limit on  $c_{\text{HW}}$ . On the other hand, sensitivity studies of future colliders such as HL-LHC show that these limits will be pushed to a few times smaller than the current limit [19].

On the experimental side, the ATLAS [12] and CMS [13] collaborations have marked yet another milestone in their quest to understand electroweak symmetry breaking: the observation of the Higgs decaying into two b-quarks. This measurement has been done by combining a challenging set of channels collectively denoted by  $VH$ , which corresponds to the Higgs produced in association with a massive vector boson  $V = Z$  or  $W^\pm$ . The final states are classified as 0L ( $Z \rightarrow \nu\bar{\nu}$ ), 1L ( $W \rightarrow \ell\nu$ ) and 2L ( $Z \rightarrow \ell^+\ell^-$ ). The combination of all the channels can be summarized as the ratio of the observed cross section by the SM expectation,  $\mu_{VH}$ . For example, the ATLAS measurement reads

$$\mu_{VH}(\text{ATLAS}) = 1.01 + 0.12(\text{stat})_{0.15}^{+0.6}(\text{syst}) \quad (5)$$

which, given the dependence of  $\mu$  with the parameter  $c_{\text{HW}}$  naively would indicate a two-sigma exclusion  $|c_{\text{HW}}| < 0.02$ .

### III. KINEMATIC INFORMATION IN $VH$

Right after the discovery of the Higgs boson in the summer of 2012, the  $VH$  channel was identified as an important source of information to search for anomalous behaviour of the Higgs. In particular, the distribution of transverse momentum of the vector boson,  $p_T^V$  was identified as very sensitive to new physics, even to the point of reviving TeVatron searches which had failed to unveil the Higgs boson [5].

When Run1 LHC data started to place limits on the  $VH$  channel, the nonobservation of deviations in the high- $p_T$  regions was also used to inform our global understanding of SMEFT theories [15,16] and the  $CP$  properties of the Higgs [20]. Moreover, the understanding and classification of the  $p_T^V$  distributions was crucial for Run2 collaborations to achieve a measurement [12,13].

As mentioned in the Introduction, the  $VH$  channel would seem qualitatively less interesting than the VBF channel, where the forward jets enrich the overall kinematic information. Nonetheless,  $VH$  with its slightly different three channels also offers interesting kinematic information, see in Fig. 2 a few examples of distributions we will use later in our analysis.

Moreover, progress is made at stages and the experimental understanding of the VBF channel is nowhere close to  $VH$ . In  $VH$ , huge SM reducible backgrounds, such as a  $Z$  and heavy flavor ( $Z + HF$ ) production, had been studied and kept under control thanks to the tremendous ingenuity of the experimental collaborations. The recent observation of the Higgs sets then a new stage for the  $VH$  channel, where new physics can be searched and tensioned against SM-Higgs production.

### IV. USING A SHALLOW NEURAL NETWORK

In this section we will describe the methodology we developed to study the SMEFT in  $VH$  using machine learning, in particular a shallow neural network.<sup>1</sup> To help the novice reader, in Appendix A we have collected a glossary of terms alongside brief explanation of their meaning. We simulate parton level (signal) events in the SMEFT framework and for an irreducible background, i.e., SM Higgs production. In the following whenever we talk

<sup>1</sup>Shallow NNs with the modified loss functions are very suitable for this task. We believe the physical reason for this behavior is related to the level of abstraction of our chosen variables. Kinematic variables, made up from combinations of measurements in the LHC detectors are already at a higher level of abstraction than other, more raw, measurements like hits on calorimeters, or lengths of displacement from the interaction vertex. If we had chosen to search for new physics using this lower-level features, we may have to increase the complexity of our NN. The deep NN might have ended up selecting, at a later stage in the NN evolution, specific combinations of the intensity of calorimeter hits and positions from the interaction points similar to our invariant mass or momenta.

about the background we mean the SM Higgs background, except when explicitly discussing the reducible SM background  $Z +$  heavy flavor.

To extract the maximum amount of information from the kinematic features, one needs to combine multidimensional information such as shown in Fig. 2 in 0D, 1D, 2D and even higher dimensionalities. The objective is to *maximize* our ability to detect new phenomena, which in HEP means maximizing the significance of an observation. Given a number of signal events  $s$ , where signal here represents SM plus a deviation like  $c_{HW}$ . For this study we consider as signal the effects of the dim-6 operators up to quadratic terms, while for further studies one might want to consider the inclusion of terms from the dim-8 operators which could compete at the same quadratic level [21]. We also evaluate a number of background events  $b$  which in our analysis correspond to the SM-only hypothesis.

One can use the Asimov estimate of significance [22] as a measure the algorithm needs to maximise, see Appendix A. Our problem then consists on building a function, inverse of the Asimov significance, and find the minimum value inside a multi-dimensional parameter space by including information from a diverse set of observables. The main difference from other loss function seen in the literature (cross-entropy) is that this specific loss function gives more priority to the correct classification of the signal, whereas in the case of cross-entropy equal weight is given to the correct classification of both categories.

Similarly to the procedure described in Ref. [23], we use a shallow neural network (NN) built from one hidden layer with number of neurons equal to the number of kinematic observables we consider.<sup>2</sup> The classification layer consists on a single unit (or neuron) with a softmax activation function. The softmax function forces the output to be in a range from 0 to 1, so the score of the classification layer can be interpreted as the probability of a given event to be a signal ( $SM + c_{HW}$ ) or background (SM only). Note that we consider a balanced data set for the signal and background process. To set the best hyperparameters, instead of performing a brute-force grid search as in [23], we make use of evolutionary algorithms in Python (DEAP) [24], in addition to the Scikit-Learn library. As activation function, we found a rectifier function [ $\max(x, 0)$ ] to perform better than the typical sigmoid and other common logistic regression options. For optimization, we found Adam was best performing. Other minor adjustments were done to the batch size and the dropout options, see Appendix A.

We then fed the algorithm with a large number of simulated events, both signal and background. The events had a number of characteristics, including  $p_T$  of the objects (b-jets, leptons, missing energy) and combinations of different objects. The identity of the event (signal or background)

<sup>2</sup>Code repository can be found at: <https://github.com/FFFreitas/Exploring-SMEFT-in-VH-with-Machine-Learning>

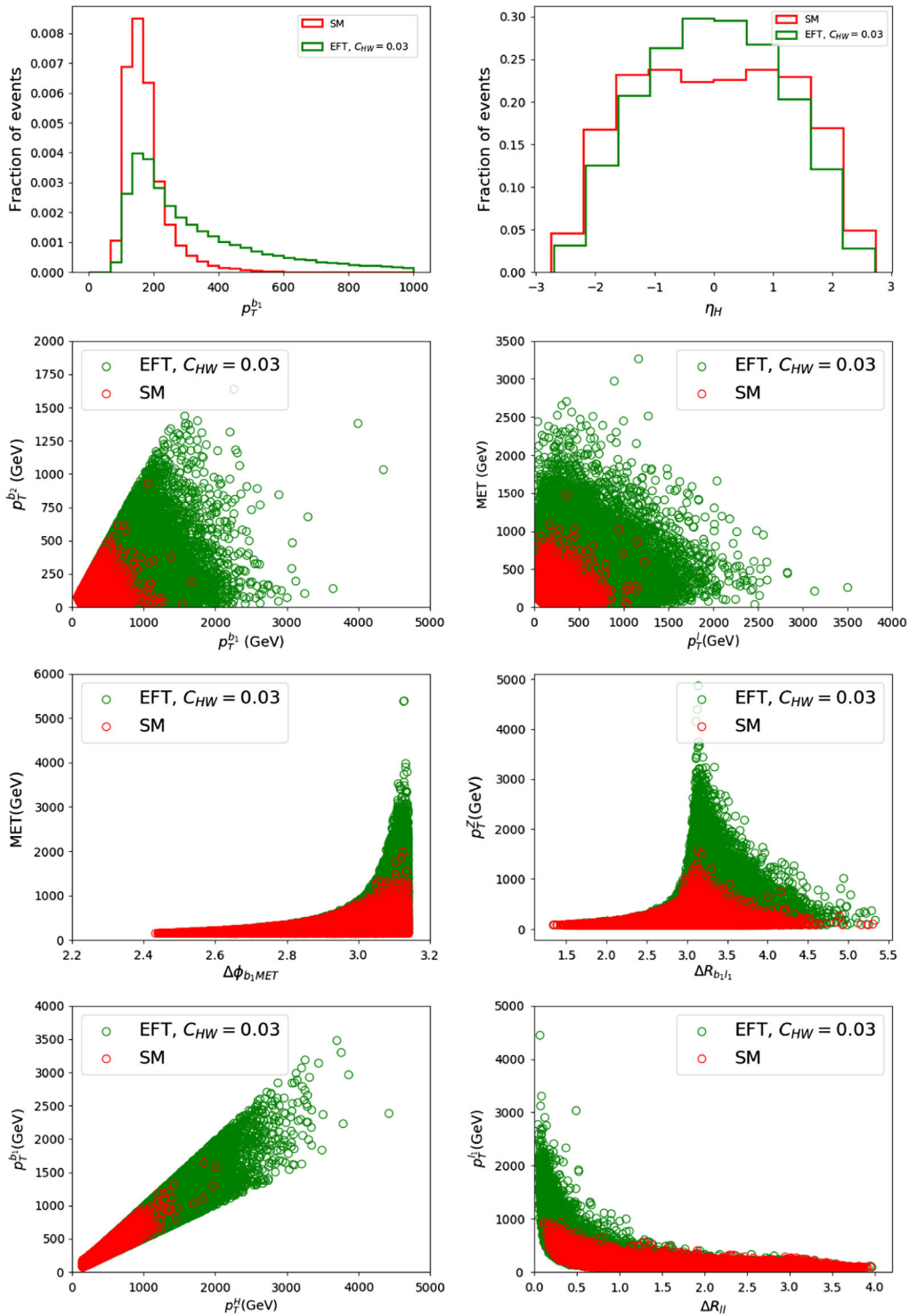


FIG. 2. Few illustrative 1D and 2D feature plots for inclusive 0L, 1L and 2L SM-Higgs production. Red dots correspond to background (SM Higgs production) and green dots to signal (Higgs SMEFT). Note the broader kinematic reach of the signal.

was used by the algorithm as part of the training, as we are dealing with a supervised machine learning problem. Before tackling minimization of the Asimov loss function, we performed a pretraining set of runs for 5 epochs, along the lines suggested in Ref. [23] using a steeper loss function. The pretraining phase speeds up the optimization algorithm

to minimize the Asimov loss function. Indeed, we found that training the NN for 5 epochs with  $l_{s/\sqrt{s+b}}$  was sufficient for the optimization algorithm to minimize that simpler loss function. A longer run, with about 20-30 epochs was then carried on with the  $l_{Asimov}$ , the Asimov significance does not increase after 30 epochs.

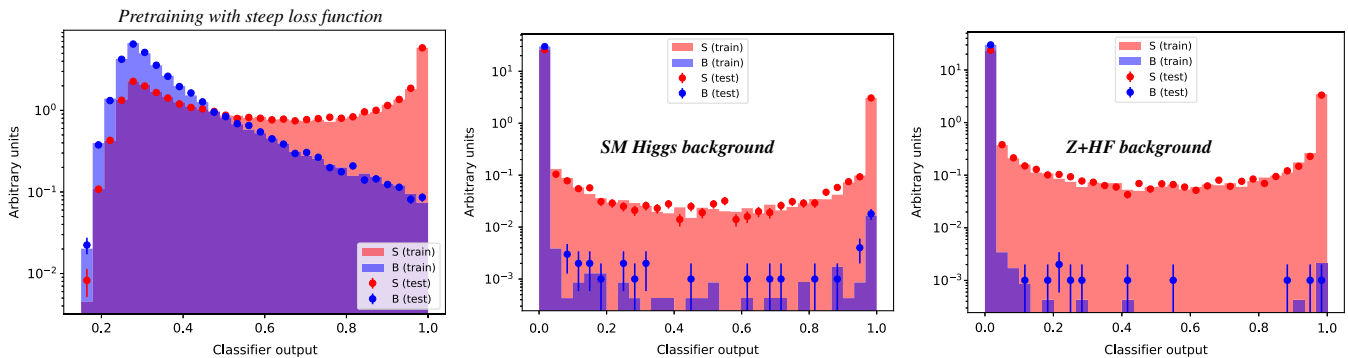


FIG. 3. Distribution of the 0L signal benchmark-point with  $c_{HW} = 0.03$  (red) and background (blue) events as a function of the classifier output. The left plot is the outcome of performing an initial pretraining run with 5 epochs. The middle (SM Higgs) and right (Z + HF) plots show the final distribution after a more precise, longer run is done. The solid distribution correspond to the outcomes on the training sample (70% of the sample), whereas the dots correspond to the test sample.

The outcome of these runs was the ability to classify events as signal or background, and to assign a Asimov significance estimate to a particular choice of  $c_{HW}$  coefficient (the strength of the deviation) and luminosity (the amount of available data).

In Fig. 3 we show the effect of pretraining in separating signal over background. The plots show the distribution of the signal benchmark-point with  $c_{HW} = 0.03$  (red) and background (blue) events as a function of the classifier output. The left plot is the outcome of performing an initial pretraining run with 5 epochs. The middle and right plots shows the final distribution after a longer run was performed. In all the plots, the solid distribution correspond to the outcomes on the training sample (70% of the sample), whereas the dots correspond to the test sample. The fact that the solid distribution (train) and dots (test) distributions are similar is an indication that the algorithm is not overfitting. The middle plot compares BSM with SM Higgs production. The right plot is the separation between BSM and a reducible background, Z + HF, where a cut on the  $m_{bb}$  variable in the Higgs mass window was done. By

comparing the middle and right plot, one sees that the reducible background is easier to remove than the genuine SM Higgs background, as expected.

## V. RESULTS

The goodness of our procedure can be first evaluated by looking at the ROC curve in Fig. 4, where we show the signal efficiency and background rejection curves. We present two examples of SMEFT effects, our benchmark value  $c_{HW} = 0.03$  and a very small value 0.001 which approaches the SM case. As expected, larger values of  $c_{HW}$  present a better AUC and higher significance.

Perhaps a more intuitive way to understand this ROC curve is to compute the predicted identity of events. In the right panel of Fig. 4, we show a kinematic distribution of true signal events, separated by their predicted identity. Unsurprisingly, events with high energy are easier to distinguish from the SM backgrounds. This can be traced back to the Feynman rule in Eq. (1), where the SMEFT effects are momentum dependent and tend to lead to higher kinematic reach than SM interactions.

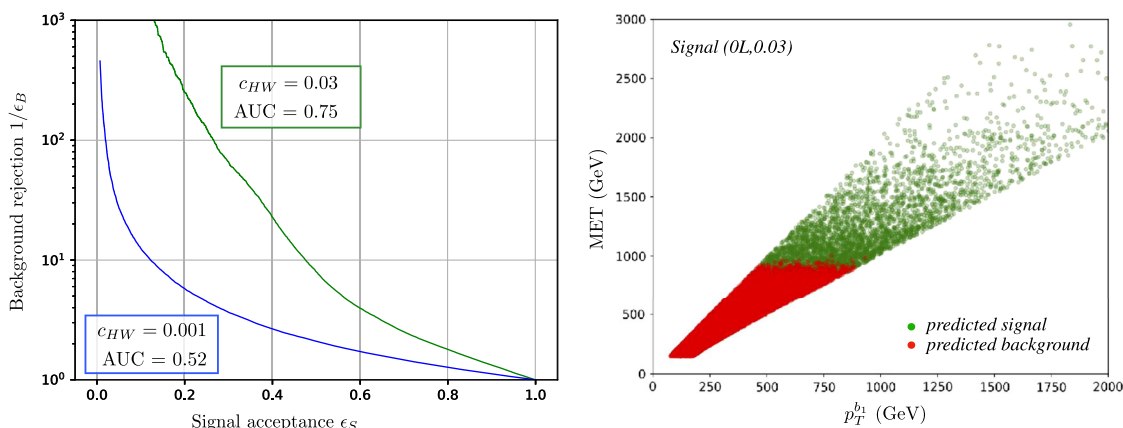


FIG. 4. Left: ROC curve for two values of the SMEFT coefficient in the 0L channel. Right: Classification of true signal events for  $c_{HW} = 0.03$  and their mapping to kinematic features.

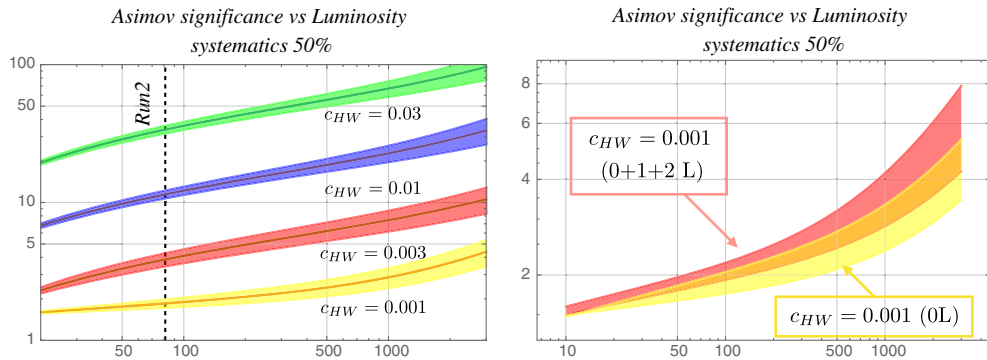


FIG. 5. Left: Luminosity ( $\text{fb}^{-1}$ ) versus Asimov significance for different values of  $c_{HW}$  in the 0-lepton channel and 50% systematic uncertainty. Right: The effect of combination of  $VH$  channels for the limiting value  $c_{HW} = 0.001$  and 50% systematic errors.

Nevertheless, quantities like the ROC curve and its AUC do not provide the answers we need in particle physics. We are interested in understanding beyond acceptance and rejection, but also the dependence with increasing luminosity and the effect of systematic uncertainties which are often disregarded in machine learning studies. In the right panel of Fig. 5 we show the Asimov significance in the 0L channel, for a choice of systematic uncertainty at 50%, as a function of luminosity for various choices of the SMEFT coefficient. The bands correspond to  $2\sigma$  ranges. In contrast with the results from global fits, we obtain that values much below the 0.03 benchmark may be excluded by the Run2 data. The extent of this current exclusion cannot be obtained in a reliable fashion from our analysis, as we did perform a simplistic leading-order parton + shower analysis. Nevertheless, one would infer that sensitivity to value of  $c_{HW}$  around 0.001 could be obtained using the CMS and ATLAS combined Run 2 data.

As one can see from the left panel of Fig. 4,  $c_{HW} = 0.001$  seems a limiting case for our algorithm in the 0L channel, as the AUC is very close to 0.5. We have chosen the 0L channel as it generally provides the best sensitivity to SMEFT [15], but one would wonder whether one could improve the sensitivity to this difficult point by combining with the other two channels (1L and 2L). The right panel in Fig. 5 shows the increase of sensitivity due to combination. For the small SMEFT deviation  $c_{HW} = 0.001$  and 50% systematics, the improvement is within the error bars of the Asimov significance. A better handle on systematics could make the combination much more effective.

## VI. OUTLOOK

With the increasing experimental understanding of the LHC data, new ways to search for new physics open up. In particular, the use of detailed kinematic information is the next frontier in terms of LHC data characterisation. More capabilities come with more ambitions, particularly in terms of the complexity of new phenomena one can hope to tackle. We have identified one channel ( $VH$ ) which is both relatively well understood and broad in terms of its

kinematic reach, and a set of machine learning techniques which could allow us to detect new physics in the behavior of the Higgs boson. We chose the SMEFT as a template of the kind of deviations one could expect in the Higgs via virtual effects of new particles.

Within the framework of our analysis, we found the 0L channel to be dominant, which was expected. We obtained a limit in the SMEFT coefficient  $c_{HW}$  of 0.001, about 30 times better than the current constraint from a global analysis [8] with the Run 2 data. This result shows the potential of incorporating these techniques to the SMEFT studies.

Our analysis could be improved in a number of ways. First and foremost, a more realistic simulation could be performed by the experiments, including NLO SMEFT effects [25]. A study including full simulation as well as a more thorough understanding of the systematic uncertainties should be carried out, hopefully by the experimental collaborations. The aim of this paper was to show the potential of this method using a shallow NN and a new choice of Asimov significance, explain the work-flow and point out the importance of including systematic effects in the NN.

Second, although we found that deep layers led to overfitting, and a shallow NN was more suitable, new algorithms could be explored to increase sensitivity. In particular, one could use outlier detection without supervision. Third, we should understand the effect of switching on more than one deviation along the lines described in Ref. [10]. This should be the stepping stone to a more global use of machine learning techniques in the area of global fits to SMEFT properties.

## ACKNOWLEDGMENTS

F. F. F. is supported by Project No. Y8Y2411B11 China Postdoctoral Science Foundation. C. K. K. wishes to acknowledge support from the Royal Society-SERB Newton International Fellowship (NF171488). The work of V. S. by the Science Technology and Facilities Council (STFC) under Grant No. ST/P000819/1.

## APPENDIX A: GLOSSARY OF TERMS

True positive rate ( $tpr \equiv \epsilon_S$ ): ratio of true positive count and total signal events. The true positive counts are the number of signal events correctly identified by the algorithm. It also corresponds to the usual notion of signal acceptance.

False positive rate ( $fpr \equiv \epsilon_B$ ): ratio of false positive counts and total number of background events. The false positive counts are the background events, predicted as signal events by the algorithm. It also corresponds to the usual notion of background selection.

Receiver operating characteristic (ROC): plot of  $tpr$  as a function of  $fpr$  for each value of the classifier threshold between 0 and 1.

Area under the curve (AUC): area under the ROC curve and a typical measure of the algorithm performance.

Accuracy: ratio of the correctly identified signal and background events versus total number of signal and background events.

Learning curve: curve with shows the performance of the algorithm with iterative runs, i.e., behavior of the loss function with iterations.

Batch: data is divided into small sets, called batches, to save time and computation efforts.

Hidden layers: intermediate layers between the input and output layers.

Asimov significance: Defined as a function of signal and background events and the uncertainty associated with the background ( $\sigma_b$ )

$$Z_A = \left[ 2((s+b) \ln \left[ \frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[ 1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right]) \right]^{1/2}. \quad (\text{A1})$$

Loss function: the function which the algorithm searches to minimize. In our analysis, we used the following loss function designed to maximize the discovery significance:

$$\ell_{s/\sqrt{s+b}} = (s+b)/s^2, \quad (\text{A2})$$

$$\ell_{\text{Asimov}} = 1/Z_A \quad (\text{A3})$$

Epochs: The period between initialization of the search for the minimum and when the batches pass the NN. Basically, number of epochs is an iteration counter of how many times complete data set is explored by the algorithm, such that learning parameters are optimized.

Dropout: mechanism to avoid the model overfitting, whereby the NN could drop few of the units (neurons) at the time of training.

Pretraining: Quick prerun with smaller number of epochs and steeper loss functions. The longer training is initialised by the pretraining hyperparameters.

Classifier output: set of predictions for test sample. Our analysis is a binary classification problem, so with the predecided (user-decided) classification threshold, the events will either belong to signal or background class.

## APPENDIX B: ANALYSIS SETUP

We generate parton level 100K events for  $WH$  and  $ZH$  processes, with  $\sqrt{s} = 14$  TeV, using MC@NLO madgraph6.3.2 and the Higgs effective theory feynrules [26] model available in the literature [27]. We consider the new physics effects up to quadratic level. Note that in  $VH$ , the typical difference between 13 TeV and 14 TeV in terms of cross section is less than 10%.

We use Pythia6 [28] to read the lhe events files but without doing the showering and hadronization. Considering leptonic decay channels of gauge boson and  $b\bar{b}$  from Higgs decay, we have the following three final states

(i) 0-lepton (0L):

$$pp \rightarrow HZ, \quad (H \rightarrow b\bar{b}, Z \rightarrow \nu\bar{\nu})$$

(ii) 1 lepton (1L):

$$pp \rightarrow HW, \quad (H \rightarrow b\bar{b}, W \rightarrow l\nu_l)$$

(iii) 2 lepton (2L):

$$pp \rightarrow HZ, \quad (H \rightarrow b\bar{b}, Z \rightarrow l^+l^-)$$

The events are generated according to the ATLAS search strategy[12] categorized as ‘‘inclusive.’’ The cuts applied are given in Table I for all the channels. We consider the two main background processes, i.e., SM associated Higgs production with vector bosons and  $V$  + heavy flavor ( $V$  + HF). The SM associated Higgs production with vector bosons background is also generated importing the same model with  $C_{HW} = 0$ .

We consider the following observables as data features:

- (i)  $p_T^{b_1}$  transverse momentum of the leading b-jet.
- (ii)  $p_T^{b_2}$  transverse momentum of the sub leading b-jet.
- (iii)  $p_T^{VH}$  transverse momentum of the  $VH$  pair.

TABLE I. Cuts applied at event generation level for both signal and background process. In case of  $Z$  + HF we apply an additional cut on  $m_{b\bar{b}}$ , i.e.,  $115 < m_{b\bar{b}} < 135$  GeV.

Channel	Inclusive
0L	$\cancel{E}_T > 150$ GeV
1L	$p_T^l > 25$ GeV, $ \eta_l  < 2.7$
2L	$\cancel{E}_T > 30$ GeV, $p_T^V > 150$ GeV $p_T^l > 7$ GeV, $ \eta_l  < 2.7$ , $p_T^V > 75$ GeV Leading lepton $p_T > 27$ GeV
0L, 1L, 2L	$p_T^b > 20$ GeV, $ \eta_b  < 2.5$ , Leading b-jet $p_T > 45$ GeV

- (iv)  $M_T^{VH}$  transverse mass of the  $VH$  pair.
  - (v)  $p_T^{W/Z}$  transverse momentum of gauge boson.
  - (vi)  $p_T^H$  transverse momentum of the reconstructed Higgs boson.
  - (vii)  $\eta^H$  pseudorapidity of the reconstructed Higgs boson.
  - (viii)  $\phi^H$  azimuthal angle of the reconstructed Higgs boson.
- which are common for all 3-channels and channel specific ones are:
- (1) 0L channel:
    - (a)  $\cancel{E}_T$  missing transverse energy
    - (b)  $\Delta\phi_{b_1, \cancel{E}_T}$  azimuthal angular separation between leading b-jet and  $\cancel{E}_T$
  - (2) 1L channel:
    - (a)  $M_T^W$  transverse mass of the  $W^\pm$
    - (b)  $p_T^l$  transverse momentum of lepton
  - (3) 2L channel:
    - (a)  $p_T^{l_1}$  transverse momentum of the leading lepton
    - (b)  $p_T^{l_2}$  transverse momentum of subleading lepton
    - (c)  $\Delta R_{ll}$  separation between two lepton in the  $\eta$ - $\phi$  plane
    - (d)  $\Delta\phi_{b_1, l_1}$  azimuthal angular separation between leading b-jet and leading lepton
    - (e)  $\Delta\phi_{b_2, l_1}$  azimuthal angular separation between subleading b-jet and leading lepton
  - (c)  $\cancel{E}_T$  missing transverse energy
  - (d)  $\Delta R_{wl}$  separation between lepton and  $W$  boson in the  $\eta$ - $\phi$  plane
  - (e)  $\Delta\phi_{b_1, l}$  azimuthal angular separation between leading b-jet and lepton
  - (f)  $\Delta\phi_{l, \cancel{E}_T}$  azimuthal angular separation between lepton and  $\cancel{E}_T$

- 
- [1] W. Buchmuller and D. Wyler, *Nucl. Phys.* **B268**, 621 (1986).
  - [2] See I. Brivio and M. Trott, *Phys. Rep.* **793**, 1 (2019) and references therein.
  - [3] M. Gorbahn, J. M. No, and V. Sanz, *J. High Energy Phys.* **10** (2015) 036.
  - [4] D. de Florian *et al.* (LHC Higgs Cross Section Working Group), [arXiv:1610.07922](https://arxiv.org/abs/1610.07922), <http://dx.doi.org/10.23731/CYRM-2017-002>.
  - [5] J. Ellis, D. S. Hwang, V. Sanz, and T. You, *J. High Energy Phys.* **11** (2012) 134; J. Ellis, V. Sanz, and T. You, *Eur. Phys. J. C* **73**, 2507 (2013).
  - [6] E. Massó and V. Sanz, *Phys. Rev. D* **87**, 033001 (2013); V. M. Abazov *et al.* (D0 Collaboration), *Phys. Rev. Lett.* **113**, 161802 (2014); K. Mimasu, V. Sanz, and C. Williams, *J. High Energy Phys.* **08** (2016) 039; S. Banerjee, C. Englert, R. S. Gupta, and M. Spannowsky, *Phys. Rev. D* **98**, 095012 (2018).
  - [7] F. Ferreira, S. Fichet, and V. Sanz, *Phys. Lett. B* **778**, 35 (2018).
  - [8] J. Ellis, C. W. Murphy, V. Sanz, and T. You, *J. High Energy Phys.* **06** (2018) 146.
  - [9] P. Mehta, M. Bukov, C. H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, *Phys. Rep.* **810**, 1 (2019).
  - [10] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *Phys. Rev. D* **98**, 052004 (2018); *Phys. Rev. Lett.* **121**, 111801 (2018).
  - [11] J. Brehmer, F. Kling, T. Plehn, and T. M. P. Tait, *Phys. Rev. D* **97**, 095017 (2018).
  - [12] M. Aaboud *et al.* (ATLAS Collaboration), *Phys. Lett. B* **786**, 59 (2018).
  - [13] A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Rev. Lett.* **121**, 121801 (2018).
  - [14] See, for example: A. Andreassen, I. Feige, C. Frye, and M. D. Schwartz, *Eur. Phys. J. C* **79**, 102 (2019); T. Roxlo and M. Reece, [arXiv:1804.09278](https://arxiv.org/abs/1804.09278); K. Fraser and M. D. Schwartz, *J. High Energy Phys.* **10** (2018) 093; M. Farina, Y. Nakai, and D. Shih, [arXiv:1808.08992](https://arxiv.org/abs/1808.08992) [*Phys. Rev. D* (to be published)]; T. Heindel, G. Kasieczka, T. Plehn, and J. M. Thompson, [arXiv:1808.08979](https://arxiv.org/abs/1808.08979); G. Kasieczka, T. Plehn, M. Russell, and T. Schell, *J. High Energy Phys.* **05** (2017) 006.
  - [15] J. Ellis, V. Sanz, and T. You, *J. High Energy Phys.* **07** (2014) 036.
  - [16] J. Ellis, V. Sanz, and T. You, *J. High Energy Phys.* **03** (2015) 157.
  - [17] E. da Silva Almeida, A. Alves, N. Rosa Agostinho, O. J. P. Éboli, and M. C. Gonzalez-Garcia, *Phys. Rev. D* **99**, 033001 (2019).
  - [18] A. Biekötter, T. Corbett, and T. Plehn, *SciPost Phys.* **6**, 064 (2019).
  - [19] M. Cepeda *et al.* (Physics of the HL-LHC Working Group), [arXiv:1902.00134](https://arxiv.org/abs/1902.00134).
  - [20] R. M. Godbole, D. J. Miller, K. A. Mohan, and C. D. White, *J. High Energy Phys.* **04** (2015) 103; F. Ferreira, B. Fuks, V. Sanz, and D. Sengupta, *Eur. Phys. J. C* **77**, 675 (2017).
  - [21] C. Hays, A. Martin, V. Sanz, and J. Setford, *J. High Energy Phys.* **02** (2019) 123.
  - [22] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Eur. Phys. J. C* **71**, 1554 (2011); **73**, 2501 (2013).
  - [23] A. Elwood and D. Krücker, [arXiv:1806.00322](https://arxiv.org/abs/1806.00322).
  - [24] <https://github.com/deap/deap>.
  - [25] C. Degrande, B. Fuks, K. Mawatari, K. Mimasu, and V. Sanz, *Eur. Phys. J. C* **77**, 262 (2017); K. Mimasu, V. Sanz, and C. Williams, *J. High Energy Phys.* **08** (2016) 039.
  - [26] N. D. Christensen and C. Duhr, *Comput. Phys. Commun.* **180**, 1614 (2009).
  - [27] A. Alloul, B. Fuks, and V. Sanz, *J. High Energy Phys.* **04** (2014) 110.
  - [28] T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, G. Miu, S. Mrenna, and E. Norrbin, *Comput. Phys. Commun.* **135**, 238 (2001).