Probing triple Higgs coupling with machine learning at the LHC

Murat Abdughani,[1,2] Daohan Wang,[1,2] Lei Wu,[3] Jin Min Yang,[1,2] and Jun Zhao[1,2]

[1] *CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics,*
Chinese Academy of Sciences, Beijing 100190, China

[2] *School of Physics, University of Chinese Academy of Sciences, Beijing 100049, China*

[3] *Department of Physics and Institute of Theoretical Physics,*
Nanjing Normal University, Nanjing, 210023, China

Abstract

Measuring the triple Higgs coupling is a crucial task in the LHC and future collider experiments. We apply the Message Passing Neural Network (MPNN) to the study of non-resonant Higgs pair production process $pp \to hh$ in the final state with $2b + 2\ell + E_{\mathrm{T}}^{\mathrm{miss}}$ at the LHC. Although the MPNN can improve the signal significance, it is still challenging to observe such a process at the LHC. We find that a 2σ upper bound (including a 10% systematic uncertainty) on the production cross section of the Higgs pair is 3.7 times the predicted SM cross section at the LHC with the luminosity of 3000 fb^{-1}, which will limit the triple Higgs coupling to the range of $[-3, 11.5]$.

1

CONTENTS

I. INTRODUCTION

The discovery of a 125 GeV Higgs boson [1, 2] is a great leap in the quest to the origin of mass. The precision measurement of the Higgs couplings is one of the primary goals of the LHC experiment, which will further reveal the electroweak symmetry breaking mechanism and shed lights on the new physics beyond the Standard Model (SM). Although the current measurements of the Higgs couplings with fermions and gauge bosons are compatible with that predicted by the SM, testing the triple and quartic Higgs self-interactions is rather challenging at the LHC (for recent reviews, see e.g., [3–5]).

In the Brout-Englert-Higgs mechanism of electroweak symmetry breaking [6–10], the Higgs boson is a massive scalar with self-interactions. The Higgs self-couplings are determined by the structure of the scalar potential,

$$V = \frac{m_h^2}{2}h^2 + \lambda_3^{\text{SM}}vh^3 + \frac{1}{4}\lambda_4^{\text{SM}}h^4 \ , \tag{1}$$

with

$$\lambda_3^{\text{SM}} = \lambda_4^{\text{SM}} = \frac{m_h^2}{2v^2}. \tag{2}$$

Here m_h is the mass of the SM Higgs boson and v is the vacuum expectation value of the SM Higgs field. λ_3^{SM} and λ_4^{SM} are the values of the SM Higgs self-couplings, respectively. In many extensions of the SM, these couplings can be altered by Higgs mixing effects or higher order corrections induced by new particles, such as Two Higgs Doublet Model [11–13] and (Next-to-)Minimal Supersymmetric Standard Model [14–17]. Since the Higgs self-coupling plays an important role in vacuum stability [18] and electroweak baryogenesis [19, 20], measuring the Higgs self-coupling will provide a crucial clue to new physics [21].
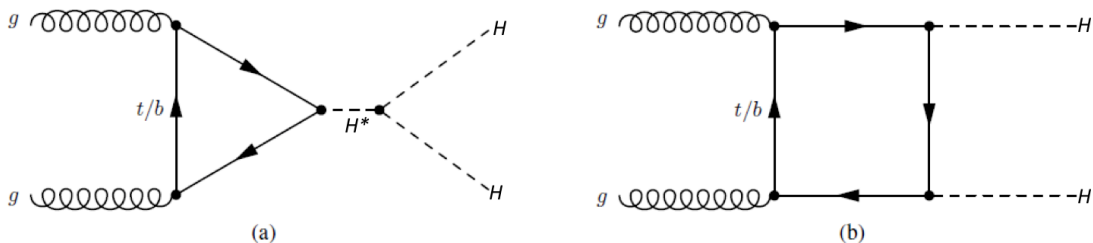

FIG. 1. The representative Feynman diagrams for the Higgs boson pair production in the SM: (a) through triple Higgs self-coupling; (b) through Higgs-fermion Yukawa interaction.

The triple Higgs coupling can be indirectly probed by using the loop effects in some observables, for example, the single Higgs production [22–24], and the electroweak precision observables [25]. With 80 fb^{-1} of LHC Run-2 data, the triple Higgs coupling has been constrained in the range $-3.2 < \lambda_3/\lambda_3^{SM} < 11.9$ at 95% C.L. [?]. On the other hand, the Higgs pair production provides a direct way to measure the triple Higgs coupling at the LHC. Such a production is dominated by the gluon-gluon fusion process, which has two main contributions: one is from the triangle diagram induced by the triple Higgs coupling, and the other is from the box diagram mediated by the top quark, as shown in Fig. 1. It should be noted that these two amplitudes interfere destructively, and thus results in a small cross section of 38.65 fb for the production process $gg \rightarrow hh$ at 14 TeV LHC, which is computed at next-to-next-to-next-to-leading order (N^{3}LO) and including finite top quark mass effects [27]. The new physics effects that can significantly modify the Higgs pair production have been intensively studied at the LHC (see, for examples [28, 30, 32–34, 37–39] and references therein).

In Refs. [3, 4, 40, 41, 43, 47–51, 53], the potential of measuring the Higgs pair production has been investigated in various decay modes: $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}WW^*$, $\gamma\gamma b\bar{b}$, $\gamma\gamma WW^*$ and WW^*WW^*. Among these channels, the process of $hh \rightarrow 4b$ has the largest branching ratio, while the process of $hh \rightarrow b\bar{b}\gamma\gamma$ has a more promising sensitivity because of the low backgrounds. Using the combination of the above six analyses, the ratio $-5 < \lambda_3/\lambda_3^{SM} < 12$ is constrained at 95% C.L. at 13 TeV LHC with the luminosity of 36.1 fb^{-1}. The sensitivity will be greatly improved at the HL-LHC [55] and future hadron colliders [56].

In addition to the conventional kinematic cut-flow analyses, the machine learning methods have been proposed to accelaerate the discovery of new physics [57–69, 71–90]. In this paper we use the Message Passing Neural Network (MPNN) [92] to explore the potential of observing the di-Higgs events through the channel $pp \rightarrow hh \rightarrow b\bar{b}WW*$, which has the second largest production rate among various final states of di-Higgs at the LHC. The MPNN inherits the generality and powerfulness of Graph Neutral Network (GNN) [93, 94]. In the MPNN, a collision event is represented as a numerical geometrical graph formed by a number of final state objects, which are non-linear models with a bunch of parameters that relates the output to the input graphs. The supervised learning is used to find optimized parameters, and will help to recognize the pattern in the collision events efficiently. Different from Deep Neural Network (DNN), MPNN is a dynamic neural network and is independent on the

number and ordering of final state particles. Therefore, the MPNN is suitable for processing the graph representation of collision event. Recently, this method has been successfully applied to collider phenomenological studies, such as jet physics [95], Higgs physics [96] and supersymmetry [97].

This paper is organized as follows. In Section II, we describe the event generation and reconstruction for the signal and backgrounds. Next, in Section III, we illustrate the event graph and network architecture for the MPNN approach. In Section IV, we present numerical results and discussions. Finally, we draw our conclusions in Section V.

II.   EVENT GENERATION AND RECONSTRUCTION

The signal and background events at parton level are generated with MadGraph5_aMC@NLO v2.6.1 [98] with the default parton distribution function (PDF) set NNPDF2.3QED [99] at the LHC with leading order with center-of-mass energy $\sqrt{s} = 14$ TeV. We employ the following cuts for parton level event generation [1]: $p_{Tj} > 20$ GeV, $p_{Tb} > 20$ GeV, $p_{T\gamma} > 10$ GeV, $p_{T\ell} > 10$ GeV, $\eta_j < 5$, $\eta_b < 5$, $\eta_\gamma < 2.5$, $\eta_\ell < 2.5$, $\Delta R_{bb} < 1.8$, $\Delta R_{\ell\ell} < 1.3$, 70 GeV $< m_{jj}, m_{bb} < 160$ GeV and $m_{\ell\ell} < 75$ GeV. We impose additionally 5 GeV $< m_{\ell\ell} < 75$ GeV for $jj\ell\ell\nu\bar{\nu}$, $\ell\ell bj$ and $tW + j$ backgrounds. The angular distance ΔR_{ij} is defined by

$$\Delta R_{ij} = \sqrt{(\Delta\phi_{ij})^2 + (\Delta\eta_{ij})^2} \; , \tag{3}$$

where $\Delta\phi_{ij} = \phi_i - \phi_j$ and $\Delta\eta_{ij} = \eta_i - \eta_j$ are the differences of the azimuthal angles and rapidities between particles i and j, respectively.

The signal cross section is normalized to the next-to-next-to-leading-order (NNLO) accuracy in QCD [100], that is $\sigma_{gg \to hh} = 40.7$ fb. The main background $t\bar{t}$ cross section is normalized to the NNLO QCD value 953.6 pb [101]. Along with the signal and $t\bar{t}$, all other backgrounds and their normalized cross sections are listed in Table I.

The events obtained after parton level cuts are hadronized with PYTHIA8243 [102], followed by detector simulation with Delphes 3.4.2 [103] and jet-reconstruction with FastJet 3.3.2 [104]. The Delphes card for ATLAS detector simulation is modified as

- Anti-k_T algorithm [105] with cone-size $\Delta R = 0.4$ is used for jet clustering, and jets with $p_{Tj} > 20$ GeV and $|\eta_j| < 2.5$ are selected.

---

[1] For the purpose of comparison with the earlier works, especially with [51, 53], we adopt the same parton level cuts, event reconstruction parameters, and *baseline selection cuts* as in [53] to get the event data.

- Flat b-tagging efficiency is $\epsilon_{b \to b} = 0.75$, mis-tagging efficiency for c quark as b is $\epsilon_{c \to b} = 0.1$, and mis-tagging rates of other jets are $\epsilon_{b \to b} = 0.01$ [106].

- Maximum transverse momenta ratio for lepton isolation is set as $\frac{\sum_i p_{Ti}}{p_{Tl}} < 0.43$, where the sum is taken over the transverse momenta p_{Ti} of all final state particles i, not including ℓ, with $p_{Ti} > 0.5$ GeV and within angular distance $\Delta R_{i\ell} < 0.3$ with lepton candidate ℓ. Leptons with $p_{Tl} > 10$ GeV and $|\eta_l| < 2.5$ are selected.

- Isolation of photons also require $\frac{\sum_i p_{Ti}}{p_{T\gamma}} < 0.12$ for particles i, without including γ, with $p_{Ti} > 0.5$ GeV and within angular distance $\Delta R_{i\gamma} < 0.3$ with photon candidate γ. Photons are require to have $p_{Tl} > 25$ GeV and $|\eta_\gamma| < 2.5$ to be selected.

After the reconstruction, the missing transverse momentum $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative vector sum of the original $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$, which equals to the negative sum of the transverse momenta of the selected photons, leptons and jets, and unused tracks as in [107]:

$$\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = - \sum_{\substack{\text{selected} \\ \text{electrons}}} \mathbf{p}_{\mathrm{T}}^e - \sum_{\substack{\text{selected} \\ \text{muons}}} \mathbf{p}_{\mathrm{T}}^\mu - \sum_{\substack{\text{selected} \\ \text{photons}}} \mathbf{p}_{\mathrm{T}}^\gamma - \sum_{\substack{\text{selected} \\ \text{jets}}} \mathbf{p}_{\mathrm{T}}^j - \sum_{\substack{\text{unused} \\ \text{tracks}}} \mathbf{p}_{\mathrm{T}}^{\text{track}} \,, \tag{4}$$

where the tracks with $p_{Tl} > 0.4$ GeV and $|\eta_l| < 2.5$ are considered.

We further apply the following cuts (same as the *baseline selection cuts* in [53]) to reduce background events sufficiently relevant to the signals:

- The two leading jets must be b-tagged, each with $p_T > 30$ GeV.

- Exactly two opposite sign leptons, each with $p_T > 20$ GeV.

- Modulus of $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ is required to be $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV.

- Angular distances for two leptons and for two b jets are $\Delta R_{\ell\ell} < 1.0$ and $\Delta R_{bb} < 1.3$, respectively.

- Invariant masses for two leptons and for two b jets respectively are $m_{\ell\ell} < 65$ GeV and 95 GeV $< m_{bb} < 140$ GeV.

We export only the four momenta (also contain the corresponding charge signs of leptons and b-jet tagging information) of those events which passed the above cuts for later network training.
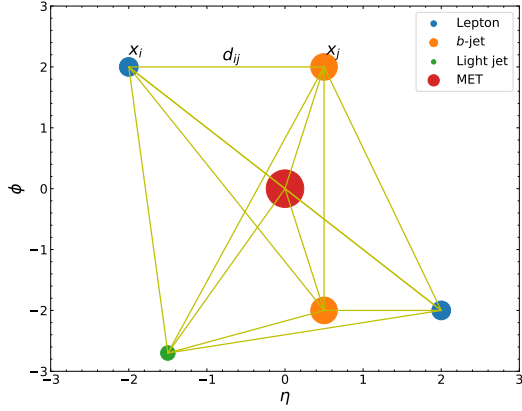
x	photon	lepton	jet	MET	p_T(TeV)	E(TeV)	m(TeV)
1	0	1	0	0	0.132	0.135	0.000
2	0	-1	0	0	0.025	0.025	0.000
3	0	0	1	0	0.163	0.227	0.012
4	0	0	1	0	0.052	0.053	0.006
5	0	0	-1	0	0.047	2.485	0.011
6	0	0	0	1	0.078	0.078	0.000

d	1	2	3	4	5	6
1	0.000	0.244	2.918	2.394	4.617	1.744
2	0.244	0.000	2.674	2.169	4.787	1.561
3	2.918	2.674	0.000	1.004	5.907	1.998
4	2.394	2.169	1.004	0.000	5.177	2.163
5	4.617	4.787	5.907	5.177	0.000	6.348
6	1.744	1.561	1.998	2.163	6.348	0.000

FIG. 2. The left figure illustrates an event graph, which includes nodes (circles) and edges (yellow lines), for a simulated signal event. A node represents a final state object passed all cuts and an edge represents the angular distance between two nodes. The upper right table shows the six objects; each of them is a seven-dimensional feature vector $x_i = (I_1, I_2, I_3, I_4, p_T, E, m)$ with I_i features identifying its type, e.g., $I_1 = 1$ for a photon, I_2 is the charge of the corresponding lepton, $I_3 = 1$ is a b-tagged jet, $I_3 = -1$ is a non-b-tagged light jet, $I_4 = 1$ is the missing transverse momentum. p_T, E and m are the transverse momentum, energy and mass of the object. The table at the bottom shows the angular distances Eq. (3) between a pair of nodes for all six objects.

III. EVENT GRAPH AND NETWORK ARCHITECTURE

Each collider event obtained in the preceding section is converted to an event graph as the input for our neural network. Fig. 2 illustrates a simulated signal event as an event graph which consists of nodes and edges. A node represents a final state object passed all the cuts and this object can be a photon, lepton, jet or missing transverse momentum (MET). Each node has a seven-dimensional feature vector $\boldsymbol{x}_i = (I_1, I_2, I_3, I_4, p_T, E, m)$ which contains the major property of the corresponding final state. For the elements of a feature vector, p_T, E and m are respectively the transverse momentum, energy and mass of the object, while the default values for I_i are 0, with $I_1 = 1$ for a photon, I_2 being the charge of the lepton, $I_3 = 1$ for a b-tagged jet, $I_3 = -1$ for a non-b-tagged jet, $I_4 = 1$ for the MET. Each pair of nodes are linked by an edge which is weighted by the angular distances (3) between the corresponding two nodes.

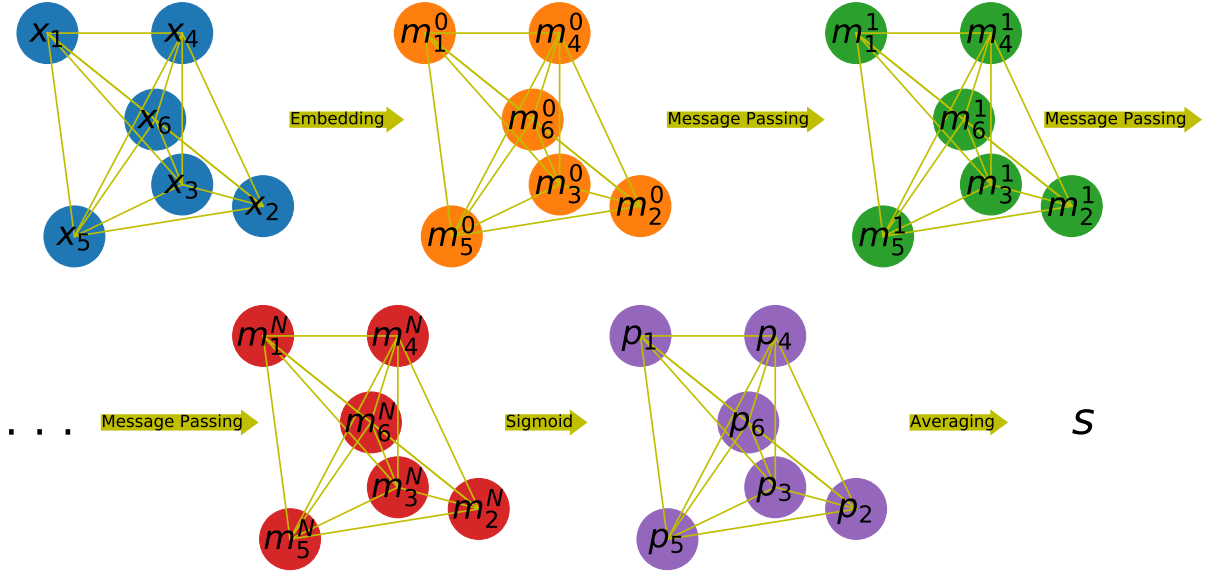Due to the rotation invariance of the differential cross section of the collider events around

FIG. 3. The schematic diagram of the MPNN classifier which consists of one input layer, one embedding layer, N message passing layer and output layer, and these layers are sequentially connected by non-linear transformations.

the beam axis, we can get rid of the information of azimuthal angle dependence of the event from the node features, and the difference of azimuthal angles is encoded in edge weights. This will make sure that the classification is not dependent on the definite azimuthal angle of the final states of an event, and stable w.r.t. the rotation of the event around the beam axis. The other two advantages of such an event graph design are: (1) The number of nodes equal to the number of final state objects, i.e., number of nodes is not fixed, which guarantees to use full information of final state objects; (2) The node features and edge weights are easily transformed by the four momenta of the object, no sophisticated discriminants are needed to be constructed, which makes the model quite general and easy to implement to other scenarios as well.

The structure of our MPNN is shown in Fig. 3. It consist of one input layer, one embedding layer, N massage passing layer and output layer. The embedding transformation between input layer and embedding layer is given by,

$$\boldsymbol{m}_i^0 = \mathrm{ReLU}(W_m^0 \boldsymbol{x}_i + \boldsymbol{b}_m^0) \,, \tag{5}$$

where W_e and $\boldsymbol{b}_e$ are learnable weight and bias vector, activation function ReLU is rectified linear unit. Dimension of $\boldsymbol{m}_i^0$ is higher than $\boldsymbol{x}_i$. ith node at the nth message passing layer is obtain by following transformation,

$$\boldsymbol{s}_i^n = \sum_j \mathrm{ReLU}(W_m^n[\boldsymbol{m}_j^{n-1}; \hat{\boldsymbol{d}}_{ij}] + \boldsymbol{b}_m^n), \tag{6}$$

$$\boldsymbol{m}_i^n = \mathrm{ReLU}(W_s^n[\boldsymbol{m}_i^{t-1}; \boldsymbol{s}_i^n] + \boldsymbol{b}_s^n) , \tag{7}$$

where i and j are indices of nodes, $\boldsymbol{s}_i^n$ is intermediate vector, the brackets are vector concatenation, Ws and $\boldsymbol{b}$s are learnable weights and bias. Above message passing transformation equations reveal that each node at the message passing layer get information from all previous nodes and distances between nodes. Expanding edge weight d_{ij} onto 21 Gaussian bases to make it more suitable for linear transformation [97], and the k-th component of this weight vector is

$$(\hat{\boldsymbol{d}}_{ij})_k = \exp\left\{\frac{(d_{ij} - \mu_k)^2}{2\sigma^2}\right\} , \tag{8}$$

where μ_k is linearly distributed in range of [0, 5] and $\sigma = 0.25$.

At the output layer, we use the sigmoid function on the vector m_i^N to get the probability p_i of the node i as

$$p_i = \sigma(W_p \boldsymbol{m}_i^N + \boldsymbol{b}_p) = \frac{1}{1 + e^{-(W_p \boldsymbol{m}_i^N + \boldsymbol{b}_p)}} . \tag{9}$$

and then average the probabilities from all nodes at the output layer by

$$s = \frac{1}{N} \sum p_i , \tag{10}$$

where N is the number of nodes in the input event.

The MPNN can be efficiently trained using supervised learning method. We adopt binary-cross-entropy as the loss function. We choose the number of message passing layers $N = 3$. The Adam [108] optimizer with a learning rate of 0.001 is used to optimize the model parameters based on the gradients calculated on mini-batch of 128 training examples. A separate set of validation examples is used to measure the generalization performance while training to prevent over-fitting using the early-stopping technique. All these are implemented in the deep learning framework of PyTorch [109] with GPU acceleration. The size of the training data set and the validation data set are 300k and 100k, respectively.
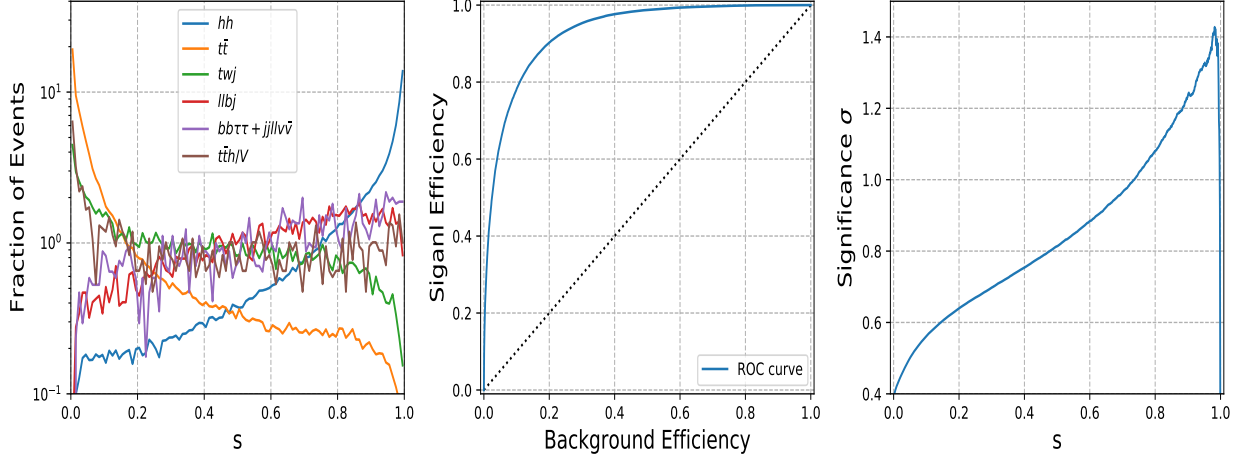
FIG. 4. The MPNN training results for the signal (hh) and backgrounds ($t\bar{t}$, tWj, $\ell^+\ell^-bj$, $b\bar{b}\tau\tau +$ $jj\ell^+\ell^-\nu\bar{\nu}$ and $t\bar{t}h/V$) in the SM at 14 TeV LHC. *Left panel:* the event fractions of signal and each background versus the final score s. *Middle panel:* the ROC curve of signal and background. *Right panel:* the signal significance versus the final score s. The luminosity $\mathcal{L} = 3000$ fb^{-1} is assumed.

IV.   RESULTS AND DISCUSSIONS

In order to estimate the observability of the signal, we calculate the signal significance (α) with the following formula,

$$\alpha = S/\sqrt{B + (\beta B)^2}, \tag{11}$$

where S and B denote number of signal and background events after our selections, respectively. $\mathcal{L}$ is the integrated luminosity of the collider. It should be mentioned that the main systematic uncertainty is parameterized by the factor of β in our calculations.

Firstly, we focus on the SM Higgs pair production process $pp \rightarrow hh \rightarrow b\bar{b}WW^* \rightarrow$ $b\bar{b}\ell^+\ell^- + E_T^{miss}$ at 14 TeV LHC with the luminosity of 3000 fb^{-1}. In Fig. 4, we show the output of the trained MPNN evaluated on the validation test. The left panel is the discrimination score s, i.e., the probability distribution in Eq. (10), for the signal and the background processes. We label the signal as "1" and the background as "0" before training. As expected, the signal peaks near the $s = 1$ and dominant background $t\bar{t}$ peaks near the score $s = 0$, which are well separated from each other. For a given value of score, s_0, we can add the signal or background events in the range of $[s_0, 1]$ in the left panel and then obtain the receiver operating characteristic (ROC) curve in the middle panel, where the signal

10

and background efficiencies are the fraction of the survival events in the initial signal and background events, respectively. We can see that the ROC curve increase steeply and show a good discrimination in the signal and background. The right panel shows the significance of signal as a function of score. Unfortunately, the maximum value of the significance for the SM Higgs pair process $pp \to hh \to b\bar{b}WW^* \to b\bar{b}\ell^+\ell^- + E_T^{miss}$ can only reach about 1.4σ at the HL-LHC.

TABLE I. The sensitivity of the SM signal process $pp \to hh \to b\bar{b}WW^* \to b\bar{b}\ell^+\ell^- + E_T^{miss}$ for MPNN, DNN and CNN at 14 TeV LHC with the luminosity of 3000 fb^{-1} are given. We require the events number of the signal $N_{\rm sig} = 20$ to have reasonable statistic. The significance α is calculated by using the Eq. 11 with $\beta = 0$ for simplicity. The cross sections of the signal and backgrounds are in the unit of fb.

	hh	$t\bar{t}$	$tW+j$	$\ell^+\ell^- bj$	$t\bar{t}h$	$\tau^+\tau^- b\bar{b}$	$t\bar{t}V$	$jj\ell^+\ell^-\nu\bar{\nu}$	$\alpha(\sigma)$	S/B
MPNN	0.0067	0.0553	0.0194	0.0161	0.0075	0.0040	0.0021	0.0020	1.12	0.06
CNN	0.00667	0.1855	0.0626	0.0243	0.0147	0.0128	0.00731	0.00786	0.65	0.021
DNN	0.00668	0.0676	0.0376	0.0163	0.0109	0.00689	0.00454	0.00418	0.95	0.045
(C+D)NN	0.00668	0.0602	0.0299	0.0133	0.00914	0.00689	0.00252	0.00344	1.03	0.053

In Table I, we compare the sensitivity of the SM signal process $pp \to hh \to b\bar{b}WW^* \to b\bar{b}\ell^+\ell^- + E_T^{miss}$ for MPNN, DNN and CNN [2] at 14 TeV LHC with the luminosity of 3000 fb^{-1}. In order to guarantee the statistic, we require to have 20 signal events after all selections for each method. We can see that the MPNN can reduce backgrounds more efficiently than DNN, CNN and other methods [110]. The signal significance given by MPNN is about 1.12σ, which is much better than that of DNN and CNN and comparable with their combined result. Besides, the similar conclusion can be seen in the values of ratio of S/B. Therefore, this channel can be an important combining channel to enhance the sensitivity of measuring the SM Higgs pair process at the HL-LHC.

In Fig. 5, we show the cross section of the signal process $pp \to hh \to b\bar{b}WW^* \to b\bar{b}\ell^+\ell^- + E_T^{miss}$ as a function of the ratio of λ_3/λ_3^{SM} for parton-level and baseline cuts at the HL-LHC. The difference of these two curves can be understood from the calculations of

---

[2] The results of DNN and CNN are taken from Ref. [53], which are also optimized by the Adam optimizer with a learning rate of 0.1%.
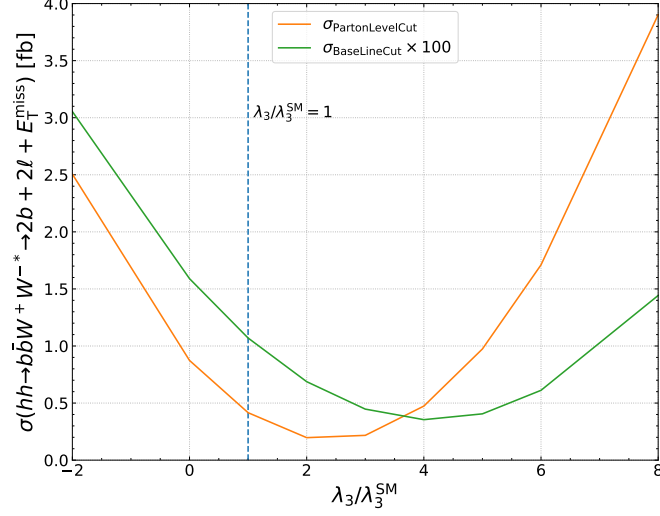
FIG. 5. The cross section of signal versus the ratio of λ_3/λ_3^{SM} after parton level cuts and baseline cuts.

partonic cross section of Higgs pair production, which is approximately given by,

$$\sigma_{gg\to hh}(\hat{s}) \sim \frac{1}{\hat{s}^2} \int d\hat{t}(|F_1|^2 + |F_2|^2) \sim c_\triangle \kappa_3^2 + c_{\triangle,\square} \kappa_3 + c_\square \tag{12}$$

where $\hat{s}$ is the center-of-mass energy of the gluon pair, and $\kappa_3 = \lambda_3/\lambda_3^{SM}$. F_1 is the parity-even contribution part of the triangle and box diagrams, which can be expressed as $F_1 = \kappa_3 F_\triangle + F_\square$. F_2 is the parity-odd contribution of the box diagrams. Thus the Higgs pair production cross section is a quadratic function of κ_3, and its coefficients rely on the triangle and box diagrams. The minimum value of the cross section after the baseline cuts moves to a larger κ_3 value, because the cuts more severely suppress the triangle-diagram process whose cross section is proportional to κ_3^2. The di-Higgs in the triangle-diagram process produced through the s-channel off-shell Higgs are not as energetic as in the box diagram, which results in lower transverse momentum of the final state particles. The baseline cuts mainly select events with boosted Higgs from the box-diagram process. In all, when the value of $c_\triangle/c_{\triangle,\square}$ becomes larger, the minimum will move to the larger κ_3 direction.

Finally, we apply our method to constrain the production cross section of the Higgs pair and the Higgs trilinear coupling in the BSM at 14 TeV LHC. We adopt the model-independent way to present the 2σ limits on the ratio of $\sigma_{hh}/\sigma_{hh}^{SM}$ in the left panel of Fig. 6, where we take the systematic uncertainty $\beta = 0, 10\%, 20\%, 30\%$ for example. It can bee
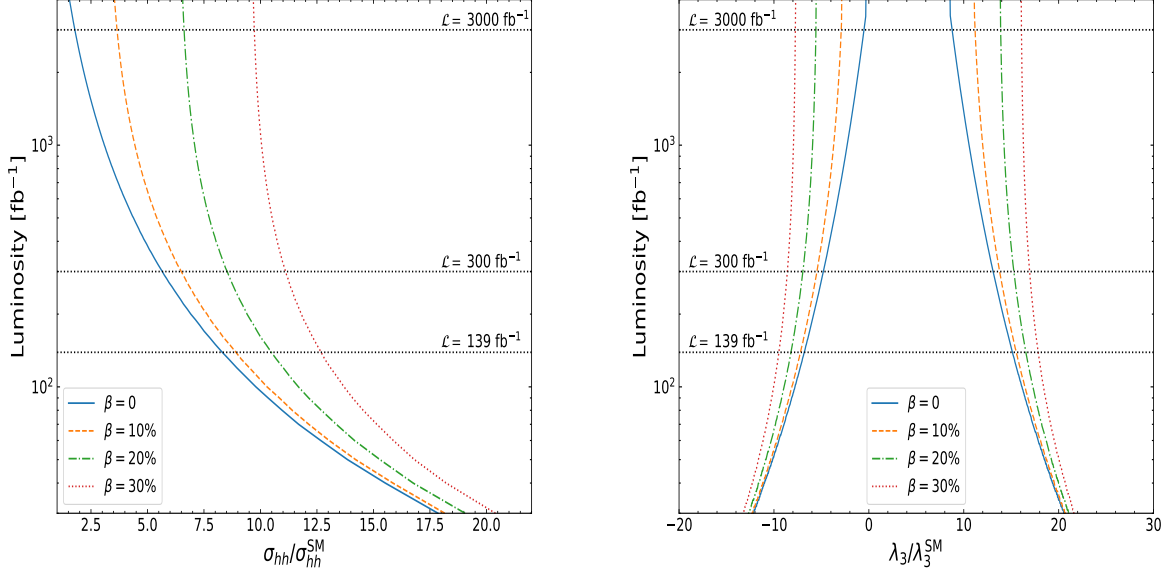
FIG. 6. The 2σ upper bounds on production cross section of the Higgs pair (*left panel*) and triple Higgs coupling (*right panel*) at 14 TeV LHC.

seen that the production cross section of the Higgs pair larger than 13.5 times of the SM prediction can be excluded for the luminosity $\mathcal{L} = 139$ fb^{-1} and systematic error $\beta = 30\%$. If β can be controlled at 10%, the 2σ upper bound on the ratio of $\sigma_{hh}/\sigma_{hh}^{SM}$ will be reduced to 9.5. Such results can be improved to be 10.2 for $\beta = 30\%$ and 3.7 for $\beta = 10\%$ at the HL-LHC. Provided $\beta = 0$, this limit on $\sigma_{hh}/\sigma_{hh}^{SM}$ will become 1.5, which is stronger than 3.3 predicted by DNN [111]. Besides, we reinterpret these bounds for triple Higgs coupling in the right panel of Fig. 6. We find that the ratio of $\lambda_{3h}/\lambda_{3h}^{SM}$ can be constrained to the range of $[-10, 18]$ for $\mathcal{L} = 139$ fb^{-1} and $\beta = 30\%$, and will be further narrowed down to the range of $[-3, 11.5]$ for $\mathcal{L} = 3000$ fb^{-1} and $\beta = 10\%$ at 2σ level.

V. CONCLUSIONS

In this paper, we explored the discovery potential of Higgs pair production process $pp \to hh \to bbWW^* \to 2b + 2\ell + E_T^{\mathrm{miss}}$ with the Message Passing Neural Network at the (HL-)LHC. In the MPNN, we can represent each collision event as an event graph that consists of the final state objects, and use the supervised learning to optimize training parameters. In

contrast with other neural networks, such as DNN, the MPNN may be more flexible because it is independent on the number and ordering of final state objects. By using the MPNN, we obtained that the significance of the SM Higgs pair production process can reach about 1.4σ at the HL-LHC, which is better than the individual results from DNN and CNN. Then, we extended our study to constrain the production cross section of the non-resonant Higgs pair and the triple Higgs trilinear coupling in a mode-independent way. We found that the production cross section of the Higgs pair larger than 10.2 times of the SM prediction can be excluded at 2σ level for the HL-LHC when a 30% systematic uncertainty is included. If the systematic error can be well controlled, such as 10%, this upper bound can be improved to 3.7 times of the predicted by the SM, which will constrain the triple Higgs coupling to the range of $[-3, 11.5]$. Therefore, we expect this channel can play an important role in enhancing the sensitivity of the combining analysis of SM Higgs pair production at the HL-LHC .

[1] G. Aad *et al.* (ATLAS), Phys. Lett. **B716**, 1 (2012), arXiv:1207.7214 [hep-ex].

[2] S. Chatrchyan *et al.* (CMS), Phys. Lett. **B716**, 30 (2012), arXiv:1207.7235 [hep-ex].

[3] J. Baglio, A. Djouadi, R. Grber, M. Mhlleitner, J. Quevillon, and M. Spira, JHEP **04**, 151 (2013), arXiv:1212.5581 [hep-ph].

[4] M. J. Dolan, C. Englert, and M. Spannowsky, JHEP **10**, 112 (2012), arXiv:1206.5001 [hep-ph].

[5] C. Englert, A. Freitas, M. Mhlleitner, T. Plehn, M. Rauch, M. Spira, and K. Walz, J. Phys. G **41**, 113001 (2014), arXiv:1403.7191 [hep-ph].

[6] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964), [,157(1964)].

[7] P. W. Higgs, Phys. Rev. Lett. **13**, 508 (1964), [,160(1964)].

[8] P. W. Higgs, Phys. Rev. **145**, 1156 (1966).

[9] T. W. B. Kibble, Phys. Rev. **155**, 1554 (1967), [,165(1967)].

[10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Phys. Rev. Lett. **13**, 585 (1964), [,162(1964)].

[11] S. Kanemura, S. Kiyoura, Y. Okada, E. Senaha, and C. Yuan, Phys. Lett. B **558**, 157 (2003), arXiv:hep-ph/0211308.

[12] S. Kanemura, Y. Okada, E. Senaha, and C.-P. Yuan, Phys. Rev. D **70**, 115002 (2004), arXiv:hep-ph/0408364.

[13] F. Arco, S. Heinemeyer, and M. Herrero, (2020), arXiv:2005.10576 [hep-ph].

[14] A. Dobado, M. J. Herrero, W. Hollik, and S. Penaranda, Phys. Rev. D **66**, 095016 (2002), arXiv:hep-ph/0208014.

[15] M. Brucherseifer, R. Gavin, and M. Spira, Phys. Rev. D **90**, 117701 (2014), arXiv:1309.3140 [hep-ph].

[16] D. T. Nhung, M. Muhlleitner, J. Streicher, and K. Walz, JHEP **11**, 181 (2013), arXiv:1306.3926 [hep-ph].

[17] L. Wu, J. M. Yang, C.-P. Yuan, and M. Zhang, Phys. Lett. B **747**, 378 (2015), arXiv:1504.06932 [hep-ph].

[18] G. Degrassi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori, and A. Strumia, JHEP **08**, 098 (2012), arXiv:1205.6497 [hep-ph].

[19] A. Kobakhidze, L. Wu, and J. Yue, JHEP **04**, 011 (2016), arXiv:1512.08922 [hep-ph].

[20] F. P. Huang, P.-H. Gu, P.-F. Yin, Z.-H. Yu, and X. Zhang, Phys. Rev. D **93**, 103515 (2016), arXiv:1511.03969 [hep-ph].

[21] A. Efrati and Y. Nir, (2014), arXiv:1401.0935 [hep-ph].

[22] M. McCullough, Phys. Rev. D **90**, 015001 (2014), [Erratum: Phys.Rev.D 92, 039903 (2015)], arXiv:1312.3322 [hep-ph].

[23] M. Gorbahn and U. Haisch, JHEP **10**, 094 (2016), arXiv:1607.03773 [hep-ph].

[24] F. Maltoni, D. Pagani, A. Shivaji, and X. Zhao, Eur. Phys. J. C **77**, 887 (2017), arXiv:1709.08649 [hep-ph].

[25] G. D. Kribs, A. Maier, H. Rzehak, M. Spannowsky, and P. Waite, Phys. Rev. D **95**, 093004 (2017), arXiv:1702.07678 [hep-ph].

[26] *Constraint of the Higgs boson self-coupling from Higgs boson differential production and decay measurements*, Tech. Rep. ATL-PHYS-PUB-2019-009 (CERN, Geneva, 2019).

[27] L.-B. Chen, H. T. Li, H.-S. Shao, and J. Wang, JHEP **03**, 072 (2020), arXiv:1912.13001 [hep-ph].

[28] M. J. Dolan, C. Englert, and M. Spannowsky, Phys. Rev. D **87**, 055002 (2013), arXiv:1210.8166 [hep-ph].

[29] T. Abe, N. Chen, and H.-J. He, JHEP **01**, 082 (2013), arXiv:1207.4103 [hep-ph].

[30] C. Han, X. Ji, L. Wu, P. Wu, and J. M. Yang, JHEP **04**, 003 (2014), arXiv:1307.3790 [hep-ph].

[31] X.-F. Wang, C. Du, and H.-J. He, Phys. Lett. B **723**, 314 (2013), arXiv:1304.2257 [hep-ph].

[32] B. Hespel, D. Lopez-Val, and E. Vryonidou, JHEP **09**, 124 (2014), arXiv:1407.0281 [hep-ph].

[33] S. Dawson, A. Ismail, and I. Low, Phys. Rev. D **91**, 115008 (2015), arXiv:1504.05596 [hep-ph].

[34] A. Kobakhidze, N. Liu, L. Wu, and J. Yue, Phys. Rev. D **95**, 015016 (2017), arXiv:1610.06676 [hep-ph].

[35] L.-C. L, C. Du, Y. Fang, H.-J. He, and H. Zhang, Phys. Lett. B **755**, 509 (2016), arXiv:1507.02644 [hep-ph].

[36] J. Ren, R.-Q. Xiao, M. Zhou, Y. Fang, H.-J. He, and W. Yao, JHEP **06**, 090 (2018), arXiv:1706.05980 [hep-ph].

[37] S. Borowka, C. Duhr, F. Maltoni, D. Pagani, A. Shivaji, and X. Zhao, JHEP **04**, 016 (2019), arXiv:1811.12366 [hep-ph].

[38] L. Wu, H. Zhang, and B. Zhu, JCAP **07**, 033 (2019), arXiv:1901.06532 [hep-ph].

[39] A. Alves, T. Ghosh, H.-K. Guo, and K. Sinha, JHEP **12**, 070 (2018), arXiv:1808.08974 [hep-ph].

[40] A. J. Barr, M. J. Dolan, C. Englert, and M. Spannowsky, Phys. Lett. B **728**, 308 (2014), arXiv:1309.6318 [hep-ph].

[41] Q. Li, Q.-S. Yan, and X. Zhao, Phys. Rev. D **89**, 033015 (2014), arXiv:1312.3830 [hep-ph].

[42] Q.-H. Cao, B. Yan, D.-M. Zhang, and H. Zhang, Phys. Lett. B **752**, 285 (2016), arXiv:1508.06512 [hep-ph].

[43] Q. Li, Z. Li, Q.-S. Yan, and X. Zhao, Phys. Rev. D **92**, 014015 (2015), arXiv:1503.07611 [hep-ph].

[44] Q.-H. Cao, Y. Liu, and B. Yan, Phys. Rev. D **95**, 073006 (2017), arXiv:1511.03311 [hep-ph].

[45] Q.-H. Cao, G. Li, B. Yan, D.-M. Zhang, and H. Zhang, Phys. Rev. D **96**, 095031 (2017), arXiv:1611.09336 [hep-ph].

[46] H.-J. He, J. Ren, and W. Yao, Phys. Rev. D **93**, 015003 (2016), arXiv:1506.03302 [hep-ph].

[47] T. Huang, J. M. No, L. Perni, M. Ramsey-Musolf, A. Safonov, M. Spannowsky, and P. Winslow, Phys. Rev. **D96**, 035007 (2017), arXiv:1701.04442 [hep-ph].

[48] C.-T. Lu, J. Chang, K. Cheung, and J. S. Lee, JHEP **08**, 133 (2015), arXiv:1505.00957 [hep-ph].

[49] J. Chang, K. Cheung, J. S. Lee, C.-T. Lu, and J. Park, Phys. Rev. D **100**, 096001 (2019), arXiv:1804.07130 [hep-ph].

[50] A. Papaefstathiou, L. L. Yang, and J. Zurita, Phys. Rev. **D87**, 011301 (2013), arXiv:1209.1489 [hep-ph].

[51] J. H. Kim, K. Kong, K. T. Matchev, and M. Park, Phys. Rev. Lett. **122**, 091801 (2019), arXiv:1807.11498 [hep-ph].

[52] G. Buchalla, M. Capozi, A. Celis, G. Heinrich, and L. Scyboz, JHEP **09**, 057 (2018), arXiv:1806.05162 [hep-ph].

[53] J. H. Kim, M. Kim, K. Kong, K. T. Matchev, and M. Park, JHEP **09**, 047 (2019), arXiv:1904.08549 [hep-ph].

[54] G. Li, L.-X. Xu, B. Yan, and C.-P. Yuan, Phys. Lett. B **800**, 135070 (2020), arXiv:1904.12006 [hep-ph].

[55] M. Cepeda *et al.*, "Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC," in *Report on the Physics at the HL-LHC,and Perspectives for the HE-LHC*, Vol. 7, edited by A. Dainese, M. Mangano, A. B. Meyer, A. Nisati, G. Salam, and M. A. Vesterinen (2019) pp. 221–584, arXiv:1902.00134 [hep-ph].

[56] R. Contino *et al.*, CERN Yellow Rep. , 255 (2017), arXiv:1606.09408 [hep-ph].

[57] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, Nucl. Instrum. Meth. **A543**, 577 (2005), arXiv:physics/0408124 [physics].

[58] P. Baldi, P. Sadowski, and D. Whiteson, Nature Commun. **5**, 4308 (2014), arXiv:1402.4735 [hep-ph].

[59] P. Baldi, P. Sadowski, and D. Whiteson, Phys. Rev. Lett. **114**, 111801 (2015), arXiv:1410.3469 [hep-ph].

[60] M. Bridges, K. Cranmer, F. Feroz, M. Hobson, R. Ruiz de Austri, and R. Trotta, JHEP **03**, 012 (2011), arXiv:1011.4306 [hep-ph].

[61] A. Buckley, A. Shilton, and M. J. White, Comput. Phys. Commun. **183**, 960 (2012), arXiv:1106.4613 [hep-ph].

[62] N. Bornhauser and M. Drees, Phys. Rev. **D88**, 075016 (2013), arXiv:1307.3383 [hep-ph].

[63] S. Caron, J. S. Kim, K. Rolbiecki, R. Ruiz de Austri, and B. Stienen, Eur. Phys. J. **C77**, 257 (2017), arXiv:1605.02797 [hep-ph].

[64] G. Bertone, M. P. Deisenroth, J. S. Kim, S. Liem, R. Ruiz de Austri, and M. Welling, Phys. Dark Univ. **24**, 100293 (2019), arXiv:1611.02704 [hep-ph].

[65] M. Abdughani, J. Ren, L. Wu, J. M. Yang, and J. Zhao, Commun. Theor. Phys. **71**, 955 (2019), arXiv:1905.06047 [hep-ph].

[66] P. C. Bhat, Ann. Rev. Nucl. Part. Sci. **61**, 281 (2011).

[67] J. Ren, L. Wu, J. M. Yang, and J. Zhao, Nucl. Phys. B **943**, 114613 (2019), arXiv:1708.06615 [hep-ph].

[68] S. H. Lim and M. M. Nojiri, JHEP **10**, 181 (2018), arXiv:1807.03312 [hep-ph].

[69] J. Amacker *et al.*, (2020), arXiv:2004.04240 [hep-ph].

[70] J. Hajer, Y.-Y. Li, T. Liu, and H. Wang, Phys. Rev. D **101**, 076015 (2020), arXiv:1807.10261 [hep-ph].

[71] A. Andreassen, B. Nachman, and D. Shih, Phys. Rev. D **101**, 095004 (2020), arXiv:2001.05001 [hep-ph].

[72] F. Bishara and M. Montull, (2019), arXiv:1912.11055 [hep-ph].

[73] L. Li, Y.-Y. Li, T. Liu, and S.-J. Xu, (2020), arXiv:2004.15013 [hep-ph].

[74] V. Mikuni and F. Canelli, (2020), arXiv:2001.05311 [physics.data-an].

[75] A. Mullin, H. Pacey, M. Parker, M. White, and S. Williams, (2019), arXiv:1912.10625 [hep-ph].

[76] C. Jin, S.-z. Chen, and H.-H. He, (2019), arXiv:1910.07160 [astro-ph.IM].

[77] E. A. Moreno, T. Q. Nguyen, J.-R. Vlimant, O. Cerri, H. B. Newman, A. Periwal, M. Spiropulu, J. M. Duarte, and M. Pierini, (2019), arXiv:1909.12285 [hep-ex].

[78] E. A. Moreno, O. Cerri, J. M. Duarte, H. B. Newman, T. Q. Nguyen, A. Periwal, M. Pierini, A. Serikova, M. Spiropulu, and J.-R. Vlimant, Eur. Phys. J. C **80**, 58 (2020), arXiv:1908.05318 [hep-ex].

[79] S. Jung, D. Lee, and K.-P. Xie, Eur. Phys. J. C **80**, 105 (2020), arXiv:1906.02810 [hep-ph].

[80] B. Bhattacherjee, S. Mukherjee, and R. Sengupta, JHEP **19**, 156 (2020), arXiv:1904.04811 [hep-ph].

[81] S. R. Qasim, J. Kieseler, Y. Iiyama, and M. Pierini, Eur. Phys. J. C **79**, 608 (2019), arXiv:1902.07987 [physics.data-an].

[82] J. Arjona Martnez, O. Cerri, M. Pierini, M. Spiropulu, and J.-R. Vlimant, Eur. Phys. J. Plus **134**, 333 (2019), arXiv:1810.07988 [hep-ph].

[83] P. T. Komiske, E. M. Metodiev, and J. Thaler, JHEP **01**, 121 (2019), arXiv:1810.05165 [hep-ph].

[84] M. Farina, Y. Nakai, and D. Shih, Phys. Rev. D **101**, 075021 (2020), arXiv:1808.08992 [hep-ph].

[85] E. Bothmann and L. Debbio, JHEP **01**, 033 (2019), arXiv:1808.07802 [hep-ph].

[86] J. Lin, M. Freytsis, I. Moult, and B. Nachman, JHEP **10**, 101 (2018), arXiv:1807.10768 [hep-ph].

[87] F. Staub, (2019), arXiv:1906.03277 [hep-ph].

[88] T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson, SciPost Phys. **6**, 030 (2019), arXiv:1808.08979 [hep-ph].

[89] H. Luo, M.-x. Luo, K. Wang, T. Xu, and G. Zhu, Sci. China Phys. Mech. Astron. **62**, 10.1007/s11433-019-9390-8, arXiv:1712.03634 [hep-ph].

[90] A. Chakraborty, S. H. Lim, M. M. Nojiri, and M. Takeuchi, (2020), arXiv:2003.11787 [hep-ph].

[91] M. Capozi and G. Heinrich, JHEP **03**, 091 (2020), arXiv:1908.08923 [hep-ph].

[92] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, CoRR **abs/1704.01212** (2017), arXiv:1704.01212.

[93] M. Gori, G. Monfardini, and F. Scarselli, in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, Vol. 2 (2005) pp. 729–734 vol. 2.

[94] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, IEEE Transactions on Neural Networks **20**, 61 (2009).

[95] I. Henrion, J. Brehmer, J. Bruna, K. Cho, K. Cranmer, G. Louppe, and G. Rochette (2017).

[96] J. Ren, L. Wu, and J. M. Yang, Phys. Lett. B **802**, 135198 (2020), arXiv:1901.05627 [hep-ph].

[97] M. Abdughani, J. Ren, L. Wu, and J. M. Yang, JHEP **08**, 055 (2019), arXiv:1807.09088 [hep-ph].

[98] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, Journal of High Energy Physics **2014** (2014), 10.1007/jhep07(2014)079.

[99] R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo (NNPDF), Nucl. Phys. **B877**, 290 (2013), arXiv:1308.0598 [hep-ph].

[100] J. Grigo, K. Melnikov, and M. Steinhauser, Nucl. Phys. **B888**, 17 (2014), arXiv:1408.2422 [hep-ph].

[101] M. Czakon, P. Fiedler, and A. Mitov, Phys. Rev. Lett. **110**, 252004 (2013), arXiv:1303.6254 [hep-ph].

[102] T. Sjstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, Comput. Phys. Commun. **191**, 159 (2015), arXiv:1410.3012 [hep-ph].

[103] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, and M. Selvaggi (DELPHES 3), JHEP **02**, 057 (2014), arXiv:1307.6346 [hep-ex].

[104] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. **C72**, 1896 (2012), arXiv:1111.6097 [hep-ph].

[105] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063 (2008), arXiv:0802.1189 [hep-ph].

[106] *Expected performance of the ATLAS detector at the High-Luminosity LHC*, Tech. Rep. ATL-PHYS-PUB-2019-005 (CERN, Geneva, 2019).

[107] M. Aaboud *et al.* (ATLAS), Eur. Phys. J. **C78**, 903 (2018), arXiv:1802.08168 [hep-ex].

[108] D. P. Kingma and J. Ba, (2014), arXiv:1412.6980 [cs.LG].

[109] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alche-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019) pp. 8024–8035.

[110] A. Adhikary, S. Banerjee, R. K. Barman, B. Bhattacherjee, and S. Niyogi, JHEP **07**, 116 (2018), arXiv:1712.05346 [hep-ph].

[111] *Prospects for HH measurements at the HL-LHC*, Tech. Rep. CMS-PAS-FTR-18-019 (CERN, Geneva, 2018).