

Probing the triple Higgs boson coupling with machine learning at the LHC

Murat Abdughani^{1,‡}, Daohan Wang^{2,3}, Lei Wu^{4,*}, Jin Min Yang^{2,3}, and Jun Zhao^{2,3,†}

¹Key Laboratory of Dark Matter and Space Astronomy, Purple Mountain Observatory,
Chinese Academy of Sciences, Nanjing 210023, China

²CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics,
Chinese Academy of Sciences, Beijing 100190, China

³School of Physics and Electronics, Henan University, Kaifeng 475001, Henan, China

⁴Department of Physics and Institute of Theoretical Physics,
Nanjing Normal University, Nanjing 210023, China



(Received 17 April 2021; accepted 2 August 2021; published 3 September 2021)

Measuring the triple Higgs boson coupling is a crucial task in the LHC and future collider experiments. We apply the message passing neural network (MPNN) to the study of nonresonant Higgs pair production process $pp \rightarrow hh$ in the final state with $2b + 2\ell + E_T^{\text{miss}}$ 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 Standard Model cross section at the LHC with the luminosity of 3000 fb^{-1} , which will limit the triple Higgs boson coupling to the range of $[-3, 11.5]$.

DOI: [10.1103/PhysRevD.104.056003](https://doi.org/10.1103/PhysRevD.104.056003)

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 reveals the electroweak symmetry breaking mechanism and sheds light 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., [3–17]).

In the Brout-Englert-Higgs mechanism of electroweak symmetry breaking [18–22], the Higgs boson is a massive scalar with self-interactions. The Higgs self-couplings are determined by the structure of the scalar potential,

$$V = \frac{m_h^2}{2} h^2 + \lambda_3 v h^3 + \frac{1}{4} \lambda_4 h^4, \quad (1)$$

where m_h is the mass of the SM Higgs boson, and v is the vacuum expectation value of the SM Higgs field. The λ_3 and λ_4 are the Higgs self-couplings, and the corresponding SM values are

$$\lambda_3^{\text{SM}} = \lambda_4^{\text{SM}} = \frac{m_h^2}{2v^2}. \quad (2)$$

The values of λ_3 and λ_4 are measured via the double and triple Higgs production processes, 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 [23–25] and (next-to-)minimal supersymmetric Standard Model [26–29]. Since the Higgs self-coupling plays an important role in vacuum stability [30] and electroweak baryogenesis [31,32], measuring the Higgs self-coupling will provide a crucial clue to new physics [33].

The triple Higgs boson coupling can be indirectly probed by using the loop effects in some observables, for example, the single Higgs production [34–36] and the electroweak precision observables [37]. With 80 fb^{-1} of the LHC Run-2 data, the triple Higgs boson coupling has been constrained in the range $-3.2 < \lambda_3/\lambda_3^{\text{SM}} < 11.9$ at 95% C.L. [38]. On the other hand, the Higgs pair production provides a direct way to measure the triple Higgs boson 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 boson coupling, and the other is from the box diagram mediated

* Corresponding author.
leiwu@njnu.edu.cn
† zhaojun@itp.ac.cn
‡ mulati@pmo.ac.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

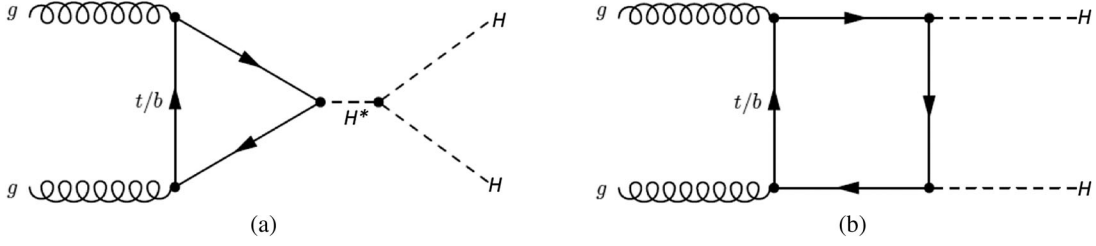


FIG. 1. Representative Feynman diagrams for the Higgs boson pair production in the SM: (a) through triple Higgs self-coupling; (b) through Higgs-fermion Yukawa interaction.

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 and including finite top quark mass effects [39]. The new physics effects that can significantly modify the Higgs pair production have been intensively studied at the LHC (see, for examples, [40–51] and references therein).

In Refs. [9,10,52–68], the potential of measuring the Higgs pair production has been investigated in various decay modes: $bbb\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}WW^*$, $\gamma\gamma b\bar{b}$, $\gamma\gamma WW^*$, 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 $-5 < \lambda_3/\lambda_3^{\text{SM}} < 12$ is constrained at 95% C.L. at 13 TeV LHC with the luminosity of 36.1 fb^{-1} [69]. The sensitivity will be greatly improved at the high-luminosity (HL)-LHC [70] and future hadron colliders [71].

In this paper, we focus on the Higgs pair production at the HL-LHC with 3 ab^{-1} luminosity, where one Higgs decays to $b\bar{b}$ and the other to WW^* . The decay branching ratio of $hh \rightarrow WW^*$ is the second largest after $hh \rightarrow b\bar{b}$, so the $b\bar{b}WW^*$ final state can thus act as an important channel to enhance the combining result if the signal can be well separated from the dominant $t\bar{t}$ background. Earlier conventional cut-flow analyses [59,62,68] and machine learning methods [67,72–74] applied to this channel get no more than 1σ significance, while recent work [65] has combined the deep neural network (DNN) and convolutional neural network methods to reach the significance of 1σ . Given the importance of the Higgs pair production, in this work we apply the machine learning method message passing neural network (MPNN) [75] to explore the potential of observing such di-Higgs events through the channel $pp \rightarrow hh \rightarrow b\bar{b}WW^*$.

In addition to the conventional kinematic cut-flow analyses, the machine learning methods have been proposed to accelerate the discovery of new physics [76–110]. The MPNN framework inherits the generality and powerfulness of the graph neural network [111,112]. It abstracts the commonalities between several of the most popular models for graph-structured data, such as spectral approaches [113–115] and nonspectral approaches [116]

in graph convolution, gated graph neural networks [117], interaction networks [118], molecular graph convolutions [119], deep tensor neural networks [120], and so on [121]. In the MPNN, a collision event is represented as a numerical geometrical graph formed by a number of final state objects, which are nonlinear models with a bunch of parameters that relate 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 the DNN, MPNN is a dynamic neural network and independent of the number and ordering of final state particles. Therefore, the MPNN is suitable for processing the graph representation of the collision event. Recently, this method has been successfully applied to collider phenomenological studies, such as jet physics [122], Higgs physics [123], and supersymmetry [124].

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

II. EVENT GENERATION AND RECONSTRUCTION

The signal and background events at parton level are generated with MadGraph5_aMC@NLO v2.6.1 [125] with the default parton distribution function (PDF) set NNPDF2.3QED [126] at the LHC with leading order with center-of-mass energy $\sqrt{s} = 14 \text{ TeV}$. We employ the following cuts for parton level event generation: $p_{Tj} > 20 \text{ GeV}$, $p_{Tb} > 20 \text{ GeV}$, $p_{T\gamma} > 10 \text{ GeV}$, $p_{T\ell} > 10 \text{ GeV}$, $\eta_j < 5$, $\eta_b < 5$, $\eta_\gamma < 2.5$, $\eta_\ell < 2.5$, $\Delta R_{bb} < 1.8$, $\Delta R_{\ell\ell} < 1.3$, $70 \text{ GeV} < m_{jj} < 160 \text{ GeV}$, $70 < m_{bb} < 160 \text{ GeV}$, and $m_{\ell\ell} < 75 \text{ GeV}$, where ℓ denotes e and μ . We impose additionally $5 \text{ GeV} < m_{\ell\ell} < 75 \text{ GeV}$ for $jj\ell\ell\nu\bar{\nu}$, $\ell\ell bj$ and $tW + j$ backgrounds. The angular distance ΔR_{ij} is defined by

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

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.

TABLE I. Signal and background cross sections in fb unit before hadron-level cuts but after baseline cuts and after the MPNN validation process requiring the signal events number $N_{\text{sig}} = 20$ to have reasonable statistics. The significance α is calculated by using the Eq. (11) with $\beta = 0$ for simplicity.

	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
No cut	40.7 [127]	953600 [128]	123200	117100 [129]	661.3 [130]	29070 [129]	1710 [131]	48200 _a	$\simeq 0$	$\simeq 0$
Baseline cuts	0.0105	1.8568	0.2189	0.0675	0.0247	0.0246	0.0153	0.0101	0.3876	0.0047
MPNN	0.0067	0.0581	0.0180	0.0152	0.0080	0.0025	0.0018	0.0017	1.13	0.06

^aApplied a NLO k factor of 2.0.

The signal cross section is normalized to the next-to-next-to-leading-order (NNLO) accuracy in QCD [127], that is $\sigma_{gg \rightarrow hh} = 40.7$ fb. The main background $t\bar{t}$ cross section is normalized to the NNLO QCD value 953.6 pb [128]. Along with the signal and $t\bar{t}$, all other backgrounds and their normalized cross sections are listed in Table I.

We generate the low- Q^2 soft QCD pile-up events and apply hadronization via package PYTHIA8243 [132], followed by detector simulation with DELPHES 3.4.2 [133]. In the ATLAS card, we consider the average amount of pile-up events per bunch crossing as 100. We take the default parametrization implemented in the ATLAS card to distribute the hard scattering events and pile-up events randomly in time and z positions. The maximum spread of pile-up events in the beam direction is 0.25 m, and the maximum spread of pile-up events in time is 8×10^{-10} s.

In this work, we follow the default parametrization of the DELPHES ATLAS card to perform the pile-up subtraction and use the spatial vertex resolution parameter $|z|$ to perform charged pile-up subtraction. We consider every charged particle originating from a reconstructed vertex with $|z| > 0.01$ cm as coming from pile-up events and only keep those tracks that pass through the TrackPileUpSubtractor in DELPHES.

Similar to the tracks, the reconstructed jets are supposed to be corrected from low- Q^2 pile-up events containing neutral particles. Jet pile-up subtraction is done via the JetPileUpSubtractor module that takes as input the jet constitutes and pile-up density ρ based on the jet area. This technique helps to correct the jet momenta by calculating pile-up density (ρ) and jet area. Jets are clustered with the calorimeter tower elements using FASTJET 3.3.2 [134] with anti- k_T jet algorithm [135] and jet radius $R = 0.4$ with $p_T > 20$ GeV, and we allow the default estimation of ρ with the calorimeter towers. As for the pile-up subtraction of missing transverse momentum, we calculate it based on the pile-up subtracted jets, photons, and leptons.

The DELPHES card for the ATLAS detector simulation is modified as follows:

- (i) Jets, including b -jets, with $p_T(j) > 20$ GeV and $|\eta_j| < 2.5$ are selected.
- (ii) Flat b -tagging efficiency is $\epsilon_{b \rightarrow b} = 0.75$, mistagging efficiency for c quark as b is $\epsilon_{c \rightarrow b} = 0.1$, and mistagging rates of other jets are $\epsilon_{j \rightarrow b} = 0.01$ [136].

- (iii) Maximum transverse momenta ratio for lepton isolation is set as $\frac{\sum_{i \neq \ell} p_{Ti}}{p_{T\ell}} < 0.15$, where the sum is taken over the transverse momenta p_{Ti} of all final state particles i , $i \neq \ell$, with $p_{Ti} > 0.5$ GeV and within angular distance $\Delta R_{i\ell} < 0.3$ with lepton candidate ℓ . Leptons with $p_T(\ell) > 10$ GeV and $|\eta_\ell| < 2.5$ are selected.
- (iv) Isolation of photons also require $\frac{\sum_{i \neq \gamma} p_{Ti}}{p_{T\gamma}} < 0.12$ 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 required to have $p_T(\gamma) > 25$ GeV and $|\eta_\gamma| < 2.5$ to be selected.

After the reconstruction, the missing transverse momentum $\mathbf{E}_T^{\text{miss}}$ is defined as the negative vector sum of the transverse momenta of the accepted photons, leptons and jets, and unused tracks as in [137],

$$\mathbf{E}_T^{\text{miss}} = - \sum_{\text{accepted electrons}} \mathbf{p}_T^\ell - \sum_{\text{accepted muons}} \mathbf{p}_T^\mu - \sum_{\text{accepted photons}} \mathbf{p}_T^\gamma - \sum_{\text{accepted jets}} \mathbf{p}_T^j - \sum_{\text{unused tracks}} \mathbf{p}_T^{\text{track}}, \quad (4)$$

where the tracks with $p_T > 0.4$ GeV and $|\eta| < 2.5$ are considered.

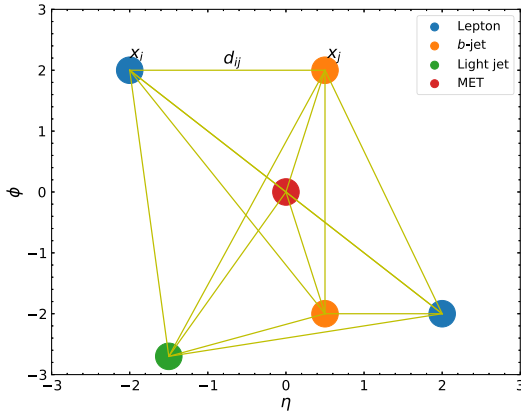
We further apply the following cuts to reduce background events sufficiently relevant to the signals:

- (i) The two leading jets must be b tagged, each with $p_T > 30$ GeV.
- (ii) Exactly two opposite sign leptons, each with $p_T > 25$ GeV.
- (iii) Modulus of $\mathbf{E}_T^{\text{miss}}$ is required to be $E_T^{\text{miss}} > 20$ GeV.
- (iv) Angular distances for two leptons and for two b jets are $\Delta R_{\ell\ell} < 1.0$ and $\Delta R_{bb} < 1.3$, respectively.
- (v) 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 (which 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.

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



x	photon	lepton charge	b -jet or light 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. 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. Note that 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.

network. Figure 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 $x_i = (I_1, I_2, I_3, I_4, p_T, E, m)$, 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, and $I_4 = 1$ 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 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 that it is stable with respect to the rotation of the event around the beam axis. The other two advantages of such an event graph design are the following: (1) The number of nodes equal to the number of final state objects, i.e., the number of nodes is not fixed, which guarantees to use full information of final state objects¹; (2) The node features and edge weights are easily transformed by the four momenta of the object, no sophisticated discriminants

¹We verified the assumption by restricting the number of light jets at the final states and obtained the best result by using full information.

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, which consists of one embedding layer, N message passing layers, and one output layer. The embedding transformation for input data is given by

$$m_i^0 = \text{ReLU}(W_m^0 x_i + b_m^0), \quad (5)$$

where W_m^0 and b_m^0 are learnable weight and bias vectors, and the activation function ReLU is the rectified linear unit [138]. The dimension of m_i^0 is higher than x_i . It can be seen that i th node m_i^0 in the embedding layer is a vector, which only contains information from input feature x_i , without including any geometrical pattern of the event graph. Then, the i th node in the n th message passing layer is obtained by the following transformation:

$$s_i^n = \sum_j \text{ReLU}(W_m^n (m_j^{n-1} \oplus \hat{d}_{ij}) + b_m^n), \quad (6)$$

$$m_i^n = \text{ReLU}(W_s^n (m_i^{n-1} \oplus s_i^n) + b_s^n), \quad (7)$$

where i and j are indices of nodes, s_i^n is the intermediate vector, \oplus represents vector concatenation, and W_s and b_s are learnable weights and biases. The message passing process is realized by two subprocesses: First, Eq. (6) collects information from all previous nodes and distances between nodes. Second, Eq. (7) passes this information, together with previous nodes, to the next one. By repeating this process, each node in the message passing layer gets knowledge of other nodes and relationships between them and updates itself. Therefore, the message-passing

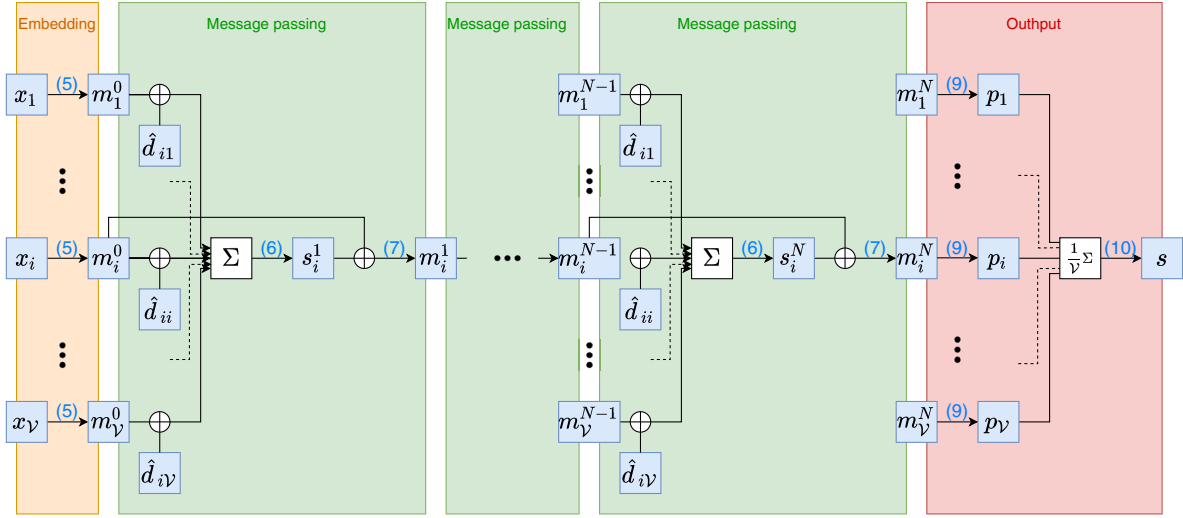


FIG. 3. Schematic diagram of the MPNN classifier which consists of one embedding layer, N pairs of message passing layers, and one output layer (these layers are sequentially connected by nonlinear transformations). The numbers in parentheses correspond to equation numbers in the text, the operators \oplus , Σ and $\frac{1}{\mathcal{V}}$ denote vector concatenation, summation, and averaging, respectively; the summation and averaging run over all \mathcal{V} nodes.

mechanism is the key for automatically extracting features of the input event graph, which efficiently disseminates the information among all the nodes taking into account the connections between nodes. After N iterations, each node state vector can be viewed as an encoding of the whole event graph representing the whole information of both the kinematic features of all final states and the geometrical relationship between them. Here, we expand edge weight d_{ij} onto 21 Gaussian bases to make it more suitable for linear transformation [124], and the k th component of this weight vector is

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

where μ_k is linearly distributed in range of $[0, 5]$ and $\sigma = 0.25$. Such an expansion is inspired by radial basis function networks [139,140], which can solve nonlinear problems by mapping input into high dimensions.

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

$$p_i = \sigma(W_p m_i^N + b_p) = \frac{1}{1 + e^{-(W_p m_i^N + b_p)}}, \quad (9)$$

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

$$s = \frac{1}{\mathcal{V}} \sum p_i, \quad (10)$$

with \mathcal{V} being the number of nodes in the input event, which is the number of final state particles in an event. It should be

mentioned that \mathcal{V} is not a constant, e.g., if there are two extra light jets and one photon in an event apart from the required two b jets, two leptons, and one MET, then we have $\mathcal{V} = 7$.

The MPNN can be efficiently trained using the supervised learning method. We adopt binary cross entropy as the loss function. Although increasing the number of hidden layers can enable the network to learn more complex features in the data, it may have disadvantages like overfitting and time consuming. We find that for our network $N = 3$ is the most optimal choice.² Note that W_m^0 , W_m^n , W_s^n , and $W_p s$ in Eqs. (5)–(7) and (9) are 30×7 , 30×51 , 30×60 , and 1×30 matrices, respectively. Thus, the overall number of learnable parameters in our MPNN model is 10441. The Adam [141] optimizer with a learning rate of 0.001 is used to optimize the model parameters based on the gradients calculated on a minibatch of 128 training examples. A separate set of validation examples is used to measure the generalization performance, while training to prevent overfitting using the early-stopping technique. All these are implemented in the deep learning framework of PYTORCH [142] with a CUDA platform and trained on a NVIDIA Titan Xp GPU with 12 Gb DDR5 memory for acceleration. One cycle of training and validation takes about half an hour when the size of the training data set and the validation data set are 300k and 100k, respectively. Note that signal and backgrounds have equal training and testing samples, while each subbackground has a number of samples proportional to the cross

²Message passing layer with $N = 3$ can increase significance by about 5% compared to $N = 2$, while $N = 4$ can only increase significance by less than 1% compared to $N = 3$.

section after the baseline cuts, e.g., $1.8568/2.2178 \times 150$ K training samples for $t\bar{t}$, $0.2189/2.2178 \times 150$ K for $tW + j$, and so on, where the sum of the cross sections of all backgrounds after the baseline cuts is 2.2178 fb (see Table. I).

IV. RESULTS AND DISCUSSIONS

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

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

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 parametrized by the factor of β in our calculations.

Firstly, we focus on the SM Higgs pair production process $pp \rightarrow hh \rightarrow b\bar{b}WW^* \rightarrow b\bar{b}\ell^+\ell^- + E_T^{\text{miss}}$ 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 $s = 1$ 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 ROC curve in the middle panel, where the signal and background efficiencies are a fraction of the survival events in the initial signal and background events, respectively. We can see that the ROC curve increases steeply and shows 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

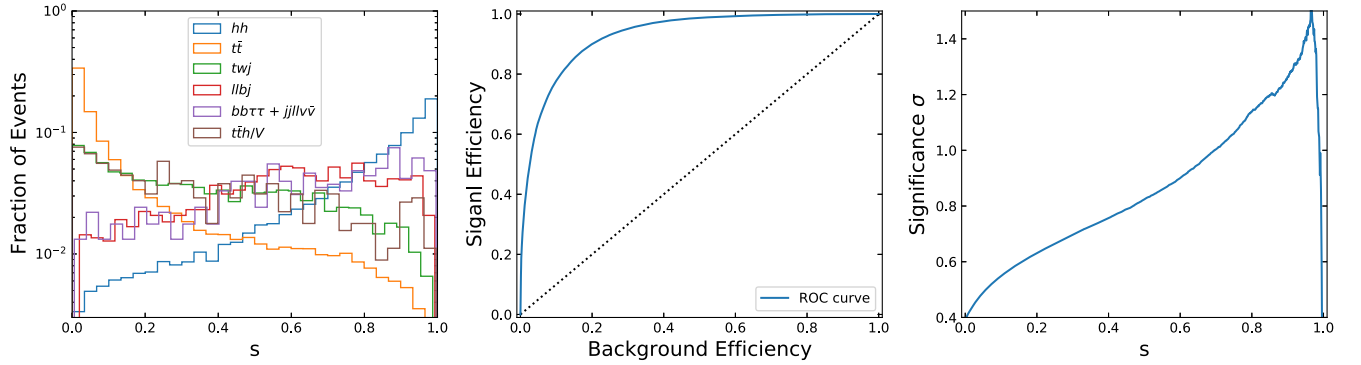


FIG. 4. The MPNN training results for the signal (hh) and backgrounds ($t\bar{t}$, tWj , $\ell^+\ell^-bj$, $b\bar{b}t\bar{t} + jj\ell^+\ell^-\nu\bar{\nu}$, and $t\bar{t}h/V$) in the SM at 14 TeV LHC. Left panel: the event fractions of the signal and each background versus the final score s . Middle panel: the receiver operating characteristic (ROC) curve of the signal and background. Right panel: the signal significance versus the final score s . The luminosity $\mathcal{L} = 3000 \text{ fb}^{-1}$ is assumed.

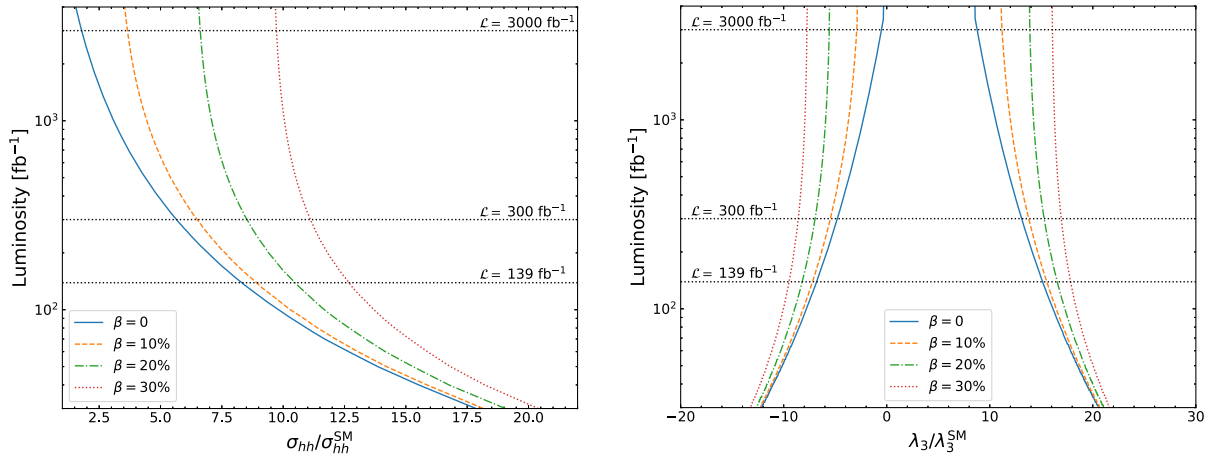


FIG. 5. The 2σ upper bounds on production cross section of the Higgs pair (left panel) and triple Higgs boson coupling (right panel) at 14 TeV LHC.

significance for the SM Higgs pair process $pp \rightarrow hh \rightarrow b\bar{b}WW^* \rightarrow b\bar{b}\ell^+\ell^- + E_T^{\text{miss}}$ can only reach about 1.5σ at the HL-LHC.

At the last row of Table I, we give the sensitivity of the SM signal process $pp \rightarrow hh \rightarrow b\bar{b}WW^* \rightarrow b\bar{b}\ell^+\ell^- + E_T^{\text{miss}}$ for the MPNN 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. The signal significance given by the MPNN is about 1.12σ .

Finally, we apply our method to constrain the production cross section of the Higgs pair and the Higgs trilinear coupling in the beyond SM at 14 TeV LHC. We adopt the model-independent way to present the 2σ limits on the ratio of $\sigma_{hh}/\sigma_{hh}^{\text{SM}}$ in the left panel of Fig. 5, where we take the systematic uncertainty $\beta = 0, 10\%, 20\%, 30\%$, for example. It can be 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 $\mathcal{L} = 139 \text{ fb}^{-1}$ and systematic error $\beta = 30\%$. If β can be controlled at 10%, then the 2σ upper bound on the ratio of $\sigma_{hh}/\sigma_{hh}^{\text{SM}}$ 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 $\sigma_{hh}/\sigma_{hh}^{\text{SM}}$ will become 1.5. Besides, we reinterpret these bounds for triple Higgs boson coupling in the right panel of Fig. 5. We find that the ratio of $\lambda_{3h}/\lambda_{3h}^{\text{SM}}$ can be constrained to the range of $[-10, 18]$ for $\mathcal{L} = 139 \text{ fb}^{-1}$ and $\beta = 30\%$ and will be further narrowed down to the range of $[-3, 11.5]$ for $\mathcal{L} = 3000 \text{ fb}^{-1}$ and $\beta = 10\%$ at 2σ level.

V. CONCLUSIONS

In this paper, we explored the discovery potential of Higgs pair production process $pp \rightarrow hh \rightarrow b\bar{b}WW^* \rightarrow 2b + 2\ell + E_T^{\text{miss}}$ 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. By using the MPNN, we obtained that the significance of the SM Higgs pair production process can reach the maximum of about 1.5σ at the HL-LHC. Then, we extended our study to constrain the production cross section of the nonresonant Higgs pair and the triple Higgs trilinear coupling in a model-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 boson 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.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (NNSFC) under Grants No. 12047560, No. 11705093, No. 12075300, and No. 11851303.

-
- [1] G. Aad *et al.* (ATLAS Collaboration), *Phys. Lett. B* **716**, 1 (2012).
 - [2] S. Chatrchyan *et al.* (CMS Collaboration), *Phys. Lett. B* **716**, 30 (2012).
 - [3] J. Alison, M. Gouzevitch, J. Mazzitelli, and C. Vernieri, *Rev. Phys.* **5**, 100045 (2020).
 - [4] S. Dawson, C. Englert, and T. Plehn, *Phys. Rep.* **816**, 1 (2019).
 - [5] M. Spira, *Prog. Part. Nucl. Phys.* **95**, 98 (2017).
 - [6] E. Torassa, *Prog. Part. Nucl. Phys.* **100**, 69 (2018).
 - [7] A. Maas, *Prog. Part. Nucl. Phys.* **106**, 132 (2019).
 - [8] S. Rappoccio, *Rev. Phys.* **4**, 100027 (2019).
 - [9] J. Baglio, A. Djouadi, R. Grber, M. Mhleitner, J. Quevillon, and M. Spira, *J. High Energy Phys.* **04** (2013) 151.
 - [10] M. J. Dolan, C. Englert, and M. Spannowsky, *J. High Energy Phys.* **10** (2012) 112.
 - [11] C. Englert, A. Freitas, M. Mhleitner, T. Plehn, M. Rauch, M. Spira, and K. Walz, *J. Phys. G* **41**, 113001 (2014).
 - [12] A. Papaefstathiou and K. Sakurai, *J. High Energy Phys.* **02** (2016) 006.
 - [13] C.-Y. Chen, Q.-S. Yan, X. Zhao, Y.-M. Zhong, and Z. Zhao, *Phys. Rev. D* **93**, 013007 (2016).
 - [14] B. Fuks, J. H. Kim, and S. J. Lee, *Phys. Rev. D* **93**, 035026 (2016).
 - [15] W. Kilian, S. Sun, Q.-S. Yan, X. Zhao, and Z. Zhao, *J. High Energy Phys.* **06** (2017) 145.
 - [16] P. Agrawal, D. Saha, and A. Shivaji, *Phys. Rev. D* **97**, 036006 (2018).
 - [17] B. Fuks, J. H. Kim, and S. J. Lee, *Phys. Lett. B* **771**, 354 (2017).
 - [18] F. Englert and R. Brout, *Phys. Rev. Lett.* **13**, 321 (1964).
 - [19] P. W. Higgs, *Phys. Rev. Lett.* **13**, 508 (1964).
 - [20] P. W. Higgs, *Phys. Rev.* **145**, 1156 (1966).
 - [21] T. W. B. Kibble, *Phys. Rev.* **155**, 1554 (1967).
 - [22] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Phys. Rev. Lett.* **13**, 585 (1964).
 - [23] S. Kanemura, S. Kiyoura, Y. Okada, E. Senaha, and C. Yuan, *Phys. Lett. B* **558**, 157 (2003).
 - [24] S. Kanemura, Y. Okada, E. Senaha, and C.-P. Yuan, *Phys. Rev. D* **70**, 115002 (2004).
 - [25] F. Arco, S. Heinemeyer, and M. Herrero, *Eur. Phys. J. C* **80**, 884 (2020).
 - [26] A. Dobado, M. J. Herrero, W. Hollik, and S. Penaranda, *Phys. Rev. D* **66**, 095016 (2002).

- [27] M. Brucherseifer, R. Gavin, and M. Spira, *Phys. Rev. D* **90**, 117701 (2014).
- [28] D. T. Nhung, M. Muhlleitner, J. Streicher, and K. Walz, *J. High Energy Phys.* **11** (2013) 181.
- [29] L. Wu, J. M. Yang, C.-P. Yuan, and M. Zhang, *Phys. Lett. B* **747**, 378 (2015).
- [30] G. Degraasi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori, and A. Strumia, *J. High Energy Phys.* **08** (2012) 098.
- [31] A. Kobakhidze, L. Wu, and J. Yue, *J. High Energy Phys.* **04** (2016) 011.
- [32] F. P. Huang, P.-H. Gu, P.-F. Yin, Z.-H. Yu, and X. Zhang, *Phys. Rev. D* **93**, 103515 (2016).
- [33] A. Efrati and Y. Nir, arXiv:1401.0935.
- [34] M. McCullough, *Phys. Rev. D* **90**, 015001 (2014); **92**, 039903(E) (2015).
- [35] M. Gorbahn and U. Haisch, *J. High Energy Phys.* **10** (2016) 094.
- [36] F. Maltoni, D. Pagani, A. Shivaji, and X. Zhao, *Eur. Phys. J. C* **77**, 887 (2017).
- [37] G. D. Kribs, A. Maier, H. Rzehak, M. Spannowsky, and P. Waite, *Phys. Rev. D* **95**, 093004 (2017).
- [38] The ATLAS Collaboration, Constraint of the Higgs boson self-coupling from Higgs boson differential production and decay measurements, Technical Report No. ATL-PHYS-PUB-2019-009, CERN, Geneva, 2019.
- [39] L.-B. Chen, H. T. Li, H.-S. Shao, and J. Wang, *J. High Energy Phys.* **03** (2020) 072.
- [40] M. J. Dolan, C. Englert, and M. Spannowsky, *Phys. Rev. D* **87**, 055002 (2013).
- [41] T. Abe, N. Chen, and H.-J. He, *J. High Energy Phys.* **01** (2013) 082.
- [42] C. Han, X. Ji, L. Wu, P. Wu, and J. M. Yang, *J. High Energy Phys.* **04** (2014) 003.
- [43] X.-F. Wang, C. Du, and H.-J. He, *Phys. Lett. B* **723**, 314 (2013).
- [44] B. Hespel, D. Lopez-Val, and E. Vryonidou, *J. High Energy Phys.* **09** (2014) 124.
- [45] S. Dawson, A. Ismail, and I. Low, *Phys. Rev. D* **91**, 115008 (2015).
- [46] A. Kobakhidze, N. Liu, L. Wu, and J. Yue, *Phys. Rev. D* **95**, 015016 (2017).
- [47] L.-C. L., C. Du, Y. Fang, H.-J. He, and H. Zhang, *Phys. Lett. B* **755**, 509 (2016).
- [48] J. Ren, R.-Q. Xiao, M. Zhou, Y. Fang, H.-J. He, and W. Yao, *J. High Energy Phys.* **06** (2018) 090.
- [49] S. Borowka, C. Duhr, F. Maltoni, D. Pagani, A. Shivaji, and X. Zhao, *J. High Energy Phys.* **04** (2019) 016.
- [50] L. Wu, H. Zhang, and B. Zhu, *J. Cosmol. Astropart. Phys.* **07** (2019) 033.
- [51] A. Alves, T. Ghosh, H.-K. Guo, and K. Sinha, *J. High Energy Phys.* **12** (2018) 070.
- [52] A. J. Barr, M. J. Dolan, C. Englert, and M. Spannowsky, *Phys. Lett. B* **728**, 308 (2014).
- [53] Q. Li, Q.-S. Yan, and X. Zhao, *Phys. Rev. D* **89**, 033015 (2014).
- [54] Q.-H. Cao, B. Yan, D.-M. Zhang, and H. Zhang, *Phys. Lett. B* **752**, 285 (2016).
- [55] Q. Li, Z. Li, Q.-S. Yan, and X. Zhao, *Phys. Rev. D* **92**, 014015 (2015).
- [56] Q.-H. Cao, Y. Liu, and B. Yan, *Phys. Rev. D* **95**, 073006 (2017).
- [57] Q.-H. Cao, G. Li, B. Yan, D.-M. Zhang, and H. Zhang, *Phys. Rev. D* **96**, 095031 (2017).
- [58] H.-J. He, J. Ren, and W. Yao, *Phys. Rev. D* **93**, 015003 (2016).
- [59] T. Huang, J. M. No, L. Perni, M. Ramsey-Musolf, A. Safonov, M. Spannowsky, and P. Winslow, *Phys. Rev. D* **96**, 035007 (2017).
- [60] C.-T. Lu, J. Chang, K. Cheung, and J. S. Lee, *J. High Energy Phys.* **08** (2015) 133.
- [61] J. Chang, K. Cheung, J. S. Lee, C.-T. Lu, and J. Park, *Phys. Rev. D* **100**, 096001 (2019).
- [62] A. Papaefstathiou, L. L. Yang, and J. Zurita, *Phys. Rev. D* **87**, 011301 (2013).
- [63] J. H. Kim, K. Kong, K. T. Matchev, and M. Park, *Phys. Rev. Lett.* **122**, 091801 (2019).
- [64] G. Buchalla, M. Capozzi, A. Celis, G. Heinrich, and L. Scyboz, *J. High Energy Phys.* **09** (2018) 057.
- [65] J. H. Kim, M. Kim, K. Kong, K. T. Matchev, and M. Park, *J. High Energy Phys.* **09** (2019) 047.
- [66] G. Li, L.-X. Xu, B. Yan, and C.-P. Yuan, *Phys. Lett. B* **800**, 135070 (2020).
- [67] A. M. Sirunyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **01** (2018) 054.
- [68] M. Aaboud *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **04** (2019) 092.
- [69] G. Aad *et al.* (ATLAS Collaboration), *Phys. Lett. B* **800**, 135103 (2020).
- [70] 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 (CERN, Geneva, 2019), pp. 221–584.
- [71] R. Contino *et al.*, CERN Yellow Rep. 255 **3** (2017).
- [72] A. Adhikary, S. Banerjee, R. K. Barman, B. Bhattacharjee, and S. Niyogi, *J. High Energy Phys.* **07** (2018) 116.
- [73] CMS Collaboration, Higgs pair production at the High Luminosity LHC, Technical Report No. CMS-PAS-FTR-15-002, CERN, Geneva, 2015.
- [74] CMS Collaboration, Projected performance of Higgs analyses at the HL-LHC for ECFA 2016, Technical Report No. CMS-PAS-FTR-16-002, CERN, Geneva, 2017.
- [75] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, arXiv:1704.01212.
- [76] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, *Nucl. Instrum. Methods Phys. Res., Sect. A* **543**, 577 (2005).
- [77] P. Baldi, P. Sadowski, and D. Whiteson, *Nat. Commun.* **5**, 4308 (2014).
- [78] P. Baldi, P. Sadowski, and D. Whiteson, *Phys. Rev. Lett.* **114**, 111801 (2015).
- [79] M. Bridges, K. Cranmer, F. Feroz, M. Hobson, R. Ruiz de Austri, and R. Trotta, *J. High Energy Phys.* **03** (2011) 012.
- [80] A. Buckley, A. Shilton, and M. J. White, *Comput. Phys. Commun.* **183**, 960 (2012).
- [81] N. Bornhauser and M. Drees, *Phys. Rev. D* **88**, 075016 (2013).

- [82] S. Caron, J. S. Kim, K. Rolbiecki, R. Ruiz de Austri, and B. Stienen, *Eur. Phys. J. C* **77**, 257 (2017).
- [83] G. Bertone, M. P. Deisenroth, J. S. Kim, S. Liem, R. Ruiz de Austri, and M. Welling, *Phys. Dark Universe* **24**, 100293 (2019).
- [84] M. Abdughani, J. Ren, L. Wu, J. M. Yang, and J. Zhao, *Commun. Theor. Phys.* **71**, 955 (2019).
- [85] P. C. Bhat, *Annu. Rev. Nucl. Part. Sci.* **61**, 281 (2011).
- [86] J. Ren, L. Wu, J. M. Yang, and J. Zhao, *Nucl. Phys.* **B943**, 114613 (2019).
- [87] S. H. Lim and M. M. Nojiri, *J. High Energy Phys.* **10** (2018) 181.
- [88] J. Amacker *et al.*, *J. High Energy Phys.* **12** (2020) 115.
- [89] J. Hajer, Y.-Y. Li, T. Liu, and H. Wang, *Phys. Rev. D* **101**, 076015 (2020).
- [90] A. Andreassen, B. Nachman, and D. Shih, *Phys. Rev. D* **101**, 095004 (2020).
- [91] F. Bishara and M. Montull, [arXiv:1912.11055](https://arxiv.org/abs/1912.11055).
- [92] L. Li, Y.-Y. Li, T. Liu, and S.-J. Xu, *J. High Energy Phys.* **10** (2020) 018.
- [93] V. Mikuni and F. Canelli, *Eur. Phys. J. Plus* **135**, 463 (2020).
- [94] A. Mullin, H. Pacey, M. Parker, M. White, and S. Williams, *J. High Energy Phys.* **02** (2021) 160.
- [95] C. Jin, S.-z. Chen, and H.-H. He, *Chin. Phys. C* **44**, 065002 (2020).
- [96] E. A. Moreno, T. Q. Nguyen, J.-R. Vlimant, O. Cerri, H. B. Newman, A. Perival, M. Spiropulu, J. M. Duarte, and M. Pierini, *Phys. Rev. D* **102**, 012010 (2020).
- [97] E. A. Moreno, O. Cerri, J. M. Duarte, H. B. Newman, T. Q. Nguyen, A. Perival, M. Pierini, A. Serikova, M. Spiropulu, and J.-R. Vlimant, *Eur. Phys. J. C* **80**, 58 (2020).
- [98] S. Jung, D. Lee, and K.-P. Xie, *Eur. Phys. J. C* **80**, 105 (2020).
- [99] B. Bhattacharjee, S. Mukherjee, and R. Sengupta, *J. High Energy Phys.* **19** (2020) 156.
- [100] S. R. Qasim, J. Kieseler, Y. Iiyama, and M. Pierini, *Eur. Phys. J. C* **79**, 608 (2019).
- [101] J. Arjona Martinez, O. Cerri, M. Pierini, M. Spiropulu, and J.-R. Vlimant, *Eur. Phys. J. Plus* **134**, 333 (2019).
- [102] P. T. Komiske, E. M. Metodiev, and J. Thaler, *J. High Energy Phys.* **01** (2019) 121.
- [103] M. Farina, Y. Nakai, and D. Shih, *Phys. Rev. D* **101**, 075021 (2020).
- [104] E. Bothmann and L. Debbio, *J. High Energy Phys.* **01** (2019) 033.
- [105] J. Lin, M. Freytsis, I. Moulton, and B. Nachman, *J. High Energy Phys.* **10** (2018) 101.
- [106] F. Staub, [arXiv:1906.03277](https://arxiv.org/abs/1906.03277).
- [107] T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson, *SciPost Phys.* **6**, 030 (2019).
- [108] H. Luo, M.-x. Luo, K. Wang, T. Xu, and G. Zhu, *Sci. China Phys. Mech. Astron.* **62**, 991011 (2019).
- [109] A. Chakraborty, S. H. Lim, M. M. Nojiri, and M. Takeuchi, *J. High Energy Phys.* **07** (2020) 111.
- [110] M. Capozzi and G. Heinrich, *J. High Energy Phys.* **03** (2020) 091.
- [111] M. Gori, G. Monfardini, and F. Scarselli, in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks*, 2005. (IEEE, Montreal, 2005), Vol. 2, pp. 729–734.
- [112] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, *IEEE Trans. Neural Networks* **20**, 61 (2009).
- [113] T. N. Kipf and M. Welling [arXiv:1609.02907](https://arxiv.org/abs/1609.02907).
- [114] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, [arXiv:1312.6203](https://arxiv.org/abs/1312.6203).
- [115] M. Defferrard, X. Bresson, and P. Vandergheynst, [arXiv:1606.09375](https://arxiv.org/abs/1606.09375).
- [116] D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, [arXiv:1509.09292](https://arxiv.org/abs/1509.09292).
- [117] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, [arXiv:1511.05493](https://arxiv.org/abs/1511.05493).
- [118] P. W. Battaglia, R. Pascanu, M. Lai, D. Rezende, and K. Kavukcuoglu, [arXiv:1612.00222](https://arxiv.org/abs/1612.00222).
- [119] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, *J. Comput.-Aided Mol. Des.* **30**, 595 (2016).
- [120] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, *Nat. Commun.* **8**, 13890 (2017).
- [121] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, [arXiv:1812.08434](https://arxiv.org/abs/1812.08434).
- [122] I. Henrion, J. Brehmer, J. Bruna, K. Cho, K. Cranmer, G. Louppe, and G. Rochette, in *Neural Message Passing for Jet Physics* (IEEE, New Jersey, 2017).
- [123] J. Ren, L. Wu, and J. M. Yang, *Phys. Lett. B* **802**, 135198 (2020).
- [124] M. Abdughani, J. Ren, L. Wu, and J. M. Yang, *J. High Energy Phys.* **08** (2019) 055.
- [125] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *J. High Energy Phys.* **07** (2014) 079.
- [126] R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo (NNPDF Collaboration), *Nucl. Phys.* **B877**, 290 (2013).
- [127] J. Grigo, K. Melnikov, and M. Steinhauser, *Nucl. Phys.* **B888**, 17 (2014).
- [128] M. Czakon, P. Fiedler, and A. Mitov, *Phys. Rev. Lett.* **110**, 252004 (2013).
- [129] D. de Florian, M. Der, and I. Fabre, *Phys. Rev. D* **98**, 094008 (2018).
- [130] S. Dittmaier *et al.* (LHC Higgs Cross Section Working Group Collaboration), [arXiv:1101.0593](https://arxiv.org/abs/1101.0593).
- [131] D. de Florian *et al.* (LHC Higgs Cross Section Working Group Collaboration), [arXiv:1610.07922](https://arxiv.org/abs/1610.07922).
- [132] 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).
- [133] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), *J. High Energy Phys.* **02** (2014) 057.
- [134] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **72**, 1896 (2012).
- [135] M. Cacciari, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 063.
- [136] The ATLAS Collaboration, Expected performance of the ATLAS detector at the High-Luminosity LHC, Technical Report No. ATL-PHYS-PUB-2019-005, CERN, Geneva, 2019.

- [137] M. Aaboud *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **78**, 903 (2018).
- [138] V. Nair and G.E. Hinton, in *Proceedings of the 27th International Conference on Machine Learning, Haifa* (2010), pp. 807–814.
- [139] D.S. Broomhead and D. Lowe, RRE **RSRE-MEMO-4148** (1998).
- [140] F. Schwenker, H. A. Kestler, and G. Palm, *Neural Netw.* **14**, 439 (2001).
- [141] D. P. Kingma and J. Ba, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [142] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimsheine, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai *et al.*, 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., New York, 2019), pp. 8024–8035.