Deep learning to improve the sensitivity of Higgs pair searches in the 4b channel at the LHC*

Yongcheng Wu (吴永成)[1,2†] Liang Xiao (肖亮)[1‡] Yan Zhang (张嫣)[1,2§]

[1]Department of Physics, Institute of Theoretical Physics and Institute of Physics Frontiers and Interdisciplinary Sciences, Nanjing Normal University, Nanjing 210023, China
[2]Nanjing Key Laboratory of Particle Physics and Astrophysics, Nanjing 210023, China

Abstract: Higgs self-coupling is crucial to understand the structure of the scalar potential and the mechanism of electroweak symmetry breaking. In this study, we utilize a deep neural network based on a particle Transformer model that relies on an attention mechanism to comprehensively analyze the measured trilinear Higgs self-coupling through Higgs pair production with subsequent decay into four b-quarks ($HH \to b\bar{b}b\bar{b}$) at the LHC. The model processes the full event-level information as input and bypasses explicit jet pairing. It also serves as an event classifier. At HL-LHC, our approach constrains κ_λ to $(-0.25, 5.41)$ at 68% CL, and thus achieved an improvement of ~44% in precision over conventional cut-based analyses. The results of a comparison against alternative machine learning architectures demonstrate the outstanding performance of the Transformer-based model owing to its ability to capture correlations in high-dimensional collision data with the help of the attention mechanism. These findings highlight the potential of attention-based networks and end-to-end event classifiers in collider phenomenology.

Keywords: Higgs physics, Higgs pair production, machine learning, collider phenomenology

DOI: 10.1088/1674-1137/ae2454 **CSTR:** 32044.14.ChinesePhysicsC.50033105

I. INTRODUCTION

The standard model (SM) is currently the most successful theoretical framework for describing fundamental particles and their interactions, with the Higgs particle considered the core of the model. Driven by the Higgs mechanism, the Higgs field breaks the electroweak symmetry spontaneously and explains the acquisition of mass by fundamental particles. Since the discovery of the 125 GeV Higgs particle by the ATLAS and CMS experiments in 2012 [1, 2], the precise measurements of the properties of the Higgs particle, including the determination of its mass, spin, parity, interactions with other particles, and Higgs self-coupling, are among the most important tasks in particle physics that provide an essential test of the SM. Despite the considerable success of the SM, some phenomena still cannot be explained using this model, such as the existence of dark matter, nonzero mass of neutrinos, and matter-antimatter asymmetry. These phenomena suggest the existence of physics beyond the SM (BSM), and the discovery of the Higgs particle has opened a window to explore new physics. Any deviation between the measured properties of the Higgs particle and predictions of the SM would be crucial for the exploration of physics beyond the SM.

Currently, the ATLAS and CMS experiments already include various measurements related to the Higgs particle. The mass of the Higgs particle has been precisely measured by both experiments [3, 4]. The width of the Higgs particle has also been precisely measured using the off-shell effect [5, 6], which was originally believed to be impossible in a hadron collider. All major production modes and decay channels of the Higgs particle have been evaluated [6, 7], with the overall signal strength in agreement with SM predictions. The interactions of the Higgs particle with other SM particles can be examined from these measurements, which is crucial to test the Higgs mechanism.

Measuring the Higgs self-coupling [8–17] is important for evaluating the properties of the Higgs boson. The Higgs self-coupling determines the shape of the Higgs potential, which is crucial for electroweak symmetry

Received 3 September 2025; Accepted 25 November 2025; Accepted manuscript online 26 November 2025
* Supported by the National Natural Science Foundation of China (NNSFC) (12305112)
† E-mail: ycwu@njnu.edu.cn
‡ E-mail: lxiao@njnu.edu.cn
§ E-mail: zyan@njnu.edu.cn

breaking [18, 19]. In the SM, the Higgs potential is expressed as

$$V = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2, \tag{1}$$

with

$$\Phi = \begin{pmatrix} \phi^+ \\ \dfrac{v + H + i\phi^0}{\sqrt{2}} \end{pmatrix}, \tag{2}$$

where $v = \sqrt{-\mu^2/\lambda}$ represents the vacuum expectation value (vev) determined from the potential. The potential for H after electroweak symmetry breaking can be expressed as

$$V = \frac{1}{2}m_H^2H^2 + \lambda_{HHH}vH^3 + \frac{1}{4}\lambda_{HHHH}H^4, \tag{3}$$

where $\lambda_{HHH} = \lambda_{HHHH} = \dfrac{m_H^2}{2v^2}$ represent the trilinear and quartic Higgs self-couplings. In the SM, the Higgs self-couplings are fully determined by the mass and vacuum expectation value of the Higgs particle, which have been measured with high precision. Any deviation from the predictions of the SM would support the validity of new physics [20–26]. In this study, we use the kappa framework to parameterize the deviation

$$V = \frac{1}{2}m_H^2H^2 + \kappa_\lambda\lambda_{HHH}vH^3 + \frac{1}{4}\kappa_{\lambda,4}\lambda_{HHHH}H^4, \tag{4}$$

where κ_λ and $\kappa_{\lambda,4}$ represent deviations in trilinear and quartic Higgs self-couplings, respectively. The effective field theory (EFT) framework provides another powerful and model-independent approach to study Higgs self-couplings [27–31].

Although this study focuses on trilinear coupling, the quartic Higgs self-coupling is also important. However, the direct measurement of this coupling is extremely challenging because of the extremely small cross-section of triple Higgs production at current or near-future colliders such as the LHC or ILC [32]. Future 100 TeV hadron colliders [33–35] or high-energy muon colliders [36] are considered promising platforms to probe this coupling. Additionally, loop corrections in certain Higgs pair production processes can offer indirect sensitivity to quartic coupling, especially those in VBF and VHH channels [32].

The direct measurement of the trilinear Higgs self-coupling relies on Higgs pair production [37–49] to which the gluon-gluon fusion (ggF) process provides the dominant contribution at the LHC [50–53]. Higgs pair production from vector boson fusion (VBF) has been investigated to measure Higgs self-coupling [18, 54]. In addition to ggF and VBF production of a pair of Higgs bosons, many other processes have been considered to probe the Higgs self-coupling, which includes the double Higgs-strahlung process [55], Higgs pairs associated with two top quarks [18], and Higgs pairs plus jet production [56, 57]. These processes have smaller cross-sections than ggF production; however, they can still contribute to the Higgs self-coupling measurement at future colliders with higher energy and luminosity. Additionally, the single Higgs production and its decay can be utilized to measure the Higgs self-coupling via its contributions at the loop level [58–65]. Meanwhile, Higgs pair production can be significantly altered in extended Higgs sectors where the Higgs pair is produced resonantly [66–72]. Studies on Higgs pair production can help probe the scalar potential in models such as xSM [73–76], 2HDM [77–82], and MSSM [83]. However, in this study, we focus on the non-resonant production of Higgs pairs and leave resonant production for future work.

Both ATLAS and CMS conducted studies on the Higgs pair production process. In the ATLAS analysis, decay channels $HH \to b\bar{b}b\bar{b}$ [84], $HH \to b\bar{b}\tau^+\tau^-$ [85], and $HH \to b\bar{b}\gamma\gamma$ [86] were examined. The results indicate that κ_λ is constrained to the range $(-0.6, 6.6)$ after combining all these channels [87]. In the CMS analysis, decay channels $HH \to b\bar{b}ZZ^*$ [88], $HH \to$ Multilepton, $HH \to b\bar{b}\gamma\gamma$ [89], $HH \to b\bar{b}\tau^+\tau^-$ [90], and $HH \to b\bar{b}b\bar{b}$ [91] were analyzed. The combined results show that the κ_λ value is constrained to $(-1.24, 6.49)$ [7]. Future colliders are expected to significantly improve the measurement precision with higher center-of-mass energy and integrated luminosity. The HL-LHC [92–95] is expected to achieve the precise measurement of the Higgs self-coupling $\kappa_\lambda = 1.0^{+0.48}_{-0.42}$, and the HE-LHC [96] and FCC-hh [97–100] are projected to achieve a precision of 5% on the Higgs self-coupling measurements. Future lepton colliders will probe the Higgs self-coupling with approximately 20% precision from FCC-ee [101, 102] and ILC [103, 104] as well as the multi-TeV muon collider [105–108].

The Higgs boson predominantly decays into $b\bar{b}$ with a branching ratio of approximately 58% in the SM. Among all possible decay channels, the $4b$ final state constitutes the largest fraction (approximately 33.6%) for a Higgs pair system [109], which makes it a promising channel for investigation. However, current constraints on the Higgs self-coupling are relatively broad from the $4b$ channel $\kappa_\lambda \in (-3.3, 11.4)$ [87], which is weak compared to that of other channels and is mainly limited by the complex background at the LHC. Improving the discriminating power between signal and background in the $4b$ channel will therefore significantly enhance sensitivity. Many studies have focused on improving the sensitivity in this channel [110–116]. In this study, we explore the possibility of using machine learning techniques to overcome the

challenge in the 4b channel of Higgs pair searches at the LHC.

The amount of data generated at colliders is enormous, and machine learning (ML) is well-suited to process such large-scale datasets [117–119]. Various ML techniques have been applied in particle physics [120, 121], such as jet-tagging [122–125], CNN models based on image data [126–129], GNNs based on graph data [130–132], ParticleNet and energy flow networks based on particle clouds [133–136], current state-of-the-art (SOTA) Transformer-based models, and particle Transformer (*ParT*) [137–141]. Similarly, ML techniques have been used to measure Higgs self-coupling [115, 116, 142–144]. Additionally, several studies investigated the performance of different machine learning algorithms for measuring Higgs self-coupling [145–147].

For the $HH \to b\bar{b}b\bar{b}$ channel, different ML architectures have been applied. For example, Ref. [115] employed a DNN architecture, which indicated that with $\mathcal{L} = 3000\,\text{fb}^{-1}$ at HL-LHC, the Higgs self-coupling can be constrained within the range $\kappa_\lambda \in (-0.8, 6.6)$ at 68% CL. A more recent work [116] utilizing a Transformer model achieved a constraint of $\kappa_\lambda \in (-1.56, 7.57)$ at 95% CL with $\mathcal{L} = 300\,\text{fb}^{-1}$ at the LHC. Another transformer-based model has also been used in a study on resonant Higgs pair production with 4b final states [148]. Considering the excellent performance of *ParT* in jet tagging tasks [137], a modified version of *ParT* including the entire event information as input for event classification was used in this work to improve the sensitivity to the Higgs self-coupling through the $HH \to b\bar{b}b\bar{b}$ channel. However, we must emphasize that the analysis in this study demonstrates the effectiveness not only of the transformer structure on event selection but also the end-to-end analysis utilizing event-level information.

The remainder of this paper is organized as follows. In Sec. II, we briefly describe all processes considered in this work. In Sec. III, after the introduction of the structure of *ParT*, we discuss the modifications introduced in this study and the details of the training process. The performance of the modified model in event classification is also presented. In Sec. IV, we present the results on the Higgs self-coupling measurement with a comparison among different methods. The interpretability of the model is discussed at the end of this section. Finally, we summarize our findings in Sec.V.

II. EVENT GENERATION

For the Higgs self-coupling measurement, we focus on the non-resonant gluon-gluon fusion (ggF) production of Higgs pairs [149–151], with both Higgs particles decaying into b-quark pair. At leading order (LO), this process includes contributions from triangle and box diagrams as well as the interference between them. The representative Feynman diagrams are shown in Fig. 1. The triangle diagrams depend on both the trilinear Higgs self-coupling and Yukawa couplings, and the box diagrams depend only on the Yukawa couplings of the fermion in the loop. In our simulations, we include contributions from third generation quarks but with fixed Yukawa coupling at the corresponding SM value. Meanwhile, the trilinear Higgs self-coupling is allowed to differ from the value obtained from the SM. The total Higgs pair production cross-section at a given energy will be a quadratic function of κ_λ and is given at both $\sqrt{s} = 13$ TeV and $\sqrt{s} = 14$ TeV.

$$\sigma^{\text{LO}}_{HH}(\kappa_\lambda) = \begin{cases} 4.15\times10^{-3}\,\kappa_\lambda^2 - 2.02\times10^{-2}\,\kappa_\lambda + 3.05\times10^{-2}\,\text{pb} \\ \qquad\qquad \sqrt{s} = 13\ \text{TeV}, \\ 4.89\times10^{-3}\,\kappa_\lambda^2 - 2.39\times10^{-2}\,\kappa_\lambda + 3.62\times10^{-2}\,\text{pb} \\ \qquad\qquad \sqrt{s} = 14\ \text{TeV}, \end{cases} \tag{5}$$

where the first term depending on κ_λ^2 in each line represents the contributions of the pure triangle diagrams, the last term without any dependence on κ_λ comes from the box diagrams, and the middle term with linear dependence on κ_λ comes from interference effects between the triangle and box diagrams. The cross-section is further normalized according to the higher order k-factor given in [115], which is approximately 2.4 for $\kappa_\lambda = 1$ and varies by 35% for $\kappa_\lambda \in [-5, 12]$, ranging from $k \approx 2.28$ at $\kappa_\lambda = 2$ to $k \approx 3.12$ at $\kappa_\lambda = 5$. The Higgs pair production cross-sections for different values of κ_λ at different energies are

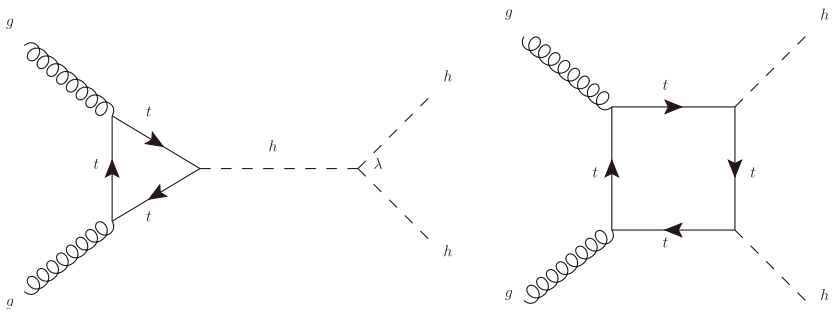


Fig. 1. Leading-order Feynman diagrams of di-Higgs production in the SM.

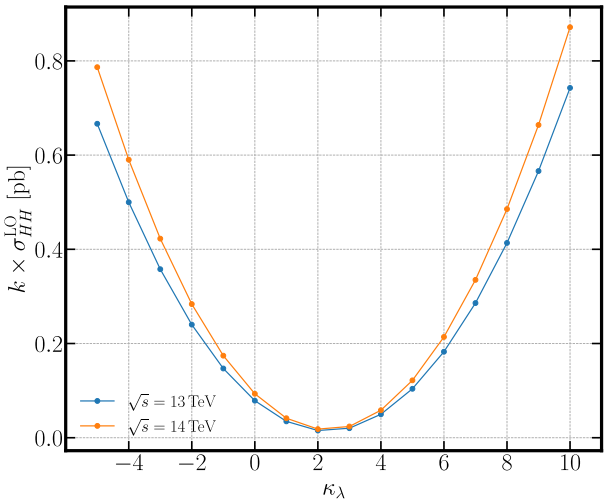


Fig. 2. (color online) Higgs pair production cross-section as a function of κ_λ at the LHC ($\sqrt{s} = 13$ TeV and 14 TeV) assuming all other couplings are fixed, with a higher order correction k-factor.

shown in Fig. 2.

For the backgrounds, we focus on the processes containing multiple b-quarks and/or multi-jets. In our analysis, the main backgrounds are the QCD production of four b quarks ($4b$) as well as two b quarks with two light quarks ($2b2j$). In addition to these two major backgrounds, we consider the production of b-quarks from the decay of the Z boson, Higgs, and top quark. In Table 1, we list all signal (given with $\kappa_\lambda = 1$) and background processes considered in this analysis together with their cross-sections at LO at both $\sqrt{s} = 13$ TeV and $\sqrt{s} = 14$ TeV. For the signal, as well as $t\bar{t}$, $4b$, and $2b2j$ background processes, NLO k-factors are listed [115, 152–155].

Table 1. All processes considered in this study, with the LO cross-section at $\sqrt{s} = 13$ TeV and $\sqrt{s} = 14$ TeV. The k-factor accounts for the NLO correction for major backgrounds $4b$, $2b2j$, and $t\bar{t}$ processes, and higher order corrections for the signal HH at $\kappa_\lambda = 1$.

Process	$\sigma^{\rm LO}$/fb		k-factor
	$\sqrt{s} = 13$ TeV	$\sqrt{s} = 14$ TeV	
HH	1.446×10^{1}	1.723×10^{1}	2.4
$b\bar{b}b\bar{b}$	2.465×10^{6}	2.840×10^{6}	1.6
$b\bar{b}jj$	5.673×10^{8}	6.470×10^{8}	1.3
$t\bar{t}$	5.058×10^{5}	5.968×10^{5}	1.5
$t\bar{t}b\bar{b}$	3.614×10^{3}	4.472×10^{3}	–
$t\bar{t}H$	3.997×10^{2}	4.794×10^{2}	–
$b\bar{b}H$	4.746×10^{1}	5.496×10^{1}	–
ZZ	9.317×10^{3}	1.026×10^{4}	–
ZH	5.799×10^{2}	6.422×10^{2}	–

The events for the signal process with possible different values of κ_λ and for background processes are simulated by `MadGraph` [156] with both $\sqrt{s} = 13$ TeV and $\sqrt{s} = 14$ TeV. To increase the efficiency of the simulation, we use moderate parton-level basic cuts: $p_T^{j,b} > 20$ GeV, $|\eta^{j,b}| < 4$. `Pythia` [157] is used for hadronization and showering, followed by the detector simulation by `Delphes` [158]. `FastJet` [159] is linked to reconstruct jets from the particle-flow output from `Delphes` using an anti-k_t algorithm [160] with two different values of ΔR: 0.5 for a slim jet originating from quarks/gluons and 1.0 for a fat jet originating from heavier objects such as Higgs, W/Z boson, and top quarks.

III. TRANSFORMER FOR EVENTS

Instead of working on individual objects (such as jets, charged leptons, and photons) within an event, we investigate the possibility of working directly on the entire event. The particle Transformer (*ParT*) [137] is a SOTA object tagging ML algorithm trained on a 10-label-classification task. By implementing *ParT*, we could, in principle, tag the Higgs, top, and W/Z bosons as a fat jet in the events and combine them to extract the event of the signal (Higgs pair) from other background processes (containing top quark and W/Z boson). Although the individual object reconstruction efficiency is sufficient, relying on reconstructed individual heavy objects inside one event may suffer from the combinatorial problem. This can considerably reduce sensitivity, especially for the $HH \to b\bar{b}b\bar{b}$ case in which the four b-jets in the final states are indistinguishable and six combinations of these b-jets form one Higgs pair candidate. Hence, we explore the possibility of training a model on the information of the entire event, which can be treated as a single extremely fat jet. Consequently, the *ParT* framework can still be used with minor modification in the event classification task. In the following sections, we first introduce the main features of *ParT*. Then, the training details of our setup are presented. Further, the performance on the event classification task is discussed. To avoid confusion with the original *ParT* model, we refer to the *ParT* working on the entire event as an event Transformer (*EvenT*). However, we emphasize that the analysis presented in this study can be performed based on any type of algorithm in addition to the Transformer. The focus here is the possibility of an end-to-end event level analysis.

A. Basics of particle Transformer

ParT [137] is a classification model based on the *Transformer* architecture [161], which is a deep learning model originally designed for natural language processing (NLP) tasks. The core of the *Transformer* is the self-attention mechanism, which establishes relationships among all elements in the input sequence rather than be-

ing limited to local information like convolutional neural networks (CNNs). The Transformer can capture global information by computing the attention weight matrix. *Transformer* models have demonstrated powerful performance across multiple domains thanks to the advantages of this mechanism.

The structure of *ParT* is shown in Fig. 3. The model receives two sets of inputs: *particles*, which include the features (e.g., kinematics, PID, and trajectory displacment) of each single object (e.g., tracks and charged leptons), and *interactions*, which indicate the relationships between two objects. In *ParT*, four such relationships of interest exist.

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}, \tag{6}$$

$$k_{T,ij} = \min(p_{T,i}, p_{T,j})\Delta R_{ij}, \tag{7}$$

$$z_{ij} = \frac{\min(p_{T,i}, p_{T,j})}{p_{T,i} + p_{T,j}}, \tag{8}$$

$$m_{ij}^2 = (p_i + p_j)^2, \tag{9}$$

where y, ϕ represent the rapidity and azimuthal angle of the individual object, respectively. Further, p_T represents the corresponding magnitude of the transverse momentum. Then, the inputs are passed through intermediate layers consisting of *particle attention blocks* and *class attention blocks*. Compared to the traditional *Transformer*, *ParT* introduces, in the *particle attention block*, a new attention mechanism referred to as particle multi-head attention (P-MHA) in which the *interactions* are included in the attention calculation.

$$\text{P-MHA}(X) = \text{concat}(H_1, \ldots, H_h)W^O, \tag{10}$$

$$H_i = \mathcal{A}_i V_i, \tag{11}$$

$$\mathcal{A}_i = \text{softmax}\left(\frac{Q_i K_i^\top}{\sqrt{d_k}} + U_i\right), \tag{12}$$

$$Q_i = XW_i^Q + b_i^Q, \tag{13}$$

$$K_i = XW_i^K + b_i^K, \tag{14}$$

$$V_i = XW_i^V + b_i^V, \tag{15}$$

where X includes *particles* features and U contains features of the *interactions* between objects. W_i^j and b_i^j are trainable parameters. Including the *interactions* U in the attention calculation enables the model to better capture relationships between particles and enhance the expressiveness of the attention mechanism. $\mathcal{A}_i$ represents the attention weight matrix, where each element represents the attention score of each pair of input *particles*.

The *class attention block* shares a similar architecture with the *particle attention block*; however, it differs in two aspects in that it uses the standard MHA mechanism instead of the particle-specific variant and that the MHA takes a global class token as input along with particle embeddings. The query, key, and value matrices are computed as

$$Q = W_q X_{\text{class}} + b_q, \tag{16}$$

$$K = W_k \mathbf{Z} + b_k, \tag{17}$$

$$V = W_v \mathbf{Z} + b_v, \tag{18}$$

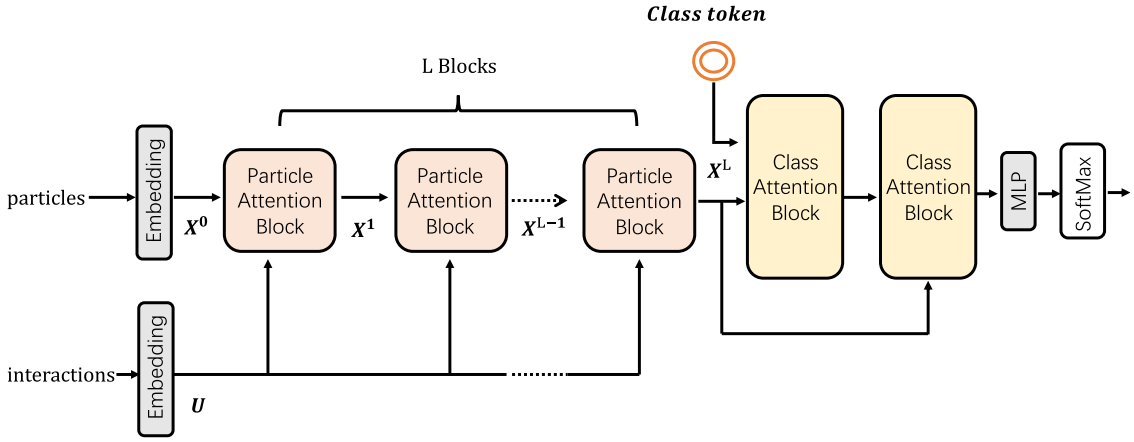


Fig. 3. (color online) *ParT* model architecture.

where $\mathbf{Z} = [X_{\text{class}}, X_{\text{PAB}}]$ represents the concatenation of the class token and particle embedding output from the last *particle attention block*. The output of the *class attention block* is passed to a Softmax layer that produces the probabilities over classes.

B. Training details

In our setup, the *EvenT* model receives information from the entire event as input features. To further facilitate establishing connections among objects within an event, information from jets reconstructed using two different $\Delta R = 0.5$ and 1.0 is also provided as discussed in the last section. Hence, the input for *EvenT* includes *particle* features of all objects within a given event, including tracks, energy deposits, charged leptons, jets, and the event itself, which extends the dimension of the input substantially. The orientation of a single event is not important. Further, *interaction* features are extended to include affiliation relations between different objects in additional original kinematic relations. To avoid the time-issue when loading all inputs, *interaction* features are now calculated in advance. The *EvenT* model is trained from scratch on a dataset with the nine classes listed in Table 1, which contains 3 million samples per class. After each training epoch, the model is validated on a separate validation set containing 3 million samples per class and its final performance is evaluated on an independent test set with an equivalent sample size per class. The model is trained with a batch size of 128 and initial learning rate of 0.001. The learning rate remains constant during the first 70% of training epochs, after which it decays exponentially, reaching 1% of its initial value by the end of training. The number of training epochs is set to 50, and the number of heads in the MHA mechanism is 8.

All processes during training can be treated equally. However, for our purpose, we focused on the Higgs pair process as much as possible from the other processes, among which the $2b2j$ process has the largest cross-section. To avoid large contamination from the $2b2j$ process, we adopt a weighted cross-entropy loss function in which the process $2b2j$ is assigned with a weight of 70, unlike ParT used for a general purpose jet-tagging task. All other processes are assigned with a weight of 1. After applying the weighted cross-entropy loss, the probability of miss-classifying $2b2j$ events as *HH* events is reduced from 5.17×10^{-3} to 5.26×10^{-5} (two orders of magnitude), drastically improving the analysis of the Higgs pair process. The measurement uncertainty of the Higgs pair production cross-section can be reduced from approximately 400% to 200% by such improvement.

Further, as our target is the measurement of κ_λ, the model is required to work properly on all values of κ_λ to obtain the optimal result. As a comparison, Fig. 4 shows the results of *HH* efficiency from training on the dataset with only $\kappa_\lambda = 1$ (blue line) and training on that containing $\kappa_\lambda = \{-1, 1, 4, 6, 8\}$ (red line). Evidently, for the training using data with only $\kappa_\lambda = 1$, the *HH* efficiency drastically decreases for other values of κ_λ. The variance of the efficiency can be as large as 40% (from 58% at $\kappa_\lambda = 2$ to 18% at $\kappa_\lambda = 3$). Meanwhile, when the data with other values of κ_λ are included, the *HH* efficiency is almost stable at approximately 50%. Such performance is therefore suitable for the further analysis of the measurement precision about κ_λ.

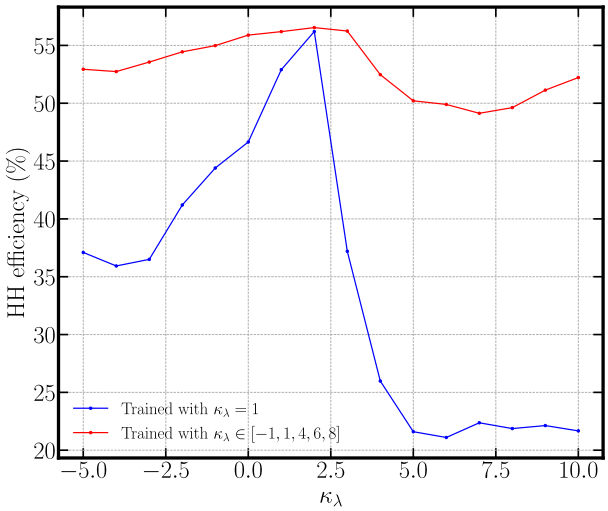


Fig. 4. (color online) *HH* efficiency under different κ_λ values for training containing only $\kappa_\lambda = 1$ *HH* data (blue line) and training containing $\kappa_\lambda = -1, 1, 4, 6, 8$ *HH* data (red line).

C. Performance of event classification

We further demonstrate the classification performance of the model. *EvenT* is now essentially a multiclass classification model with nine event categories. For multiclass models, the performance can be evaluated using the confusion matrix, ROC curves, and AUC scores.

A confusion matrix is structured as an $N\times N$ matrix, where N represents the number of target classes with rows representing true labels and columns indicating predicted labels to quantify the performance of the classification model. The confusion matrix can be presented as

$$\text{Confusion Matrix} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1N} \\ C_{21} & C_{22} & \cdots & C_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1} & C_{N2} & \cdots & C_{NN} \end{bmatrix}. \tag{19}$$

Specifically,

- C_{ii} represents the probability that an event with a true class i is correctly predicted as class i.

- C_{ij} (for $i \neq j$) represents the probability that an

event with true class i is incorrectly predicted as class j.

The confusion matrix of *EvenT* on the testing set, where the HH class only includes events with $\kappa_\lambda = 1$, is visualized in the left panel of Fig. 5, where the color intensity indicates corresponding probabilities.

Several comments can be made regarding the confusion matrix in order. First, the $2b2j$ column shows that every class has a relatively high efficiency tagging as $2b2j$ because we enhance the weight of $2b2j$ events during the training, as discussed above. As may be observed from the diagonal elements, excluding the $2b2j$ class, all remaining classes have a good self-tagging efficiency. Additionally, processes similar to each other form a diagonal block in the tagging efficiency, such as ($ttbb$, ttH) and (ZZ, ZH). For our purpose of measuring the Higgs pair production process, the first column of the confusion matrix indicates that all other processes have been heavily suppressed while maintaining sufficient efficiency for the HH process.

The receiver operating characteristic (ROC) curve and the area under the curve (AUC) are commonly used measures of the performance of binary classification tasks, and they can also be extended to multiclass classification problems. In a binary classification setting with only positive and negative classes, the ROC curve is generated from the true positive rate (TPR) and false positive rate (FPR) at various classification thresholds. The AUC is the area under the ROC curve, and it is used to quantify the overall classification performance.

For multiclass classification tasks, the one-vs-rest (OvR) strategy can be used to simplify the problem into a series of binary classification tasks, allowing the computation of ROC curves and AUC scores. For each class, it is considered as the positive class while all other classes are treated as the negative class. The right panel of Fig. 5 shows the ROC curves and corresponding AUC values for each class. For most of the classes, the AUC is approximately 0.9, which indicates a relatively good performance of *EvenT*. Further, the AUC for distinguishing the HH process from other processes is approximately 0.9, which provides a solid foundation for the analysis of the Higgs self-coupling via Higgs pair production in the next section. Meanwhile, the information contained in the confusion matrix as well as the OvR ROC curves can be used not only in the HH analysis but also in other relevant analyses. In this study, we use HH as a benchmark study to demonstrate the effectiveness of the end-to-end analysis based on event-level information, especially when involving pairing or other event topology analysis that will reduce selection efficiency.

IV. HIGGS SELF-COUPLING MEASUREMENT

A. *EvenT*-based analysis

Based on the *EvenT* discussed earlier, we consider the measurement of κ_λ through the $HH \to 4b$ channel in this section. The corresponding χ^2, which indicates the deviation from the SM case, is constructed according to

$$\chi^2(\sigma_{HH},\kappa_\lambda) = \frac{(S(\sigma_{HH},\kappa_\lambda) - S_{\rm SM})^2}{S_{\rm SM}}, \tag{20}$$

$$S(\sigma_{HH},\kappa_\lambda) = \mathcal{L} \times \left(C_{11}(\kappa_\lambda)\sigma_{HH} + \sum_{i=2}^{9} C_{i1}\sigma_i \right), \tag{21}$$

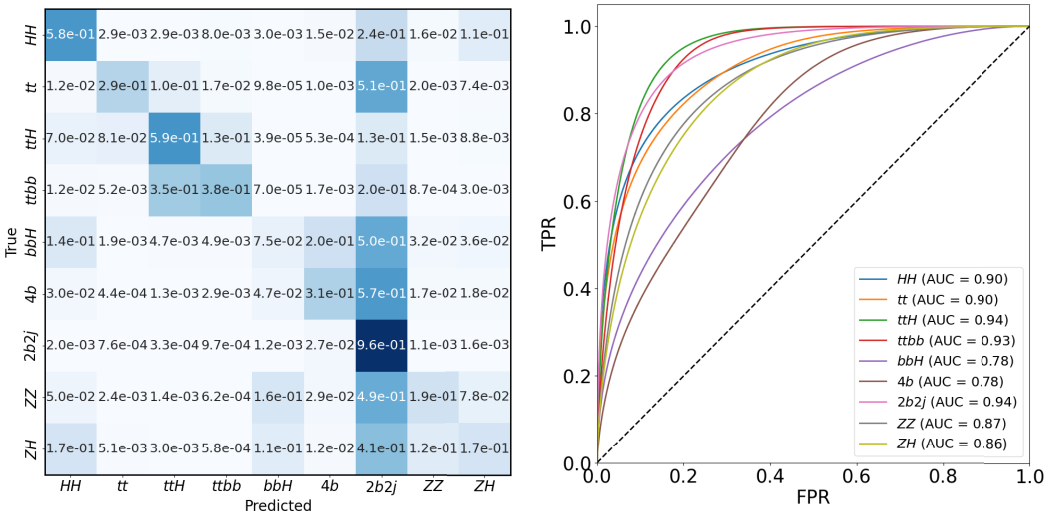


Fig. 5. (color online) Left: Confusion matrix of *EvenT* on a nine-class classification task. Right: OvR ROC curves for each class of *EvenT* and corresponding AUC. All results were obtained with the weighted cross-entropy loss.

$$S_{\rm SM} = S(\sigma_{HH}|_{\kappa_\lambda=1}, \kappa_\lambda = 1), \tag{22}$$

where $i = 1, \cdots, 9$ corresponds to all processes considered in the confusion matrix shown in the left panel of Fig. 5, including HH, tt, ttH, $ttbb$, Hbb, $4b$, $2b2j$, ZZ, and ZH. C_{ij} represents elements in the confusion matrix. For the HH process, the dependence of C_{11} on κ_λ is also included, as shown in Fig. 4. σ_i represents the cross-section of each process listed in Table 1. Notably, σ_{HH} is a free parameter in the χ^2 calculation. In our analysis, we consider the LHC experiment with $\sqrt{s} = 14$ TeV and integrated luminosities of $\mathcal{L} = 300/3000\,\text{fb}^{-1}$. Based on the above mentioned setup for a fixed value of κ_λ, χ^2 depends only on the corresponding Higgs pair production cross-section σ_{HH}. Then, the upper limit on the cross-section σ_{HH} can be obtained for a given κ_λ by solving $\chi^2(\sigma_{HH}, \kappa_\lambda) = 1(3.84)$ for a 68% (95%) confidence level (CL).

However, as discussed earlier, the performance of *EvenT* depends on the classification threshold. The effect of the threshold on the Higgs pair production measurement is shown in Fig. 6. It may be observed that the signal efficiency decreases smoothly with an increase in the threshold for both $\kappa_\lambda = 1$ (solid green line) and $\kappa_\lambda = 5$ (dashed green line), whereas the misclassification rate of the dominant background $2b2j$ (solid orange line) drops more sharply. Hence, the relative error of measuring the Higgs pair production cross-section decreases with higher threshold, as shown by the blue curves in Fig. 6. However, we must emphasize that when the threshold is close to 1, the events passing the threshold decrease drastically. Hence, the corresponding analysis suffers from large uncertainties. In our analysis, we use three benchmark thresholds $p_{\rm th} = 0.5$, 0.7, and 0.9 (indicated by the vertical purple dashed lines in Fig. 6) together with $p_{\rm th} = 0.0$, where we rely entirely on the raw output of the *EvenT* for the classification.

The upper limits on the Higgs pair production cross-section as a function of κ_λ obtained from *EvenT* with different classification thresholds are shown in Fig. 7. The upper limits closely follow the HH efficiency shown in Fig. 4. The theoretical prediction of the Higgs pair production cross-section is also shown in Fig. 7. We compare the *EvenT*-derived upper limits with the theoretical prediction to extract the constraints on the Higgs self-coupling κ_λ. At the HL-LHC, with $\sqrt{s} = 14$ TeV and $\mathcal{L} = 3000\,\text{fb}^{-1}$, the constraint at 68% CL is given as $\kappa_\lambda \in [-0.25, 5.41]$ for $p_{\rm th} = 0.9$. The constraints are slightly weaker for other choices of the threshold, as indicated in Fig. 7.

B. Cut-based analysis

For comparison, we followed the cut-based analysis presented in [84] to single out $HH \to b\bar{b}b\bar{b}$ events. In this analysis, events first need to pass several general selections including a requirement of at least 4 b-tagged (with tagging efficiency of approximately 77%) jets with $p_T > 40$ GeV and $|\eta_b| < 2.5$. The four b-jets with the highest p_T are used to reconstruct the Higgs pair system. Among the three possible pairings of these four b-jets, the one in which the higher p_T jet pair has the smallest ΔR is used for further analysis. To further suppress the background, $|\Delta\eta_{HH}| < 1.5$ is also imposed for the two reconstructed Higgs.

Events are further required to satisfy additional selection criteria designed to reduce the background and improve the analysis sensitivity. A top veto cut is required

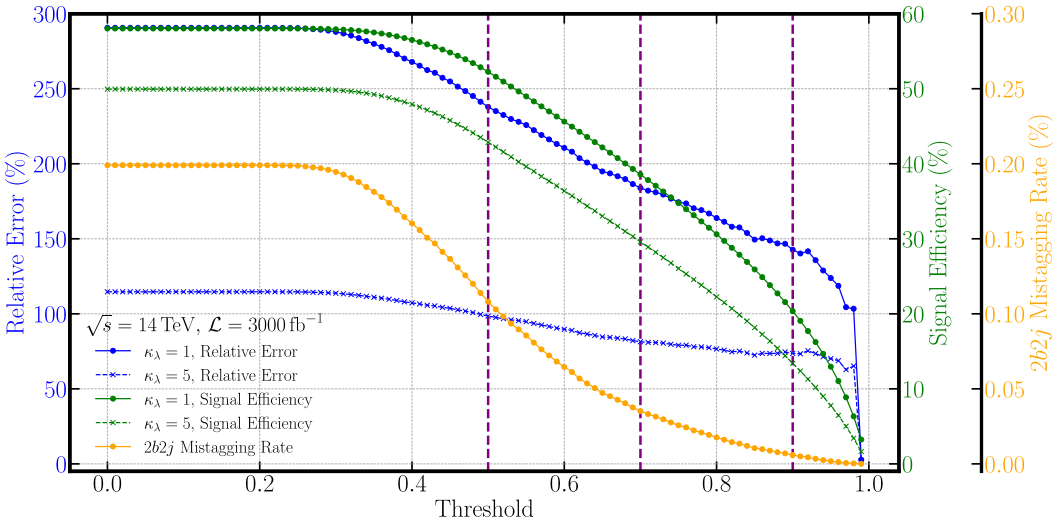


Fig. 6. (color online) Signal efficiency (green lines) and relative error of the Higgs pair production cross-section measurement (blue lines) for $\kappa_\lambda = 1$ (solid lines) and $\kappa_\lambda = 5$ (dashed lines) at different thresholds. The misclassification rate of the dominant background $2b2j$ is shown in orange.

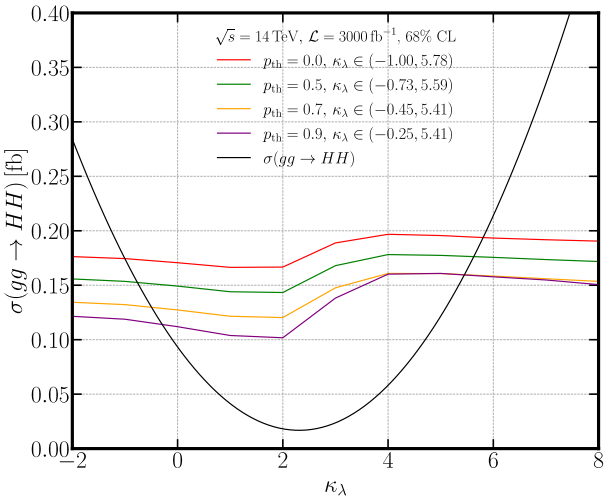


Fig. 7. (color online) Upper limit on the Higgs pair production cross-section as a function of κ_λ for different classification thresholds $p_{\rm th}$ (four colored lines) with $\sqrt{s} = 14$ TeV and $\mathcal{L} = 3000\,\mathrm{fb}^{-1}$ at the LHC. The Higgs pair production cross-section as a function of κ_λ is shown as a black line.

to suppress the $t\bar{t}$ background. The top veto discriminant X_{Wt} is defined as

$$X_{Wt} = \min_{jjb}\left\{\sqrt{\left(\frac{m_{jj}-m_W}{0.1m_{jj}}\right)^2+\left(\frac{m_{jjb}-m_t}{0.1m_{jjb}}\right)^2}\right\}, \quad (23)$$

where $m_W = 80.4$ GeV and $m_t = 172.5$ GeV represent the nominal W boson and top quark mass, respectively. m_{jj} represents the invariant mass of the two jets assumed to come from the W boson decay. Together with one of the leading b-jets, m_{jjb} represents the invariant mass of the reconstructed top. Then, X_{Wt} is obtained by minimizing over all combinations of two normal jets and one b-jet. During the minimization, 10% uncertainties are used to approximate the invariant mass resolution. Then, events with $X_{Wt} < 1.5$ are excluded from the analysis.

A discriminant X_{HH} is defined to further test the compatibility of events with the $HH \to b\bar{b}b\bar{b}$.

$$X_{HH} = \sqrt{\left(\frac{m_{H_1}-124\,\mathrm{GeV}}{0.1m_{H_1}}\right)^2+\left(\frac{m_{H_2}-117\,\mathrm{GeV}}{0.1m_{H_2}}\right)^2}, \quad (24)$$

where m_{H_1} and m_{H_2} are the masses of the leading and subleading reconstructed Higgs boson candidates, respectively. The reference masses 124 GeV and 117 GeV are obtained from the m_{H_1} and m_{H_2} distribution from the simulation, respectively [84]. Notably, when calculating X_{HH}, 10% uncertainties are also used for the reconstructed invariant masses. Events with $X_{HH} < 1.6$ are considered $HH \to b\bar{b}b\bar{b}$ signal. All the aforementioned selections are used on our events, including signal events with different κ_λ, as well as all backgrounds including the most important QCD backgrounds 4b and 2b2j, along with tt, $ttbb$, ttH, bbH, ZZ, and ZH, as listed in Table 1. The results are shown in Fig. 8 and will be discussed in the following section.

C. Performance

To demonstrate the performance of *EvenT*, we compare the results obtained from the *EvenT*-based analysis with those from a cut-based analysis using kinematic variables discussed earlier, as well as with two ML-based studies: one using DNN [115] and the other using SPA-NET [116]. The comparison is shown in Fig. 8.

A previous study [115] employed a DNN-based model to single out Higgs pair signals and included comprehensive background studies. The model contains two fully connected hidden layers with 250 hidden nodes with the `ReLU` activation function. To prevent overfitting, a `Dropout` layer is added between the two hidden layers, which randomly drops 30% of the nodes during training. The results are given at 68% CL with an integrated luminosity of 3000 fb^{-1} at the HL-LHC. The comparison is shown in the left group of Fig. 8. When $\kappa_\lambda > 0$, even when we do not deal with the threshold to enhance sensitivity, the constraint from *EvenT* is stronger than that from the DNN method. For the negative side, as we only include one negative value in training $\kappa_\lambda = -1$, without imposing the threshold, constraint from *EvenT* is slightly weaker than that from the DNN method. However, when we include a moderate threshold to enhance sensitivity, the result becomes better than that of the DNN.

SPA-NET is an attention-based model [162–164] that employs a stack of transformer encoders to embed input features. These embeddings are then used for jet assignment and event classification via a symmetric tensor attention module, which preserves the permutation symmetry inherent to the input data. SPA-NET was used in the $HH \to b\bar{b}b\bar{b}$ analysis in [116], where the architecture was used both in pairing the b-jets into two Higgs as well as in signal-background discrimination. However, the analysis in [116] considered only the 4b background. Hence, the comparison is made by including only 4b background in *EvenT* analysis. The results are shown in the middle group of Fig. 8, which are given at 95% CL with 300 fb^{-1} luminosity at the LHC. Evidently, with only the 4b background, the results from *EvenT* are better than those from SPA-NET from both sides, even without imposing any threshold. Both SPA-NET and *EvenT* are based on the Transformer framework. Therefore, the results demonstrate the effectiveness of the event-level analysis.

The comparison with the cut-based analysis discussed earlier is also presented in the right group of Fig. 8. Notably, the cut-based and *EvenT*-based analyses are performed on the same set of testing data. The con-

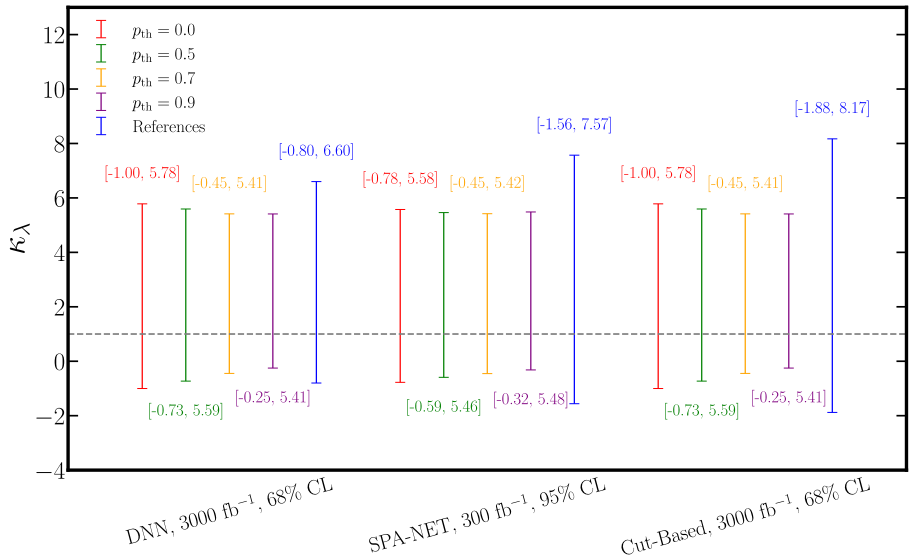


Fig. 8. (color online) Comparison of constraints on κ_λ with three benchmark studies.

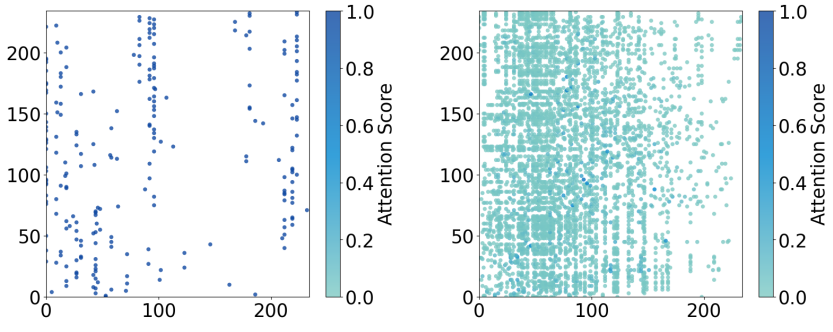
straints on κ_λ are given at 68% CL with 3000 fb^{-1} at 14 TeV. Evidently, the *EvenT* result is already better than that from the cut-based analysis without imposing any threshold to enhance sensitivity. With a stronger threshold in the classification, the result from *EvenT* becomes considerably stronger than that from the general cut-based analysis, which clearly demonstrates the advantage of the ML models in the analysis.

D. Interpretability of the Model

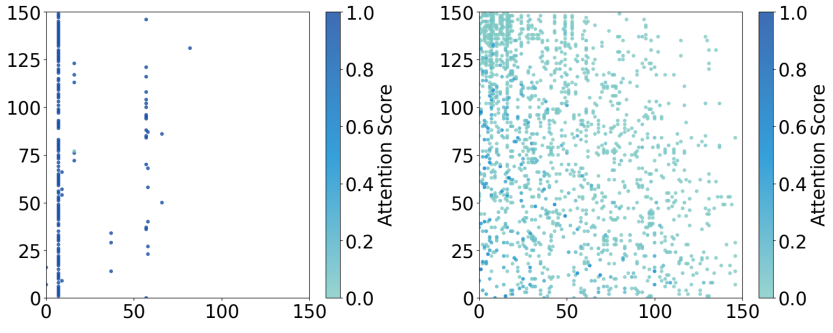
Before we conclude, we discuss the interpretability of the *EvenT* model to understand the internal connections of the model. As mentioned previously, the outstanding performance of Transformer-based models is attributed to the attention mechanism. Hence, we focus on the visualization of the attention mechanism of the *EvenT* model following the method in [165] using the attention matrix in Eq. (12), especially in our case where the inputs contain the information from the entire event instead of fully reconstructed objects. We visualized attention scores from the first attention head of the last particle attention block. The results are shown in Fig. 9, which presents the attention score (element of the attention matrix) of the HH (upper panels) and $2b2j$ (lower panels) processes. As a comparison, we present the attention matrix after (left panels) and before (right panels) training together. Attention scores lower than 0.01 are ignored in the interest of clarity. The attention score is more concentrated for both the HH and $2b2j$ processes with training, which strongly indicates that the model does learn the important relationship between particles. Further, the model focuses on different parts of the particle space for different processes, which is key to distinguish between different processes.

To incorporate the spatial position of particles, particle types, and jets to which they belong during visualization, we visualize the *particle attention graph*, focusing on the final layer of the P-MHA module. Each particle is represented as a point in the η-ϕ plane, as shown in Figs. 10 and 11 for the HH and $2b2j$ processes, respectively. Different marker shapes are used to distinguish particle types: $\times$ for mesons, $\triangle$ for charged hadrons, $\blacktriangledown$ for neutral hadrons, $\bullet$ for photons, and $+$ for electrons. Particles within the same jet, which is reconstructed using the anti-k_t algorithm, are shown in the same color, and particles outside any jet are shown in gray. This color information is used only for visualization. The opacity of each point is proportional to the transverse momentum p_T of the corresponding particle. The color of the connecting lines represents the attention score between particles. Solid lines indicate connections between particles inside jets, and dashed lines indicate connections involving out-of-jet particles. To make the plot clear without excessive low-weight connections, we only include connections with a weight higher than 0.5.

In the upper panels of both Figs. 10 and 11, the connections in the entire η-ϕ plane are shown after (upper left) and before (upper right) the training for the HH and $2b2j$ processes, respectively. Similar to Fig. 9, connections among particles become stronger after training for both the HH and $2b2j$ processes. To be more clear about connections established during training, for each process, we provide the corresponding zoomed-in views of the attention graph in the lower panels of Figs. 10 and 11. Evidently, the training strengthens the connections within jets from the information about the entire event. By comparing these visualized attention graphs, we conclude that the trained model can focus on the important particle pairs in an event and learning relationships within and between jets by training on the event-level instead of reconstructed objects. This capability underpins the

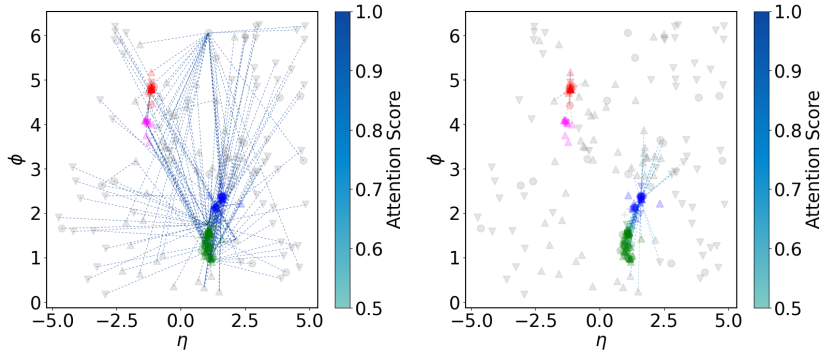


(a) HH Attention Score (Trained) (b) HH Attention Score (Untrained)

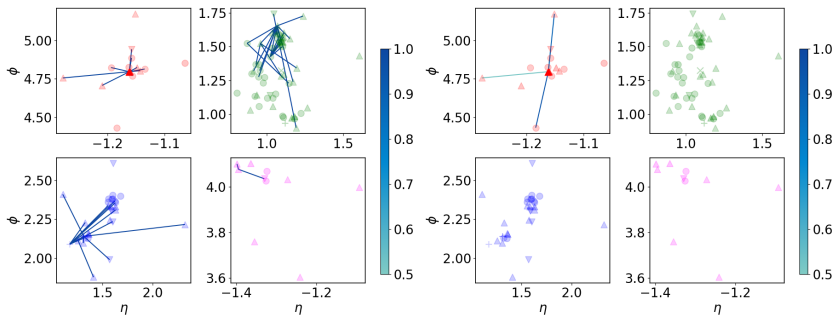


(c) $2b2j$ Attention Score (Trained) (d) $2b2j$ Attention Score (Untrained)

Fig. 9. (color online) Attention score (elements in attention matrix) for HH (upper panels) and $2b2j$ (lower panels) processes after (left panels) and before (right panels) training. Notably, we removed attention scores below 0.01 to make the plot clearer.

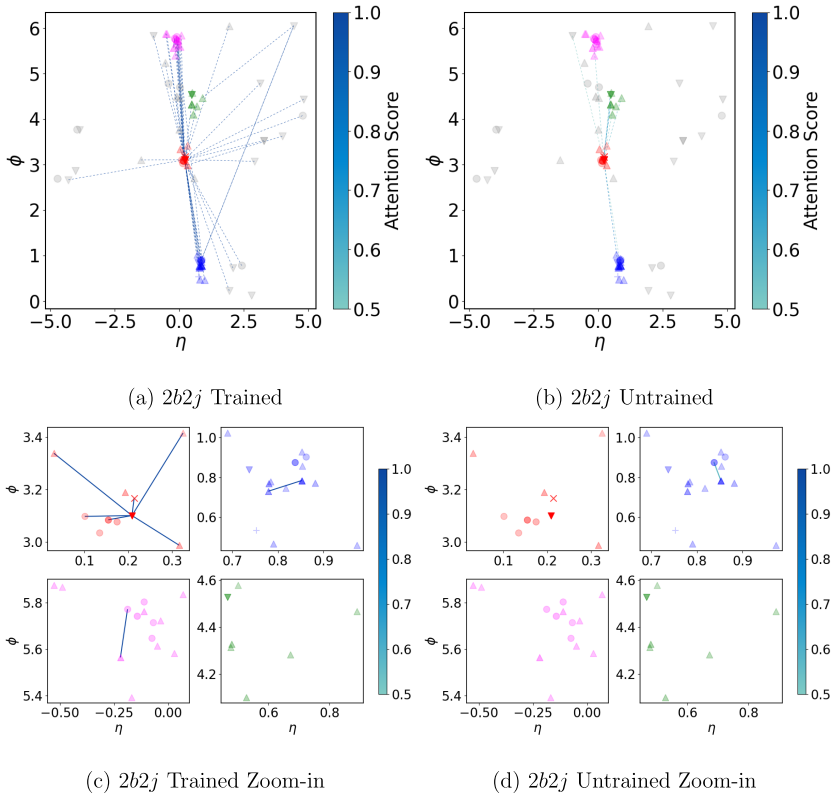


(a) HH Trained (b) HH Untrained



(c) HH Trained Zoom-in (d) HH Untrained Zoom-in

Fig. 10. (color online) Attention graphs for HH processes before and after model training. The lower panels show connections within a single jet.



(a) $2b2j$ Trained (b) $2b2j$ Untrained

(c) $2b2j$ Trained Zoom-in (d) $2b2j$ Untrained Zoom-in

Fig. 11. (color online) Attention graphs for $2b2j$ processes before and after model training. The lower panels show the connections within a single jet.

outstanding performance of the model and the possibility of end-to-end event classification by event-level training. However, in depth studies of what specific feature and relations have been learned during the training are also very important to fully understand the underlying mechanism of the algorithm. We leave this challenging task as a key topic of interest for future research.

V. CONCLUSIONS

In this study, we employed a deep neural network architecture based on *ParT* to enhance the sensitivity of the Higgs self-coupling measurement through $HH \to b\bar{b}b\bar{b}$ at the LHC, which suffers from complex QCD backgrounds. With the help of the attention mechanism, the model can focus on important relationships among input features, thereby enabling robust event classification when trained on full event-level information. The performance of this classifier was evaluated across nine event categories, and it achieved an AUC ≈ 0.9 for most categories, significantly distinguishing one process from the rest. In the current study, we focused on $HH \to b\bar{b}b\bar{b}$ analysis. However, it can be applied to other analyses with very minor modifications.

This study demonstrates the possibility of using the *ParT* model beyond the jet-tagging task. By treating the entire event as a single fat jet, the model achieves high classification accuracy while circumventing the error-prone explicit jet-pairing process inherent in traditional event reconstruction. Such an approach streamlines the analysis by directly utilizing the full event information and demonstrates the potential of such end-to-end analysis in stuides on collider phenomenology.

Applying the model to research on $HH \to b\bar{b}b\bar{b}$ at the HL-LHC, the model constrains the Higgs self-coupling κ_λ to $(-0.25, 5.41)$ at 68% CL, which represents an improvement of approximately 44% in precision compared with the traditional cut-based approach. The results of the comparison against other ML methods further highlight the advantages of the Transformer-based architecture and event-level analysis, particularly in capturing correlations within high-dimensional event data. Further, the attention mechanism naturally supports the interpretability of the model, which can also help to identify the most important features and correlations in attempting to separate different processes.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Huilin Qu, Congqiao Li, Sitian Qian, and Kun Wang for their insightful discussion on the ParT, as well as Huifang Lv for her valuable contributions during the early stages of this project. The authors gratefully acknowledge the valuable

discussions and insights provided by the members of the China Collaboration of Precision Testing and New Physics (CPTNP).

References

[1] ATLAS Collaboration, Phys. Lett. B **716**, 1 (2012), arXiv: 1207.7214
[2] CMS Collaboration, Phys. Lett. B **716**, 30 (2012), arXiv: 1207.7235
[3] CMS Collaboration, Phys. Lett. B **805**, 135425 (2020), arXiv: 2002.06398
[4] ATLAS Collaboration, Phys. Rev. Lett. **131**, 251802 (2023), arXiv: 2308.04775
[5] CMS Collaboration, Nature Phys. **18**, 1329 (2022), arXiv: 2202.06923
[6] ATLAS Collaboration, Nature **607**, 52 (2022), arXiv: 2207.00092
[7] CMS Collaboration, Nature **607**, 60 (2022), arXiv: 2207.00043
[8] H. Abouabid *et al.*, Eur. Phys. J. C **84**, 1183 (2024), arXiv: 2407.03015
[9] C. T. Lu, J. Chang, K. Cheung *et al.*, JHEP **08**, 133 (2015), arXiv: 1505.00957
[10] F. Kling, T. Plehn, and P. Schichtel, Phys. Rev. D **95**, 035026 (2017), arXiv: 1607.07441
[11] J. Alison *et al.*, Rev. Phys. **5**, 100045 (2020), arXiv: 1910.00012
[12] U. Baur, T. Plehn, and D.L. Rainwater, Phys. Rev. D **69**, 053004 (2004), arXiv: hep-ph/0310056
[13] G. Degrassi, B. Di Micco, P.P. Giardino *et al.*, Phys. Lett. B **817**, 136307 (2021), arXiv: 2102.07651
[14] H. Bahl, J. Braathen, M. Gabelmann *et al.*, PoS **EPS-HEP2023**, 407 (2024), arXiv: 2311.01134
[15] F. Arco, S. Heinemeyer, M. Mühlleitner *et al.*, Eur. Phys. J. C **83**, 1019 (2023), arXiv: 2212.11242
[16] A. Papaefstathiou, L.L. Yang, and J. Zurita, Phys. Rev. D **87**, 011301 (2013), arXiv: 1209.1489
[17] U. Baur, T. Plehn, and D.L. Rainwater, Phys. Rev. Lett. **89**, 151801 (2002), arXiv: hep-ph/0206024
[18] J. Baglio, A. Djouadi, R. Gröber *et al.*, JHEP **04**, 151 (2013), arXiv: 1212.5581
[19] M.J. Dolan, C. Englert, and M. Spannowsky, Phys. Rev. D **87**, 055002 (2013), arXiv: 1210.8166
[20] G. Durieux, M. McCullough, and E. Salvioni, JHEP **12**, 148 (2022), arXiv: 2209.00666
[21] K. Cheung, A. Jueid, C. T. Lu *et al.*, Phys. Rev. D **103**, 015019 (2021), arXiv: 2003.11043
[22] S. Chang and M.A. Luty, JHEP **03**, 140 (2020), arXiv: 1902.05556
[23] R.S. Gupta, H. Rzehak, and J.D. Wells, Phys. Rev. D **88**, 055024 (2013), arXiv: 1305.6397
[24] P.N. Bhattiprolu and J.D. Wells, Phys. Lett. B **861**, 139263 (2025), arXiv: 2407.11847
[25] S. Dawson, A. Ismail, and I. Low, Phys. Rev. D **91**, 115008 (2015), arXiv: 1504.05596
[26] S. Banerjee, B. Batell, and M. Spannowsky, Phys. Rev. D **95**, 035009 (2017), arXiv: 1608.08601
[27] L. Alasfar *et al.*, SciPost Phys. Comm. Rep. **2024**, 2 (2024), arXiv: 2304.01968
[28] F. Goertz, A. Papaefstathiou, L.L. Yang *et al.*, JHEP **04**, 167 (2015), arXiv: 1410.3471
[29] A. Azatov, R. Contino, G. Panico *et al.*, Phys. Rev. D **92**, 035001 (2015), arXiv: 1502.00539
[30] Q. H. Cao, B. Yan, D. M. Zhang *et al.*, Phys. Lett. B **752**, 285 (2016), arXiv: 1508.06512
[31] G. Li, L. X. Xu, B. Yan *et al.*, Phys. Lett. B **800**, 135070 (2020), arXiv: 1904.12006
[32] T. Liu, K. F. Lyu, J. Ren *et al.*, Phys. Rev. D **98**, 093004 (2018), arXiv: 1803.04359
[33] W. Bizoń, U. Haisch, and L. Rottoli, JHEP **10**, 267 (2019), arXiv: 1810.04665
[34] C. Y. Chen, Q. S. Yan, X. Zhao *et al.*, Phys. Rev. D **93**, 013007 (2016), arXiv: 1510.04013
[35] B. Fuks, J.H. Kim, and S.J. Lee, Phys. Lett. B **771**, 354 (2017), arXiv: 1704.04298
[36] M. Chiesa, F. Maltoni, L. Mantani *et al.*, JHEP **09**, 098 (2020), arXiv: 2003.13628
[37] J. M. Torndal, J. List, D. Ntounis *et al.*, PoS **EPS-HEP2023**, 406 (2024), arXiv: 2311.16774
[38] J.M. Torndal and J. List, *Higgs self-coupling measurement at the International Linear Collider*, in *International Workshop on Future Linear Colliders*, 7, (2023), arXiv: 2307.16515
[39] J. Davies, K. Schönwald, M. Steinhauser *et al.*, JHEP **10**, 033 (2023), arXiv: 2308.01355
[40] J. Davies, PoS **LL2024**, 015 (2024), arXiv: 2407.08264
[41] H. Zhang, K. Schönwald, M. Steinhauser *et al.*, PoS **LL2024**, 014 (2024), arXiv: 2407.05787
[42] G. Heinrich, S. Jones, M. Kerner *et al.*, JHEP **11**, 040 (2024), arXiv: 2407.04653
[43] A. Carvalho, M. Dall'Osso, T. Dorigo *et al.*, JHEP **04**, 126 (2016), arXiv: 1507.02245
[44] J. Davies, K. Schönwald, M. Steinhauser *et al.*, JHEP **04**, 193 (2025), arXiv: 2501.17920
[45] A. Adhikary, S. Banerjee, R.K. Barman *et al.*, JHEP **07**, 116 (2018), arXiv: 1712.05346
[46] J.H. Kim, K. Kong, K.T. Matchev *et al.*, Phys. Rev. Lett. **122**, 091801 (2019), arXiv: 1807.11498
[47] W. Bizoń, U. Haisch, L. Rottoli *et al.*, JHEP **02**, 170 (2024), arXiv: 2402.03463
[48] K. Nakamura, K. Nishiwaki, K.-y. Oda *et al.*, Eur. Phys. J. C **77**, 273 (2017), arXiv: 1701.06137
[49] A. Adhikary, R.K. Barman, and B. Bhattacherjee, JHEP **12**, 179 (2020), arXiv: 2006.11879
[50] L. B. Chen, H.T. Li, H. S. Shao *et al.*, Phys. Lett. B **803**, 135292 (2020), arXiv: 1909.06808
[51] J. Baglio, F. Campanario, S. Glaus *et al.*, JHEP **04**, 181 (2020), arXiv: 2003.03227
[52] J. Baglio, F. Campanario, S. Glaus *et al.*, Eur. Phys. J. C **79**, 459 (2019), arXiv: 1811.05692
[53] S. Borowka, N. Greiner, G. Heinrich, S.P. Jones, M. Kerner, J. Schlenk, *et al.*, Phys. Rev. Lett. **117**, 012001 (2016), arXiv: 1604.06447
[54] M.M. Czuryło, *Search for the non-resonant Vector Boson Fusion production of the Higgs boson pairs decaying to $b\bar{b}b\bar{b}$ final state using the ATLAS detector*, Ph.D. thesis, U. Heidelberg (main), (2023). 10.11588/heidok.00032906
[55] Q. H. Cao, Y. Liu, and B. Yan, Phys. Rev. D **95**, 073006 (2017), arXiv: 1511.03311
[56] K. Chai, J. H. Yu, and H. Zhang, Phys. Rev. D **107**, 055031 (2023), arXiv: 2210.14929

[57] M.J. Dolan, C. Englert, N. Greiner *et al.*, Phys. Rev. Lett. **112**, 101802 (2014), arXiv: 1310.1084

[58] J. Chen, T. Li, C. T. Lu *et al.*, Phys. Rev. D **105**, 053009 (2022), arXiv: 2112.12507

[59] B. Henning, D. Lombardo, M. Riembau *et al.*, Phys. Rev. Lett. **123**, 181801 (2019), arXiv: 1812.09299

[60] M. McCullough, Phys. Rev. D **90**, 015001 (2014), arXiv: 1312.3322

[61] M. Gorbahn and U. Haisch, JHEP **10**, 094 (2016), arXiv: 1607.03773

[62] G. Degrassi, P. P. Giardino, F. Maltoni *et al.*, JHEP **12**, 080 (2016), arXiv: 1607.04251

[63] J. Gao, X. M. Shen, G. Wang *et al.*, Phys. Rev. D **107**, 115017 (2023), arXiv: 2302.04160

[64] G. Degrassi, M. Fedele, and P.P. Giardino, JHEP **04**, 155 (2017), arXiv: 1702.01737

[65] L. Alasfar, J. de Blas, and R. Gröber, JHEP **05**, 111 (2022), arXiv: 2202.02333

[66] H. Abouabid, A. Arhrib, D. Azevedo, J.E. Falaki, P.M. Ferreira, M. Mühlleitner, *et al.*, JHEP **09**, 011 (2022), arXiv: 2112.12515

[67] S.I. Godunov, M.I. Vysotsky, and E.V. Zhemchugov, J. Exp. Theor. Phys. **120**, 369 (2015), arXiv: 1408.0184

[68] A.V. Kotwal, S. Chekanov, and M. Low, Phys. Rev. D **91**, 114018 (2015), arXiv: 1504.08042

[69] L. Di Luzio, R. Gröber, and M. Spannowsky, Eur. Phys. J. C **77**, 788 (2017), arXiv: 1704.02311

[70] R. Grober, M. Muhlleitner, and M. Spira, Nucl. Phys. B **925**, 1 (2017), arXiv: 1705.05314

[71] A. Adhikary, S. Banerjee, R. Kumar Barman *et al.*, JHEP **09**, 068 (2019), arXiv: 1812.05640

[72] E. Arganda, M. Epele, N.I. Mileo *et al.*, Eur. Phys. J. Plus **139**, 615 (2024), arXiv: 2401.03178

[73] T. Huang, J. M. No, L. Pernié *et al.*, Phys. Rev. D **96**, 035007 (2017), arXiv: 1701.04442

[74] J. M. No and M. Ramsey-Musolf, Phys. Rev. D **89**, 095031 (2014), arXiv: 1310.6035

[75] V. Barger, L. L. Everett, C. B. Jackson *et al.*, Phys. Rev. Lett. **114**, 011801 (2015), arXiv: 1408.0003

[76] C. Y. Chen, S. Dawson, and I. M. Lewis, Phys. Rev. D **91**, 035015 (2015), arXiv: 1410.5488

[77] D. Barducci, K. Mimasu, J. M. No *et al.*, JHEP **02**, 002 (2020), arXiv: 1910.08574

[78] S. Heinemeyer, M. Mühlleitner, K. Radchenko *et al.*, Eur. Phys. J. C **85**, 437 (2025), arXiv: 2403.14776

[79] A. Arhrib, S. Moretti, S. Semlali *et al.*, *Probing a 2HDM Type-I light Higgs state via* $H_{\rm SM} \to hh \to b\bar{b}\gamma\gamma$ *at the LHC*, arXiv: 2412.06052

[80] J. Baglio, O. Eberhardt, U. Nierste *et al.*, Phys. Rev. D **90**, 015008 (2014), arXiv: 1403.1264

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

[82] M. Bauer, M. Carena, and A. Carmona, Phys. Rev. Lett. **121**, 021801 (2018), arXiv: 1801.00363

[83] E. Arganda, J. L. Díaz-Cruz, N. Mileo *et al.*, Nucl. Phys. B **929**, 171 (2018), arXiv: 1710.07254

[84] ATLAS Collaboration, Phys. Rev. D **108**, 052003 (2023), arXiv: 2301.03212

[85] ATLAS Collaboration, JHEP **07**, 040 (2023), arXiv: 2209.10910

[86] ATLAS Collaboration, Phys. Rev. D **106**, 052001 (2022), arXiv: 2112.11876

[87] ATLAS Collaboration, *Probing the nature of electroweak symmetry breaking with Higgs boson pairs in ATLAS*, in *30th International Workshop on Deep-Inelastic Scattering and Related Subjects*, 7, (2023), arXiv: 2307.11467

[88] CMS Collaboration, JHEP **06**, 130 (2023), arXiv: 2206.10657

[89] CMS Collaboration, JHEP **03**, 257 (2021), arXiv: 2011.12373

[90] CMS Collaboration, Phys. Lett. B **842**, 137531 (2023), arXiv: 2206.09401

[91] CMS Collaboration, Phys. Rev. Lett. **129**, 081802 (2022), arXiv: 2202.09617

[92] J. Chang, K. Cheung, J. S. Lee *et al.*, Phys. Rev. D **100**, 096001 (2019), arXiv: 1804.07130

[93] ATLAS Collaboration, *Projected sensitivity of measurements of Higgs boson pair production with the ATLAS experiment at the HL-LHC*, ATL-PHYS-PUB-2025-006

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

[95] A.J. Barr, M.J. Dolan, C. Englert *et al.*, Phys. Lett. B **728**, 308 (2014), arXiv: 1309.6318

[96] D. Gonçalves, T. Han, F. Kling *et al.*, Phys. Rev. D **97**, 113004 (2018), arXiv: 1802.04319

[97] A. Taliercio, P. Mastrapasqua, C. Caputo *et al.*, *Higgs Self Couplings Measurements at Future proton-proton Colliders: a Snowmass White Paper*, in Snowmass 2021, 3, (2022), arXiv: 2203.08042

[98] S. Braibant, PoS **EPS-HEP2021**, 616 (2022)

[99] M. L. Mangano, G. Ortona, and M. Selvaggi, Eur. Phys. J. C **80**, 1030 (2020), arXiv: 2004.03505

[100] B. Stapf, A. Taliercio, E. Gallo *et al.*, PoS **EPS-HEP2023**, 413 (2024), arXiv: 2312.03513

[101] FCC Collaboration, PoS **PANIC2021**, 420 (2022)

[102] FCC Collaboration, PoS **EPS-HEP2023**, 391 (2024)

[103] S.-i. Kawada, PoS **PANIC2021**, 396 (2022)

[104] J. List, B. Bliewert, D. Ntounis *et al.*, PoS **ICHEP2024**, 079 (2025), arXiv: 2411.01507

[105] Muon Collider Physics and Detector Working Group Collaboration, PoS **EPS-HEP2021**, 619 (2022)

[106] L. Buonincontri, Nuovo Cim. C **44**, 31 (2021)

[107] T. Han, D. Liu, I. Low *et al.*, Phys. Rev. D **103**, 013002 (2021), arXiv: 2008.12204

[108] L. Buonincontri, M. Casarsa, L. Giambastiani *et al.*, Nuovo Cim. C **47**, 288 (2024)

[109] F. Goertz, A. Papaefstathiou, L.L. Yang *et al.*, JHEP **06**, 016 (2013), arXiv: 1301.3492

[110] J. K. Behr, D. Bortoletto, J. A. Frost *et al.*, Eur. Phys. J. C **76**, 386 (2016), arXiv: 1512.08928

[111] D. Wardrope, E. Jansen, N. Konstantinidis *et al.*, Eur. Phys. J. C **75**, 219 (2015), arXiv: 1410.2794

[112] D.E. Ferreira de Lima, A. Papaefstathiou, and M. Spannowsky, JHEP **08**, 030 (2014), arXiv: 1404.7139

[113] H. L. Li, M. Ramsey-Musolf, and S. Willocq, Phys. Rev. D **100**, 075035 (2019), arXiv: 1906.05289

[114] A. Alves, D. Gonçalves, T. Ghosh *et al.*, JHEP **03**, 053 (2020), arXiv: 1909.05268

[115] J. Amacker *et al.*, JHEP **12**, 115 (2020), arXiv: 2004.04240

[116] C. W. Chiang, F. Y. Hsieh, S. C. Hsu *et al.*, JHEP **09**, 139 (2024), arXiv: 2401.14198

[117] J. Alison, S. An, P. Bryant *et al.*, *End-to-end particle and event identification at the Large Hadron Collider with CMS Open Data*, in *Meeting of the Division of Particles and Fields of the American Physical Society*, 10, (2019),

arXiv: 1910.07029
[118] M. Andrews, J. Alison, S. An, P. Bryant, B. Burkle, S. Gleyzer, *et al*., Nucl. Instrum. Meth. A **977**, 164304 (2020), arXiv: 1902.08276
[119] M. Andrews, M. Paulini, S. Gleyzer *et al*., Comput. Softw. Big Sci. **4**, 6 (2020), arXiv: 1807.11916
[120] P. Baldi, P. Sadowski, and D. Whiteson, Nature Commun. **5**, 4308 (2014), arXiv: 1402.4735
[121] M. Abdughani, J. Ren, L. Wu *et al*., Commun. Theor. Phys. **71**, 955 (2019), arXiv: 1905.06047
[122] D. Guest, J. Collado, P. Baldi *et al*., Phys. Rev. D **94**, 112002 (2016), arXiv: 1607.08633
[123] A. Butter *et al*., SciPost Phys. **7**, 014 (2019), arXiv: 1902.09914
[124] N. S. Woodward, S. E. Park, G. Grosso *et al*., *Product Manifold Machine Learning for Physics*, arXiv: 2412.07033
[125] J. Lin, M. Freytsis, I. Moult *et al*., JHEP **10**, 101 (2018), arXiv: 1807.10768
[126] J. Cogan, M. Kagan, E. Strauss *et al*., JHEP **02**, 118 (2015), arXiv: 1407.5675
[127] L. de Oliveira, M. Kagan, L. Mackey *et al*., JHEP **07**, 069 (2016), arXiv: 1511.05190
[128] J. S.H. Lee, I. Park, I. J. Watson *et al*., J. Korean Phys. Soc. **74**, 219 (2019), arXiv: 2012.02531
[129] C. F. Madrazo, I. H. Cacha, L. L. Iglesias *et al*., EPJ Web Conf. **214**, 06017 (2019), arXiv: 1708.07034
[130] F. A. Dreyer and H. Qu, JHEP **03**, 052 (2021), arXiv: 2012.08526
[131] M. Abdughani, J. Ren, L. Wu *et al*., JHEP **08**, 055 (2019), arXiv: 1807.09088
[132] S. Gong, Q. Meng, J. Zhang, H. Qu, C. Li, S. Qian, *et al*., JHEP **07**, 030 (2022), arXiv: 2201.08187
[133] H. Qu and L. Gouskos, Phys. Rev. D **101**, 056019 (2020), arXiv: 1902.08570
[134] V. Mikuni and F. Canelli, Mach. Learn. Sci. Tech. **2**, 035027 (2021), arXiv: 2102.05073
[135] P. T. Komiske, E. M. Metodiev, and J. Thaler, JHEP **01**, 121 (2019), arXiv: 1810.05165
[136] M. Usman, M. H. Shahid, M. Ejaz *et al*., *Particle Multi-Axis Transformer for Jet Tagging*, arXiv: 2406.06638
[137] H. Qu, C. Li, and S. Qian, *Particle Transformer for Jet Tagging*, arXiv: 2202.03772
[138] R. Tagami, T. Suehara, and M. Ishino, EPJ Web Conf. **315**, 03011 (2024), arXiv: 2410.11322
[139] CMS Collaboration, Nuovo Cim. C **47**, 250 (2024)
[140] Y. Wu, K. Wang, C. Li *et al*., Chin. Phys. C **49**, 013110 (2025), arXiv: 2407.08682
[141] M. He and D. Wang, Eur. Phys. J. C **83**, 1116 (2023), arXiv: 2307.04723
[142] A. Apresyan *et al*., *Improving Di-Higgs Sensitivity at Future Colliders in Hadronic Final States with Machine Learning*, in *Snowmass 2021*, 3, (2022), arXiv: 2203.07353
[143] P. Stylianou and G. Weiglein, Eur. Phys. J. C **84**, 366 (2024), arXiv: 2312.04646
[144] L. Alasfar, R. Gröber, C. Grojean *et al*., JHEP **11**, 045 (2022), arXiv: 2207.04157
[145] L. Huang, S.-b. Kang, J.H. Kim *et al*., JHEP **08**, 114 (2022), arXiv: 2203.11951
[146] B. Tannenwald, C. Neu, A. Li *et al*., *Benchmarking Machine Learning Techniques with Di-Higgs Production at the LHC*, arXiv: 2009.06754
[147] M. Abdughani, D. Wang, L. Wu *et al*., Phys. Rev. D **104**, 056003 (2021), arXiv: 2005.11086
[148] A. Hammad, S. Moretti, and M. Nojiri, JHEP **03**, 144 (2024), arXiv: 2401.00452
[149] M. Kerner, PoS **LL2024**, 016 (2024)
[150] H.T. Li, Z. G. Si, J. Wang *et al*., Chin. Phys. C **49**, 023107 (2025), arXiv: 2407.14716
[151] R. Frederix, S. Frixione, V. Hirschi *et al*., Phys. Lett. B **732**, 142 (2014), arXiv: 1401.7340
[152] G. Heinrich, S.P. Jones, M. Kerner *et al*., JHEP **06**, 066 (2019), arXiv: 1903.08137
[153] D. de Florian and J. Mazzitelli, Phys. Rev. Lett. **111**, 201801 (2013), arXiv: 1309.6594
[154] D. de Florian and J. Mazzitelli, JHEP **09**, 053 (2015), arXiv: 1505.07122
[155] D. de Florian and J. Mazzitelli, Phys. Lett. B **724**, 306 (2013), arXiv: 1305.5206
[156] J. Alwall, R. Frederix, S. Frixione *et al*., JHEP **07**, 079 (2014), arXiv: 1405.0301
[157] T. Sjöstrand, S. Ask, J. R. Christiansen *et al*., Comput. Phys. Commun. **191**, 159 (2015), arXiv: 1410.3012
[158] DELPHES 3 Collaboration, JHEP **02**, 057 (2014), arXiv: 1307.6346
[159] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012), arXiv: 1111.6097
[160] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063 (2008), arXiv: 0802.1189
[161] A. Vaswani, N. Shazeer, N. Parmar *et al*., *Attention is all you need*, arXiv: 1706.03762
[162] M.J. Fenton, A. Shmakov, T. W. Ho *et al*., Phys. Rev. D **105**, 112008 (2022), arXiv: 2010.09206
[163] A. Shmakov, M. J. Fenton, T. W. Ho *et al*., SciPost Phys. **12**, 178 (2022), arXiv: 2106.03898
[164] M. J. Fenton, A. Shmakov, H. Okawa *et al*., Commun. Phys. **7**, 139 (2024), arXiv: 2309.01886
[165] A. Wang, A. Gandrakota, J. Ngadiuba *et al*., *Interpreting Transformers for Jet Tagging*, 12, (2024), arXiv: 2412.03673