

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§}

¹Department of Physics, Institute of Theoretical Physics and Institute of Physics Frontiers and Interdisciplinary Sciences, Nanjing Normal University, Nanjing, 210023, China

²Nanjing Key Laboratory of Particle Physics and Astrophysics, Nanjing, 210023, China

Abstract: The Higgs self-coupling is crucial for understanding the structure of the scalar potential and the mechanism of electroweak symmetry breaking. In this work, utilizing deep neural network based on Particle Transformer that relies on attention mechanism, we present a comprehensive analysis of the measurement of the trilinear Higgs self-coupling through the Higgs pair production with subsequent decay into four b -quarks ($HH \rightarrow b\bar{b}b\bar{b}$) at the LHC. The model processes full event-level information as input, bypassing explicit jet pairing and can serve as an event classifier. At HL-LHC, our approach constrains the κ_λ to $(-0.25, 5.41)$ at 68% CL achieving around 44% improvement in precision over conventional cut-based analyses. Comparison against alternative machine learning architectures also shows the outstanding performance of the Transformer-based model, which is mainly due to its ability to capture the correlations in the high-dimensional collision data with the help of attention mechanism. The result highlights the potential of attention-based networks and the end-to-end event classifier in collider phenomenology.

Keywords: Higgs Physics, Higgs Pair Production, Machine Learning, Collider Phenomenology

DOI: CSTR:

I. INTRODUCTION

The Standard Model (SM) is currently the most successful theoretical framework for describing fundamental particles and their interactions, with the Higgs particle being the core of the model. The Higgs field, through the Higgs mechanism, breaks the Electroweak symmetry spontaneously and explains how fundamental particles acquire mass. Since the discovery of the 125 GeV Higgs particle by the ATLAS and CMS experiments in 2012 [1, 2], precise measurements of the properties of the Higgs, 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 and provides an essential test of the SM. Despite the great success of the SM, there are still phenomena that cannot be explained in SM, such as the existence of dark matter, the nonzero mass of neutrinos, and the matter-antimatter asymmetry etc. These phenomena suggest the existence of the physics beyond the SM (BSM), and the discovery of the Higgs particle has opened a window for exploring

new physics. Any deviation between the measured properties of the Higgs particle and the predictions of the SM would be crucial for us to explore the physics beyond the SM.

Currently, the ATLAS and CMS experiments have already performed various measurements about the Higgs. The mass of the Higgs has been precisely measured by both experiments [3, 4]. The width of the Higgs has also been precisely measured through the off-shell effect [5, 6] which was originally thought not possible at hadron collider. All major production modes and decay channels of the Higgs have been measured [6, 7] with the overall signal strength in agreement with the SM predictions. From these measurements, the interactions of the Higgs with other SM particles can be extracted which are crucial to test the Higgs mechanism.

Measurement of the Higgs self-coupling [8–17] is also an important task in probing the properties of the Higgs boson. The Higgs self-coupling determines the shape of the Higgs potential, which is crucial for electroweak symmetry breaking [18, 19]. In the SM, the

Received 3 September 2025; Accepted 25 November 2025

* This work is supported by the National Natural Science Foundation of China (NNSFC) under grant No. 12305112

[†] E-mail: ycwu@njnu.edu.cn

[‡] E-mail: lxiao@njnu.edu.cn

[§] E-mail: zyan@njnu.edu.cn



Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Article funded by SCOAP³ and published under licence by Chinese Physical Society and the Institute of High Energy Physics of the Chinese Academy of Sciences and the Institute of Modern Physics of the Chinese Academy of Sciences and IOP Publishing Ltd

Higgs potential is expressed as:

$$V = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \quad (1)$$

with

$$\Phi = \begin{pmatrix} \phi^+ \\ \frac{v + H + i\phi^0}{\sqrt{2}} \end{pmatrix}. \quad (2)$$

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

$$V = \frac{1}{2} m_H^2 H^2 + \lambda_{HHH} v H^3 + \frac{1}{4} \lambda_{HHHH} H^4, \quad (3)$$

where $\lambda_{HHH} = \lambda_{HHHH} = \frac{m_H^2}{2v^2}$ represent the trilinear and quartic Higgs self-couplings. It is clear that in the SM, the Higgs self-couplings are fully determined by the mass and vev of the Higgs which have been measured with a high precision. Any deviation from the SM prediction indicates the presence of new physics [20–26]. In this work, we use the kappa framework to parameterize the deviation

$$V = \frac{1}{2} m_H^2 H^2 + \kappa_\lambda \lambda_{HHH} v H^3 + \frac{1}{4} \kappa_{\lambda,4} \lambda_{HHHH} H^4. \quad (4)$$

where κ_λ and $\kappa_{\lambda,4}$ indicate the deviations in trilinear and quartic Higgs self-couplings respectively. The Effective Field Theory (EFT) framework provides another powerful and model-independent approach to studying the Higgs self-couplings [27–31].

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

Direct measurement of the trilinear Higgs self-coupling relies on the Higgs pair productions [37–49] to which the gluon-gluon fusion (ggF) process provides the dominant contribution at the LHC [50–53]. The Higgs pair production from vector boson fusion (VBF) has also been investigated for the measurement of Higgs self-couplings [18, 54]. In addition to ggF and VBF production of a pair

of Higgs, many other processes are considered to probe the Higgs self-coupling including the double Higgs-strahlung process [55], Higgs pair associated with two top quarks [18], Higgs pair plus jets production [56, 57]. These processes have smaller cross sections than ggF production, but they can still contribute to the Higgs self-coupling measurement at future colliders with higher energy and luminosity. Further, the single Higgs production and its decay can also be utilized to measure the Higgs self-coupling through its contributions at loop level [58–65]. On the other hand, the Higgs pair production can be significantly altered in extended Higgs sectors where the Higgs pair is produced resonantly [66–72]. The studies of the Higgs pair production can help probing the scalar potential in these models, e.g. xSM [73–76], 2HDM [77–82] and MSSM [83]. However, in this work, we will focus on the non-resonant production of Higgs pair and leave the resonant production to future work.

Both ATLAS and CMS have conducted studies on the Higgs pair production process. In the ATLAS analysis, the decay channels $HH \rightarrow b\bar{b}b\bar{b}$ [84], $HH \rightarrow b\bar{b}\tau^+\tau^-$ [85], and $HH \rightarrow b\bar{b}\gamma\gamma$ [86] were examined. The results indicate that the κ_λ is constrained to the range $(-0.6, 6.6)$ after combining all these channels [87]. In the CMS analysis, the decay channels $HH \rightarrow b\bar{b}ZZ^*$ [88], $HH \rightarrow \text{Multilepton}$, $HH \rightarrow b\bar{b}\gamma\gamma$ [89], $HH \rightarrow b\bar{b}\tau^+\tau^-$ [90] and $HH \rightarrow b\bar{b}b\bar{b}$ [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 under higher center-of-mass energy and integrated luminosity. The HL-LHC [92–95] is expected to have a precise measurement of the Higgs self-coupling $\kappa_\lambda = 1.0^{+0.48}_{-0.42}$, while the HE-LHC [96] and FCC-hh [97–100] are projected to achieve a precision of 5% on the Higgs self-coupling. The future lepton collider can also probe the Higgs self-coupling with about 20% precision from FCC-ee [101, 102], ILC [103, 104] as well as 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 (about 33.6%) for a Higgs pair system [109], making it a promising channel for investigation. The current constraints on the Higgs self-coupling are however relatively broad from $4b$ channel $\kappa_\lambda \in (-3.3, 11.4)$ [87], which is weak compared with other channels and mainly limited by the complex background at the LHC. Improving the discriminating power between signal and background in $4b$ channel will hence significantly enhance the sensitivity. There are already many studies focusing on improving the sensitivity in this channel [110–116]. In this work, we explore the possibility of using machine learning techniques to tackle the challenge in $4b$ channel of Higgs pair searches at the LHC.

The amount of data generated at colliders is enorm-

ous, and machine learning (ML) is particularly well-suited for handling such large-scale datasets [117–119]. There have been various ML applications in particle physics [120, 121], such as jet-tagging [122–125] including CNN based on image data [126–129], GNN based on graph data [130–132], ParticleNet and Energy Flow Network based on particle clouds [133–136], and the current state-of-the-art (SOTA) Transformer-based model, Particle Transformer (*ParT*) [137–141]. Similarly, for the measurements of Higgs self-coupling, ML techniques are also used [115, 116, 142–144]. Additionally, there are studies that investigate the performance of different machine learning algorithms in Higgs self-coupling measurement [145–147].

For the $HH \rightarrow b\bar{b}b\bar{b}$ channel, different machine learning architectures have been applied. For instance, Ref. [115] employing a DNN architecture demonstrated 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 is also used for the resonant Higgs pair production study with $4b$ final states [148]. Considering the excellent performance of *ParT* in jet tagging tasks [137], in this work, a modified version of *ParT* including entire event information as input for event classification will be used to improve the sensitivity on the Higgs self-coupling through $HH \rightarrow b\bar{b}b\bar{b}$ channel. However, we need to emphasize that the analysis included in this study demonstrates the effectiveness of not only the transformer structure on the event selection, but also the end-to-end analysis utilizing the event-level information.

This paper is organized as follows. In Sec. II, we briefly describe all the processes we will consider in this work. In Sec. III, after the introduction of the structure of *ParT*, we will discuss the modification we will make and the details of the training process. The performance of the modified model on event classification will also be 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 also

discussed at the end of this section. Finally, we summarize 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 pair [149–151], with both Higgs decays into b -quark pair, which contains, at leading order (LO), the 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, while the box diagrams depend only on the Yukawa couplings of the fermion in the loop. In our simulations, we include the contributions from the third generation quarks, but with fixed Yukawa coupling at corresponding SM value. On the other hand, the trilinear Higgs self-coupling is allowed to be different with the SM value. The total Higgs pair production cross section at given energy will be a quadratic function of κ_λ , and is given as at both $\sqrt{s} = 13 \text{ TeV}$ and $\sqrt{s} = 14 \text{ TeV}$:

$$\sigma_{HH}^{\text{LO}}(\kappa_\lambda) = \begin{cases} 4.15 \times 10^{-3} \kappa_\lambda^2 - 2.02 \times 10^{-2} \kappa_\lambda + 3.05 \times 10^{-2} \text{ pb} & \sqrt{s} = 13 \text{ TeV}, \\ 4.89 \times 10^{-3} \kappa_\lambda^2 - 2.39 \times 10^{-2} \kappa_\lambda + 3.62 \times 10^{-2} \text{ pb} & \sqrt{s} = 14 \text{ TeV}. \end{cases} \quad (5)$$

where the first term depending on κ_λ^2 in each line is the pure triangle contributions, the last term without any dependence on κ_λ comes from the box diagrams, while the middle term with linear dependence on κ_λ comes from the 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 about 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 choices of κ_λ at different energies are shown in Fig. 2.

For the backgrounds, we will mainly focus on the processes containing multi b -quarks and/or multi-jets. In

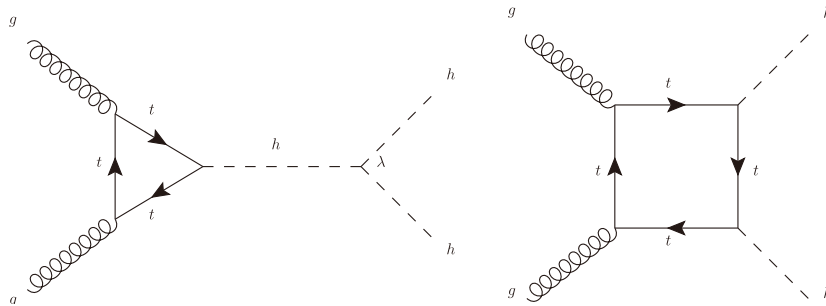


Fig. 1. The leading-order Feynman diagrams of the di-Higgs production in the SM.

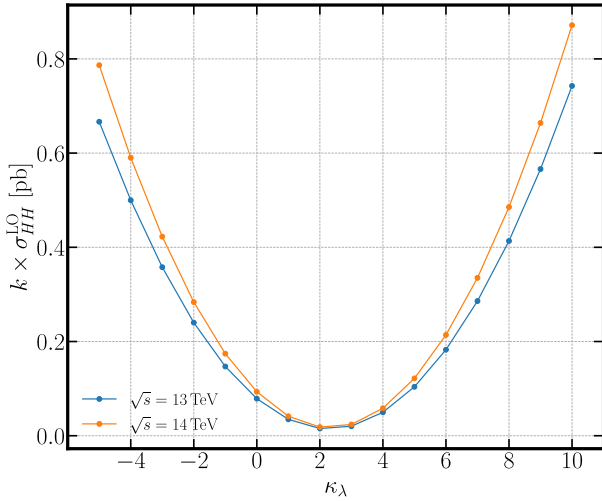


Fig. 2. (color online) The Higgs pair production cross section as function of κ_λ at the LHC ($\sqrt{s} = 13$ TeV and 14 TeV), assuming all other couplings are fixed, with higher order correction k-factor included.

our analysis, the main backgrounds are the QCD production of four b quarks ($4b$) as well as two b quarks together with two light quarks ($2b2j$). Other than these two major backgrounds, we also consider the production of b -quarks from the decay of 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 section 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, the NLO k-factors are also listed [115, 152–155].

The events for both signal process with possible different values of κ_λ and background processes are simulated by MadGraph [156] with both $\sqrt{s} = 13$ TeV and

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

Process	σ^{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	—

$\sqrt{s} = 14$ TeV. To increase the simulation efficiency, we use a moderate parton-level basic cuts: $p_T^{j,b} > 20$ GeV, $|\eta^{j,b}| < 4$. The Pythia [157] is used for the hadronization and showering followed by the detector simulation by Delphes [158]. FastJet [159] is further linked to reconstruct the Jets from the particle-flow output from Delphes using anti- k_t algorithm [160] with two different values of ΔR , 0.5 for a slim jet originated from quarks/gluon, 1.0 for a fat jet originated mainly from heavier objects, e.g. Higgs, W/Z boson, top quarks.

III. TRANSFORMER ON EVENT

In this study, instead of working on the individual objects (jets, charged leptons, photons etc.) within an event, we investigate the possibility of directly working on the whole event. The Particle Transformer (*ParT*) [137] is the state-of-the-arts (SOTA) object tagging machine learning (ML) algorithm, which was trained on a 10-label-classification task. By implementing *ParT*, we could in principle tag the Higgs, top, W/Z boson as a fat jet in the events, and then combine them to extract the event of signal (Higgs pair) from other background processes (containing top quark, W/Z boson etc.). Although, the individual object reconstruction efficiency is sufficient, relying on reconstructed individual heavy objects inside one events may suffer from the combinatorial problem which can heavily reduce the sensitivity, especially for $HH \rightarrow b\bar{b}b\bar{b}$ case where the four b -jets in the final states are indistinguishable and there are 6 combinations of these b -jets to form one Higgs pair candidate. Hence, we explore the possibility of training on the information of 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, we will first briefly introduce the main features of *ParT*. Then the training details in our setup will be presented. The performance on the event classification is also discussed. To avoid confusion with the original *ParT* model, we will refer to the *ParT* working on the entire event in our setup as Event Transformer (*EventT*). However, we also want to emphasize that the analysis presented in this work can be performed based on any kind of algorithms other than Transformer. The focus here is the possibility of an end-to-end event level analysis.

A. Basics about Particle Transformer

The *ParT* [137] is basically 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 *Transformer* is the self-attention mechanism, which establishes relationships among all elements in the input sequence, rather than being limited to local information like CNN. By

computing the attention weight matrix, *Transformer* can capture global information. Due to the advantages of this mechanism, *Transformer* has demonstrated powerful performance across multiple domains.

The structure of *ParT* is shown in Fig. 3. The model receives two sets of inputs: *particles* which include the features (kinematics, PID, trajectory displacement etc.) of each single object (tracks, charged leptons etc.) and *interactions* which indicate the relationships between two objects. In *ParT*, there are mainly 4 such relationships:

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

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

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

$$m_{ij}^2 = (p_i + p_j)^2. \quad (9)$$

where y, ϕ are the rapidity and azimuthal angle of the individual object respectively. p_T is the corresponding magnitude of the transverse momentum. Then the inputs are passed through the intermediate layers consisting of *Particle Attention Blocks* and *Class Attention Blocks*. Compared to the traditional *Transformer*, *ParT* introduces, in the *Particle Attention Block*, new attention mechanism referred as particle multihead attention (P-MHA) where the *interactions* are included in the attention calculation:

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

$$H_i = \mathcal{A}_i V_i, \quad (11)$$

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

$$Q_i = X W_i^Q + b_i^Q, \quad (13)$$

$$K_i = X W_i^K + b_i^K, \quad (14)$$

$$V_i = X W_i^V + b_i^V. \quad (15)$$

where X includes the *particles* features, 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 the relationships between particles and enhance the expressiveness of the attention mechanism. $\mathcal{A}_i$ is the attention weight matrix of which 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* but differs in two main aspects: it uses the standard Multi-Head Attention (MHA) mechanism instead of the particle-specific variant, and the MHA takes a global class token as input along with the particle embeddings. The query, key, and value matrices are then computed as:

$$Q = W_q X_{\text{class}} + b_q, \quad (16)$$

$$K = W_k Z + b_k, \quad (17)$$

$$V = W_v Z + b_v, \quad (18)$$

where $Z = [X_{\text{class}}, X_{\text{PAB}}]$ is the concatenation of the class token and the particle embedding output from the last

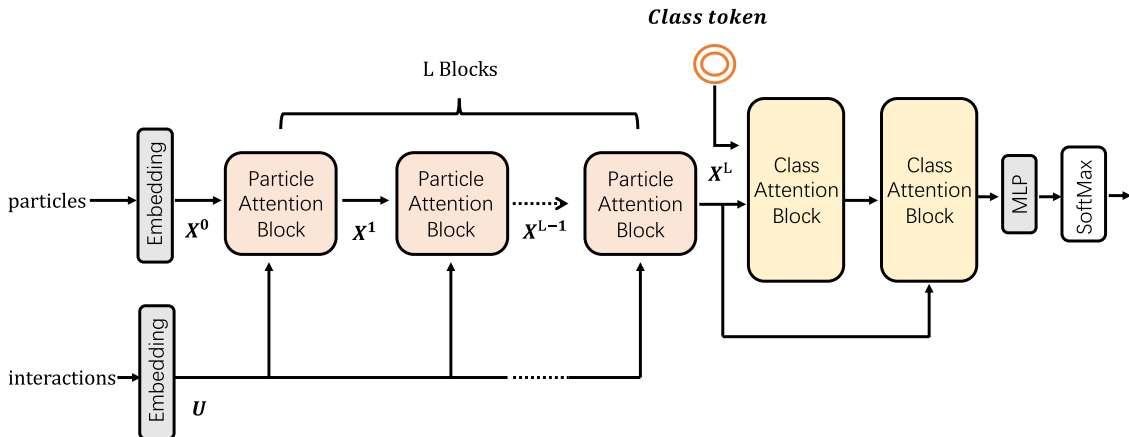


Fig. 3. (color online) The *ParT* Model Architecture.

Particle Attention Block. The output of the *Class Attention Block* are then passed to a Softmax layer which produces the probabilities over the classes.

B. Training Details

In our setup, the *EventT* model receives the information from the whole 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 , as discussed in last section, is also provided. Hence, the input for *EventT* now includes *particle* features of all objects within one event including tracks, energy deposits, charged leptons, jets and event itself, which extend the dimension of the input substantially. Note that, the orientation of a single event is not important. Further, the *interaction* features are also extended to include the affiliation relations between different objects in additional original kinematic relations. To avoid the time-issue when loading all inputs, the *interaction* features are now calculated in advance. The *EventT* model is trained from scratch on a dataset with 9 classes listed in Table 1 containing 3 million samples per class. After each training epoch, the model is validated on a separate validation set containing 3 million samples per class, while 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 multi-head attention mechanism is 8.

In general, all processes during the training shall be treated equally. However, for our purpose, we would like to single out the Higgs pair process as much as possible from the other processes among which $2b2j$ process has the largest cross section. In order to avoid large contamination from $2b2j$ process, unlike *ParT* which is used for a general purpose jet-tagging task, we adopt a weighted cross-entropy loss function, where the process $2b2j$ is assigned with a weight of 70, while 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} by two orders of magnitude which dramatically improves the analysis about the Higgs pair process. The measurement uncertainty of the Higgs pair production cross section can be reduced from about 400% to around 200% by such improvement.

Further, as our target is the measurement of κ_λ , the model is required to work properly on all different value of κ_λ in order to obtain the optimal result. As a comparison, we show in Fig. 4 the results of *HH* efficiency from training on data set with only $\kappa_\lambda = 1$ (blue line) and train-

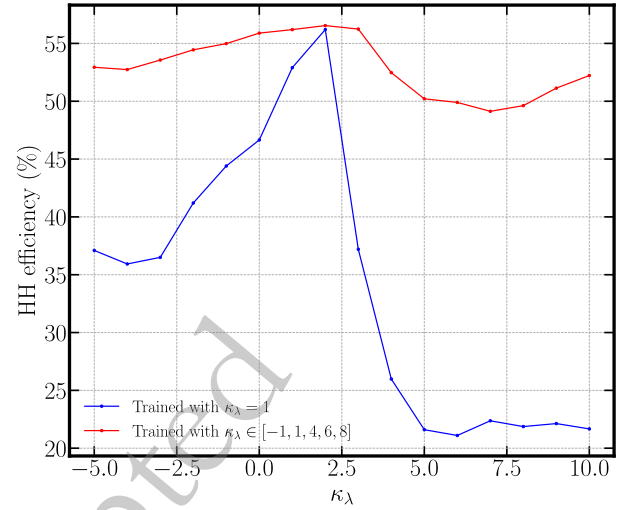


Fig. 4. (color online) *HH* efficiency under different κ_λ values for training containing only $\kappa_\lambda = 1$ *HH* data (blue line) and training containing $\kappa_\lambda = -1, 1, 4, 6, 8$ *HH* data (red line).

ing on data set containing $\kappa_\lambda = \{-1, 1, 4, 6, 8\}$ (red line). It is clear that for the training using data with only $\kappa_\lambda = 1$, the *HH* efficiency will dramatically decrease for other value of κ_λ . The variance of the efficiency can be as large as 40% (from 58% at $\kappa_\lambda = 2$ to 18% at $\kappa_\lambda = 3$). On the other hand, when we include the data with other value of κ_λ , the *HH* efficiency is roughly stable at around 50%. Such performance is hence suitable for the further analysis on the measurement precision about κ_λ .

C. Performance of Event Classification

In this section, we further demonstrate the classification performance of the model. *EventT* is now essentially a multi-class classification model with 9 event categories. For multi-class 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 denotes 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:

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

Specifically:

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

- C_{ij} (for $i \neq j$) represents the probability that an event of true class i is incorrectly predicted as class j .

The confusion matrix of *EventT* on the test 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 the corresponding probabilities.

There are several comments about the confusion matrix in order. First, it is clear from the $2b2j$ column, every class has a relatively high efficiency tagging as $2b2j$ since we enhance the weight of $2b2j$ events during the training as discussed above. Second, as can be seen from the diagonal elements, when not considering the $2b2j$ class, the rest classes all have a good self-tagging efficiency. Further, processes similar to each other will form a diagonal block in the tagging efficiency, e.g. ($ttbb, ttH$) and (ZZ, ZH). For our purpose of measuring Higgs pair production process, the first column of the confusion matrix indicates that all other processes have been heavily suppressed while keep a sufficient efficiency of HH process.

The ROC (Receiver Operating Characteristic) curve and AUC (Area Under Curve) are commonly used metrics for binary classification tasks. They can also be extended to multi-class classification problems for model evaluation. In a binary classification setting with only positive and negative classes, the ROC curve is generated from the True Positive Rate (TPR) and the False Positive Rate (FPR) at various classification thresholds. The AUC is the area under the ROC curve and is used to quantify the overall classification performance.

For multi-class 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 treated 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 the corresponding AUC values for each class. For most of the classes, the AUC is around 0.9 indicating a relatively good performance of the *EventT*. Further, the AUC for distinguishing HH process from other processes is about 0.9, this provides a solid foundation for the analysis of the Higgs self-coupling via Higgs pair production in the next section. On the other hand, the information contained in the confusion matrix as well as those OvR ROC curves can be used not only on the HH analysis, but also on other relevant analysis. In this work, we use the 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 the selection efficiency.

IV. HIGGS SELF-COUPLING MEASUREMENT

A. *EventT*-Based Analysis

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

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

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

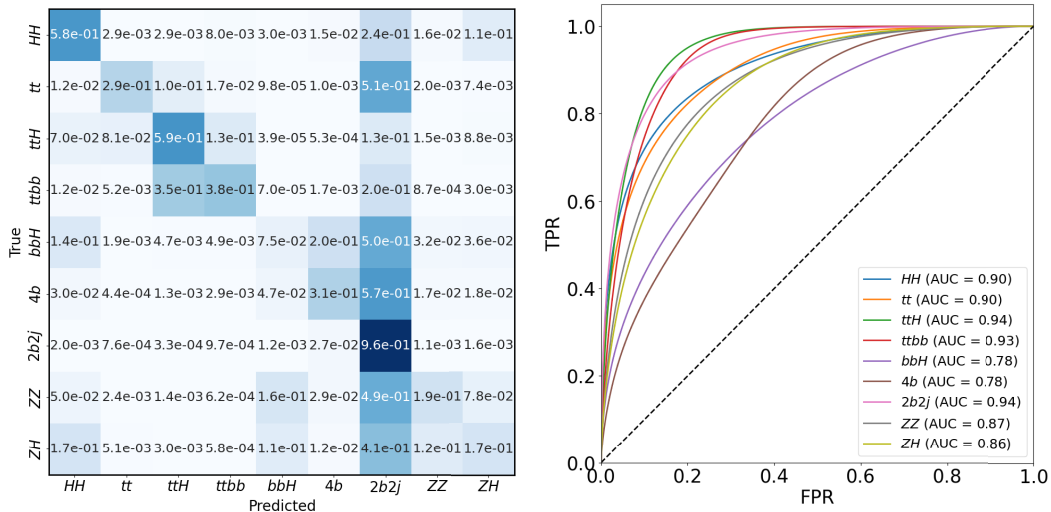


Fig. 5. (color online) Left: Confusion matrix of the *EventT* on 9-class classification task. Right: OvR ROC curves for each class of *EventT* and the corresponding AUC. All the results are obtained with the weighted cross-entropy loss.

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

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

However, as we discussed in above section, the performance of *EventT* also depends on the classification threshold. The effect of the threshold on the Higgs pair production measurement is shown in Fig. 6. We can see that the signal efficiency decrease smoothly with the increase of the threshold for both $\kappa_\lambda = 1$ (solid green line) and $\kappa_\lambda = 5$ (dashed green line). While 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 will decrease with higher threshold as shown by the blue curves in Fig. 6. However, we need to emphasize that when the threshold is close to 1, the events passing the threshold will decrease dramatically. The corresponding analysis hence suffers from large uncertainties. In our analysis, we will use three benchmark thresholds

$p_{th} = 0.5, 0.7, 0.9$ (indicated by the vertical purple dashed lines in Fig. 6) together with $p_{th} = 0.0$ where we rely entirely on the raw output of *EventT* for the classification.

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

B. Cut-Based Analysis

As a comparison, we followed the cut-based analysis presented in [84] to single out $HH \rightarrow b\bar{b}b\bar{b}$ events. In this analysis, the events first need to pass several general selections including: requiring at least 4 b -tagged (with tagging efficiency about 77%) jets with $p_T > 40\text{GeV}$, and $|\eta_b| < 2.5$. The four b -jets with the highest p_T will be 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. To suppress the $t\bar{t}$ background, a top veto cut is needed. The top veto discriminant X_{W_t} is defined as:

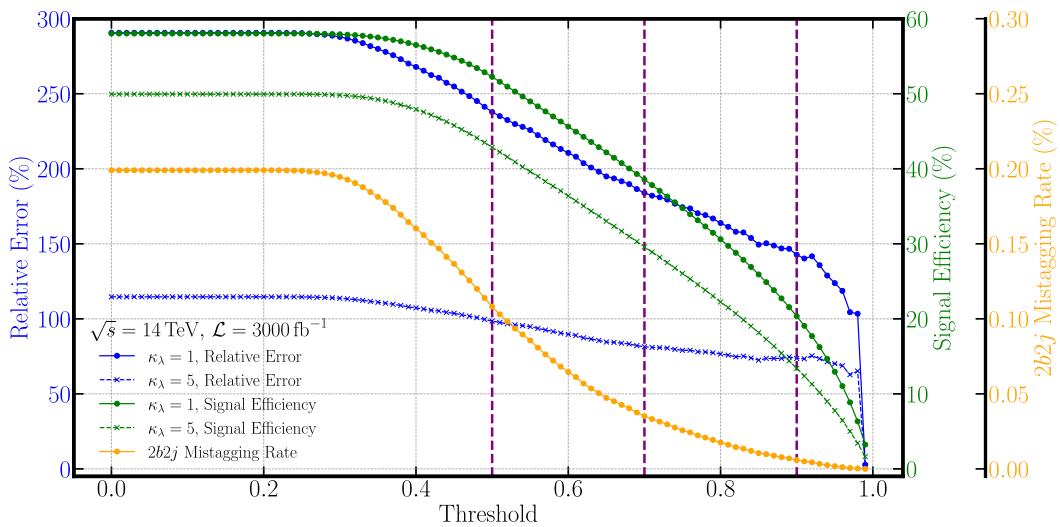


Fig. 6. (color online) Signal efficiency (green lines) and relative error of the Higgs pair production cross-section measurement (blue lines) for $\kappa_\lambda = 1$ (solid lines) and $\kappa_\lambda = 5$ (dashed lines) at different thresholds. The misclassification rate of the dominant background $2b2j$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

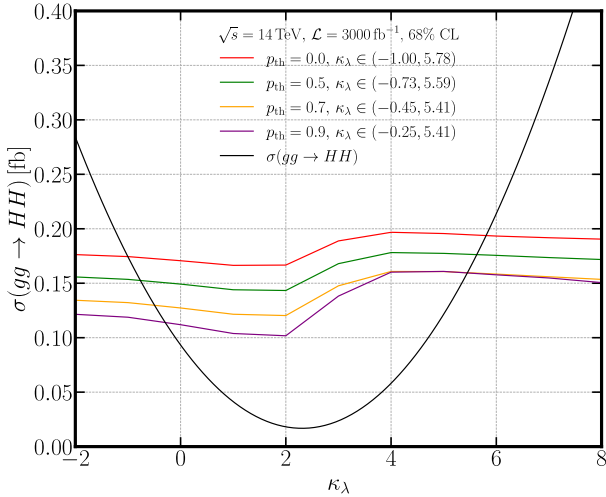


Fig. 7. (color online) Upper limit on the Higgs pair production cross section as a function of κ_λ for different classification thresholds p_{th} (four colored lines) with $\sqrt{s} = 14\text{TeV}$ and $\mathcal{L} = 3000\text{fb}^{-1}$ at the LHC. The Higgs pair production cross section as a function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

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

where $m_W = 80.4\text{GeV}$ and $m_t = 172.5\text{GeV}$ are the nominal W boson and top quark mass. m_{jj} is the invariant mass of two jets that are assumed to come from the W boson decay. Together with one of the leading b -jet, m_{jib} represent the invariant mass of the reconstructed top. The X_{Wt} is then obtained by minimizing over all the 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 in the analysis.

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

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

where m_{H_1} and m_{H_2} are the masses of the leading and sub-leading reconstructed Higgs boson candidates, respectively. The reference mass 124GeV and 117GeV are obtained from the m_{H_1} and m_{H_2} distribution from the simulation [84]. Note that, when calculating X_{HH} , 10% uncertainties are also used for the reconstructed invariant masses. Events with $X_{HH} < 1.6$ are considered as $HH \rightarrow b\bar{b}b\bar{b}$ signal. All above 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 $2b2j$ as well as $t\bar{t}$, $t\bar{t}b\bar{b}$, $t\bar{t}H$, $b\bar{b}H$, ZZ , 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 *EventT*, we compare the results obtained from the *EventT*-based analysis with those from a cut-based analysis using kinematic variables discussed above, as well as with two ML-based studies, one using DNN [115], the other using SPA-NET [116]. The comparison are shown in Fig. 8.

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

SPA-NET is also 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. The SPA-NET has also been used in the $HH \rightarrow b\bar{b}b\bar{b}$ analysis in [116] where the architecture is used both in pairing of 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 *EventT* analysis. The results are shown in the middle group of Fig. 8 which are given at 95% CL with 300fb^{-1} luminosity at the LHC. It is clear that, with only $4b$ background, the results from *EventT* are better than that from SPA-NET from both side even without imposing any threshold. Note that both SPA-NET and *EventT* are based on Transformer framework, the results hence demonstrate the effectiveness of the event-level analysis.

The comparison with the cut-based analysis discussed above is also presented in the right group of Fig. 8. Note that the cut-based analysis and *EventT*-based analysis are performed on the same set of testing data. The constraints on κ_λ are given at 68% CL with 3000fb^{-1} at 14 TeV. It is clear that, the *EventT* result is already better than that from the cut-based analysis without imposing

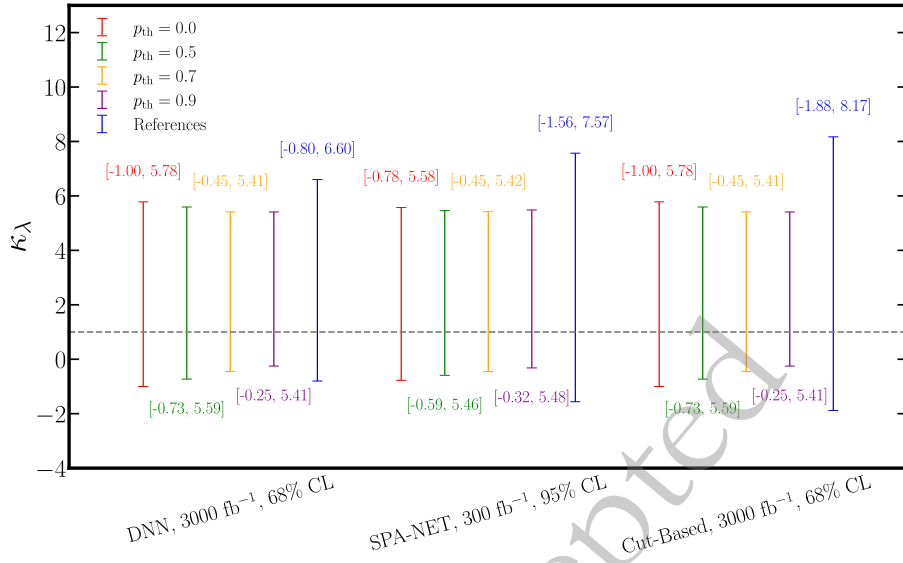


Fig. 8. (color online) The comparison of the constraints on κ_λ with three benchmark studies.

any threshold to enhance the sensitivity. With stronger threshold in the classification, the result from *EventT* becomes much stronger than that from the general cut-based analysis which clearly demonstrates the advantage of the ML models in the analysis.

D. Model Interpretability

Before we conclude, in this section, we also would like to discuss the interpretability of the *EventT* model to understand the internal connections of the model. As we have mentioned previously, the outstanding performance of Transformer based models is mainly attributed to the attention mechanism. Hence, we will mainly focus on the visualization of the attention mechanism of the *EventT* 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 whole event instead of fully reconstructed objects. We visualized the 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 *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. In order to make the plot clear, we ignore the attention score that is lower than 0.01. It is clear that the attention score is more concentrated for both *HH* and *2b2j* processes with the training which strongly indicates that the model does learn the important relationship between particles. Further, the model will pay attention to different part in the particle space for different processes, which will be the key for the model to distinguish different processes.

To incorporate the spatial position of particles, particle types, and the jets to which they belong during

visualization, we further 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 Fig. 10 and Fig. 11 for *HH* and *2b2j* processes respectively. Different marker shapes are used to distinguish particle types: $\times$ for mesons, Δ for charged hadrons, ∇ for neutral hadrons, $\bullet$ for photons, and $+$ for electrons. Particles within the same jet, which is reconstructed using anti- k_t algorithm, are shown in the same color, while particles outside any jet are shown in gray. Note that this color information is just used 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, while dashed lines indicate connections involving out-of-jet particles. Note that to make the plot clear without full of low-weight connections, we only include the connections with weight higher than 0.5.

In the upper panels of both Fig. 10 and Fig. 11, the connections in the whole η - ϕ plane are shown after (upper left) and before (upper right) the training for *HH* and *2b2j* processes respectively. Similar to that in Fig. 9, the connections among particles become stronger after the training for both *HH* and *2b2j* processes. To be more clear about the connections that are established during the training, for each process, we also provide the corresponding zoom-in view of the attention graph in the lower panels of Fig. 10 and Fig. 11. It is clear that the training also strengthen the connections within jets from the information about the whole event. By comparing these visualized attention graph, we conclude that the trained model is capable of focusing on the important particle pairs in an event and learning relationships within and

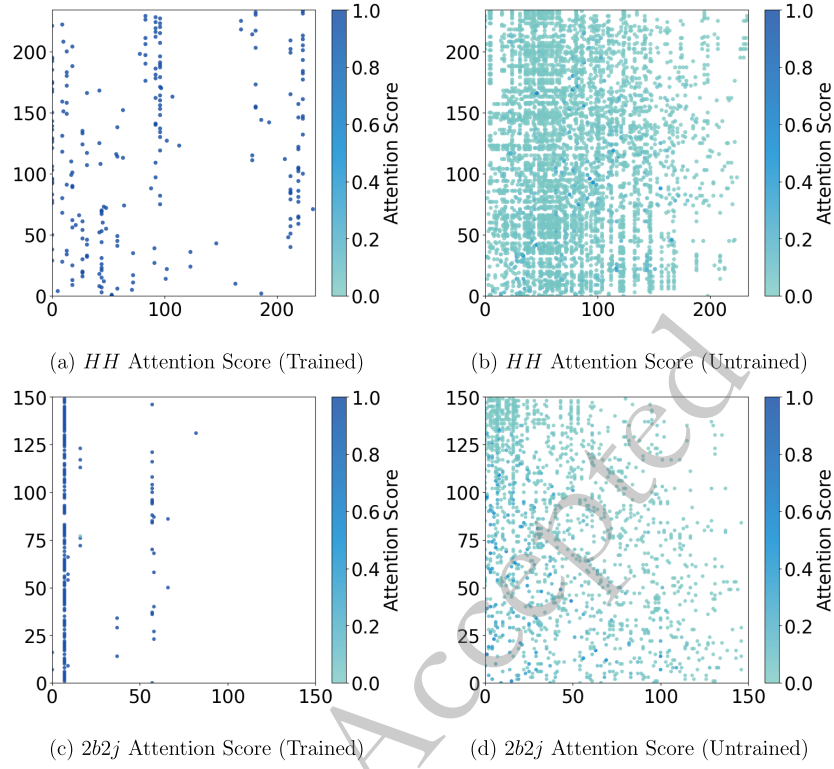


Fig. 9. (color online) The attention score (elements in attention matrix) for HH (upper panels) and $2b2j$ (lower panels) processes after (left panels) and before (right panels) training. Note that we have removed the attention score below 0.01 to make the plot clearer.

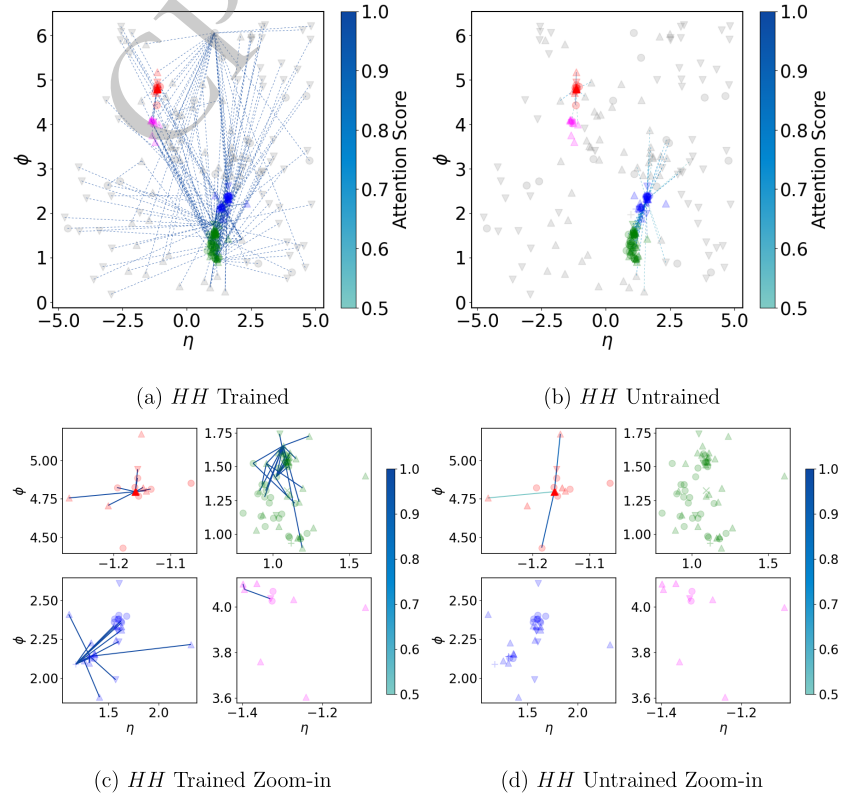


Fig. 10. (color online) Attention graphs for HH processes before and after model training. The lower panels show the connections within a single jet.

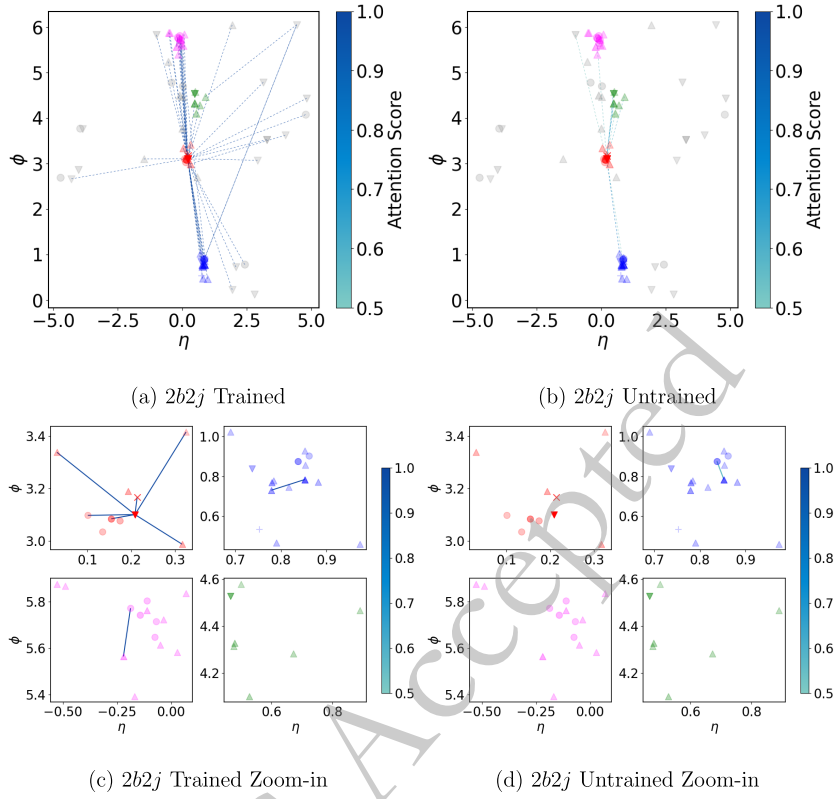


Fig. 11. (color online) Attention graphs for $2b2j$ processes before and after model training. The lower panels show the connections within a single jet.

between jets by training on the event-level instead of reconstructed objects. This capability underpins the model's outstanding performance and the possibility to the end-to-end event classification by event-level training. However in depth studies of what specific feature and relations have been learnt during the training is also very important to fully understand the underlying mechanism of the algorithm. This is a challenging task which we will leave for future works.

V. CONCLUSIONS

In this work, we employ a deep neural network architecture based on *ParT* to enhance the sensitivity of the Higgs self-coupling measurement through $HH \rightarrow b\bar{b}b\bar{b}$ 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, enabling robust event classification when trained on full event-level information. The performance of such classifier is evaluated across 9 event categories achieving $AUC \approx 0.9$ for most categories, significantly distinguishing one process from the rest. In current study, we focus on $HH \rightarrow b\bar{b}b\bar{b}$ analysis. However, it can actually be applied to other analysis with very minor modifications.

The studies in this work demonstrate 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 approach streamlines analysis by directly utilizing the full event information and demonstrates the potential of such end-to-end analysis in collider phenomenology studies.

By applying the model to $HH \rightarrow b\bar{b}b\bar{b}$ studies at the HL-LHC, the model constrains the Higgs self-coupling κ_λ to $(-0.25, 5.41)$ at 68% CL, which achieves around 44% improvement in precision compared with the traditional cut-based approach. The comparison against other ML methods further highlights the advantages of the Transformer-based architecture and the event-level analysis, particularly in capturing the correlations within high-dimensional event data. Further, the attention mechanism naturally provide the interpretability of the model which can also help us understanding what are the most important features and correlations when trying 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 Huiyang 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 and J.S. Lee, *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. Degrandi, B. Di Micco, P.P. Giardino and E. Rossi, *Phys. Lett. B* **817**, 136307 (2021), arXiv: 2102.07651
- [14] H. Bahl, J. Braathen, M. Gabelmann and G.R. Weiglein, *PoS EPS-HEP2023*, 407 (2024), arXiv: 2311.01134
- [15] F. Arco, S. Heinemeyer, M. Mühlleitner and K. Radchenko, *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, M.M. Mühlleitner, J. Quevillon and M. Spira, *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, J. Song and Y.W. Yoon, *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 and J. Zurita, *JHEP* **04**, 167 (2015), arXiv: 1410.3471
- [29] A. Azatov, R. Contino, G. Panico and M. Son, *Phys. Rev. D* **92**, 035001 (2015), arXiv: 1502.00539
- [30] Q.-H. Cao, B. Yan, D.-M. Zhang and H. Zhang, *Phys. Lett. B* **752**, 285 (2016), arXiv: 1508.06512
- [31] G. Li, L.-X. Xu, B. Yan and C.P. Yuan, *Phys. Lett. B* **800**, 135070 (2020), arXiv: 1904.12006
- [32] T. Liu, K.-F. Lyu, J. Ren and H.X. Zhu, *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, Y.-M. Zhong and Z. Zhao, *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, B. Mele, F. Piccinini and X. Zhao, *JHEP* **09**, 098 (2020), arXiv: 2003.13628
- [37] J.M. Torndal, J. List, D. Ntounis and C. Vernieri, *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 [2307.16515].
- [39] J. Davies, K. Schönwald, M. Steinhauser and H. Zhang, *JHEP* **10**, 033 (2023), arXiv: 2308.01355
- [40] J. Davies, *PoS LL2024*, 015 (2024), arXiv: 2407.08264
- [41] H. Zhang, K. Schönwald, M. Steinhauser and J. Davies, *PoS LL2024*, 014 (2024), arXiv: 2407.05787
- [42] G. Heinrich, S. Jones, M. Kerner, T. Stone and A. Vestner, *JHEP* **11**, 040 (2024), arXiv: 2407.04653
- [43] A. Carvalho, M. Dall'Osso, T. Dorigo, F. Goertz, C.A. Gottardo and M. Tosi, *JHEP* **04**, 126 (2016), arXiv: 1507.02245
- [44] J. Davies, K. Schönwald, M. Steinhauser and H. Zhang, *JHEP* **04**, 193 (2025), arXiv: 2501.17920
- [45] A. Adhikary, S. Banerjee, R.K. Barman, B. Bhattacharjee and S. Niyogi, *JHEP* **07**, 116 (2018), arXiv: 1712.05346
- [46] J.H. Kim, K. Kong, K.T. Matchev and M. Park, *Phys. Rev. Lett.* **122**, 091801 (2019), arXiv: 1807.11498
- [47] W. Bizoń, U. Haisch, L. Rottoli, Z. Gillis, B. Moser and P. Windischhofer, *JHEP* **02**, 170 (2024), arXiv: 2402.03463
- [48] K. Nakamura, K. Nishiwaki, K.-y. Oda, S.C. Park and Y. Yamamoto, *Eur. Phys. J. C* **77**, 273 (2017), arXiv: 1701.06137
- [49] A. Adhikary, R.K. Barman and B. Bhattacharjee, *JHEP* **12**, 179 (2020), arXiv: 2006.11879
- [50] L.-B. Chen, H.T. Li, H.-S. Shao and J. Wang, *Phys. Lett. B* **803**, 135292 (2020), arXiv: 1909.06808
- [51] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, J. Ronca, M. Spira, *et al.*, *JHEP* **04**, 181 (2020), arXiv: 2003.03227
- [52] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, M. Spira and J. Streicher, *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. Czurzyło, *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 and M. Spannowsky, *Phys. Rev. Lett.* **112**, 101802 (2014), arXiv: 1310.1084
- [58] J. Chen, T. Li, C.-T. Lu, Y. Wu and C.-Y. Yao, *Phys. Rev. D* **105**, 053009 (2022), arXiv: 2112.12507
- [59] B. Henning, D. Lombardo, M. Riembau and F. Riva, *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. Degrossi, P.P. Giardino, F. Maltoni and D. Pagani, *JHEP* **12**, 080 (2016), arXiv: 1607.04251
- [63] J. Gao, X.-M. Shen, G. Wang, L.L. Yang and B. Zhou, *Phys. Rev. D* **107**, 115017 (2023), arXiv: 2302.04160
- [64] G. Degrossi, 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. Mühlleitner and M. Spira, *Nucl. Phys. B* **925**, 1 (2017), arXiv: 1705.05314
- [71] A. Adhikary, S. Banerjee, R. Kumar Barman and B. Bhattacharjee, *JHEP* **09**, 068 (2019), arXiv: 1812.05640
- [72] E. Arganda, M. Epele, N.I. Mileo and R.A. Morales, *Eur. Phys. J. Plus* **139**, 615 (2024), arXiv: 2401.03178
- [73] T. Huang, J.M. No, L. Pernié, M. Ramsey-Musolf, A. Safonov, M. Spannowsky, *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, A.D. Peterson and G. Shaughnessy, *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, C. Vernieri and J. Zurita, *JHEP* **02**, 002 (2020), arXiv: 1910.08574
- [78] S. Heinemeyer, M. Mühlleitner, K. Radchenko and G. Weiglein, *Eur. Phys. J. C* **85**, 437 (2025), arXiv: 2403.14776
- [79] A. Arhrib, S. Moretti, S. Semlali, C.H. Shepherd-Themistocleous, Y. Wang and Q.S. Yan, *Probing a 2HDM Type-I light Higgs state via $H_{SM} \rightarrow hh \rightarrow b\bar{b}\gamma\gamma$ at the LHC*, [2412.06052].
- [80] J. Baglio, O. Eberhardt, U. Nierste and M. Wiebusch, *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. Diaz-Cruz, N. Mileo, R.A. Morales and A. Szykman, *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 [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, C.-T. Lu and J. Park, *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 and M. Spannowsky, *Phys. Lett. B* **728**, 308 (2014), arXiv: 1309.6318
- [96] D. Gonçalves, T. Han, F. Kling, T. Plehn and M. Takeuchi, *Phys. Rev. D* **97**, 113004 (2018), arXiv: 1802.04319
- [97] A. Taliencio, P. Mastropasqua, C. Caputo, P. Vischia, N. De Filippis and P. Bhat, *Higgs Self Couplings Measurements at Future proton-proton Colliders: a Snowmass White Paper*, in *Snowmass 2021*, 3, 2022 [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. Taliencio, E. Gallo, K. Tackmann and P. Mastropasqua, *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, J. Tian, C. Vernieri and J.M. Torndal, *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 and X. Wang, *Phys. Rev. D* **103**, 013002 (2021), arXiv: 2008.12204
- [108] L. Buonincontri, M. Casarsa, L. Giambastiani, D. Lucchesi, L. Sestini and D. Zuliani, *Nuovo Cim. C* **47**, 288 (2024)
- [109] F. Goertz, A. Papaefstathiou, L.L. Yang and J. Zurita, *JHEP* **06**, 016 (2013), arXiv: 1301.3492
- [110] J.K. Behr, D. Bortoletto, J.A. Frost, N.P. Hartland, C. Issever and J. Rojo, *Eur. Phys. J. C* **76**, 386 (2016), arXiv: 1512.08928

- [111] D. Wardrope, E. Jansen, N. Konstantinidis, B. Cooper, R. Falla and N. Norjoharuddeen, *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, H.-K. Guo and K. Sinha, *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 and I. Low, *JHEP* **09**, 139 (2024), arXiv: 2401.14198
- [117] J. Alison, S. An, P. Bryant, B. Burkle, S. Gleyzer, M. Narain *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 [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 and B. Poczos, *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, J.M. Yang and J. Zhao, *Commun. Theor. Phys.* **71**, 955 (2019), arXiv: 1905.06047
- [122] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban and D. Whiteson, *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, J. Krupa and P. Harris, *Product Manifold Machine Learning for Physics*, [2412.07033].
- [125] J. Lin, M. Freytsis, I. Moulton and B. Nachman, *JHEP* **10**, 101 (2018), arXiv: 1807.10768
- [126] J. Cogan, M. Kagan, E. Strauss and A. Schwartzman, *JHEP* **02**, 118 (2015), arXiv: 1407.5675
- [127] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *JHEP* **07**, 069 (2016), arXiv: 1511.05190
- [128] J.S.H. Lee, I. Park, I.J. Watson and S. Yang, *J. Korean Phys. Soc.* **74**, 219 (2019), arXiv: 2012.02531
- [129] C.F. Madrazo, I.H. Cacha, L.L. Iglesias and J.M. de Lucas, *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 and J.M. Yang, *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, U. Hani, N. Fatima, A.R. Khan *et al.*, *Particle Multi-Axis Transformer for Jet Tagging*, 2406.06638.
- [137] H. Qu, C. Li and S. Qian, *Particle Transformer for Jet Tagging*, 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, H. Qu and J. Zhu, *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 [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, A. Paul and Z. Qian, *JHEP* **11**, 045 (2022), arXiv: 2207.04157
- [145] L. Huang, S.-b. Kang, J.H. Kim, K. Kong and J.S. Pi, *JHEP* **08**, 114 (2022), arXiv: 2203.11951
- [146] B. Tannenwald, C. Neu, A. Li, G. Buehlmann, A. Cuddeback, L. Hatfield *et al.*, *Benchmarking Machine Learning Techniques with Di-Higgs Production at the LHC*, 2009.06754.
- [147] M. Abdughani, D. Wang, L. Wu, J.M. Yang and J. Zhao, *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, X. Zhang and D. Zhao, *Chin. Phys. C* **49**, 023107 (2025), arXiv: 2407.14716
- [151] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, P. Torrielli, *et al.*, *Phys. Lett. B* **732**, 142 (2014), arXiv: 1401.7340
- [152] G. Heinrich, S.P. Jones, M. Kerner, G. Luisoni and L. Scyboz, *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, V. Hirschi, F. Maltoni, O. Mattelaer, *et al.*, *JHEP* **07**, 079 (2014), arXiv: 1405.0301
- [157] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, *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, J. Uszkoreit, L. Jones, A.N. Gomez *et al.*, *Attention is all you need*, 1706.03762.
- [162] M.J. Fenton, A. Shmakov, T.-W. Ho, S.-C. Hsu, D. Whiteson and P. Baldi, *Phys. Rev. D* **105**, 112008 (2022), arXiv: 2010.09206
- [163] A. Shmakov, M.J. Fenton, T.-W. Ho, S.-C. Hsu, D. Whiteson and P. Baldi, *SciPost Phys.* **12**, 178 (2022), arXiv: 2106.03898
- [164] M.J. Fenton, A. Shmakov, H. Okawa, Y. Li, K.-Y. Hsiao, S.-C. Hsu, *et al.*, *Commun. Phys.* **7**, 139 (2024), arXiv: 2309.01886
- [165] A. Wang, A. Gandrakota, J. Ngadiuba, V. Sahu, P. Bhatnagar, E.E. Khoda *et al.*, *Interpreting Transformers for Jet Tagging*, 12, 2024 [2412.03673].