Published for SISSA by Springer

Received: March 19, 2024
Revised: July 15, 2024
Accepted: August 14, 2024
Published: September 20, 2024

Deep learning to improve the sensitivity of Di-Higgs searches in the $4b$ channel

Cheng-Wei Chiang,[a,b] **Feng-Yang Hsieh**,[a] **Shih-Chieh Hsu**[c] **and Ian Low**[d,e]

[a] *Department of Physics, National Taiwan University, Taipei 10617, Taiwan*
[b] *Physics Division, National Center for Theoretical Sciences, Taipei 10617, Taiwan*
[c] *Department of Physics, University of Washington, Seattle, WA 98195, U.S.A.*
[d] *High Energy Physics Division, Argonne National Laboratory, Argonne, IL 60439, U.S.A.*
[e] *Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, U.S.A.*

E-mail: chengwei@phys.ntu.edu.tw, f10222035@ntu.edu.tw, schsu@uw.edu, ilow@northwestern.edu

Abstract: The study of di-Higgs events, both resonant and non-resonant, plays a crucial role in understanding the fundamental interactions of the Higgs boson. In this work we consider di-Higgs events decaying into four b-quarks and propose to improve the experimental sensitivity by utilizing a novel machine learning algorithm known as Symmetry Preserving Attention Network (Spa-Net) — a neural network structure whose architecture is designed to incorporate the inherent symmetries in particle reconstruction tasks. We demonstrate that the Spa-Net can enhance the experimental reach over baseline methods such as the cut-based and the Dense Neural Network-based analyses. At the Large Hadron Collider, with a 14-TeV center-of-mass energy and an integrated luminosity of 300 fb^{-1}, the Spa-Net allows us to establish 95% C.L. upper limits in resonant production cross-sections that are 10% to 45% stronger than baseline methods. For non-resonant di-Higgs production, Spa-Net enables us to constrain the self-coupling that is 9% more stringent than the baseline method.

Keywords: Higgs Production, Higgs Properties

ArXiv ePrint: 2401.14198

https://doi.org/10.1007/JHEP09(2024)139

Contents

1 Introduction

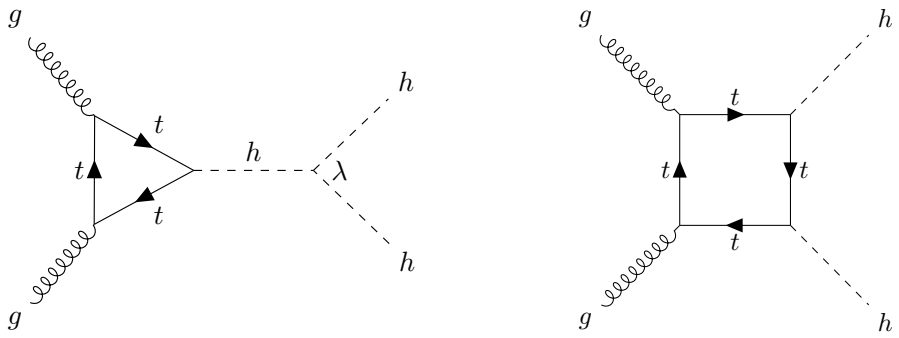
Since the discovery of the 125-GeV Higgs boson, h, an immediate and pressing task has been to determine the Higgs potential and self-interactions, in addition to measuring its couplings to other SM particles, thereby verifying whether electroweak symmetry breaking (EWSB) is achieved in the same way as the Standard Model (SM) prescribes [1, 2]. An important parameter in the Higgs potential is the trilinear coupling λ, which enters the potential as:

$$V(h) = \frac{1}{2}m_h^2 h^2 + \lambda v h^3 + \lambda_{4h} h^4 \,. \tag{1.1}$$

Here $m_h = 125\,\text{GeV}$ and the Higgs vacuum expectation value is $v = 246\,\text{GeV}$. In the SM, $\lambda = \lambda^{\text{SM}} = m_h^2/(2v^2)$ and $\lambda_{4h} = \lambda_{4h}^{\text{SM}} = m_h^2/(8v^2)$. Phenomenologically the Higgs trilinear coupling contributes to the pair production of Higgs bosons, which has not been observed experimentally. In the SM, the leading-order Feynman diagrams contributing to the di-Higgs production at a hadron collider, such as the Large Hadron Collider (LHC) at CERN, are shown in figure 1, where λ features prominently.

An experimental verification of the Higgs potential has important implications, as it is well known that the potential in the SM cannot induce a strong first-order electroweak phase transition in the early Universe, which is a crucial ingredient to explain the observed matter-antimatter asymmetry [3–5]. New physics should enter at an energy scale slightly higher than the weak scale to modify the Higgs potential at finite temperatures, to facilitate a

JHEP09(2024)139



(a) Triangle diagram (b) Box diagram

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

sufficiently strong first-order phase transition. In this case, the λ coupling would be modified from its SM value, which could have a large impact on the di-Higgs production rate at the LHC [6]. In addition, kinematic distributions of the Higgs pair could offer a unique window into new particles and new interactions above the weak scale [7, 8].

With a mass of 125 GeV, the Higgs boson predominantly decays into a $b\bar{b}$ pair. Therefore $hh \to 4b$ channel offers the largest rate among all possible decays of the Higgs pair which, nevertheless, suffers from the much larger background from multijets and multi-b's production of QCD. Furthermore, because of our inability to distinguish a b-jet from a $\bar{b}$-jet, it is a very challenging task experimentally to form the correct pairing among the 4 b's to reconstruct the Higgs mass and the associated kinematic distributions. In this work, we would like to propose the use of machine learning algorithms from deep neural networks to help improve the sensitivity of experimental searches in the $4b$ channel. More specifically, we study the possibility of using a new neural network architecture called the Symmetry Preserving Attention Network (SPA-NET) [9–11] to simultaneously perform signal/background separation and identify the correct pairings among the 4 b-jets in the final states. To utilize the advantages of SPA-NET, our analysis is applied in the resolved regime; a different model structure would be required for the boosted regime [12]. We will demonstrate that SPA-NET offers improved sensitivity over existing experimental techniques employed in the $4b$ channel [13–17], as well as over an analysis invoking the Dense Neural Network (Dense-NN) machine learning algorithm [18].

We will consider two types of analyses in di-Higgs productions. One involves the on-shell, resonant production of a hypothetical new scalar particle, which subsequently decays into the Higgs pair. As an explicit example, we consider the two-Higgs doublet models (2HDMs) in the alignment limit [19–21], where the properties of the 125 GeV Higgs boson are SM-like. In this case, the Higgs pair is produced through the decay product of the heavy scalar [22, 23]. We do not modify the SM contribution to the di-Higgs production in this scenario. The other analysis, on the other hand, involves "non-resonant" production in the sense that we alter the SM trilinear coupling λ, of which the production cross-section is a function, and does not invoke direct production of new particles.

This paper is organized as follows. In section 2, we describe the procedures employed for generating both signal and background samples utilized in the training of neural networks.

JHEP09(2024)139

Section 3 discusses three different jet pairing methods. The first two are cut-based methods inspired by an ATLAS analysis, while the last is done with SPA-NET. In section 4, we provide the training procedures of Dense-NN and SPA-NET classifiers and how to employ these neural network classifiers on the event selection task. In section 5, we perform the resonant analysis and discuss how SPA-NET can improve the sensitivity of the search. In section 6, we build classifiers to discriminate non-resonant signal and background events and then demonstrate how SPA-NET performs better in constraining the Higgs trilinear coupling in comparison with more traditional methods. Finally, we summarize in section 7.

2 Event generation

We use `MadGraph5_aMC@NLO 3.3.1` [24] to generate both signal and background events at the centre-of-mass (CM) energy $\sqrt{s} = 13\,\text{TeV}$ with the `NNPDF23_nlo_as_0119` PDF set [25]. The LO matrix elements are considered. For parton showering and hadronization, we employ `Pythia 8.306` [26] with `NNPDF2.3 LO` PDF set. The detector simulation is performed using `Delphes 3.5.0` [27]. The jets are reconstructed with `FastJet 3.3.4` [28] using the anti-k_t [29] algorithm with radius $R = 0.4$. Only jets with a transverse momentum of $p_{\text{T}} \geq 20\,\text{GeV}$ are considered.

2.1 Signal event generation

We consider two types of di-Higgs events: resonant production through an on-shell new scalar boson and non-resonant production in the SM with, however, a rescaled Higgs self-coupling $\lambda = \kappa_\lambda \lambda^{\text{SM}}$.

For the resonant signal, Higgs boson pairs hh are produced via the decay of heavy CP-even scalar H in the 2HDM, which itself is produced through the gluon-fusion channel. We consider m_H ranging from 300 GeV to 1200 GeV. In this mass range, the b-jets can be reconstructed into four distinct energetic jets. The 2HDMC [30] calculator with `HiggsBounds-5.10.2` [31–35] and `HiggsSignal-2.6.2` [36–39] extensions is used to compute the parameters at these benchmark points, which are submitted to `MadGraph5_aMC@NLO` through the parameter card.

The non-resonant signal is produced at one-loop at the leading order, via the Feynman diagrams shown in figure 1. The process is simulated using the `MadGraph5_aMC@NLO 3.3.1` with the `loop sm` model. When considering non-resonant productions, we leave κ_λ as a free parameter varying over the domain of $[-10, 15]$.

The decays $H \to hh, h \to b\bar{b}$ are implemented by `MadSpin` [40]. For resonant analysis, the b-tagging efficiency in `Delphes` is modified based on the ATLAS MV2c10 b-tagger at the 70% working point [41, 42]. At this working point, the light-jet (charm-jet) rejection is about 385 (12), which is the reciprocal of the false positive rate. For non-resonant analysis, the b-tagging efficiency is modified based on the ATLAS DL1r 77% working point [43]. At this working point, the light-jet (charm-jet) rejection is about 130 (4.9).

2.2 Background event generation

The main background is QCD multijet production: $pp \to b\bar{b}b\bar{b}$. The resonant and non-resonant analysis background is simulated using different b-tagging settings as described in section 2.1. The sub-leading background is top-quark pair production, which contributes to less than 10% of the dominant background and is not included in the analysis. It is important to note

JHEP09(2024)139

	Training	Validation	Testing
Resonant	950k	50k	100k
Non-resonant	171k	9k	18k

Table 1. Sizes of various samples used for neural network study in resonant and non-resonant analyses. Each category consists of an equal size of signal and background samples.

that `Pythia` considers the initial-state radiation (ISR) and final-state radiation (FSR) at the parton level. Therefore, there might be more than 4 jets after the jet clustering.

In generating both the signal and background events, we implement a basic "four-tag cut", which requires at least four b-tagged $R = 0.4$ anti-k_t jets with $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ and pseudorapidity η within the range $|\eta| < 2.5$. The numbers of events that passed this cut are given in table 1.

3 Jet pairing

To reconstruct two Higgs boson candidates, we use three different jet assignment methods. The first two, ΔR+min-D_{hh} and min-ΔR, are cut-based and inspired by the ATLAS analysis in refs. [13, 15]. The third one makes use of the SPA-NET neural network [9–11], a novel architecture specifically designed for the jet assignment task.

3.1 Cut-based pairing

In the ΔR + min-D_{hh} pairing method, the four b-jets with the highest p_{T} are paired to construct two Higgs boson candidates. There are three possible pairings for the jets. Only the pairing that satisfies the following ΔR requirements is accepted:

$$\left.\begin{aligned} \frac{360\,\mathrm{GeV}}{m_{4\mathrm{j}}} - 0.5 < \Delta R_{\mathrm{jj}}^{\mathrm{lead}} < \frac{653\,\mathrm{GeV}}{m_{4\mathrm{j}}} + 0.475 \\ \frac{235\,\mathrm{GeV}}{m_{4\mathrm{j}}} < \Delta R_{\mathrm{jj}}^{\mathrm{subl}} < \frac{875\,\mathrm{GeV}}{m_{4\mathrm{j}}} + 0.35 \end{aligned}\right\} \text{ if } m_{4\mathrm{j}} < 1250\,\mathrm{GeV}\,, \tag{3.1}$$

$$\left.\begin{aligned} 0 < \Delta R_{\mathrm{jj}}^{\mathrm{lead}} < 1 \\ 0 < \Delta R_{\mathrm{jj}}^{\mathrm{subl}} < 1 \end{aligned}\right\} \text{ if } m_{4\mathrm{j}} > 1250\,\mathrm{GeV}\,,$$

where the $\Delta R_{\mathrm{jj}}^{\mathrm{lead}}$ is the angular distance between the jets in the p_{T}-leading Higgs boson candidate and $\Delta R_{\mathrm{jj}}^{\mathrm{subl}}$ for the sub-leading candidate, and $m_{4\mathrm{j}}$ is the total invariant mass of the four jets. The angular distance is calculated using the formula $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, where $\Delta\eta$ and $\Delta\phi$ represent the pseudorapidity and azimuthal angle differences between the two jets, respectively.

If no pairing satisfies the above ΔR requirements, the event is dropped. If more than one pairing satisfies the ΔR requirements, we choose the one with the minimum D_{hh}, defined as

$$D_{hh} = \frac{\left| m_{h_1} - \frac{120}{110} m_{h_2} \right|}{\sqrt{1 + \left(\frac{120}{110}\right)^2}}\,, \tag{3.2}$$

where m_{h_1}, m_{h_2} are the masses of the leading Higgs candidate and sub-leading Higgs candidate, respectively. The quantity D_{hh} is the distance from (m_{h_1}, m_{h_2}) to the line connecting

JHEP09(2024)139

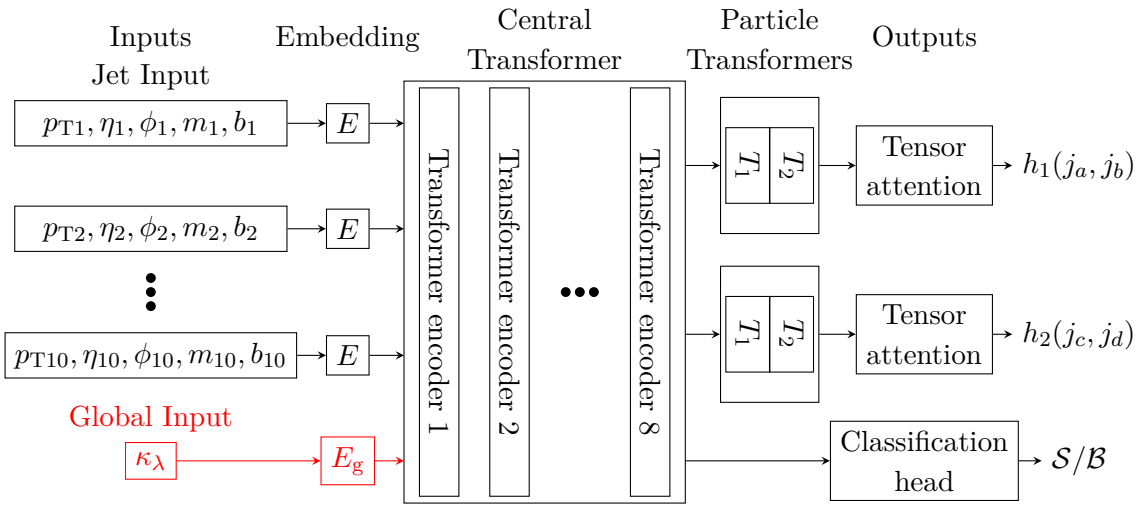


Figure 2. The high-level model structure of SPA-NET. Each E is an embedding layer, T_i is the transformer encoder, and h_i is the jet assignment result which contains two jets j_i for the Higgs decay. The particle transformer is a stack of transformer encoders. The global input is only used in the non-resonant analysis.

(0 GeV, 0 GeV) and (120 GeV, 110 GeV). The values of 120 GeV and 110 GeV account for energy loss.

In the min-ΔR pairing method, the four b-tagged jets with the highest p_T are used to form the two Higgs boson candidates. The min-ΔR method selects the pairing configuration in which the higher-p_T jet pair has the smallest ΔR separation.

3.2 Spa-Net pairing

In this subsection, we provide an overview of the SPA-NET's model structure and explain how this architecture is particularly well-suited for the jet assignment/pairing task. We also describe our approach to constructing the training samples for SPA-NET. Our goal is to train SPA-NET to recognize which jets arise from the decay of a given Higgs boson, thereby identifying the correct pairing of the Higgs boson candidates.

Figure 2 shows the high-level model structure of SPA-NET [9–11]. The embedding blocks encode the input features to the embedding vectors living in the latent space. These embedding vectors are fed into the central transformer, which is a stack of transformer encoders. The central transformer then outputs the event embedding vector, which is used in the jet assignment and the classification tasks. For the jet assignment, the event embedding vector is encoded by the particle transformers and the tensor attentions. Finally, SPA-NET constructs the jet assignment results from these outputs. The network architecture has a feed-forward structure for the classification head.

The key concept underlying SPA-NET is that it can encode inputs into an abstract latent space and utilize attention mechanisms to learn contextual information about jet relationships, which is used in subsequent tasks, thereby avoiding the permutation approach in the classical methods.

There are several inherent symmetries in the jet assignment task. For example, the detector signatures of quarks and anti-quarks are nearly indistinguishable. Consequently,

JHEP09(2024)139

it is important to consider all possible combinations of jets originating from these partons. Moreover, the reconstruction task is insensitive to the swapping of specific labels. For instance, in the di-Higgs case, swapping the pairing results of two Higgs would lead to the same event reconstruction. The design of SPA-NET's model structure and loss function takes these symmetries into account.

Due to the properties of the transformer, the event embedding vector in SPA-NET is independent of the order of the input jets. Moreover, SPA-NET utilizes the technique of symmetric tensor attention [10], which constructs a tensor with permutation symmetries of labels (e.g., the $b\bar{b}$ and hh pairs). Therefore, the output also contains label permutation symmetries. For the pairing loss function, SPA-NET utilizes the combined symmetric loss that incorporates permutation symmetries with the cross-entropy loss, as defined in equation 6 of ref. [10]. These properties enable SPA-NET to preserve the symmetry inherent to the jet assignment problems.

In our context, it is essential to emphasize that SPA-NET is not restricted to using only the b-tagged jets for the jet assignment task, but considers all jets in an event. This allows the network to make a correct prediction even in a situation where some of the jets are mistagged. Therefore, SPA-NET can utilize a larger dataset in the pairing task than the traditional methods.

The input features for the SPA-NET are a list of jets, each represented by its 4-component vector $(p_{\mathrm{T}}, \eta, \phi, m)$ as well as a boolean b-tag, where ϕ is the azimuthal angle and m is the invariant mass. To be specific, we only keep the 10 highest p_{T} jets in each event. For each event, we define the correct jet assignments by matching the jets to the simulated truth quarks within an angular distance of $\Delta R < 0.4$. In case a simulated truth quark is matched to more than one jet, such an event will be dropped. Furthermore, some simulated truth quarks may not be matched to any jet, in which case the event will not be used in training either. The percentage of samples satisfying these matching conditions to the samples passing the four-tag cut varies from 77% to 89%. Note that these matching conditions are only required for training SPA-NET. However, in the final analysis, SPA-NET is applied to all events passing the four-tag cut.

4 Neural network classifiers

After jet pairing, the next step is to distinguish the signal from the background. In addition to a cut-based approach, two types of neural network classifiers are employed in this study: the Dense-NN, which is a conventional deep learning architecture used as the baseline neural network approach, and SPA-NET, which could also be used as a classifier to separate signal from background.

To perform a Dense-NN-based analysis, we construct a Dense-NN classifier to distinguish between signal and background events. We implement our Dense-NN using the library of the `Tensorflow` [44]. The network consists of simple dense layers and the internal node uses the rectified linear unit (ReLU) as the activation function. The categorical cross-entropy is used as the loss function, which is then minimized by the `Adam` algorithm. Hyperparameters of the Dense-NN are selected by utilizing the `Optuna` [45] hyperparameter optimization package. The learning rate, hidden dimension, and the number of layers are optimized by performing

JHEP09(2024)139

100 iterations of hyperparameter optimization, and the set of hyperparameters that produces the best classification accuracy is selected for full training.

Upon training, the Dense-NN is used to determine whether an event is a signal or a background. The Dense-NN classifier assigns a signal score p_{signal} to every event, which represents the confidence that this event is a signal. An event is classified as a signal if its p_{signal} is larger than p_{th}, a threshold score determined through the maximization of sensitivity $S/\sqrt{B}$, where S and B represent the number of signal and background events, respectively.

For the SPA-NET classifier, it is important to note that the SPA-NET classification head does not take the results from the jet assignment part. Using the transformer outputs alone produces better performance compared to including the jet assignment results because errors in the jet assignment part can affect the overall performance. However, even if we only use the results from the classification head, we still train both the jet assignment and classification tasks simultaneously by adding losses in both parts.

When training SPA-NET on both tasks, it is essential to provide more information than just focusing on the classification task. The jet assignment information can help SPA-NET build the embedding space structure. As a result, training on both tasks allows us to achieve better performance while using the same size of training samples compared to only training on the classification head.

The hyperparameters of SPA-NET are selected using the `Optuna` hyperparameter optimization package. We optimize the learning rate, dropout rate, gradient clipping, L2 penalty, hidden dimension, number of encoder layers, number of branch encoder layers, and number of classification layers. Each set of hyperparameters is trained for 10 epochs. We perform 100 iterations of hyperparameter optimization, and the set of hyperparameters that produces the best classification accuracy is selected for full training. The hyperparameter optimization results are summarized in appendix A.

Following the hyperparameter optimization process, SPA-NET is trained for 50 epochs using the `AdamW` optimizer with L2 regularization. The total loss in SPA-NET combines the contributions from both jet assignment and classification parts, indicating that these parts are not trained independently. The loss from each part is computed separately and summed with equal weights.

Similarly, the SPA-NET assigns a signal score p_{signal} to each event. To select the di-Higgs candidate events, we set a requirement $p_{\text{signal}} > p_{\text{th}}$ and the threshold p_{th} is determined through the maximization of sensitivity $S/\sqrt{B}$.

Additionally, in the resonant (non-resonant) analysis, SPA-NET consists of approximately 37.9 million (541 thousand) trainable parameters, which are significantly more than the Dense-NN's 268 thousand (136 thousand) parameters. Although such a larger number of parameters allows the model to capture more complex patterns, it could also make the training more challenging. SPA-NET is well-structured to handle these parameters, making the training possible. Note in particular that SPA-NET is trained on both jet pairing and classification simultaneously, while Dense-NN is trained only on the classification task.

5 Search for resonant Di-Higgs production

For resonant Higgs boson pairs, we describe the steps to set the 95% confidence level (CL) upper limits on the cross-section of the resonant production of a new heavy scalar H decaying

JHEP09(2024)139

into two Higgs bosons hh and demonstrate that SPA-NET gives the best limit among all three methods.

5.1 Event selection in resonant analysis

5.1.1 Cut-based selection

After jet pairing as described in section 3, we define the leading Higgs boson candidate h_1 to be the one with the highest scalar sum of jet p_{T}. The sub-leading Higgs is denoted by h_2. The following transverse momentum cuts are further applied to the leading and sub-leading Higgs candidates [13]:

$$\begin{aligned} p_{\mathrm{T}}^{\mathrm{lead}} &> m_{4\mathrm{j}} \times 0.5 - 103\,\mathrm{GeV}\,, \\ p_{\mathrm{T}}^{\mathrm{subl}} &> m_{4\mathrm{j}} \times 0.33 - 73\,\mathrm{GeV}\,, \end{aligned} \tag{5.1}$$

where $m_{4\mathrm{j}}$ is the total invariant mass of the two Higgs candidates, $p_{\mathrm{T}}^{\mathrm{lead}}$ is the transverse momentum of the leading Higgs boson candidate, and $p_{\mathrm{T}}^{\mathrm{subl}}$ is for the sub-leading Higgs boson candidate.

For background rejection, we first apply a cut on the pseudorapidity difference between the two Higgs candidates $|\Delta\eta_{hh}| < 1.5$. Next, we define the quantity X_{hh} [13]

$$X_{hh} = \sqrt{\left(\frac{m_{h_1} - 120\,\mathrm{GeV}}{0.1 m_{h_1}}\right)^2 + \left(\frac{m_{h_2} - 110\,\mathrm{GeV}}{0.1 m_{h_2}}\right)^2}\,, \tag{5.2}$$

where m_{h_1} is the mass of the leading Higgs candidate, and m_{h_2} is the mass of the sub-leading Higgs candidate. Events with $X_{hh} < 1.6$ are considered as in the signal region. The reference masses of 120 GeV and 110 GeV account for energy losses in the detector.

A top veto cut is needed to suppress the $t\bar{t}$ background. We form "W candidates" by pairing every possible pair of jets with $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ and $|\eta| < 2.5$, including those that are not selected as the h candidates. We then build "top quark candidates" by pairing the W candidates with each remaining jet selected for the h candidates. For each possible top quark candidate, we calculate the quantity X_{Wt} defined as [13]

$$X_{Wt} = \sqrt{\left(\frac{m_W - 80\,\mathrm{GeV}}{0.1 m_W}\right)^2 + \left(\frac{m_t - 173\,\mathrm{GeV}}{0.1 m_t}\right)^2}\,, \tag{5.3}$$

where m_W is the mass of the W candidate, and m_t is the mass of the top quark candidate. For each event, there are multiple X_{Wt} values, corresponding to different possible top quark candidates. An event is vetoed if any of its X_{Wt} is less than 1.5.

5.1.2 Dense-NN selection

To use the Dense-NN classifier, we need first to apply jet assignments. We employ the min-ΔR and SPA-NET pairing methods to construct the Higgs candidates and generate two separate training datasets. Subsequently, we utilize these datasets to train two separate Dense-NN classifiers and use the signal scores to separate the signal from the background. The input variables utilized by the Dense-NN classifiers are summarized in table 2, as inspired by ref. [18]. These features include the 4-vector of the two Higgs candidates, the angular distance

JHEP09(2024)139

Reconstructed objects	Input variables	#
Higgs candidate	$(p_{\mathrm{T}}, \eta, \phi, m)$	8
Jet	$\Delta R(j_1, j_2)$	2
b-tagging	Boolean for $j_i \in h_{1,2}^{\mathrm{cand}}$	4
Di-Higgs system	$p_{\mathrm{T}}^{hh}, m_{hh}$	2

Table 2. Input variables for the dense neural network in the resonant analysis.

ΔR between the two jets associated with each Higgs candidate, the b-tagging information of the four jets, and the transverse momentum of the di-Higgs system. The threshold values are determined through the maximization of the sensitivity $S/\sqrt{B}$ at $m_H = 1000$ GeV events.

5.1.3 Spa-Net selection

Similar to the Dense-NN selection approach, in the SPA-NET approach we utilize the signal scores generated by SPA-NET and set a specific threshold value. Events with a signal score greater than or equal to this threshold will be considered candidate resonant events. The threshold values are determined through the maximization of the sensitivity $S/\sqrt{B}$ at $m_H = 1000$ GeV events.

5.2 Results from resonant searches

For the cut-based and Dense-NN selection methods, it is necessary to construct the Higgs boson candidates. Figure 3 shows the pairing efficiency of various methods. All pairing methods exhibit better performance in the higher mass region, while the pairing efficiency declines more significantly as the mass goes below $\sim 500\,\mathrm{GeV}$. This effect is especially noticeable in the min-ΔR method. This is because, in the low-resonance region, the Higgs boson obtains lower energy, causing the b-jet pair to have a larger ΔR separation, resulting in reduced performance. The SPA-NET pairing method outperforms other methods for all mass values.

Figure 4 shows the selection efficiency, defined to be the ratio of the number of events that pass the final cut to the total number of events without applying any cuts, for the cut-based selection with different pairing methods. The corresponding selection efficiencies of background samples range from 9.4×10^{-5} to 2.82×10^{-4}. All three pairing methods exhibit similar performance. In the low-mass region, the efficiency is reduced due to the lower energy. The min-ΔR pairing method has even lower efficiency in this region due to its inferior pairing performance.

Table 3 presents the training results for the neural network classifiers, where we use the accuracy (ACC) and the area under the Receiver Operating Characteristic (ROC) curve (AUC) as two evaluation metrics. The Dense-NN classifier with the SPA-NET pairing method shows better performance than the Dense-NN with the min-ΔR pairing method. The SPA-NET classifier has the best performance among the three classifiers.

The difference between the Dense-NN and the SPA-NET classifiers arises from the input features. While the Dense-NN employs well-known physical observables as the input features, the SPA-NET classifier uses event embedding vectors. These vectors are generated using all the information in the event, including both Higgs and background jets. Even though the

JHEP09(2024)139

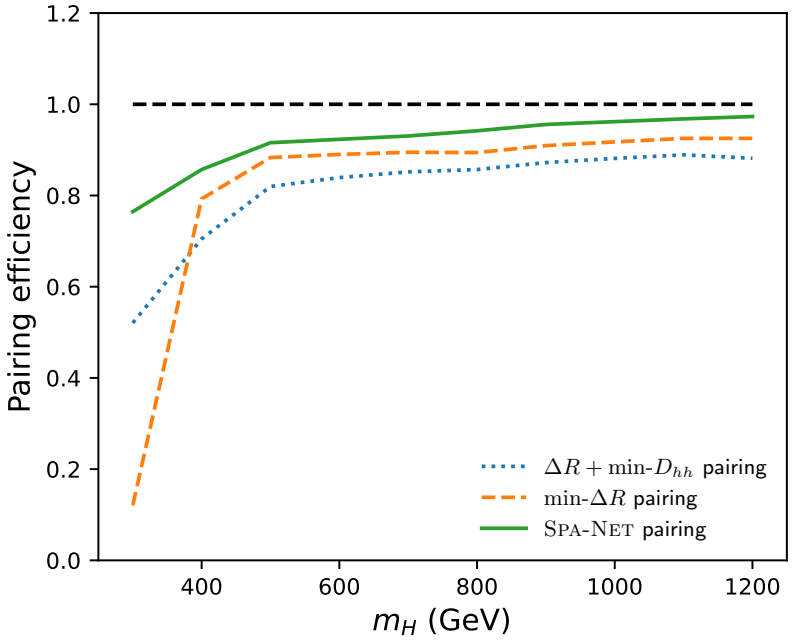


Figure 3. The pairing performance for samples with different m_H.

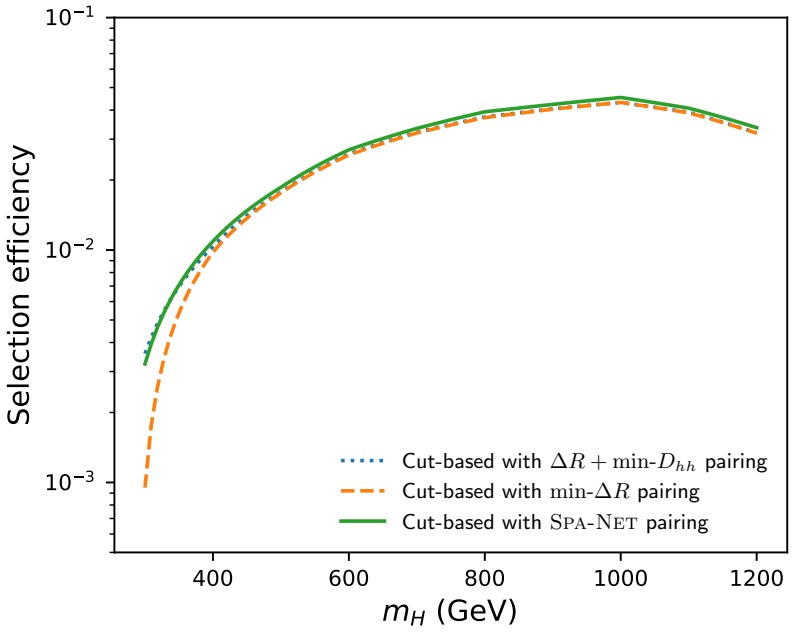


Figure 4. The selection efficiency for samples with different m_H.

JHEP09(2024)139

Classifier	ACC	AUC
Dense-NN with min-ΔR pairing	0.865 ± 0.001	0.938 ± 0.001
Dense-NN with SPA-NET pairing	0.876 ± 0.001	0.946 ± 0.001
SPA-NET	0.894 ± 0.002	0.961 ± 0.001

Table 3. The classification performance of different neural network classifiers. The ACC and AUC are evaluated based on 10 trainings.

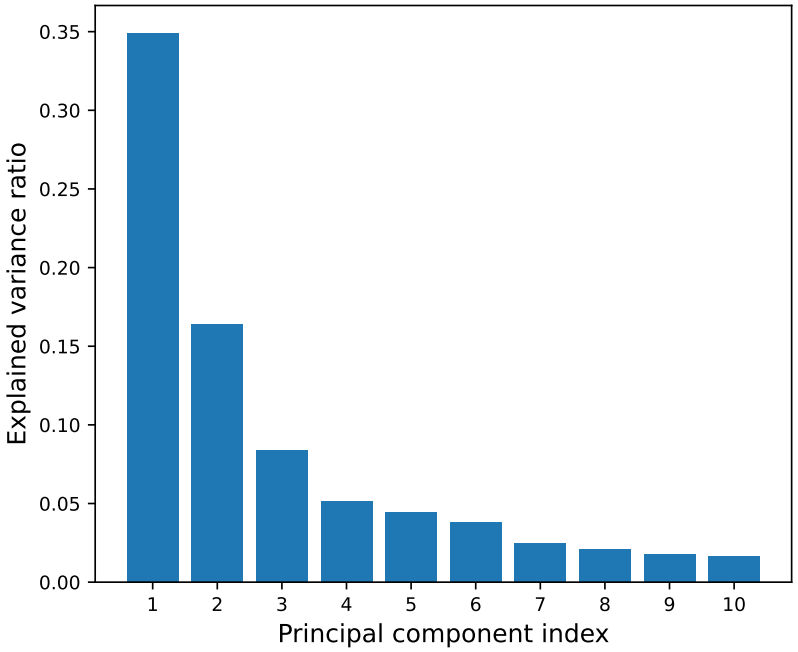


Figure 5. The variance ratio of the first ten principal components in resonant analysis.

event embedding vector cannot be readily interpreted as traditional physical observables, it is found to be better suited for the event classification task. To understand the physical information encoded in these event embedding vectors, further analysis is needed to find out the relationship between the high-level physical observables and the components of the event embedding vectors.

To better understand the embedding vectors, we have performed the Principal Component Analysis (PCA), which uses an orthogonal linear transformation that transforms the data to a new basis. (The PCA class from the `scikit-learn` [46] package is used.) In the new basis, the components are ordered by their variance. Figure 5 shows the variance importance for the first 10 principal components. The first three components can explain about 60% of the total variance. This indicates that these components capture significant information from the event embedding vectors. Therefore, in the following analysis, we employ only these first three principal components.

To find the correlation between the first three principal components and selected Dense-NN input features, we calculated the correlation coefficients, with the results shown in figure 6. The features such as Higgs candidates' p_{T}, ΔR, di-Higgs system's p_{T}^{hh}, and m_{hh} have

JHEP09(2024)139

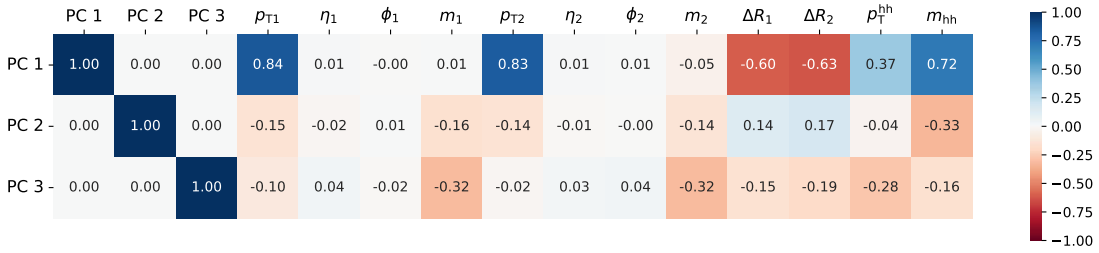


Figure 6. The correlation coefficients between the first three principal components and high-level observables.

higher correlation coefficients with the principal components. This suggests that the event embedding vectors indeed contain important and meaningful physical information. Also, the transformer block in SPA-NET can learn and extract relevant physical information from the input features and encode it into event embedding vectors.

Even though there is a high correlation between the embedding vector and high-level observables, the better training results of SPA-NET imply that the embedding vector contains additional information. This information might not correspond to the familiar high-level physical parameters, which nonetheless proves effective for the classification task. These findings highlight the fact that the SPA-NET can extract the relevant physical information from the input data efficiently. In addition, the flexibility of SPA-NET allows it to construct more suitable variables for the classification task. As a consequence, the training performance exceeds that of utilizing only well-known physical parameters.

To obtain the upper limits on the cross-section of resonant Higgs boson pair production, we use the reconstructed m_{hh} distribution. From the binned m_{hh} distribution, we can obtain the likelihood function L, which consists of a product of Poisson distributions for the number of events in each bin.

The signal strength μ is chosen as the parameter of interest (POI). The profile likelihood ratio takes the following form:

$$-2\Delta \ln L(\mu) \equiv -2 \ln \left(\frac{L(\mu)}{L(\hat{\mu})} \right) , \tag{5.4}$$

where $\hat{\mu}$ is the maximum likelihood estimate of μ. The upper limit on the signal strength is computed by the package `pyhf` [47, 48], which is based on the CL_s method [49]. The POI is excluded at the 95% CL when CL_s is less than 0.05. Then we can convert the results to the upper limit of the cross-section.

In setting the upper limit, we consider a luminosity of $\mathcal{L} = 300\,\text{fb}^{-1}$ for the 14-TeV LHC. Since the kinematics of the 13-TeV and 14-TeV samples are similar, we can scale the cross-sections to those of the 14-TeV samples. Figure 7 shows the upper limits on the resonant Higgs pair production as a function of m_H for different selection methods. While all methods give similar results in the high resonance region, the SPA-NET selection has superior performance, providing the most stringent constraints throughout the considered mass range. The Dense-NN selection methods give higher upper limits than SPA-NET, and the cut-based selection methods give the worst results. Specifically, SPA-NET selection

JHEP09(2024)139

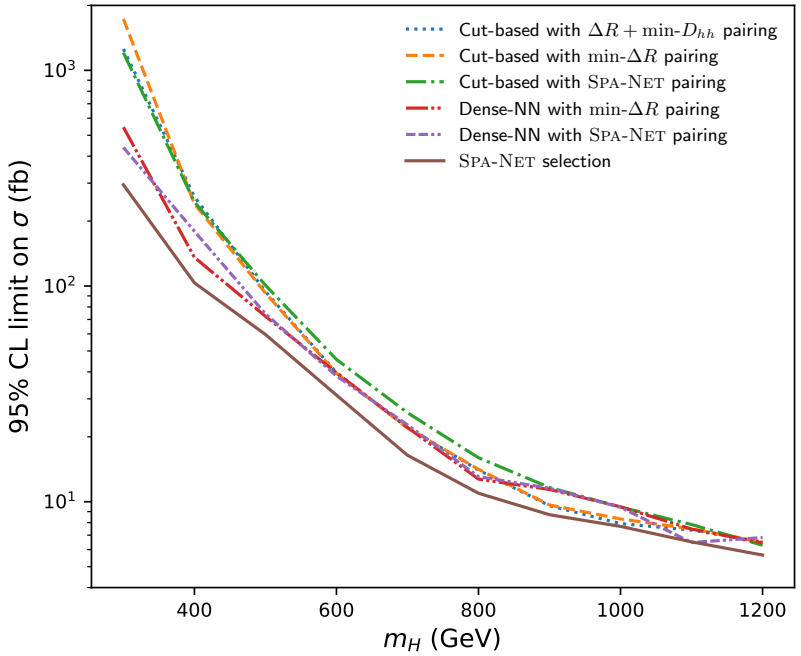


Figure 7. The upper limit on the cross-section of resonant di-Higg production for different m_H.

enables us to establish cross-section upper limits that are 10% to 45% stronger compared to Dense-NN with min-ΔR pairing.

Figure 8 shows the invariant mass m_{hh} distributions using various selection methods. In the high-mass region, the Dense-NN and SPA-NET selection methods let more background and signal events pass. Consequently, the results are similar to those obtained by the cut-based selection. In the low-mass region, the SPA-NET selection method can cut more background events, which accounts for why the SPA-NET selection achieves more stringent upper limits in this specific mass range.

6 Higgs self-coupling constraints

In this section, our signal events consist of non-resonant Higgs boson pairs with different Higgs self-coupling scale factor κ_λ. Similar to the analysis on resonant production, we compare the cut-based method with two distinct neural network classifiers, the Dense-NN and the SPA-NET, which are used to identify candidate signal events. We will show that the SPA-NET architecture is better suited for the event classification task, and the improved classification results can yield stronger constraints on the coupling scale factor κ_λ.

6.1 Event selection in non-resonant analysis

6.1.1 Cut-based selection

The cut-based event selection methods are similar to that of the resonant analysis. In the non-resonant analysis, we utilize the min-ΔR and SPA-NET pairing methods. The high-level physical variables of two Higgs boson candidates are used in subsequent selection steps and analysis.

JHEP09(2024)139

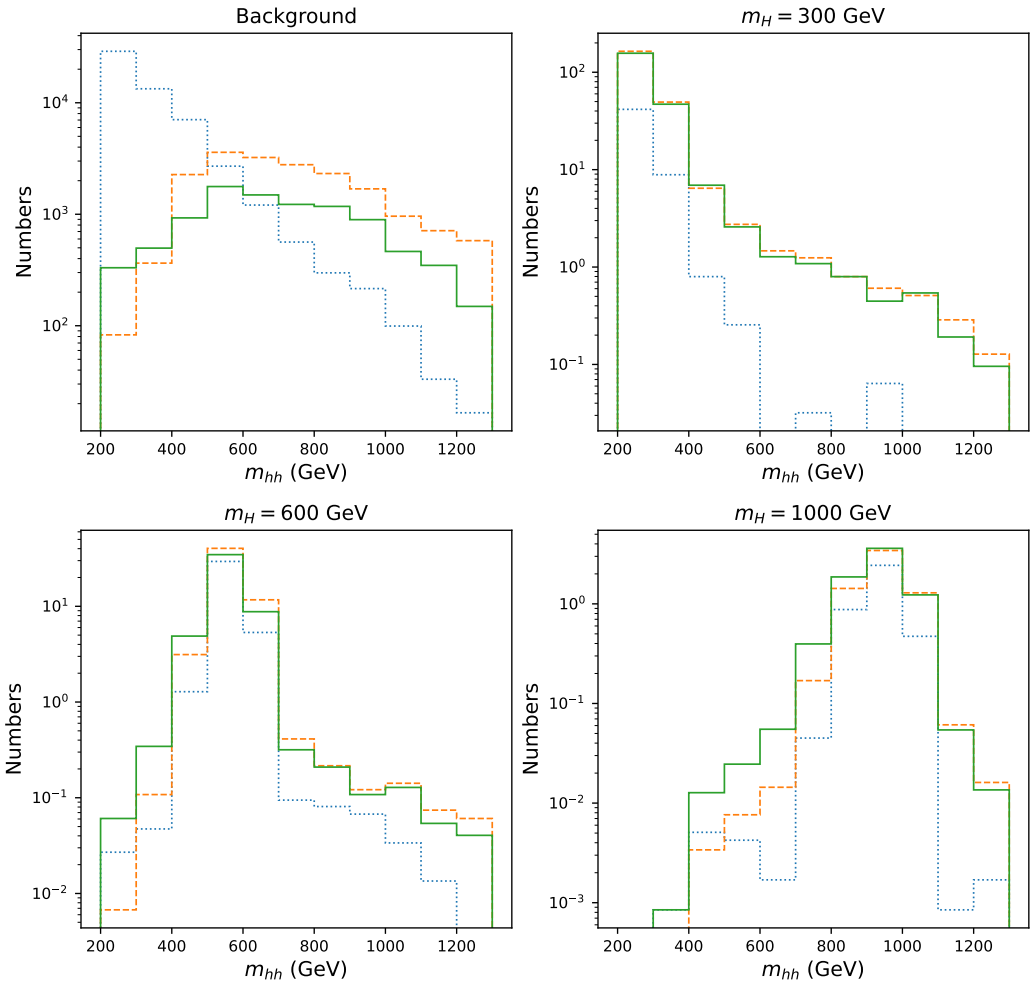


Figure 8. The invariant mass m_{hh} distribution for the cut-based selection with SPA-NET pairing (blue dotted bins), the Dense-NN with SPA-NET pairing (orange dashed bins), and the SPA-NET selection (green solid bins). The number of events is normalized to the target luminosity $\mathcal{L} = 300\,\text{fb}^{-1}$. Other pairing methods would give similar results to the corresponding selection method with SPA-NET pairing.

To further reduce the multijet background, we apply a cut on the pseudorapidity difference between the two Higgs candidates $|\Delta\eta_{hh}| < 1.5$. To reduce the $t\bar{t}$ background, we employ the top veto cut. We compute the quantity X_{Wt}, corresponding to different possible top quark candidates. An event is vetoed if any of its X_{Wt} is less than 1.5.

Finally, events with $X_{hh} < 1.6$ are considered as in the signal region (see eq. (5.2) for the definition of X_{hh}). Those events would be used to determine the Higgs self-coupling constraints.

6.1.2 Dense-NN selection

The signal samples are non-resonant samples with different κ_λ values. The network's input features are summarized in table 4, as inspired by ref. [18]. To obtain these input features, the Higgs pairing needs to be determined first. We use the min-ΔR pairing method and the

JHEP09(2024)139

Reconstructed objects	Variables used for training	#
Higgs candidate	$(p_{\mathrm{T}}, \eta, \phi, m)$	8
Jet	$\Delta R(j_1, j_2)$	2
Missing transverse momentum	$E_{\mathrm{T}}^{\mathrm{miss}}, \phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	2
Leptons	N_e, N_μ	2
b-tagging	Boolean for $j_i \in h_{1,2}^{\mathrm{cand}}$	4
Di-Higgs system	$p_{\mathrm{T}}^{hh}, m_{hh}$	2
Self coupling	κ_λ	1

Table 4. Input variables for the dense neural network in the non-resonant analysis.

SPA-NET pairing method to construct the Higgs candidates and generate two separate training datasets, which are then used to train two different Dense-NN classifiers. For training, we set $\kappa_\lambda = [-5, -3, -1, 1, 2, 3, 5, 7, 9, 12]$ and generate 9,000 samples for each value in the list. For simplicity, we do not generate samples across the full range $[-10, 15]$ mentioned in section 2.1, as samples away from the boundaries do not improve the Dense-NN's performance in the κ_λ constraint setting. For background samples, since they lack specific κ_λ information for the input feature, a κ_λ value is randomly chosen from the above list. We use the same sizes of signal and background samples as the scheme given in table 1 in the neural network training.

For testing, the κ_λ input is scanned over the full range, and the average of outputs is utilized as the final signal score p_{signal}. An event is classified as a signal if its p_{signal} is larger than p_{th}, which is determined by maximizing the sensitivity $S/\sqrt{B}$ on the $\kappa_\lambda = 1$ sample.

6.1.3 Spa-Net selection

Here we utilize SPA-NET to perform both the jet assignment and signal/background classification tasks. The input data contains information about the reconstructed jets and the global event features. SPA-NET would output both the jet pairing and the type of the events (signal vs. background). In contrast to the Dense-NN classifier, the key advantage is that SPA-NET can process the jet-level information and can preserve the permutation symmetry inherent in the problem.

The input features contain a list of jets, where each jet is represented by its 4-component vector $(p_{\mathrm{T}}, \eta, \phi, m)$ as well as a Boolean b-tag. Additionally, we input the self-coupling scale factor κ_λ as the global feature of each event. The sample preparation and testing steps follow the same procedure as described for the Dense-NN case in the previous section.

6.2 Non-resonant analysis results

To obtain the high-level physical observables for cut-based selection and Dense-NN training, we need to construct the Higgs boson candidates with different jet assignment algorithms. Figure 9 shows the pairing efficiency of the two jet pairing methods: min-ΔR and SPA-NET. Both methods exhibit their best performance around $\kappa_\lambda = 2$. For the min-ΔR method, the pairing efficiency ranges from 40% to 80% while for SPA-NET the pairing efficiency ranges from 70% to 90%. Therefore, the SPA-NET pairing method outperforms the min-ΔR method for all coupling values. Figure 10 shows the selection efficiency for the cut-based

JHEP09(2024)139

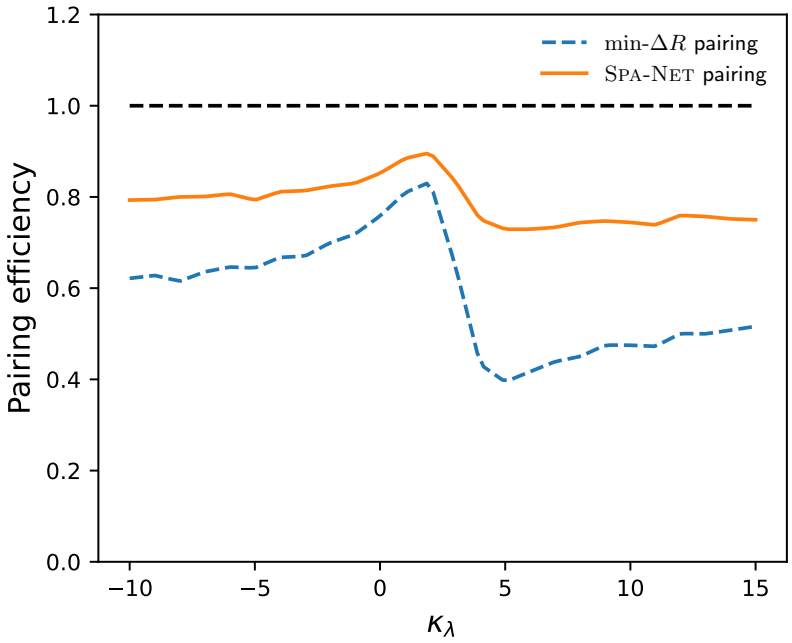


Figure 9. The pairing performance for different κ_λ samples. The SPA-NET is trained on pairing and classification tasks at the same time.

Classifier	ACC	AUC
Dense-NN with min-ΔR pairing	0.799 ± 0.011	0.881 ± 0.012
Dense-NN with SPA-NET pairing	0.803 ± 0.004	0.884 ± 0.004
SPA-NET	0.828 ± 0.002	0.911 ± 0.001

Table 5. The classification performance of different selection methods. The ACC and AUC are evaluated based on 10 trainings.

selection using different pairing methods. The curves of the selection efficiency are similar to the ones of pairing efficiency, meaning that the selection efficiency and the pairing results are highly correlated.

In table 5, we present the training results for the classifiers. The Dense-NN classifiers perform similarly for both pairing methods, while the SPA-NET classifier performs best. To better understand the event embedding vectors and find out the relationship between the input features of Dense-NN and those of SPA-NET, we perform a similar analysis as in the resonant case.

First, we performed the PCA on the event embedding vectors. Figure 11 shows the variance importance for the first 10 principal components, with the first three components being able to explain about 50% of the total variance. Since these components capture significant information from the event embedding vectors, we focus on these first three principal components in subsequent analysis.

Next, we compute the correlation coefficients between these principal components and high-level physical observables. The results are shown in figure 12. The features such as

JHEP09(2024)139

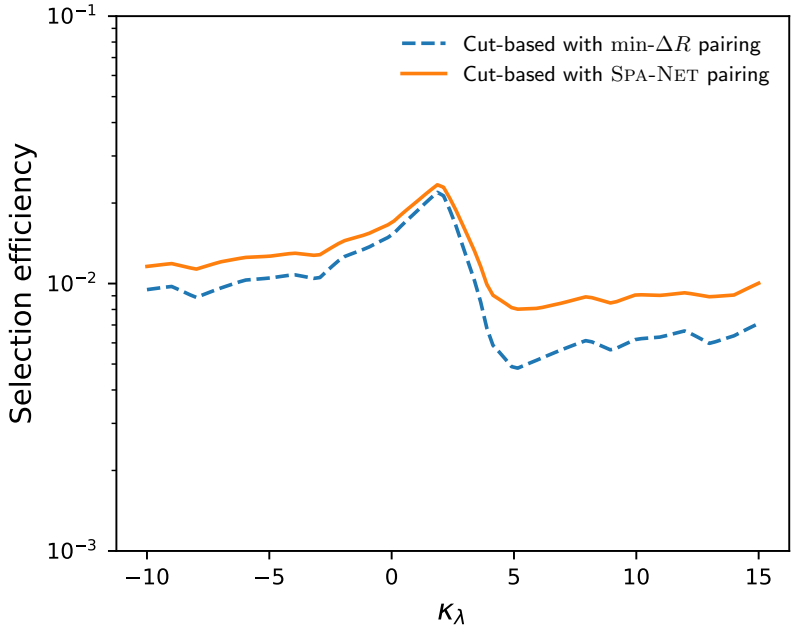


Figure 10. The selection efficiency for different κ_λ samples. The selection efficiency is the ratio of the number of events that pass the final cut to the total number of events without applying any cuts. The corresponding selection efficiencies of background samples range from 9.7×10^{-5} to 2.72×10^{-4}.

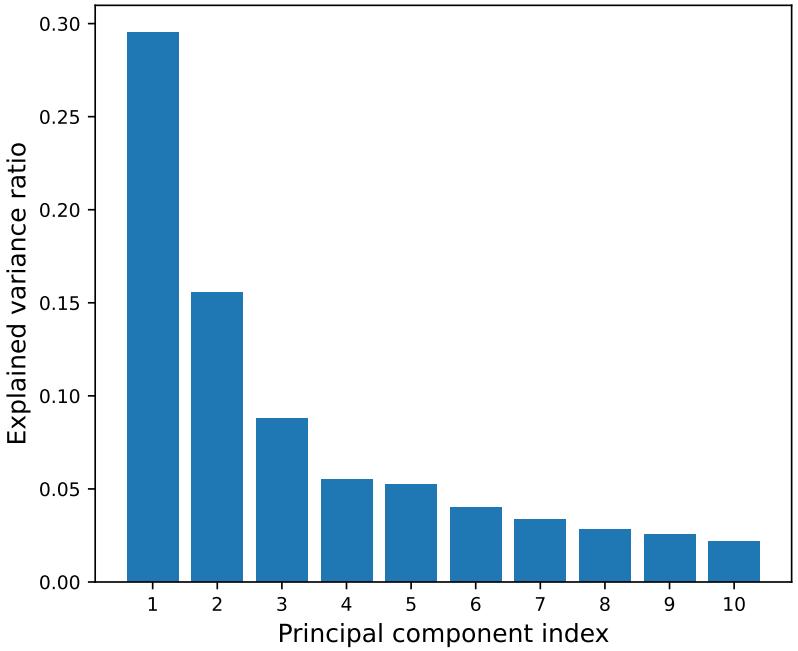


Figure 11. The variance ratio of the first ten principal components in the non-resonant analysis.

JHEP09(2024)139

Figure 12. The correlation coefficients between the first three principal components and high-level observables.

	Expected Constraints					
POI	Signal strength μ			Self-coupling κ_λ		
S/B selection	Lower	Upper	$\Delta\kappa_\lambda$	Lower	Upper	$\Delta\kappa_\lambda$
Cut-based with min-ΔR pairing	−3.30	10.38	13.68	−3.27	10.33	13.60
Cut-based with SPA-NET pairing	−3.58	10.47	14.05	−3.55	10.43	13.98
Dense-NN with min-ΔR pairing	−1.44	8.49	9.93	−1.44	8.49	9.93
Dense-NN with SPA-NET pairing	−1.72	8.94	10.66	−1.72	8.99	10.71
SPA-NET selection	−1.57	7.49	9.06	−1.56	7.57	9.13

Table 6. Constraints on κ_λ using different selection methods. We consider a luminosity of $\mathcal{L} = 300\,\text{fb}^{-1}$ for the 14-TeV LHC.

Higgs candidates' p_T, ΔR, di-Higgs system's p_T^{hh}, and m_{hh} have higher correlation coefficients with the principal components. These results are similar to the resonant case. These findings suggest that the event embedding vectors indeed contain meaningful physical information. Moreover, the flexibility of SPA-NET enables it to explore beyond the familiar physical parameters, thereby allowing SPA-NET to obtain superior performance compared to the traditional Dense-NN structure in classification tasks.

For the κ_λ constraints setting, we consider a luminosity of $\mathcal{L} = 300\,\text{fb}^{-1}$ and use the 14-TeV cross-sections. Since the kinematics of the 13-TeV and 14-TeV samples are similar, we can scale the cross-sections to those of the 14-TeV samples.

Similar to the resonant analysis, to obtain constraints on the Higgs self-coupling scale factor κ_λ, we use the reconstructed m_{hh} distribution, from which we compute the likelihood function L consisting of a product of Poisson distributions for the number of events in each bin. The signal strength μ is chosen as the POI. The values of the coupling κ_λ are excluded at the 95% CL if the predicted cross-section of the signal model with that configuration is excluded with $\text{CL}_\text{s} < 0.05$. Alternatively, we can obtain the exclusion limit by using the profile likelihood with the coupling κ_λ as POI. A scan of the profile likelihood ratio is taken as a function of the coupling, and from this, we can set the 1.96σ-level constraints.

When using the signal strength μ as the POI, figure 13 displays the upper limits of the hh cross-section as a function of κ_λ. The dip of the curves around $\kappa_\lambda = 2.45$ is due to the complete destructive interference [50]. On the other hand, when using κ_λ as the POI, the profile likelihood ratio scan for κ_λ is shown in figure 14. All results are presented in table 6. The SPA-NET classifier shows a significant improvement in the upper bound

JHEP09(2024)139

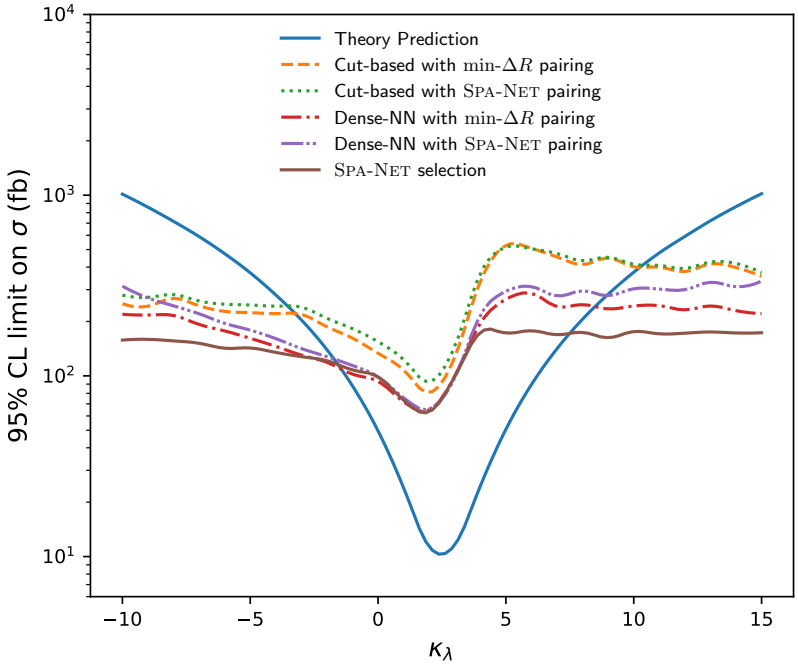


Figure 13. The upper limits of the cross-section with different κ_λ. The theory prediction is the cross-section computed from `MadGraph5_aMC@NLO`. The coupling κ_λ with the cross-section greater than the upper limit would be excluded.

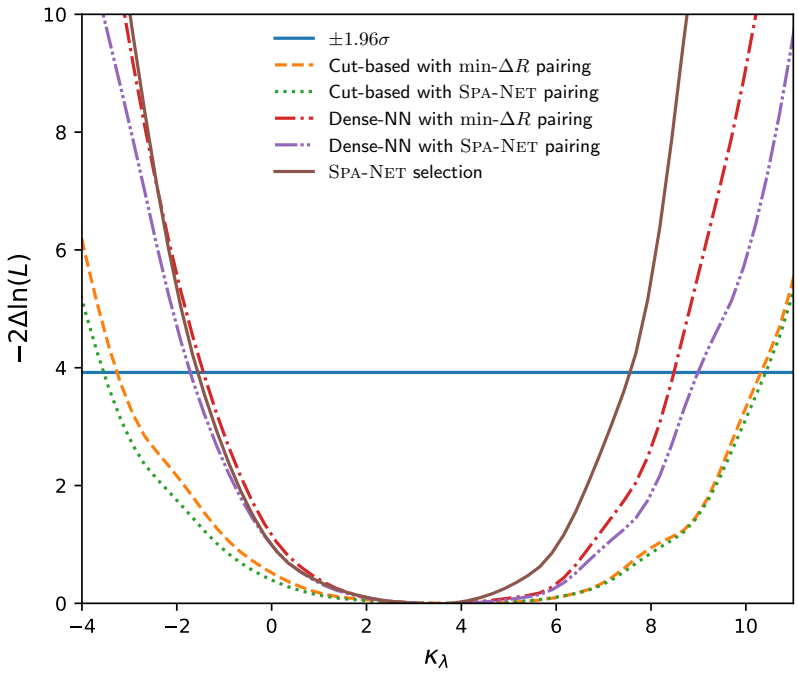


Figure 14. The profile likelihood ratio scans for κ_λ. The blue line indicates the 1.96σ exclusion boundary.

JHEP09(2024)139

of the κ_λ constraints. Regardless of the selected POI, the SPA-NET classifier provides the strongest constraints on κ_λ. When using μ (κ_λ) as the POI, values of κ_λ beyond $[-1.57, 7.49]$ ($[-1.56, 7.57]$) would be excluded at the 95% CL (1.96σ) level.

To achieve similar performance using the min-ΔR Dense-NN classifier, the luminosity would need to be scaled to $\mathcal{L} \simeq 400\,\text{fb}^{-1}$. In this sense, the SPA-NET classifier offers a luminosity gain of $\mathcal{L} = 100\,\text{fb}^{-1}$. These results demonstrate that the SPA-NET classifier outperforms other methods in constraining Higgs self-coupling scale factor κ_λ.

7 Conclusions

In this paper, we have utilized a novel deep neural network architecture called SPA-NET to improve the analysis for di-Higgs events in the $4b$ final state. By comparing SPA-NET with the traditional cut-based method, we have shown its better pairing efficiency in both resonant and non-resonant scenarios. Moreover, SPA-NET also outperforms the Dense-NN classifier in the classification task.

At the LHC with a 14-TeV C.M. energy and an integrated luminosity of $300\,\text{fb}^{-1}$, SPA-NET selection allows us to establish 95% C.L. cross-section upper limits in resonant analysis that are 10% to 45% stronger compared to Dense-NN with min-ΔR pairing. In non-resonant analysis, the SPA-NET selection has provided 9% stronger bounds on κ_λ couplings when compared to the baseline method, the min-ΔR Dense-NN selection. To achieve similar performance using the min-ΔR Dense-NN selection, the luminosity would need to be scaled to $\mathcal{L} = 400\,\text{fb}^{-1}$. Therefore, the SPA-NET selection offers a luminosity gain of $\mathcal{L} = 100\,\text{fb}^{-1}$.

The key difference between Dense-NN and SPA-NET classifiers is in the input features. The SPA-NET classifier uses the embedding vectors as its input features. We have demonstrated that embedding vectors are related to high-level physical observables and they can capture the physical information from the events. This feature makes the embedding vectors well-suited for event classification tasks.

Our current work focuses primarily on the efficacy of SPA-NET in the di-Higgs analysis. However, we acknowledge the potential of alternative architectures such as Deep Sets [51]. A possible research direction is to compare the performance of Deep Sets with that of SPA-NET and traditional Dense-NN models. This could potentially shed more light on the relative contributions of jet permutation invariance and particle label symmetries and explore the impact of different input features. This analysis would deepen our understanding of the underlying mechanisms of neural network architectures in physics analyses.

In conclusion, our work highlights the remarkable potential of SPA-NET in di-Higgs event analysis. Through the superior performance in both jet pairing and classification tasks, SPA-NET improves upper limits on production cross-sections and Higgs self-couplings. Moreover, SPA-NET has been proven to be not biased on simulator-specific information [10], suggesting its potential applicability to experiments. Beyond Higgs physics, the methodologies developed here offer a promising approach for various analyses suffering from complex pairing tasks, paving the way for future advancements in particle physics research.

Data and code availability. The training and testing datasets used in this study are publicly available on Zenodo [52]. The code used for SPA-NET training in this paper is accessible on GitHub [53].

JHEP09(2024)139

Parameter	Resonant		Non-resonant	
	min-ΔR	SPA-NET	min-ΔR	SPA-NET
Learning rate	0.000227	0.000262	0.00495	0.000948
Batch Size	512	512	512	512
Hidden Dimensionality	256	256	256	256
Layer Count	5	5	3	2

Table 7. The hyperparameters used in Dense-NN training. The results of different pairing methods are presented.

Parameter	Resonant	Non-resonant
Learning Rate	0.00049	0.00659
Training Epochs	50	50
Batch Size	2048	2048
Dropout	0.061	0.0059
L_2 Gradient Clipping	0.445	0.425
L_2 Penalty	0.000382	0.000374
Hidden Dimensionality	256	32
Central Encoder Count	8	8
Branch Encoder Count	4	2
Classification Layers	1	1
Assignment Loss Scale	1	1
Classification Loss Scale	1	1

Table 8. The hyperparameters used in SPA-NET training.

Acknowledgments

We are grateful to Alexander Shmakov for the assistance with the SPA-NET package. Additionally, we extend special thanks to David Shih and Alexander Shmakov for their valuable comments on our manuscript. C.-W. Chiang and F.-Y. Hsieh are supported in part by the National Science and Technology Council of Taiwan under Grant No. NSTC-111-2112-M-002-018-MY3. S.-C. Hsu is supported by the U.S. National Science Foundation grants No. 2110963. Work at Argonne is supported in part by the U.S. Department of Energy under contract DE-AC02-06CH11357. I. Low acknowledges the hospitality of the Phenomenology Group at National Taiwan University during the completion of this work.

A Hyperparameters

The neural network hyperparameters are optimized by the `Optuna` hyperparameter optimization package. Table 7 and table 8 provide the hyperparameters utilized in the Dense-NN and SPA-NET training, respectively.

JHEP09(2024)139

References

[1] ATLAS collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, *Nature* **607** (2022) 52 [*Erratum ibid.* **612** (2022) E24] [arXiv:2207.00092] [INSPIRE].

[2] CMS collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, *Nature* **607** (2022) 60 [arXiv:2207.00043] [INSPIRE].

[3] A.D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32 [INSPIRE].

[4] V.A. Kuzmin, V.A. Rubakov and M.E. Shaposhnikov, *On the Anomalous Electroweak Baryon Number Nonconservation in the Early Universe*, *Phys. Lett. B* **155** (1985) 36 [INSPIRE].

[5] A.G. Cohen, D.B. Kaplan and A.E. Nelson, *Progress in electroweak baryogenesis*, *Ann. Rev. Nucl. Part. Sci.* **43** (1993) 27 [hep-ph/9302210] [INSPIRE].

[6] S. Kanemura, Y. Okada and E. Senaha, *Electroweak baryogenesis and quantum corrections to the triple Higgs boson coupling*, *Phys. Lett. B* **606** (2005) 361 [hep-ph/0411354] [INSPIRE].

[7] S. Dawson, A. Ismail and I. Low, *What's in the loop? The anatomy of double Higgs production*, *Phys. Rev. D* **91** (2015) 115008 [arXiv:1504.05596] [INSPIRE].

[8] C.-R. Chen and I. Low, *Double take on new physics in double Higgs boson production*, *Phys. Rev. D* **90** (2014) 013018 [arXiv:1405.7040] [INSPIRE].

[9] M.J. Fenton et al., *Permutationless many-jet event reconstruction with symmetry preserving attention networks*, *Phys. Rev. D* **105** (2022) 112008 [arXiv:2010.09206] [INSPIRE].

[10] A. Shmakov et al., *SPANet: Generalized permutationless set assignment for particle physics using symmetry preserving attention*, *SciPost Phys.* **12** (2022) 178 [arXiv:2106.03898] [INSPIRE].

[11] M.J. Fenton et al., *Reconstruction of unstable heavy particles using deep symmetry-preserving attention networks*, *Commun. Phys.* **7** (2024) 139 [arXiv:2309.01886] [INSPIRE].

[12] A. Hammad, S. Moretti and M. Nojiri, *Multi-scale cross-attention transformer encoder for event classification*, *JHEP* **03** (2024) 144 [arXiv:2401.00452] [INSPIRE].

[13] ATLAS collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **01** (2019) 030 [arXiv:1804.06174] [INSPIRE].

[14] ATLAS collaboration, *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **105** (2022) 092002 [arXiv:2202.07288] [INSPIRE].

[15] ATLAS collaboration, *Search for nonresonant pair production of Higgs bosons in the $bb^{-}bb^{-}$ final state in pp collisions at s=13 TeV with the ATLAS detector*, *Phys. Rev. D* **108** (2023) 052003 [arXiv:2301.03212] [INSPIRE].

[16] CMS collaboration, *Search for resonant pair production of Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at 13 TeV*, *JHEP* **08** (2018) 152 [arXiv:1806.03548] [INSPIRE].

[17] CMS collaboration, *Search for Higgs Boson Pair Production in the Four b Quark Final State in Proton-Proton Collisions at s=13 TeV*, *Phys. Rev. Lett.* **129** (2022) 081802 [arXiv:2202.09617] [INSPIRE].

JHEP09(2024)139

[18] J. Amacker et al., *Higgs self-coupling measurements using deep learning in the $b\bar{b}b\bar{b}$ final state*, *JHEP* **12** (2020) 115 [arXiv:2004.04240] [INSPIRE].

[19] M. Carena, I. Low, N.R. Shah and C.E.M. Wagner, *Impersonating the Standard Model Higgs Boson: Alignment without Decoupling*, *JHEP* **04** (2014) 015 [arXiv:1310.2248] [INSPIRE].

[20] M. Carena et al., *Complementarity between Nonstandard Higgs Boson Searches and Precision Higgs Boson Measurements in the MSSM*, *Phys. Rev. D* **91** (2015) 035003 [arXiv:1410.4969] [INSPIRE].

[21] M. Carena et al., *Alignment limit of the NMSSM Higgs sector*, *Phys. Rev. D* **93** (2016) 035013 [arXiv:1510.09137] [INSPIRE].

[22] I. Low, N.R. Shah and X.-P. Wang, *Higgs alignment and novel CP-violating observables in two-Higgs-doublet models*, *Phys. Rev. D* **105** (2022) 035009 [arXiv:2012.00773] [INSPIRE].

[23] T.-K. Chen, C.-W. Chiang and I. Low, *Simple model of dark matter and CP violation*, *Phys. Rev. D* **105** (2022) 075025 [arXiv:2202.02954] [INSPIRE].

[24] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[25] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244 [arXiv:1207.1303] [INSPIRE].

[26] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].

[27] DELPHES 3 collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [arXiv:1307.6346] [INSPIRE].

[28] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[29] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[30] D. Eriksson, J. Rathsman and O. Stål, *2HDMC: Two-Higgs-Doublet Model Calculator Physics and Manual*, *Comput. Phys. Commun.* **181** (2010) 189 [arXiv:0902.0851] [INSPIRE].

[31] P. Bechtle et al., *HiggsBounds: Confronting Arbitrary Higgs Sectors with Exclusion Bounds from LEP and the Tevatron*, *Comput. Phys. Commun.* **181** (2010) 138 [arXiv:0811.4169] [INSPIRE].

[32] P. Bechtle et al., *HiggsBounds 2.0.0: Confronting Neutral and Charged Higgs Sector Predictions with Exclusion Bounds from LEP and the Tevatron*, *Comput. Phys. Commun.* **182** (2011) 2605 [arXiv:1102.1898] [INSPIRE].

[33] P. Bechtle et al., *Recent Developments in HiggsBounds and a Preview of HiggsSignals*, *PoS* **CHARGED2012** (2012) 024 [arXiv:1301.2345] [INSPIRE].

[34] P. Bechtle et al., HiggsBounds – 4*: Improved Tests of Extended Higgs Sectors against Exclusion Bounds from LEP, the Tevatron and the LHC*, *Eur. Phys. J. C* **74** (2014) 2693 [arXiv:1311.0055] [INSPIRE].

[35] P. Bechtle et al., *Applying Exclusion Likelihoods from LHC Searches to Extended Higgs Sectors*, *Eur. Phys. J. C* **75** (2015) 421 [arXiv:1507.06706] [INSPIRE].

[36] O. Stål and T. Stefaniak, *Constraining extended Higgs sectors with HiggsSignals*, *PoS* **EPS-HEP2013** (2013) 314 [arXiv:1310.4039] [INSPIRE].

JHEP09(2024)139

[37] P. Bechtle et al., *HiggsSignals: Confronting arbitrary Higgs sectors with measurements at the Tevatron and the LHC*, *Eur. Phys. J. C* **74** (2014) 2711 [arXiv:1305.1933] [INSPIRE].

[38] P. Bechtle et al., *Probing the Standard Model with Higgs signal rates from the Tevatron, the LHC and a future ILC*, *JHEP* **11** (2014) 039 [arXiv:1403.1582] [INSPIRE].

[39] P. Bechtle et al., *HiggsSignals-2: Probing new physics with precision Higgs measurements in the LHC 13 TeV era*, *Eur. Phys. J. C* **81** (2021) 145 [arXiv:2012.09197] [INSPIRE].

[40] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, *JHEP* **03** (2013) 015 [arXiv:1212.3460] [INSPIRE].

[41] ATLAS collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, Tech. Rep. ATL-PHYS-PUB-2016-012, CERN, Geneva (2016).

[42] ATLAS collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, 2016 *JINST* **11** P04008 [arXiv:1512.01094] [INSPIRE].

[43] ATLAS collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, *Eur. Phys. J. C* **83** (2023) 681 [arXiv:2211.16345] [INSPIRE].

[44] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*, arXiv:1603.04467.

[45] T. Akiba et al., *Optuna: A Next-generation Hyperparameter Optimization Framework*, arXiv:1907.10902 [INSPIRE].

[46] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *J. Machine Learning Res.* **12** (2011) 2825 [arXiv:1201.0490] [INSPIRE].

[47] L. Heinrich, M. Feickert and G. Stark, *pyhf: v0.7.3*, DOI:10.5281/zenodo.1169739.

[48] L. Heinrich, M. Feickert, G. Stark and K. Cranmer, *pyhf: pure-Python implementation of HistFactory statistical models*, *J. Open Source Softw.* **6** (2021) 2823 [INSPIRE].

[49] A.L. Read, *Presentation of search results: The CL_s technique*, *J. Phys. G* **28** (2002) 2693 [INSPIRE].

[50] V. Barger et al., *New physics in resonant production of Higgs boson pairs*, *Phys. Rev. Lett.* **114** (2015) 011801 [arXiv:1408.0003] [INSPIRE].

[51] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121 [arXiv:1810.05165] [INSPIRE].

[52] A. Dobrescu, M.V. Giuffrida and S.A. Tsaftaris, *Dataset for "Deep Learning to Improve the Sensitivity of Di-Higgs Searches in the 4b Channel"*, DOI:10.5281/zenodo.10952296.

[53] F.-Y. Hsieh, *hh4b Analysis with SPANet*, GitHub (2024).

JHEP09(2024)139