Searching for gluon quartic gauge couplings at muon colliders using the autoencoder

Yu-Ting Zhang[†] and Xin-Tong Wang[‡]
Department of Physics, Liaoning Normal University, Dalian 116029, China

Ji-Chong Yang[*]
Department of Physics, Liaoning Normal University, Dalian 116029, China
and Center for Theoretical and Experimental High Energy Physics,
Liaoning Normal University, Dalian 116029, China

(Received 8 January 2024; accepted 1 May 2024; published 20 May 2024)

One of the difficulties one has to face in the future phenomenological studies of new physics (NP), is the need to deal with increasing amounts of data. It is therefore increasingly important to improve the efficiency in the phenomenological study of NP. Whether it is the use of the Standard Model effective field theory (SMEFT), the use of machine learning (ML) algorithms, or the use of quantum computing, all are means of improving the efficiency. In this paper, we use a ML algorithm, the autoencoder (AE), to study the dimension-8 operators in the SMEFT which contribute to the gluon quartic gauge couplings (gQGCs) at muon colliders. The AE is one of the ML algorithms that has the potential to be accelerated by the quantum computing. It is found that the AE-based anomaly detection algorithm can be used as event selection strategy to study the gQGCs at the muon colliders, and is effective compared with traditional event selection strategies.

DOI: 10.1103/PhysRevD.109.095028

I. INTRODUCTION

Supported by the large amount of experimental evidences, it can be concluded that the Standard Model (SM) is able to describe and explain the vast majority of phenomena in particle physics, with a few rare exceptions. These exceptions include experimental results such as the neutrino mass [1–3], the W-boson mass problem [4,5], the muon $g-2$ problem [6–8], and more [9]. Besides, the SM cannot describe dark matter, gravity, etc. As a result, the existence of new physics (NP) beyond the SM has been widely believed, and the search of NP as well as precision measurements have been at the forefront of interest in the high-energy physics (HEP) community [10].

Both the search for NP and precision measurements require dealing with a large number of events. With more and more data to be processed, more efficient ways to search for NP are called for. One of the reasons why the SM effective field theory (SMEFT) [11–14] has been widely used in the phenomenological study of NP in recent years is that SMEFT has an outstanding advantage of searching for NP signals with high efficiency. In the SMEFT, the NP particles are integrated out, and the NP effects become new interactions of known particles. Formally, the new interactions appear as higher-dimension operators with Wilson coefficients suppressed by powers of a NP scale Λ. The operators that are most likely to be found correspond to the operators with Wilson coefficients that are least suppressed by Λ. The high efficiency of searching for NP using SMEFT is demonstrated by the fact that, instead of dealing with various NP models, the number of operators to be considered at a specific order of Λ in the SMEFT is finite. Not only that, but if an operator is not found, then multiple NP models contributing to this operator will also be constrained. However, as the importance of the dimension-8 operators in the SMEFT has been realized [15–18], more and more phenomenological studies have been devoted to the dimension-8 operators in recent years [19–45]. For one generation of fermions, there are 895 baryon number conserving dimension-8 operators [46,47], and there are even more operators if one considers operators with dimension more than eight. A procedure to select the events which does not rely on operators to be searched for can further improve efficiency.

[*]Corresponding author: yangjichong@lnnu.edu.cn
[†]976982345@qq.com
[‡]2226769965@qq.com

Machine learning (ML) algorithms are one of the ways to process the data efficiently. ML is a multidisciplinary cross-discipline for studying how computers can mimic and implement human learning behaviors in order to acquire new knowledge, and has already been used in HEP studies [48–57]. ML algorithms have an advantage in processing complex data, and one of its common applications is anomaly detection (AD). When using AD to search for NP models, its implementation is often independent of the NP model to be searched for. While hyperparameters in ML algorithms will often be NP model-dependent, the procedure of tuning of hyperparameters is usually common for different NP models. Based on this, an event selection strategy utilizing AD can be viewed as an automated event selection strategy. The use of AD algorithms in phenomenological studies is a hot topic in recent years [58–67].

Meanwhile, quantum computing is another effective way to deal with large amounts of data. Many ML algorithms can be implemented by quantum computing [68–70], an example of which is the autoencoder (AE) algorithm [71,72]. AE is an unsupervised learning dimensionality reduction algorithm using artificial neural networks (ANN), which is at the same time capable of AD. Therefore, it can be expected that AE can be used to detect NPs, and AE has already been used in phenomenological studies in HEP [65,66,73,74]. Analogous to the principal component analysis (PCA) algorithm, the most common scenario for AE is data dimensionality reduction. While the PCA is a linear dimensionality reduction by solving the feature vector, the AE algorithm is a nonlinear dimensionality reduction. It is verified that, AD based on PCA is able to search for NP [64], therefore it can be expected that a similar approach works also for AE. Similar to PCA, AE has the potential for quantum acceleration [75–78], and thus AE holds great promise for processing large amounts of data. To verify the feasibility of the AE algorithm in the search of high dimensional operators in the SMEFT, we use an AE algorithm-based AD (AEAD) to study the gluon quartic gauge couplings (gQGCs) [79,80] in this paper. In the SMEFT, operators contributing to gQGCs start from dimension-8, therefore we concentrate on those dimension-8 operators.

As an arena, the processes affected by gQGCs at muon colliders are considered. The muon colliders have been hotly discussed in recent years for searching for NP signals [81–91]. A muon collider has the advantage of being able to explore both high-luminosity and high-energy frontiers, while at the same time being less affected by the QCD background as a lepton collider. On the one hand, higher collision energies are better if one is committed to studying dimension-8 or even higher-dimension operators, and on the other hand, higher luminosities require more efficient means of processing data. Thus, the processes at muon colliders that are affected by gQGCs are both worth studying and suitable for exploring the AEAD algorithm. For comparison with a conventional event selection strategy, in this paper the process $\mu^+\mu^- \to jj\nu\bar{\nu}$ is studied which has been studied in Ref. [92]. It has been shown that this process is sensitive to the gQGCs.

The rest of the paper is organized as follows. In Sec. II, the dimension-8 operators contributing to the gQGCs are introduced. In Sec. III, the event selection strategy based on AEAD algorithm is presented. In Sec. IV, the numerical results and expected constraints on the operator coefficients are presented. Section V is a summary.

II. DIMENSION-8 OPERATORS CONTRIBUTING TO THE GQGCS

The gQGCs arise from the Born-Infeld (BI) extension of the SM, which was originally motivated by the idea that there should be an upper limit on the strength of the electromagnetic field [93]. It has been shown that, the BI model is also related to the M-theory inspired models [79,80,94–96]. In the SMEFT, the operators contributing to gQGCs appear at dimension-8,

$$
\begin{aligned}
O_{gT,0} &= \frac{1}{16M_0^4}\sum_a G^a_{\mu\nu}G^{a,\mu\nu} \times \sum_i W^i_{\alpha\beta}W^{i,\alpha\beta},\\
O_{gT,1} &= \frac{1}{16M_1^4}\sum_a G^a_{\alpha\nu}G^{a,\mu\beta} \times \sum_i W^i_{\mu\beta}W^{i,\alpha\nu},\\
O_{gT,2} &= \frac{1}{16M_2^4}\sum_a G^a_{\alpha\mu}G^{a,\mu\beta} \times \sum_i W^i_{\nu\beta}W^{i,\alpha\nu},\\
O_{gT,3} &= \frac{1}{16M_3^4}\sum_a G^a_{\alpha\mu}G^a_{\beta\nu} \times \sum_i W^{i,\mu\beta}W^{i,\nu\alpha},\\
O_{gT,4} &= \frac{1}{16M_4^4}\sum_a G^a_{\mu\nu}G^{a,\mu\nu} \times B_{\alpha\beta}B^{\alpha\beta},\\
O_{gT,5} &= \frac{1}{16M_5^4}\sum_a G^a_{\alpha\nu}G^{a,\mu\beta} \times B_{\mu\beta}B^{\alpha\nu},\\
O_{gT,6} &= \frac{1}{16M_6^4}\sum_a G^a_{\alpha\mu}G^{a,\mu\beta} \times B_{\nu\beta}B^{\alpha\nu},\\
O_{gT,7} &= \frac{1}{16M_7^4}\sum_a G^a_{\alpha\mu}G^a_{\beta\nu} \times B^{\mu\beta}B^{\nu\alpha},
\end{aligned}
\tag{1}
$$

where $G^a_{\mu\nu}$ are gluon field strengths, $W^i_{\mu\nu}$ and $B_{\mu\nu}$ denote electroweak field strengths, and M_i are mass scales associated with NP particles. For convenience, we define $f_i \equiv 1/(16M_i^4)$. The expected constraints on M_i at the Large Hadron Collider (LHC) with the center-of-mass (c.m.) energy $\sqrt{s} = 13$ TeV, and luminosity $L = 36.7\ \text{fb}^{-1}$ obtained by using the process $gg \to \gamma\gamma$ are shown in Table I. The combined sensitivities of the $Z\gamma$ and $\gamma\gamma$ channels at the LHC with $\sqrt{s} = 13$ TeV and $L = 137\ \text{fb}^{-1}$ [80] are about three times of the ones shown in Table I.

TABLE I. The constraints on the operator coefficients at 95% CL obtained by the process $gg \to \gamma\gamma$ at the LHC with $\sqrt{s} = 13$ TeV and $L = 36.7$ fb^{-1} [79].

$M_0 \geq 1040$ GeV	$M_1 \geq 777$ GeV
$M_2 \geq 750$ GeV	$M_3 \geq 709$ GeV
$M_4 \geq 1399$ GeV	$M_5 \geq 1046$ GeV
$M_6 \geq 1010$ GeV	$M_7 \geq 954$ GeV

The process $\mu^+\mu^- \to jj\nu\bar{\nu}$ at muon colliders can also be affected by the gQGCs [92]. Different from the case of a hadron collider that the operators are classified into four pairs with same Lorentz structures in phenomenological studies, at the muon colliders, the pairs can be decoupled. In particular, the process $\mu^+\mu^- \to jj\nu\bar{\nu}$ can be affected by the vector boson scattering (VBS) subprocess $W^+W^- \to gg$, which is associated with only $O_{gT,0,1,2,3}$ operators. The Feynman diagrams are shown in Fig. 1. Since the VBS contribution is logarithmically enhanced at large c.m. energies compared with the triboson process, we concentrate on the $O_{gT,0,1,2,3}$ operators. It has been shown that, at muon colliders the process $\mu^+\mu^- \to jj\nu\bar{\nu}$ can archive a competitive sensitivity compared with the hadron colliders [92].

In the process $\mu^+\mu^- \to jj\nu\bar{\nu}$, there is no interference between the SM and gQGCs, which simplifies the procedure to obtain the expected constraints. However, there are two (anti-)neutrinos in the final state. This usually results in some loss of information, which in turn affects the efficiency of the event selection strategy. This just provides a place to test whether the AEAD algorithm is effective or not.

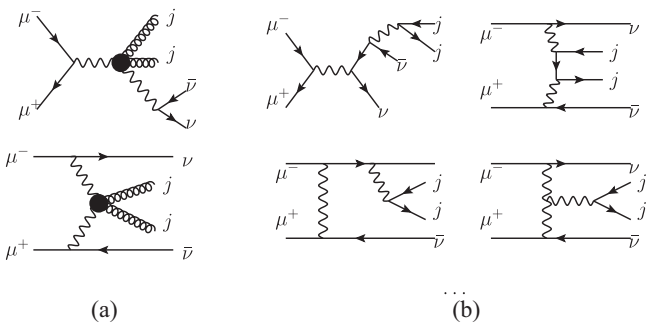

FIG. 1. Feynman diagrams for the process $\mu^+\mu^- \to jj\nu\bar{\nu}$, where (a) shows the Feynman diagrams of the gQGC contribution, and (b) shows the typical Feynman diagrams of the SM background.

III. AUTOENCODER ANOMALY DETECTION

A. A brief introduction of the autoencoder algorithm

AE is a type of ANN, which can be used in various applications, including data compression, feature extraction, denoising, data generation, etc. The structure of an AE

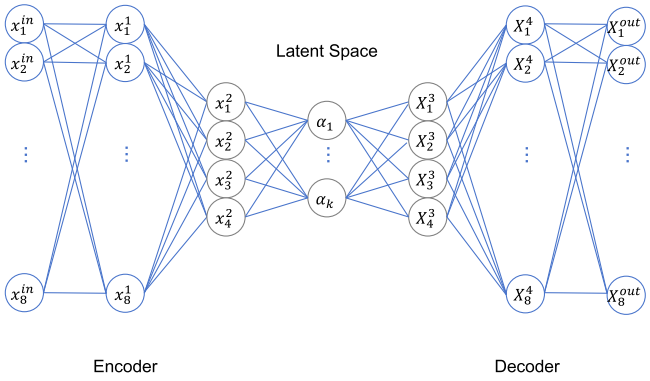

FIG. 2. The graphical representation of the AE. The AE network can be decomposed into two parts, where the encoder is consist of the $x^{\text{in},1,2}$ and the latent layers, the decoder is consist of the latent, $X^{3,4,\text{out}}$ layers. In this paper, we use a dense connected network.

is shown in Fig. 2, which is primarily composed of two parts, the encoder and the decoder. Both the encoder and the decoder can consist of multiple layers, with the neurons in the input layer denoted as x_i^{in}, and those in the output layer denoted as X_i^{out}. The data input to the input neurons and the data obtained from the output neurons are also notated as x_i^{in} and X_i^{out}, respectively. The number of neurons in the input layer is as same as the one in the output layer and the dimension of the input vector (denoted as l). In this paper, $l = 8$. The goal to train an AE is to reconstruct the input data, i.e., the labels of the training data are just the input data, therefore the training of an AE is unsupervised learning. The training objective is to minimize the reconstruction error between the $\vec{x}^{\text{in}}$ and $\vec{X}^{\text{out}}$, aiming to obtain as similar a set of vectors from the output layer as possible after inputting a set of vectors into the input layer.

A well-trained AE can reproduce the input by utilizing a few variables, $\alpha_{i=1,\ldots,k}$ with $k < l$, along with the decoder network. Being able to reconstruct the input data indicates that, the information in α_i is enough to describe the data with the help of the decoder. Also, this means that the encoder is able to compress the information in the input data into k numbers, α_i (which is often called the latent space), i.e., data dimensionality reduction.

The reason an AE can achieve data dimensionality reduction lies in the fact that, the features of the input data are not independent. The encoder learns the relationships among the features and compresses them into α_i variables, which can then be used by the decoder to reconstruct the $\vec{X}^{\text{out}}$. This mechanism of AE can be used for AD. It is well-known that, a challenge for ML methods is the ability to extrapolate into unknown phase-space regions, AEAD just turned the disadvantage in extrapolating into a feature, i.e., the artificial intelligence trained using the SM cannot recognize NP and will treat NP as anomalies.

At first, AE is trained on the SM events. Since the events are generated according to the physical laws of the SM, the relationships between the features learned by the AE are also related to the physical laws of the SM shaded behind them. At the same time, it can be expected that the reconstruction of the events generated according to the physical laws of the NP will be less accurate than that of the SM, because the AE has not learned the relationships between the features of the events of the NP. The mean-squared error (mse) defined as $d = \sum_n^N (\vec{x}^{\text{in},n} - \vec{X}^{\text{out},n})^2/N$ can be used to represent the reconstruction error, where N is the number of events. In AEAD, d can also been used as an anomaly score. It can be expected that d is larger for the NP events compared with those of the SM events, therefore can be used to select the NP events.

B. The event selection strategy based on AEAD

Following the above idea, the AEAD event selection strategy can be summarized as follows:

(i) Generate the training and validation datasets using Monte Carlo (MC) simulation, which consist of the events from the SM;
(ii) Train AE, and use the validation dataset to avoid overfitting;
(iii) For a test dataset, which can be either obtained from the experiment or from MC simulation, calculate d for each event;
(iv) Use a threshold d_{th} to select events, i.e. select the events with $d > d_{\text{th}}$, where d_{th} can be tuned according to the signal significance or expected constraints on parameters of NP.

Note that, although the AE is unsupervised ML algorithm, in AEAD the SM dataset is used in the training phase, therefore the AEAD is no longer unsupervised. However, the NP datasets are not used in the training phase. The test dataset can be from the experiment, and there is no need to know whether it contains NP or what kind of NP it might contain. It can be expected that, the signal of NP can be traced by the AEAD whenever the test dataset contains events that differ from the law of the SM.

IV. NUMERICAL RESULTS

A. Data preparation

To compare with the traditional event selection strategy, in this paper, we use the same events generated as in Ref. [92]. The events are generated using MC simulation with the MadGraph5_aMC@NLO toolkit [97–99], where the standard cuts are set as the default. The parton shower is applied using PYTHIA8 [100] with default settings. A fast detector simulation is performed using DELPHES [101] with the muon collider card. The data cleaning and preparation phase is applied using MLAnalysis [102]. The ANN is constructed and trained using the KERAS with the TensorFlow backend [103]. The events for the NP are generated with one operator at a time. The operator coefficients are set as the same as Ref. [92]. As an EFT, the SMEFT is only valid under a certain energy scale. One of the signals that the SMEFT is no longer valid is the violation of the unitarity [104–106], and the partial wave unitarity is often used in the phenomenological studies of the SMEFT to check whether the SMEFT is valid, which can sets bounds on the operator coefficients [107–113]. The operator coefficients used in the MC simulation are within the constraints set by the partial wave unitarity bounds [92]. The partial wave unitarity bounds (denoted as f_i^U) are listed in Table II. Denoting σ_{SM} as the cross section of the SM contribution, and $\sigma_{gT,i}(\tilde{f}_i)$ as NP contributions with the operator coefficients to be $\tilde{f}_i$, respectively, the cross sections and operator coefficients $\tilde{f}_i$ are listed in Table II.

TABLE II. The cross sections of the SM contribution and the NP contributions [92]. The NP contributions are cross sections when the operator coefficients are $\tilde{f}_i, \tilde{f}_i$ used in the simulation are also shown. The cross sections after N_j cut are denoted as $\hat{\sigma}$, which are also shown. The f_i^U are partial wave unitarity bounds on the operator coefficients.

$\sqrt{s}$ (TeV)	3	10	14	30
σ_{SM} (fb)	868.8	1454.8	1608.7	1898.8
$\hat{\sigma}_{\text{SM}}$ (fb)	722.9	1191.9	1315.5	1542.4
f_0^U (TeV^{-4})	3.5	0.028	0.0074	0.00035
$\tilde{f}_0$ (TeV^{-4})	1	0.012	0.004	0.00035
$\sigma_{gT,0}(\tilde{f}_0)$ (fb)	3.21	1.15	1.12	1.14
$\hat{\sigma}_{gT,0}(\tilde{f}_0)$ (fb)	3.21	1.15	1.12	1.14
f_1^U (TeV^{-4})	10.5	0.085	0.022	0.001
$\tilde{f}_1$ (TeV^{-4})	1.5	0.02	0.007	0.0006
$\sigma_{gT,1}(\tilde{f}_1)$ (fb)	3.55	1.38	1.44	1.34
$\hat{\sigma}_{gT,1}(\tilde{f}_1)$ (fb)	3.52	1.37	1.43	1.33
$f_{2,3}^U$ (TeV^{-4})	14.0	0.114	0.030	0.004
$\tilde{f}_{2,3}$ (TeV^{-4})	3	0.03	0.012	0.0012
$\sigma_{gT,2}(\tilde{f}_2)$ (fb)	3.83	0.956	1.34	1.78
$\hat{\sigma}_{gT,2}(\tilde{f}_2)$ (fb)	3.82	0.955	1.34	1.78
$\sigma_{gT,3}(\tilde{f}_3)$ (fb)	4.12	0.924	1.27	1.61
$\hat{\sigma}_{gT,3}(\tilde{f}_3)$ (fb)	4.10	0.921	1.26	1.61

In DELPHES we use the fast jet finder with anti-k_T algorithm, and with $R = 0.5$, $p_T^j > 20$ GeV, where R is cone radius, and p_T^j are the transverse momenta of jets. After the events are generated, we require that the final states to have at least two jets. This requirement is denoted as the N_j cut, the cross sections after N_j cut are denoted as $\hat{\sigma}_{\text{SM}}$ and $\hat{\sigma}_{gT,i}(\tilde{f}_i)$ which are also listed in Table II. Then an eight dimensional vector is made to represent each event such that $\vec{v} = (p_t^{(1)}, p_x^{(1)}, p_y^{(1)}, p_z^{(1)}, p_t^{(2)}, p_x^{(2)}, p_y^{(2)}, p_z^{(2)})$,

TABLE III. The means $\bar{v}_i$ and standard deviations ϵ_i of the components of the vectors $\vec{v}$ over the training datasets.

$\sqrt{s}$ TeV	$\bar{v}_1$ GeV	$\bar{v}_2$ GeV	$\bar{v}_3$ GeV	$\bar{v}_4$ GeV
3	271.00	−0.14	−0.19	0.48
10	576.24	−0.01	0.14	3.00
14	709.14	−0.21	−0.01	0.57
30	1053.94	0.15	0.01	4.64
$\sqrt{s}$ TeV	$\bar{v}_5$ GeV	$\bar{v}_6$ GeV	$\bar{v}_7$ GeV	$\bar{v}_8$ GeV
3	110.13	0.06	0.04	0.33
10	225.00	−0.04	0.08	1.45
14	275.42	−0.02	−0.03	0.03
30	398.96	−0.10	−0.13	1.74
$\sqrt{s}$ TeV	ϵ_1 GeV	ϵ_2 GeV	ϵ_3 GeV	ϵ_4 GeV
3	226.27	94.94	94.65	326.25
10	665.88	109.03	108.73	886.88
14	879.40	111.35	112.47	1118.43
30	1504.96	120.43	122.45	1829.17
$\sqrt{s}$ TeV	ϵ_5 GeV	ϵ_6 GeV	ϵ_7 GeV	ϵ_8 GeV
3	99.21	42.48	42.84	135.12
10	284.41	48.77	48.20	355.99
14	371.30	50.10	52.11	456.52
30	603.26	56.48	56.66	718.74

where $p^{(1)}$ and $p^{(2)}$ are the four momenta of hardest and second hardest jets. Note that $p_t^{(1)}$ and $p_t^{(2)}$ are the energies of jets in this paper. In the following, we consider only the 8-dimensional vectors described above, ignoring the physical meaning behind them.

For the SM contribution, we generate 1,000,000 events for each c.m. energy, taking the case of $\sqrt{s} = 3$ TeV as an example, 832,056 events are left after the N_j cut, of which 400,000 events consist the training dataset, 100,000 events consist the validation dataset, and rest events consist the test dataset. For each operator, 300,000 events are generated, and all the events after the N_j cut are used as the test dataset. The numbers of events after the N_j cut for the NP are generally more than 297,000. Before the datasets are fed with the AE, a z-score standardization [114] is applied, i.e., the vectors $\vec{v}^n$ are replaced by $\vec{x}^{\text{in},n}$ such that $x_i^{\text{in},n} = (v_i^n - \bar{v}_i)/\epsilon_i$, where $x_i^{\text{in},n}, v_i^n$ are the ith components of the $\vec{x}^{\text{in},n}$ and $\vec{v}^n$ vectors, $\bar{v}_i$ and ϵ_i are the mean value and standard deviation of the ith component over the training dataset. $\bar{v}_i$ and ϵ_i used in this paper are listed in Table III.

The AE is to reproduce the input vectors, therefore the labels of the training datasets are just as same as the input of the datasets.

B. Structure of the AE network

Since the events are represented by eight dimensional vectors, the number of neurons in the input layer of the encoder, and in the output layer of the decoder are both eight. The number of layers and the number of neurons in each layer is depicted in Fig. 2. In this paper, we use a densely connected ANN, taking the hidden layer x_i^1 as an example, values at neural x_i^1 can be calculated as $\vec{x}^1 = g^{\text{in},1}(W^{\text{in},1}\vec{x}^{\text{in}} + \vec{b}^{\text{in}})$, where $W^{i,j}$ is the weight matrix stored in the links between ith and jth layers, and $\vec{b}^i$ is the basis vector stored in the neurons in ith layer, and $g^{i,j}$ is the activation function between ith and jth layers. In this paper, we use the "LeakyReLU" function [115] between the layers,

$$g(x) = \begin{cases} x, & x > 0, \\ \alpha x, & x \leq 0, \end{cases} \tag{2}$$

where $\alpha = 0.01$. Specifically, $g^{\text{in},1}, g^{1,2}$, $g^{\text{latent},3}$, and $g^{3,4}$ are LeakyReLU functions, $g^{2,\text{latent}}$ and $g^{4,\text{out}}$ are linear activation functions. The loss function defines how well the input vectors can be reproduced by the AE. In this paper, we use the mse as the loss function.

Since it is expected that, the reason AE to be able to distinguish between the SM and NP is because the AE is able to find the patterns of the SM while being unable to learn the patterns of NP. Therefore we need the AE to have weak generalization properties and only reproduce events of the SM accurately, and smaller k the better performance of AEAD is expected. To investigate the effect of the dimension of the latent space, four cases are considered, they are $k = 1$, 2, 3, and 4.

C. Early stopping

Overfitting is a situation where the model performs well on the training set but relatively poorly on the test set. In that case, the model is weak in predicting unknown data. One of the methods to avoid overfitting is early stop.

The process of training the model is the process of updating the model parameters (i.e., $W^{i,j}$ and $\vec{b}^i$) through learning. In the training phase, the dataset is divided into a training dataset and a validation dataset, and only the training dataset is used to update the parameters. During the training process, the errors for the training set and validation set gradually decreases, and after reaching a critical point, the errors for the training set continue to decrease and the ones for the validation set start to increase. Early stopping is to prevent overfitting by stopping the training before the critical point, i.e., the number of iterations is truncated.

We use early stopping method to avoid overfitting in this paper. As an example, the mses for the training dataset and the validation dataset as functions of number of epochs for $k = 1$ are shown in Fig. 3. We stop when the mean of mse

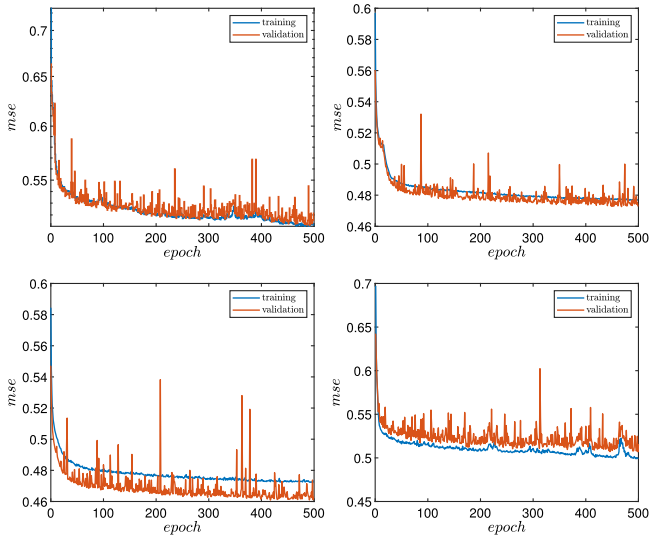

FIG. 3. The learning curves in the training phase for $k = 1$. The top-left panel corresponds to $\sqrt{s} = 3$ TeV, the top-right panel corresponds to $\sqrt{s} = 10$ TeV, the bottom-left panel corresponds to $\sqrt{s} = 14$ TeV, and the bottom-right panel corresponds to $\sqrt{s} = 30$ TeV.

of last 50 epochs of the validation set is larger than then the one of the last 100 epochs, which is checked every 100 epochs. In the training, we preserve the networks by training to 500 epochs while keeping only the one with the smallest mses for validation datasets. So when mse of the validation dataset plateaus early, or starts to rise, the result with the smallest mse will not be missed.

It has been noticed that the mse of the validation dataset is smaller than the one of the training dataset, for example the case at $\sqrt{s} = 14$ TeV in Fig. 3. It is found that whether or not validation is better than training occurs mainly depends on the segmentation of the training and validation datasets. Meanwhile, if one looks at the mean absolute error (mae) of the training and validation sets, which is defined as $\sum_n^N \sum_i^8 |x_i^{\text{in},n} - X_i^{\text{out},n}|/N$, one will find that the two are always close to each other, for example as the case at $\sqrt{s} =$ 14 TeV shown in Fig. 4. Since mse is more sensitive to anomalous samples, it can inferred that, the above phenomenon is due to the fact that there are events in the SM dataset that are more anomalous relative to the rest of the SM events, and that where they are partitioned determines the mse. Normally, this is a situation more appropriate to use mae as a loss function, however, since mse is more sensitive to anomalous samples and our goal is not to reproduce the SM events but to find anomalies, we still use mse as a loss function.

D. Distribution of the anomaly score

As described in the previous section, the AEAD uses how well the AE can reproduce the input as the anomaly score. That is, one can use d as the anomaly score. Note that

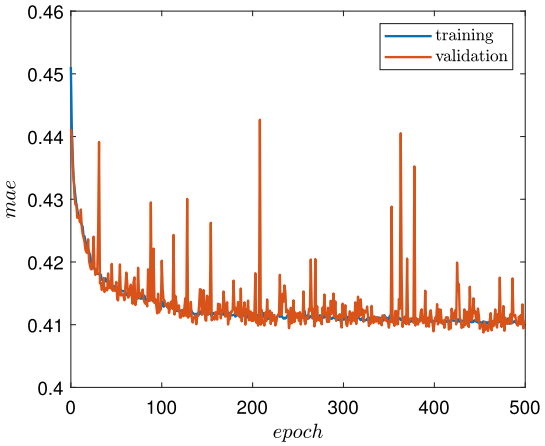

FIG. 4. The mean absolute errors of training and validation datasets at $\sqrt{s} = 14$ TeV.

d is defined using the data after z-score standardization, therefore is dimensionless.

In this subsection and in the following, we concentrate on the test datasets. As an example, the normalized distributions of d when $k = 1$ are shown in Fig. 5. It can be seen that the d for the SM background at different energies are generally small, thanks to the well trained AE. Meanwhile, the d for the NP signals are generally larger, and the larger the $\sqrt{s}$, the larger d. From the distributions it can be conclude that d can provide a good discriminate ability to select the NP signals as expected.

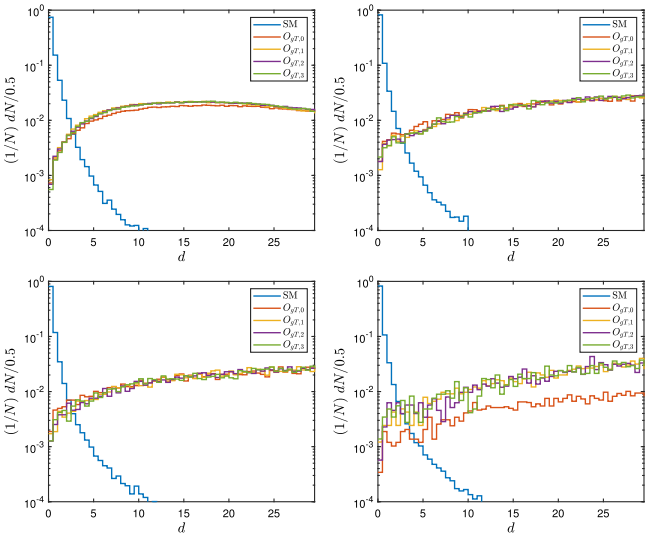

FIG. 5. The normalized distributions of d when $k = 1$ for the SM events and for the NP events in the test datasets. The top-left panel corresponds to $\sqrt{s} = 3$ TeV, the top-right panel corresponds to $\sqrt{s} = 10$ TeV, the bottom-left panel corresponds to $\sqrt{s} = 14$ TeV, and the bottom-right panel corresponds to $\sqrt{s} = 30$ TeV.

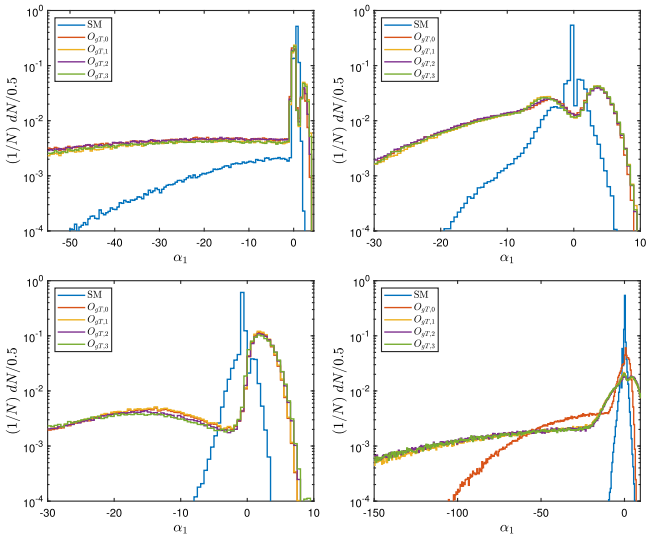

FIG. 6. The distributions of the test dataset events in the latent space when $k=1$. The top-left panel corresponds to $\sqrt{s}=3$ TeV, the top-right panel corresponds to $\sqrt{s}=10$ TeV, the bottom-left panel corresponds to $\sqrt{s}=14$ TeV, and the bottom-right panel corresponds to $\sqrt{s}=30$ TeV.

E. Latent space distribution

It is known that AE can also be used as a data dimensionality reduction algorithm, a scheme that uses AE for classification uses AE as a data preparation stage. In combination with AE, other classification algorithms or AD algorithms can be applied in the latent space, i.e., the space consisting of the α_i values that are in the middle layer. Since the AE is trained to approximately reproduce events of the SM, this means that α_i contain the major information needed to reconstruct the events. For this to happen, there are hidden relationships between the components of the vectors input to the AE, as a result, the components of the input vectors are not independent of each other. After dimension reduction, the events can be represented by a smaller number (k in our case) of variables. Therefore, with the help of latent space, the features of the SM and NP are more easily presented visually.

The distributions of the events in the latent spaces when $k=1$ are shown in Fig. 6, and the case for $k=2$ are shown in Fig. 7. It can be seen that, in the latent spaces, the NP events already distribute differently from the SM events. Due to the nonlinearity of the activation functions (segmented functions are used as the activation functions), different regions in the latent space often represent different functions that have been imposed to α_i in the decoder, and therefore represent different relationships among the components of the input/output vectors. The fact that the SM and NP events distributed differently in the latent space reflects the conjecture that, since the hidden relationships between the components are obtained by having the AE trained on the SM events, the events of the NP are not described by these relationships. This explains why d can be used to search for the NP signals.

F. Cut efficiency

Since there is no interference between the SM and NP, the cross section after cut can be written as

$$\sigma(f_i)=\varepsilon_{\rm SM}\hat{\sigma}_{\rm SM}+\varepsilon_{O_{gT,i}}\frac{f_i^2}{\tilde{f}_i^2}\hat{\sigma}_{gT,i}(\tilde{f}_i), \tag{3}$$

where $\hat{\sigma}_{\rm SM}$ and $\hat{\sigma}_{gT,i}(\tilde{f}_i)$ are listed in Table II, $\varepsilon_{\rm SM}$ is the cut efficiency of selecting events with $d>d_{\rm th}$ for the SM, and

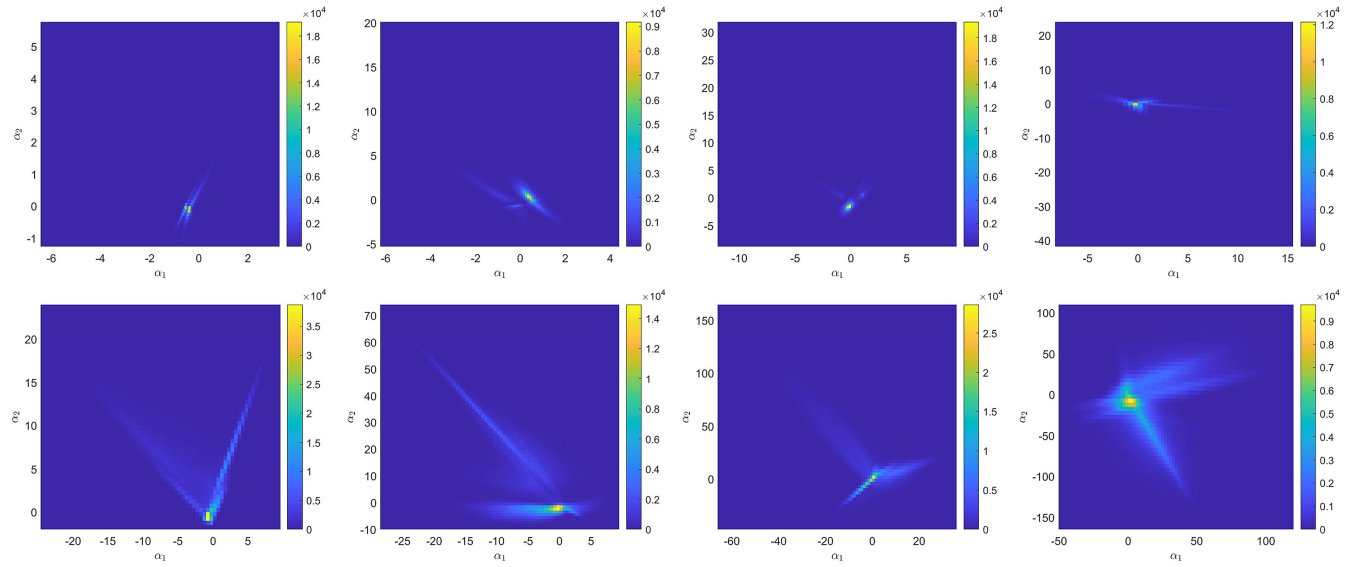

FIG. 7. Same as Fig. 6 but for $k=2$. The panels in the first row are for the SM test dataset, and those in the second row are for the $O_{gT,0}$ test datasets. The first, second, third and the fourth columns correspond to $\sqrt{s}=3$ TeV, 10 TeV, 14 TeV, and 30 TeV, respectively.

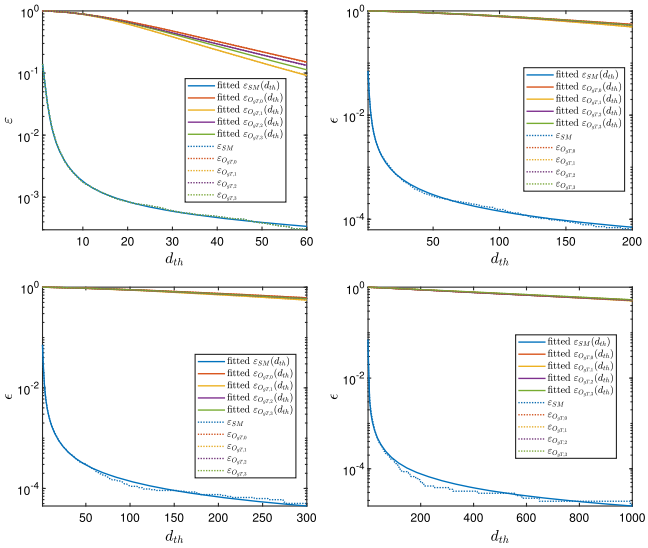

FIG. 8. ε as functions of $d_{\rm th}$ compared with the fitted $\varepsilon(d_{\rm th})$. The top-left panel corresponds to $\sqrt{s}=3$ TeV, the top-right panel corresponds to 10 TeV, the bottom-left panel corresponds to 14 TeV, and the bottom-right panel corresponds to 30 TeV.

$\varepsilon_{O_{gT,i}}$ are cut efficiencies for the NP signals. Note that, ε does not include the effect of the N_j cut (denoted as ε_{N_j}), which is already included in $\hat{\sigma}=\varepsilon_{N_j}\sigma$. $\varepsilon_{\rm SM}$ and $\varepsilon_{O_{gT,i}}$ are functions of $d_{\rm th}$, and to facilitate the study of expected coefficient constraints, we fit the cut efficiencies using rational functions,

$$\varepsilon_{\rm SM}(d_{\rm th})=\frac{1+a_1 d_{\rm th}+a_5 d_{\rm th}^2}{a_2+a_3 d_{\rm th}+a_4 d_{\rm th}^2+a_6 d_{\rm th}^3},$$
$$\varepsilon(d_{\rm th})=\frac{1+a_1 d_{\rm th}}{a_2+a_3 d_{\rm th}+a_4 d_{\rm th}^2}. \quad (4)$$

For the case of $\varepsilon_{\rm SM}$ when $k=1$ and $\sqrt{s}=3$ TeV, we use a rational function with six parameters, and for the other cases we use a rational function with four parameters.

As an example, the results of the fits in the case of $k=1$ are shown in Fig. 8. It can be seen that, $\varepsilon_{\rm SM}$ is much smaller than $\varepsilon_{O_{gT,i}}$, indicates that the event selection strategy using $d_{\rm th}$ can be used to suppress the background.

G. Expected constraints on the operator coefficients

Usually, when the NP signals are not found, the task is to set constraints on the operator coefficients. This can be done by using the statistical signal significance which is defined as [116,117]

$$\mathcal{S}_{\rm stat}=\sqrt{2[(N_{\rm bg}+N_s)\ln(1+N_s/N_{\rm bg})-N_s]}, \quad (5)$$

where $N_{\rm bg}$ is the number of background events, N_s is the number of signal events. Since there is no interference between the SM and NP, the number of events after cuts can be obtained by $N_s=\varepsilon_{O_{gT,i}}L\times\hat{\sigma}_{gT,i}$ and $N_{\rm bg}=\varepsilon_{\rm SM}L\times\hat{\sigma}_{\rm SM}$, where ε is the cut efficiency, $\hat{\sigma}$ is the cross section after N_j cut, and L is the luminosity.

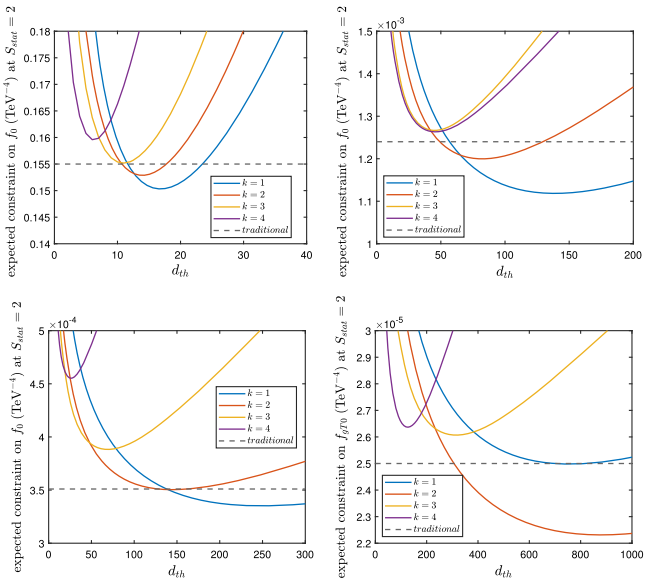

FIG. 9. The expected constraints on f_0 at different energies as functions of $d_{\rm th}$ when $\mathcal{S}_{\rm stat}=2$ in the "conservative" case. The top-left panel corresponds to $\sqrt{s}=3$ TeV, the top-right panel corresponds to $\sqrt{s}=10$ TeV, the bottom-left panel corresponds to $\sqrt{s}=14$ TeV, and the bottom-right panel corresponds to $\sqrt{s}=30$ TeV.

The luminosities in the "conservative" case for the muon colliders at $\sqrt{s}=3$ TeV, 10 TeV, 14 TeV, and 30 TeV are $L=1$ ab^{-1}, 10 ab^{-1}, 10 ab^{-1} and 10 ab^{-1}, respectively [88]. Using the fitted $\varepsilon(d_{\rm th})$, the expected constraints on the operator coefficients can be directly obtained. The expected constraints for $\mathcal{S}_{\rm stat}=2$ in the "conservative" case are shown in Fig. 9. It can be shown that, the cases for $k=1$ and $k=2$ can perform better than the traditional event selection strategy. At $\sqrt{s}=3$, 10 and 14 TeV, $k=1$ works best, and at $\sqrt{s}=30$ TeV, $k=2$ works best. In the following, we choose d_{th} which minimize the expect constraints according to Fig. 9. The results of $d_{\rm th}$ are listed in Table IV.

TABLE IV. The values of $d_{\rm th}$ which minimize the expected constraints on f_0 when $\mathcal{S}_{\rm stat}=2$ in the "conservative" case.

	k	3 TeV 1 ab^{-1}	10 TeV 10 ab^{-1}	14 TeV 10 ab^{-1}	30 TeV 10 ab^{-1}
$d_{\rm th}$	1	16.8	139.1	248.8	745.1
	2	13.9	81.9	144.9	879.7
	3	10.9	44.8	68.7	313.3
	4	6.3	45.5	26.1	125.6

TABLE V. Expected constraints on $f_i(10^{-3}\ \mathrm{TeV}^{-4})$ and M_i (TeV) in the "conservative" case at muon colliders.

		$k=1$				$k=2$
	$\mathcal{S}_{\mathrm{stat}}$	3 TeV 1 ab^{-1}	10 TeV 10 ab^{-1}	14 TeV 10 ab^{-1}	30 TeV 10 ab^{-1}	30 TeV 10 ab^{-1}
$\lvert f_0\rvert$	2	<149	<1.11	<0.333	<0.0248	<0.0222
	3	<184	<1.36	<0.410	<0.0306	<0.0273
	5	<239	<1.77	<0.532	<0.0498	<0.0356
M_0	2	>0.804	>2.74	>3.70	>7.08	>7.28
	3	>0.764	>2.60	>3.51	>6.72	>6.92
	5	>0.715	>2.44	>3.29	>6.29	>6.47
$\lvert f_1\rvert$	2	<221	<1.75	<0.538	<0.0398	<0.0382
	3	<272	<2.14	<0.661	<0.0490	<0.0471
	5	<353	<2.78	<0.858	<0.0638	<0.0614
M_1	2	>0.729	>2.45	>3.28	>6.29	>6.36
	3	>0.692	>2.32	>3.12	>5.98	>6.04
	5	>0.649	>2.18	>2.92	>5.59	>5.65
$\lvert f_2\rvert$	2	<416	<3.09	<0.930	<0.0686	<0.0623
	3	<511	<3.79	<1.142	<0.0845	<0.0768
	5	<664	<4.92	<1.484	<0.1100	<0.1001
M_2	2	>0.622	>2.12	>2.86	>5.49	>5.63
	3	>0.591	>2.01	>2.72	>5.22	>5.34
	5	>0.554	>1.89	>2.55	>4.88	>5.00
$\lvert f_3\rvert$	2	<402	<3.12	<0.957	<0.0710	<0.0667
	3	<494	<3.83	<1.176	<0.0873	<0.0821
	5	<641	<4.97	<1.528	<0.1137	<0.1071
M_3	2	>0.628	>2.12	>2.84	>5.45	>5.53
	3	>0.596	>2.01	>2.70	>5.17	>5.25
	5	>0.559	>1.88	>2.53	>4.84	>4.91

TABLE VI. Same as Table V but for the "optimistic" case.

		$k=1$		$k=2$
	$\mathcal{S}_{\mathrm{stat}}$	14 TeV 20 ab^{-1}	30 TeV 90 ab^{-1}	30 TeV 90 ab^{-1}
$\lvert f_0\rvert$ $(10^{-4}\ \mathrm{TeV}^{-4})$	2	<2.80	<0.142	<0.127
	3	<3.43	<0.175	<0.156
	5	<4.45	<0.226	<0.202
M_0 (TeV)	2	>3.87	>8.14	>8.37
	3	>3.67	>7.73	>7.95
	5	>3.44	>7.25	>7.46
$\lvert f_1\rvert$ $(10^{-4}\ \mathrm{TeV}^{-4})$	2	<4.51	<0.228	<0.219
	3	<5.54	<0.280	<0.269
	5	<7.18	<0.363	<0.349
M_1 (TeV)	2	>3.43	>7.23	>7.31
	3	>3.26	>6.87	>6.94
	5	>3.05	>6.44	>6.51
$\lvert f_2\rvert$ $(10^{-4}\ \mathrm{TeV}^{-4})$	2	<7.80	<0.394	<0.357
	3	<9.58	<0.483	<0.438
	5	<12.42	<0.626	<0.568
M_2 (TeV)	2	>2.99	>6.31	>6.47
	3	>2.84	>6.00	>6.14
	5	>2.66	>5.62	>5.76
$\lvert f_3\rvert$ $(10^{-4}\ \mathrm{TeV}^{-4})$	2	<8.04	<0.407	<0.382
	3	<9.86	<0.499	<0.469
	5	<12.79	<0.647	<0.608
M_3 (TeV)	2	>2.97	>6.26	>6.36
	3	>2.82	>5.95	>6.04
	5	>2.64	>5.57	>5.66

The expected constraints on f_i and M_i in both the "conservative" and "optimistic" cases [88] are calculated, and listed in Tables V and VI. Taking the case of $\sqrt{s} = 3$ TeV as an example, in this paper, the 2σ constraint on M_0 is about 27% of $\sqrt{s}$ which is smaller than $\sqrt{s}$. From the point of view of partial wave unitarity bounds, there is no sign that the validity of EFT has been violated. However, if we assume that $f_i = c_i/\Lambda^4$ and assume that $\Lambda \geq \sqrt{s}$, then $c_i \sim \mathcal{O}(4\pi)$ (for example $|f_0/\Lambda^4|s^2 < 12.1$ for $\mathcal{S}_{\mathrm{stat}} = 2$ and $\sqrt{s} = 3$ TeV). Thus, the constraints are still in the strongly coupled scenarios, and one can expect that combined constraints of multiple processes or higher luminosities can further tighten the constraints. Compared with the HL-LHC, where the combined (combined of $pp \to \gamma\gamma$, $pp \to \ell^+\ell^-\gamma$, $pp \to \nu\bar{\nu}\gamma$, and $pp \to q\bar{q}\gamma$ channels) constraint on M_0 is about 22% of $\sqrt{s}$ [80], our result indicates that the muon collider is sensitive to gQGCs.

The uncertainties in this paper come from different aspects. In MC simulation, higher-order contributions are not included. The beam induced background is also important, which typically leads to particles tangent to the trajectories, and is not yet included in DELPHES at this stage, therefore difficult to be considered. Another uncertainty comes from the stochastic nature of training, e.g., randomly dividing the data into training, validation and test sets, random network initialization, random selection of a portion of the training data when calculating the gradient during training. To study the projected sensitivities, one operator is considered at a time. There are possible contributions from other high dimensional contributions. Note that if the jets were from quarks, there will be possible interference between the SM and NP, and the interference of a dimension-12 operator is at the same order of gQGCs assuming the Wilson coefficients are at the same order.

H. Compare AEAD with traditional method

A comparison is made between the AEAD and the traditional method. It can be seen that, the best expected constraints derived by the AEAD are generally better than those derived using the traditional method used in Ref. [92], which is listed in Table VII. The expected constraints on f_i at 95% CL level ($S_{\mathrm{stat}} = 2$) are compared with those from the traditional method [92] in Fig. 10, it can

TABLE VII. The traditional event selection strategies at different energies [92]. $\not{p}_T$ is the transverse missing momentum and m_{jj} is the invariant mass of two hardest jets. It is also required that $N_j \geq 2$ to calculate m_{jj}.

$\sqrt{s}$ (TeV)	$\not{p}_T$	m_{jj}
3	>50 GeV	>1 TeV
10	>100 GeV	>3 TeV
14	>100 GeV	>5 TeV
30	>200 GeV	>10 TeV

be shown that in all cases, the AEAD performs better. It can be concluded that the AEAD can archive better results. In particular, the AEAD always works better for $O_{gT,1,3}$ than $O_{gT,0,2}$, and the case of $\sqrt{s} = 14$ TeV is the worst case.

It is also noted that, only marginal gains are achieved by the AEAD with respect to the 'traditional' event selection strategy. However, compared with a traditional event selection strategy, AEAD is an AD algorithm. It does not need the information of NP, in the training of the AE, only the information of the SM is used. Although the AEAD does not utilize the information of NP, because of this it brings advantages. AEAD can find signals of NP without knowing what NP signals to look for. Except for that, AEAD simply find out the anomalous signals. Even if there was no NP, anomalous signals are noteworthy, they could be rare processes in the SM, or possible errors.

Another reason is that the final state of the process in this paper is simple. The ANN is good at finding potential patterns from a large and complicated feature space. It is expected that as the precision tests of the SM get more intensive, rarer processes with more final state particles will be considered, it will be more difficult for a traditional method to reveal NP signals from analysis of a large amount of possible observables, which is a better case to utilize the advantages of ANNs.

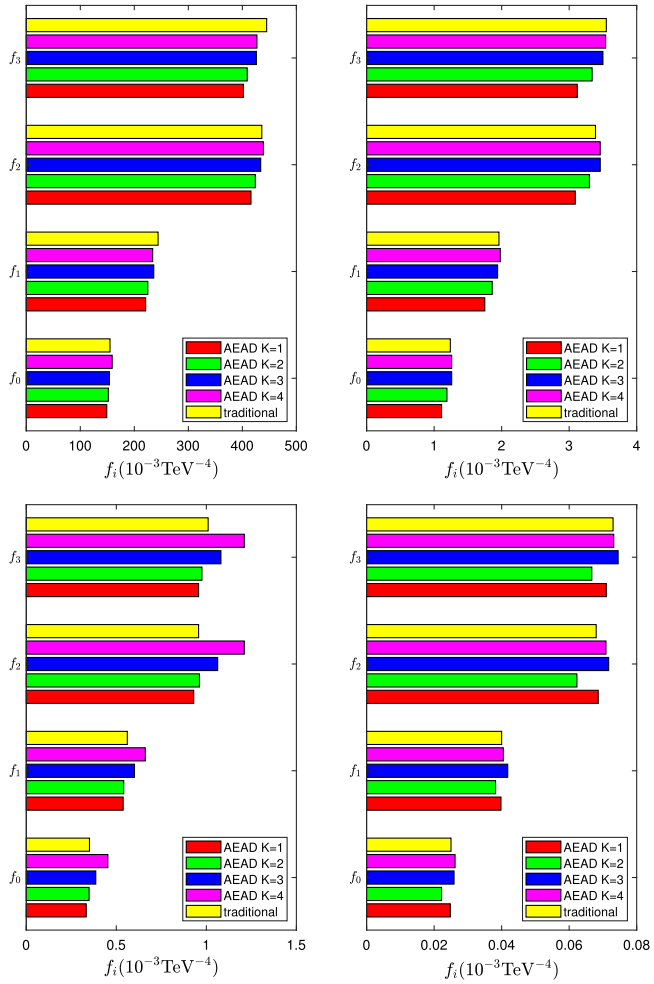

FIG. 10. The comparison of the expected constraints at 95% C.L. level obtained by a traditional event selection strategy [92], and the AEAD event selection strategy. The top-left panel corresponds to $\sqrt{s} = 3$ TeV, the top-right panel corresponds to 10 TeV, the bottom-left panel corresponds to 14 TeV, and the bottom-right panel corresponds to 30 TeV.

V. SUMMARY

With the potential for acceleration using quantum computing, the role of AE in searching for signals of NP becomes important, especially since NP has yet to show clear signs, and the search for NP cannot avoid dealing with increasing amounts of data for the foreseeable future. In this paper, the process of searching for NP signals using AEAD is proposed. The procedure is independent of the content of the NP to be searched since only the SM dataset is used in the training phase.

As an example, the process $\mu^+\mu^- \to \nu\bar{\nu}jj$ at muon colliders is considered, which is sensitive to the dimension-8 operators contributing to gQGCs. The event selection strategy based on AEAD is studied, and the expected constraints on the operator coefficients are calculated. It can be shown that, the constraints are generally tighter than those obtained by using a traditional event selection strategy. Therefore, it can be concluded that the AEAD is effective in the phenomenological study of the SMEFT. It is expected that the AE algorithm accelerated by quantum computers can be even more efficient in the future.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant No. 12147214, and the Natural Science Foundation of the Liaoning Scientific Committee No. LJKZ0978.

[1] Y. Farzan and M. Tortola, Neutrino oscillations and non-standard interactions, Front. Phys. (Beijing) **6**, 10 (2018).

[2] P. S. Bhupal Dev *et al.*, Neutrino non-standard interactions: A status report, SciPost Phys. Proc. **2**, 001 (2019).

[3] C. A. Argüelles *et al.*, Snowmass white paper: Beyond the standard model effects on neutrino flavor, Eur. Phys. J. C **83**, 15 (2023).

[4] T. Aaltonen *et al.* (CDF Collaboration), High-precision measurement of the W boson mass with the CDF II detector, Science **376**, 170 (2022).

[5] J. de Blas, M. Pierini, L. Reina, and L. Silvestrini, Impact of the recent measurements of the top-quark and W-boson masses on electroweak precision fits, Phys. Rev. Lett. **129**, 271801 (2022).

[6] B. Abi *et al.* (Muon $g-2$ Collaboration), Measurement of the positive muon anomalous magnetic moment to 0.46 ppm, Phys. Rev. Lett. **126**, 141801 (2021).

[7] D. P. Aguillard *et al.* (Muon $g-2$ Collaboration), Measurement of the positive muon anomalous magnetic moment to 0.20 ppm, Phys. Rev. Lett. **131**, 161802 (2023).

[8] T. Aoyama *et al.*, The anomalous magnetic moment of the muon in the standard model, Phys. Rep. **887**, 1 (2020).

[9] A. Crivellin and B. Mellado, Anomalies in particle physics, Nat. Rev. Phys. **6**, 294 (2024).

[10] J. Ellis, Outstanding questions: Physics beyond the standard model, Phil. Trans. R. Soc. A **370**, 818 (2012).

[11] S. Weinberg, Baryon and lepton nonconserving processes, Phys. Rev. Lett. **43**, 1566 (1979).

[12] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, Dimension-six terms in the standard model Lagrangian, J. High Energy Phys. 10 (2010) 085.

[13] S. Willenbrock and C. Zhang, Effective field theory beyond the standard model, Annu. Rev. Nucl. Part. Sci. **64**, 83 (2014).

[14] E. Masso, An effective guide to beyond the standard model physics, J. High Energy Phys. 10 (2014) 128.

[15] D. R. Green, P. Meade, and M.-A. Pleier, Multiboson interactions at the LHC, Rev. Mod. Phys. **89**, 035008 (2017).

[16] C. Zhang and S.-Y. Zhou, Convex geometry perspective on the (Standard Model) effective field theory space, Phys. Rev. Lett. **125**, 201601 (2020).

[17] C. W. Murphy, Dimension-8 operators in the standard model effective field theory, J. High Energy Phys. 10 (2020) 174.

[18] H.-L. Li, Z. Ren, J. Shu, M.-L. Xiao, J.-H. Yu, and Y.-H. Zheng, Complete set of dimension-eight operators in the Standard Model effective field theory, Phys. Rev. D **104**, 015026 (2021).

[19] C. Zhang and S.-Y. Zhou, Positivity bounds on vector boson scattering at the LHC, Phys. Rev. D **100**, 095003 (2019).

[20] Q. Bi, C. Zhang, and S.-Y. Zhou, Positivity constraints on aQGC: Carving out the physical parameter space, J. High Energy Phys. 06 (2019) 137.

[21] M. Aaboud *et al.* (ATLAS Collaboration), Studies of $Z\gamma$ production in association with a high-mass dijet system in pp collisions at $\sqrt{s}=8$ TeV with the ATLAS detector, J. High Energy Phys. 07 (2017) 107.

[22] V. Khachatryan *et al.* (CMS Collaboration), Measurement of the cross section for electroweak production of $Z\gamma$ in association with two jets and constraints on anomalous quartic gauge couplings in proton–proton collisions at $\sqrt{s}=8$ TeV, Phys. Lett. B **770**, 380 (2017).

[23] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of the cross section for electroweak production of a Z boson, a photon and two jets in proton-proton collisions at $\sqrt{s}=13$ TeV and constraints on anomalous quartic couplings, J. High Energy Phys. 06 (2020) 076.

[24] V. Khachatryan *et al.* (CMS Collaboration), Measurement of electroweak-induced production of $W\gamma$ with two jets in pp collisions at $\sqrt{s}=8$ TeV and constraints on anomalous quartic gauge couplings, J. High Energy Phys. 06 (2017) 106.

[25] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of vector boson scattering and constraints on anomalous quartic couplings from events with four leptons and two jets in proton–proton collisions at $\sqrt{s}=13$ TeV, Phys. Lett. B **774**, 682 (2017).

[26] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of differential cross sections for Z boson pair production in association with jets at $\sqrt{s}=8$ and 13 TeV, Phys. Lett. B **789**, 19 (2019).

[27] M. Aaboud *et al.* (ATLAS Collaboration), Observation of electroweak $W^{\pm}Z$ boson pair production in association with two jets in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector, Phys. Lett. B **793**, 469 (2019).

[28] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of electroweak WZ boson production and search for new physics in $WZ+$ two jets events in pp collisions at $\sqrt{s}=13$ TeV, Phys. Lett. B **795**, 281 (2019).

[29] V. Khachatryan *et al.* (CMS Collaboration), Evidence for exclusive $\gamma\gamma\to W^+W^-$ production and constraints on anomalous quartic gauge couplings in pp collisions at $\sqrt{s}=7$ and 8 TeV, J. High Energy Phys. 08 (2016) 119.

[30] A. M. Sirunyan *et al.* (CMS Collaboration), Observation of electroweak production of same-sign W boson pairs in the two jet and two same-sign lepton final state in proton-proton collisions at $\sqrt{s}=13$ TeV, Phys. Rev. Lett. **120**, 081801 (2018).

[31] J. Ellis, H.-J. He, and R.-Q. Xiao, Probing neutral triple gauge couplings with $Z^*\gamma(\nu\bar{\nu}\gamma)$ production at hadron colliders, Phys. Rev. D **108**, L111704 (2023).

[32] S. Spor, Probe of the anomalous neutral triple gauge couplings in photon-induced collision at future muon colliders, Nucl. Phys. **B991**, 116198 (2023).

[33] S. Spor, E. Gurkanli, and M. Köksal, Search for the anomalous $ZZ\gamma$ and $Z\gamma\gamma$ couplings via $\nu\nu\gamma$ production at the CLIC, Nucl. Phys. **B979**, 115785 (2022).

[34] A. Yilmaz, Search for the limits on anomalous neutral triple gauge couplings via ZZ production in the $\ell\ell\nu\nu$ channel at FCC-hh, Nucl. Phys. **B969**, 115471 (2021).

[35] J. Ellis, H.-J. He, and R.-Q. Xiao, Probing new physics in dimension-8 neutral gauge couplings at e^+e^- colliders, Sci. China Phys. Mech. Astron. **64**, 221062 (2021).

[36] A. Senol, H. Denizli, A. Yilmaz, I. Turk Cakir, and O. Cakir, Study on anomalous neutral triple gauge boson couplings from dimension-eight operators at the HL-LHC, Acta Phys. Pol. B **50**, 1597 (2019).

[37] A. Yilmaz, A. Senol, H. Denizli, I. Turk Cakir, and O. Cakir, Sensitivity on anomalous neutral triple gauge couplings via ZZ production at FCC-hh, Eur. Phys. J. C **80**, 173 (2020).

[38] J. Ellis, S.-F. Ge, H.-J. He, and R.-Q. Xiao, Probing the scale of new physics in the $ZZ\gamma$ coupling at e^+e^- colliders, Chin. Phys. C **44**, 063106 (2020).

[39] Y.-C. Guo, Y.-Y. Wang, J.-C. Yang, and C.-X. Yue, Constraints on anomalous quartic gauge couplings via $W\gamma jj$ production at the LHC, Chin. Phys. C **44**, 123105 (2020).

[40] Y.-C. Guo, Y.-Y. Wang, and J.-C. Yang, Constraints on anomalous quartic gauge couplings by $\gamma\gamma \to W^+W^-$ scattering, Nucl. Phys. **B961**, 115222 (2020).

[41] J.-C. Yang, Y.-C. Guo, C.-X. Yue, and Q. Fu, Constraints on anomalous quartic gauge couplings via $Z\gamma jj$ production at the LHC, Phys. Rev. D **104**, 035015 (2021).

[42] Q. Fu, J.-C. Yang, C.-X. Yue, and Y.-C. Guo, The study of neutral triple gauge couplings in the process $e^+e^- \to Z\gamma$ including unitarity bounds, Nucl. Phys. **B972**, 115543 (2021).

[43] J.-C. Yang, Z.-B. Qing, X.-Y. Han, Y.-C. Guo, and T. Li, Tri-photon at muon collider: A new process to probe the anomalous quartic gauge couplings, J. High Energy Phys. 07 (2022) 053.

[44] S. Jahedi and J. Lahiri, Probing anomalous $ZZ\gamma$ and $Z\gamma\gamma$ couplings at the e^+e^- colliders using optimal observable technique, J. High Energy Phys. 04 (2023) 085.

[45] S. Jahedi, Optimal estimation of dimension-8 neutral triple gauge couplings at e^+e^- colliders, J. High Energy Phys. 12 (2023) 031.

[46] C. F. Anders *et al.*, Vector boson scattering: Recent experimental and theory developments, Rev. Phys. **3**, 44 (2018).

[47] B. Henning, X. Lu, T. Melia, and H. Murayama, 2, 84, 30, 993, 560, 15456, 11962, 261485, …: Higher dimension operators in the SM EFT, J. High Energy Phys. 08 (2017) 016; 09 (2019) 19.

[48] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, Machine learning at the energy and intensity frontiers of particle physics, Nature (London) **560**, 41 (2018).

[49] P. Baldi, P. Sadowski, and D. Whiteson, Searching for exotic particles in high-energy physics with deep learning, Nat. Commun. **5**, 4308 (2014).

[50] J. Ren, L. Wu, J. M. Yang, and J. Zhao, Exploring supersymmetry with machine learning, Nucl. Phys. **B943**, 114613 (2019).

[51] M. Abdughani, J. Ren, L. Wu, and J. M. Yang, Probing stop pair production at the LHC with graph neural networks, J. High Energy Phys. 08 (2019) 055.

[52] J. Ren, L. Wu, and J. M. Yang, Unveiling CP property of top-Higgs coupling with graph neural networks at the LHC, Phys. Lett. B **802**, 135198 (2020).

[53] M. Letizia, G. Losapio, M. Rando, G. Grosso, A. Wulzer, M. Pierini, M. Zanetti, and L. Rosasco, Learning new physics efficiently with nonparametric methods, Eur. Phys. J. C **82**, 879 (2022).

[54] R. T. D'Agnolo, G. Grosso, M. Pierini, A. Wulzer, and M. Zanetti, Learning multivariate new physics, Eur. Phys. J. C **81**, 89 (2021).

[55] R. T. D'Agnolo and A. Wulzer, Learning new physics from a machine, Phys. Rev. D **99**, 015014 (2019).

[56] J.-C. Yang, J.-H. Chen, and Y.-C. Guo, Extract the energy scale of anomalous $\gamma\gamma \to W^+W^-$ scattering in the vector boson scattering process using artificial neural networks, J. High Energy Phys. 09 (2021) 085.

[57] J.-C. Yang, X.-Y. Han, Z.-B. Qin, T. Li, and Y.-C. Guo, Measuring the anomalous quartic gauge couplings in the $W^+W^- \to W^+W^-$ process at muon collider using artificial neural networks, J. High Energy Phys. 09 (2022) 074.

[58] A. De Simone and T. Jacques, Guiding new physics searches with unsupervised learning, Eur. Phys. J. C **79**, 289 (2019).

[59] M. A. Md Ali, N. Badrud'din, H. Abdullah, and F. Kemi, Alternate methods for anomaly detection in high-energy physics via semi-supervised learning, Int. J. Mod. Phys. A **35**, 2050131 (2020).

[60] E. Fol, R. Tomás, J. Coello de Portugal, and G. Franchetti, Detection of faulty beam position monitors using unsupervised learning, Phys. Rev. Accel. Beams **23**, 102805 (2020).

[61] G. Kasieczka *et al.*, The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics, Rep. Prog. Phys. **84**, 124201 (2021).

[62] L. Jiang, Y.-C. Guo, and J.-C. Yang, Detecting anomalous quartic gauge couplings using the isolation forest machine learning algorithm, Phys. Rev. D **104**, 035021 (2021).

[63] J.-C. Yang, Y.-C. Guo, and L.-H. Cai, Using a nested anomaly detection machine learning algorithm to study the neutral triple gauge couplings at an e^+e^- collider, Nucl. Phys. **B977**, 115735 (2022).

[64] Y.-F. Dong, Y.-C. Mao, and J.-C. Yang, Searching for anomalous quartic gauge couplings at muon colliders using principal component analysis, Eur. Phys. J. C **83**, 555 (2023).

[65] M. Crispim Romão, N. F. Castro, and R. Pedro, Finding new physics without learning about it: Anomaly Detection as a tool for searches at colliders, Eur. Phys. J. C **81**, 27 (2021); **81**, 1020(E) (2021).

[66] M. van Beekveld, S. Caron, L. Hendriks, P. Jackson, A. Leinweber, S. Otten, R. Patrick, R. Ruiz De Austri, M. Santoni, and M. White, Combining outlier analysis algorithms to identify new physics at the LHC, J. High Energy Phys. 09 (2021) 024.

[67] M. Kuusela, T. Vatanen, E. Malmi, T. Raiko, T. Aaltonen, and Y. Nagai, Semi-supervised anomaly detection—Towards model-independent searches of new physics, J. Phys. Conf. Ser. **368**, 012032 (2012).

[68] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[69] M. Schuld, I. Sinayskiy, and F. Petruccione, An introduction to quantum machine learning, Contemp. Phys. **56**, 172 (2015).

[70] D. P. García, J. Cruz-Benito, and F. J. García-Peñalvo, Systematic literature review: Quantum machine learning and its applications, arXiv:2201.04093.

[71] C.-Y. Liou, J.-C. Huang, and W.-C. Yang, Modeling word perception using the elman network, Neurocomputing; Variable Star Bulletin **71**, 3150 (2008), Advances in Neural Information Processing (ICONIP 2006) / Brazilian Symposium on Neural Networks (SBRN 2006).
[72] C.-Y. Liou, W.-C. Cheng, J.-W. Liou, and D.-R. Liou, Autoencoder for words, Neurocomputing;Variable Star Bulletin **139**, 84 (2014).
[73] M. Farina, Y. Nakai, and D. Shih, Searching for new physics with deep autoencoders, Phys. Rev. D **101**, 075021 (2020).
[74] O. Cerri, T. Q. Nguyen, M. Pierini, M. Spiropulu, and J.-R. Vlimant, Variational autoencoders for new physics mining at the Large Hadron Collider, J. High Energy Phys. 05 (2019) 036.
[75] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum autoencoders for efficient compression of quantum data, Quantum Sci. Technol. **2**, 045001 (2017).
[76] C. Bravo-Prieto, Quantum autoencoders with enhanced data encoding, Mach. Learn. Sci. Technol. **2**, 035028 (2021).
[77] D. Bondarenko and P. Feldmann, Quantum autoencoders to denoise quantum data, Phys. Rev. Lett. **124**, 130502 (2020).
[78] A. Khoshaman, W. Vinci, B. Denis, E. Andriyash, H. Sadeghi, and M. H. Amin, Quantum variational autoencoder, Quantum Sci. Technol. **4**, 014001 (2019).
[79] J. Ellis and S.-F. Ge, Constraining gluonic quartic gauge coupling Ooperators with $gg \to \gamma\gamma$, Phys. Rev. Lett. **121**, 041801 (2018).
[80] J. Ellis, S.-F. Ge, and K. Ma, Hadron collider probes of the quartic couplings of gluons to the photon and Z boson, J. High Energy Phys. 04 (2022) 123.
[81] D. Buttazzo, D. Redigolo, F. Sala, and A. Tesi, Fusing vectors into scalars at high energy lepton colliders, J. High Energy Phys. 11 (2018) 144.
[82] J. P. Delahaye, M. Diemoz, K. Long, B. Mansoulié, N. Pastrone, L. Rivkin, D. Schulte, A. Skrinsky, and A. Wulzer, Muon colliders, arXiv:1901.06150.
[83] M. Lu, A. M. Levin, C. Li, A. Agapitos, Q. Li, F. Meng, S. Qian, J. Xiao, and T. Yang, The physics case for an electron-muon collider, Adv. High Energy Phys. **2021**, 6693618 (2021).
[84] R. Franceschini and M. Greco, Higgs and BSM physics at the future muon collider, Symmetry **13**, 851 (2021).
[85] R. Palmer *et al.*, Muon collider design, Nucl. Phys. B, Proc. Suppl. **51**, 61 (1996).
[86] S. D. Holmes and V. D. Shiltsev, Muon collider, in *Outlook for the Future*, edited by C. Joshi, A. Caldwell, P. Muggli, S. D. Holmes, and V. D. Shiltsev (Springer-Verlag Berlin Heidelberg, Germany, 2013), pp. 816–822, 10.1007/978-3-642-23053-0_48.
[87] A. Costantini, F. De Lillo, F. Maltoni, L. Mantani, O. Mattelaer, R. Ruiz, and X. Zhao, Vector boson fusion at multi-TeV muon colliders, J. High Energy Phys. 09 (2020) 080.
[88] H. Al Ali *et al.*, The muon Smasher's guide, Rep. Prog. Phys. **85**, 084201 (2022).
[89] T. Han, Y. Ma, and K. Xie, High energy leptonic collisions and electroweak parton distribution functions, Phys. Rev. D **103**, L031301 (2021).
[90] T. Han, Y. Ma, and K. Xie, Quark and gluon contents of a lepton at high energies, J. High Energy Phys. 02 (2022) 154.
[91] C. Aime *et al.*, Muon collider physics summary, arXiv:2203.07256.
[92] J.-C. Yang, Y.-C. Guo, and Y.-F. Dong, Study of the gluonic quartic gauge couplings at muon colliders, Commun. Theor. Phys. **75**, 115201 (2023).
[93] M. Born and L. Infeld, Foundations of the new field theory, Proc. R. Soc. A **144**, 425 (1934).
[94] E. S. Fradkin and A. A. Tseytlin, Nonlinear electrodynamics from quantized strings, Phys. Lett. B **163B**, 123 (1985).
[95] A. A. Tseytlin, Born-Infeld action, supersymmetry and string theory, in *The Many Faces of the Superworld* (World Scientific, Singapore, 2000), pp. 417–452.
[96] C. Cheung, K. Kampf, J. Novotny, C.-H. Shen, J. Trnka, and C. Wen, Vector effective field theories from soft limits, Phys. Rev. Lett. **120**, 261602 (2018).
[97] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, J. High Energy Phys. 07 (2014) 079.
[98] N. D. Christensen and C. Duhr, FeynRules—Feynman rules made easy, Comput. Phys. Commun. **180**, 1614 (2009).
[99] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, and T. Reiter, UFO—The Universal FeynRules Output, Comput. Phys. Commun. **183**, 1201 (2012).
[100] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, An introduction to PYTHIA 8.2, Comput. Phys. Commun. **191**, 159 (2015).
[101] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), DELPHES 3, A modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057.
[102] Y.-C. Guo, F. Feng, A. Di, S.-Q. Lu, and J.-C. Yang, MLAnalysis: An open-source program for high energy physics analyses, Comput. Phys. Commun. **294**, 108957 (2024).
[103] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous distributed systems, arXiv:1603.04467.
[104] T. D. Lee and C.-N. Yang, Theoretical discussions on possible high-energy neutrino experiments, Phys. Rev. Lett. **4**, 307 (1960).
[105] M. Froissart, Asymptotic behavior and subtractions in the Mandelstam representation, Phys. Rev. **123**, 1053 (1961).
[106] G. Passarino, W W scattering and perturbative unitarity, Nucl. Phys. **B343**, 31 (1990).
[107] T. Corbett, O. J. P. Éboli, and M. C. Gonzalez-Garcia, Unitarity constraints on dimension-six operators, Phys. Rev. D **91**, 035014 (2015).
[108] J. Layssac, F. M. Renard, and G. J. Gounaris, Unitarity constraints for transverse gauge bosons at LEP and supercolliders, Phys. Lett. B **332**, 146 (1994).
[109] T. Corbett, O. J. P. Éboli, and M. C. Gonzalez-Garcia, Unitarity constraints on dimension-six operators II:

Including fermionic operators, Phys. Rev. D **96**, 035006 (2017).
[110] R. Gomez-Ambrosio, Vector boson scattering studies in CMS: The $pp \to ZZjj$ channel, Acta Phys. Pol. B Proc. Suppl. **11**, 239 (2018).
[111] G. Perez, M. Sekulla, and D. Zeppenfeld, Anomalous quartic gauge couplings and unitarization for the vector boson scattering process $pp \to W^+W^+jjX \to \ell^+\nu_\ell\ell^+\nu_\ell jjX$, Eur. Phys. J. C **78**, 759 (2018).
[112] E. d. S. Almeida, O. J. P. Éboli, and M. C. Gonzalez–Garcia, Unitarity constraints on anomalous quartic couplings, Phys. Rev. D **101**, 113003 (2020).
[113] W. Kilian, S. Sun, Q.-S. Yan, X. Zhao, and Z. Zhao, Multi-Higgs boson production and unitarity in vector-boson fusion at future hadron colliders, Phys. Rev. D **101**, 076012 (2020).
[114] D. Donoho and J. Jin, Higher criticism for detecting sparse heterogeneous mixtures, Ann. Stat. **32**, 962 (2004).
[115] A. Graves, Supervised sequence labelling with recurrent neural networks, in Studies in Computational Intelligence (Springer, Heidelberg, 2012), 10.1007/978-3-642-24797-2.
[116] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, Eur. Phys. J. C **71**, 1554 (2011); **73**, 2501(E) (2013).
[117] P. A. Zyla *et al.* (Particle Data Group), Review of particle physics, Prog. Theor. Exp. Phys. **2020**, 083C01 (2020).