

Weakly supervised anomaly detection in events with a Higgs boson and exotic physics

Chi Lung Cheng^{1,2,*} Sarah Demers^{3,†} Sascha Diefenbacher^{1,‡}
Runze Li^{3,§} Benjamin Nachman^{4,5,||} and Dennis Noll^{1,5,¶}

¹Physics Division, *Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

²Department of Physics, *University of Wisconsin, Madison, Wisconsin 53706, USA*

³Department of Physics, *Yale University, New Haven, Connecticut 06511, USA*

⁴Department of Particle Physics and Astrophysics, *Stanford University, Stanford, California 94305, USA*

⁵Fundamental Physics Directorate, *SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA*



(Received 6 November 2025; accepted 20 May 2026; published 6 July 2026)

We present a machine-learning-based anomaly detection strategy designed to identify anomalous physics in events containing resonant Standard Model physics and demonstrate this method on the final state of a Higgs boson decaying to two photons. The demonstration targets high-dimensional deviations in the region of phase space containing the Higgs mass peak in a fully signal-agnostic manner. A latent-space embedding, learned from event kinematics, enables the use of a large set of potentially sensitive features. Backgrounds are estimated using a hybrid approach that combines machine-learning-based generative modeling with traditional simulation, and a discriminator is trained in the latent space to distinguish data from background estimates. After applying a selection on the classifier output, the invariant mass distribution of the diphoton system is examined for localized excesses above the simulated Higgs peak. We benchmark the sensitivity of this strategy using simplified simulated proton-proton collisions corresponding to data recorded during Run 2 of the LHC and show that the method can provide significant improvements in sensitivity, even for small signal injections that could remain undetected in an inclusive analysis. These results demonstrate that the proposed strategy is a promising and viable approach for future searches and should be applied to recorded collider data.

DOI: [10.1103/7582-jt9n](https://doi.org/10.1103/7582-jt9n)

I. INTRODUCTION

The LHC has enabled the exploration of the Standard Model (SM) with unprecedented precision. Nevertheless, there remain compelling motivations to search for physics beyond the Standard Model (BSM), such as the existence of dark matter [1], the observed matter-antimatter asymmetry, or the hierarchy problem [2]. A major challenge in searches for BSM physics is the vast number of possible models, which far exceeds the number of feasible targeted search efforts.

Anomaly detection (AD) [3–6] offers a complementary search paradigm, enabling the study of signatures that correspond to a wide range of BSM scenarios simultaneously. This approach is designed to complement dedicated, model-specific searches by providing sensitivity across broad model spaces. AD searches have seen use at both the ATLAS [7–10] and CMS [11] experiments. Different AD methods can be categorized in part on the prior knowledge of the signal. In this work, we present a weakly supervised AD analysis strategy, following the Classification Without Labels (CWoLa) [12–14] and Classifying Anomalies THrough Outer Density Estimation (CATHODE) [15] approaches. This method can achieve close-to-optimal sensitivity depending on the signals assumed to be present in the data.

Our new AD strategy is aimed at identifying anomalies associated with an SM resonance. This extends CATHODE to the case where the background has two components: a nonresonant part estimated from data sidebands (as in standard CATHODE) and a resonant part estimated from simulation or alternative methods. As a demonstrator, we apply it to final states containing a Higgs boson H in

*Contact author: ccheng84@wisc.edu

†Contact author: sarah.demers@yale.edu

‡Contact author: sdiefenbacher@lbl.gov

§Contact author: runze.li@yale.edu

||Contact author: nachman@stanford.edu

¶Contact author: nollde@stanford.edu

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

association with any other particles X , a setup we denote as *Higgs And X Anomaly Detection* (HAXAD).

The Higgs boson plays a unique role in the SM as the only elementary scalar particle and poses a potential portal to new physics. The presence of BSM effects in this sector could alter the inclusive and differential production cross sections of the Higgs boson, modify its decay rates, or introduce entirely new production mechanisms. The Higgs boson could appear in various scenarios, including: supersymmetry (SUSY) cascades [16–21], exotic top-quark decays [22–26], or decays of vector-like quarks [27,28], as motivated in Ref. [29]. It could also arise from extended Higgs sector models, such as those featuring a complex singlet [30] or two real singlets [31], as well as a complex two-Higgs-doublet (2HDM) model [32], a 2HDM model extended by a real scalar singlet [33,34], or the Next-to-Minimal Supersymmetric Standard Model [35,36], as motivated in Ref. [37]. To date, no significant deviations from the SM have been observed in the Higgs sector—but most analyses have been tailored to specific signal models, offering high sensitivity only within limited regions of parameter space.

Within the Higgs sector, our search focuses on the $H \rightarrow \gamma\gamma$ decay channel. While the branching ratio for $H \rightarrow \gamma\gamma$ is relatively small, this channel is highly competitive due to its excellent diphoton mass resolution, strong background rejection, and the ability to model the dominant background processes with analytic functions.

In this work, we demonstrate our proposed strategy on simplified simulated datasets and aim to illustrate how the method could be applied to real data from experiments such as ATLAS [38] or CMS [39]. A classical AD analysis was previously performed by ATLAS in the $H \rightarrow \gamma\gamma$ channel [29], following the approach of earlier searches [40,41] that look for excesses across a large number of exclusive final states, none of which are optimized for a specific signal model. In contrast, the present strategy incorporates machine learning (ML) techniques and is optimized directly on the target data, enabling it to adjust to the characteristics of any potential new physics signal. Additionally, while the new strategy is demonstrated here for the $H \rightarrow \gamma\gamma$ decay channel, the proposed strategy is applicable to any known resonance and is also not restricted to collider-based experiments.

The paper is organized as follows. Section II describes the used dataset, including the event preselection and reconstruction of event features. Section III explains the analysis strategy, and Sec. IV presents the results of the demonstrative analysis. Finally, Sec. V provides our conclusions and outlook.

II. DATASETS

The analysis uses simulated data corresponding to proton-proton collision at a center-of-mass energy of $\sqrt{s} = 13$ TeV with an integrated luminosity of $\mathcal{L} = 137 \text{ fb}^{-1}$. The event generation is performed with MADGRAPH5_AMC@NLO3.5.9 [42] and PYTHIA8.312 [43,44]. All processes are generated at

leading order in QCD, without matching or merging [45]. The simulation does not include contributions from additional simultaneous proton-proton interactions (pileup) or multiparton interactions. A Higgs boson mass of $m_H = 125$ GeV is used for all processes involving Higgs bosons, and all Higgs decays are simulated at the parton-shower level. A simplified detector simulation is performed using DELPHESV3.5.0 [46,47] with a configuration corresponding to a simplified ATLAS detector layout. A link to the repository containing the full simulation framework is given at the end of the paper. The use of leading-order approximations and the simplified detector simulation will result in a less accurate modeling than using higher-order calculations and a dedicated full detector simulation; however, we do not expect a more accurate simulation to change the quantitative results. Likewise, any classical benchmark results may differ from ones obtained using full simulation. Therefore, we do not aim to quantify the performance of any classical benchmark. Instead, we use them as a relative comparison point.

The SM background is composed of two main categories: the nonresonant continuum and the resonant Higgs backgrounds. The continuum background includes nonresonant processes producing two prompt photons and up to three jets in the hard process. To regulate divergences in the generation, minimal phase-space cuts are applied to the transverse momentum of the photons, requiring $p_T > 32$ GeV.

The resonant background includes the production of a SM Higgs boson with its subsequent decay into $H \rightarrow \gamma\gamma$. It includes the four leading Higgs production channels, gluon-gluon fusion (ggF), vector boson fusion, associated production with a weak boson (VH , with $V = W, Z$), production of the Higgs boson together with a top-quark-antiquark pair ($t\bar{t}H$), and other subleading production processes.

Two classes of BSM signal processes are considered to benchmark the sensitivity of the *signal-agnostic* analysis: SUSY signals and signals from an extended Higgs sector. The SUSY signals involve the production of a chargino ($\tilde{\chi}_1^\pm$) in association with the next-to-lightest neutralino ($\tilde{\chi}_2^0$), with prompt decays $\tilde{\chi}_1^\pm \rightarrow W^\pm \tilde{\chi}_1^0$ and $\tilde{\chi}_2^0 \rightarrow H \tilde{\chi}_1^0$, where $W^\pm$ is a SM W boson, H is the SM Higgs boson, and $\tilde{\chi}_1^0$ is the lightest neutralino [48–53]. For the benchmark signal, denoted as SUSY150, the masses of the chargino and the next-to-lightest neutralino are set to $m_{\tilde{\chi}_1^\pm} = m_{\tilde{\chi}_2^0} = 150$ GeV and the lightest neutralino mass is set to $m_{\tilde{\chi}_1^0} = 0.5$ GeV. This mass point is excluded by Ref. [53] and chosen only to demonstrate the new method. The extended Higgs sector signals are aligned with Ref. [37] and contain an additional heavy BSM Higgs boson (X) and a lighter BSM Higgs boson (S). The X boson is exclusively produced via ggF and exclusively decays to the S boson and the SM Higgs boson (H) as $X \rightarrow SH$ [31,37,54–59]. The lighter BSM Higgs boson is constrained to decay exclusively to light quarks, to artificially make its missing transverse energy (E_T^{Miss})

distribution significantly different from the SUSY signals. For the benchmark signal, the mass of X is set to 750 GeV, and the mass of S is set to 100 GeV. Only $H \rightarrow \gamma\gamma$ decays are realized for the Higgs boson in all signal processes.

The simulated events are used to construct datasets corresponding to the targeted integrated luminosity. Events are randomly sampled from all simulated events according to weights derived from their production cross section and generator-level event weights, ensuring that the final datasets correspond to the expected number of recorded events. To test different signal hypotheses, signal events are injected into the background at various levels of signal significance $s = S/\sqrt{B}$ with the expected inclusive number of signal (S) and background (B) events in the respective signal region.

The event preselection broadly aligns with the standard event preselection of $H \rightarrow \gamma\gamma$ analyses from the ATLAS Collaboration. As our primary analysis objects, we select the two highest transverse momentum (p_T) photons with an invariant mass $m_{\gamma\gamma}$ in the range of $105 < m_{\gamma\gamma} < 160$ GeV. The leading and subleading photons are required to have a minimum p_T of 35 GeV and 25 GeV, respectively, to ensure the diphoton trigger is fully efficient and to reduce potential turn-on effects. To further enhance the selection of photons from a resonant decay, the leading and subleading photons must satisfy $p_T/m_{\gamma\gamma} > 0.4$ and $p_T/m_{\gamma\gamma} > 0.3$, respectively. To suppress backgrounds from jets misidentified as photons, both photons are required to be isolated by ensuring the sum of charged-particle transverse momenta in a cone of $\Delta R = 0.5$ around the photon is less than 50% of its p_T . A similar isolation criterion is applied to leptons. Jets are reconstructed using the anti- k_T algorithm [60–62] with $R = 0.4$. A threshold of $p_T > 20$ GeV is applied to reduce contributions from pileup and underlying event. Figure 1 shows the $m_{\gamma\gamma}$ distribution of all preselected events.

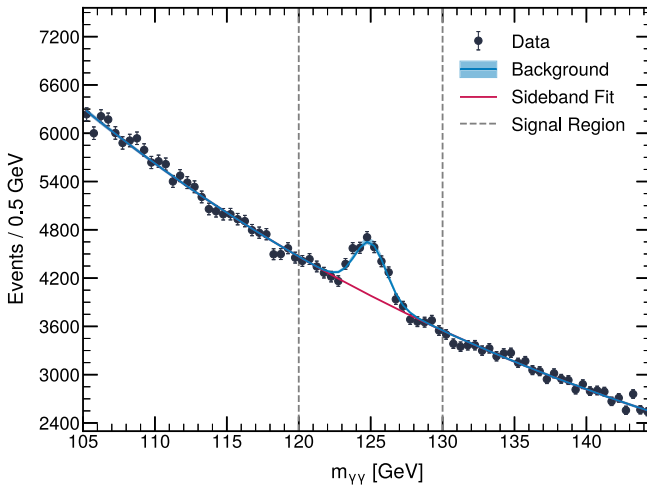


FIG. 1. Distribution of the invariant mass of the two selected photons ($m_{\gamma\gamma}$) for all preselected events.

To preserve the model-agnostic nature of the analysis, each event is characterized by nine high-level kinematic features. These variables are chosen for their general sensitivity to a broad range of BSM physics processes, rather than being optimized for any specific signal signature. The features comprise the diphoton transverse momentum ($p_T^{\gamma\gamma}$) and angular separation ($\Delta R_{\gamma\gamma}$); the transverse momentum of the two leading jets ($p_T^{J_1}, p_T^{J_2}$), their invariant mass (m_{JJ}), and their angular separation (ΔR_{JJ}); a flag indicating the presence of one or more isolated leptons; the scalar sum of jet transverse momentum, H_T , for jets with $p_T > 30$ GeV; and the missing transverse energy, E_T^{Miss} . If an event contains fewer than two jets, the kinematic variables for any nonexistent jets, as well as dijet features, are assigned a default value of zero. There are many more features that characterize events and are potentially sensitive to BSM physics, and the use of an even more extended set of features is left for future work.

III. ANALYSIS STRATEGY

In this analysis, we employ a ML-based AD strategy to search for discrepancies between observed data and estimated background events. The background estimation is performed using a hybrid approach that combines ML with traditional simulation-based techniques. Classifiers, trained to distinguish between data and background estimation, operate in a high-dimensional embedded feature space, enabling the use of a large number of variables that are potentially sensitive to BSM physics. The final sensitivity is extracted with a bump hunt in the invariant mass distribution of the diphoton system.

Similar to other CATHODE-like AD approaches, we need to define a signal region (SR) in which we search for a potential signal and a sideband region (SB) which we assume to be signal depleted. As we search for anomalies related to the Higgs boson, the SR is constrained to be centered around the Higgs mass. Therefore, our SR is fixed and does not need to be moved—a new feature of our approach that is anchored on the particular resonance. For this analysis, we chose a SR width of 10 GeV, centered around the approximate Higgs mass, resulting in SR boundaries of [120 GeV, 130 GeV].

The background estimation in the SR is complicated by unphysical artifacts in the data, such as hard cutoffs and default placeholder values. To mitigate these effects, we encode our data into a latent space using a variational autoencoder (VAE) [63], which smooths over such artifacts while preserving physics-relevant information. The two key innovations of this analysis are the use of a high-dimensional embedding space for better sensitivity and the hybrid background estimation method, enabling a completely new analysis scope.

All steps of the analysis—including the latent space embedding, background estimation, signal classification,

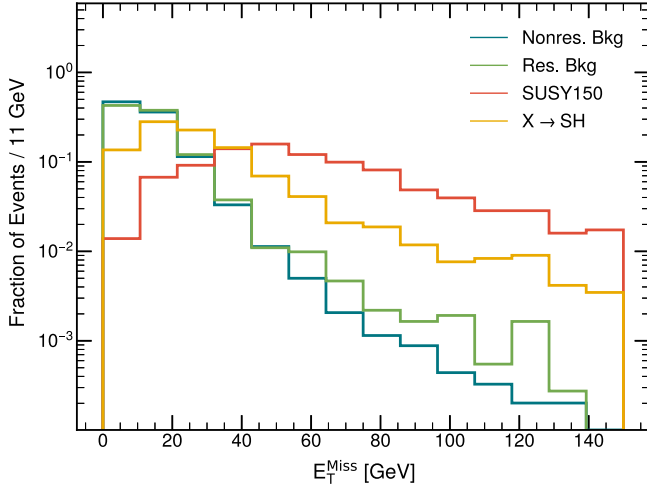


FIG. 2. Distribution of the missing transverse momentum E_T^{Miss} for the nonresonant $\gamma\gamma$ + jets background, the SM Higgs background, and the benchmark signals SUSY150 and $X \rightarrow \text{SH}$.

and statistical inference—are described in the following sections.

A. Latent space embedding

The nine selected kinematic features from the SB data are preprocessed by applying a logarithmic transformation to the p_T and mass features given in units of GeV. All features are then standardized to have zero mean and unit variance. The original distributions of the missing transverse momentum E_T^{Miss} in the SR are shown in Fig. 2 separately for the nonresonant $\gamma\gamma$ + jets background, the resonant SM Higgs background, and the benchmark signals SUSY150 and $X \rightarrow \text{SH}$.

To remain as general as possible, we do not impose a jet multiplicity requirement. As a result, some events lack defined values for jet-related observables such as m_{JJ} , $\Delta R_{J_1 J_2}$, p_{T,J_1} , and p_{T,J_2} . For such cases, these observables are set to zero, corresponding to the mean of their distributions after preprocessing. This introduces large peaks at singular values in the respective feature distributions. Generative models, being inherently probabilistic, struggle to reproduce such discrete spikes, which complicates the estimation of a continuous SR background. This can degrade the performance of downstream weakly supervised classifiers, reducing signal sensitivity or even causing false discoveries due to background sculpting.

To address this, we employ a VAE [63] to encode each event into a lower-dimensional latent representation. The VAE smooths over unphysical artifacts while preserving physics-relevant information for distinguishing signal from background. Therefore, the VAE latent space provides a simpler but still effective feature space for generative modeling.

The VAE is trained on SB data, mapping the nine selected input features into a four-dimensional latent space.

The model is implemented in PYTORCH [64]. Its architecture consists of an encoder and decoder, each with three hidden layers of 128 nodes. The training uses the ADAM optimizer [65] with an initial learning rate of 0.001 and a cosine annealing scheduler. The SB dataset is randomly split into 80% training and 20% validation subsets. The maximum number of epochs is 200, with early stopping if the validation loss does not improve for 20 epochs.

Once trained, the VAE encoder maps both SR and SB events into the latent space by sampling from the Gaussian distributions parametrized by the mean and variance outputs of the encoder. The latent feature distributions for the events in the SR are shown in Fig. 3. Compared to the raw physical features, the encoded features are significantly smoother and have no singular peaks from default values. In the SR, both benchmark signals exhibit distinct latent-space distributions compared to the $\gamma\gamma$ + jets and SM Higgs backgrounds, indicating that the VAE effectively captures and represents the underlying physics.

Our embedding approach is comparable to the Latent-CATHODE concept [66]. However, it differs since we use a VAE for the embedding, rather than a normalizing flow (NF), and we use a generative model trained on the embedding space, rather than approximating it with a Gaussian distribution, as explained in the next section.

B. Background estimation

When estimating the SR events, we need to distinguish between two contributions: (i) diphoton events produced in nonresonant electromagnetic processes and (ii) diphoton events produced by decays of the SM Higgs boson. Both contributions require separate treatment.

The nonresonant part is expected to be continuous between SR and SB. As a result, we use a generative model, conditioned on the diphoton mass and trained on the SB, to interpolate into the SR. This approach is purely data driven and does not rely on simulation.

The generative model we use is a NF [67]. Our NF is implemented in PYTORCH [64] using the NFLOWS package [68]. The model itself consists of six rational quadratic spline (RQS) [69] transformation layers. Each transformation layer has a backend consisting of a two-block residual network with a hidden layer width of $8 \times \text{\#features} = 72$ nodes. Interspersed between each RQS layer is a random permutation layer. The NF is trained on the data points encoded by the VAE model. As this is an already well-regularized representation, only minor mean and width scaling are required for the NF training.

For the training, the full SB data is split into train and validation sets, with the training set comprising 80% of the data. The NF training uses an ADAM [65] optimizer with an initial learning rate of 0.001. We use an early stopping procedure that monitors the validation loss and halves the learning rate every time the validation loss does not improve for two epochs, with an epoch being defined as

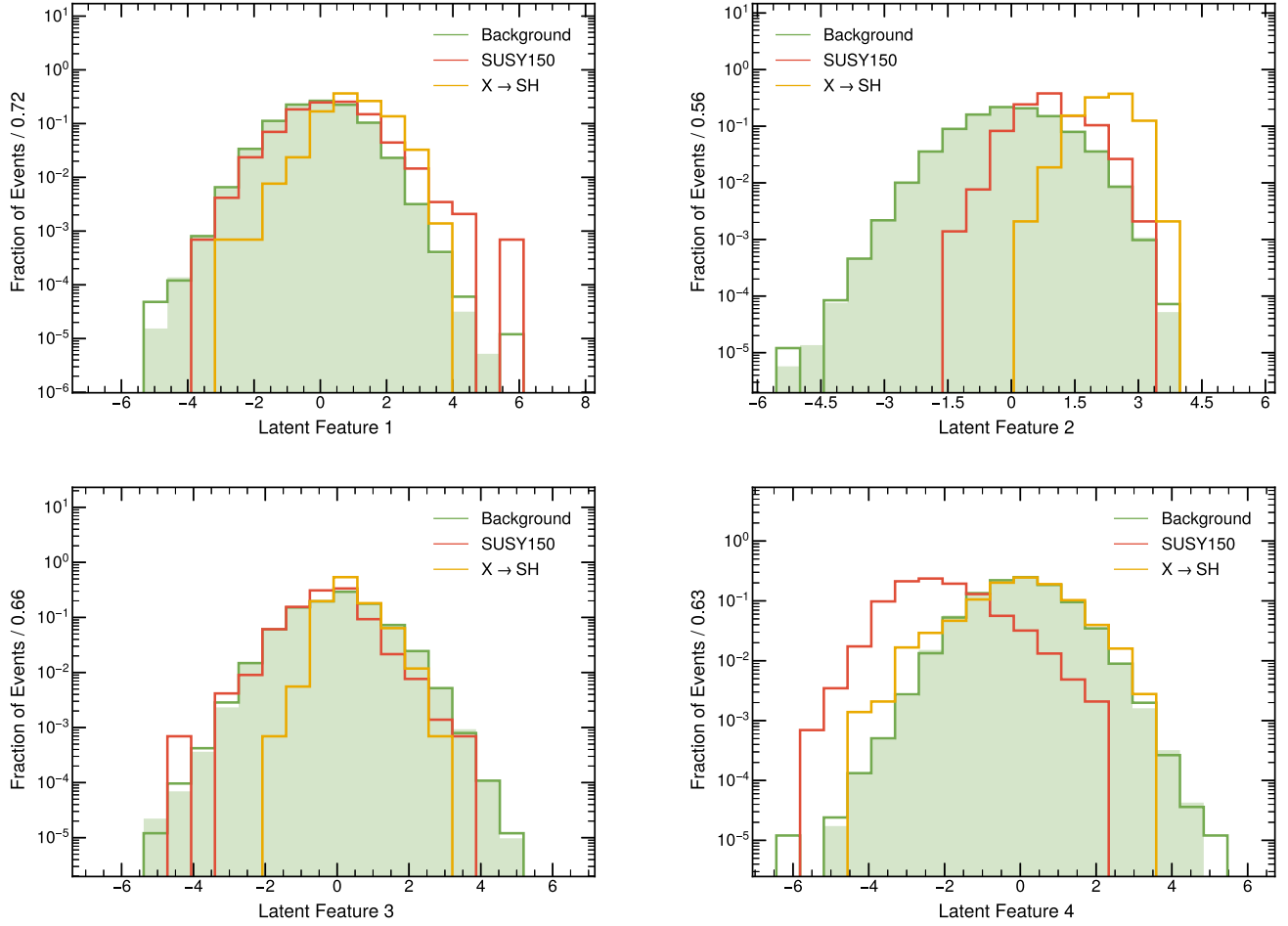


FIG. 3. Comparison between the distributions in the four-dimensional latent space for events in the SR. The shapes of the background in the data (line) and the estimated background (filled) indicate a high accuracy of the background estimation. The shapes of the SUSY150 and $X \rightarrow SH$ signals indicate the separation from the backgrounds in the latent space.

1000 iterations. After the learning rate has been reduced below 10^{-8} , the training terminates. To minimize the impact of random model initialization and batch ordering, we train an ensemble of four NF models, each with a different train and validation split.

Several alternative approaches were tried to improve the performance and accuracy of the generative model. Notably, we investigated the use of conditional flow matching models [70] as well as the use of a classifier network to reweight the generated samples [71]. However, none of the attempted approaches led to a performance improvement over the NF model, and in some instances, the alternate models were found to be more prone to background sculpting.

After the model is trained, we use it to generate VAE-latent space events in the SR. To this end, we need to sample diphoton masses from the SR. We achieve this by fitting an exponential falling function to the diphoton mass spectrum in the SB [72]. The fitted function has the form ae^{-x^c} with the normalized diphoton mass x and the free parameters a and c . The conditional inputs to our model are

then sampled from the fitted function in the SR using rejection sampling. Each NF in the ensemble is used to sample 250,000 events, which are combined into a set of 1 million generated events. This oversampling has been shown to improve the performance of the weakly supervised classifiers. The generated events are then weighted such that the sum of all generated event weights corresponds to the number of expected events we obtain by integrating the fitted diphoton mass function in the SR.

The resonant contribution from the SM Higgs is only present in the SR and can therefore not be modeled with an ML-based and data-driven approach from the SB, like the nonresonant background. Instead, we estimate the contribution of the Higgs process directly from classical simulations as described in Sec. II. In this demonstrator, the resonant background is modeled with different events from the same simulations in the pseudodata and the simulated data—to model the effect of potential differences which might occur in an analysis on recorded data, a 5% normalization uncertainty is added on the total cross section of the simulated Higgs process as described in Sec. III D.

Figure 3 shows a comparison of the distributions of pseudodata and background estimation in the VAE space. It can be seen that the ML technique can model the distribution of the pseudodata. The estimation of the resonant background agrees with the data by construction in this demonstrator analysis, as previously described. Classifiers trained to distinguish between the pseudo-data without any injected signal and the background estimation achieve an area under the curve close to 0.5. This value is consistent with random guessing, meaning the classifier cannot separate the data from generated samples, indicating that the background is accurately modeled.

C. Signal classification

After generating the background reference sample, weakly supervised classifiers are trained to distinguish data from the background reference. As other recent AD strategies [73,74], we employ boosted decision tree (BDT) classifiers for this task. The classifiers are trained with a binary cross-entropy loss and implemented using the XGBOOST package [75]. We observed that reducing the BDT size helps prevent overfitting to small discrepancies between the generated background samples and the data. This avoids background sculpting and improves signal sensitivity. Consequently, the maximum number of estimators is set to 50, and the maximum tree depth is limited to 3. A learning rate of 0.01 is used, and early stopping is applied with a patience of five epochs.

We further employ a five-fold cross-validation strategy to prevent overfitting and use our data as efficiently as possible. We randomly split our data set into five folds, and for each fold, a BDT is trained using three folds as the training set, one fold as the validation set, and one fold as the test set. After training, each BDT is applied to its corresponding test set, and the five test sets are then merged to reconstruct the full dataset. As a result, every data event receives a classifier score from a BDT that was not trained using that event. To reduce fluctuations from limited training statistics and model initialization, we use the mean of ensembles of classifiers. Each classifier ensemble consists of four BDTs, each trained using the five-fold cross-validation, with different data splits and background samples generated from different VAE-generative model pairs.

D. Inference

The statistical interpretation of the data is performed using a PYTHON implementation of the BumpHunter algorithm [76], which searches for localized excesses in the $m_{\gamma\gamma}$ distribution. Since the analysis targets anomalies associated with the Higgs boson, the search window is fixed to the SR of $120 < m_{\gamma\gamma} < 130$ GeV, which avoids the look-elsewhere effect from scanning multiple windows. We perform the bump hunt on $m_{\gamma\gamma}$ histograms with a bin width of 0.5 GeV. To enhance the separation between potential

new physics signals and the SM background, a tight selection is applied. To this end, we place a cut on the classifier score such that we retain the 0.5% most signal-like parts of data in the SR. When no signal is present in the data, weakly supervised classifiers tend to select random regions of the phase space, which leads to nonsmooth distributions. To mitigate this effect, we use the average output from five classifier ensembles, as explained below. Here, a classifier ensemble refers to the ensemble of four individual classifiers as explained in Sec. III C. All classifiers of every classifier ensemble are trained with different data splitting and random model initializations. For each classifier ensemble, we apply the score-cut and build a histogram from the remaining events. The final histogram for the hypothesis test is then obtained by averaging over the histograms from each classifier ensemble. The nonresonant background after the score cut is estimated by fitting the same exponential function as used in the background modeling [72], $ae^{-x \cdot c}$ with the normalized diphoton mass x and the free parameters a and c , to the postcut data in the SB. The resulting function is then interpolated into the SR.

We quantify the significance of any observed excess using 3.5×10^6 pseudoexperiments, incorporating both statistical and systematic uncertainties. The dominant systematic contributions arise from the classifier ensemble and the SM background modeling. A total of 100,000 bootstrapped variations of the classifier ensemble are constructed and used in equal proportion across all pseudoexperiments, to limit computational resources. For each bootstrapped ensemble, the full procedure for deriving the postselection histograms is repeated to obtain the corresponding varied background predictions. The statistical uncertainty on the nonresonant background is incorporated by resampling the sideband data before performing the fit. This results in variations in the fit result, effectively propagating this uncertainty into the background estimate in the SR. To account for theoretical uncertainties in the SM Higgs production cross section and decay branching fractions, the normalization of the resonant background is varied according to a log-normal distribution. For the purposes of this demonstrator, we assume a standard deviation of 5%. A detailed uncertainty calculation of this is beyond the scope of this study. The final background prediction for each pseudoexperiment is the sum of the nonresonant and resonant components, with an additional Poisson fluctuation applied to simulate statistical variations. Finally, the compatibility of the observed data with the background-only hypothesis is determined using a frequentist approach: For each pseudoexperiment, we consider its individual local p-value. The global p-value is calculated as the fraction of pseudoexperiments with a p-value smaller than that observed in the data. This represents the probability of observing such a deviation by chance after considering all statistical and systematic uncertainties.

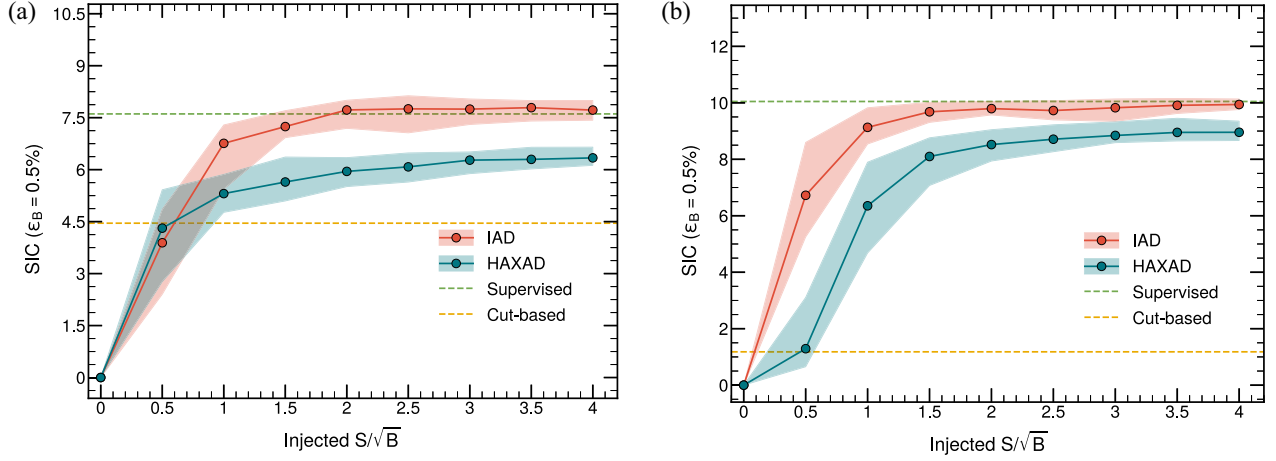


FIG. 4. SIC from a cut on the classifier output over the respective TPR for (a) the SUSY150 and (b) the $X \rightarrow SH$ benchmark signal. The used dataset contains the respective signal with an inclusive preclassifier significance of 1σ . The plot shows the median SIC and the 1σ quantile spread of 50 classifier ensembles, of which each five are trained on a different random instance of the training data.

IV. RESULTS

The HAXAD method is compared against two other methods, the Ideal Anomaly Detector (IAD) [15] and a cut-based approach. In the IAD setup, the background samples are directly drawn from the same simulated samples from which the pseudodata are constructed. In this case, aside from statistical fluctuations, the only distinguishing factor between the data and the background reference is the presence of signal. The IAD setup also uses the nine physical features instead of the four encoded latent space features to train the CWoLa classifiers. It therefore defines the theoretical upper bound of the performance of the HAXAD setup under the assumption of a perfect background estimation and a latent space embedding that is fully efficient to the assumed signal model. The workflows following the classifier training are identical to the ones used in the HAXAD method.

The cut-based method is similar to what was done in the categorical ATLAS $H + X$ analysis [29], which applies one-dimensional cuts on selected kinematic features to define regions with increased signal fraction. Instead of multiple cuts, only one cut is employed for simplicity as $E_T^{\text{Miss}} > 80$ GeV. This cut is optimized for the SUSY150 but not for the $X \rightarrow SH$ benchmark signal. The inference workflow of the cut-based method after the cut is identical to the HAXAD method.

The significance improvement characteristic (SIC) between signals and backgrounds in the data is used to evaluate the performance of the classifiers. Figure 4(a) [Fig. 4(b)] shows the maximum SIC of the respective classifier over the injected amount of the SUSY150 ($X \rightarrow SH$) benchmark signal. The plot shows the median and the 1σ quantile range for 50 classifier ensembles. These ensembles are grouped into sets of 5, with each set trained on one of 10 different random splits of the original dataset.

For each set, the background samples were generated using a unique combination of VAE and generative model. For both tested benchmark models, the significance improvement achieved by the IAD classifier approaches that of the fully supervised classifier. The HAXAD classifier yields a slightly smaller improvement than the IAD classifier, possibly indicating a small information loss in the VAE embedding. Nevertheless, both IAD and HAXAD outperform the cut-based selection for a 1σ signal injection in the SUSY150 scenario, where the cut is optimized, and for even smaller signal injections in the $X \rightarrow SH$ scenario, where the cut is not optimized.

The $m_{\gamma\gamma}$ spectrum after the classifier-based selection, including the bump hunting results, using the SUSY150 ($X \rightarrow SH$) signal model at 1σ signal injection significances, is shown in Fig. 5(a) [Fig. 5(b)], where a global significance of 4.6σ (5.0σ) is observed.

The global significance values are scanned over different signal injection significances using the SUSY150 ($X \rightarrow SH$) benchmark signal are shown in Fig. 6(a) [Fig. 6(b)]. At each signal injection strength, 10 different random instances of the initial datasets are used to get the expected global significance with uncertainty. The HAXAD method is compared with the IAD method and the cut-based method. All three methods show similar performance, with the HAXAD and IAD methods performing slightly better than the cut-based method for signal injection strengths larger than 1σ . The picture changes for the $X \rightarrow SH$ benchmark signal for which the cut-based analysis is not optimized. For this signal model, both the IAD method and the HAXAD method exceed the cut-based method by a large amount for signal injections greater than 0.75σ . For the considered amount of pseudoexperiments in the inference, the upper limit of the global significance is 5σ , and both HAXAD and the IAD are reaching this limit around the 1σ injection level.

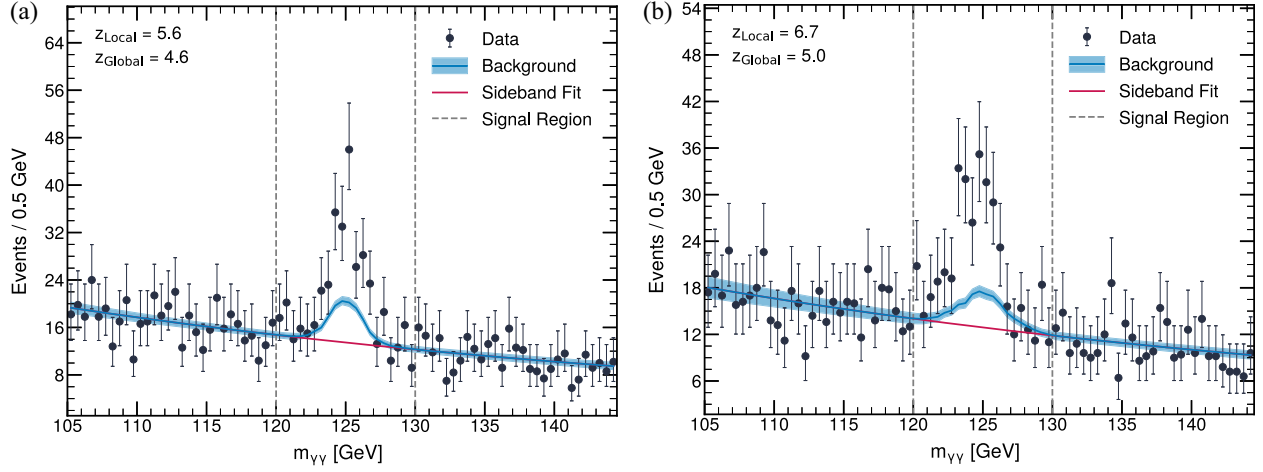


FIG. 5. Distribution of the invariant mass of the two selected photons ($m_{\gamma\gamma}$) after the classifier-based selection in the case of an injected signal with an inclusive 1σ pre-selection significance of (a) the SUSY150 and (b) the $X \rightarrow \text{SH}$ benchmark signal.

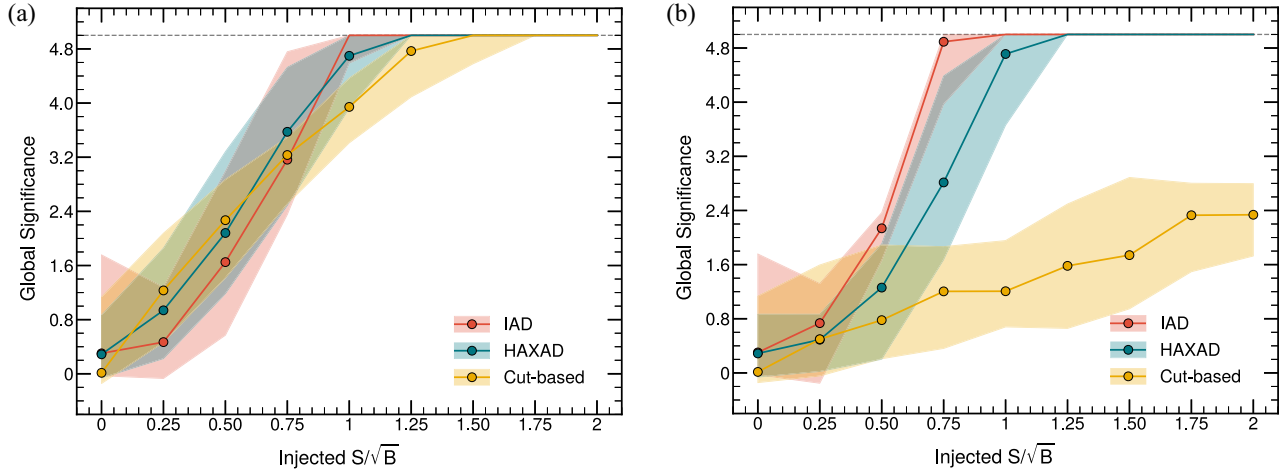


FIG. 6. Observed significances of the HAXAD method in comparison to the IAD and the cut-based method, with $E_T^{\text{Miss}} > 80$ GeV, over different injected pre-selection significances. Shown are the results for (a) the SUSY150 and (b) the $X \rightarrow \text{SH}$ benchmark signal.

for both benchmark signals. On both signal models, and across all signal injection strengths, the HAXAD method reaches a similar performance as the IAD method.

V. CONCLUSION

Model-agnostic AD methods provide an important complement to dedicated searches in high-energy physics. A wide range of AD approaches exist; however, many struggle in the presence of rare processes or resonances that are already accounted for in the SM, as the method can interpret the SM signatures as potential anomalies. This has limited the deployment of AD searches in interesting regions, such as in proximity to the Higgs mass.

We introduced HAXAD, a method for incorporating limited information about the SM into AD searches, and demonstrated its use on BSM signals associated with a SM Higgs boson. We achieved this by combining the data-driven

background estimation of CATHODE-like methods with simulations of SM Higgs events. We were able to demonstrate that a VAE-based encoding can be used to overcome challenges in the generative background estimation, with a limited sacrifice to the sensitivity. Our approach uses a large number of features from each event for the classification and paves the way for future searches to use a feature set that is extended even beyond that, up to and including the use of low-level jet constituents as inputs.

While being fully model agnostic, the new method was benchmarked using two different BSM signals. When injecting the benchmark signals into the dataset with a preselection significance of 1σ , our method was able to achieve a classification with an average significance improvement of 5.3 (6.4) and an average total significance of 4.6σ (4.8σ) for the SUSY150 ($X \rightarrow \text{SH}$) benchmark signal, respectively. For both benchmark signals, HAXAD remains competitive with, and outperforms, a dedicated

cut-based search, while requiring no assumption about the respective signal. As such, HAXAD provides a blueprint for future AD searches around the Higgs peak or other resonances.

The simulation framework, including all used configurations, is available at <https://github.com/hep-lbd/EventGen>. The analysis code is available at https://gitlab.cern.ch/haxad/haxad_demonstrator.

ACKNOWLEDGMENTS

C. L. C., S. Di., B. N., and D. N. are supported by the U.S. Department of Energy (DOE), Office of Science under Contract No. DE-AC02-05CH11231. B. N. and D. N. are supported by U.S. DOE Grant No. DE-AC02-76SF00515. S. D. and R. L. are supported by U.S. DOE Grant

No. DE-SC0017660. B. N. is additionally supported by the John Templeton Foundation. C. L. C. is also supported by the U.S. DOE Office of Science under Contract No. DE-SC0017647. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy User Facility using NERSC Grant No. HEP-ERCAP0032304.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

-
- [1] G. Bertone, D. Hooper, and J. Silk, Particle dark matter: Evidence, candidates and constraints, *Phys. Rep.* **405**, 279 (2005).
 - [2] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, The hierarchy problem and new dimensions at a millimeter, *Phys. Lett. B* **429**, 263 (1998).
 - [3] G. Kasieczka *et al.*, The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics, *Rep. Prog. Phys.* **84**, 124201 (2021).
 - [4] G. Karagiorgi, G. Kasieczka, S. Kravitz, B. Nachman, and D. Shih, Machine learning in the search for new fundamental physics, *Nat. Rev. Phys.* **4**, 399 (2021).
 - [5] T. Aarrestad *et al.*, The dark machines anomaly score challenge: Benchmark data and model independent event classification for the large hadron collider, *SciPost Phys.* **12**, 043 (2022).
 - [6] V. Belis, P. Odagiu, and T. K. Aarrestad, Machine learning for anomaly detection in particle physics, *Rev. Phys.* **12**, 100091 (2024).
 - [7] G. Aad *et al.* (ATLAS Collaboration), Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states using $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector, *Phys. Rev. D* **108**, 052009 (2023).
 - [8] G. Aad *et al.* (ATLAS Collaboration), Search for new phenomena in two-body invariant mass distributions using unsupervised machine learning for anomaly detection at $s = 13$ TeV with the ATLAS detector, *Phys. Rev. Lett.* **132**, 081801 (2024).
 - [9] G. Aad *et al.* (ATLAS Collaboration), Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector, *Phys. Rev. Lett.* **125**, 131801 (2020).
 - [10] G. Aad *et al.* (ATLAS Collaboration), Weakly supervised anomaly detection for resonant new physics in the dijet final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *Phys. Rev. D* **112**, 072009 (2025).
 - [11] V. Chekhovsky *et al.* (CMS Collaboration), Model-agnostic search for dijet resonances with anomalous jet substructure in proton–proton collisions at $\sqrt{s} = 13$ TeV, *Rep. Prog. Phys.* **88**, 067802 (2025).
 - [12] E. M. Metodiev, B. Nachman, and J. Thaler, Classification without labels: Learning from mixed samples in high energy physics, *J. High Energy Phys.* **10** (2017) 174.
 - [13] J. H. Collins, K. Howe, and B. Nachman, Anomaly detection for resonant new physics with machine learning, *Phys. Rev. Lett.* **121**, 241803 (2018).
 - [14] J. H. Collins, K. Howe, and B. Nachman, Extending the search for new resonances with machine learning, *Phys. Rev. D* **99**, 014038 (2019).
 - [15] A. Hallin, J. Isaacson, G. Kasieczka, C. Krause, B. Nachman, T. Quadfasel, M. Schlaffer, D. Shih, and M. Sommerhalder, Classifying anomalies through outer density estimation, *Phys. Rev. D* **106**, 055006 (2022).
 - [16] Y. A. Gol'fand and E. P. Likhtman, Extension of the algebra of poincare group generators and violation of p-invariance, *JETP Lett.* **13**, 323 (1971).
 - [17] D. V. Volkov and V. P. Akulov, Is the neutrino a Goldstone particle?, *Phys. Lett.* **46B**, 109 (1973).
 - [18] J. Wess and B. Zumino, Supergauge transformations in four-dimensions, *Nucl. Phys.* **B70**, 39 (1974).
 - [19] J. Wess and B. Zumino, Supergauge invariant extension of quantum electrodynamics, *Nucl. Phys.* **B78**, 1 (1974).
 - [20] S. Ferrara and B. Zumino, Supergauge invariant Yang-Mills theories, *Nucl. Phys.* **B79**, 413 (1974).
 - [21] A. Salam and J. A. Strathdee, Supersymmetry and non-Abelian gauges, *Phys. Lett.* **51B**, 353 (1974).
 - [22] J. Guasch and J. Sola, FCNC top quark decays: A door to SUSY physics in high luminosity colliders?, *Nucl. Phys.* **B562**, 3 (1999).
 - [23] S. Bejar, J. Guasch, and J. Sola, Loop induced flavor changing neutral decays of the top quark in a general two Higgs doublet model, *Nucl. Phys.* **B600**, 21 (2001).

- [24] G. Eilam, A. Gemintern, T. Han, J. M. Yang, and X. Zhang, Top quark rare decay $t \rightarrow ch$ in R-parity violating SUSY, *Phys. Lett. B* **510**, 227 (2001).
- [25] J. A. Aguilar-Saavedra, Effects of mixing with quark singlets, *Phys. Rev. D* **67**, 035003 (2003); **69**, 099901(E) (2004).
- [26] J. J. Cao, G. Eilam, M. Frank, K. Hikasa, G. L. Liu, I. Turan, and J. M. Yang, SUSY-induced FCNC top-quark processes at the large hadron collider, *Phys. Rev. D* **75**, 075021 (2007).
- [27] F. del Aguila and M. J. Bowick, The possibility of new fermions with $\Delta I = 0$ mass, *Nucl. Phys. B* **224**, 107 (1983).
- [28] J. A. Aguilar-Saavedra, Identifying top partners at LHC, *J. High Energy Phys.* **11** (2009) 030.
- [29] G. Aad *et al.* (ATLAS Collaboration), Model-independent search for the presence of new physics in events including $H \rightarrow \gamma\gamma$ with $\sqrt{s} = 13$ TeV pp data recorded by the ATLAS detector at the LHC, *J. High Energy Phys.* **07** (2023) 176.
- [30] V. Barger, P. Langacker, M. McCaskey, M. Ramsey-Musolf, and G. Shaughnessy, Complex singlet extension of the standard model, *Phys. Rev. D* **79**, 015018 (2009).
- [31] T. Robens, T. Stefaniak, and J. Wittbrodt, Two-real-scalar-singlet extension of the SM: LHC phenomenology and benchmark scenarios, *Eur. Phys. J. C* **80**, 151 (2020).
- [32] I. F. Ginzburg, M. Krawczyk, and P. Osland, Two-Higgs-doublet models with CP violation, Report No. CERN-TH-2002-330, IFT-40-2002, 2002.
- [33] X.-G. He, T. Li, X.-Q. Li, J. Tandean, and H.-C. Tsai, Constraints on scalar dark matter from direct experimental searches, *Phys. Rev. D* **79**, 023521 (2009).
- [34] M. Mühlleitner, M. O. P. Sampaio, R. Santos, and J. Wittbrodt, The N2HDM under theoretical and experimental scrutiny, *J. High Energy Phys.* **03** (2017) 094.
- [35] U. Ellwanger, C. Hugonie, and A. M. Teixeira, The next-to-minimal supersymmetric standard model, *Phys. Rep.* **496**, 1 (2010).
- [36] M. Maniatis, The next-to-minimal supersymmetric extension of the standard model reviewed, *Int. J. Mod. Phys. A* **25**, 3505 (2010).
- [37] G. Aad *et al.* (ATLAS Collaboration), Search for a resonance decaying into a scalar particle and a Higgs boson in the final state with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *J. High Energy Phys.* **11** (2024) 047.
- [38] G. Aad *et al.* (ATLAS Collaboration), The ATLAS experiment at the CERN large hadron collider, *J. Instrum.* **3**, S08003 (2008).
- [39] S. Chatrchyan *et al.* (CMS Collaboration), The CMS experiment at the CERN LHC, *J. Instrum.* **3**, S08004 (2008).
- [40] M. Aaboud *et al.* (ATLAS Collaboration), A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment, *Eur. Phys. J. C* **79**, 120 (2019).
- [41] A. M. Sirunyan *et al.* (CMS Collaboration), MUSiC: A model-unspecific search for new physics in proton-proton collisions at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **81**, 629 (2021).
- [42] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, MadGraph 5: Going beyond, *J. High Energy Phys.* **06** (2011) 128.
- [43] T. Sjostrand, S. Mrenna, and P. Z. Skands, PYTHIA6.4 physics and manual, *J. High Energy Phys.* **05** (2006) 026.
- [44] 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 PYTHIA8.2, *Comput. Phys. Commun.* **191**, 159 (2015).
- [45] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, Matching matrix elements and shower evolution for top-pair production in hadronic collisions, *J. High Energy Phys.* **01** (2007) 013.
- [46] J. de Favereau (DELPHES 3 Collaboration), C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, DELPHES 3, A modular framework for fast simulation of a generic collider experiment, *J. High Energy Phys.* **02** (2014) 057.
- [47] A. Mertens, New features in Delphes 3, *J. Phys. Conf. Ser.* **608**, 012045 (2015).
- [48] P. Fayet, Supersymmetry and weak, electromagnetic and strong interactions, *Phys. Lett.* **64B**, 159 (1976).
- [49] P. Fayet, Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions, *Phys. Lett.* **69B**, 489 (1977).
- [50] G. R. Farrar and P. Fayet, Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry, *Phys. Lett.* **76B**, 575 (1978).
- [51] J. Alwall, P. Schuster, and N. Toro, Simplified models for a first characterization of new physics at the LHC, *Phys. Rev. D* **79**, 075020 (2009).
- [52] D. Alves (LHC New Physics Working Group Collaboration), Simplified models for LHC new physics searches, *J. Phys. G* **39**, 105005 (2012).
- [53] G. Aad *et al.* (ATLAS Collaboration), Search for direct production of electroweakinos in final states with missing transverse momentum and a Higgs boson decaying into photons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *J. High Energy Phys.* **10** (2020) 005.
- [54] P. Basler, S. Dawson, C. Englert, and M. Mühlleitner, Showcasing HH production: Benchmarks for the LHC and HL-LHC, *Phys. Rev. D* **99**, 055048 (2019).
- [55] S. Baum and N. R. Shah, Benchmark suggestions for resonant double Higgs production at the LHC for extended Higgs sectors, *arXiv:1904.10810*.
- [56] A. Tumasyan *et al.* (CMS Collaboration), Search for a massive scalar resonance decaying to a light scalar and a Higgs boson in the four b quarks final state with boosted topology, *Phys. Lett. B* **842**, 137392 (2023).
- [57] A. Tumasyan *et al.* (CMS Collaboration), Search for a heavy Higgs boson decaying into two lighter Higgs bosons in the $\tau\tau b\bar{b}$ final state at 13 TeV, *J. High Energy Phys.* **11** (2021) 057.
- [58] A. Tumasyan *et al.* (CMS Collaboration), Search for a new resonance decaying into two spin-0 bosons in a final state with two photons and two bottom quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV, *J. High Energy Phys.* **05** (2024) 316.
- [59] G. Aad *et al.* (ATLAS Collaboration), Search for a new heavy scalar particle decaying into a Higgs boson and a new scalar singlet in final states with one or two light leptons and a pair of τ -leptons with the ATLAS detector, *J. High Energy Phys.* **10** (2023) 009.

- [60] M. Cacciari and G. P. Salam, Dispelling the N^3 myth for the k_t jet-finder, *Phys. Lett. B* **641**, 57 (2006).
- [61] M. Cacciari, G. P. Salam, and G. Soyez, FastJet user manual, *Eur. Phys. J. C* **72**, 1896 (2012).
- [62] M. Cacciari, G. P. Salam, and G. Soyez, The anti- k_t jet clustering algorithm, *J. High Energy Phys.* **04** (2008) 063.
- [63] D. P. Kingma and M. Welling, Auto-encoding variational Bayes, [arXiv:1312.6114](#).
- [64] A. Paszke *et al.*, Pytorch: An imperative style, high-performance deep learning library, [arXiv:1912.01703](#).
- [65] D. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](#).
- [66] A. Hallin, G. Kasieczka, T. Quadfasel, D. Shih, and M. Sommerhalder, Resonant anomaly detection without background sculpting, *Phys. Rev. D* **107**, 114012 (2023).
- [67] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, *J. Mach. Learn. Res.* **22**, 2617 (2021).
- [68] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, nflows: Normalizing flows in PyTorch (Zenodo, 2020), [10.5281/zenodo.4296287](#).
- [69] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, Neural spline flows, [arXiv:1906.04032](#).
- [70] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, Flow matching for generative modeling, [arXiv:2210.02747](#).
- [71] S. Diefenbacher, E. Eren, G. Kasieczka, A. Korol, B. Nachman, and D. Shih, DCTRGAN: Improving the precision of generative models with reweighting, *J. Instrum.* **15**, P11004 (2020).
- [72] ATLAS Collaboration, Recommendations for the modeling of smooth backgrounds, ATL-PHYS-PUB-2020-028, 2020.
- [73] T. Finke, M. Hein, G. Kasieczka, M. Krämer, A. Mück, P. Prangchaikul, T. Quadfasel, D. Shih, and M. Sommerhalder, Tree-based algorithms for weakly supervised anomaly detection, *Phys. Rev. D* **109**, 034033 (2024).
- [74] M. Freytsis, M. Perelstein, and Y. C. San, Anomaly detection in the presence of irrelevant features, *J. High Energy Phys.* **02** (2024) 220.
- [75] T. Chen and C. Guestrin, Xgboost: A scalable tree boosting system, *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), [arXiv:1603.02754](#).
- [76] L. Vaslin, S. Calvet, V. Barra, and J. Donini, pyBump-Hunter: A model independent bump hunting tool in PYTHON for high energy physics analyses, *SciPost Phys. Codebases* **2023**, 15 (2023).