



OPEN ACCESS

EDITED BY

Thea Aarrestad,
European Organization for Nuclear Research
(CERN), Switzerland

REVIEWED BY

David Shih,
Rutgers University, United States
Raffaele D'Agno,
Commissariat à l'Energie Atomique et aux
Energies Alternatives (CEA), France

*CORRESPONDENCE

John Andrew Raine
✉ john.raine@unige.ch

†These authors have contributed equally to this work

SPECIALTY SECTION

This article was submitted to
Big Data and AI in High Energy Physics,
a section of the journal
Frontiers in Big Data

RECEIVED 18 March 2022

ACCEPTED 28 February 2023

PUBLISHED 21 March 2023

CITATION

Raine JA, Klein S, Sengupta D and Golling T
(2023) CURTAINS for your sliding window:
Constructing unobserved regions by
transforming adjacent intervals.
Front. Big Data 6:899345.
doi: 10.3389/fdata.2023.899345

COPYRIGHT

© 2023 Raine, Klein, Sengupta and Golling. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CURTAINS for your sliding window: Constructing unobserved regions by transforming adjacent intervals

John Andrew Raine^{*†}, Samuel Klein[†], Debajyoti Sengupta[†] and Tobias Golling

Département de Physique Nucléaire et Corpusculaire, Université de Genève, Geneva, Switzerland

We propose a new model independent technique for constructing background data templates for use in searches for new physics processes at the LHC. This method, called Curtains, uses invertible neural networks to parameterise the distribution of side band data as a function of the resonant observable. The network learns a transformation to map any data point from its value of the resonant observable to another chosen value. Using Curtains, a template for the background data in the signal window is constructed by mapping the data from the side-bands into the signal region. We perform anomaly detection using the Curtains background template to enhance the sensitivity to new physics in a bump hunt. We demonstrate its performance in a sliding window search across a wide range of mass values. Using the LHC Olympics dataset, we demonstrate that Curtains matches the performance of other leading approaches which aim to improve the sensitivity of bump hunts, can be trained on a much smaller range of the invariant mass, and is fully data driven.

KEYWORDS

machine learning, anomaly detection, invertible neural network, particle physics, bump hunting, new physics, model independent, unsupervised learning

1. Introduction

In the ongoing search for new physics phenomena to explain the fundamental nature of the universe, particle colliders such as the Large Hadron Collider (LHC) provide an unparalleled window into the energy and intensity frontiers in particle physics. Searches for new particles not contained within the Standard Model of particle physics (SM) are a core focus of the physics programme, with the hope to explain observations in the universe which are inconsistent with predictions from the SM, such as dark matter, gravity, and the observed matter anti-matter asymmetry.

Many searches at the LHC target specific models built upon theories which contain new particles with particular attributes. However, these searches are only sensitive to a specific model. Due to the vast space of models which could extend the SM, it is unfeasible to perform dedicated searches for all of them.

One of the cornerstones in the model independent hunt for new physics phenomena at the LHC is the bump hunt, a search for a localised excess on top of a smooth background. The most sensitive observable for the bump hunt is in an invariant mass spectrum which corresponds to the mass of the particle produced at resonance in particle collisions or decays. The invariant mass spectrum comprises non-resonant events, which produce a falling background across all mass values, with

particles appearing as bumps on top of this background. The width of a bump is driven by the decay width of the particle and the detectors resolution. At the ATLAS and CMS Collaborations (ATLAS Collaboration, 2008; CMS Collaboration, 2008) bump hunt techniques are employed to search for new fundamental particles, and were a crucial in the observation of the Higgs boson (ATLAS Collaboration, 2012; CMS Collaboration, 2012). At the LHCb experiment (LHCb Collaboration, 2008), these techniques have also been successfully employed to observe new resonances in composite particles (LHCb Collaboration, 2020, 2021, 2022).

In a bump hunt, the assumption is made that any resonant signal will be localised. With this assumption, a sliding window fit can be performed using a signal region with a side-band region on either side. As the signal is assumed to be localised, the expected background contribution in the signal region can be extrapolated from the two side-bands. The data in the signal region can be compared to the extrapolated background to test for a significant excess. This test is performed across the whole spectrum by sliding the window. In a standard bump hunt, only the resonant observable is used in the sliding window fit to extrapolate the background and test for localised excesses.

However, with the incredible amounts of data collected by the ATLAS and CMS Experiments, and lack of evidence for new particles (ATLAS Collaboration, 2021a,b,c; CMS Collaboration, 2022a,b,c), the prospect of observing a bump on a single spectrum as more data is collected is growing ever more unlikely. Therefore, attention has turned to using advanced machine learning techniques to improve the sensitivity of searches for new physics, and in particular to improving the reach of the bump hunt approach. Such approaches typically utilise additional discriminatory variables for separating signal from background.

If an accurate background template over discriminatory features can be constructed for the signal region, then the classification without labels method (CWOLA) (Metodiev et al., 2017) can be used to extend the bump hunt. As shown in Collins et al. (2019), the data in the side-bands can be used to construct the template for training the classifier if the discriminatory features are uncorrelated with the resonant variable.

In this paper we introduce a new method, Constructing Unobserved Regions by Transforming Adjacent Intervals (CURTAINS). By combining invertible neural networks (INNs) with an optimal transport loss (Rubner et al., 2000; Villani, 2009; Cuturi, 2013), we learn the optimal transport function between the two side-bands, and use this trained network (henceforth, referred to as the “transformer”) to construct a background template by transforming the data from each side-band into the signal region.

CURTAINS is able to construct a background template for any set of observables, thus classifiers can be constructed using strongly correlated observables. These variables provide additional information and are often the best variables for discriminating signal from background and therefore increase the sensitivity of the search. Furthermore, CURTAINS is a fully data driven approach, requiring no simulated data.

In this paper, we apply CURTAINS to a search for new physics processes in dijet events produced at the LHC and recorded by a general purpose detector, similar to the ATLAS or CMS

experiments. We demonstrate the performance of this method using the R&D dataset provided from the LHC Olympics (LHCO) (Kasieczka et al., 2019), a community challenge for applying anomaly detection and other machine learning approaches to the search for new physics (Kasieczka et al., 2021).

We demonstrate that CURTAINS can accurately learn the conditional transformation of background data given the original and target invariant mass of the events. Classifiers trained using the background template provided by CURTAINS outperform leading approaches, and the improved sensitivity to signal processes matches or improves upon the performance in an idealised anomaly detection scenario.

Finally, to demonstrate its applicability to a bump hunt and observing potential new signals, we apply the CURTAINS method in a sliding window approach for various levels of injected signal data and show that excesses above the expected background can be observed without biases or spurious excesses in the absence of a signal process.

2. The dataset

The LHCO R&D dataset comprises two sets of labelled data. Background data from the Standard Model is produced through QCD dijet production, and signal events from the decay of a new particle to two lighter new particles, which each decay to two quarks $W' \rightarrow X (\rightarrow q\bar{q}) Y (\rightarrow q\bar{q})$, where the three new particles have mass $m_{W'} = 3.5$ TeV, $m_X = 500$ GeV, and $m_Y = 100$ GeV. Both samples are generated with Pythia 8.219 (Sjöstrand et al., 2008) and interfaced to Delphes 3.4.1 (de Favereau et al., 2014) for the detector simulation. The reconstructed particles are clustered into jets using the anti- k_r algorithm (Cacciari et al., 2008) using the FastJet package (Cacciari et al., 2012), with a radius parameter $R = 1.0$. Each event is required to have two jets, with at least one jet passing a cut on its transverse momentum $p_T^J > 1.2$ TeV to simulate a jet trigger in the detector.

In total 1 million QCD dijet events and 100,000 signal events are generated. CURTAINS uses all the QCD dijet events as the standard background sample, and in addition doped samples are constructed using all the QCD events and a small number of events from the signal sample from the 100,000 available W' events. The standard benchmark datasets used to assess the performance of CURTAINS comprise the full background dataset, with 0, 500, 667, 1,000, or 8,000 injected signal events.

All event observables are constructed from the two highest p_T jets, with the two jets ordered by their invariant mass, such that J_1 has $m_{J_1} > m_{J_2}$. The studied features include the base set of variables introduced in Nachman and Shih (2020) and applied in Hallin et al. (2022),

$$m_{JJ}, m_{J_1}, \Delta m_J = m_{J_1} - m_{J_2}, \tau_{21}^1, \tau_{21}^2,$$

where τ_{21} is the n -subjettiness ratio of a large radius jet (Thaler and Van Tilburg, 2011), measuring whether a jet has underlying substructure more like a two prong or one decay, and m_{JJ} is the invariant mass of the dijet system. As an additional feature we include

$$\Delta R_{JJ},$$

which is the angular separation between the two jets in $\eta - \phi$ space. This additional feature is included as it can bring additional sensitivity to some signal models. Furthermore, it is strongly correlated with the resonant feature, m_{JJ} , and so including it when training the transformer and classifier provides a stringent test of the CURTAINS method.

The width of the signal region in the sliding window is set to 200 GeV by default, with 200 GeV wide side-bands either side. In this paper, we simplify the sliding window approach by shifting the window by 200 GeV such that there is no overlap between signal regions. This would reduce the sensitivity for cases where the signal peak falls on the boundary of the signal region. We avoid this by defining our bins such that the signal is centred within a signal region. Where the signal is unknown overlapping windows would need to be employed, with a strategy in place to avoid selecting the same data twice in the final analysis. The turn on in the dijet invariance mass spectrum caused by the trigger requirements of both jets is removed by only performing the sliding window scan with signal regions above 3.0 TeV. The full range used for the sliding window scan is up to a dijet invariant mass of 4.6 TeV.

To evaluate the performance of classifiers using this dataset, a k -fold procedure with five folds is employed, using three fifths of the dataset for training, one fifth for validation and one fifth as a hold out set per fold. No optimisation is performed on the hold out sets, and all optimisation criteria are satisfied using the validation set per fold. This ensures all available data are used in a statistical analysis, which is even more crucial in data driven approaches, where statistical precision is key in the search for new physics. The remaining 92,000 signal events not used to construct the doped datasets are used to evaluate the classifier performance, maximising the statistical precision.

3. Method

3.1. CURTAINS

In CURTAINS conditional invertible neural networks (cINNs) (Ardizzone et al., 2019a,b) are employed to transform data points from an input distribution to those from the target distribution. The transformation is conditioned on a function f of the resonant feature m_{JJ} of the input and target data points. Unlike flows (Rezende and Mohamed, 2016; Kobyzev et al., 2021), which use the exact maximum likelihood of transforming data to a desired distribution, usually a multivariate normal distribution, we use an optimal transport loss to train the network to transform data between the two desired distributions.

As the cINN can be used in both directions, the inputs to the conditional function are referred to as the lower and higher values m_{JJ}^{low} and m_{JJ}^{hi} . In the case of a forward pass through the network, m_{JJ}^{low} are the true values of the input data, with m_{JJ}^{hi} the target values, and vice versa in the case of an inverse pass. Furthermore, instead of training the cINN in only the forward direction, we iterate between both the forward and inverse directions to ensure better closure between the output and target distributions and to prevent a bias toward transformations in one direction.

Several different network architectures for the transformer were studied in the development of CURTAINS. The transformers

presented in this paper are built on the invertible transformations introduced in Durkan et al. (2019) which use rational-quadratic (RQ) splines, which are found to be very expressive and easy to train. The conditioning function f is chosen to be

$$f(m_{JJ}^{low}, m_{JJ}^{hi}) = m_{JJ}^{hi} - m_{JJ}^{low}. \quad (1)$$

The features which are to be transformed determine the input and target dimensions of the CURTAINS transformer.

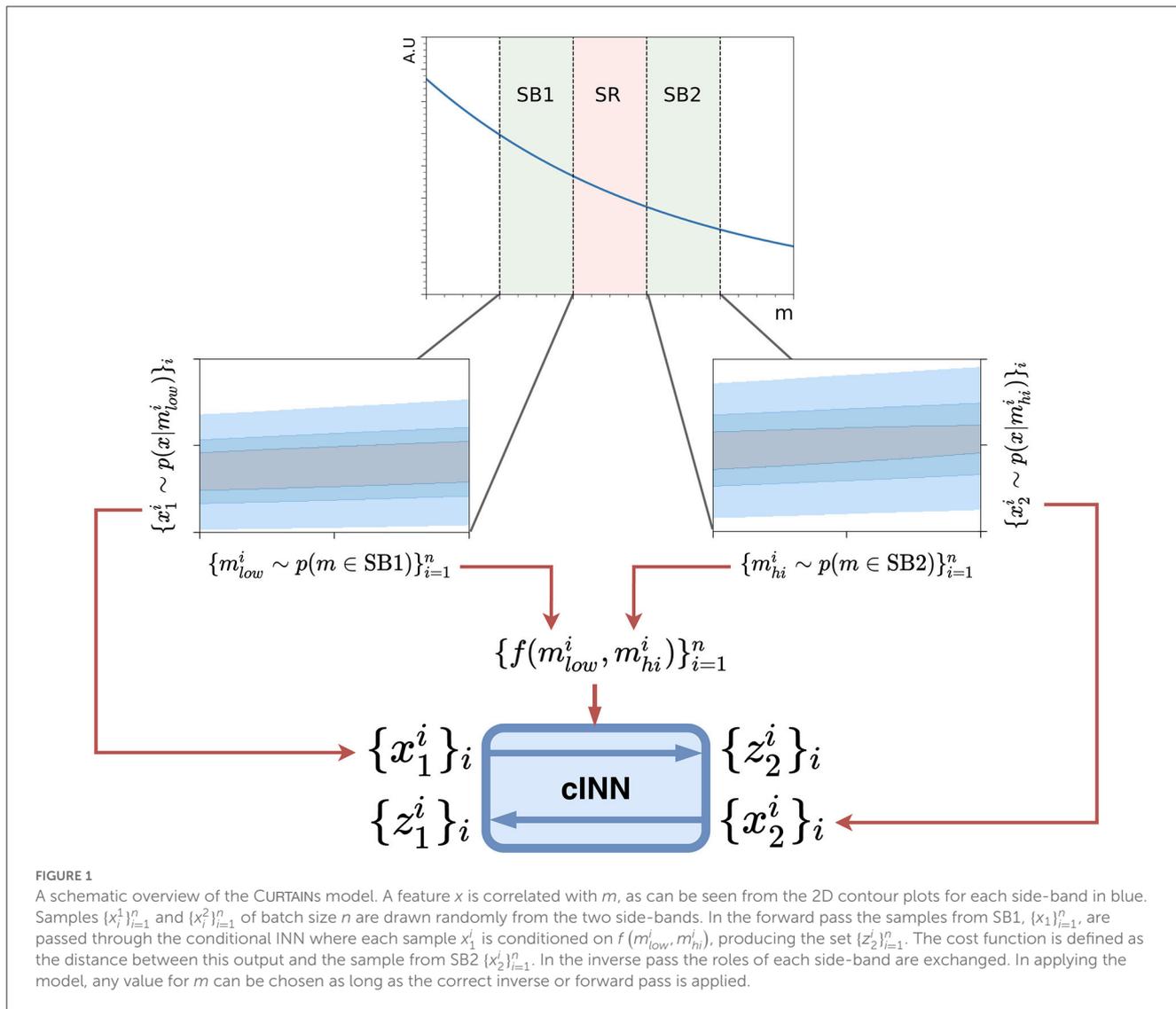
To train the network, batches of data are drawn from the low-mass side-band SB1 and the high-mass side-band SB2. The data from SB1 is first fed through the network in a forward pass, conditioned using m_{JJ}^{low} and m_{JJ}^{hi} , with target values for each event assigned by randomly pairing the masses drawn from each side-band. The loss between the transformed data and target data is calculated using the sinkhorn divergence (Cuturi, 2013) across the whole batch, in order to measure the distance between the distributions of the two sets of data. The gradient of this loss is used to update the network weights. In the next batch the data from SB2 is fed through the network in an inverse pass and the same procedure is performed. This alternating procedure is repeated throughout the training. A schematic overview of the CURTAINS transformer model is shown in Figure 1.

With this training procedure the optimal transport function is not exactly derived as the conditional information is only implicit and a transformed event will not necessarily be paired to the event with the mass to which it was mapped in the loss calculation. However, after training the network we observe that the learned transformation is a good approximation of the true optimal transformation.

In order to improve the closure of the transformed data to regions other than the side-bands, an additional training step is performed. After an epoch of training the network between SB1 and SB2, each side-band itself is split into two equal width sub side-bands. The network is then trained for an epoch of each intra side-band training, following the same procedure as for the inter side-band training. Although not necessary for the CURTAINS method, this extra step is performed in order to extend the range of values of the conditional information used to train the network. Instead of having a minimum value of $f(m_{JJ}^{low}, m_{JJ}^{hi})$ equal to the width of the signal region separating SB1 and SB2, its minimum values is now zero. This ensures that the conditioning variables used to map data to the signal region always lie in the distribution of values used during training.

The CURTAINS transformer is trained for 1,000 epochs with a batch size of 256 using the Adam optimiser (Kingma and Ba, 2017). A cosine annealing learning rate schedule is used with an initial learning rate of 10^{-4} . A typical training time of 6 h using an NVIDIA[®] 3080 RTX GPU is required for a central window encompassing 10^5 samples across the two side-bands.

The CURTAINS transformers are trained separately for each step in the sliding window, using all the available data in the side-bands. In order to construct a background template in another region, all the data from SB1 and SB2 are transformed in either a forward or inverse pass to mass values sampled from the target window. To create the background template in the signal region, the data from SB1 (SB2) are transformed to values of m_{JJ} corresponding to the signal region in a forward (inverse) pass with the CURTAINS



transformer. These two transformed datasets are combined to create the background template in the signal region.

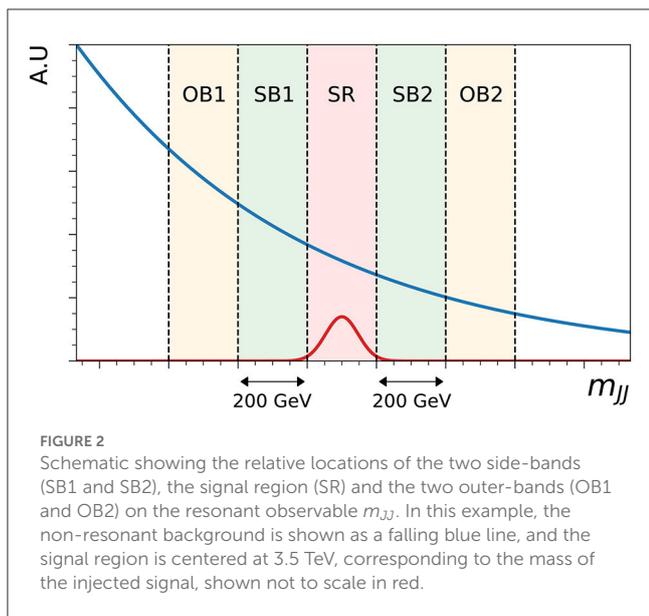
In the case of validating the CURTAINS transformer, the side-band data can be transformed to a target window with the same width as the signal region but going in the opposite direction in m_{JJ} , defining outer-band regions for SB1 (OB1) and SB2 (OB2). These regions can be used to validate and tune the CURTAINS method in a real world setting. The five bands of one sliding window are illustrated in Figure 2, with the depicted signal region centred on the invariant mass of the injected signal. In the studies presented in this paper the width of the side-bands and validation regions is set to 200 GeV by default, unless otherwise specified. To increase the statistics of the constructed datasets the transformer can be applied many times to the same data with different m_{JJ} target values in each pass.

The hyperparameters and architecture of the CURTAINS transformer were optimised in a grid search by measuring the agreement between data transformed into the two outer-band regions from the two side-bands for one step of the sliding window without any doping of signal events. The agreement is

measured by training a classifier to separate the two datasets and ensuring the Receiver Operator Characteristic (ROC) curve had a linear response with an area under the curve close to 0.5, which suggests the network was unable to differentiate between real and transformed data in this region. The optimal CURTAINS transformer is made up of eight stacked RQ spline coupling layers. Each coupling layer is constructed from three residual blocks each of two hidden layers of 32 nodes with LEAKY RELU activations, resulting in an output spline with four bins. The n-flows package (Durkan et al., 2020) is used to implement the network architecture in Pytorch 1.8.0 (Paszke et al., 2019). These settings are then used to train all CURTAINS transformers for each step of the sliding window, and for all doping levels.

3.2. Mass fitting and sampler

In order to sample target values for the CURTAINS transformer and not be biased to the presence of any excess of events in the signal region, the distribution of the resonant feature in the signal



region needs to be extrapolated from the side-band data. Here we model the QCD dijet background with the functional form

$$f(z) = p_1(1 - z)^{p_2} z^{p_3}, \tag{2}$$

where $z = m_{JJ}/\sqrt{s}$ with the centre of mass energy of the collision $\sqrt{s} = 13$ TeV. The parameters p_1 , p_2 , and p_3 are obtained from an unbinned fit to the side-band data in using the `zfit` package (Eschle et al., 2020). This Ansatz has been used previously in analyses performed at the LHC (ATLAS Collaboration, 2016) and is similar to that used in more recent searches with the omission of the last free parameter (CMS Collaboration, 2018; ATLAS Collaboration, 2020). Once fit to the side-band data, the learned parameters are used in the PDF from which to sample target m_{JJ} values for the transformer.

3.3. Anomaly detection

Once the background data has been transformed into the signal region from the side-bands, it is possible to use them as the background template to test for the presence of signal in the data from the region. There are several approaches which could be used for anomaly detection with the data transformed with the CURTAINS method, however in this paper we will focus on the CWOLA classifier, as applied also in Collins et al. (2019), Benkendorfer et al. (2021), and Hallin et al. (2022) on this dataset.

For a CWOLA classifier, it can be shown that the performance of a classifier trained on two sets of data, each containing a different mixture of signal and background data will result in the optimal classifier trained on pure sets of signal and background data. Here, we assume our transformed data represents a sample of pure background events, and test the hypothesis that in our signal region data there is a mixture of signal and background data. In the presence of signal events in the signal region, the

classifier will be able to separate the signal region data from the background template, with the true signal events having higher classification scores than the true background data. By applying a cut on the classifier output to reject a given fraction of the background, calculated from the scores of the background template, the significance of the signal events can be enhanced.

In cases where there is signal contamination in at least one of the side-bands of the sliding window, the background template constructed with CURTAINS will also contain a non-zero signal to background fraction. With the assumption that the signal is localised, and the bin widths are not too small, the relative fraction of signal in the signal region will be different from the background template. As such, the CWOLA method will still be able to approach the performance of the ideal classifier. The background template provided by CURTAINS will have a lower signal to background ratio than the signal region in at least one step of the sliding window, and in this bin an excess can be expected.

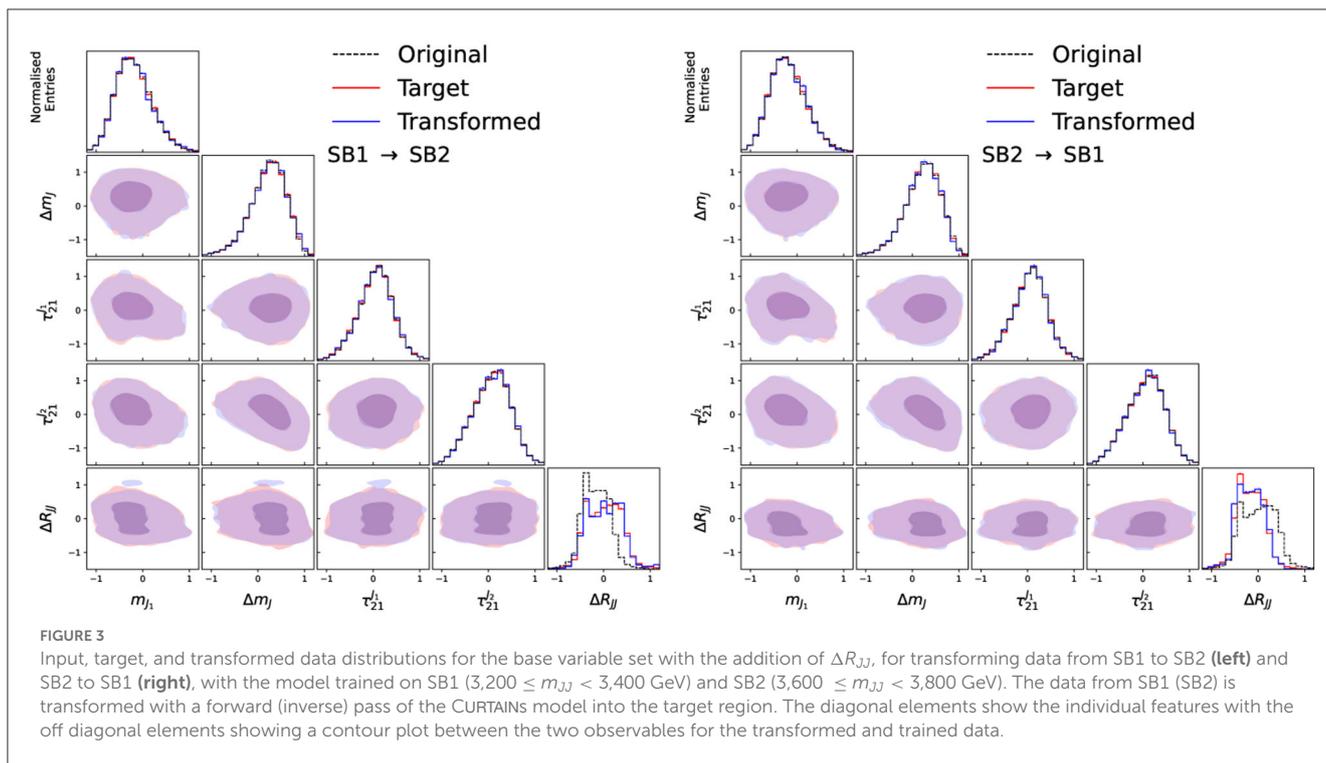
In the event of the signal being fully localised within a side-band, this will result in the opposite labels being used in the training of the CWOLA classifier with regards to which dataset contains the higher fraction of signal. After applying a cut on the classifier a slight reduction in events with respect to the prediction could therefore be expected. However, in practise we observe no significant deviation with the dataset under consideration.

The values used as acceptance thresholds on the output of the classifier are independently determined for each classifier in the signal regions across all sliding windows and levels of doping. These cuts are used to enhance the sensitivity to the presence of signal data in each window of the fit. The amount of data which remains after the cut can be compared to the expected background, determined by taking the total number of data in each signal region multiplied by the background rejection factor. In the presence of a signal, a significant excess of data will be observed above the expected background.

A further test of the performance when using the CURTAINS method is to compare against three benchmark classifiers. The first is a fully supervised classifier, trained with knowledge of which events were from the signal process and which were QCD background. Two further classifiers, the idealised classifiers, are trained in the same manner as with the CURTAINS background template, except that the background template comprises true background data from the signal region itself.

Both the supervised and idealised classifiers are only trained for the window in which the signal region is aligned with the peak of the signal data. The supervised classifier provides an upper bound on the achievable performance on the dataset. The idealised classifier sets the target level of performance which can be achieved with a perfect background template, and can be used to validate the performance of CURTAINS for use in anomaly detection.

All the classifiers for all signal regions and all levels of signal doping share the same architecture and hyperparameters. The classifiers used in this paper have been chosen as they are robust to changes in datasets and initial conditions, in particular when using k -fold training and low training statistics. The classifiers are constructed from multilayer perceptrons with three hidden layers with 32 nodes and RELU activations. The classifiers are trained for 20 epochs using the Adam optimiser with a batch size of 128, and



an initial learning rate of 0.001 which anneals to zero following a cosine curve over 20 epochs.

4. Comparison to other work

Our method is one of several approaches with aims to enhance the sensitivity to new physics processes coming from the resonant production of a new particle using machine learning (Collins et al., 2019; Andreassen et al., 2020; Nachman and Shih, 2020; Benkendorfer et al., 2021; Hallin et al., 2022).

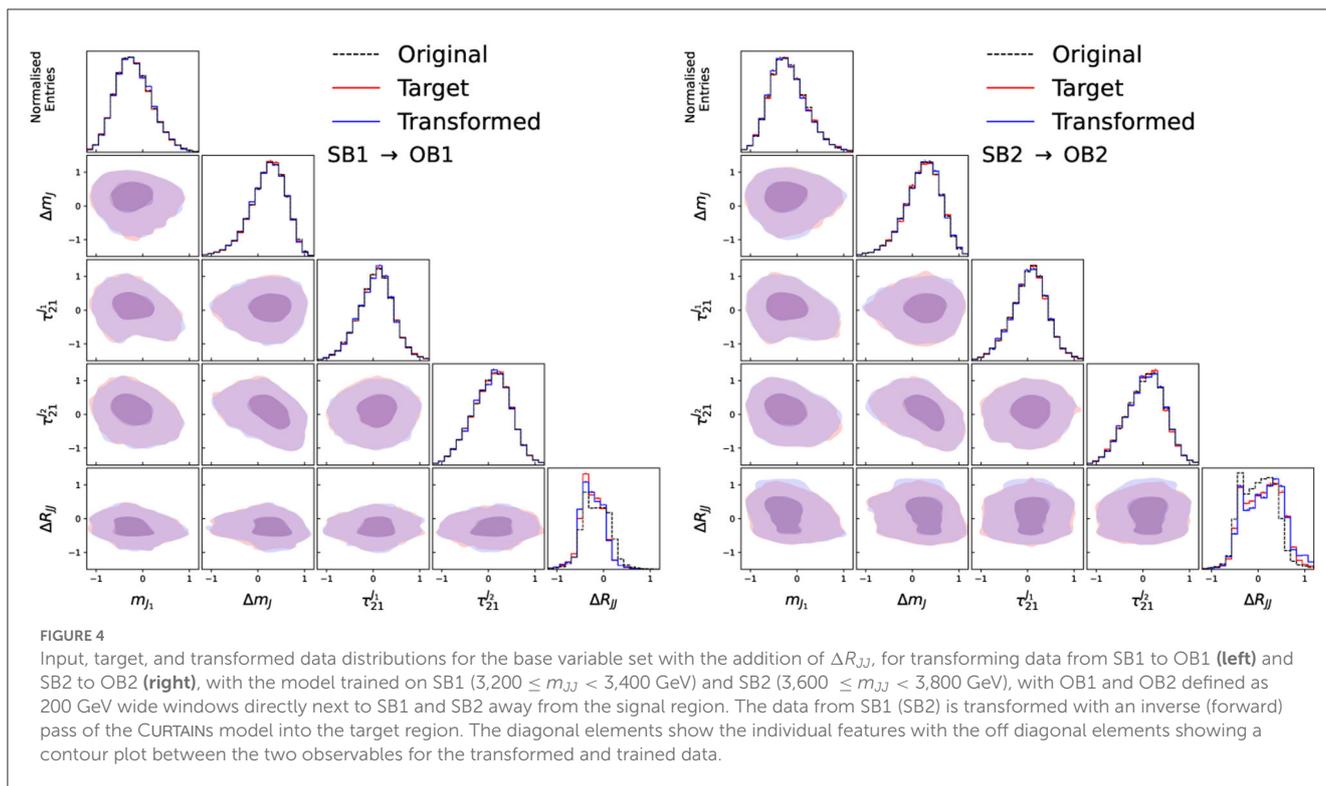
In comparison to the CATHODE method introduced in Hallin et al. (2022), which is one of the current best anomaly detection methods for resonant signals using the CWOLA approach, our method shares some similarities but differs on key points. Although both approaches make use of INNs, CURTAINS does not train a flow with maximum likelihood but instead uses an optimal transport loss in order to minimise the distance between the output of the model and the target data, with the aim to approximate the optimal transport function between two points in feature space when moving along the resonant spectrum. As a result, CURTAINS does not generate new samples to construct the background template, but instead transforms the data in the side-bands to equivalent datapoints with a mass in the signal region. This approach avoids the need to match data encodings to an intermediate prior distribution, normally a multidimensional gaussian distribution, which can lead to mismodelling of underlying correlations between the observables in the data if the trained posterior is not in perfect agreement with the prior distribution. The CATHODE method has no regularisation on the model’s dependence on the resonant variable, and this dependence is non trivial, so extrapolating to unseen datapoints—such as the signal region—can be unreliable.

In contrast, the CURTAINS method can be constructed such that at evaluation the conditioning variable is never outside of the values seen from the training data.

Furthermore, in comparison to CATHODE, CURTAINS is designed to be trained only in the sliding window with all information extracted over a narrow range of the resonant observable, as is standard in a bump hunt. This means CURTAINS is less sensitive to effects from multiple resonances on the same spectrum, and is not dominated by areas of the distribution with more data. Furthermore, thanks to the optimal transformation learned between the side-bands, it can also be applied to transform side-band data into additional validation regions and not just to construct the background template in the signal region.

In contrast to the methods proposed in Andreassen et al. (2020) (SALAD) and Benkendorfer et al. (2021) (SA-CWOLA), CURTAINS does not rely on any simulation and is a completely data-driven technique. In CURTAINS the side-band data is able to be transformed directly into the signal region, instead of deriving a reweighting between the data and simulated data from the side-bands, which is subsequently applied to transform the simulated data in the signal region into a background template. Due to the resampling of the value of the resonant observable, CURTAINS is also able to produce a background template with additional statistics, rather than being limited by the number of events in the signal region from the simulated sample.

There are also a wide range of approaches looking for new physics that do not rely on resonant signals. Many techniques are built on autoencoders (Aguilar-Saavedra et al., 2017; Blance et al., 2019; Cerri et al., 2019; Heimel et al., 2019; Roy and Vijay, 2019; Farina et al., 2020; Hajer et al., 2020; Jawahar et al., 2022), looking to identify uncommon events or objects. These models are subsequently used to reject SM-like processes in favour of



potential new physics. Other approaches are motivated from the ratio of probability densities and directly measure a test statistic from the comparison of a sample of events with respect to a set of reference distributed events (D’Agnolo and Wulzer, 2019; Simone and Jacques, 2019; D’Agnolo et al., 2021; Letizia et al., 2022). A comparison of a wide range of methods is also performed in Kasieczka et al. (2021), which summarises a community challenge for anomaly detection in high energy physics.

5. Results

5.1. Validating CURTAINS transformer

The first test of performance in CURTAINS is to demonstrate that the transformation learned between the two side-bands is accurate, and further to determine whether the learned transformation can extrapolate well to the validation regions. As Monte Carlo simulation is being used for the studies, we can control the composition of the samples in the studies. The performance of the approach is evaluated using a sample containing only background data, as well as various levels of signal doping. The same model configuration is used for all samples, and the sliding window is chosen such that it is centred on the true signal peak with a signal region width of 200 GeV.

The input features and their correlations for the input, target and transformed data distributions are shown in Figure 3 for the two side-bands trained in the case of no signal, and in Figure 4 for the two validation regions. As can be seen, the transformed data distributions are well reproduced with the CURTAINS approach. The ability of CURTAINS to handle features which are strongly correlated with m_{JJ} can be seen from the agreement of the ΔR_{JJ}

distributions between SB1 and SB2, which exhibit very different shapes.

In the case of no signal being present, we can also verify whether the background template constructed by transforming data from the side-bands with CURTAINS matches the target data in the signal region. The performance of the CURTAINS method can be seen in Figure 5, with the transformed data closely matching the data distributions and correlations.

To quantify the level of agreement between the transformed distributions and the target data, classifiers are trained to separate the two datasets, and the area under the ROC curve is measured. The level of agreement between the CURTAINS transformed data and target data can be seen for several levels of signal doping in Table 1. We can see that CURTAINS has very good agreement with the target distribution in all signal regions and in all cases is seen to be better than in the validation region. The reduced performance in OB1 and OB2 is a result of the transformer extrapolating outside of the trained sliding window. This demonstrates their ability to be used for validating the CURTAINS method and future classification architectures.

5.2. Application to anomaly detection

To demonstrate the performance of CURTAINS to produce a robust background template, the sliding window is centred on the resonant mass of the signal events, and the performance of the CWOLA classifier is compared against a background template produced using the CATHODE method. The signal region width is set to 400 GeV to contain the majority of the signal events, resulting in 120,000 background events. The

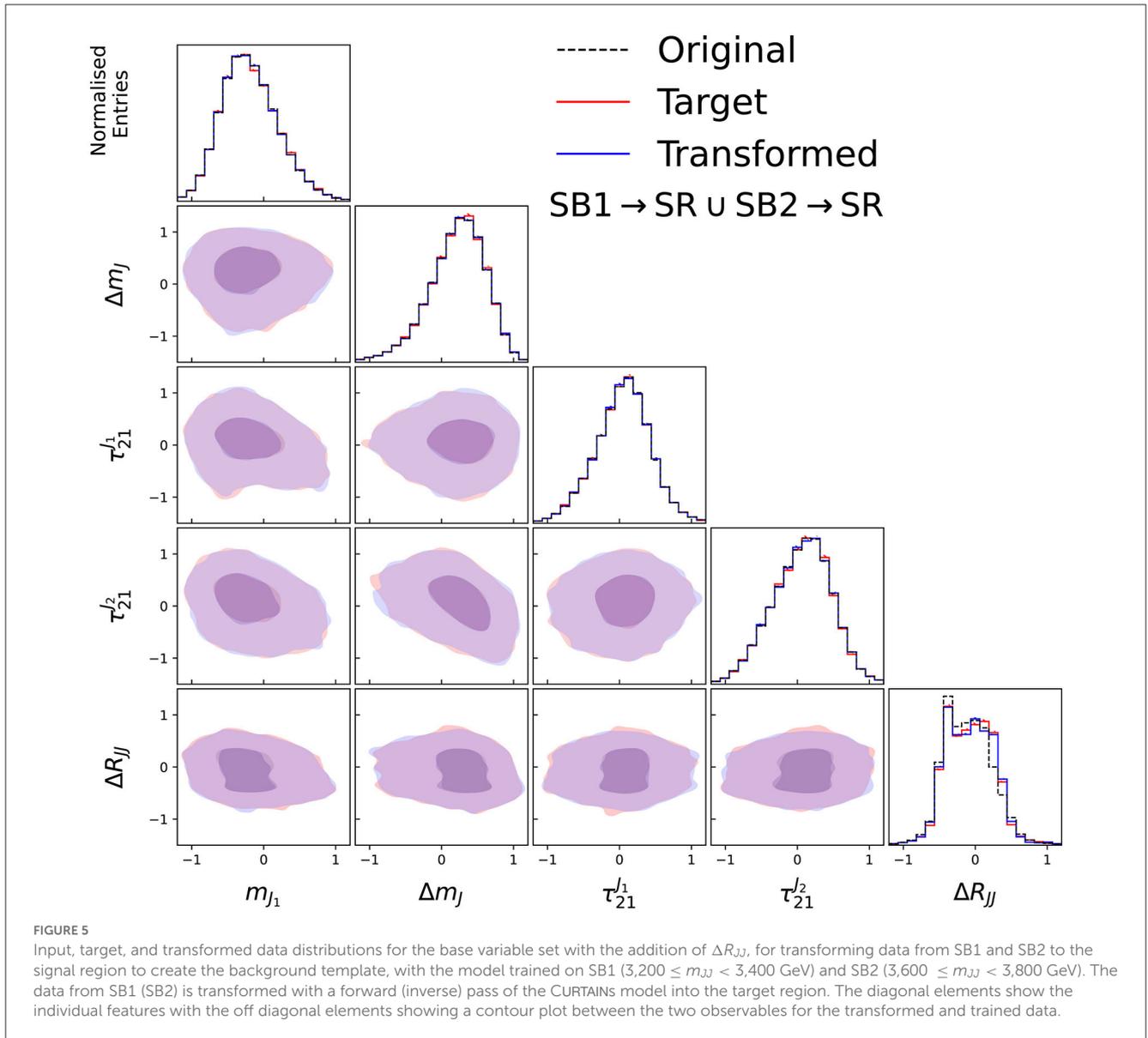


TABLE 1 Quantitative agreement between the data distributions of the transformed data and the target data as measured by the AUC of the ROC curve trained on the two samples, as measured for various levels of signal doping with a 200 GeV wide signal region.

| | SB1→SB2 | SB2→SB1 | SB1→OB1 | SB2→OB2 | SB1→SR ∪ SB2→SR |
|--------------|---------|---------|---------|---------|-----------------|
| 0 signal | 0.504 | 0.504 | 0.519 | 0.512 | 0.509 |
| 500 signal | 0.503 | 0.503 | 0.518 | 0.506 | 0.506 |
| 667 signal | 0.505 | 0.504 | 0.516 | 0.514 | 0.505 |
| 1,000 signal | 0.499 | 0.502 | 0.520 | 0.502 | 0.512 |
| 8,000 signal | 0.508 | 0.511 | 0.523 | 0.521 | 0.522 |

background template is produced with oversampling, with a total of nine times the number of expected events in the signal region. Two comparisons to CATHODE can be performed, one using the same training windows as for CURTAINS, which we refer to as CATHODE (local), and one using the full invariant mass distribution outside the signal region, as

presented in Hallin et al. (2022), which we refer to as CATHODE (full).

For reference, the methods are compared to a classifier trained using an idealised background template and to a fully supervised classifier. The idealised background template constructed using true background events from the signal region, and the supervised

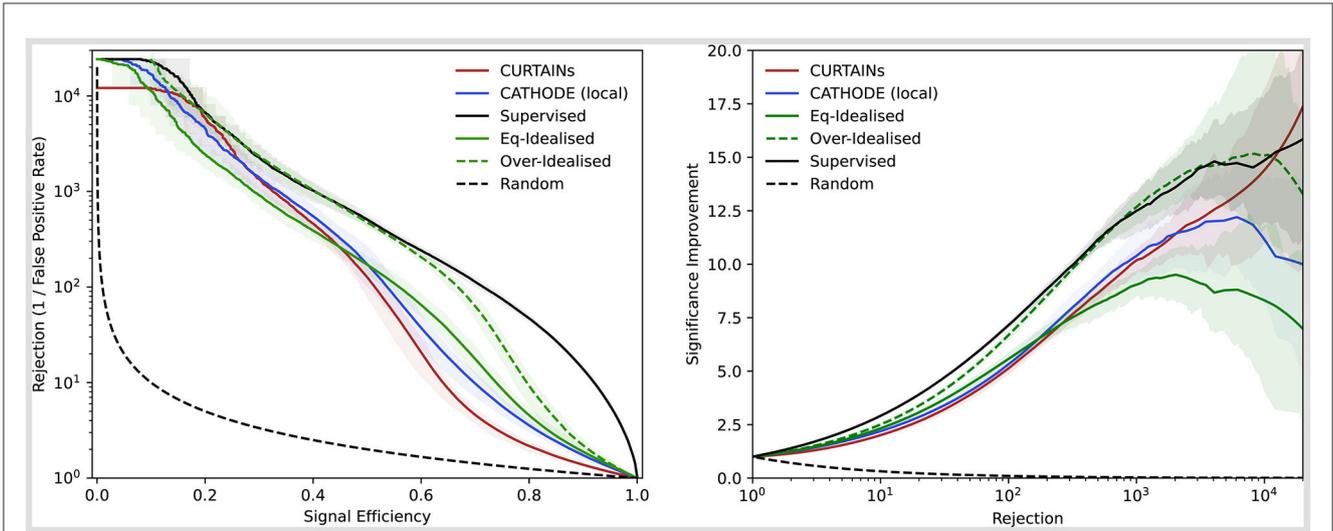


FIGURE 6 Background rejection as a function of signal efficiency (**left**) and signal improvement as a function of background rejection (**right**) for the different background template models (CURTAINS—red, CATHODE—blue, Eq-Idealised—green, Over-Idealised—dashed green) and a fully supervised classifier (black). The sample with 3,000 injected signal events in used to train all classifiers in the signal region $3,300 \leq m_{JJ} < 3,700$ GeV. The solid lines show the mean value of fifty classifier trainings with different random seeds. The uncertainty encompasses 68% of the runs either side of the mean.

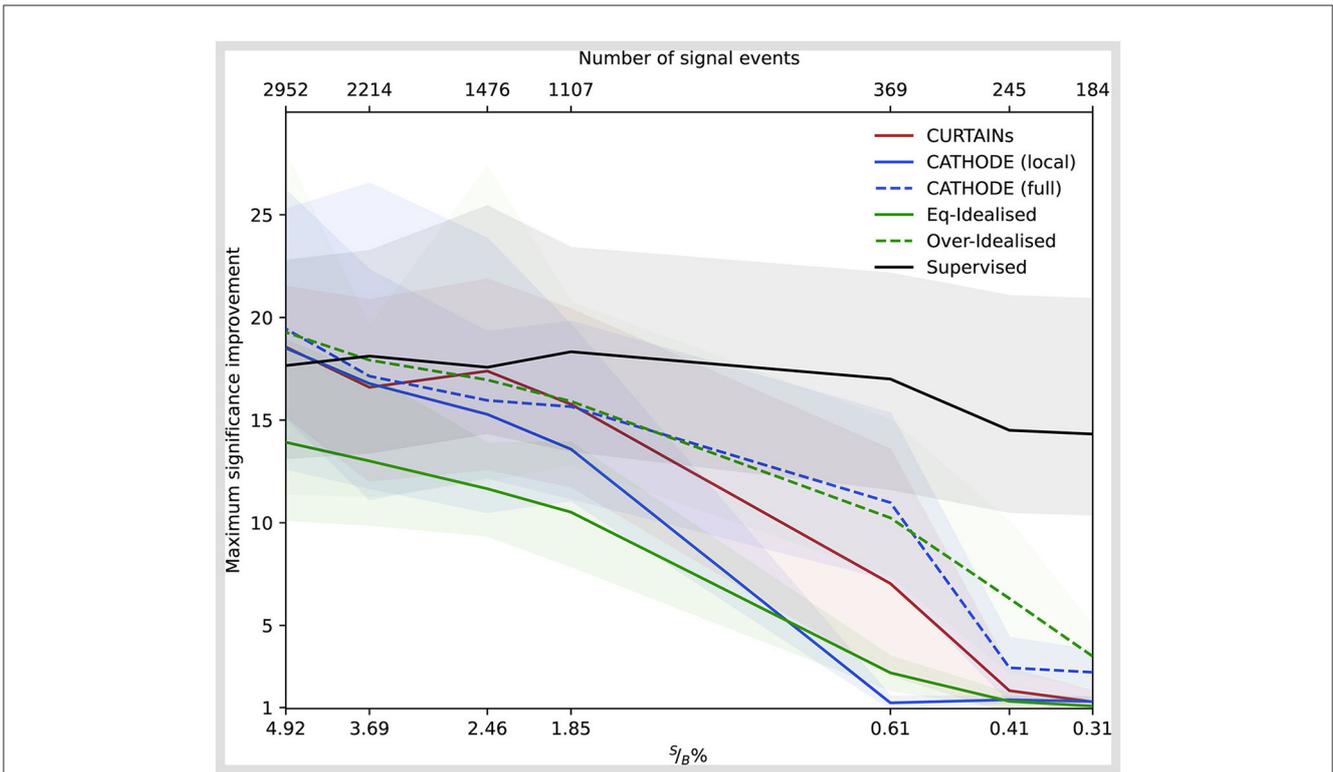
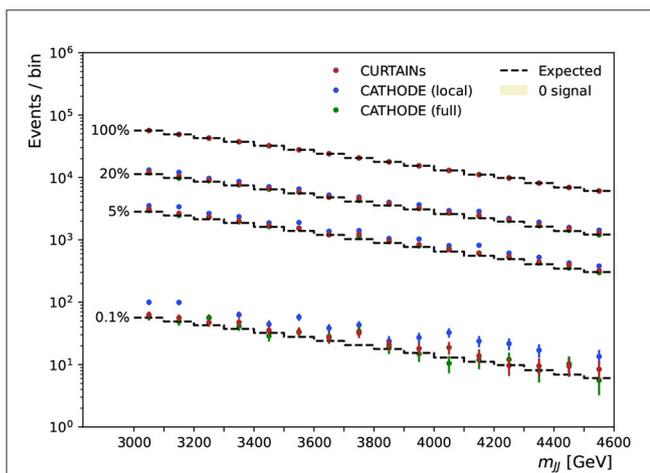
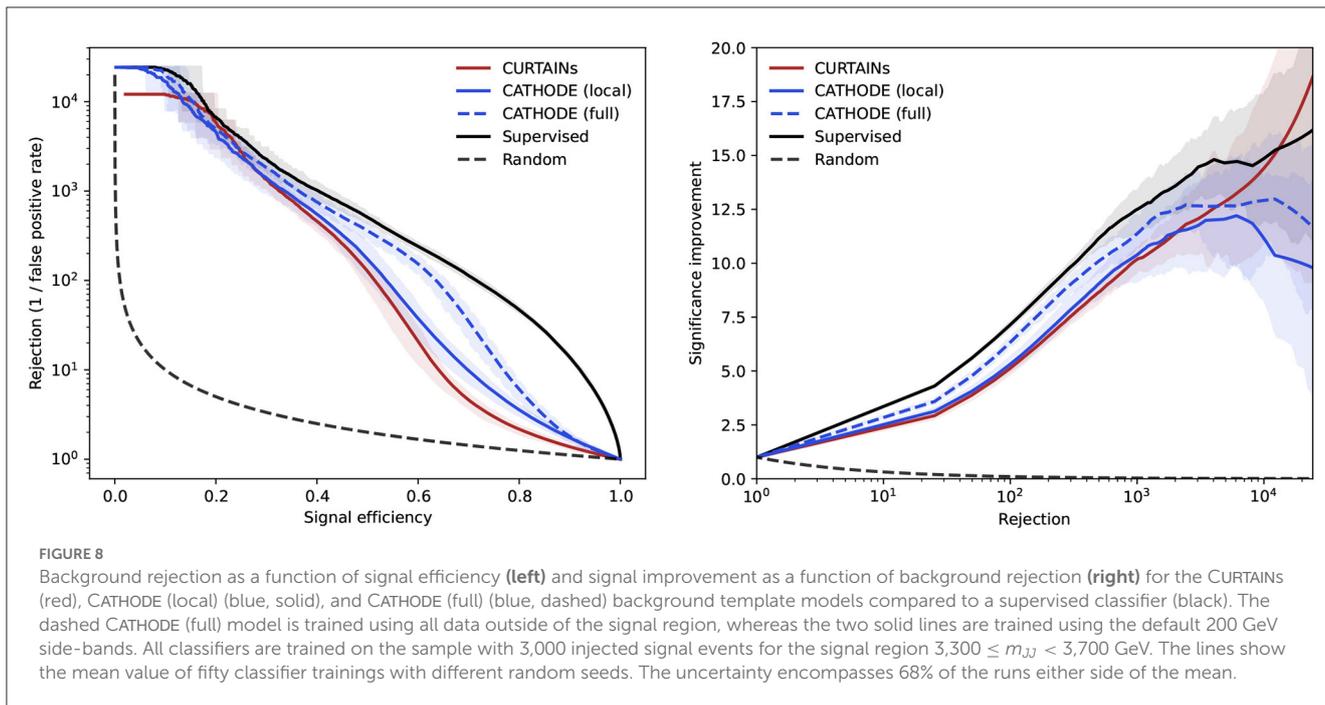


FIGURE 7 The significance improvement as a function of decreasing signal purity (raw signal events) for the different background template models [CURTAINS—red, CATHODE (local)—blue, Eq-Idealised—green, Over-Idealised—dashed green] and a fully supervised classifier (black). All classifiers trained in the signal region $3,300 \leq m_{JJ} < 3,700$ GeV for varying levels of signal doping. The solid lines show the mean value of fifty classifier trainings with different random seeds. The uncertainty encompasses 68% of the runs either side of the mean.

classifier is trained to separate the signal data from the background data using class labels. To construct the idealised background dataset we either use an equal number of background data as there

are in the signal region (Eq-Idealised) to measure the performance assuming we had access to a perfect model of the background data, or the same number of data points as are produced with



the CURTAINS and CATHODE approaches (Over-Idealised), which should approach the best possible performance for models which can oversample.

The performance of the classifiers with the different methods are shown for the doped sample with 3,000 injected signal events (of which 2,214 are in the signal region) in Figure 6, comparing the background rejection as a function of signal efficiency and

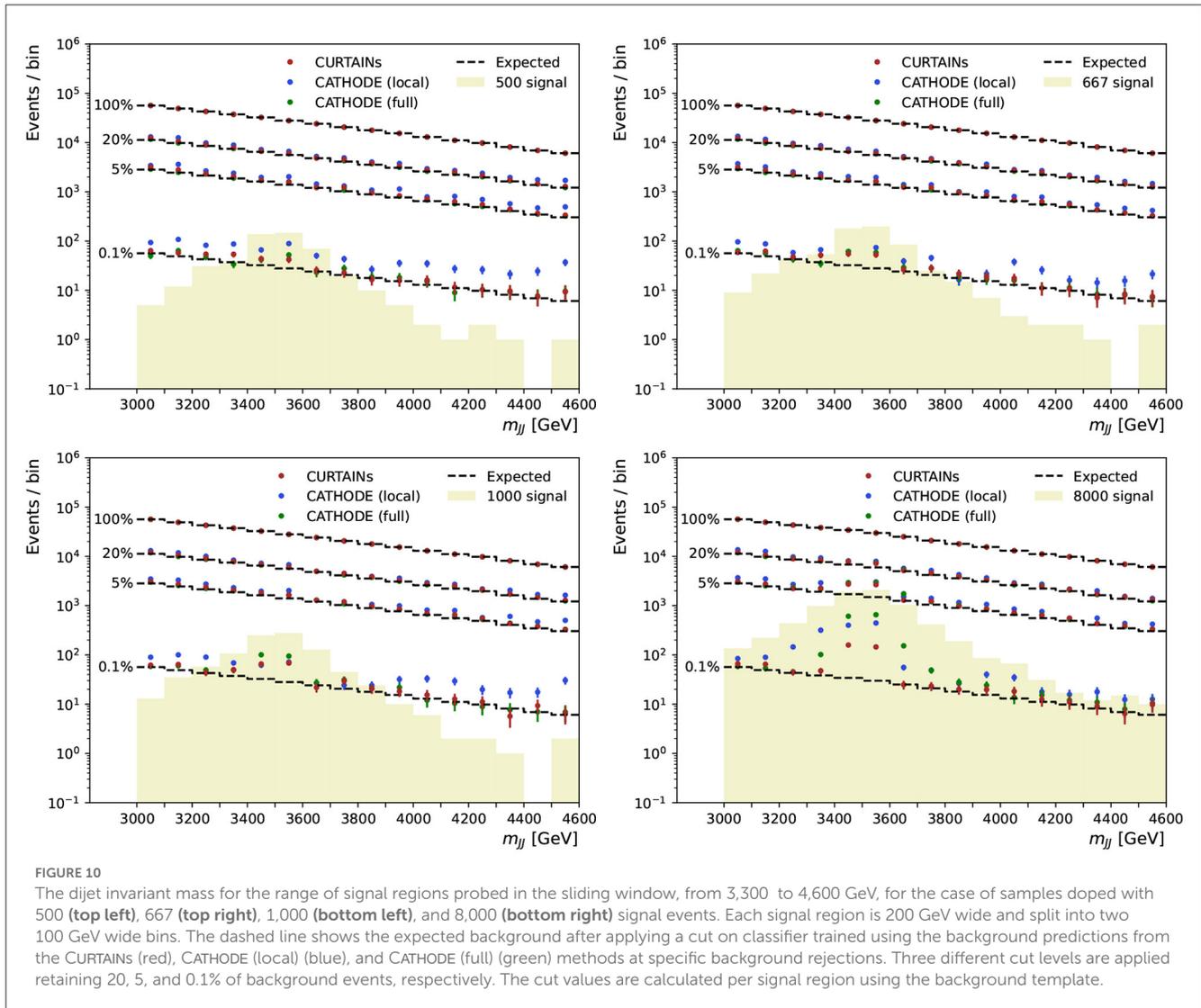
the significance improvement as a function of the background rejection. In order to maintain a fair comparison to the Eq-Idealised classifier, which requires true background data from the signal region for the background template, only half of the available data in the signal region is used for training with the k -fold strategy for all other approaches. The maximum significance improvement is shown for a wide range of doping levels in Figure 7. This metric is a good measure of performance for anomaly detection, rather than the area under the ROC curve, as it translates to the expected performance gain when applying an optimal cut on a classifier.

We can see that CURTAINS not only outperforms CATHODE (local), but also approaches the performance of the Over-Idealised and supervised scenarios. When using the full range outside of the signal regions to train CATHODE (full) the performance recovers and CURTAINS is only able to match the performance at high levels of background rejection, as seen in Figure 8. However, this demonstrates that CURTAINS is able to reach a higher level of performance when trained on lower numbers of events.

5.3. Application in a sliding window

As it is not possible to know the location of the signal events when applying CURTAINS to data, the real test of the performance and robustness of the method is in the sliding window setting.

Both CURTAINS and CATHODE (local and full) are used to generate the background templates in a sliding window scan in the range 3,000–4,600 GeV, with steps of 200 GeV and equal 200 GeV wide signal regions. Classifiers are trained to separate the signal region data from the background template, and cuts are applied to retain 20, 5, and 0.1% of the background events.



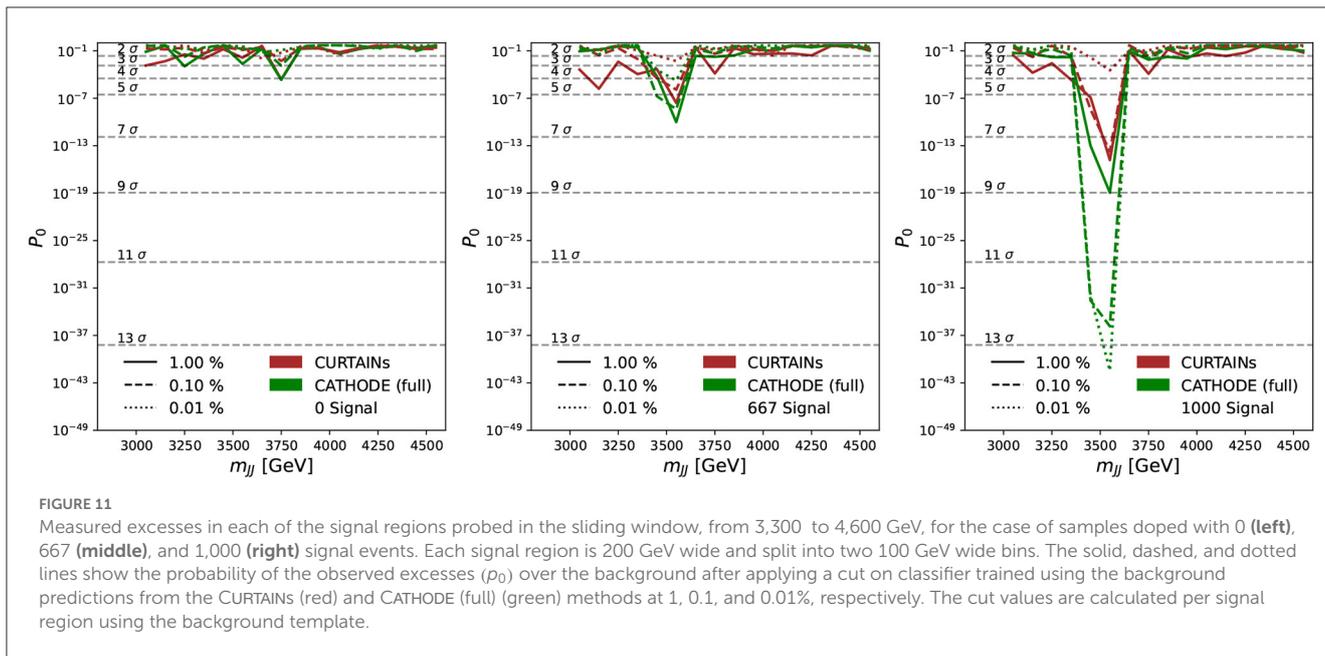
These scans are performed for several levels of signal doping and are shown in Figure 9 for the case where there is no signal present, and in Figure 10 for doped samples with 500, 667, 1,000, and 8,000 injected signal events. Each signal region is subdivided into two bins of equal width in m_{JJ} for the plot. The expected background is determined by multiplying the original yield of each bin by the chosen background retention factor.

In contrast to the CWOLA bump hunt approach introduced in Collins et al. (2019), which uses the classifier trained in the signal region to apply a cut on all events on the invariant mass spectrum before performing a traditional bump hunt, we treat each signal region as an independent region and do not apply the classifiers outside of the regions in which they are trained. This sliding window approach tests how the CURTAINS and CATHODE approaches perform as the sideband windows used to train the networks as well as the signal region transition between the presence of signal, to signal in one of the sidebands as well as the case where there is perfect alignment of signal in the signal region. This approach does not test the ability of the trained classifiers to extrapolate outside of the values of invariant mass used to train them. Were they to be applied outside of the respective regions it is

expected sculpting of the invariant mass distribution would occur after applying cuts on the classifier due to the strong correlation between ΔR_{JJ} and m_{JJ} .

As can be seen from the sliding window scans in Figures 9, 10, using CURTAINS and CATHODE (full) we are able to correctly identify the location of the signal events for even reasonably low levels of signal. Where there are no or very few signal events the yields after each cut do not deviate too far from expected background. The corresponding significance of excesses seen in each bin for three cuts on background efficiency are shown in Figure 11. In the case where there is no signal injected into the sample, both CURTAINS and CATHODE (full) have relatively low local excesses reaching a 4σ deviation only in one bin at the 1% background efficiency and not exceeding 3σ deviations for tighter cuts when considering only statistical uncertainties on the yields in each bin.

However, at lower levels of background rejection, significant local excesses over the expectation are observed. At 5% background efficiency a maximum local deviation is observed at 4σ for CATHODE (full) and 5σ for CURTAINS. In the presence of signal both CURTAINS and CATHODE (full) have observed excesses at



the signal mass peak for each cut level for the lower levels of signal events, with CATHODE (full) approaches results in a more prominent excess.

The CATHODE (local) approach yields an excess across the whole spectrum in the absence of signal and for all levels of injected signal. However, it also finds an excess under the signal peak in the cases where signal is injected, which at higher levels of signal exceeds that found by CURTAINS.

In an analysis a systematic excess over the expectation calculated from the original yields per bin would not necessarily be problematic, as the expectation could be determined from a side band fit in m_{JJ} after applying the cut. Additionally, these values do not take any systematic uncertainties into account and only consider statistical uncertainties on the number of events passing each cut from the yields.

Although the ability to isolate the signal events when using CURTAINS in the window scan decreases at low numbers of signal events and signal purity, this is also seen for both idealised cases in Figure 7 and suggests this as rather an area where the classifier architecture and anomaly detection method need to be optimised. The performance of CURTAINS in this setting could also be further improved by optimising the binning used in the sliding window, and the number of subdivisions within each signal region.

6. Conclusions

In this paper we have proposed a new method, CURTAINS, for use in weakly supervised anomaly detection which can be used to extend the sensitivity of bump hunt searches for new resonances. This method stays true to the bump hunt approach by remaining completely data driven, and with all templates and signal extraction performed on a local region in a sliding window configuration.

CURTAINS is able to produce a background template in the signal region which closely matches the true background

distributions. When applied in conjunction with anomaly detection techniques to identify signal events, CURTAINS matches the performance of an idealised setting in which the background template is defined using background events from the signal region. It also does not produce spurious excesses in the absence of signal events.

As real data points are used with the CURTAINS transformer to produce the background template, we avoid problems which can arise from sampling a prior distribution leading to non perfect agreement over distributions of features and their correlations. By conditioning the transformation on the difference in input and target m_{JJ} , we also avoid the need to interpolate or extrapolate outside of the values seen in training. Using this approach we see CURTAINS is able to reach similar levels of performance in comparison to state-of-the-art methods. CURTAINS delivers this performance even when using much less training data, as seen when using side-bands as opposed to the full data distribution outside of the signal region.

Another key advantage of CURTAINS over other proposed techniques is the ability to apply it to validation regions. By transforming the side-bands data to other regions than the signal region, validation regions can be defined in which the transformer and classifier architectures can be optimised on real data. Here the CURTAINS transformer can be validated and optimised by ensuring the agreement between the transformed data and target data distributions is as close as possible, and the classifier architecture can be optimised to make sure it does not pick up on residual differences between transformed and target data. In this paper, only the former optimisation procedure was performed, with the classifier architecture instead chosen for its robustness to variability in initial conditions.

However, care must be taken to optimise the width of the signal region when training the CURTAINS model to make sure that the signal to background ratio is not constant across the side-band and signal regions.

It may be possible to extend CURTAINS to extrapolation tasks, where a model would be trained on one control region and applied to all other regions. This could allow one model to be trained per bump hunt, or a model could be trained to extrapolate to the tails of distributions, allowing these regions to be probed in a model independent fashion. Thanks to its performance and ability to be applied to a sliding window fit, CURTAINS is simple to apply to current sliding window fits and should bring significant gains in sensitivity in the search for new physics at the LHC and other domains.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: <https://zenodo.org/record/4536377>.

Author contributions

SK and DS: training, optimisation, and modelling studies. JR and TG: conceptualisation. JR: strategy, approach, and editor. All authors have read and agreed on the content this draft and are accountable for the content of the work.

Funding

The authors would like to acknowledge funding through the SNSF Sinergia grant called Robust Deep Density Models for High-Energy Particle Physics and Solar Flare Analysis (RODEM) with funding number CRSII5_193716, and the SNSF project grant 200020_181984 called Exploiting LHC data with machine learning and preparations for HL-LHC.

References

- Aguilar-Saavedra, J. A., Collins, J. H., and Mishra, R. K. (2017). A generic anti-QCD jet tagger. *J. High Energy Phys.* 11:163. doi: 10.1007/JHEP11(2017)163
- Andreassen, A., Nachman, B., and Shih, D. (2020). Simulation assisted likelihood-free anomaly detection. *Phys. Rev. D* 101:095004. doi: 10.1103/PhysRevD.101.095004
- Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E. W., Klessen, R. S., et al. (2019a). *Analyzing Inverse Problems With Invertible Neural Networks*. Available online at: <https://arxiv.org/abs/1808.04730> (accessed March 15, 2022).
- Ardizzone, L., LÄijth, C., Kruse, J., Rother, C., and KÄüthe, U. (2019b). *Guided Image Generation With Conditional Invertible Neural Networks*. Available online at: <https://arxiv.org/abs/1907.02392> (accessed March 15, 2022).
- ATLAS Collaboration (2008). The ATLAS experiment at the CERN large Hadron collider. *J. Instrum.* 3, S08003. doi: 10.1088/1748-0221/3/08/S08003
- ATLAS Collaboration (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* 716, 1–29. doi: 10.1016/j.physletb.2012.08.020
- ATLAS Collaboration (2016). Search for new phenomena in dijet mass and angular distributions from *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B* 754, 302–322. doi: 10.48550/arXiv.1512.01530
- ATLAS Collaboration (2020). Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *J. High Energy Phys.* 3:145.
- ATLAS Collaboration (2021a). *Summary Plots for Heavy Particle Searches and Long-lived Particle Searches - July 2021*. Available

Acknowledgments

The authors would like to thank Matthias Schlaffer, our resident CATHODE Guru, for his invaluable input in establishing a reliable baseline for comparisons and useful discussions, and Knut Zoch for input on the initial studies and samples used. Both Knut and Matthias are also thanked for their feedback on this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fdata.2023.899345/full#supplementary-material>

online at: <https://cds.cern.ch/record/2777015> (accessed March 15, 2022).

ATLAS Collaboration (2021b). *Summary Plots from ATLAS Searches for Pair-Produced Leptoquarks - June 2021*. Available online at: <https://cds.cern.ch/record/2771726> (accessed March 15, 2022).

ATLAS Collaboration (2021c). *SUSY Summary Plots - June 2021*. Available online at: <https://cds.cern.ch/record/2771785> (accessed March 15, 2022).

Benkendorfer, K., Pottier, L. L., and Nachman, B. (2021). Simulation-assisted decorrelation for resonant anomaly detection. *Phys. Rev. D* 104:035003. doi: 10.1103/PhysRevD.104.035003

Blance, A., Spannowsky, M., and Waite, P. (2019). Adversarially-trained autoencoders for robust unsupervised new physics searches. *J. High Energy Phys.* 2019:47. doi: 10.1007/JHEP10(2019)047

Cacciari, M., Salam, G. P., and Soyez, G. (2008). The anti- k_t jet clustering algorithm. *J. High Energy Phys.* 4:63. doi: 10.1088/1126-6708/2008/04/063

Cacciari, M., Salam, G. P., and Soyez, G. (2012). FastJet user manual. *Eur. Phys. J. C* 72:1896. doi: 10.1140/epjc/s10052-012-1896-2

Cerri, O., Nguyen, T. Q., Pierini, M., Spiropulu, M., and Vlimant, J.-R. (2019). Variational autoencoders for new physics mining at the large hadron collider. *J. High Energy Phys.* 2019:36. doi: 10.1007/JHEP05(2019)036

CMS Collaboration (2008). The CMS experiment at the CERN LHC. *J. Instrum.* 3, S08004. doi: 10.1088/1748-0221/3/08/S08004

- CMS Collaboration (2012). Observation of a New Boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B* 716, 30–61. doi: 10.1016/j.physletb.2012.08.021
- CMS Collaboration (2018). Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s}=13$ TeV and constraints on dark matter mediators and other new particles. *J. High Energy Phys.* 8:130.
- CMS Collaboration (2022a). *CMS Summary Plots EXO 13 TeV*. Available online at: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryPlotsEXO13TeV> (accessed September 14, 2022).
- CMS Collaboration (2022b). *CMS Physics Results B2G*. Available online at: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsB2G> (accessed September 14, 2022).
- CMS Collaboration (2022c). *CMS Physics Results SUS*. Available online at: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS> (accessed September 14, 2022).
- Collins, J. H., Howe, K., and Nachman, B. (2019). Extending the search for new resonances with machine learning. *Phys. Rev. D* 99, 014038. doi: 10.1103/PhysRevD.99.014038
- Cuturi, M. (2013). *Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances*. Available online at: <https://arxiv.org/abs/1306.0895> (accessed March 15, 2022).
- D'Agnolo, R. T., Grosso, G., Pierini, M., Wulzer, A., and Zanetti, M. (2021). Learning multivariate new physics. *Eur. Phys. J. C* 81:89. doi: 10.1140/epjc/s10052-021-08853-y
- D'Agnolo, R. T., and Wulzer, A. (2019). Learning new physics from a machine. *Phys. Rev. D* 99:015014. doi: 10.1103/PhysRevD.99.015014
- de Favereau, J., Delaere, C., Demin, P., Giammanco, A., Lemaître, V., Mertens, A., et al. (2014). DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *J. High Energy Phys.* 2:57. doi: 10.1007/JHEP02(2014)057
- Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. (2019). *Neural Spline Flows*. Available online at: <https://arxiv.org/abs/1906.04032> (accessed March 15, 2022).
- Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. (2020). *nflows: normalizing flows in PyTorch* (Zenodo). doi: 10.5281/zenodo.4296287
- Eschle, J., Puig Navarro, A., Silva Coutinho, R., and Serra, N. (2020). zfit: Scalable pythonic fitting. *SoftwareX* 11:100508. doi: 10.1016/j.softx.2020.100508
- Farina, M., Nakai, Y., and Shih, D. (2020). Searching for new physics with deep autoencoders. *Phys. Rev. D* 101:075021. doi: 10.1103/PhysRevD.101.075021
- Hajer, J., Li, Y.-Y., Liu, T., and Wang, H. (2020). Novelty detection meets collider physics. *Phys. Rev. D* 101:076015. doi: 10.1103/PhysRevD.101.076015
- Hallin, A., Isaacson, J., Kasieczka, G., Krause, C., Nachman, B., Quadfasel, T., et al. (2022). Classifying anomalies through outer density estimation. *Phys. Rev. D* 106, 055006. doi: 10.1103/PhysRevD.106.055006
- Heimel, T., Kasieczka, G., Plehn, T., and Thompson, J. M. (2019). QCD or what? *SciPost Phys.* 6:30. doi: 10.21468/SciPostPhys.6.3.030
- Jawahar, P., Aarrestad, T., Chernyavskaya, N., Pierini, M., Wozniak, K. A., Ngadiuba, J., et al. (2022). Improving variational autoencoders for new physics detection at the LHC with normalizing flows. *Front. Big Data* 5:803685. doi: 10.3389/fdata.2022.803685
- Kasieczka, G. et al. (2021). The LHC olympics 2020 a community challenge for anomaly detection in high energy physics. *Rept. Prog. Phys.* 84, 124201. doi: 10.1088/1361-6633/ac36b9
- Kasieczka, G., Nachman, B., and Shih, D. (2019). *R&D Dataset for LHC Olympics 2020 Anomaly Detection Challenge* (Zenodo). doi: 10.5281/zenodo.4536377
- Kingma, D. P., and Ba, J. (2017). *Adam: A Method for Stochastic Optimization*. Available online at: <https://arxiv.org/abs/1412.6980> (accessed March 15, 2022).
- Kobyzev, I., Prince, S. J., and Brubaker, M. A. (2021). Normalizing flows: an introduction and review of current methods. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 3964–3979. doi: 10.1109/TPAMI.2020.2992934
- Letizia, M., Losapio, G., Rando, M., Grosso, G., Wulzer, A., Pierini, M., et al. (2022). Learning new physics efficiently with nonparametric methods. *arXiv*. 82, 879. doi: 10.1140/epjc/s10052-022-10830-y
- LHCb Collaboration (2008). The LHCb Detector at the LHC. *J. Instrum.* 3:S08005. doi: 10.1088/1748-0221/3/08/S08005
- LHCb Collaboration (2020). Observation of structure in the J/ψ -pair mass spectrum. *Sci. Bull.* 65, 1983–1993. doi: 10.1016/j.scib.2020.08.032
- LHCb Collaboration (2021). *Observation of an Exotic Narrow Doubly Charmed Tetraquark*. CERN-EP-2021-165, LHCb-PAPER-2021-031. LHCb Collaboration.
- LHCb Collaboration (2022). *Observation of the Doubly Charmed Baryon Decay Ξ_c^{++} to $\Xi_c^+ \pi^+$* . LHCb-PAPER-2021-052, CERN-EP-2022-016. LHCb Collaboration.
- Metodiev, E. M., Nachman, B., and Thaler, J. (2017). Classification without labels: learning from mixed samples in high energy physics. *J. High Energy Phys.* 10, 174. doi: 10.1007/JHEP10(2017)174
- Nachman, B., and Shih, D. (2020). Anomaly detection with density estimation. *Phys. Rev. D* 101:075042. doi: 10.1103/PhysRevD.101.075042
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). "Pytorch: an imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Associates, Inc.), 8024–8035.
- Rezende, D. J., and Mohamed, S. (2016). *Variational Inference With Normalizing Flows*. Available online at: <https://arxiv.org/abs/1505.05770> (accessed March 15, 2022).
- Roy, T. S., and Vijay, A. H. (2019). A robust anomaly finder based on autoencoders. *arXiv*. Available online at: <https://arxiv.org/abs/1903.02032> (accessed March 14, 2022).
- Rubner, Y., Tomasi, C., and Guibas, L. J. (2000). The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.* 40, 99–121. doi: 10.1023/A:1026543900054
- Simone, A. D., and Jacques, T. (2019). Guiding new physics searches with unsupervised learning. *Eur. Phys. J. C* 79:289. doi: 10.1140/epjc/s10052-019-6787-3
- Sjöstrand, T., Mrenna, S., and Skands, P. Z. (2008). A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.* 178:852. doi: 10.1016/j.cpc.2008.01.036
- Thaler, J., and Van Tilburg, K. (2011). Identifying boosted objects with n-subjettiness. *J. High Energy Phys.* 2011:15. doi: 10.1007/JHEP03(2011)015
- Villani, C. (2009). *Optimal Transport: Old and New*, Vol. 338. Berlin; Heidelberg: Springer, 93–111. doi: 10.1007/978-3-540-71050-9