MACHINE LEARNING
Science and Technology

PURPOSE-LED PUBLISHING™

Part of the MACHINE LEARNING Series

PAPER • OPEN ACCESS

Machine-learning techniques for model-independent searches in dijet final states

To cite this article: The CMS Collaboration 2026 *Mach. Learn.: Sci. Technol.* **7** 045008

IOP Publishing Mach. Learn.: Sci. Technol. 7 (2026) 045008 https://doi.org/10.1088/2632-2153/ae7d87

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
14 January 2026

REVISED
30 May 2026

ACCEPTED FOR PUBLICATION
15 June 2026

PUBLISHED
6 July 2026

PAPER

Machine-learning techniques for model-independent searches in dijet final states

The CMS Collaboration

CERN,Geneva, Switzerland

E-mail: cms-publication-committee-chair@cern.ch

Keywords: CMS, machine learning, anomaly, dijet, resonance

Abstract

Anomaly detection methods used in a recent search for new phenomena by CMS at the CERN LHC are presented. The methods use machine learning to detect anomalous jets produced in the decay of new massive particles without depending on a specific theory model. The effectiveness of these approaches in enhancing sensitivity to various simulated signal samples is studied and compared using data collected in proton–proton collisions at a center-of-mass energy of 13 TeV. In an example analysis, the capabilities of anomaly detection methods are further demonstrated by identifying large-radius jets consistent with Lorentz-boosted hadronically decaying top quarks in a model-agnostic framework.

1. Introduction

Modern physics is remarkably successful at describing the world around us, from the largest to the smallest scales. Yet, our understanding of the Universe is known to be incomplete. Important questions include, among others, the nature of dark matter and the origin of dark energy, as well as the imbalance between matter and antimatter in the Universe. From a different perspective, general relativity, the theory of the gravitational force, is incompatible with the description of electromagnetism and interactions given by the standard model (SM) of particle physics. Many of the proposed solutions to these problems involve the existence of new particles that could be produced at particle accelerators such as the CERN LHC. However, searches conducted so far at the ATLAS [1] and CMS [2] experiments have yielded negative results.

By far the most common events produced in collisions at the LHC originate in processes mediated by the strong interaction. Together with other SM processes, they form an overwhelming background from which the (typically few) signal events resulting from the production of new particles have to be carefully separated. Machine learning (ML) is commonly employed for this task [3, 4], mainly in a supervised approach: classifiers are trained to distinguish between examples of events predicted by the SM and another set of events following a beyond-the-SM (BSM) theory [5], requiring an explicit formulation for each alternative theory.

ML has also been proposed [6] as a tool to suppress the background in a largely model-agnostic way—that is, without prior knowledge about the kind of signal events that may be present in the data (anomaly detection [7]). This paper reports on the techniques used in a recent search for BSM particles [8].

The BSM particles (noted A) considered in this paper are not observed directly in the detectors. Instead, they decay almost instantly into a pair of lighter (less massive) particles B and C, which in turn decay into SM particles, as shown in figure 1. The end result, as measured by the detector, consists of two collimated highly-energetic sprays of SM particles, called jets. The distribution of individual particles within the jets, also called the jet ‘substructure’, depends on the properties of the intermediate particles B and C: each quark or gluon appearing in the decay chain of B and C typically leads to one ‘subjet’. The substructure can be thus used to identify the particles initiating the jets. In addition to decaying to a pair of jets with nontrivial substructure, we assume that particle A is heavy (2–6 TeV) and has a narrow

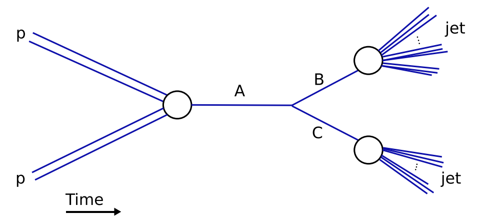


Figure 1. Diagram of the A $\rightarrow$ BC $\rightarrow$ 2 jets signal topology targeted in this work. The particle A is produced in a collision between two protons, with time flowing from left to right. Adapted from [8].

intrinsic width compared to the detector resolution, allowing to construct signal-depleted and signal-enriched regions.

The results in [8] are derived using a 'bump hunt' technique based on the invariant mass of the two jets, m_{jj} . In the SM, m_{jj} is expected to follow a smoothly falling distribution, whereas the new particle A would manifest itself as a sharp peak. However, the amount of signal may be too small to produce a visible peak. In this paper, we consider five ML methods that use jet substructure to detect signal events and separate them from the background.

The use of five very different methods raises the question of their relative and absolute performance, as well as their complementarity in finding new particles. We perform a detailed comparison in this paper, studying in particular the impact of the set of input features used by each method and the performance when using a normalized feature set. We also assess the complementarity of the methods by studying the correlations between the anomaly scores they assign to each event.

The usefulness of anomaly detection techniques is best demonstrated by using them to find known particles in experimental data. In this paper, we showcase anomaly detection in a real-world scenario where we enhance the signal produced by hadronically-decaying Lorentz-boosted top quarks in large-radius jets and analyze the selected events to recover key properties of the top quark.

This paper is structured as follows. In section 2, we review related prior work on anomaly detection. We introduce the datasets used in this study in section 3 and describe the five anomaly detection techniques in section 4. In section 5, we investigate their properties and the correlation between their results. The application to Lorentz-boosted top quarks is presented in section 6. Finally, a summary of the paper can be found in section 7.

2. Related work

There is a long history of quasi-model-independent, or model-agnostic, searches for new particles at colliders [9–20]. Such searches have traditionally been performed at collider experiments by comparing distributions measured in data to predictions of the SM. Observing a significant difference in the number of events or the shape of the distributions would be interpreted as a discovery. This approach is limited by the accuracy of the background predictions, to which large uncertainties are often assigned. In addition, the number of bins grows exponentially with the dimensionality, limiting comparisons to a relatively small number of observables. Fine categorization of the data further reduces the sensitivity.

The new class of model-agnostic searches explored in this paper uses machine learning to address these issues. In contrast to traditional model-agnostic searches, these anomaly detection methods generally focus on a single final state (e.g., dijet events) and use estimates of the background based on control samples in data. Many methods have been demonstrated using simulated events [6, 21–48], including by CMS [49], but thus far there have been few applications to collider data.

Among the proposed techniques, autoencoders [50, 51] were deployed by ATLAS in three searches for BSM physics [52–54]. The signature considered by the first search consisted of a resonance decaying to a Higgs boson and an unspecified particle Y reconstructed as a large-radius jet. A recurrent neural network trained on SM backgrounds was used to select Y candidates. In the second search, ATLAS looked for BSM particles decaying to jet+X, where X could be a jet, b-tagged jet, lepton, or photon. Events were required to contain at least one lepton to satisfy trigger requirements. An autoencoder based on event-level properties, such as object multiplicities and momenta, was trained on a fraction of the data and subsequently used to select potentially anomalous events. Finally, in [54], a variational autoencoder (VAE) [55] is used to identify anomalous jets in a search for all-hadronic BSM signatures.

In a slightly different context, autoencoders are used for data quality monitoring of the CMS electromagnetic calorimeters [56], as proposed in [57].

A second class of algorithms is based on weak supervision [24], a training mode that does not use explicit examples of the targeted signal, while still producing classifiers capable of dramatically reducing the SM background. Weak supervision was first applied to collider data in a measurement of the $t\bar{t}b\bar{b}$ cross section by the CMS experiment [58] to estimate the multijet background. The first search using this technique was performed by the ATLAS experiment [19], seeking the same dijet signal topology as we consider in this paper . However, as a first-of-its-kind search, it only used a two-dimensional (2D) feature space for anomaly detection, consisting of the masses of the two jets. This limitation was lifted in a recent iteration [20] of this search, in which the feature set was expanded to a total of six substructure variables.

An extensive comparison of anomaly detection methods on simulated events was already performed in [6]. In contrast to the simulations used in [6], in this work, all the methods have been used in a search using collider data. We also compare, for the first time, methods developed after the publication of [6].

An anomaly detection technique based on weak supervision was already used [59] to 'search' for a known particle, namely the **Υ** meson, based on CMS open data. This work focuses on a different SM particle, the top quark, which most commonly gets produced in pairs. In addition, we consider the hadronic decay of the top quark, whereas [59] focuses on a leptonic decay channel—two vastly different experimental signatures.

3. Datasets

The analysis is performed using proton-proton collision data collected with the CMS detector at $\sqrt{s} = 13\,\text{TeV}$ in the 2016–2018 data-taking period. The detector is a cylindrical apparatus built around the interaction point of the LHC beams and designed to identify the particles produced in the collisions by combining the signal they leave while crossing several specialized detector subsystems. The information collected by all subsystems is merged and processed to reconstruct particle properties and cluster them into jets. More detailed descriptions of CMS and the data processing used in this paper can be found in [2, 60] and appendix A.

The collision data used in this paper correspond to an integrated luminosity of $138\,fb^{-1}$ [61–63]. Events are selected using triggers based on the transverse momentum p_T of the jets or the scalar sum of their transverse momenta, H_T, with thresholds between 450 and 500 GeV for jet p_T and between 800 and 1050 GeV for H_T, depending on the data-taking year. The events used in the analysis are required to contain two jets with $|\eta| < 2.5$, $p_\mathrm{T} > 300$ GeV, and a pseudorapidity difference $|\Delta\eta_\mathrm{jj}| < 1.3$. Jet pairs with $|\Delta\eta_\mathrm{jj}| > 1.3$ are also used when defining various control regions. These two jets are referred to collectively as the 'dijet system'. The invariant mass m_jj of this system is required to be larger than 1455 GeV, which guarantees a trigger efficiency above 99%. We refer to this set of requirements as the 'baseline preselection'.

Since jets are composite objects containing many particles, it is natural to use the properties $(p_\mathrm{T}, \eta, \phi)$ of their constituents as inputs for anomaly detection; this strategy is referred to as using 'low-level features'. In practice, it is often desirable to use a small set of 'high-level' variables derived from the constituents' properties and known to contain physically relevant information. The following variables are used in this paper : the number of jet constituents n_PF ; the mass m_SD of the jet calculated using the soft-drop algorithm [64]; the N-subjettiness variables τ_N and their ratios $\tau_{NM} = \tau_N/\tau_M$ [65], where N and M stand for the number of subjets the variables are sensitive to; DeepB, the maximum DeepCSV [66] b tagging score of the two leading subjets of the large-radius jet selected using the soft-drop algorithm; and the lepton subjet fraction LSF_3 [67], encoding how much of the jet p_T is carried by the jet's leading lepton.

A second dataset is constructed by combining Monte Carlo samples for the main physics processes contributing to the analysis phase space. The dominant background process is the production of multiple jets through processes governed by quantum chromodynamics (QCD). QCD multijet events, as well as hadronically decaying W and Z bosons, are simulated at leading order using MadGraph 5_aMC@NLO version 2.6.5 [68] with the MLM merging scheme [69]. Processes involving top quarks are generated at next-to-leading order using the POWHEG BOX v2.0 [70–72]: top quark pair production [73], single top quark production in the t channel [74], and associated production of a top quark and W boson [75]. Parton shower and hadronization are simulated using PYTHIA 8.240 [76] with the CP5 tune [77]. Events are sampled according to their respective cross sections and generator weights to obtain an unweighted

Table 1. Signal processes considered in the analysis, categorized according to the number of partons produced in the decay of each jet. In addition to the A $\rightarrow$ BC process of figure 1, the decay products of particles B and C are indicated. The notation follows [8].

Category	Process
$1+2$	$Q^* \rightarrow qW' \rightarrow 3q$
$2+2$	$X \rightarrow YY' \rightarrow 4q$
$2+4$	$W_{KK} \rightarrow RW \rightarrow 3W$
$3+3$	$W' \rightarrow B't \rightarrow bZt$
$5+5$	$Z' \rightarrow T'T' \rightarrow tZtZ$
$6+6$	$G_{KK} \rightarrow HH \rightarrow 4t$

sample and passed through a detailed GEANT4 [78] simulation of the CMS detector before undergoing the selection described above. The equivalent luminosity of this dataset is about 27 fb^{-1}, limited by the size of the QCD multijet datasets.

A selection of Monte Carlo samples describing representative BSM models is also used in the analysis. They all involve the production of a heavy resonance (masses of 2, 3, and 5 TeV are considered) decaying to two daughter particles each producing a large-radius jet. They are categorized according to the number of partons produced in the decay of intermediate particles: for instance, the process $X \rightarrow YY' \rightarrow 4q$ is categorized as $2+2$ because each jet contains two partons. The considered topologies are listed in table 1. Where relevant, masses of 25, 80, 170, and 400 GeV are used for the intermediate particles in the decay chains. The samples are generated at leading order using MADGRAPH and undergo the same showering, hadronization, detector simulation, and selection process as the background samples discussed earlier. A more complete description of the signal samples can be found in [8].

The simulated samples are made available on Zenodo [79] in the data format described in appendix B.

4. Machine-learning based anomaly detection

This section describes the five machine-learning techniques used in the model-agnostic search, focusing on their architecture, training procedure, and statistical validation. We sort them roughly by the strength of the assumptions made about the signal. The method with the least assumptions, *VAE-QR* (section 4.1), is based on a VAE followed by an additional quantile regression (QR) step. In section 4.2, we then cover three methods, *CWoLa Hunting*, *TNT*, and *CATHODE* that use ‘weakly supervised’ classifiers to deduce properties of the signal directly from data. The construction used to train the classifiers assumes that the signal m_{jj} distribution has a sharp peak. Finally, section 4.3 covers a method, *QUAK*, that incorporates information from simulated signal events to help focus on anomalies created by new particles—as opposed to, for instance, defects in the detector.

The methods described in this section do not test directly for the existence of signal; a two-step procedure is followed instead. Each method is first used to select a set of events in which SM backgrounds are suppressed compared to a putative unknown signal. The m_{jj} distribution of this set of selected events is then used, in a second step, to derive statistical statements about the presence or absence of a signal. Specifically, the distribution is fitted using the sum of two generic functional forms representing the background and signal contributions (as discussed in [8] in more detail). For this to be possible, the selection performed by the anomaly detection methods must not significantly alter the shape of the m_{jj} spectrum, which we refer to as ‘sculpting’ the mass distribution. As detailed below, several methods implement explicit countermeasures to combat sculpting.

A brief summary of the methods is shown in table 2. Before applying them to collision data, all methods were validated in simulation and in a background-dominated control region. More details about the validation strategy can be found in [8].

4.1. VAE

The first method explored in this search makes use of a VAE and QR, and we call it *VAE-QR*. Autoencoders are a class of neural network architectures that utilize dimensionality reduction and are commonly applied to low-level input data. An autoencoder model is composed of two parts, an encoder

Table 2. Summary of the methods used in this paper . The type column distinguishes between unsupervised (Uns.), weakly supervised (Weak.), and semi-supervised (Semi.) methods. Depending on the method, the ML models use single jets or complete dijet events as inputs. We also list the input features and types of ML models used, with the following abbreviations: 'AE' for autoencoders, 'VAE' for variational autoencoders, 'DNN' for fully connected deep neural networks, and 'NF' for normalizing flows. The two variants of *CATHODE* are shown separately. Details are given in the text.

Method	Type	Input	Features	Model types
VAE-QR	Uns.	Jet	Particle p_x, p_y, p_z	VAE, DNN
CWoLa Hunting	Weak.	Jet	$m_{\mathrm{SD}}, \tau_{21}, \tau_{32}, \tau_{43}, n_{\mathrm{PF}}, \mathrm{LSF}_3, \mathrm{DeepB}$	DNN
TNT	Weak.	Jet	$m_{\mathrm{SD}}, \tau_{21}, \tau_{32}, \tau_{43}, n_{\mathrm{PF}}, \mathrm{LSF}_3, \mathrm{DeepB}$	AE, DNN
CATHODE	Weak.	Event	$m_{\mathrm{j1}}, \Delta m_{\mathrm{j1j2}}, \tau_{41,\mathrm{j1}}, \tau_{41,\mathrm{j2}}$	NF, DNN
CATHODE-b	Weak.	Event	$m_{\mathrm{j1}}, \Delta m_{\mathrm{j1j2}}, \tau_{41,\mathrm{j1}}, \tau_{41,\mathrm{j2}}, \mathrm{DeepB}_{\mathrm{j1}}, \mathrm{DeepB}_{\mathrm{j2}}$	NF, DNN
QUAK	Semi.	Event	$\tau_s, \tau_{21}, \tau_{32}, \tau_{43}, n_{\mathrm{PF}}, \rho$, DeepB for both jets	NF

and a decoder. The encoder takes the input data and compresses it into a lower-dimensional representation called the latent space. The decoder then attempts to decompress the latent representation to recover the original data. The two components are trained simultaneously to minimize the difference between the decoder output and the original data. The use of autoencoders to detect anomalous jets was first proposed in [50, 51]. This approach relies on the observation that reconstruction quality is generally poorer for events that were either not encountered during training, or seen in insufficient numbers.

VAEs extend the basic autoencoder concept to a probabilistic model, increasing the model expressivity and adding generative capabilities. Compared to an autoencoder, a VAE differs by turning the latent space into a probabilistic distribution, typically a multivariate Gaussian, whose mean μ and standard deviation σ are controlled by the encoder. The decoder inputs are sampled from the latent-space distribution.

In our application, the VAE is trained to encode and decode the momentum of the 100 highest-p_{T} constituents of individual jets, encoded in Cartesian coordinates as a zero-padded 100×3 matrix. The network architecture is shown in figure 2. Processing starts by normalizing the inputs to a mean of zero and a standard deviation of one, which is followed by a series of convolutional layers. The first convolution, with 12 2D filters of size $(1,3)$, generates an embedding for individual particles. Correlations between particles are learned by two subsequent one-dimensional convolution layers acting on three particles at a time, with 16 and 20 filters respectively. After these convolutions, the number of latent-space dimensions is reduced to 12 by two fully connected layers. The decoder uses the same operations in reverse order.

The VAE is trained using jets from a sideband region in $\Delta\eta_{\mathrm{jj}}$ in which signal is expected to be suppressed compared to the SM background. We select events by requiring $2 < |\Delta\eta_{\mathrm{jj}}| < 2.5$, a veto on additional jets with $p_{\mathrm{T}} > 300$ GeV, and p_{T} imbalance between the two jets. The two highest-p_{T} jets in each event in this region are kept for training, leading to a set of 7.4 million jets. This collection is resampled to correct the three-dimensional $(p_{\mathrm{T}}, \eta, \phi)$ distribution, which would otherwise differ significantly from that in the search region, leading to a reduced set of 1.9 million jets. This ensures that the VAE learns the distribution appropriate for jets in the signal region (SR; for instance, jets in the $\Delta\eta_{\mathrm{jj}}$ sideband region have, on average, a smaller p_{T}). The sampling uses simulated background events in the SR. Each simulated jet is replaced by the closest jet from the sideband, selected using a k-dimensional tree based on the jets' p_{T}, η, and ϕ. The resulting η and p_{T} distributions are shown in figure 3.

Training is performed by minimizing a two-component loss function. The first term is the permutation-invariant Chamfer distance [80] calculated between the input and the output jets, rewarding the network for correctly encoding and decoding its input. The latent-space distribution is enforced by adding a weighted regularization term to the loss function, computed as the (closed-form) Kullback–Leibler divergence [81] between the μ and σ vectors of the latent space and a unit Gaussian prior ($\mu = 0$, $\sigma = 1$). The latter term rewards if the Gaussians in the latent space (the bottleneck of the architecture) are not too disjoint and provide a good compromise between representing the different modes found in the data and overlapping with one another. The weight of the regularization term [82] is set to 0.5, which allows the VAE to focus more on the reconstruction quality while still enforcing a well-behaved latent space.

The network is implemented using TensorFlow /Keras 2.11.0 and trained on batches of 256 jets with the Adam optimizer [83] and an initial learning rate of 10^{-3}. The learning rate is multiplied by 0.3 if the improvement in the validation loss over 4 epochs is smaller than 0.03. Training is ended once this threshold is not reached over 8 epochs, typically after 30 epochs.

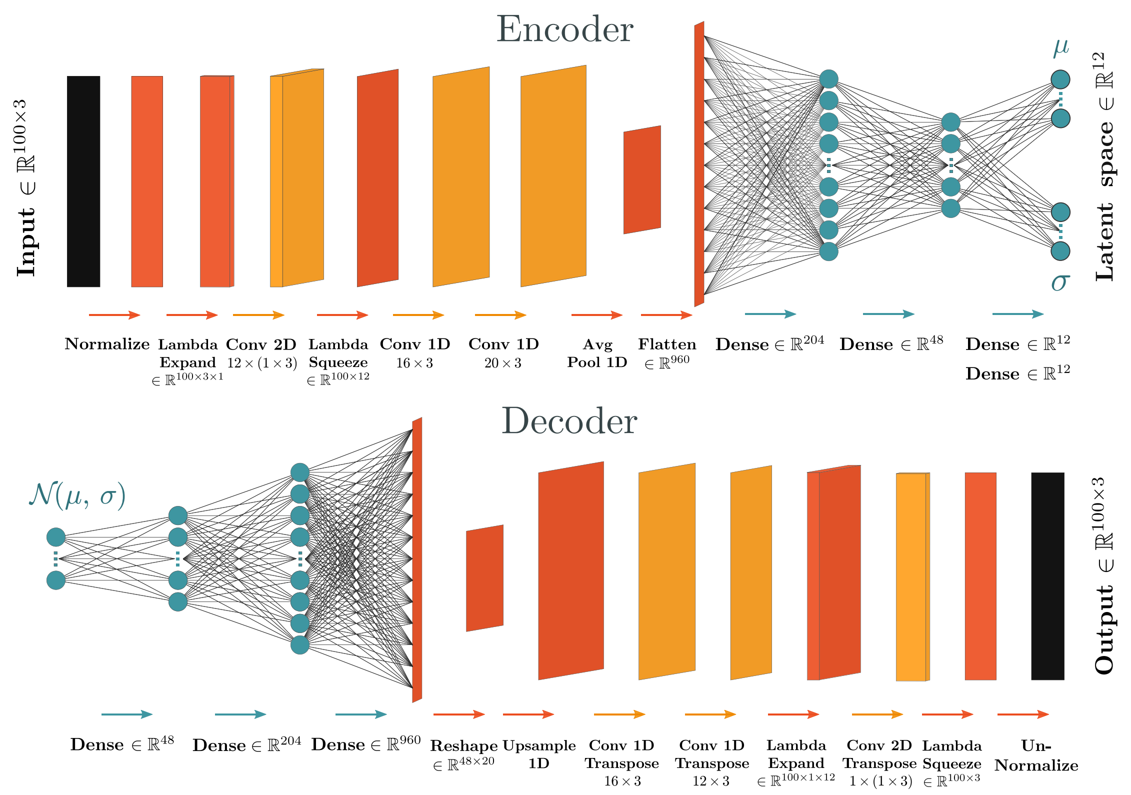


Figure 2. The VAE architecture used for jet anomaly detection. The output dimensions of layers are indicated with the $\in$ symbol. For convolution layers, the numbers of filters f and kernel sizes s are written as $f \times s$, with multidimensional s for 2D layers.

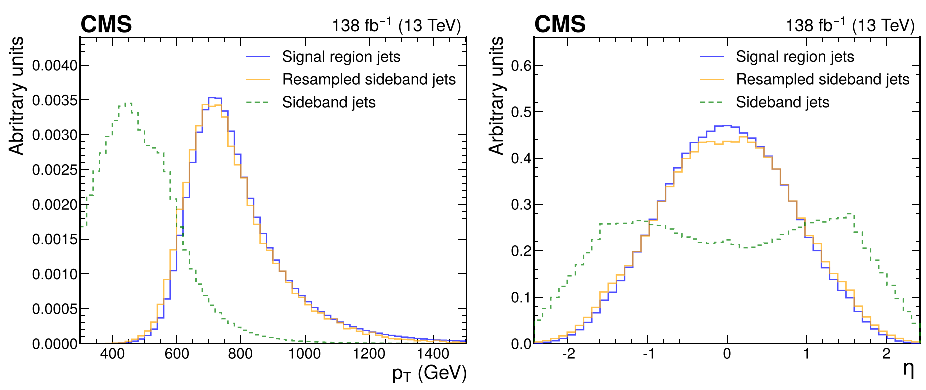


Figure 3. Jet p_T (upper) and η (right) distributions in data before and after resampling jets from the $2 < |\Delta\eta_{jj}| < 2.5$ sideband (green) to match the distribution in the signal region (blue). The distributions of resampled jets are shown in orange. Histograms are normalized to unity. The details are given in the text.

In figure 4, we show the performance of the VAE in reconstructing the p_T of particles inside the leading jet from an SM background event, a $W' \to B't \to bZt$ event, and an $X \to YY' \to 4q$ event. For this figure, p_T is calculated from the x and y components of the momentum vector, requiring the network to learn two variables and their correlation. The quality of the reconstruction, as measured by the Pearson correlation coefficient R between the input and output p_T, is better for background jets than for signal jets. To reduce sensitivity to outliers, the 2% lowest-p_T and the 2% highest-p_T jets are excluded from the calculation of R. Focusing first on background jets, we observe that the network reconstructs the leading particle well. The reconstructed p_T still matches the original p_T for the fourth particle, the last one for which an exact representation would fit in the 12-dimensional latent space. For background jets, the network is also capable of recovering the seventh particle, indicating that it successfully learned correlations between its inputs. This knowledge is, however, only useful for background, and reconstruction quality is visibly worse for signal jets. This property enables the use of the autoencoder as an anomaly detector.

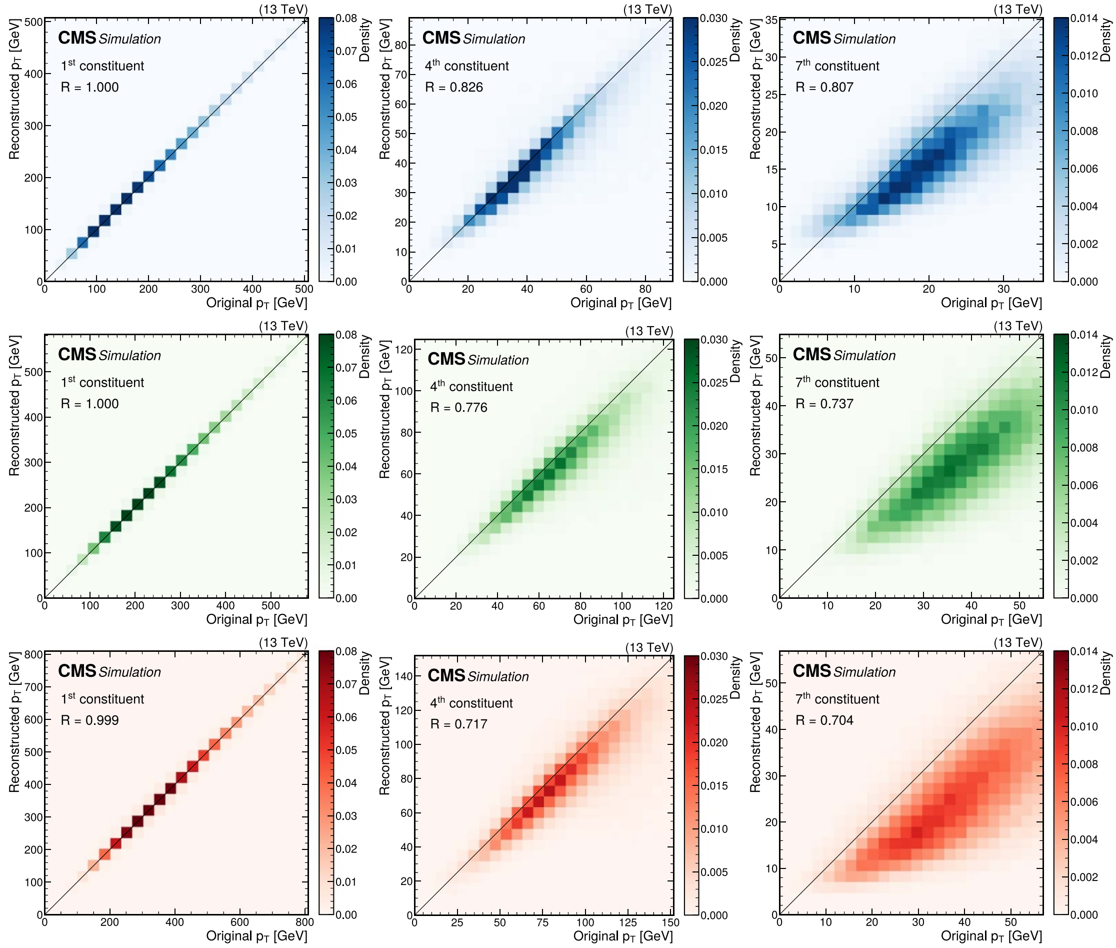


Figure 4. Quality of jet constituent p_{T} reconstruction for simulated SM background jets (upper row), jets from a $\mathrm{W}' \to \mathrm{B}'\mathrm{t} \to b\mathrm{Zt}$ decay (middle row), and jets from an $\mathrm{X} \to \mathrm{YY}' \to 4\mathrm{q}$ decay (lower row). The first (left column), fourth (middle column), and seventh (right column) constituents are shown. The Pearson correlation coefficient R between the input and output is also shown.

Once trained, the VAE is used to compute the anomaly score of SR jets, for which we exploit both reconstruction quality and latent-space information [40] by reusing the loss function described above. The SR jets with a large loss are considered anomalous. The anomaly scores for each of the two jets of every event are considered, and the smallest is taken. This is equivalent to considering that, for an event to be anomalous, both jets have to pass the same anomaly threshold. Other choices, such as the maximum and the average were considered and shown not to perform as well in conjunction with the QR described below.

The construction outlined in the previous paragraph leads to a strong correlation between the anomaly score and m_{jj}. This is not desired, as a selection based on the anomaly score then results in an arbitrarily modified m_{jj} spectrum even in the absence of signal, hindering the statistical analysis. We solve this problem using QR. That is, we construct the anomaly threshold s used in the event selection to have dependence on m_{jj} such that we select a constant fraction q of all events regardless of m_{jj}. This is achieved by training a neural network to approximate the function $s_q(m_{\mathrm{jj}})$ having this property for values of $q = 10$, 5, and 1%.

The QR models are fully connected networks with five hidden layers and 30 nodes per layer using the exponential linear unit activation function, implemented in KERAS. They have one output node per value of q. Network weights are initialized following a uniform distribution with variance according to [84]. The networks are trained using events from the SR. To avoid absorbing a signal into the regression, the dataset is split in four subsets, each of which is used to train one model. When evaluating the selection for an event from a given subset, we use the average of the regressions derived in the three other subsets, so that no event is used twice. To further reduce the risk of absorbing a signal, the average is smoothed by fitting it with a third-order polynomial.

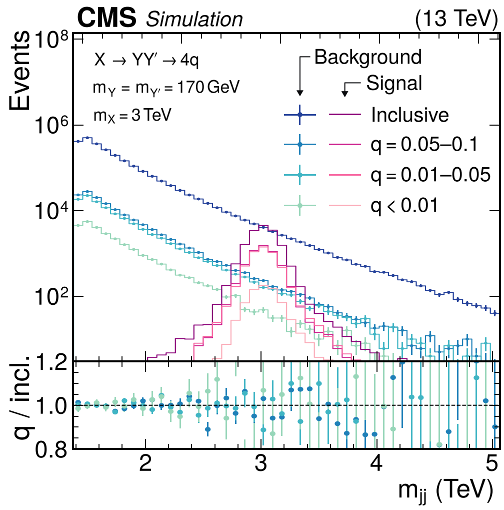


Figure 5. Simulated dijet invariant mass spectrum after selecting events based on the score of the *VAE-QR* method. The distributions are shown in the upper panel for SM background (blue to green) and the $X \rightarrow YY' \rightarrow 4q$ signal model with $m_X = 3$ TeV and $m_Y = m_{Y'} = 170$ GeV (purple to pink). The inclusive selection is compared to the three quantile ranges used in the statistical analysis. The signal-to-background ratio increases when applying a tighter selection. No significant difference is observed between the shapes, showing that the quantile regression performs as expected. The lower panel shows the ratio of the quantile ranges to the inclusive distribution for the SM background, scaled by the respective quantile probabilities. The error bars shown for background represent the statistical uncertainty.

After the QR procedure, the m_{jj} distributions in each quantile are smooth and their shapes closely follow the inclusive distribution. Figure 5 shows the m_{jj} distributions obtained after the m_{jj} -dependent selection corresponding to each of the quantiles in simulated QCD multijet background and for an $X \rightarrow YY' \rightarrow 4q$ benchmark model: one can observe a good agreement among these distributions within statistical fluctuations, and in particular with the inclusive distribution.

The final results use events selected with $q < 10\%$. For model-independent results such as p-values [85], a single fit of the m_{jj} distribution is performed for all events meeting this criterion. A different approach is followed when the signal shape is known, such as when setting limits. In these cases, the three orthogonal categories defined by $q < 1\%$, $1 \leqslant q < 5\%$, and $5 \leqslant q < 10\%$ are fitted simultaneously, using the m_{jj} shape and yield distribution among the categories for the considered signal. This more sensitive method cannot be applied in a model-agnostic way because the shapes and yields were found to differ between signals.

4.2. Weakly supervised methods

Three anomaly detection methods used in this paper (*CWoLa Hunting*, *TNT*, and *CATHODE*) use weakly supervised learning [24]. The goal of all of these methods is to learn, directly from data, with no input from simulations, an approximation to the likelihood ratio (LR) $R(x)$ between the underlying probability densities of background, $p_{bg}(x)$, and data (possibly including signal), $p_{data}(x)$, as a function of some input variables x:

$$R(x) = \frac{p_{data}(x)}{p_{bg}(x)}. \tag{1}$$

This LR, if it could be learned exactly, would be the most powerful model-agnostic anomaly detector. Indeed, if data contains a fraction f_s of signal, its probability density can be rewritten as:

$$p_{data}(x) = (1 - f_s)\, p_{bg}(x) + f_s p_{sig}(x), \tag{2}$$

where $p_{sig}(x)$ is the probability density of signal. Injecting this expression into equation (1), we find that $R(x)$ is monotonically related to the signal-to-background LR $p_{sig}(x)/p_{bg}(x)$. By learning $R(x)$, weakly supervised methods obtain an equivalent classifier without prior knowledge of $p_{sig}(x)$.

The strategy for learning a good approximation to $R(x)$ followed by all weakly supervised approaches is to train a classifier between data and samples drawn from the background model fully based on control samples in data. If the background model is accurate and the classifier is well-trained, the classifier will approach $R(x)$ asymptotically. In practice, the performance typically degrades with smaller samples sizes available for training and lower fractions of signal events in the data sample.

Table 3. Mass regions used by the weakly supervised methods and corresponding observed number of events. Signal regions are required to have sideband mass regions on either side for reliable background estimation. This means only the A1–A6 and B1–B6 regions are used to seek signals, with the A0, B0, A7, and B7 regions used solely as sideband control regions.

Name	Mass region (GeV)	Num. of events
A0	1350–1650	1.4×10^7
A1	1650–2017	4.5×10^6
A2	2017–2465	1.4×10^6
A3	2465–3013	4.0×10^5
A4	3013–3682	1.0×10^5
A5	3682–4500	2.2×10^4
A6	4500–5500	3.9×10^3
A7	5500–8000	479
B0	1492–1824	6.6×10^6
B1	1824–2230	2.1×10^6
B2	2230–2725	6.3×10^5
B3	2725–3331	1.7×10^5
B4	3331–4071	4.2×10^4
B5	4071–4975	8.5×10^3
B6	4975–6081	1.3×10^3
B7	6081–8000	144

All three weakly supervised methods train the classifier in an interval in m_{jj} called the SR and use events from outside the interval to model the background distribution $p_{bg}(x)$ in the SR. In order to cover all possible signal masses, the training is repeated in twelve partially overlapping SRs, listed in table 3. The regions cover the whole available m_{jj} range with approximately uniform logarithmic spacing and are arranged in a staggered manner so that signals located at the boundary between two regions would not be missed. The SRs are required to have sideband mass regions on either side for reliable background estimation, meaning the upper and lower extremes of the m_{jj} distribution are used solely as sideband control regions.

The three methods rely on m_{jj} sideband regions but differ from each other with respect to which events are used to train the classifier and in the modeling of $p_{bg}(x)$. As we will review further in the sections below, *CWoLa Hunting* and *TNT* use m_{jj} sideband regions directly as the background samples; these can serve as a good model for the background if the x features are statistically independent (decorrelated) from m_{jj} . Meanwhile, *CATHODE* first trains a generative model on the sidebands and interpolates this into the SR; samples from this model are used as the background model and correctly take into account any correlation between x and m_{jj} .

In order to make use of all available data events, the weakly supervised methods implement cross validation. The data and the method-specific background samples are split into five equally populated k-folds, one of which is set apart as the holdout set, in which the trained model will be applied to data to select the most anomalous events. Of the remaining four k-folds, one is used for validation and the other three for training; training is repeated for each possible choice of validation fold. When evaluating the models on the holdout set, we use the average prediction of the four models trained with varied validation sets.

After training, the classifiers trained in each SR are used to select events from their respective holdout k-folds, without restriction on m_{jj} . The events selected from all holdout k-folds are merged together and the resulting m_{jj} distribution is used as input to the fit.

This procedure is repeated for every mass region from table 3, yielding a different m_{jj} spectrum depending on the signal mass hypothesis. This results in as many smoothly falling m_{jj} spectra as there are SRs. A resonant signal identified by the classifiers would manifest itself as an excess of events localized in the corresponding SR.

4.2.1. Per-jet CWoLa hunting

Classification without labels (*CWoLa Hunting*) [25, 86] was the first proposed anomaly detection method based on weak supervision. A classifier is trained to learn the LR $R(x)$ from equation (1), using data selected from an interval in m_{jj} for p_{data} and a reweighted sample of events from sidebands on each side of the interval as a model of $p_{bg}(x)$. If a resonant signal is present in the chosen interval and not in the sidebands, the classifier will learn to tag it. However, if there is no signal present, the classifier scores

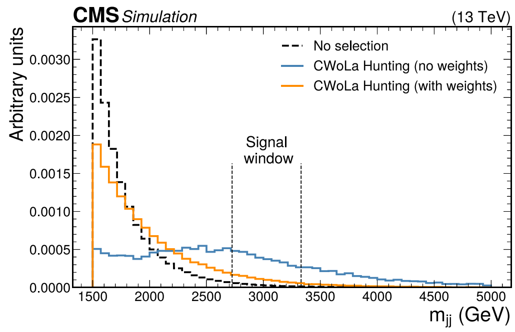


Figure 6. A comparison of the dijet mass distribution of simulated background events selected by the *CWoLa Hunting* algorithm, performed with and without the reweighting procedure described in the text. The events selected without the reweighting procedure (solid blue) have a significantly distorted mass distribution as compared to the distribution before selection (dashed black). In contrast, the events selected when the reweighting procedure is employed (solid orange) retain a smoothly falling mass distribution which allows a reliable signal extraction.

will not be meaningful. The procedure is repeated for different choices of the signal mass region in order to cover the full m_{jj} range.

One very important aspect of this paradigm is that the features used for classification should not be correlated with m_{jj} . Otherwise, the classifier will be able to differentiate background events in the SR from those in the sidebands. It will then learn to consider background events within the SR as signal-like, which could introduce artificial bumps in the m_{jj} distribution even in the absence of a signal. If, for example, the p_{T} of each jet were used as input features to the network, then it would be able to identify events in the SR based on the p_{T} difference between the sidebands and SR, leading to sculpting. Due to the power of neural networks, subtle correlations between multiple features and m_{jj} can potentially lead to sculpting. In the present work, the training strategy is modified to reduce sculpting.

In the original *CWoLa Hunting* proposal, a single classifier is used for both jets in the event, allowing the classifier to capture correlations between the jets. In this work, two distinct classifiers are used instead, trained on the heaviest and lightest of the two jets in each event, respectively. This allows additional measures to be taken to mitigate background sculpting. Each classifier uses seven input features:

$$m_{\mathrm{SD}}, \quad \tau_{21}, \quad \tau_{32}, \quad \tau_{43}, \quad n_{\mathrm{PF}}, \quad \mathrm{LSF}_3, \quad \mathrm{DeepB}.$$

The sidebands used to model $p_{\mathrm{bg}}(x)$ are the two mass regions adjacent to the one used for training. Three reweightings of the jets are performed to enhance the fidelity of this model and to minimize correlations between the learned anomaly score and m_{jj} . The relative weight of jets from the upper sideband is first increased to balance their contribution with that of the lower sideband. This is necessary because the m_{jj} spectrum is steeply falling, and thus the upper sidebands are much less populated than the lower sidebands. Additionally, to prevent class imbalance, jets in the SR are reweighted so that their total weight is the same as the total sideband weight. Finally, jets in the SR are reweighted to ensure that their p_{T} distribution corresponds to the one observed in the sidebands after the previous reweighting steps. This reweighting is accomplished by constructing a histogram of jet p_{T}'s with 20 bins and then taking the ratio of the signal window and sideband jets. To remove the effect of outliers from bins with low numbers of jets, the p_{T} reweighting factors are clipped to the range $[0.1, 5.0]$.

The reduced correlation between the learned anomaly score and jet p_{T} then mitigates mass sculpting effects in simulation studies. This is shown in figure 6, where the mass distribution of events selected by the *CWoLa Hunting* algorithm is compared with and without the reweighting procedure. Without the reweighting procedure, the algorithm learns a significant mass bias, strongly distorting the mass shape. With the reweighting procedure, the smoothly falling nature of the mass distribution is retained, allowing a reliable signal extraction.

The classifiers are simple fully connected networks using the exponential linear unit activation function [87] and dropout [88] with a drop rate of 0.2 between each layer to mitigate overtraining. Because the amount of data used in the training for different SRs varies considerably, two network architectures are used for the classifiers: a network with 6 hidden layers of sizes 64, 128, 128, 32, 16, and 8 is used for SRs A1-A4 and B1-B4, and a smaller 4-layer network with 32, 54, 16, and 8 nodes is used for the rest.

The networks are trained in TensorFlow 2.12.0 and Keras 2.11.0 using the Adam optimizer with a learning rate of 10^{-4} and the default hyperparameters. The binary cross-entropy loss function is used,

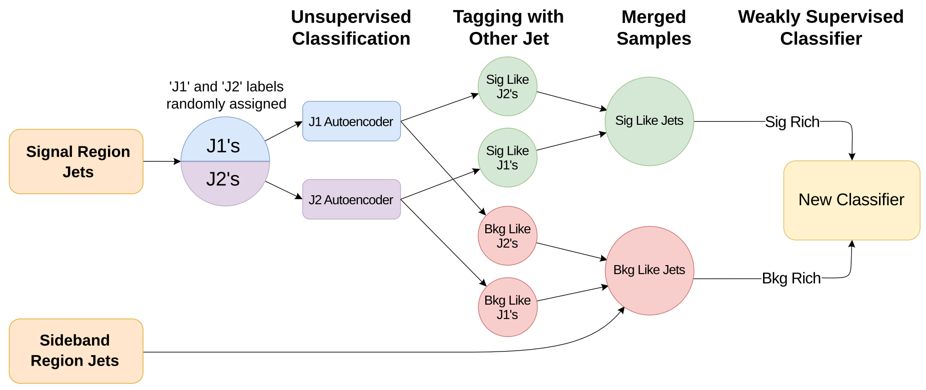


Figure 7. A schematic showing the training algorithm for *TNT*. The two jets in the event are randomly assigned to subsets J1 and J2. The dijet invariant masses and the autoencoder scores are used to construct signal- and background-like subsets of J1 and J2. These subsets are then merged, and a single classifier is trained to distinguish between signal- and background-like jets. Adapted from [8].

with jets from the sidebands labeled as 0 and jets from the SR labeled as 1. The batch size is 256 and up to 100 epochs are performed, with early stopping if the validation loss does not improve after 10 epochs. Training is repeated three times with different random initialization of the weights. We keep the network that selects the largest number of jets in the SR of the validation k-fold at the selection efficiency later used to select events, and discard the other two[1]. Within each k-fold, the validation set used to select the best-performing network is disjoint from the holdout set used to select anomalous events for final statistical analysis. Fluctuations are uncorrelated between k-folds, ensuring the model selection criterion does not amplify statistical fluctuations in the final results. Additionally, any bias to preferentially select events from the SR based only on the mass is mitigated by the reweighting procedure described above, and is validated to not introduce significant mass bias in simulation and control region studies.

Once the networks for each jet are trained, their scores need to be combined to select events. Jet scores are first converted to percentiles separately for the heavier and lighter jet in each event, and the per-event score is taken as the larger of the two. This method was chosen as it yielded the most robust performance across the benchmark signals as compared to alternatives, such as multiplying, adding, or taking the minimum of the jet scores. The classifiers trained with different validation sets are combined after this step.

Once the event-level anomaly score is defined, top-scoring events in each k-fold are selected across the whole m_{jj} range for use in the fitting procedure. The selection threshold is chosen to reach a target efficiency in the two m_{jj} sidebands, with reweighting applied to compensate for the difference in the number of events. This efficiency is chosen to ensure sufficient event counts for the fitting procedure. It is set to 1% up to SR B3, 3% for SRs A4–B5, and 5% above.

4.2.2. Tag N' Train (TNT)

The *TNT* [90] approach to anomaly detection combines weak supervision with an initial selection based on an autoencoder. The algorithm specifically targets signals in which both jets are anomalous, using the autoencoder anomaly score of one jet to obtain a signal-enriched sample of the other jet. They are then combined for the weakly supervised training. A schematic of the algorithm is shown in figure 7.

The version of *TNT* implemented in this paper starts by grouping the jets in two sets J1 and J2, randomly assigning one jet in every event to J1 or J2 and the other jet to the other set. This contrasts with the original proposal, in which one sample was made of the heavier jet in each event and the other of the lighter jet. This change reduces correlation between the two samples, improving the performance of the *TNT* algorithm.

The autoencoders used by *TNT* are based on low-level features formatted as jet images, as proposed in [50, 51]. The images are constructed as in [91], with pre-processing applied before pixelating, and centered on the p_{T} -weighted centers of the jet constituents. Constituents within an η and ϕ range of -0.6 to 0.6 around the center of the jets are cast as images of 32 by 32 pixels. In order to reduce dependence on the p_{T} of the jet, each image is normalized so that the sum of all the pixel intensities is equal to one. The autoencoders are based on convolutional networks with filter sizes of 3×3, where

[1] A similar idea has since been proposed in [89].

the image's dimensionality is reduced through max-pooling layers after each convolutional layer. Three convolutional layers with 32, 24, and 16 filters are used. The output is then fed through two dense layers of sizes 128 and 32, before reaching a latent space dimension of 6. The ReLU activation function is used after every convolutional and dense layer. The architecture is then mirrored, with two fully connected layers and then three convolutional layers using 2D sampling layers in place of the max-pooling layers to produce an image of the same dimensions as the input.

For each considered SR and test k-fold, an autoencoder is trained to reconstruct jets in the two m_{jj} sidebands. Both jets in each event are used for training the autoencoder. The network is trained for 100 epochs using the mean square error loss and the Adam optimizer with a learning rate of 10^{-4} and the default hyperparameters.

Once trained, the loss of the autoencoder is used in conjuction with m_{jj} to construct the mixed samples for the weakly supervised training. Within the SR, the loss of the J1 jet is used to tag the J2 jet as background- or signal-like, and vice-versa. Jets with a loss in the top 20% result in a signal-like tag for the other jet, and jets with a loss in the lowest 40% result in a background-like tag. Jets tagged in this way are then collected into a signal-like and a background-like sample, with all jets from the sidebands added to the background-like sample to increase its statistical power. The remaining 40% of events in the SR are not used further.

After the background-like and signal-like samples have been constructed, a weakly supervised classifier is trained to distinguish between them using the same architecture, input variables, and hyperparameters as for *CWoLa Hunting* . Reweighting during training is slightly different: background-like jets from the SR and upper sideband are each reweighted to match the weight of the lower sideband. As in *CWoLa Hunting*, the background- and signal-like classes are balanced with each other and the p_{T} distribution of the signal-like class is adjusted to match the background-like class in order to mitigate any potential m_{jj} sculpting when selecting events.

Like *CWoLa Hunting*, the distribution of anomaly scores for each jet is used to convert the raw *TNT* anomaly score into a percentile. The final event-level anomaly score is obtained as the product of the percentiles of the two jets. Compared to the maximum used for *CWoLa Hunting*, this combination requires both jets to be anomalous. The rest of the procedure is identical to *CWoLa Hunting* : top-scoring events are selected across the whole m_{jj} range and combined to obtain the m_{jj} distribution used in the fitting procedure.

4.2.3. CATHODE and CATHODE-b

Classifying anomalies through outer density estimation *(CATHODE*) [23] also follows the weak supervision paradigm. An overview of the three main steps of the method is shown in figure 8. First, the SR in which the LR $R(x)$ from equation (1) will be learned is defined as an interval in m_{jj}, similar to *CWoLa Hunting* . The second step is to estimate the background in the SR, which is done by learning the background density in the sidebands ($m_{\mathrm{jj}} \notin$ SR) with a conditional generative model, which is then sampled for $m_{\mathrm{jj}} \in$ SR. Finally, the anomaly score is obtained by training a weakly supervised classifier between the data and sampled background events.

Modeling $p_{\mathrm{bg}}(x)$ with a generative architecture is a significant upgrade over *CWoLa Hunting*, to which *CATHODE* is otherwise similar. It allows the modeling of all correlations between m_{jj} and the features x used for classification. This allows the training of a single event-level classifier instead of the jet-wise tagging done in *CWoLa Hunting* . The input features used in this paper are inspired by [23]. With j1 designating the heavier of the two jets and j2 the lighter, they are:

$$m_{\mathrm{j1}}, \quad \Delta m_{\mathrm{j1j2}} = m_{\mathrm{j1}} - m_{\mathrm{j2}}, \quad \tau_{41,\mathrm{j1}}, \quad \tau_{41,\mathrm{j2}}.$$

The τ_{41} ratio was chosen instead of τ_{21} as in [23] because it was found to provide sensitivity to a broader set of signals. Compared to *CWoLa Hunting*, this smaller feature set is better suited for training normalizing flows and minimizes performance degradation caused by the presence of irrelevant features [92].

A second advantage of the generative model is the possibility to use the complete m_{jj} interval to constrain the background distribution, instead of the comparatively narrow sidebands used in *CWoLa Hunting* . This results in a better estimate of $p_{\mathrm{bg}}(x)$. The generative model is also not limited in the number of background samples it can produce, which is used to reduce statistical fluctuations in the classifier training. Based on studies from the original *CATHODE* paper, we sample four times as many background events as there are data events in the SR.

Our implementation of *CATHODE* closely follows [23], with slightly larger networks owing to the larger dataset. The generative models are inverse autoregressive normalizing flows [93] with 15

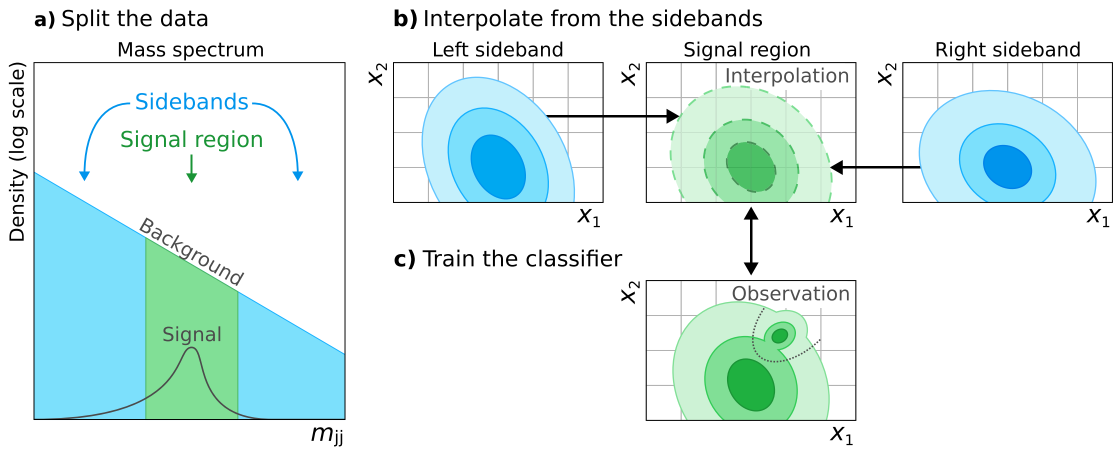


Figure 8. The three main steps of *CATHODE*. (a) Given a signal mass hypothesis, m_{jj} is used to define the signal region in which the signal would be localized. (b) A background estimate for the input features x is obtained from events outside the signal region by training a generative model and interpolating it. (c) A weakly supervised classifier is trained to distinguish between the interpolated background and the observed signal region data, singling out any difference between them, e.g., caused by signal events.

MADE [94] blocks, implemented in PyTorch version 2.0.0 [95] with Pyro version 1.8.4 [96, 97]. Each block has a single hidden layer with 128 nodes and uses ReLU activation. Batch normalization layers [98], with the momentum parameter set to 0.1, are inserted between blocks to stabilize the training. Optimization is performed using the Adam algorithm with a learning rate of 10^{-4} and a weight decay (L2 regularization) parameter of 10^{-6}. The loss function is the negative log-likelihood (NLL) of the transformed samples under a standard Gaussian distribution. The network is trained for 100 epochs and an ensemble of the five best epochs is used for sampling.

The flows are conditioned on m_{jj} and learn $p(x|m_{jj})$ for m_{jj} in the sidebands. To sample background events, we sample m_{jj} values from a kernel density estimate [99] of the m_{jj} distribution in the SR and pass them to the flows to obtain x. The sampled events are then used to train the weakly supervised classifiers that predict the final anomaly score.

The classifiers are simple fully connected networks with three hidden layers of 128 nodes using ReLU activation and a single output node with sigmoid activation, implemented in the same PyTorch release as the flows. The binary cross entropy loss function is used with sampled events having label 0 and data events having label 1. The two classes are weighted to compensate for the oversampling. Training uses the Adam optimizer with a learning rate of 10^{-3} for a maximum of 100 epochs. The results are stabilized by taking the average of the network output over the 10 epochs with the smallest validation loss.

The full training procedure is repeated for each possible choice of holdout and validation folds. The final anomaly score in a given holdout fold is obtained by averaging the output of the four classifiers trained with different validation k-folds. This is evaluated on all events from the SR and the sidebands, of which we keep the 1% most anomalous. As in *CWoLa Hunting* and *TNT*, events selected from the five test k-folds are merged together and used in the fit. The complete procedure is repeated for SRs A1 to A6 and B1 to B5. Region B6 is omitted because the number of events in region B7 alone is insufficient to train the flows.

This paper also features a second instance of *CATHODE*, called *CATHODE-b*, that differs from *CATHODE* only in the input variables, adding the DeepB scores of each jet to the previous list. This modification is motivated by the many new physics models involving couplings to third-generation quarks and the additional sensitivity to these models brought by b tagging. *CATHODE-b* shares its methodology with *CATHODE* but is otherwise treated as a separate method.

4.2.4. Signal selection efficiency

The classifiers obtained from weakly supervised methods are only approximations of the ideal case, equation (1), because of, among others, the finite number of signal events in the training sample. This can be understood from two extreme cases. When the data contains no signal, the classifier cannot learn anything meaningful and its score is a random function of the input variables. If, on the other hand, the data contains only signal, we recover the maximally powerful case of the supervised classifier. Any intermediate signal contamination, such as the one that would originate from a BSM signal, results in a classification power between these two extremes. In practice, a signal fraction of the order of 1% or less is often sufficient to achieve maximal performance.

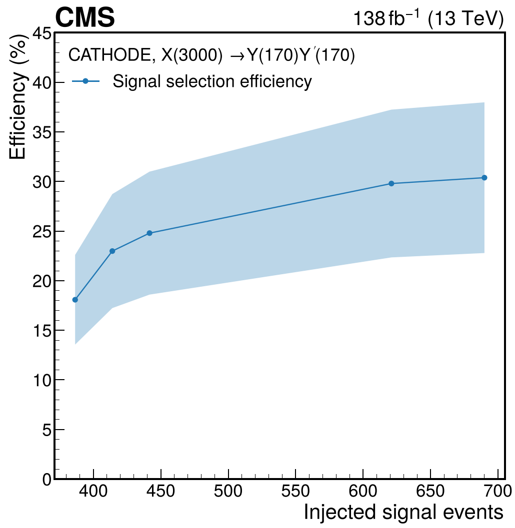


Figure 9. Signal selection efficiency of the weakly supervised classifier in the *CATHODE* method, evaluated by injecting simulated signal events in data for the $\mathrm{X} \to \mathrm{YY}' \to 4\mathrm{q}$ signal at 3 TeV in SR B3. The shaded region represents the total statistical and systematic uncertainty evaluated as described in the text. The uncertainty is largely correlated between the different signal injection points, such that the rise in the signal efficiency is statistically meaningful despite the change over the considered range being smaller than the uncertainty band.

The fraction ϵ of signal events that survives the event selection is directly related to the classification power of the classifier, and therefore also depends on the total number N_{sig} of signal events present in the data. This dependence is illustrated in figure 9, which shows how efficiency increases as a function of N_{sig}. As discussed in appendix C of [8], obtaining this $\epsilon(N_{\mathrm{sig}})$ curve and estimating the associated uncertainties is essential to set a limit on the signal cross section, one of the core results of searches for new particles. This section details the procedure followed in [8].

The N_{sig} values at which the efficiency ϵ needs to be estimated are determined by the limit setting procedure and correspond to the region where the search starts being sensitive to the chosen signal. Assuming a given N_{sig}, we start by randomly sampling this number of events from the signal sample of interest and mixing them with data from the SR. Using data directly instead of simulated background is necessary because of the limited size of the background samples. It also removes the need to consider systematic uncertainties related to background simulation, which can be sizeable. On the other hand, injecting signal into the data may induce a bias in case of signal presence in the data. We direct the interested reader to [8] for more details on this effect, whose magnitude was deemed acceptable.

Having injected signal events into the data, we repeat the training of the classifiers with the mixed datasets. The trained classifiers and the remaining signal events are used to compute $\epsilon(N_{\mathrm{sig}})$. This estimate is, however, noisy: the classifier performance depends on the specific set of signal events injected and, to a smaller extent, on the random initialization of network parameters. To obtain stable results, the procedure is repeated five times with different injected events and the results are averaged. The envelope of the five runs is used as a systematic uncertainty.

As discussed in [8], the distribution of signal events is affected by systematic uncertainties related to the imperfect modeling of the underlying physics process and the detector. Significant changes to the features of signal events would affect the training of the weakly supervised classifiers and may result in a change in ϵ. In principle, this can be accounted for by repeating the injection and retraining procedure described above with versions of the signal adjusted for the uncertainties. However, repeating the training for many sources of systematic uncertainty quickly becomes computationally prohibitive. For small uncertainties, the residual stochastic fluctuations caused by retraining can also be larger than the uncertainty. A selection procedure is therefore used to perform the retraining only for the most impactful sources.

The selection of these sources is based on the distributions of the classifier input features for signal events. A histogram is first constructed for every variable under consideration, dividing the range in ten equally sized bins, excluding the upper and lower 1% of the distributions for stability. Then, the histograms corresponding to variations of the systematic uncertainties by ±1 standard deviations are obtained within the same binning. A source of uncertainty is selected for retraining if, for at least one bin in any of the input features, the number of events that would be injected for the up or down variation deviates from the nominal count n by more than one fifth of its statistical uncertainty $\sqrt{n}$.

Table 4. Signal models used to train the six signal normalizing flows used in the *QUAK* method. Each flow uses signals with specific masses for the daughter particles B and C. For signals decaying to an SM and a BSM particle, the BSM particle is always the heavier of the two.

Daughter masses (GeV)	Processes
80, 80	$\mathrm{X} \to \mathrm{YY}' \to 4\mathrm{q}$
80, 170	$W_{\mathrm{KK}} \to RW \to 3W$
	$\mathrm{W}' \to \mathrm{B}'\mathrm{t} \to \mathrm{bZt}$
	$\mathrm{X} \to \mathrm{YY}' \to 4\mathrm{q}$
80, 400	$W_{\mathrm{KK}} \to RW \to 3W$
	$\mathrm{X} \to \mathrm{YY}' \to 4\mathrm{q}$
170, 170	$\mathrm{W}' \to \mathrm{B}'\mathrm{t} \to \mathrm{bZt}$
	$\mathrm{X} \to \mathrm{YY}' \to 4\mathrm{q}$
170, 400	$\mathrm{W}' \to \mathrm{B}'\mathrm{t} \to \mathrm{bZt}$
	$\mathrm{X} \to \mathrm{YY}' \to 4\mathrm{q}$
400, 400	$\mathrm{G}_{\mathrm{KK}} \to \mathrm{HH} \to 4\mathrm{t}$
	$\mathrm{Z}' \to \mathrm{T}'\mathrm{T}' \to \mathrm{tZtZ}$

This is a conservative criterion and the impact on the classifier is found to still be small for several of the selected sources.

The final estimate of the uncertainty on $\epsilon(N_{\mathrm{sig}})$ is obtained by summing in quadrature the contributions from all sources: the injection and retraining uncertainty, and the systematic uncertainties in the signal description, whether or not they were selected for retraining. The uncertainty band shown in figure 9 is obtained using this procedure.

4.3. QUAK

The quasi anomalous knowledge (*QUAK)* method [100] is the most model-dependent method employed in this analysis. *QUAK* uses a set of example signals as a prior to help enhance sensitivity to physical anomalies. The method seeks a balance between the purely model-agnostic approaches of the prior methods and fully supervised methods in which the model is optimized for a specific chosen signal. The use of signal priors could help filter out physically irrelevant anomalies, which could arise from bugs in the event reconstruction. *QUAK* uses two unsupervised models. The first is trained on simulated background events and the second on the signal prior. The final event selection makes use of both models. Events which are dissimilar to background events and similar to the signal prior are selected as anomalous.

Our application of *QUAK* uses autoregressive rational quadratic spline (RQS) normalizing flows [101] implemented in PyTorch 1.9 with nflows [102] 0.14 as the unsupervised model. We use six RQS transformations with 80 control points. The parameters controlling the splines are obtained from a network comprising five fully connected layers with 160 nodes and ReLU activation, evaluated on the output of the previous layer. The flows are trained with the Adam optimizer with a learning rate of 10^{-3} and a batch size of 10 000. Training stops when the loss does not improve over 10 consecutive epochs, with a maximum of 100 epochs. The NLL predicted by the flows for every event is used to construct one 'signal' and one 'background' score. This architecture takes advantage of the progress in generative models since the publication of [100] and was found to improve performance over the original autoencoder-based proposal.

The benchmark signals discussed in section 3 are used as the signal prior. To ensure sensitivity to a wide range of signals, the signal prior is split into six separate groups based on the masses of the B and C particles. The groups are listed in table 4. A separate flow model is trained for each group. The final signal score is the combination of the NLLs obtained from the six different flow models, calculated as the signed L^5 norm $||\vec{x}||_5 = (x_1^5 + x_2^5 + \cdots + x_6^5)^{1/5}$. This relatively large power was chosen so that the final signal score is sensitive to high signal-like scores from any of the six flows. The background-like score is simply the NLL obtained from the corresponding flow.

Each flow estimates the joint probability density of 14 event-level input variables, seven for each of the two jets, in its training dataset. The two jets in each event are sorted by mass. The following jet features are used:

$$\tau_s, \tau_{21}, \tau_{32}, \tau_{43}, n_{\mathrm{PF}}, \rho, \mathrm{DeepB},$$

where the modified N-subjettiness variable τ_s is defined as $\tau_s = \sqrt{\tau_{21}}/\tau_1$ and ρ is the ratio of the jet mass and transverse momentum, $\rho = m_{\mathrm{SD}}/p_{\mathrm{T}}$.

A 2D '*QUAK* space' is constructed and used to define the event selection. The two dimensions of the space encode an event's background-like and signal-like scores, respectively. The most anomalous events are expected to appear in the region characterized by a low background-like score and a high signal-like score. As the exact location of an anomaly in the *QUAK* space could vary, the event selection procedure is varied for different resonance mass hypotheses, m_{A}. First a SR is defined as $m_{\mathrm{A}} - 400\mathrm{GeV} < m_{\mathrm{jj}} < m_{\mathrm{A}} + 200$ GeV. The width is motivated by the m_{jj} resolution of the detector, and a slight shift towards low m_{jj} is added to account for longer tails observed on the left side of signal m_{jj} distributions. Two 500 GeV -wide sidebands, on each side of the SR, are used to estimate the background distribution in the *QUAK* space. These events are used to populate a 2D histogram representation of the *QUAK* space with 500 bins. Bins exhibiting an excess of SR events relative to the sidebands are selected. Events from the full dijet mass spectrum that fall within these chosen *QUAK* bins are then used in the resonance signal fit.

A broad search is conducted using *QUAK* with a signal prior consisting of a combination of all benchmark signal samples, combining six flows as described above. Additionally, targeted searches for each benchmark signal are performed using *QUAK* with a single signal flow, trained solely on the specific signal being tested. Since this approach is optimized for detecting each benchmark signal individually, it is expected to perform better for known signals but generalize less effectively to unknown signals. This method serves as a point of comparison when assessing the performance of anomaly detection techniques.

5. Performance analysis

The anomaly detection methods show varying performance across signal types. Understanding these variations can guide the development of more generic and performant methods that combine different approaches. In this section, we explore the performance characteristics of our methods using the benchmark signals. First, section 5.1 examines the ability of all methods to identify the benchmark signals. In section 5.2, we then investigate whether the same events are found by every method. Finally, section 5.3 focuses on the impact of the choice of input features on the performance and genericity of the methods.

A widely used metric to assess the performance of anomaly detection algorithms in high-energy physics is the significance improvement characteristic (SIC) [103], defined from the true and false positive rates (TPR and FPR) as $\mathrm{SIC} = \mathrm{TPR}/\sqrt{\mathrm{FPR}}$. The SIC quantifies how much the statistical significance (z-score) of an excess is enhanced after applying a selection based on an anomaly detection algorithm, as compared to the baseline selection only. For instance, a SIC of 3 means that an initial excess of two standard deviations becomes a six standard deviation excess after event selection.

5.1. Anomaly tagging performance

Unlike traditional classification problems, it is nontrivial to compare the different anomaly detection techniques. For the weakly supervised methods the tagging performance changes as a function of the amount of signal present in the data. Therefore, comparing models on a SIC curve has limited utility as one will have to pick an arbitrary signal cross section on which to evaluate the performance. Because the tagging performance can vary substantially across different cross section ranges it is difficult to draw conclusions from such a comparison.

A more appropriate performance comparison evaluates the search sensitivity achieved by each algorithm. For this study, we construct a mock dataset by injecting signal at a chosen cross section into the simulated background. The complete version of each anomaly detection algorithm, including training, selection and signal extraction, is run on this mock dataset, and the final expected signal significance (or, equivalently, p-value) is reported. Better anomaly detection algorithms are those that can achieve a higher signal significance (smaller p-value) at a given cross section.

Two signals are used as benchmarks for this study. The first signal used as a benchmark is the $2+2$ prong $\mathrm{X} \to \mathrm{YY}' \to 4\mathrm{q}$ signal with $m_{\mathrm{X}} = 3\,\mathrm{TeV}$ and $m_{\mathrm{Y}} = m_{\mathrm{Y}}' = 170$ GeV. This is a relatively straightforward signal in which the two large-area jets are both 2-pronged. The second signal used as a benchmark is the $3+3$ prong $\mathrm{W}' \to \mathrm{B}'\mathrm{t} \to \mathrm{bZt}$ signal with $m_{\mathrm{W}}' = 3\,\mathrm{TeV}$ and $m_{\mathrm{B}'} = 400$ GeV . This signal has more complicated 3-prong jets on both sides (bZ and t), each of which also contains a b quark jet.

To contextualize the performance of the anomaly detection methods, their sensitivities are compared to those of several standard methods. The standard comparison methods are chosen to differ only in their event selections, and utilize the same preselection criteria, fitting procedure, and statistical analysis as employed by the anomaly detection methods. The first comparison method is an inclusive search, which does not use any substructure-based selection criteria. By not using any requirements on the jet substructure, this strategy is maximally inclusive to all signals. However, its sensitivity is limited by the

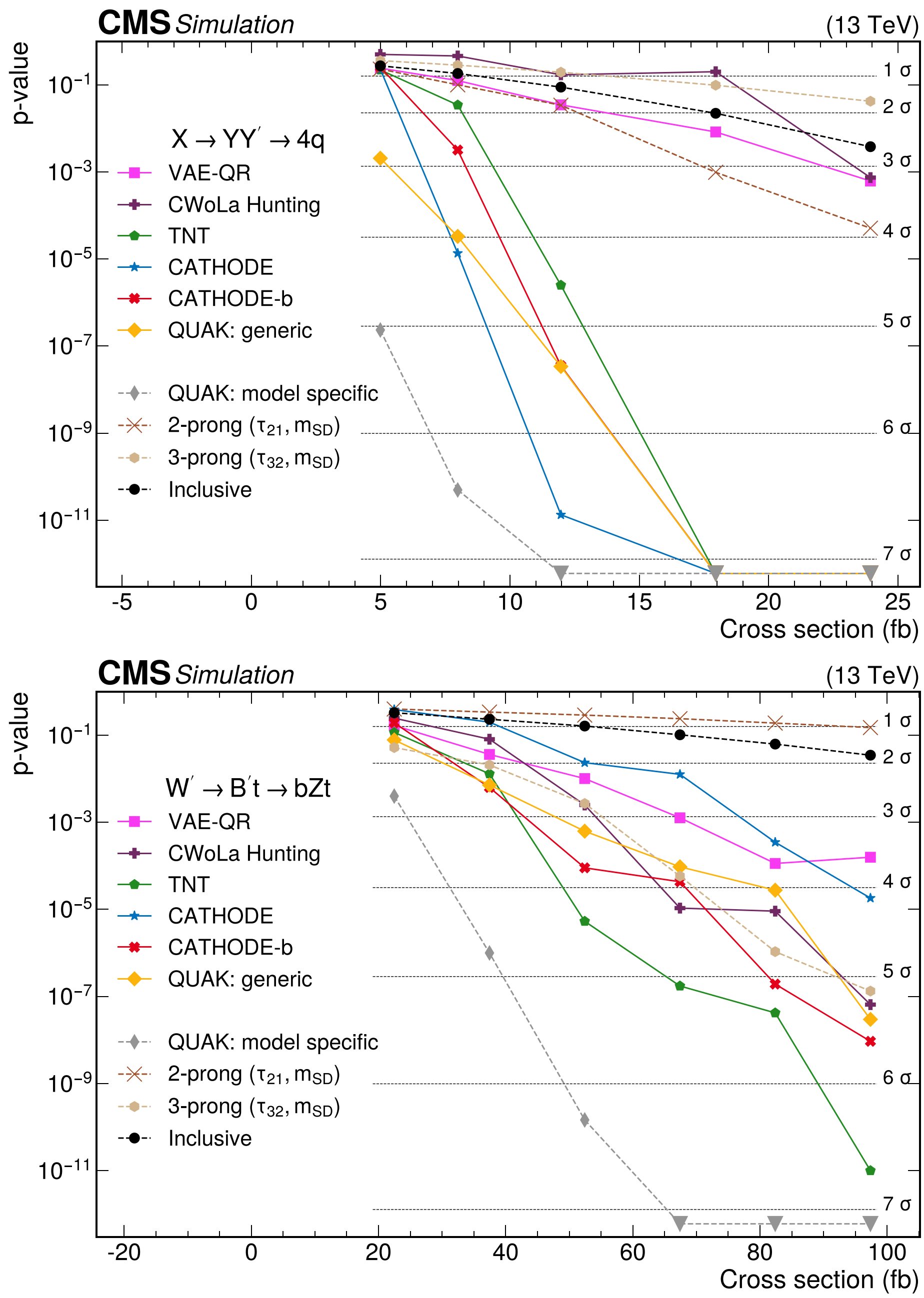


Figure 10. The p-values as a function of the injected signal cross sections for the different analysis procedures for two different signals, using different input features for each method as listed in table 2 and in the legend. The upper panel shows results for the 2-prong $X \to YY' \to 4q$ signal with $m_X = 3$ TeV and $m_Y = m_{Y'} = 170$ GeV, while the lower panel shows results for the 3-prong $W' \to B't \to bZt$ signal with $m'_W = 3$ TeV and $m_{B'} = 400$ GeV. Significance values larger than 7 standard deviations (σ) are denoted with downwards facing triangles. Adapted from [8]. © the authors. Published by IOP Publishing Ltd. CC BY 4.0.

lack of a strategy to reduce the background yield. Two model-specific event selections based on typical selections targeting jet substructure are also used. The first (second) selection targets 2-prong (3-prong) signals by requiring both jets in the event to have $\tau_{21} < 0.4$ ($\tau_{32} < 0.65$) and $m_{SD} > 50$ GeV . Finally, model-specific selections using *QUAK* are employed to maximize sensitivity, with the signal prior set to the specific signal model being tested.

Figure 10 shows the sensitivity of the search methods in simulation, evaluated as explained above. We compare the expected p-value extracted from the m_{jj} fit as a function of the signal cross section for two benchmark signals. This figure is reproduced from [8] and included here for completeness.

As expected, the inclusive search is sensitive to both models, but reaches only low significances at the considered signal cross sections. The model-specific selections targeting 2- and 3-prong signals have better sensitivity than the inclusive search for the matching signal model, but significantly worse sensitivity for the other one. This underscores the main drawback of traditional selection criteria: they introduce model dependence. We thus use the inclusive baseline preselection as the model-agnostic baseline when evaluating the performance of anomaly detection methods. This choice is, for instance, present in the definition of the SIC.

In contrast to the 2- and 3-prong selections, all anomaly detection methods demonstrate increased sensitivity compared to the inclusive search, for both signals. The relative performance of the anomaly detection methods differs between the two signals and no single method is optimal for both. As expected, the model-specific *QUAK* search yields the best sensitivity for both signals, reaching the discovery level (larger than five standard deviations) at significantly lower cross sections than other methods.

Because the methods use different feature sets, comparing them with common input features provides a fairer assessment of algorithmic performance. We therefore repeat this study, with each method using a common feature set, that of *CATHODE-b* . The *VAE-QR* method was not included in this study because it is based on high-dimensional low-level input features and cannot readily be adapted to use a small set of high level features. For *TNT*, we only modify the inputs of the classifier without changing the unsupervised preselection. Timeline constraints prevented the inclusion of *QUAK* . The results are shown in figure 11. *CATHODE-b* achieves the best results because, unlike *TNT* and *CWoLa Hunting*, it combines information from both jets to tag anomalous events. Among the per-jet methods, *TNT* outperforms *CWoLa Hunting* because of its additional preselection criteria [90].

5.2. Anomaly score correlations

We also explore the correlation between the anomaly scores of the different methods on these two benchmark signals as well as background events. We train the methods in simulation and compare their results on identical events. For the weakly supervised methods, an amount of signal equivalent to the limit obtained from the inclusive search performed on the same dataset is used in the comparison.

Because the anomaly scores of the different methods have varying ranges and distributions, we apply a preprocessing to standardize them before making a comparison. From the distribution of anomaly scores of a given method on a given sample, we use a quantile transformation to convert the distribution to a Gaussian of mean zero and variance one. For example, the median anomaly score in the original distribution is mapped to a value of 0, and a score at the 16th percentile is mapped to a value of -1. We then investigate the correlations between the scores assigned by the different methods for background, $\mathrm{X} \rightarrow \mathrm{YY}' \rightarrow 4\mathrm{q}$, and $\mathrm{W}' \rightarrow \mathrm{B}'\mathrm{t} \rightarrow \mathrm{bZt}$ events. Direct comparisons are shown as scatter plots in appendix C. In each case we compute the Pearson correlation coefficient, which measures the linear correlation between the two distributions, as well as the distance correlation (DisCo [104]), which also captures nonlinear relationships. The Pearson correlation matrices between the different methods are shown in figure 12. We assume that *CATHODE* and *CATHODE-b* are strongly correlated and only include *CATHODE* .

In general, we find the correlations between the different methods to be relatively low. The two methods with the strongest correlations are *CWoLa Hunting* and *TNT*, which have Pearson correlations of $\sim$60%. This is to be expected because they are the only two methods that use exactly the same input features, and they have similar training procedures. A similar correlation also exists between *CATHODE* and *QUAK*, which both use information from two jets during training. The low correlations indicate that the different approaches are complementary: they identify anomalous events based on different criteria, meaning each method captures anomalies that others might miss. This also indicates that more powerful search strategies could be obtained by combining the strengths of the ones presented here.

We stress that because the normalization is done separately for each sample, one cannot compare scores as shown in these plots across the different samples. Additionally, because the comparison is done separately with 'pure' samples of signal or background, this is an investigation of which of the signal events or background events look the most anomalous to each method. This means that there can be low correlation between two performant methods. For example, they may disagree about which is the 'more anomalous' of two signal events but still agree that both are much more anomalous than background events.

A different perspective is obtained by considering the selected events directly. For each method, we choose the threshold leading to a 1% background efficiency, and we compare the selected signal events. The results of this study are shown in figure 13. For the $\mathrm{X} \rightarrow \mathrm{YY}' \rightarrow 4\mathrm{q}$ signal, *CATHODE* and *QUAK* select half of the events and reach discovery-level significance. *TNT* follows with close to 40% of the signal events correctly selected. The *VAE-QR* only finds 12% of the signal and *CWoLa Hunting* is not sensitive. Unsurprisingly, methods with correlated anomaly scores also tend to have the largest overlap, with up to 82% between *CATHODE* and *QUAK* . Most events found with *VAE-QR* are also identified by *CATHODE* and *QUAK*; the overlap with *TNT* is generally lower. For the more complex $\mathrm{W}' \rightarrow \mathrm{B}'\mathrm{t} \rightarrow b\mathrm{Zt}$ signal, all methods achieve comparable sensitivity by selecting 20%–40% of the signal events, without strong overlap. This confirms our earlier observation that the different methods are complementary.

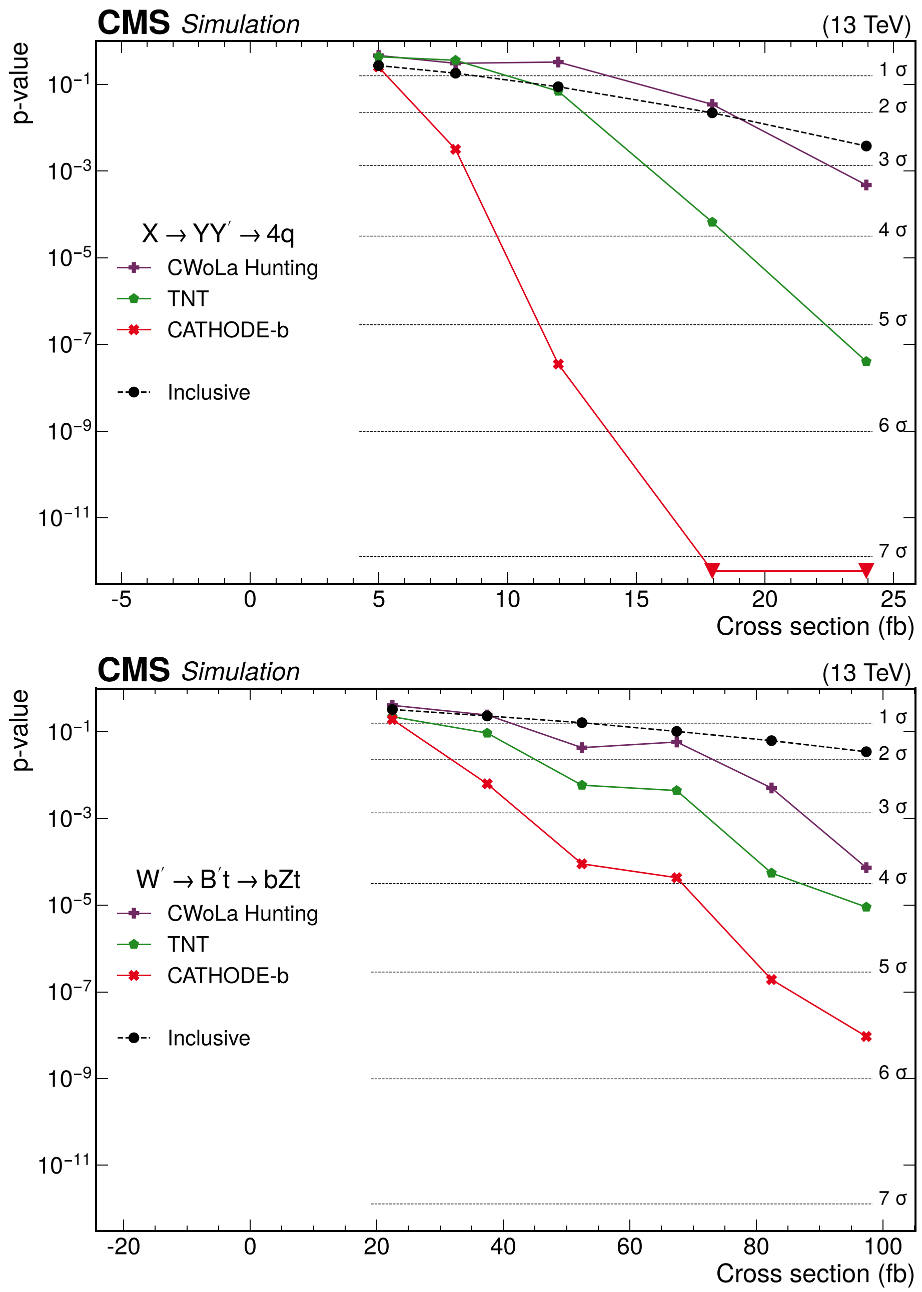


Figure 11. The p-values as a function of the injected signal cross sections for the different analysis procedures modified to use a common set of input features, for two different signals: (upper) the 2-prong $X \to YY' \to 4q$ signal with $m_X = 3$ TeV and $m_Y = m_{Y'} = 170$ GeV, and (lower) 3-prong $W' \to B't \to bZt$ signal with $m'_W = 3$ TeV and $m_{B'} = 400$ GeV. Significance values larger than 7 standard deviations (σ) are denoted with downwards facing triangles.

5.3. Impact of input features

Except for *TNT* and *CWoLa Hunting*, the five methods described in this paper all use different input features. It was already seen in section 5.1 that this has an impact on their performance. Furthermore, the most correlated methods share part of their input features. This motivates a more complete study of the impact of the choice of inputs. In this section, we perform an experiment to investigate the extent to which differences in sensitivity can be attributed to the set of input features, by training supervised classifiers to distinguish between signal and background.

Ideally, a supervised classifier trained with the binary cross-entropy loss learns the LR between signal and background, resulting in optimal classification performance. Thus, we can use such classifiers as proxies to study the separation power in a given feature set. When possible, we reuse architectures already used in the anomaly detection methods: this is the case for *CWoLa Hunting*, *TNT*, *CATHODE*, and *CATHODE-b* . The *CWoLa Hunting* and *TNT* classifiers operate at the level of individual jets. The scores from both jets are combined differently, as described in sections 4.2.1 and 4.2.2. We reflect this difference by training one set of jet-level supervised classifiers and combining their scores following the

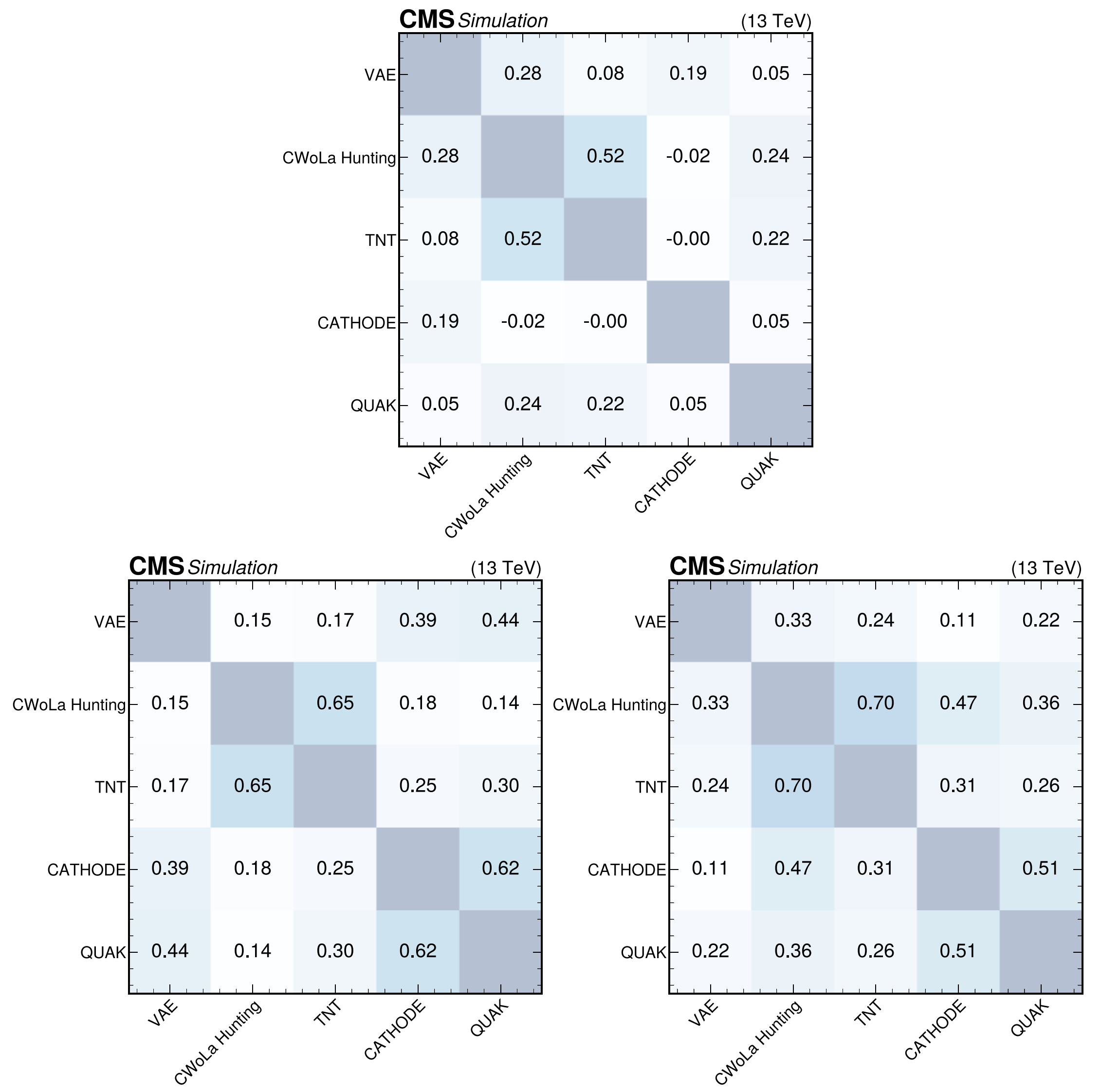


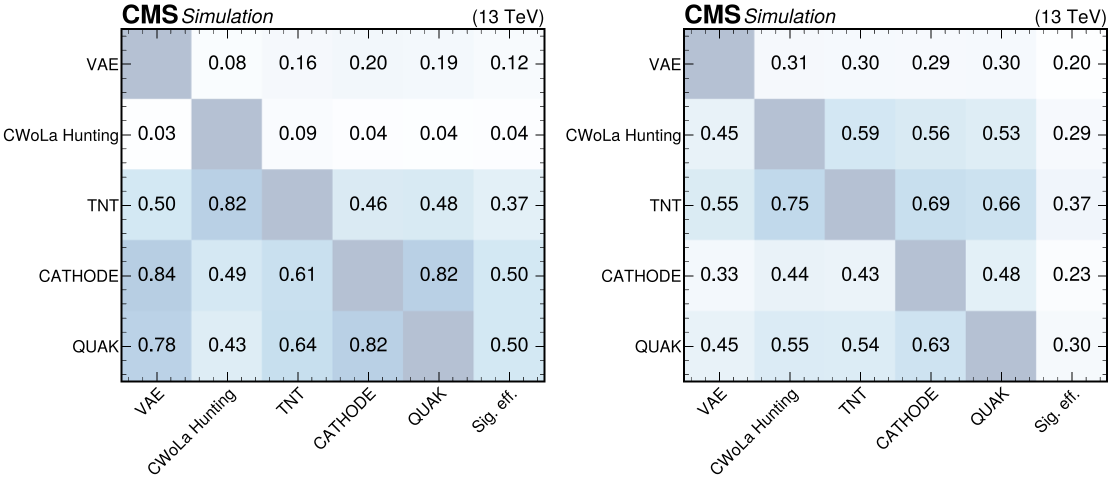
Figure 12. Summary plots showing the Pearson correlation coefficient for each pair of anomaly detection algorithms as evaluated on simulated events from the SM backgrounds (upper), $X \to YY' \to 4q$ signal (lower left), and $W' \to B't \to bZt$ signal (lower right). In many cases, the correlations are weak, indicating complementarity between the different approaches.

respective method. For *QUAK*, we use a fully connected network with 64, 64, and 32 nodes with ReLU activation in the three hidden layers, and a single node with sigmoid activation for the classification output. Owing to the difficulty of training a supervised classifier on the low-level input features used by the *VAE-QR*, it is not included in this study.

Supervised classifiers were trained for all signal models described in section 3, using data in the corresponding SR (table 3) as a background proxy. Using data instead of simulation guarantees the absence of modeling issues. A contamination of data by a signal strong enough to affect classifier trainings significantly is ruled out by the negative result of the weakly supervised searches [8].

We choose the SIC at a FPR of 1% as the main metric for comparisons. Indeed, this corresponds most closely to the gain in sensitivity achieved after selecting events based on the classifier score. The working point of 1% corresponds to the one used by all weakly supervised methods for $m_{\mathrm{A}} = 3\,\mathrm{TeV}$, and by *CATHODE* and *CATHODE-b* for $m_{\mathrm{A}} = 5\,\mathrm{TeV}$.

The SIC obtained by the supervised classifiers is shown in figure 14. The *CWoLa Hunting* -like supervised classifiers achieve a SIC of two for most signals. Combining the jet scores as in *TNT* leads to a higher SIC (about four at 3 TeV and six at 5 TeV) for all signals with two anomalous jets, but performs worse for the $Q^* \to qW' \to 3q$ signal. This is expected since *TNT* uses the product of the scores, enhancing events with two anomalous jets, unlike in this signal model. For most signals, the *CATHODE*, *CATHODE-b*, and *QUAK* feature sets yield similar results. They outperform the *CWoLa Hunting*—and



	VAE	CWoLa Hunting	TNT	CATHODE	QUAK	Sig. eff.
VAE		0.08	0.16	0.20	0.19	0.12
CWoLa Hunting	0.03		0.09	0.04	0.04	0.04
TNT	0.50	0.82		0.46	0.48	0.37
CATHODE	0.84	0.49	0.61		0.82	0.50
QUAK	0.78	0.43	0.64	0.82		0.50

	VAE	CWoLa Hunting	TNT	CATHODE	QUAK	Sig. eff.
VAE		0.31	0.30	0.29	0.30	0.20
CWoLa Hunting	0.45		0.59	0.56	0.53	0.29
TNT	0.55	0.75		0.69	0.66	0.37
CATHODE	0.33	0.44	0.43		0.48	0.23
QUAK	0.45	0.55	0.54	0.63		0.30

Figure 13. Overlap between the events selected using the anomaly detection methods, for simulated $X \to YY' \to 4q$ (left) and $W' \to B't \to bZt$ (right) signal events, with a false positive rate of 1%. Each cell corresponds to the fraction of the events selected by the method shown on the horizontal axis that are also selected by the method on the vertical axis. The last column denotes the overall efficiency of selecting signal events for each method. For instance, the *CATHODE* method selects 50% of all $X \to YY' \to 4q$ events. Of those events selected by *CATHODE*, 46% are also selected by *TNT*, but only 4% are found by *CWoLa Hunting*.

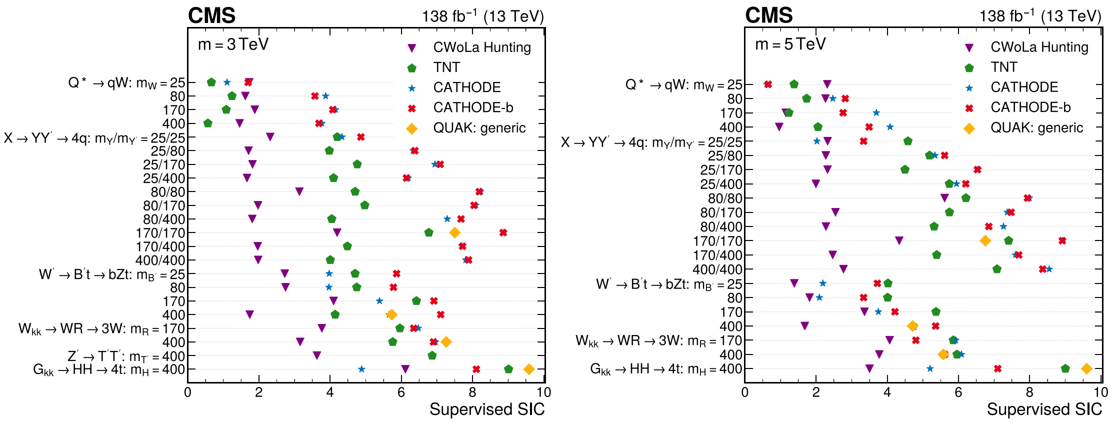


Figure 14. Significance improvement obtained from supervised classifiers trained using the same inputs as the anomaly detection methods to distinguish between simulated signal and data. The left (right) panel shows the 3 (5) TeV mass point for all signal models. The masses of intermediate decay particles are listed in GeV . The *QUAK* method was not tested on all models.

TNT -like classifiers for 3 TeV signals, while the difference with *TNT* is less marked at 5 TeV . A comparison of *CATHODE* and *CATHODE-b* shows very similar results except for signals involving b quark jets, to which the additional DeepB inputs of *CATHODE-b* add extra sensitivity.

We now turn to investigating to what extent the differences in the input features can explain differences between the sensitivity of the anomaly detection methods. We measure the improvement in sensitivity by comparing the expected upper limits on signal cross sections obtained in [8] in data, without and with anomaly detection. We calculate the limit improvement factor by taking the ratio of the limit without, to the limit with, anomaly detection. A value larger than one for this metric indicates that using anomaly detection improves the expected limit by the same factor with respect to the inclusive search.

Figure 15 compares the SIC and the limit improvement factor. We observe a clear correlation between the two variables. In particular, for each method we observe a clear correlation between the separation power of the input features for a given signal and the performance on the same signal. Depending on the method, the improvements in the limits from the anomaly detection methods are between two and eight times smaller than the SIC obtained from the supervised classifiers. This performance gap is mainly caused by the inferior tagging of anomaly detection methods owing to their model-agnostic approaches. A difference of at most ∼30% can also be attributed to effects accounted for in the

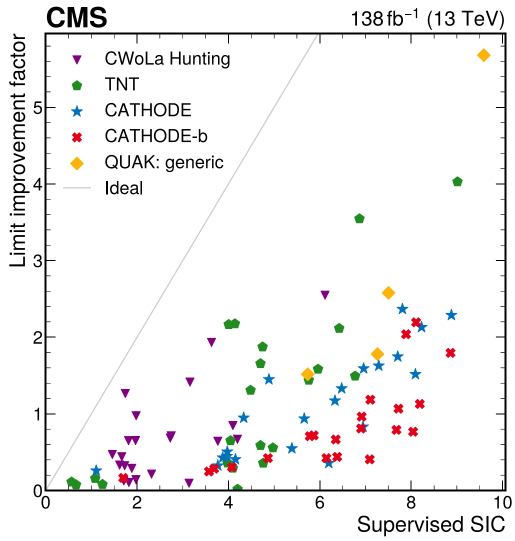


Figure 15. Comparison of anomaly detection and supervised classification performance. The factors by which the anomaly detection methods improve the upper limits on the signal cross sections are compared to the SIC of classifiers trained with the same inputs. The 3 TeV mass point is shown for all models.

limit-setting procedure but not in the SIC: systematic uncertainties (section 4.2.4) and the uncertainty in the expected background yield in the SR.

The size of the performance gap is seen to vary significantly between different anomaly detection methods. The best performing methods in this comparison are *CWoLa Hunting* and *TNT*, which reach half of the improvement of the corresponding supervised classifier for several signals. Despite its use of information about the signals during training, *QUAK* does not reach the same performance. *CATHODE* and *CATHODE-b* follow, being the methods least successful at extracting the separation power present in their datasets. This is partially compensated by stronger input features for *CATHODE* and *CATHODE-b*, especially for the $\mathrm{X} \to \mathrm{YY}' \to 4\mathrm{q}$ signal family at 3 TeV .

The direct comparison of *CATHODE* and *CATHODE-b*, for which only the input features are different, also highlights an effect already described in [92]: input features that carry no information about the signal can reduce the performance of weakly supervised methods. Indeed, it can be seen in figure 15 that *CATHODE* almost always outperforms *CATHODE-b* . A careful investigation unveils that *CATHODE-b* surpasses *CATHODE* for all signal models that contain b jets, but is otherwise less performant. Using classifiers based on large ensembles of boosted decision trees, rather than neural networks as employed in this study, has been seen to mitigate the performance loss due to uninformative features [92], and may be employed in future searches.

6. Validation in data: identification of $\mathrm{t\bar{t}}$ events

Despite the tests on simulation, it is useful to demonstrate that the anomaly detection methods can successfully find rare anomalies in data. However, no SM particles fit the topology of the original search strategy: the decay of a heavy resonance to daughters producing large-radius jets with substructure differing from the SM backgrounds. Therefore, a variation of the search strategy is developed and employed on data to identify $\mathrm{t\bar{t}}$ events. This modified search strategy, based on weak supervision, targets nonresonant pair production of two particles producing large-radius jets. It assumes that the two jets have identical masses and substructure characteristics. This strategy is able to successfully identify Lorentz-boosted $\mathrm{t\bar{t}}$ production, initially a $\sim$1% fraction of the sample, from the SM backgrounds.

Identification of $\mathrm{t\bar{t}}$ using anomaly detection methods has already been demonstrated in the semileptonic decay channel [48], using open data released by the CMS Collaboration. The algorithm used [105] was based on unsupervised learning. In contrast, we target the more challenging fully hadronic decay channel and use weak supervision.

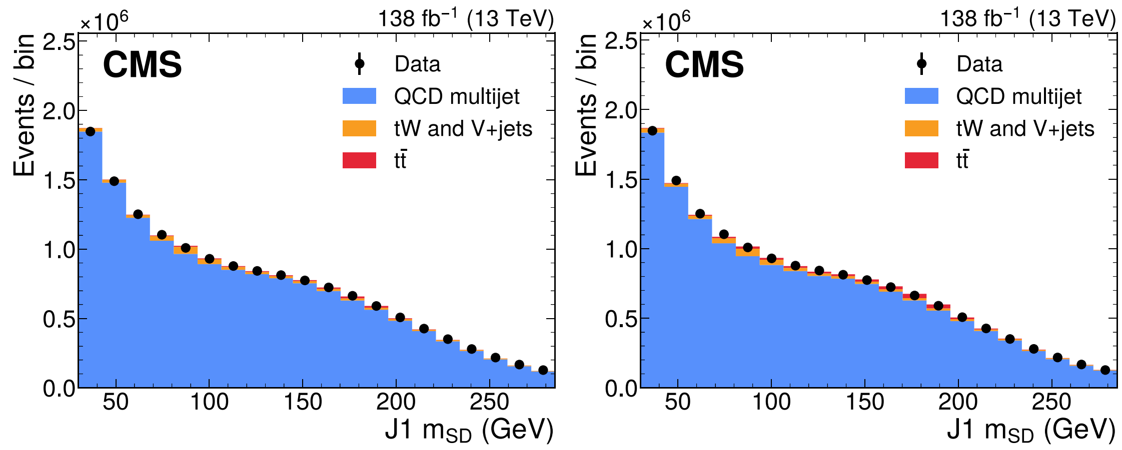


Figure 16. Distribution of the J1 soft-drop mass after the basic (upper) and b-tagged (right) preselections. The black dots represent the data and the colored histograms correspond to simulated events. In both cases the sample is dominated by the QCD multijet background (blue). While the b tagging preselection increases the relative contribution of $t\bar{t}$ (red) from 1 to 2.5% around the top quark mass (105–220 GeV), its contribution remains small. Other background processes are shown in yellow.

6.1. Anomaly detection setup

The study begins from the same basic selection criteria as the resonant anomaly search outlined in section 3. The two leading jets in the event are additionally required to have $p_{\mathrm{T}} > 400 GeV$. The requirement on the pseudorapidity difference between the two jets, $|\Delta\eta_{\mathrm{jj}}| < 1.3$, is kept unchanged. At this p_{T}, fully hadronic top quark decays are partially merged: sometimes all the decay products from the top quark are contained in the jet, but in other events only the quarks from the W decay are contained and the b quark forms a separate jet. These selection criteria were not optimized and a true nonresonant pair production search would potentially perform better. They are, however, sufficient for the desired validation on Lorentz-boosted $t\bar{t}$ events.

The natural variable to use to extract the signal events is the soft-drop mass of one of the jets, which should peak at the resonance mass (~175 GeV for top quark jets), but is smoothly falling for the QCD multijet background. This contrasts with the rest of this paper, where the invariant mass of the dijet system is used. A split strategy is employed to circumvent sculpting of the m_{SD} distribution by the selection based on the anomaly score: the two jets in the event are randomly assigned to the two categories J1 and J2 and anomaly detection is applied to J2, while the signal contribution is extracted using J1, ensuring that there is no correlation between the anomaly score of J2 and the mass of J1 for background events. This enables the estimation of the contribution of QCD multijet events in the final SR based on sidebands in data.

After the preselection described above, about 1% of the events with J1 m_{SD} around the top quark mass (105–220 GeV) are expected to originate from $t\bar{t}$ production. As discussed in section 4.2.4, the performance of weakly supervised anomaly detection methods depends on the fraction of $t\bar{t}$ events in the base dataset. In order to explore this effect in data, we also consider a more restrictive preselection in which J1 is required to contain a b-tagged subjet, using the 'loose' working point of DeepCSV [66]. This additional selection increases the expected $t\bar{t}$ fraction in events with J1 m_{SD} around the top quark mass to about 2.5%. The jet mass distributions after the two versions of the preselection are shown in figure 16.

We chose to use the same input variables for this study as in the *CWoLa Hunting* and *TNT* methods, since they are well-suited to discriminate between top quark jets and the background. There is a strong correlation between m_{SD} and the other substructure features for the background jets, which could cause problems for the weakly supervised training if the network can use the correlation to distinguish between the SR and the sidebands. Inspired by the *TNT* approach, the signal-rich and background-rich regions are defined using the J1 jet mass, but the J2 jet features are used as inputs to the anomaly tagger. For the QCD multijet background, the masses and substructure features are generally uncorrelated between the two jets. By defining signal-rich and background-rich regions based on the mass of J1, we obtain samples of J2 that exhibit identical fractions of QCD multijet background events in both regions.

In a true model-independent search, where the mass of the pair produced resonance is unknown, one would want to repeat the procedure with different signal windows defined on the J1 jet mass, similar to the different m_{jj} windows used in the dijet search. For this demonstration we take a simplified approach, employing the procedure on three partially overlapping signal windows defined on the J1 m_{SD} : 65–150, 105–220, and 145–250 GeV . The corresponding background-rich regions is defined as all events outside the signal window.

The weakly supervised training is performed using the same model architecture and hyperparameters as the *TNT* method employed in the main search. For each considered m_{SD} window, the training is performed separately for both the baseline and b-tagged preselections. As a simplification, a two-fold cross validation scheme is used rather than five-fold in the dijet search. The dataset is split into two halves, the first of which is used for training the anomaly tagger. After training, the tagger is applied to the second half of the data to select the anomalous jets.

The anomaly score of the J2 jets is evaluated, and the 0.1% most anomalous jets are selected to define the 'pass' region. The events which fail this selection define the 'fail' region. For presentation purposes, the selected data are separated into background and signal contributions. Their normalization is extracted from a simultaneous maximum likelihood fit of the two regions performed using the COMBINE package [106]. The m_{SD} distribution of the QCD multijet background in the pass region is estimated from data, using the m_{SD} distribution of events in the fail region multiplied by a freely floating transfer factor. It was verified in simulation that the shapes of the QCD background in the pass and fail regions are sufficiently similar for the transfer factor to be parameterized by a single constant, without any dependence on the jet mass. The shapes of resonant backgrounds from W and Z boson production in association with jets, as well as the $t\bar{t}$ signal shape, are estimated using simulation. The normalization of all contributions is determined entirely from the fit, without any prior constraint.

6.2. Results

The procedure is performed separately for the two versions of the preselection and for the three considered signal windows. The results using the baseline preselection are shown in figure 17. For the 65–150 GeV mass window, which is not well aligned with the top quark mass peak, the weakly supervised training does not learn to identify top quark jets, and the pass region contains no significant $t\bar{t}$ contribution. In contrast, for the 105–220 and 145–250 GeV mass windows, the procedure is successful and large contributions from $t\bar{t}$, significantly enhanced with respect to figure 16, are clearly visible in the pass region. In the $t\bar{t}$ signal, separate contributions from large-radius jets containing merged W or t decay products can be seen at masses of $\sim$80 and $\sim$175 GeV, respectively. The former can occur if the $t\bar{t}$ pair is produced with moderate p_T such that only the W decay, and not the additional b quark, is merged into a single large-radius jet. In these mass windows, naive estimates of the statistical significance, based on an asymptotic approximation [107], are very large (>10 standard deviations).

To further study the performance at identifying $t\bar{t}$, each trained model is tested in simulation. In addition, we compare them to a fully supervised model trained with labeled simulations of $t\bar{t}$ and QCD multijet events, and using the same input features and network architecture as the weakly supervised models. The SIC curves for the 105–220 and 145–250 GeV mass windows are shown in figure 18. The models trained from the 65–150 GeV region are found to be unsuccessful top taggers with significance improvement smaller than one and therefore not shown.

For the 105–220 GeV mass window, the model trained with the b tagging preselection outperforms the one trained without, which is explained by the different signal fractions. However, even this worse model is capable of identifying $t\bar{t}$ events, reaching a significance improvement larger than two at the background efficiency of 10^{-3} employed in this section. Both models trained in the 145–250 GeV mass window achieve a performance close to the supervised classifier. We emphasize that even though the weakly supervised models were trained solely on data and used no truth-labeled examples, the best performing model achieves a significance improvement only $\sim$15% worse than that of the supervised classifier. This illustrates the power of anomaly detection approaches.

6.3. Signal characterization

In the case of any significant excess of events found by an anomaly detection algorithm, one would like to infer its features, both to validate that the models are picking up a genuine physics-based signal and as input for further analysis. In the following, we employ two strategies to determine the reasons that cause jets to be classified as anomalous and demonstrate them in the context of the $t\bar{t}$ validation study. They are discussed in appendix B of [8], and a brief summary is provided here for completeness. We apply these techniques to the model trained with the b tagging preselection in the 105–220 GeV mass window.

The first strategy is used to interpret which input features are most important for the computed anomaly score. An interpretability technique called the permutation feature importance [108], suitable

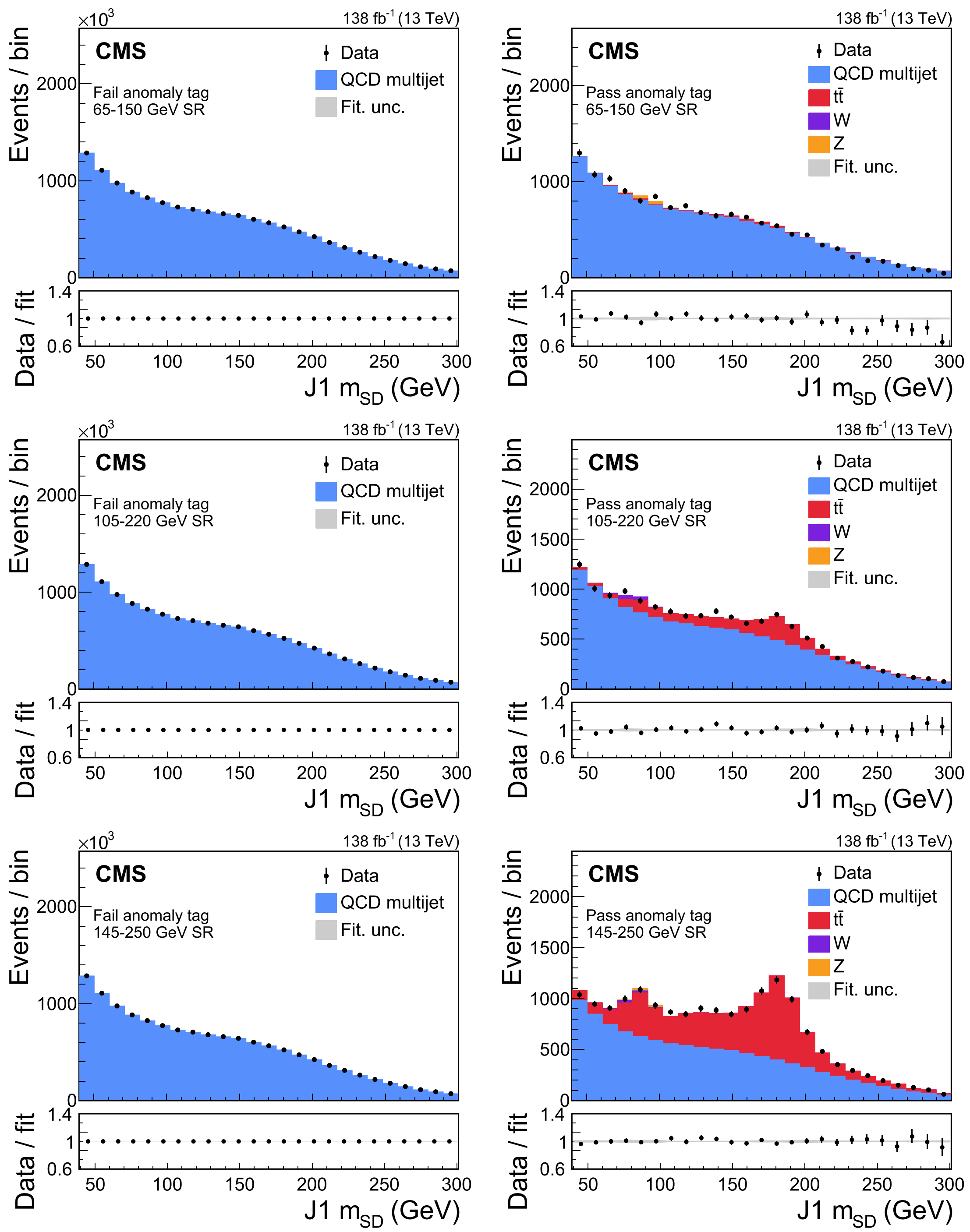


Figure 17. Post-fit plots of the fail (left) and pass (right) regions for the $t\bar{t}$ extraction procedure performed for the 65–150 (upper), 105–220 (middle), and 145–250 GeV (lower) signal windows. Data (black points with error bars) is compared to the fitted estimates of the QCD multijet (blue), Z+jets (orange), W +jets (purple), and $t\bar{t}$ (red) processes. The lower panel shows the ratio between the observed data points and the fitted estimates. The gray shading denotes the systematic uncertainty. The contribution from $t\bar{t}$ is clearly visible in the pass region of the 105–220 and 145–250 GeV signal windows.

for interpretation without labeled examples, is used. To assess the importance of a feature, we randomly swap its value for the events with the 1% largest anomaly scores with the values of other events chosen at random. We then compute the average absolute change of the classifier output with respect to the original anomalous events; a large change indicates that the value of the feature was important for the classifier score. The results are stabilized by taking the average of 100 repetitions of the procedure with different random replacements. The results of this study are shown in figure 19. They indicate that the three most important features are m_{SD}, the DEEPCSV score, and τ_{32}. The number of jet constituents also seems to play a role in the network's decision.

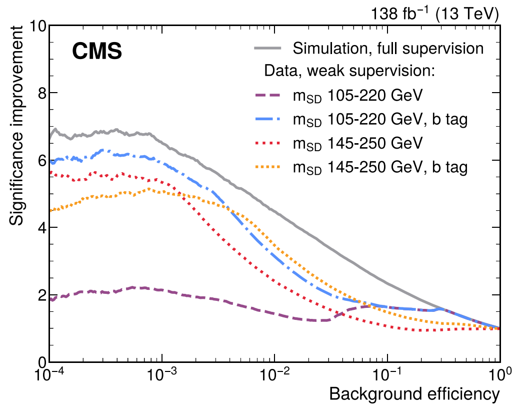


Figure 18. A comparison of the top quark identification performance of classifiers trained in different ways, evaluated in simulation. The two models trained in the 145–250 GeV mass window (red and yellow), as well as the one trained using the b tagging preselection in the in the 105–220 GeV mass window (blue), nearly match the performance of a supervised classifier (gray). The classifier trained with the baseline preselection in the 105–220 GeV mass window (purple) exhibits an improvement that is smaller than the others, yet larger than one.

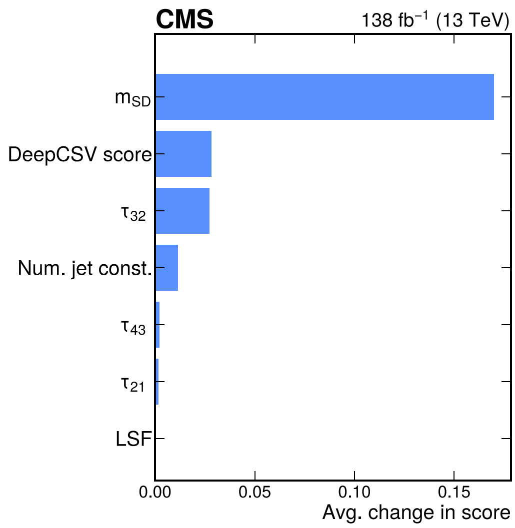


Figure 19. Excess characterization for the weakly supervised anomaly detection strategy applied to the $t\bar{t}$ region with the b tagging preselection. The sensitivity of the anomaly score to the different input observables is assessed to aid in the determination of the properties of the excess. The jet mass, b tagging score, and τ_{32} are seen to be the most important observables, consistent with the properties of large-radius top quark jets.

The second strategy we use is based on investigating the features that differ between the excess events and typical events. The distributions of features of events with high anomaly scores are compared to the distributions for all events in the region of the excess, without any anomaly score selection. Differences between these two distributions give insight into how anomalous events and background events differ.

In the case of weakly supervised anomaly detection, caution must be used when reporting properties of the signal from the distributions of the high-anomaly-score sample. Because the weakly supervised methods learn to identify the specific characteristics of the signal, they preferentially select background events that have the same characteristics as the signal. Thus, the high-anomaly-score sample has a biased selection of background events, which would have to be accounted for in a true measurement of signal properties. However, the qualitative features of the distributions are still useful in characterizing the basic phenomenological properties of the excess.

The results of this comparison are shown in figure 20. Significant differences between the two samples are visible for features already flagged by the importance study. A jet mass of ∼190 GeV is visible in the m_{SD} distribution (which was not explicitly calibrated), the DeepCSV score indicates the presence of b quarks, and the τ_{32} distribution clearly highlights the three-pronged nature of large-radius top quark jets. No significant difference is observed for the number of jet constituents, which the network may

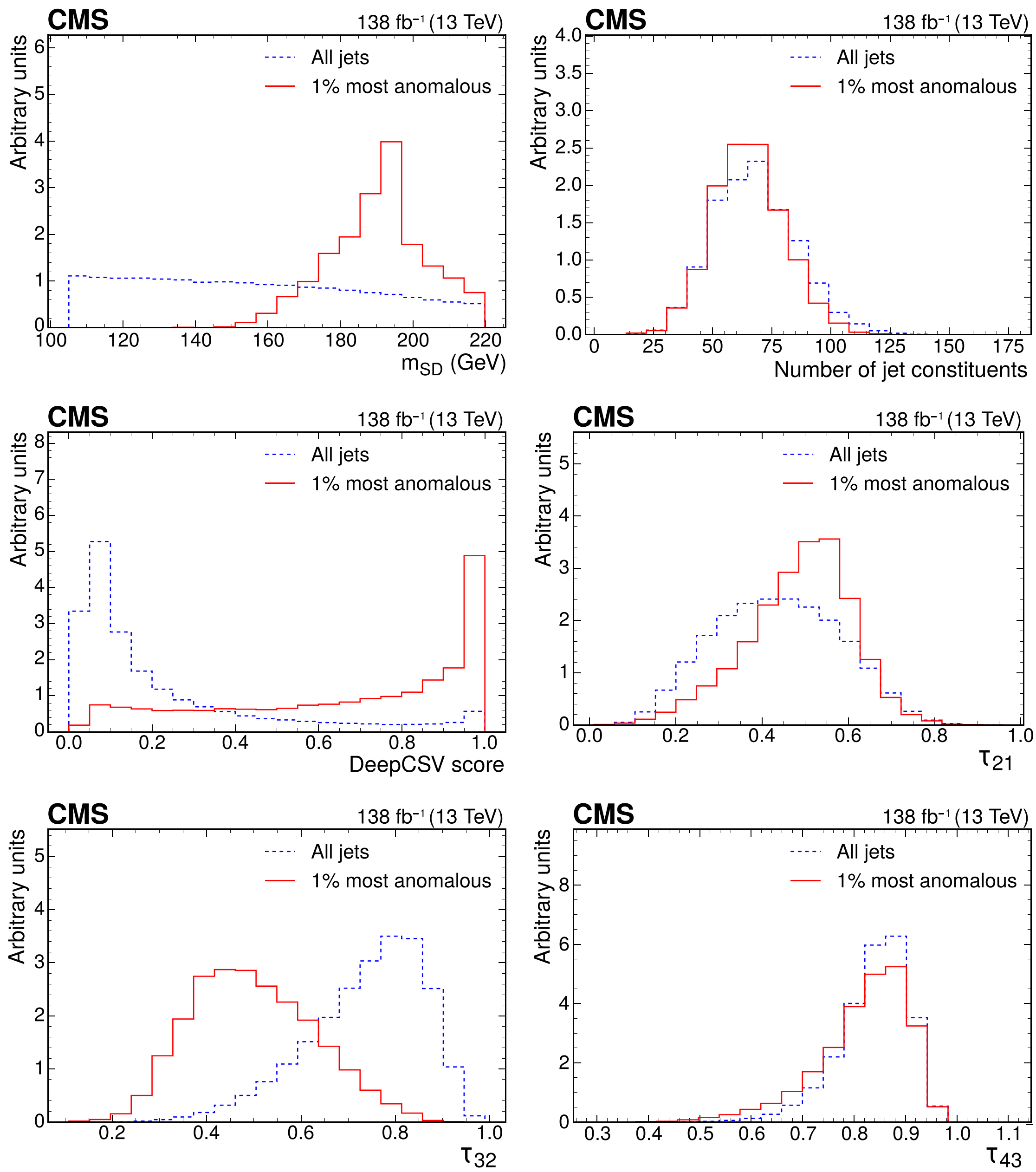


Figure 20. Excess characterization for the weakly supervised anomaly detection strategy applied to the $t\bar{t}$ region with the b tagging preselection. The plots compare the properties of the jets with the highest anomaly score (red) to those for all jets in the region of the excess (blue). The variables shown are the soft-drop mass m_{SD} (upper left), the number of jet constituents n_{PF} (upper right), the DeepCSV score (middle left), and the three subjettiness ratios τ_{21} (middle right), τ_{32} (lower left), and τ_{43} (lower right). The three-pronged nature of the signal is clear from the low τ_{32} scores, the presence of b tags from the high DeepCSV score, and the jet mass (m_{SD}) distribution peaks close to the top quark mass.

only be using in combination with another feature. Recovering these properties of the top quark shows how an anomaly search could be interpreted if it yields positive results.

7. Summary

We have presented a detailed description of the five anomaly detection methods used to search for new particles decaying to two anomalous jets in [8], using data collected by CMS between 2016 and 2018. Approaches based on weakly supervised, unsupervised, and semi-supervised training paradigms have been explored. All methods were successfully able to identify some classes of anomalous jets as distinct from SM backgrounds, and were therefore able to enhance the sensitivity to the signatures of new particles in CMS data in a model-agnostic fashion.

The sensitivity of the methods to the presence of an anomalous signal in the data has been shown to be higher than an inclusive search or simple selections based on jet substructure. The performance of the five methods has been found to depend on the considered signal model, with no single method outperforming all others. Further investigation has shown that correlations between the anomaly scores are low, indicating that methods identify anomalous signal events in different ways. Combining the different approaches could lead to more powerful methods. Furthermore, the impact of differences in the input features has been explored and shown to explain performance differences between signal models, but not between methods.

Finally, a weakly supervised anomaly detection method has been used to separate top quark jets from jets produced in other standard model processes. The achieved separation power comes very close to that of a fully supervised classifier trained with the same input variables. This constitutes a validation of resonant anomaly detection in collider data. In addition, interpretation techniques have been used to successfully retrieve some of the main properties of the top quark. This would enable further investigation and confirmation of the signal in the event of a positive result from an anomaly search.

Acknowledgments

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid and other centers for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC, the CMS detector, and the supporting computing infrastructure provided by the following funding agencies: SC (Armenia), BMBWF and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, FAPERGS, and FAPESP (Brazil); MES and BNSF (Bulgaria); CERN; CAS, MoST, and NSFC (China); MINCIENCIAS (Colombia); MSES and CSF (Croatia); RIF (Cyprus); SENESCYT (Ecuador); ERC PRG, TARISTU24-TK10 and MoER TK202 (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); SRNSF (Georgia); BMFTR, DFG, and HGF (Germany); GSRI (Greece); NKFIH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); MSIT and NRF (Republic of Korea); MES (Latvia); LMTLT (Lithuania); MOE and UM (Malaysia); BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI (Mexico); MOS (Montenegro); MBIE (New Zealand); PAEC (Pakistan); MES, NSC, and NAWA (Poland); FCT (Portugal); MESTD (Serbia); MICIU/AEI and PCTI (Spain); MOSTR (Sri Lanka); Swiss Funding Agencies (Switzerland); MST (Taipei); MHESI (Thailand); TUBITAK and TENMAK (Türkiye); NASU (Ukraine); STFC (United Kingdom); DOE and NSF (USA).

Individuals have received support from the Marie-Curie program and the European Research Council and Horizon 2020 Grant, Contract Nos. 675440, 724704, 752730, 758316, 765710, 824093, 101115353, 101002207, 101001205, and COST Action CA16108 (European Union); the Leventis Foundation; the Alfred P. Sloan Foundation; the Alexander von Humboldt Foundation; the Science Committee, Project No. 22rl-037 (Armenia); the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA) and Fonds voor Wetenschappelijk Onderzoek Contract No. 1228724 N (Belgium); the Beijing Municipal Science & Technology Commission, No. Z191100007219010, the Fundamental Research Funds for the Central Universities, the Ministry of Science and Technology of China under Grant No. 2023YFA1605804, the Natural Science Foundation of China under Grant No. 12061141002, 12535004, and USTC Research Funds of the Double First-Class Initiative No. YD2030002017 (China); the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Shota Rustaveli National Science Foundation, grant FR-22-985 (Georgia); the Deutsche Forschungsgemeinschaft (DFG), among others, under Germany's Excellence Strategy—EXC 2121 'Quantum Universe' – 390833306, and under Project Number 400140256 - GRK2497; the Hellenic Foundation for Research and Innovation (HFRI), Project Number 2288 (Greece); the Hungarian Academy of Sciences, the New National Excellence Program - ÚNKP, the NKFIH research Grants K 131991, K 133046, K 138136, K 143460, K 143477, K 146913, K 146914, K 147048, 2020-2.2.1-ED-2021-00181, TKP2021-NKTA-64, and 2025-1.1.5-NEMZ_KI-2025-00004 (Hungary); the Council of Science and Industrial Research, India; ICSC—National Research Center for High Performance Computing, Big Data and Quantum Computing, FAIR—Future Artificial Intelligence Research, and CUP I53D23001070006 (Mission 4 Component 1), funded by the NextGenerationEU program (Italy); the Latvian Council of Science; the Ministry of Education and Science, Project No. 2022/WK/14, and the National Science Center, Contracts Opus 2021/41/B/ST2/01369, 2021/43/B/ST2/01552, 2023/49/B/ST2/03273, and the NAWA contract BPN/PPO/2021/1/00011 (Poland); the Fundação para a Ciência e a Tecnologia, Grant CEECIND/01334/2018 (Portugal); the National Priorities Research Program by Qatar National Research

Fund; MICIU/AEI/10.13039/501100011033, ERDF/EU, 'European Union NextGenerationEU/PRTR', and Programa Severo Ochoa del Principado de Asturias (Spain); the Chulalongkorn Academic into Its 2nd Century Project Advancement Project, the National Science, Research and Innovation Fund program IND_FF_68_369_2300_097, and the Program Management Unit for Human Resources & Institutional Development, Research and Innovation, grant B39G680009 (Thailand); the Kavli Foundation; the Nvidia Corporation; the SuperMicro Corporation; the Welch Foundation, Contract C-1845; and the Weston Havens Foundation (USA).

Data availability statement

Release and preservation of data used by the CMS Collaboration as the basis for publications is guided by the CMS data preservation, re-use and open access policy: https://doi.org/10.7483/OPENDATA.CMS.1BNU.8V1W. The dataset used for simulation studies is available at DOI:10.5281/zenodo.16656501 [124].

Appendix A. The CMS detector and event reconstruction

The CMS apparatus [2, 60] is a multipurpose, nearly hermetic detector, designed to trigger on [109–111] and identify electrons, muons, photons, and (charged and neutral) hadrons [112–114]. Its central feature is a superconducting solenoid of 6m internal diameter, providing a magnetic field of 3.8T . Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter, each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are reconstructed using gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. More detailed descriptions of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in [2, 60].

Events of interest are selected using a two-tiered trigger system. The first level, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100kHz within a fixed latency of $4\,\mu$s [109]. The second level, known as the high-level trigger, consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to a few kHz before data storage [110].

The primary vertex is taken to be the vertex corresponding to the hardest scattering in the event, evaluated using tracking information alone, as described in section 9.4.1 of [115]. A particle-flow (PF) algorithm [116] aims to reconstruct and identify each individual particle in an event, with an optimized combination of information from the various elements of the CMS detector. Jets are clustered from the PF candidates in an event using the anti-k_T jet finding algorithm [117, 118]. In this analysis, large-radius jets with a distance parameter of $R = 0.8$ are used. Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be, on average, within 5% to 10% of the true momentum over the whole p_T spectrum and detector acceptance. Additional proton-proton interactions within the same or nearby bunch crossings (pileup) can contribute additional tracks and calorimetric energy depositions, increasing the apparent jet momentum. The pileup per particle identification algorithm [119, 120] is used to mitigate the effect of pileup at the reconstructed particle level, making use of local shape information, event pileup properties, and tracking information. Jet energy corrections are derived from simulation studies and control samples in data so that the average measured energy of jets becomes identical to that of particle level jets [119, 121].

The missing transverse momentum vector $\vec{p}_\mathrm{T}^{\,\mathrm{miss}}$ is defined as the projection onto the plane perpendicular to the beam axis of the negative vector sum of the momenta of all reconstructed PF objects in an event. Its magnitude is referred to as $p_\mathrm{T}^\mathrm{miss}$.

Appendix B. Companion dataset

This appendix describes the companion dataset [79] of this paper made available on Zenodo to support the development of new anomaly detection algorithms. It consists of an unweighted dataset of 24 million simulated SM background events and 52 different BSM signal samples. The considered background and signal processes, as well as the event generators used, are the ones described in section 3. In the following, we focus on file contents.

Events included in the datasets undergo a loose preselection. They are required to pass the trigger, contain two jets with $p_T > 300$ GeV and $|\eta| < 2.5$, and have $m_{jj} > 1200$ GeV . For signal samples, the efficiency of this selection is recorded in the `preselection_eff` variable. Two systematic variations, corresponding to the uncertainties in the jet energy scale (JES) and resolution (JER), are also stored as up/down pairs.

Basic event information is stored in the `event_info` array, which contains for every event the event number, p_T^{miss}, its angle ϕ, the original generator weight, whether the event contains a leptonic decay, an empty placeholder field, the year of the simulated data-taking conditions, and the total number of jets in the event. The generated physics process is encoded in the `truth_label` column, with signal events having positive labels. The QCD multijet events have label 0; -1 is used for background processes involving top quarks, -2 for processes involving W bosons, and -3 for Z boson production. The kinematic information of the two leading jets is stored as follows in the `jet_kinematics` array: the first two entries correspond to m_{jj} and $\Delta\eta_{jj}$. This is followed by the four-momenta of the two jets saved as p_T, η, ϕ, and m_{SD} . All masses and momenta use GeV as the default unit.

Substructure information is provided in three tables per jet. The first, `extraInfo`, contains the high-level variables τ_1, τ_2, τ_3, τ_4, LSF$_3$, DeepB, and n_{PF} . The second table stores the four-momenta of the 100 highest-p_T particles in each jet sorted by descending p_T and encoded as p_x, p_y, p_z, and E. Finally, the properties of up to six secondary vertices (SVs) contained in the jet are saved in the `SVs` table as the mass, p_T, the number of tracks, the normalized χ^2 of the SV fit, and the two- and three-dimensional displacement significances.

For signal samples, we provide limited generator-level information, experimental corrections, and variables needed to evaluate systematic uncertainties. The `gen_info` table contains the p_T, η, ϕ, and particle type ID [5] of the quarks produced in the decay of the intermediate BSM particles, the number of which depends on the signal topology.

An important correction to the event weights is stored in the `sys_weights` array. The first entry contains the product of experimental correction factors. The rest of the vector contains 10 up/down pairs of systematic variations that apply multiplicatively to the event weights, corresponding to experimental and theory uncertainties.

A second weight aiming to correct the description of jets in boosted topologies [122] is provided in the `lund_weights` table. Statistical uncertainties in this correction are provided as twice 100 replicas, `pt_var` and `stat_var`, covering p_T extrapolation and statistical uncertainties, respectively. Each uncertainty is given by the standard deviation of the replicas. In addition, five sources of systematic uncertainty are encoded in the `sys_var` list as up/down pairs.

Finally, uncertainties in jet kinematic variables are also included in one table per jet (`JME_vars`). Each table contains the jet p_T and m_{SD} varied up and down under the JES and JER uncertainties, as well as variations of the jet mass scale (JMS) and resolution (JMR) [123]. These uncertainties should be applied together with the corresponding variations of the preselection efficiency. The table layout is as follows: ${p_T}^{\text{JES up}}$, ${m_{SD}}^{\text{JES up}}$, ${p_T}^{\text{JES down}}$, ${m_{SD}}^{\text{JES down}}$, ${p_T}^{\text{JER up}}$, ${m_{SD}}^{\text{JER up}}$, ${p_T}^{\text{JER down}}$, ${m_{SD}}^{\text{JER down}}$, ${m_{SD}}^{\text{JMS up}}$, ${m_{SD}}^{\text{JMS down}}$, ${m_{SD}}^{\text{JMR up}}$, ${m_{SD}}^{\text{JMR down}}$.

Appendix C. Anomaly score comparisons

In this appendix, we show comparisons of the anomaly scores assigned by different methods to background (figure C1), $X \to YY' \to 4q$ (figure C2), and $W' \to B't \to bZt$ (figure C3) events. Following the procedure in section 5.2, each method is applied to a simulated dataset and the scores are transformed to follow a standard Gaussian distribution. We then create a series of 2D scatter plots comparing the normalized distributions of one method with another. These plots are used as input to compute the correlations shown in section 5.2.

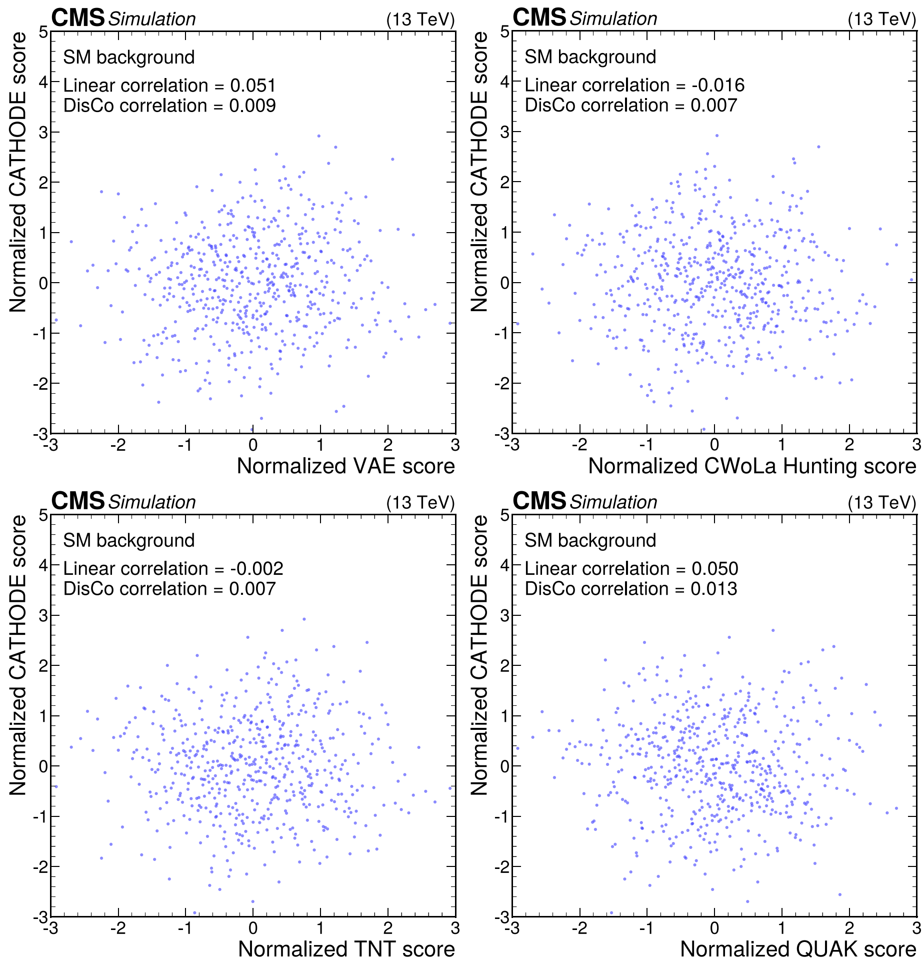


Figure C1. Anomaly score correlations of different methods on the simulated background sample. Scores are transformed to follow a normal distribution. The Pearson linear correlation coefficient and distance correlation (DisCo) are listed for each pairing. Details are given in the text.

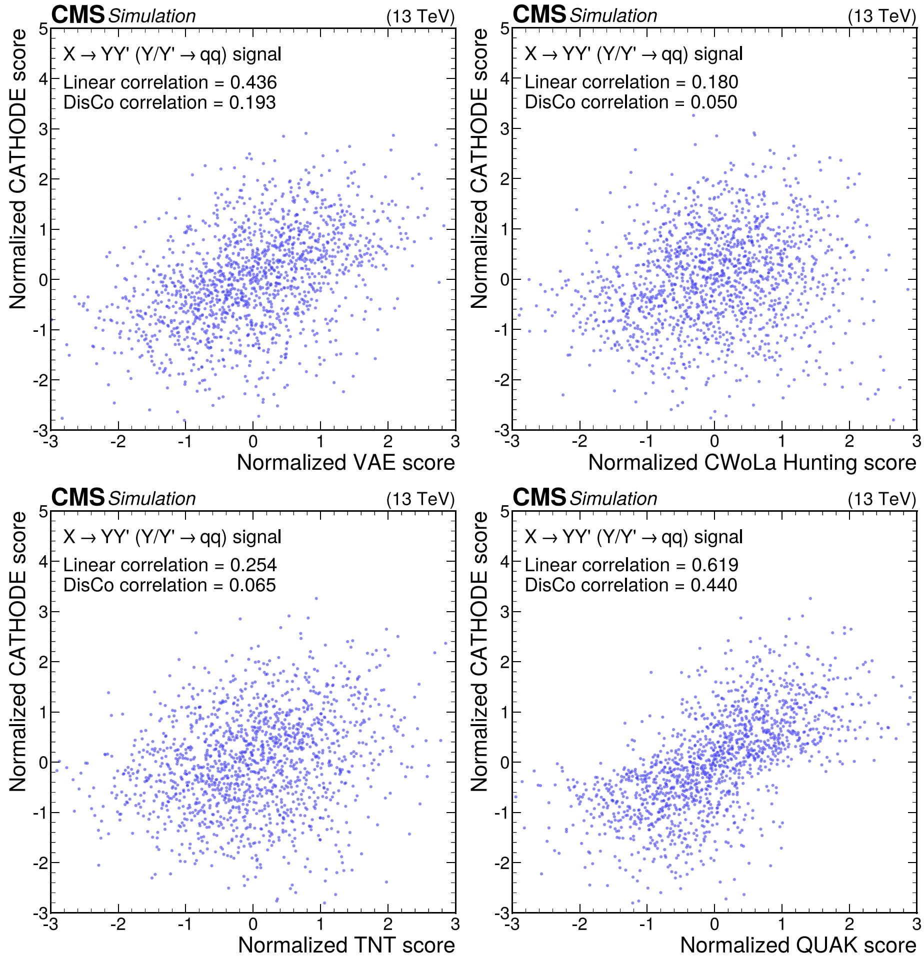


Figure C2. Anomaly score correlations of different methods on the $X \to YY' \to 4q$ signal model. Scores transformed to follow a normal distribution. The Pearson linear correlation coefficient and distance correlation (DisCo) are listed for each pairing. Details are given in the text.

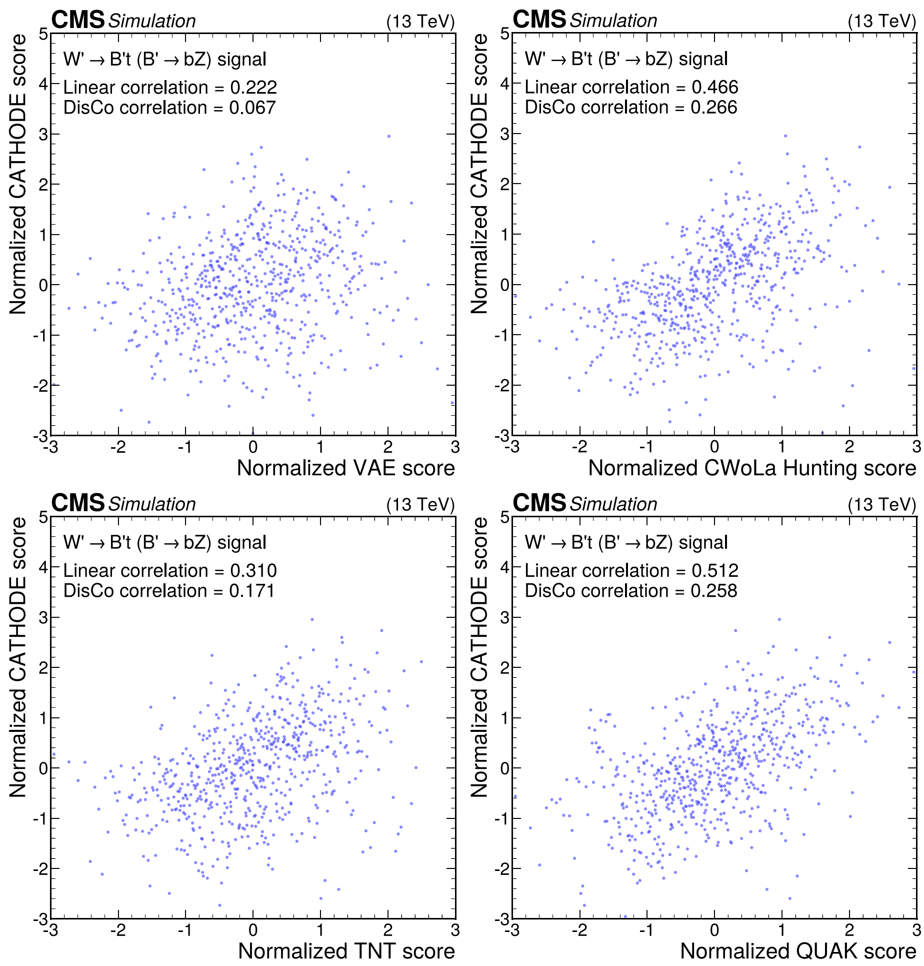


Figure C3. Anomaly score correlations of different methods on the $W' \to B't \to bZt$ signal model. Scores are transformed to follow a normal distribution. The Pearson linear correlation coefficient and distance correlation (DisCo) are listed for each pairing. Details are given in the text.

The CMS Collaboration

A Hayrapetyan, V Makarenko, A Tumasyan[1]
Yerevan Physics Institute, Yerevan, Armenia
W Adam, L Benato, T Bergauer, M Dragicevic, C Giordano, P S Hussain, M Jeitler[2], N Krammer, A Li, D Liko, M Matthewman, J Schieck[2], R Schöfbeck[2], M Shooshtari, M Sonawane, W Waltenberger, C -E Wulz[2]
Institut für Hochenergiephysik, Vienna, Austria
T Janssen, H Kwon, D Ocampo Henao, T Van Laer, P Van Mechelen
Universiteit Antwerpen, Antwerpen, Belgium
J Bierkens, N Breugelmans, J D'Hondt, S Dansana, A De Moor, M Delcourt, F Heyen, Y Hong, P Kashko, S Lowette, I Makarenko, D Müller, J Song, S Tavernier, M Tytgat[3], G P Van Onsem, S Van Putte, D Vannerom
Vrije Universiteit Brussel, Brussel, Belgium
B Bilin, B Clerbaux, A K Das, I De Bruyn, G De Lentdecker, H Evard, L Favart, P Gianneios, A Khalilzadeh, F A Khan, A Malara, M A Shahzad, A Sharma, L Thomas, M Vanden Bemden, C Vander Velde, P Vanlaer, F Zhang
Université Libre de Bruxelles, Bruxelles, Belgium
M De Coen, D Dobur, G Gokbulut, D Marckx, K Skovpen, A M Tomaru, N Van Den Bossche, J van der Linden, J Vandenbroeck, L Wezenbeek
Ghent University, Ghent, Belgium
H Aarup Petersen, S Bein, A Benecke, A Bethani, G Bruno, A Cappati, J De Favereau De Jeneret, C Delaere, F Gameiro Casalinho, A Giammanco, A O Guzel, V Lemaitre, J Lidrych, P Malek, P Mastrapasqua, S Turkcapar

Université Catholique de Louvain, Louvain-la-Neuve, Belgium
G A Alves, M Barroso Ferreira Filho, E Coelho, C Hensel, D Matos Figueiredo, T Menezes De Oliveira, C Mora Herrera, P Rebello Teles, M Soeiro, E J Tonelli Manganote[4], A Vilela Pereira
Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil
W L Aldá Júnior, H Brandao Malbouisson, W Carvalho, J Chinellato[5], M Costa Reis, E M Da Costa, G G Da Silveira[6], D De Jesus Damiao, S Fonseca De Souza, R Gomes De Souza, S S Jesus, T Laux Kuhn[6], M Macedo, K Mota Amarilo, L Mundim, H Nogima, J P Pinheiro, A Santoro, A Sznajder, M Thiel, F Torres Da Silva De Araujo[7]
Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
C A Bernardes[6], L Calligaris, F Damas, T R Fernandez Perez Tomei, E M Gregores, B Lopes Da Costa, I Maietto Silverio, P G Mercadante, S F Novaes, Sandra S Padula, V Scheurer
Universidade Estadual Paulista, Universidade Federal do ABC, São Paulo, Brazil
A Aleksandrov, G Antchev, P Danev, R Hadjiiska, P Iaydjiev, M Shopova, G Sultanov
Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia, Bulgaria
A Dimitrov, L Litov, B Pavlov, P Petkov, A Petrov
University of Sofia, Sofia, Bulgaria
S Keshri, D Laroze, S Thakur
Instituto De Alta Investigación, Universidad de Tarapacá, Casilla 7 D, Arica, Chile
W Brooks
Universidad Tecnica Federico Santa Maria, Valparaiso, Chile
T Cheng, T Javaid, L Wang, L Yuan
Beihang University, Beijing, People's Republic of China
Z Hu, Z Liang, J Liu, X Wang, H Yang
Department of Physics, Tsinghua University, Beijing, People's Republic of China
G M Chen[8], H S Chen[8], M Chen[8], Y Chen, Q Hou, X Hou, F Iemmi, C H Jiang, H Liao, G Liu, Z -A Liu[9], J N Song[9], S Song, J Tao, C Wang[8], J Wang, H Zhang, J Zhao
Institute of High Energy Physics, Beijing, People's Republic of China
A Agapitos, Y Ban, A Carvalho Antunes De Oliveira, S Deng, B Guo, Q Guo, C Jiang, A Levin, C Li, Q Li, Y Mao, S Qian, S J Qian, X Qin, C Quaranta, X Sun, D Wang, J Wang, M Zhang, Y Zhao, C Zhou
State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, People's Republic of China
S Yang
State Key Laboratory of Nuclear Physics and Technology, Institute of Quantum Matter, South China Normal University, Guangzhou, People's Republic of China
Z You
Sun Yat-Sen University, Guangzhou, People's Republic of China
K Jaffel, N Lu
University of Science and Technology of China, Hefei, People's Republic of China
G Bauer[10,11], Z Cui[11], B Li[12], H Wang, K Yi[13], J Zhang
Nanjing Normal University, Nanjing, People's Republic of China
Y Li, Y Zhou[14]
Institute of Modern Physics and Key Laboratory of Nuclear Physics and Ion-beam Application (MOE)—Fudan University, Shanghai, People's Republic of China
Z Lin, C Lu, M Xiao[15]
Zhejiang University, Hangzhou, Zhejiang, People's Republic of China
C Avila, D A Barbosa Trujillo, A Cabrera, C Florez, J Fraga, J A Reyes Vega
Universidad de Los Andes, Bogota, Colombia
C Rendón, M Rodriguez, A A Ruales Barbosa, J D Ruiz Alvarez
Universidad de Antioquia, Medellin, Colombia
N Godinovic, D Lelas, A Sculac
University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Split, Croatia
M Kovac, A Petkovic, T Sculac
University of Split, Faculty of Science, Split, Croatia
P Bargassa, V Brigljevic, B K Chitroda, D Ferencek, K Jakovcic, A Starodumov, T Susa
Institute Rudjer Boskovic, Zagreb, Croatia

A Attikis, K Christoforou, S Konstantinou, C Leonidou, L Paizanos, F Ptochos, P A Razis, H Rykaczewski, H Saka, A Stepennov
University of Cyprus, Nicosia, Cyprus
M Finger†, M Finger Jr
Charles University, Prague, Czech Republic
E Ayala
Escuela Politecnica Nacional, Quito, Ecuador
E Carrera Jarrin
Universidad San Francisco de Quito, Quito, Ecuador
H Abdalla[16], Y Assran[17,18]
Academy of Scientific Research and Technology of the Arab Republic of Egypt, Egyptian Network of High Energy Physics, Cairo, Egypt
A Hussein, H Mohammed
Center for High Energy Physics (CHEP-FU), Fayoum University, El-Fayoum, Egypt
M Kadastik, T Lange, C Nielsen, J Pata, M Raidal, N Seeba, L Tani
National Institute of Chemical Physics and Biophysics, Tallinn, Estonia
E Brücken, A Milieva, K Osterberg, M Voutilainen
Department of Physics, University of Helsinki, Helsinki, Finland
F Garcia, P Inkaew, K T S Kallonen, R Kumar Verma, T Lampén, K Lassila-Perini, B Lehtela, S Lehti, T Lindén, N R Mancilla Xinto, M Myllymäki, M m Rantanen, S Saariokari, N T Toikka, J Tuominiemi
Helsinki Institute of Physics, Helsinki, Finland
N Bin Norjoharuddeen, H Kirschenmann, P Luukka, H Petrow
Lappeenranta-Lahti University of Technology, Lappeenranta, Finland
M Besancon, F Couderc, M Dejardin, D Denegri, P Devouge, J L Faure, F Ferri, P Gaigne, S Ganjour, P Gras, G Hamel de Monchenault, M Kumar, V Lohezic, Y Maidannyk, J Malcles, F Orlandi, L Portales, S Ronchi, M Ö Sahin, A Savoy-Navarro[19], P Simkina, M Titov, M Tornago
IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
R Amella Ranz, F Beaudette, G Boldrini, P Busson, C Charlot, M Chiusi, T D Cuisset, O Davignon, A De Wit, T Debnath, I T Ehle, S Ghosh, A Gilbert, R Granier de Cassagnac, L Kalipoliti, M Manoni, M Nguyen, S Obraztsov, C Ochando, R Salerno, J B Sauvan, Y Sirois, G Sokmen, L Urda Gómez, A Zabi, A Zghiche
Laboratoire Leprince-Ringuet, CNRS/IN2P3, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, France
J -L Agram[20], J Andrea, D Bloch, J -M Brom, E C Chabert, C Collard, G Coulon, S Falke, U Goerlach, R Haeberle, A -C Le Bihan, M Meena, O Poncet, G Saha, P Vaucelle
Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France
A Di Florio, B Orzari
Centre de Calcul de l'Institut National de Physique Nucleaire et de Physique des Particules, CNRS/IN2P3, Villeurbanne, France
D Amram, S Beauceron, B Blancon, G Boudoul, N Chanon, D Contardo, P Depasse, H El Mamouni, J Fay, S Gascon, M Gouzevitch, C Greenberg, G Grenier, B Ille, E Jourd'Huy, M Lethuillier, B Massoteau, L Mirabito, A Purohit, M Vander Donckt, J Xiao
Institut de Physique des 2 Infinis de Lyon (IP2I), Villeurbanne, France
I Lomidze, T Toriashvili[21], Z Tsamalaidze[22]
Georgian Technical University, Tbilisi, Georgia
V Botta, S Consuegra Rodríguez, L Feld, K Klein, M Lipinski, P Nattland, V Oppenländer, A Pauls, D Pérez Adán, N Röwert
RWTH Aachen University, I. Physikalisches Institut, Aachen, Germany
C Daumann, S Diekmann, N Eich, D Eliseev, F Engelke, J Erdmann, M Erdmann, B Fischer, T Hebbeker, K Hoepfner, F Ivone, A Jung, N Kumar, M y Lee, F Mausolf, M Merschmeyer, A Meyer, A Pozdnyakov, W Redjeb, H Reithler, U Sarkar, V Sarkisovi, A Schmidt, C Seth, A Sharma, J L Spah, V Vaulin, S Zaleski
RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany
M R Beckers, C Dziwok, G Flügge, N Hoeflich, T Kress, A Nowack, O Pooth, A Stahl, A Zotz
RWTH Aachen University, III. Physikalisches Institut B, Aachen, Germany

A Abel, M Aldaya Martin, J Alimena, S Amoroso, Y An, I Andreev, J Bach, S Baxter, M Bayatmakou, H Becerril Gonzalez, O Behnke, A Belvedere, F Blekman[23], K Borras[24], A Campbell, S Chatterjee, L X Coll Saravia, G Eckerlin, D Eckstein, E Gallo[23], A Geiser, M Guthoff, A Hinzmann, L Jeppe, M Kasemann, C Kleinwort, R Kogler, M Komm, D Krücker, W Lange, D Leyva Pernia, K -Y Lin, K Lipka[25], W Lohmann[26], J Malvaso, R Mankel, I -A Melzer-Pellmann, M Mendizabal Morentin, A B Meyer, G Milella, K Moral Figueroa, A Mussgiller, L P Nair, J Niedziela, A Nürnberg, J Park, E Ranken, A Raspereza, D Rastorguev, L Rygaard, M Scham[27,24], S Schnake[24], P Schütze, C Schwanenberger[23], D Schwarz, D Selivanova, K Sharko, M Shchedrolosiev, D Stafford, M Torkian, A Ventura Barroso, R Walsh, D Wang, Q Wang, K Wichmann, L Wiens[24], C Wissing, Y Yang, S Zakharov, A Zimermmane Castro Santos
Deutsches Elektronen-Synchrotron, Hamburg, Germany
A R Alves Andrade, M Antonello, S Bollweg, M Bonanomi, L Ebeling, K El Morabit, Y Fischer, M Frahm, E Garutti, A Grohsjean, A A Guvenli, J Haller, D Hundhausen, G Kasieczka, P Keicher, R Klanner, W Korcari, T Kramer, C c Kuo, F Labe, J Lange, A Lobanov, J Matthiesen, L Moureaux, K Nikolopoulos, A Paasch, K J Pena Rodriguez, N Prouvost, T Quadfasel, B Raciti, M Rieger, D Savoiu, P Schleper, M Schröder, J Schwandt, M Sommerhalder, H Stadie, G Steinbrück, T Von Schwartz, R Ward, B Wiederspan, M Wolf, C Yede
University of Hamburg, Hamburg, Germany
A Bal, S Brommer, A Brusamolino, E Butz, Y M Chen, T Chwalek, A Dierlamm, G G Dincer, D Druzhkin, U Elicabuk, N Faltermann, M Giffels, A Gottmann, F Hartmann[28], M Horzela, F Hummer, U Husemann, J Kieseler, M Klute, J Knolle, R Kunnilan Muhammed Rafeek, O Lavoryk, J M Lawhorn, A Lintuluoto, S Maier, A A Monsch, M Mormile, Th Müller, E Pfeffer, M Presilla, G Quast, K Rabbertz, B Regnery, R Schmieder, N Shadskiy, I Shvetsov, H J Simonis, L Sowa, L Stockmeier, K Tauqeer, M Toms, B Topko, N Trevisani, C Verstege, T Voigtländer, R F Von Cube, J Von Den Driesch, M Wassmer, C Winter, R Wolf, W D Zeuner, X Zuo
Karlsruher Institut fuer Technologie, Karlsruhe, Germany
G Anagnostou, G Daskalakis, A Kyriakis
Institute of Nuclear and Particle Physics (INPP), NCSR Demokritos, Aghia Paraskevi, Greece
G Melachroinos, Z Painesis, I Paraskevas, N Saoulidou, K Theofilatos, E Tziaferi, E Tzovara, K Vellidis, I Zisopoulos
National and Kapodistrian University of Athens, Athens, Greece
T Chatzistavrou, G Karapostoli, K Kousouris, E Siamarkou, G Tsipolitis
National Technical University of Athens, Athens, Greece
I Bestintzanos, I Evangelou, C Foudas, P Katsoulis, P Kokkas, P G Kosmoglou Kioseoglou, N Manthos, I Papadopoulos, J Strologas
University of Ioánnina, Ioánnina, Greece
C Hajdu, D Horvath[29,30], K Márton, A J Rádl[31], F Sikler, V Veszpremi
HUN-REN Wigner Research Centre for Physics, Budapest, Hungary
M Csanád, K Farkas, A Fehérkuti[32], M M A Gadallah[33], Á Kadlecsik, M León Coello, G Pásztor, G I Veres
MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary
B Ujvari, G Zilizi
Faculty of Informatics, University of Debrecen, Debrecen, Hungary
G Bencze, S Czellar, J Molnar, Z Szillasi
HUN-REN ATOMKI—Institute of Nuclear Research, Debrecen, Hungary
T Csorgo[32], F Nemes[32], T Novak, I Szanyi[34]
Karoly Robert Campus, MATE Institute of Technology, Gyongyos, Hungary
S Bahinipati, S Nayak, R Raturi
IIT Bhubaneswar, Bhubaneswar, India
S Bansal, S B Beri, V Bhatnagar, S Chauhan, N Dhingra[35], A Kaur, H Kaur, M Kaur, S Kumar, T Sheokand, J B Singh, A Singla
Panjab University, Chandigarh, India
A Bhardwaj, A Chhetri, B C Choudhary, A Kumar, A Kumar, M Naimuddin, S Phor, K Ranjan, M K Saini
University of Delhi, Delhi, India

P Palni
Indian Institute of Technology Mandi (IIT-Mandi), Himachal Pradesh, India
S Acharya[36], B Gomber
University of Hyderabad, Hyderabad, India
S Mukherjee
Indian Institute of Technology Kanpur, Kanpur, India
S Bhattacharya, S Das Gupta, S Dutta, S Dutta, S Sarkar
Saha Institute of Nuclear Physics, HBNI, Kolkata, India
M M Ameen, P K Behera, S Chatterjee, G Dash, A Dattamunsi, P Jana, P Kalbhor, S Kamble, J R Komaragiri[37], T Mishra, P R Pujahari, A K Sikdar, R K Singh, P Verma, S Verma, A Vijay
Indian Institute of Technology Madras, Madras, India
B K Sirasva
IISER Mohali, India, Mohali, India
L Bhatt, S Dugad, G B Mohanty, M Shelake, P Suryadevara
Tata Institute of Fundamental Research-A, Mumbai, India
A Bala, S Banerjee, S Barman[38], R M Chatterjee, M Guchait, Sh Jain, A Jaiswal, S Kumar, M Maity[38], G Majumder, K Mazumdar, S Parolia, R Saxena, A Thachayath
Tata Institute of Fundamental Research-B, Mumbai, India
D Maity[39], P Mal, K Naskar[39], A Nayak[39], K Pal, P Sadangi, S K Swain, S Varghese[39], D Vats[39]
National Institute of Science Education and Research, An OCC of Homi Bhabha National Institute, Bhubaneswar, Odisha, India
S Dube, P Hazarika, B Kansal, A Laha, R Sharma, S Sharma, K Y Vaish
Indian Institute of Science Education and Research (IISER), Pune, India
S Ghosh
Indian Institute of Technology Hyderabad, Telangana, India
H Bakhshiansohi[40], A Jafari[41], V Sedighzadeh Dalavi, M Zeinali[42]
Isfahan University of Technology, Isfahan, Iran
S Bashiri, S Chenarani[43], S M Etesami, Y Hosseini, M Khakzad, E Khazaie, M Mohammadi Najafabadi, S Tizchang[44]
Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
M Felcini, M Grunewald
University College Dublin, Dublin, Ireland
M Abbrescia[a,b], M Barbieri[a,b], M Buonsante[a,b], A Colaleo[a,b], D Creanza[a,c], N De Filippis[a,c], M De Palma[a,b], W Elmetenawee[a,b,45], N Ferrara[a,c], L Fiore[a], L Generoso[a,b], L Longo[a], M Louka[a,b], G Maggi[a,c], M Maggi[a], I Margjeka[a], V Mastrapasqua[a,b], S My[a,b], F Nenna[a,b], S Nuzzo[a,b], A Pellecchia[a,b], A Pompili[a,b], G Pugliese[a,c], R Radogna[a,b], D Ramos[a], A Ranieri[a], L Silvestris[a], F M Simone[a,c], Ü Sözbilir[a], A Stamerra[a,b], D Troiano[a,b], R Venditti[a,b], P Verwilligen[a], A Zaza[a,b]
INFN Sezione di Bari[a], Università di Bari[b], Politecnico di Bari[c], Bari, Italy
G Abbiendi[a], C Battilana[a,b], D Bonacorsi[a,b], P Capiluppi[a,b], F R Cavallo[a], M Cuffiani[a,b], G M Dallavalle[a], T Diotalevi[a,b], F Fabbri[a], A Fanfani[a,b], R Farinelli[a], D Fasanella[a], P Giacomelli[a], L Giommi[a,b], C Grandi[a], L Guiducci[a,b], M Lorusso[a,b], L Lunerti[a], G Masetti[a], F L Navarria[a,b], G Paggi[a,b], A Perrotta[a], A M Rossi[a,b], S Rossi Tisbeni[a,b], T Rovelli[a,b]
INFN Sezione di Bologna[a], Università di Bologna[b], Bologna, Italy
S Costa[a,b,46], A Di Mattia[a], A Lapertosa[a], R Potenza[a,b], A Tricomi[a,b,46]
INFN Sezione di Catania[a], Università di Catania[b], Catania, Italy
J Altork[a,b], P Assiouras[a], G Barbagli[a], G Bardelli[a], M Bartolini[a,b], A Calandri[a,b], B Camaiani[a,b], A Cassese[a], R Ceccarelli[a], V Ciulli[a,b], C Civinini[a], R D'Alessandro[a,b], L Damenti[a,b], E Focardi[a,b], T Kello[a], G Latino[a,b], P Lenzi[a,b], M Lizzo[a], M Meschini[a], S Paoletti[a], A Papanastassiou[a,b], G Sguazzoni[a], L Viliani[a]
INFN Sezione di Firenze[a], Università di Firenze[b], Firenze, Italy
L Benussi, S Bianco, S Meola[47], D Piccolo
INFN Laboratori Nazionali di Frascati, Frascati, Italy
M Alves Gallo Pereira[a], F Ferro[a], E Robutti[a], S Tosi[a,b]
INFN Sezione di Genova[a], Università di Genova[b], Genova, Italy

A Benaglia[a], F Brivio[a], V Camagni[a,b], F Cetorelli[a,b], F De Guio[a,b], M E Dinardo[a,b], P Dini[a], S Gennai[a], R Gerosa[a,b], A Ghezzi[a,b], P Govoni[a,b], L Guzzi[a], M R Kim[a], G Lavizzari[a,b], M T Lucchini[a,b], M Malberti[a], S Malvezzi[a], A Massironi[a], D Menasce[a], L Moroni[a], M Paganoni[a,b], S Palluotto[a,b], D Pedrini[a], A Perego[a,b], G Pizzati[a,b], T Tabarelli de Fatis[a,b]

INFN Sezione di Milano-Bicocca[a], Università di Milano-Bicocca[b], Milano, Italy

S Buontempo[a], C Di Fraia[a,b], F Fabozzi[a,c], L Favilla[a,d], A O M Iorio[a,b], L Lista[a,b,48], P Paolucci[a,28], B Rossi[a]

INFN Sezione di Napoli[a], Università di Napoli 'Federico II'[b], Napoli, Italy; Università della Basilicata[c], Potenza, Italy; Scuola Superiore Meridionale (SSM)[d], Napoli, Italy

P Azzi[a], N Bacchetta[a,49], D Bisello[a,b], L Borella[a], P Bortignon[a,c], G Bortolato[a,b], A C M Bulla[a,c], R Carlin[a,b], T Dorigo[a,50], F Gasparini[a,b], S Giorgetti[a], E Lusiani[a], M Margoni[a,b], A T Meneguzzo[a,b], F Montecassiano[a], J Pazzini[a,b], F Primavera[a,b], P Ronchese[a,b], R Rossin[a,b], F Simonetto[a,b], M Tosi[a,b], A Triossi[a,b], S Ventura[a], M Zanetti[a,b], P Zotto[a,b], A Zucchetta[a,b], G Zumerle[a,b]

INFN Sezione di Padova[a], Università di Padova[b], Padova, Italy; Universita degli Studi di Cagliari[c], Cagliari, Italy

A Braghieri[a], M Brunoldi[a,b], S Calzaferri[a,b], P Montagna[a,b], M Pelliccioni[a,b], V Re[a], C Riccardi[a,b], P Salvini[a], I Vai[a,b], P Vitulo[a,b]

INFN Sezione di Pavia[a], Università di Pavia[b], Pavia, Italy

S Ajmal[a,b], M E Ascioti[a,b], G M Bilei[†,a], C Carrivale[a,b], D Ciangottini[a,b], L Della Penna[a,b], L Fanò[a,b], V Mariani[a,b], M Menichelli[a], F Moscatelli[a,51], A Rossi[a,b], A Santocchia[a,b], D Spiga[a], T Tedeschi[a,b]

INFN Sezione di Perugia[a], Università di Perugia[b], Perugia, Italy

C Aimè[a,b], C A Alexe[a,c], P Asenov[a,b], P Azzurri[a], G Bagliesi[a], L Bianchini[a,b], T Boccali[a], E Bossini[a], D Bruschini[a,c], R Castaldi[a], F Cattafesta[a,c], M A Ciocci[a,d], M Cipriani[a,b], R Dell'Orso[a], S Donato[a,b], R Forti[a,b], A Giassi[a], F Ligabue[a,c], A C Marini[a,b], A Messineo[a,b], S Mishra[a], V K Muraleedharan Nair Bindhu[a,b], S Nandan[a], F Palla[a], M Riggirello[a,c], A Rizzi[a,b], G Rolandi[a,c], S Roy Chowdhury[a,52], T Sarkar[a], A Scribano[a], P Solanki[a,b], P Spagnolo[a], F Tenchini[a,b], R Tenchini[a], G Tonelli[a,b], N Turini[a,d], F Vaselli[a,c], A Venturi[a], P G Verdini[a]

INFN Sezione di Pisa[a], Università di Pisa[b], Scuola Normale Superiore di Pisa[c], Pisa, Italy; Università di Siena[d], Siena, Italy

P Akrap[a,b], C Basile[a,b], S C Behera[a], F Cavallari[a], L Cunqueiro Mendez[a,b], F De Riggi[a,b], D Del Re[a,b], E Di Marco[a], M Diemoz[a], F Errico[a], L Frosina[a,b], R Gargiulo[a,b], B Harikrishnan[a,b], F Lombardi[a,b], E Longo[a,b], L Martikainen[a,b], J Mijuskovic[a,b], G Organtini[a,b], N Palmeri[a,b], R Paramatti[a,b], S Rahatlou[a,b], C Rovelli[a], F Santanastasio[a,b], L Soffi[a], V Vladimirov[a,b]

INFN Sezione di Roma[a], Sapienza Università di Roma[b], Roma, Italy

N Amapane[a,b], R Arcidiacono[a,c], S Argiro[a,b], M Arneodo[†,a,c], N Bartosik[a,c], R Bellan[a,b], A Bellora[a,b], C Biino[a], C Borca[a,b], N Cartiglia[a], M Costa[a,b], R Covarelli[a,b], N Demaria[a], L Finco[a], M Grippo[a,b], B Kiani[a,b], L Lanteri[a,b], F Legger[a], F Luongo[a,b], C Mariotti[a], S Maselli[a], A Mecca[a,b], L Menzio[a,b], P Meridiani[a], E Migliore[a,b], M Monteno[a], M M Obertino[a,b], G Ortona[a], L Pacher[a,b], N Pastrone[a], M Ruspa[a,c], F Siviero[a,b], V Sola[a,b], A Solano[a,b], A Staiano[a], C Tarricone[a,b], D Trocino[a], G Umoret[a,b], E Vlasov[a,b], R White[a,b]

INFN Sezione di Torino[a], Università di Torino[b], Torino, Italy; Università del Piemonte Orientale[c], Novara, Italy

J Babbar[a,b,52], S Belforte[a], V Candelise[a,b], M Casarsa[a], F Cossutti[a], K De Leo[a], G Della Ricca[a,b], R Delli Gatti[a,b]

INFN Sezione di Trieste[a], Università di Trieste[b], Trieste, Italy

S Dogra, J Hong, J Kim, T Kim, D Lee, H Lee, J Lee, S W Lee, C S Moon, Y D Oh, S Sekmen, B Tae, Y C Yang

Kyungpook National University, Daegu, Korea

M S Kim

Department of Mathematics and Physics—GWNU, Gangneung, Korea

G Bak, P Gwak, H Kim, D H Moon, J Seo

Chonnam National University, Institute for Universe and Elementary Particles, Kwangju, Korea

E Asilar, F Carnevali, J Choi[53], T J Kim, Y Ryou

Hanyang University, Seoul, Korea

S Ha, S Han, B Hong, J Kim, K Lee, K S Lee, S Lee, J Yoo
Korea University, Seoul, Korea
J Goh, J Shin, S Yang
Kyung Hee University, Department of Physics, Seoul, Korea
Y Kang, H S Kim, Y Kim, B Ko, S Lee
Sejong University, Seoul, Korea
J Almond, J H Bhyun, J Choi, J Choi, W Jun, H Kim, J Kim, T Kim, Y Kim, Y W Kim, S Ko, H Lee, J Lee, J Lee, B H Oh, J Shin, U K Yang, I Yoon
Seoul National University, Seoul, Korea
W Jang, D Y Kang, D Kim, S Kim, J S H Lee, Y Lee, I C Park, Y Roh, I J Watson
University of Seoul, Seoul, Korea
G Cho, K Hwang, B Kim, S Kim, K Lee, H D Yoo
Yonsei University, Department of Physics, Seoul, Korea
Y Lee, I Yu
Sungkyunkwan University, Suwon, Korea
T Beyrouthy, Y Gharbia
College of Engineering and Technology, American University of the Middle East (AUM), Dasman, Kuwait
F Alazemi
Kuwait University—College of Science—Department of Physics, Safat, Kuwait
K Dreimanis, O M Eberlins, A Gaile, C Munoz Diaz, D Osite, G Pikurs, R Plese, A Potrebko, M Seidel, D Sidiropoulos Kontos
Riga Technical University, Riga, Latvia
N R Strautnieks
University of Latvia (LU), Riga, Latvia
M Ambrozas, A Juodagalvis, S Nargelas, A Rinkevicius, G Tamulaitis
Vilnius University, Vilnius, Lithuania
I Yusuff[54], Z Zolkapli
National Centre for Particle Physics, Universiti Malaya, Kuala Lumpur, Malaysia
J F Benitez, A Castaneda Hernandez, A Cota Rodriguez, L E Cuevas Picos, H A Encinas Acosta, L G Gallegos Maríñez, J A Murillo Quijada, L Valencia Palomo
Universidad de Sonora (UNISON), Hermosillo, Mexico
G Ayala, H Castilla-Valdez, H Crotte Ledesma, R Lopez-Fernandez, J Mejia Guisao, R Reyes-Almanza, A Sánchez Hernández
Centro de Investigacion y de Estudios Avanzados del IPN, Mexico City, Mexico
C Oropeza Barrera, D L Ramirez Guadarrama, M Ramírez García
Universidad Iberoamericana, Mexico City, Mexico
I Bautista, F E Neri Huerta, I Pedraza, H A Salazar Ibarguen, C Uribe Estrada
Benemerita Universidad Autonoma de Puebla, Puebla, Mexico
I Bubanja, N Raicevic
University of Montenegro, Podgorica, Montenegro
P H Butler
University of Canterbury, Christchurch, New Zealand
A Ahmad, M I Asghar, A Awais, M I M Awan, W A Khan
National Centre for Physics, Quaid-I-Azam University, Islamabad, Pakistan
V Avati, L Forthomme, L Grzanka, M Malawski, K Piotrzkowski
AGH University of Krakow, Krakow, Poland
M Bluj, M Górski, M Kazana, M Szleper, P Zalewski
National Centre for Nuclear Research, Swierk, Poland
K Bunkowski, K Doroba, A Kalinowski, M Konecki, J Krolikowski, A Muhammad
Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Warsaw, Poland
P Fokow, K Pozniak, W Zabolotny
Warsaw University of Technology, Warsaw, Poland
M Araujo, D Bastos, C Beirão Da Cruz E Silva, A Boletti, M Bozzo, T Camporesi, G Da Molin, M Gallinaro, J Hollar, N Leonardo, G B Marozzo, A Petrilli, M Pisano, J Seixas, J Varela, J W Wulff
Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal
P Adzic, L Markovic, P Milenovic, V Milosevic
Faculty of Physics, University of Belgrade, Belgrade, Serbia

D Devetak, M Dordevic, J Milosevic, L Nadderd, V Rekovic, M Stojanovic
VINCA Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia
M Alcalde Martinez, J Alcaraz Maestre, Cristina F Bedoya, J A Brochero Cifuentes, Oliver M Carretero, M Cepeda, M Cerrada, N Colino, B De La Cruz, A Delgado Peris, A Escalante Del Valle, D Fernández Del Val, J P Fernández Ramos, J Flix, M C Fouz, M Gonzalez Hernandez, O Gonzalez Lopez, S Goy Lopez, J M Hernandez, M I Josa, J Llorente Merino, C Martin Perez, E Martin Viscasillas, D Moran, C M Morcillo Perez, Á Navarro Tobar, R Paz Herrera, A Pérez-Calero Yzquierdo, J Puerta Pelayo, I Redondo, J Vazquez Escobar
Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
J F de Trocóniz
Universidad Autónoma de Madrid, Madrid, Spain
B Alvarez Gonzalez, J Ayllon Torresano, A Cardini, J Cuevas, J Del Riego Badas, D Estrada Acevedo, J Fernandez Menendez, S Folgueras, I Gonzalez Caballero, P Leguina, M Obeso Menendez, E Palencia Cortezon, J Prado Pico, A Soto Rodríguez, P Vischia
Universidad de Oviedo, Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), Oviedo, Spain
S Blanco Fernández, I J Cabrillo, A Calderon, J Duarte Campderros, M Fernandez, G Gomez, C Lasaosa García, R Lopez Ruiz, C Martinez Rivero, P Martinez Ruiz del Arbol, F Matorras, P Matorras Cuevas, E Navarrete Ramos, J Piedra Gomez, C Quintana San Emeterio, L Scodellaro, I Vila, R Vilar Cortabitarte, J M Vizan Garcia
Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain
B Kailasapathy[55], D D C Wickramarathna
University of Colombo, Colombo, Sri Lanka
W G D Dharmaratna[56], K Liyanage, N Perera
University of Ruhuna, Department of Physics, Matara, Sri Lanka
D Abbaneo, C Amendola, R Ardino, E Auffray, J Baechler, D Barney, J Bendavid, M Bianco, A Bocci, L Borgonovi, C Botta, A Bragagnolo, C E Brown, C Caillol, G Cerminara, P Connor, K Cormier, D d'Enterria, A Dabrowski, A David, A De Roeck, M M Defranchis, M Deile, M Dobson, P J Fernández Manteca, B A Fontana Santos Alves, E Fontanesi, W Funk, A Gaddi, S Giani, D Gigi, K Gill, F Glege, M Glowacki, A Gruber, J Hegeman, J K Heikkilä, R Hofsaess, B Huber, T James, P Janot, O Kaluzinska, O Karacheban[26], G Karathanasis, S Laurila, P Lecoq, E Leutgeb, C Lourenço, A -M Lyon, M Magherini, L Malgeri, M Mannelli, A Mehta, F Meijers, J A Merlin, S Mersi, E Meschi, M Migliorini, F Monti, F Moortgat, M Mulders, M Musich, I Neutelings, S Orfanelli, F Pantaleo, M Pari, G Petrucciani, A Pfeiffer, M Pierini, M Pitt, H Qu, D Rabady, A Reimers, B Ribeiro Lopes, F Riti, P Rosado, M Rovere, H Sakulin, R Salvatico, S Sanchez Cruz, S Scarfi, M Selvaggi, K Shchelina, P Silva, P Sphicas[57], A G Stahl Leiton, A Steen, S Summers, D Treille, P Tropea, E Vernazza, J Wanczyk[58], S Wuchterl, M Zarucki, P Zehetner, P Zejdl, G Zevi Della Porta
CERN, European Organization for Nuclear Research, Geneva, Switzerland
T Bevilacqua[59], L Caminada[59], W Erdmann, R Horisberger, Q Ingram, H C Kaestli, D Kotlinski, C Lange, U Langenegger, A Nigamova, L Noehte[59], T Rohe, A Samalan
PSI Center for Neutron and Muon Sciences, Villigen, Switzerland
T K Aarrestad, M Backhaus, G Bonomelli, C Cazzaniga, K Datta, P De Bryas Dexmiers D'Archiacchiac[58], A De Cosa, G Dissertori, M Dittmar, M Donegà, F Glessgen, C Grab, N Härringer, T G Harte, W Lustermann, M Malucchi, R A Manzoni, L Marchese, A Mascellani[58], F Nessi-Tedaldi, F Pauss, B Ristic, R Seidita, J Steggemann[58], A Tarabini, D Valsecchi, R Wallny
ETH Zurich—Institute for Particle Physics and Astrophysics (IPA), Zurich, Switzerland
C Amsler[60], P Bärtschi, F Bilandzija, M F Canelli, G Celotto, V Guglielmi, A Jofrehei, B Kilminster, T H Kwok, S Leontsinis, V Lukashenko, A Macchiolo, F Meng, M Missiroli, J Motta, P Robmann, E Shokr, F Stäger, R Tramontano, P Viscone
Universität Zürich, Zurich, Switzerland
D Bhowmik, C M Kuo, P K Rout, S Taj, P C Tiwari[37]
National Central University, Chung-Li, Taiwan
L Ceard, K F Chen, Z g Chen, A De Iorio, W -S Hou, T h Hsu, Y w Kao, S Karmakar, G Kole, Y y Li, R -S Lu, E Paganis, X f Su, J Thomas-Wilsker, L s Tsai, D Tsionou, H y Wu, E Yazgan
National Taiwan University (NTU), Taipei, Taiwan

C Asawatangtrakuldee, N Srimanobhas
High Energy Physics Research Unit, Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok, Thailand
Y Maghrbi
Tunis El Manar University, Tunis, Tunisia
D Agyel, F Dolek, I Dumanoglu[61], Y Guler[62], E Gurpinar Guler[62], C Isik, O Kara[63], A Kayis Topaksu, Y Komurcu, G Onengut, K Ozdemir[64], B Tali[65], U G Tok, E Uslan, I S Zorbakir
Çukurova University, Physics Department, Science and Art Faculty, Adana, Turkey
S Sen
Hacettepe University, Ankara, Turkey
M Yalvac[66]
Middle East Technical University, Physics Department, Ankara, Turkey
B Akgun, I O Atakisi[67], E Gülmez, M Kaya[68], O Kaya[69], M A Sarkisla[70], S Tekten[71]
Bogazici University, Istanbul, Turkey
D Boncukcu, A Cakir, K Cankocak[61,72]
Istanbul Technical University, Istanbul, Turkey
B Hacisahinoglu, I Hos[73], B Kaynak, S Ozkorucuklu, O Potok, H Sert, C Simsek, C Zorbilmez
Istanbul University, Istanbul, Turkey
S Cerci, C Dozen[74], B Isildak, E Simsek, D Sunar Cerci, T Yetkin[74]
Yildiz Technical University, Istanbul, Turkey
A Boyaryntsev, O Dadazhanova, B Grynyov
Institute for Scintillation Materials of National Academy of Science of Ukraine, Kharkiv, Ukraine
L Levchuk
National Science Centre, Kharkiv Institute of Physics and Technology, Kharkiv, Ukraine
J J Brooke, A Bundock, F Bury, E Clement, D Cussans, D Dharmender, H Flacher, J Goldstein, H F Heath, M -L Holmberg, L Kreczko, S Paramesvaran, L Robertshaw, M S Sanjrani[40], J Segal, V J Smith
University of Bristol, Bristol, United Kingdom
A H Ball, K W Bell, A Belyaev[75], C Brew, R M Brown, D J A Cockerill, A Elliot, K V Ellis, J Gajownik, K Harder, S Harper, J Linacre, K Manolopoulos, M Moallemi, D M Newbold, E Olaiya, D Petyt, T Reis, A R Sahasransu, G Salvi, T Schuh, C H Shepherd-Themistocleous, I R Tomalin, K C Whalen, T Williams
Rutherford Appleton Laboratory, Didcot, United Kingdom
I Andreou, R Bainbridge, P Bloch, O Buchmuller, C A Carrillo Montoya, D Colling, I Das, P Dauncey, G Davies, M Della Negra, S Fayer, G Fedi, G Hall, H R Hoorani, A Howard, G Iles, C R Knight, P Krueper, J Langford, K H Law, J León Holgado, L Lyons, A -M Magnan, B Maier, S Mallios, A Mastronikolis, M Mieskolainen, J Nash[76], M Pesaresi, P B Pradeep, B C Radburn-Smith, A Richards, A Rose, L Russell, K Savva, C Seez, R Shukla, A Tapper, K Uchida, G P Uttley, T Virdee[28], M Vojinovic, N Wardle, D Winterbottom
Imperial College, London, United Kingdom
J E Cole, A Khan, P Kyberd, I D Reid
Brunel University, Uxbridge, United Kingdom
S Abdullin, A Brinkerhoff, E Collins, M R Darwish, J Dittmann, K Hatakeyama, V Hegde, J Hiltbrand, B McMaster, J Samudio, S Sawant, C Sutantawibul, J Wilson
Baylor University, Waco, Texas, United States of America
J M Hogan
Bethel University, St. Paul, Minnesota, United States of America
R Bartek, A Dominguez, S Raj, B Sahu, A E Simsek, S S Yu
Catholic University of America, Washington, DC, United States of America
B Bam, A Buchot Perraguin, S Campbell, R Chudasama, S I Cooper, C Crovella, G Fidalgo, S V Gleyzer, A Khukhunaishvili, K Matchev, E Pearson, P Rumerio[77], E Usai, R Yi
The University of Alabama, Tuscaloosa, Alabama, United States of America
S Cholak, G De Castro, Z Demiragli, C Erice, C Fangmeier, C Fernandez Madrazo, J Fulcher, F Golf, S Jeon, J O'Cain, I Reed, J Rohlf, K Salyer, D Sperka, D Spitzbart, I Suarez, A Tsatsos, E Wurtz, A G Zecchinelli
Boston University, Boston, Massachusetts, United States of America

G Barone, G Benelli, D Cutts, S Ellis, L Gouskos, M Hadley, U Heintz, K W Ho, T Kwon, L Lambrecht, G Landsberg, K T Lau, J Luo, S Mondal, J Roloff, T Russell, S Sagir[78], X Shen, M Stamenkovic, N Venkatasubramanian
Brown University, Providence, Rhode Island, United States of America
S Abbott, S Baradia, B Barton, R Breedon, H Cai, M Calderon De La Barca Sanchez, E Cannaert, M Chertok, M Citron, J Conway, P T Cox, F Eble, R Erbacher, O Kukral, G Mocellin, S Ostrom, I Salazar Segovia, J S Tafoya Vargas, W Wei, S Yoo
University of California, Davis, Davis, California, United States of America
K Adamidis, M Bachtis, D Campos, R Cousins, S Crossley, G Flores Avila, J Hauser, M Ignatenko, M A Iqbal, T Lam, Y f Lo, E Manca, A Nunez Del Prado, D Saltzberg, V Valuev
University of California, Los Angeles, California, United States of America
R Clare, J W Gary, G Hanson
University of California, Riverside, Riverside, California, United States of America
A Aportela, A Arora, J G Branson, S Cittolin, S Cooperstein, B D'Anzi, D Diaz, J Duarte, L Giannini, Y Gu, J Guiang, V Krutelyov, R Lee, J Letts, H Li, M Masciovecchio, F Mokhtar, S Mukherjee, M Pieri, D Primosch, M Quinnan, V Sharma, M Tadel, E Vourliotis, F Würthwein, A Yagil, Z Zhao
University of California, San Diego, La Jolla, California, United States of America
A Barzdukas, L Brennan, C Campagnari, S Carron Montero[79], K Downham, C Grieco, M M Hussain, J Incandela, M W K Lai, A J Li, P Masterson, J Richman, S N Santpur, R Schmitz, D Stuart, T Á Vámi, X Yan, D Zhang
University of California, Santa Barbara—Department of Physics, Santa Barbara, California, United States of America
A Albert, S Bhattacharya, A Bornheim, O Cerri, R Kansal, J Mao, H B Newman, G Reales Gutiérrez, T Sievert, M Spiropulu, J R Vlimant, R A Wynne, S Xie
California Institute of Technology, Pasadena, California, United States of America
J Alison, S An, M Cremonesi, V Dutta, E Y Ertorer, T Ferguson, T A Gómez Espinosa, A Harilal, A Kallil Tharayil, M Kanemura, C Liu, M Marchegiani, P Meiring, S Murthy, P Palit, K Park, M Paulini, A Roberts, A Sanchez, W Terrill
Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America
J P Cumalat, W T Ford, A Hart, S Kwan, J Pearkes, C Savard, N Schonbeck, K Stenson, K A Ulmer, S R Wagner, N Zipper, D Zuolo
University of Colorado Boulder, Boulder, Colorado, United States of America
J Alexander, X Chen, J Dickinson, A Duquette, J Fan, X Fan, J Grassi, S Hogan, P Kotamnives, J Monroy, G Niendorf, M Oshiro, J R Patterson, A Ryd, J Thom, P Wittich, R Zou, L Zygala
Cornell University, Ithaca, New York, United States of America
M Albrow, M Alyari, O Amram, G Apollinari, A Apresyan, L A T Bauerdick, D Berry, J Berryhill, P C Bhat, K Burkett, J N Butler, A Canepa, G B Cerati, H W K Cheung, F Chlebana, C Cosby, G Cummings, I Dutta, V D Elvira, J Freeman, A Gandrakota, Z Gecse, L Gray, D Green, A Grummer, S Grünendahl, D Guerrero, O Gutsche, R M Harris, J Hirschauer, V Innocente, B Jayatilaka, S Jindariani, M Johnson, U Joshi, B Klima, S Lammel, C Lee, D Lincoln, R Lipton, T Liu, K Maeshima, D Mason, P McBride, P Merkel, S Mrenna, S Nahn, J Ngadiuba, D Noonan, S Norberg, V Papadimitriou, N Pastika, K Pedro, C Pena[80], C E Perez Lara, V Perovic, F Ravera, A Reinsvold Hall[81], L Ristori, M Safdari, E Sexton-Kennedy, E Smith, N Smith, A Soha, L Spiegel, S Stoynev, J Strait, L Taylor, S Tkaczyk, N V Tran, L Uplegger, E W Vaandering, C Wang, I Zoi
Fermi National Accelerator Laboratory, Batavia, Illinois, United States of America
C Aruta, P Avery, D Bourilkov, P Chang, V Cherepanov, R D Field, C Huh, E Koenig, M Kolosova, J Konigsberg, A Korytov, G Mitselmakher, K Mohrman, A Muthirakalayil Madhu, N Rawal, S Rosenzweig, V Sulimov, Y Takahashi, J Wang
University of Florida, Gainesville, Florida, United States of America
T Adams, A Al Kadhim, A Askew, S Bower, R Goff, R Hashmi, A Hassani, R S Kim, T Kolberg, G Martinez, M Mazza, H Prosper, P R Prova, R Yohay
Florida State University, Tallahassee, Florida, United States of America
B Alsufyani, S Butalla, S Das, M Hohlmann, M Lavinsky, E Yanes
Florida Institute of Technology, Melbourne, Florida, United States of America

M R Adams, N Barnett, A Baty, C Bennett, R Cavanaugh, R Escobar Franco, O Evdokimov, C E Gerber, H Gupta, M Hawksworth, A Hingrajiya, D J Hofman, Z Huang, J h Lee, C Mills, S Nanda, G Nigmatkulov, B Ozek, T Phan, D Pilipovic, R Pradhan, E Prifti, P Roy, T Roy, D Shekar, N Singh, A Thielen, M B Tonjes, N Varelas, M A Wadud, J Yoo
University of Illinois Chicago, Chicago, Illinois, United States of America
M Alhusseini, D Blend, K Dilsiz[82], O K Köseyan, A Mestvirishvili[83], O Neogi, H Ogul[84], Y Onel, A Penzo, C Snyder, E Tiras[85]
The University of Iowa, Iowa City, Iowa, United States of America
B Blumenfeld, J Davis, A V Gritsan, L Kang, S Kyriacou, P Maksimovic, M Roguljic, S Sekhar, M V Srivastav, M Swartz
Johns Hopkins University, Baltimore, Maryland, United States of America
A Abreu, L F Alcerro Alcerro, J Anguiano, S Arteaga Escatel, P Baringer, A Bean, R Bhattacharya, Z Flowers, D Grove, J King, G Krintiras, M Lazarovits, C Le Mahieu, J Marquez, M Murray, M Nickel, S Popescu[86], C Rogan, C Royon, S Rudrabhatla, S Sanders, C Smith, G Wilson
The University of Kansas, Lawrence, Kansas, United States of America
B Allmond, N Islam, A Ivanov, K Kaadze, Y Maravin, J Natoli, G G Reddy, D Roy, G Sorrentino
Kansas State University, Manhattan, Kansas, United States of America
A Baden, A Belloni, J Bistany-riebman, S C Eno, N J Hadley, S Jabeen, R G Kellogg, T Koeth, B Kronheim, S Lascio, P Major, A C Mignerey, C Palmer, C Papageorgakis, M M Paranjpe, E Popova[87], A Shevelev, L Zhang
University of Maryland, College Park, Maryland, United States of America
C Baldenegro Barrera, H Bossi, S Bright-Thonney, I A Cali, Y c Chen, P c Chou, M D'Alfonso, J Eysermans, C Freer, G Gomez-Ceballos, M Goncharov, G Grosso, P Harris, D Hoang, G M Innocenti, K Ivanov, D Kovalskyi, J Krupa, L Lavezzo, Y -J Lee, K Long, C Mcginn, A Novak, M I Park, C Paus, C Reissel, C Roland, G Roland, S Rothman, T a Sheng, G S F Stephans, D Walter, J Wang, Z Wang, B Wyslouch, T J Yang
Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
A Alpana, B Crossman, W J Jackson, C Kapsiak, M Krohn, D Mahon, J Mans, B Marzocchi, R Rusack, O Sancar, R Saradhy, N Strobbe
University of Minnesota, Minneapolis, Minnesota, United States of America
K Bloom, D R Claes, G Haza, J Hossain, C Joo, I Kravchenko, K H M Kwok, A Rohilla, J E Siado, W Tabb, A Vagnerini, A Wightman, F Yan
University of Nebraska-Lincoln, Lincoln, Nebraska, United States of America
H Bandyopadhyay, L Hay, H w Hsia, I Iashvili, A Kalogeropoulos, A Kharchilava, A Mandal, M Morris, D Nguyen, S Rappoccio, H Rejeb Sfar, A Williams, D Yu
State University of New York at Buffalo, Buffalo, New York, United States of America
A Aarif, G Alverson, E Barberis, J Bonilla, B Bylsma, M Campana, J Dervan, Y Haddad, Y Han, I Israr, A Krishna, M Lu, N Manganelli, R Mccarthy, D M Morse, T Orimoto, L Skinnari, C S Thoreson, E Tsai, D Wood
Northeastern University, Boston, Massachusetts, United States of America
S Dittmer, K A Hahn, M Mcginnis, Y Miao, D G Monk, M H Schmitt, A Taliercio, M Velasco, J Wang
Northwestern University, Evanston, Illinois, United States of America
G Agarwal, R Band, R Bucci, S Castells, A Das, A Datta, A Ehnis, R Goldouzian, M Hildreth, K Hurtado Anampa, T Ivanov, C Jessop, A Karneyeu, K Lannon, J Lawrence, N Loukas, L Lutton, J Mariano, N Marinelli, T McCauley, C Mcgrady, C Moore, Y Musienko[22], H Nelson, M Osherson, A Piccinelli, R Ruchti, A Townsend, Y Wan, M Wayne, H Yockey
University of Notre Dame, Notre Dame, Indiana, United States of America
M Carrigan, R De Los Santos, L S Durkin, C Hill, M Joyce, D A Wenzl, B L Winer, B R Yates
The Ohio State University, Columbus, Ohio, United States of America
H Bouchamaoui, G Dezoort, P Elmer, A Frankenthal, M Galli, B Greenberg, N Haubrich, K Kennedy, G Kopp, Y Lai, D Lange, A Loeliger, D Marlow, I Ojalvo, J Olsen, F Simpson, D Stickland, C Tully
Princeton University, Princeton, New Jersey, United States of America

S Malik, R Sharma
University of Puerto Rico, Mayaguez, Puerto Rico, United States of America
S Chandra, A Gu, L Gutay, M Huwiler, M Jones, A W Jung, D Kondratyev, J Li, M Liu, G Negro, N Neumeister, G Paspalaki, S Piperov, N R Saha, J F Schulte, F Wang, A Wildridge, W Xie, Y Yao, Y Zhong
Purdue University, West Lafayette, Indiana, United States of America
N Parashar, A Pathak, E Shumka
Purdue University Northwest, Hammond, Indiana, United States of America
D Acosta, A Agrawal, C Arbour, T Carnahan, P Das, K M Ecklund, F J M Geurts, T Huang, I Krommydas, N Lewis, W Li, J Lin, O Miguel Colin, B P Padley, R Redjimi, J Rotter, C Vico Villalba, M Wulansatiti, E Yigitbasi, Y Zhang
Rice University, Houston, Texas, United States of America
O Bessidskaia Bylund, A Bodek, P de Barbaro†, R Demina, A Garcia-Bellido, H S Hare, O Hindrichs, N Parmar, P Parygin[87], H Seo, R Taus
University of Rochester, Rochester, New York, United States of America
B Chiarito, J P Chou, S V Clark, S Donnelly, D Gadkari, Y Gershtein, E Halkiadakis, C Houghton, D Jaroslawski, A Kobert, I Laflotte, A Lath, J Martins, M Perez Prada, B Rand, J Reichert, P Saha, S Salur, S Schnetzer, D Shih, S Somalwar, R Stone, S A Thayil, S Thomas, J Vora
Rutgers, The State University of New Jersey, Piscataway, New Jersey, United States of America
D Ally, A G Delannoy, S Fiorendi, J Harris, T Holmes, A R Kanuganti, N Karunarathna, J Lawless, L Lee, E Nibigira, B Skipworth, S Spanier
University of Tennessee, Knoxville, Tennessee, United States of America
D Aebi, M Ahmad, T Akhter, K Androsov, A Basnet, A Bolshov, O Bouhali[88], A Cagnotta, V D'Amante, R Eusebi, P Flanagan, J Gilmore, Y Guo, T Kamon, S Luo, R Mueller, A Safonov
Texas A&M University, College Station, Texas, United States of America
N Akchurin, J Damgov, Y Feng, N Gogate, W Jin, Y Kazhykarim, K Lamichhane, S W Lee, C Madrid, A Mankel, T Peltola, I Volobouev
Texas Tech University, Lubbock, Texas, United States of America
E Appelt, Y Chen, S Greene, A Gurrola, W Johns, R Kunnawalkam Elayavalli, A Melo, D Rathjens, F Romeo, P Sheldon, S Tuo, J Velkovska, J Viinikainen, J Zhang
Vanderbilt University, Nashville, Tennessee, United States of America
B Cardwell, H Chung, B Cox, J Hakala, G Hamilton Ilha Machado, R Hirosky, M Jose, A Ledovskoy, C Mantilla, C Neu, C Ramón Álvarez, Z Wu
University of Virginia, Charlottesville, Virginia, United States of America
S Bhattacharya, P E Karchin
Wayne State University, Detroit, Michigan, United States of America
A Aravind, S Banerjee, K Black, T Bose, E Chavez, S Dasu, P Everaerts, C Galloni, H He, M Herndon, A Herve, C K Koraka, S Lomte, R Loveless, A Mallampalli, A Mohammadi, S Mondal, T Nelson, G Parida, L Pétré, D Pinna, A Savin, V Shang, V Sharma, W H Smith, D Teague, H F Tsoi, W Vetens, A Warden
University of Wisconsin—Madison, Madison, Wisconsin, United States of America
S Afanasiev, V Alexakhin, Yu Andreev, T Aushev, D Budkouski, R Chistov, M Danilov, T Dimova, A Ershov, S Gninenko, I Gorbunov, A Gribushin, A Kamenev, V Karjavine, M Kirsanov, V Klyukhin, O Kodolova[89], V Korenkov, I Korsakov, A Kozyrev, N Krasnikov, A Lanev, A Malakhov, V Matveev, A Nikitenko[90,89], V Palichik, V Perelygin, S Petrushanko, O Radchenko, M Savina, V Shalaev, S Shmatov, S Shulha, Y Skovpen, K Slizhevskiy, V Smirnov, O Teryaev, I Tlisova, A Toropin, N Voytishin, A Zarubin, I Zhizhin
Authors affiliated with an international laboratory covered by a cooperation agreement with CERN
E Boos, V Bunichev, M Dubinin[80], V Savrin, A Snigirev, L Dudko, V Kim[22], V Murzin, V Oreshkin, D Sosnov
Authors affiliated with an institute formerly covered by a cooperation agreement with CERN
†Deceased
[1]Also at Yerevan State University, Yerevan, Armenia
[2]Also at TU Wien, Vienna, Austria
[3]Also at Ghent University, Ghent, Belgium
[4]Also at FACAMP—Faculdades de Campinas, Sao Paulo, Brazil
[5]Also at Universidade Estadual de Campinas, Campinas, Brazil

[6]Also at Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[7]Also at The University of the State of Amazonas, Manaus, Brazil
[8]Also at University of Chinese Academy of Sciences, Beijing, People's Republic of China
[9]Also at University of Chinese Academy of Sciences, Beijing, People's Republic of China
[10]Also at School of Physics, Zhengzhou University, Zhengzhou, People's Republic of China
[11]Now at Henan Normal University, Xinxiang, People's Republic of China
[12]Also at University of Shanghai for Science and Technology, Shanghai, People's Republic of China
[13]Also at The University of Iowa, Iowa City, Iowa, United States of America
[14]Also at Nanjing Normal University, Nanjing, People's Republic of China
[15]Also at Center for High Energy Physics, Peking University, Beijing, People's Republic of China
[16]Also at Cairo University, Cairo, Egypt
[17]Also at Suez University, Suez, Egypt
[18]Now at British University in Egypt, Cairo, Egypt
[19]Also at Purdue University, West Lafayette, Indiana, United States of America
[20]Also at Université de Haute Alsace, Mulhouse, France
[21]Also at Tbilisi State University, Tbilisi, Georgia
[22]Also at an institute formerly covered by a cooperation agreement with CERN
[23]Also at University of Hamburg, Hamburg, Germany
[24]Also at RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany
[25]Also at Bergische University Wuppertal (BUW), Wuppertal, Germany
[26]Also at Brandenburg University of Technology, Cottbus, Germany
[27]Also at Forschungszentrum Jülich, Juelich, Germany
[28]Also at CERN, European Organization for Nuclear Research, Geneva, Switzerland
[29]Also at HUN-REN ATOMKI—Institute of Nuclear Research, Debrecen, Hungary
[30]Now at Universitatea Babes-Bolyai—Facultatea de Fizica, Cluj-Napoca, Romania
[31]Also at MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary
[32]Also at HUN-REN Wigner Research Centre for Physics, Budapest, Hungary
[33]Also at Physics Department, Faculty of Science, Assiut University, Assiut, Egypt
[34]Also at The University of Kansas, Lawrence, Kansas, United States of America
[35]Also at Punjab Agricultural University, Ludhiana, India
[36]Also at University of Hyderabad, Hyderabad, India
[37]Also at Indian Institute of Science (IISc), Bangalore, India
[38]Also at University of Visva-Bharati, Santiniketan, India
[39]Also at Institute of Physics, Bhubaneswar, India
[40]Also at Deutsches Elektronen-Synchrotron, Hamburg, Germany
[41]Also at Isfahan University of Technology, Isfahan, Iran
[42]Also at Sharif University of Technology, Tehran, Iran
[43]Also at Department of Physics, University of Science and Technology of Mazandaran, Behshahr, Iran
[44]Also at Department of Physics, Faculty of Science, Arak University, ARAK, Iran
[45]Also at Helwan University, Cairo, Egypt
[46]Also at Centro Siciliano di Fisica Nucleare e di Struttura Della Materia, Catania, Italy
[47]Also at Università degli Studi Guglielmo Marconi, Roma, Italy
[48]Also at Scuola Superiore Meridionale, Università di Napoli 'Federico II', Napoli, Italy
[49]Also at Fermi National Accelerator Laboratory, Batavia, Illinois, United States of America
[50]Also at Lulea University of Technology, Lulea, Sweden
[51]Also at Consiglio Nazionale delle Ricerche—Istituto Officina dei Materiali, Perugia, Italy
[52]Also at UPES—University of Petroleum and Energy Studies, Dehradun, India
[53]Also at Institut de Physique des 2 Infinis de Lyon (IP2I), Villeurbanne, France
[54]Also at Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia
[55]Also at Trincomalee Campus, Eastern University, Sri Lanka, Nilaveli, Sri Lanka
[56]Also at Saegis Campus, Nugegoda, Sri Lanka
[57]Also at National and Kapodistrian University of Athens, Athens, Greece
[58]Also at Ecole Polytechnique Fédérale Lausanne, Lausanne, Switzerland
[59]Also at Universität Zürich, Zurich, Switzerland
[60]Also at Stefan Meyer Institute for Subatomic Physics, Vienna, Austria
[61]Also at Near East University, Research Center of Experimental Health Science, Mersin, Turkey
[62]Also at Konya Technical University, Konya, Turkey

[63]Also at Istanbul Topkapi University, Istanbul, Turkey
[64]Also at Izmir Bakircay University, Izmir, Turkey
[65]Also at Adiyaman University, Adiyaman, Turkey
[66]Also at Bozok Universitetesi Rektörlügü, Yozgat, Turkey
[67]Also at Istanbul Sabahattin Zaim University, Istanbul, Turkey
[68]Also at Marmara University, Istanbul, Turkey
[69]Also at Milli Savunma University, Istanbul, Turkey
[70]Also at Informatics and Information Security Research Center, Gebze/Kocaeli, Turkey
[71]Also at Kafkas University, Kars, Turkey
[72]Now at Istanbul Okan University, Istanbul, Turkey
[73]Also at Istanbul University—Cerrahpasa, Faculty of Engineering, Istanbul, Turkey
[74]Also at Istinye University, Istanbul, Turkey
[75]Also at School of Physics and Astronomy, University of Southampton, Southampton, United Kingdom
[76]Also at Monash University, Faculty of Science, Clayton, Australia
[77]Also at Università di Torino, Torino, Italy
[78]Also at Karamanoğlu Mehmetbey University, Karaman, Turkey
[79]Also at California Lutheran University, Thousand Oaks, California, United States of America
[80]Also at California Institute of Technology, Pasadena, California, United States of America
[81]Also at United States Naval Academy, Annapolis, Maryland, United States of America
[82]Also at Bingol University, Bingol, Turkey
[83]Also at Georgian Technical University, Tbilisi, Georgia
[84]Also at Sinop University, Sinop, Turkey
[85]Also at Erciyes University, Kayseri, Turkey
[86]Also at Horia Hulubei National Institute of Physics and Nuclear Engineering (IFIN-HH), Bucharest, Romania
[87]Now at another institute formerly covered by a cooperation agreement with CERN
[88]Also at Hamad Bin Khalifa University (HBKU), Doha, Qatar
[89]Also at Yerevan Physics Institute, Yerevan, Armenia
[90]Also at Imperial College, London, United Kingdom

References

[1] ATLAS Collaboration 2008 The ATLAS experiment at the CERN large hadron collider *J. Instrum.* **3** S08003
[2] CMS Collaboration 2008 The CMS experiment at the CERN LHC *J. Instrum.* **3** S08004
[3] CMS Collaboration 2025 Performance of heavy-flavour jet identification in Lorentz-boosted topologies in proton-proton collisions at $\sqrt{s} = 13$ TeV *J. Instrum.* **20** 11006
[4] Feickert M and Nachman B 2021 A living review of machine learning for particle physics (arXiv:2102.02770)
[5] Navas S *et al* (Particle Data Group) 2024 Review of particle physics *Phys. Rev. D* **110** 030001
[6] Kasieczka G *et al* 2021 The LHC olympics 2020 a community challenge for anomaly detection in high energy physics *Rept. Prog. Phys.* **84** 124201
[7] Belis V, Odagiu P and Aarrestad T K 2024 Machine learning for anomaly detection in particle physics *Rev. Phys.* **12** 100091
[8] CMS Collaboration 2025 Model-agnostic search for dijet resonances with anomalous jet substructure in proton–proton collisions at $\sqrt{s} = 13$ TeV *Rep. Prog. Phys.* **88** 067802
[9] D0 Collaboration 2000 Search for new physics in eμX data at DØ using SLEUTH: a quasi-model-independent search strategy for new physics *Phys. Rev. D* **62** 092004
[10] D0 Collaboration 2001 Quasi-model-independent search for new physics at large transverse momentum *Phys. Rev. D* **64** 012004
[11] D0 Collaboration 2001 Quasi-Model-Independent Search for New High p_t Pphysics at D0 *Phys. Rev. Lett.* **86** 3712
[12] H1 Collaboration 2009 A general search for new phenomena at HERA *Phys. Lett. B* **674** 257
[13] H1 Collaboration 2004 A general search for new phenomena in ep scattering at HERA *Phys. Lett. B* **602** 14
[14] CDF Collaboration 2008 Model-independent and quasi-model-independent search for new physics at CDF *Phys. Rev. D* **78** 012002
[15] CDF Collaboration 2007 Model-independent global search for new high-p_T physics at CDF (arXiv:0712.2534)
[16] CDF Collaboration 2009 Global search for new physics with 2.0 fb^{-1} at CDF *Phys. Rev. D* **79** 011101
[17] CMS Collaboration 2021 MUSiC: a model-unspecific search for new physics in proton-proton collisions at $\sqrt{s} = 13$ TeV *Eur. Phys. J. C* **81** 629
[18] ATLAS Collaboration 2019 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
[19] ATLAS Collaboration 2020 Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV *pp* collisions in the ATLAS detector *Phys. Rev. Lett.* **125** 131801
[20] ATLAS Collaboration 2025 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
[21] Aarrestad T *et al* 2022 The dark machines anomaly score challenge: benchmark data and model independent event classification for the Large Hadron Collider *SciPost Phys.* **12** 043
[22] Aguilar-Saavedra J A, Collins J and Mishra R K 2017 A generic anti-QCD jet tagger *J. High Energy Phys.* JHEP11(2017)163

[23] Hallin A, Isaacson J, Kasieczka G, Krause C, Nachman B, Quadfasel T, Schlaffer M, Shih D and Sommerhalder M 2022 Classifying anomalies through outer density estimation *Phys. Rev. D* **106** 055006
[24] Metodiev E M, Nachman B and Thaler J 2017 Classification without labels: learning from mixed samples in high energy physics *J. High Energy Phys.* JHEP10(2017)174
[25] Collins J H, Howe K and Nachman B 2019 Extending the search for new resonances with machine learning *Phys. Rev. D* **99** 014038
[26] Finke T, Krämer M, Lipp M and Mück A 2022 Boosting mono-jet searches with model-agnostic machine learning *J. High Energy Phys.* JHEP08(2022)015
[27] Nachman B and Shih D 2020 Anomaly detection with density estimation *Phys. Rev. D* **101** 075042
[28] Raine J A, 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
[29] Klein S, Raine J A and Golling T 2022 Flows for flows: training normalizing flows between arbitrary distributions with maximum likelihood estimation (arXiv:2211.02487)
[30] Sengupta D, Klein S, Raine J A and Golling T 2024 CURTAINs flows for flows: constructing unobserved regions with maximum likelihood estimation *SciPost Phys.* **17** 046
[31] Sengupta D, Leigh M, Raine J A, Klein S and Golling T 2024 Improving new physics searches with diffusion models for event observables and jet constituents *J. High Energy Phys.* JHEP04(2024)109
[32] Mikuni V and Nachman B 2024 High-dimensional and permutation invariant anomaly detection *SciPost Phys* **16** 062
[33] Golling T, Klein S, Mastandrea R and Nachman B 2023 Flow-enhanced transportation for anomaly detection *Phys. Rev. D* **107** 096025
[34] Jawahar P, Aarrestad T, Chernyavskaya N, Pierini M, Wozniak K A, Ngadiuba J, Duarte J and Tsan S 2022 Improving variational autoencoders for new physics detection at the LHC with normalizing flows *Front. Big Data* **5** 803685
[35] Tsan S *et al* 2021 Particle graph autoencoders and differentiable, learned energy mover's distance *35th Conf. on Neural Information Processing Systems* (arXiv:2111.12849)
[36] Finke T, Krämer M, Morandini A, Mück A and Oleksiyuk I 2021 Autoencoders for unsupervised anomaly detection in high energy physics *J. High Energy Phys.* JHEP06(2021)161
[37] Vaslin L, Barra V and Donini J 2023 GAN-AE: an anomaly detection algorithm for new physics search in LHC data *Eur. Phys. J. C* **83** 1008
[38] Anzalone L, Singh Chhibra S, Maier B, Chernyavskaya N and Pierini M 2024 Triggering dark showers with conditional dual auto-encoders *Mach. Learn. Sci. Tech.* **5** 035064
[39] 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.* JHEP05(2019)036
[40] Dillon B M, Plehn T, Sauer C and Sorrenson P 2021 Better latent spaces for better autoencoders *SciPost Phys.* **11** 061
[41] Cheng T, Arguin J-F, Leissner-Martin J, Pilette J and Golling T 2023 Variational autoencoders for anomalous jet tagging *Phys. Rev. D* **107** 016002
[42] Dillon B M, Favaro L, Plehn T, Sorrenson P and Krämer M 2023 A normalized autoencoder for LHC triggers *SciPost Phys. Core* **6** 074
[43] van Beekveld M *et al* 2021 Combining outlier analysis algorithms to identify new physics at the LHC *J. High Energy Phys.* JHEP09(2021)024
[44] Kuusela M, Vatanen T, Malmi E, Raiko T, Aaltonen T and Nagai Y 2012 Semi-supervised anomaly detection—towards model-independent searches of new physics *J. Phys.: Conf. Ser.* **368** 012032
[45] Govorkova E *et al* 2022 Autoencoders on field-programmable gate arrays for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider *Nat. Mach. Intell.* **4** 154
[46] Mikuni V, Nachman B and Shih D 2022 Online-compatible unsupervised nonresonant anomaly detection *Phys. Rev. D* **105** 055006
[47] Woźniak K A, Cerri O, Duarte J M, Möller T, Ngadiuba J, Nguyen T Q, Pierini M, Spiropulu M and Vlimant J-R 2020 New physics agnostic selections for new physics searches *EPJ Web Conf.* **245** 06039
[48] Knapp O, Cerri O, Dissertori G, Nguyen T Q, Pierini M and Vlimant J R 2021 Adversarially learned anomaly detection on CMS open data: re-discovering the top quark *Eur. Phys. J. Plus* **136** 236
[49] CMS Collaboration 2025 Wasserstein normalized autoencoder for anomaly detection *Mach. Learn.: Sci. Technol.* **7** 035030
[50] Farina M, Nakai Y and Shih D 2020 Searching for new physics with deep autoencoders *Phys. Rev. D* **101** 075021
[51] Heimel T, Kasieczka G, Plehn T and Thompson J M 2019 QCD or what? *SciPost Phys.* **6** 030
[52] ATLAS Collaboration 2023 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
[53] ATLAS Collaboration 2024 Search for new phenomena in two-body invariant mass distributions using unsupervised machine learning for anomaly detection at $\sqrt{s} = 13$ tev with the ATLAS Detector *Phys. Rev. Lett.* **132** 081801
[54] ATLAS Collaboration 2025 Search for new physics in final states with semivisible jets or anomalous signatures using the ATLAS detector *Phys. Rev. D* **112** 012021
[55] Kingma D P and Welling M 2014 Auto-encoding variational Bayes (arXiv:1312.6114)
[56] CMS ECAL Collaboration 2024 Autoencoder-based anomaly detection system for online data quality monitoring of the CMS electromagnetic calorimeter *Comput. Softw. Big Sci.* **8** 11
[57] Bourilkov D 2020 Machine and deep learning applications in particle physics *Int. J. Mod. Phys. A* **34** 1930019
[58] CMS Collaboration 2020 Measurement of the $\mathrm{t\bar{t}b\bar{b}}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13$ TeV *Phys. Lett. B* **803** 135285
[59] Gambhir R, Mastandrea R, Nachman B and Thaler J 2025 Isolating unisolated upsilons with anomaly detection in CMS open data *Phys. Rev. Lett.* **135** 021902
[60] CMS Collaboration 2024 Development of the CMS detector for the CERN LHC Run 3 *J. Instrum.* **19** 05064
[61] CMS Collaboration 2021 Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS *Eur. Phys. J. C* **81** 800
[62] CMS Collaboration 2018 CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV CMS Physics Analysis Summary CMS-PAS-LUM-17-004

[63] CMS Collaboration 2019 CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV CMS Physics Analysis Summary CMS-PAS-LUM-18-002
[64] Larkoski A J, Marzani S, Soyez G and Thaler J 2014 Soft drop *J. High Energy Phys.* JHEP05(2014)146
[65] Thaler J and Van Tilburg K 2011 Identifying boosted objects with N-subjettiness *J. High Energy Phys.* JHEP03(2011)015
[66] CMS Collaboration 2018 Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV *J. Instrum.* **13** P05011
[67] Brust C, Maksimovic P, Sady A, Saraswat P, Walters M T and Xin Y 2015 Identifying boosted new physics with non-isolated leptons *J. High Energy Phys.* JHEP04(2015)079
[68] Alwall J *et al* 2014 The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations *J. High Energy Phys.* JHEP07(2014)079
[69] Alwall J *et al* 2008 Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions *Eur. Phys. J. C* **53** 473
[70] Nason P 2004 A new method for combining NLO QCD with shower Monte Carlo algorithms *J. High Energy Phys.* JHEP11(2004)040
[71] Frixione S, Nason P and Oleari C 2007 Matching NLO QCD computations with parton shower simulations: the POWHEG method *J. High Energy Phys.* JHEP11(2007)070
[72] Alioli S, Nason P, Oleari C and Re E 2010 A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX *J. High Energy Phys.* JHEP06(2010)043
[73] Frixione S, Nason P and Ridolfi G 2007 A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction *J. High Energy Phys.* JHEP09(2007)126
[74] Frederix R, Re E and Torrielli P 2012 Single-top *t*-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO *J. High Energy Phys.* JHEP09(2012)130
[75] Re E 2011 Single-top *Wt*-channel production matched with parton showers using the POWHEG method *Eur. Phys. J. C* **71** 1547
[76] Sjöstrand T, Ask S, Christiansen J R, Corke R, Desai N, Ilten P, Mrenna S, Prestel S, Rasmussen C O and Skands P Z 2015 An introduction to PYTHIA 8.2 *Comput. Phys. Commun.* **191** 159
[77] CMS Collaboration 2020 Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements *Eur. Phys. J. C* **80** 4
[78] GEANT4 Collaboration 2003 GEANT4—a simulation toolkit *Nucl. Instrum. Meth. A* **506** 250
[79] Zenodo record for this analysis 2025 (available at: http://dx.doi.org/10.5281/zenodo.16656501)
[80] Fan H, Su H and Guibas L 2017 A point set generation network for 3D object reconstruction from a single image *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (arXiv:1612.00603)
[81] Kullback S and Leibler R A 1951 On information and sufficiency *Ann. Math. Stat.* **22** 79
[82] Higgins I *et al* 2017 beta-VAE: learning basic visual concepts with a constrained variational framework *Int. Conf. on Learning Representations* (available at: https://openreview.net/forum?id=Sy2fzU9gl)
[83] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)
[84] He K, Zhang X, Ren S and Sun J 2015 Delving deep into rectifiers: surpassing human-level performance on imagenet classification *IEEE Int. Conf. on Computer Vision (ICCV)* p 1026
[85] Demortier L 2008 P values and nuisance parameters *StatIstical Issues For LHC Physics. Proc., Workshop, PHYSTAT-LHC, (Geneva, Switzerland, 27-29 June 2007)* p 23
[86] Collins J H, Howe K and Nachman B 2018 Anomaly detection for resonant new physics with machine learning *Phys. Rev. Lett.* **121** 241803
[87] Clevert D, Unterthiner T and Hochreiter S 2016 Fast and accurate deep network learning by exponential linear units (ELUs) *4th Int. Conf. on Learning Representations (ICLR)* (arXiv:1511.07289)
[88] Srivastava N *et al* 2014 Dropout: a simple way to prevent neural networks from overfitting *J. Mach. Learn. Res.* **15** 1929 (available at: https://jmlr.org/papers/v15/srivastava14a.html)
[89] Hein M et al 2025 How to pick the best anomaly detector? (arXiv:2511.14832)
[90] Amram O and Suarez C M 2021 Tag N' Train: a technique to train improved classifiers on unlabeled data *J. High Energy Phys.* JHEP01(2021)153
[91] Macaluso S and Shih D 2018 Pulling out all the tops with computer vision and deep learning *J. High Energy Phys.* JHEP10(2018)121
[92] Finke T, Hein M, Kasieczka G, Krämer M, Mück A, Prangchaikul P, Quadfasel T, Shih D and Sommerhalder M 2024 Tree-based algorithms for weakly supervised anomaly detection *Phys. Rev. D* **109** 034033
[93] Kingma D P *et al* 2016 Improving variational inference with inverse autoregressive flow *Proc. 30th Int. Conf. on Neural Information Processing Systems* NeurIPS p 4743 (arXiv:1606.04934)
[94] Germain M, Gregor K, Murray I and Larochelle H 2015 MADE: masked autoencoder for distribution estimation *Proc. of the 32nd Int. Conf. on Machine Learning* (PMLR) vol 37 p 881 (arXiv:1502.03509)
[95] Ansel J *et al* 2024 PyTorch 2: faster machine learning through dynamic python bytecode transformation and graph compilation *29th ACM Int. Conf. on Architectural Support for Programming Languages and Operating Systems (ASPLOS '24)* vol 2 p 929
[96] Bingham E *et al* 2019 Pyro: deep universal probabilistic programming *J. Mach. Learn. Res.* **20** 1 (available at: https://jmlr.org/papers/v20/18-403.html)
[97] Phan D, Pradhan N, and Jankowiak M 2019 Composable effects for flexible and accelerated probabilistic programming in NumPyro (arXiv:1912.11554)
[98] Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift *Proc. of the 32nd Int. Conf. on Machine Learning* (arXiv:1502.03167 [PMLR 37 (2015) 448])
[99] Parzen E 1962 On estimation of a probability density function and mode *Ann. Math. Stat.* **33** 1065
[100] Park S E *et al* 2020 Quasi anomalous knowledge: searching for new physics with embedded knowledge *J. High Energy Phys.* JHEP06(2021)030
[101] Durkan C, Bekasov A, Murray I and Papamakarios G 2019 Neural spline flows *Proc. 33rd Int. Conf. on Neural Information Processing Systems NeurIPS* pp 7511 (arXiv:1906.04032)
[102] Durkan C, Bekasov A, Murray I and Papamakarios G 2020 nflows: normalizing flows in PyTorch (available at: http://dx.doi.org/10.5281/zenodo.4296287)
[103] Gallicchio J, Huth J, Kagan M, Schwartz M D, Black K and Tweedie B 2011 Multivariate discrimination and the Higgs + W/Z search *J. High Energy Phys.* JHEP04(2011)069

[104] Kasieczka G and Shih D 2020 Robust jet classifiers through distance correlation *Phys. Rev. Lett.* **125** 122001

[105] Zenati H *et al* 2018 Adversarially learned anomaly detection *20th IEEE Int. Conf. on Data Mining (ICDM)* (arXiv:1812.02288)

[106] CMS Collaboration 2024 The CMS statistical analysis and combination tool: combine *Comput. Softw. Big Sci.* **8** 19

[107] Cowan G, Cranmer K, Gross E and Vitells O 2011 Asymptotic formulae for likelihood-based tests of new physics *Eur. Phys. J. C* **71** 1554

[108] Breimen L 2001 Random forests *Mach. Learn.* **45** 5

[109] CMS Collaboration 2020 Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV *J. Instrum.* **15** 10017

[110] CMS Collaboration 2017 The CMS trigger system *J. Instrum.* **12** 01020

[111] CMS Collaboration 2024 Performance of the CMS high-level trigger during LHC run 2 *J. Instrum.* **19** 11021

[112] CMS Collaboration 2021 Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC *J. Instrum.* **16** 05014

[113] CMS Collaboration 2018 Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV *J. Instrum.* **13** 06015

[114] CMS Collaboration 2014 Description and performance of track and primary-vertex reconstruction with the CMS tracker *J. Instrum.* **9** 10009

[115] CMS Collaboration 2015 Technical proposal for the Phase-II upgrade of the Compact Muon Solenoid CMS Technical Proposal CERN-LHCC-2015-010, CMS-TDR-15-02

[116] CMS Collaboration 2017 Particle-flow reconstruction and global event description with the CMS detector *J. Instrum.* **12** 10003

[117] Cacciari M, Salam G P and Soyez G 2008 The anti-k_{T} jet clustering algorithm *J. High Energy Phys.* JHEP04(2008)063

[118] Cacciari M, Salam G P and Soyez G 2012 FastJet user manual *Eur. Phys. J. C* **72** 1896

[119] CMS Collaboration 2020 Pileup mitigation at CMS in 13 TeV data *J. Instrum.* **15** 09018

[120] Bertolini D, Harris P, Low M and Tran N 2014 Pileup per particle identification *J. High Energy Phys.* JHEP10(2014)059

[121] CMS Collaboration 2017 Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV *J. Instrum.* **12** 02014

[122] CMS Collaboration 2025 A method for correcting the substructure of multiprong jets using the Lund jet plane *J. High Energy Phys.* JHEP11(2025)038

[123] CMS Collaboration 2017 Jet algorithms performance in 13 TeV data, CMS Physics Analysis Summary CMS-PAS-JME-16-003

[124] CMS Collaboration 2026 CMS dijet anomaly detection dataset *Zenodo* (available at: http://dx.doi.org/10.5281/zenodo.16656501)