Anomaly detection in collider physics via factorized observables

Eric M. Metodiev,[1,2,*] Jesse Thaler,[1,2,†] and Raymond Wynne[2,3,4,‡]

[1]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[2]*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, USA*

[3]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[4]*Division of Physics, Mathematics and Astronomy, Caltech, Pasadena, California, USA*

(Received 10 January 2024; accepted 29 July 2024; published 9 September 2024)

To maximize the discovery potential of high-energy colliders, experimental searches should be sensitive to unforeseen new physics scenarios. This goal has motivated the use of machine learning for unsupervised anomaly detection. In this paper, we introduce a new anomaly detection strategy called FORCE: factorized observables for regressing conditional expectations. Our approach is based on the inductive bias of factorization, which is the idea that the physics governing different energy scales can be treated as approximately independent. Assuming factorization holds separately for signal and background processes, the appearance of nontrivial correlations between low- and high-energy observables is a robust indicator of new physics. Under the most restrictive form of factorization, a machine-learned model trained to identify such correlations will in fact converge to the optimal new physics classifier. We test FORCE on a benchmark anomaly detection task for the Large Hadron Collider involving collimated sprays of particles called jets. By teasing out correlations between the kinematics and substructure of jets, our method can reliably extract percent-level signal fractions. This strategy for uncovering new physics adds to the growing toolbox of anomaly detection methods for collider physics with a complementary set of assumptions.

DOI: 10.1103/PhysRevD.110.055012

Despite the excellent targeted search efforts of multiple experiments, no conclusive evidence for new physics has been seen at the Large Hadron Collider (LHC) since the Higgs boson discovery in 2012 [1,2]. It is difficult, however, to exclude the possibility that new physics might exist in a form that has yet to be theoretically predicted. Although targeted searches for a specific scenario (or class of scenarios) might yield a serendipitous discovery, they could lack sensitivity to even sizeable amounts of unforeseen new physics in LHC data. To enable the broadest coverage for collider searches, robust techniques are needed to probe generic deviations from the Standard Model. This goal has inspired the development of several anomaly detection approaches for collider physics [3–103], which have recently found experimental applications [24,95].

Any anomaly detection technique must make assumptions about what constitutes an anomaly, which then implies limitations on its sensitivity. One class of techniques uses comparisons between data and simulation to detect anomalous events [3,5,6]; this approach is susceptible to detector or generator mismodeling and may confuse poorly modeled regions of phase space for new physics. A more data-driven approach assumes that new physics will appear as a localized cluster in phase space [4,12,32]; this is an excellent inductive bias to detect mass resonances, but limits the types of models that can be probed. The most unstructured techniques, such as autoencoder reconstruction losses, operationally define the notion of anomalous events via the choice of machine learning architecture [8,9,13]; since they lack controlled assumptions, it is challenging to determine the applicability of such methods to particular new physics scenarios.

In this paper, we introduce an anomaly detection strategy called FORCE—factorized observables for regressing conditional expectations—based on the inductive bias of factorization. Factorization occurs when the physics governing high-energy scales is approximately independent from those governing low-energy scales. Jet production offers a canonical example of factorization at colliders, where the processes that determine the kinematics and flavors of high-energy partons are approximately independent of the dynamics that yield collimated sprays of low-energy

[*]Contact author: metodiev@mit.edu
[†]Contact author: jthaler@mit.edu
[‡]Contact author: rwynne@caltech.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP^3.

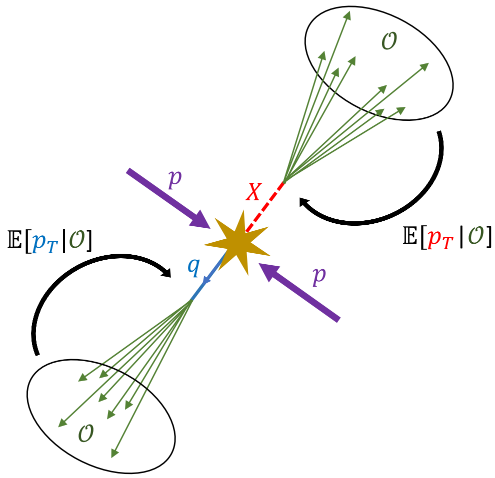


FIG. 1. Illustration of the FORCE anomaly detection approach, applied to a dijet search. A machine-learning model is trained to predict the kinematics of a jet from its substructure. The model output converges to the optimal new physics jet classifier assuming factorization holds for both the signal and background processes.

hadrons. As illustrated in Fig. 1, a machine learning model can be trained to predict the kinematics of a jet from its boost-invariant substructure. Kinematics and substructure are approximately independent in the absence of new physics, so if the model learns nontrivial correlations, then this indicates a possible anomaly. Our approach does not require simulated data, works even if the new physics is nonresonant, and provably converges to the optimal classifier assuming strict factorization. FORCE builds upon previous uses of factorized structures to estimate backgrounds [104,105], train data-driven collider classifiers [21,106–110], and disentangle particle flavors using topic modeling [111–114].

To demonstrate the FORCE approach, we perform a case study involving jets [115–118]. Jets are proxies for the partons or resonances produced in high-energy collisions, with the kinematics of a jet reflecting the kinematics of its initiating particle. Jets then acquire substructure through lower-energy processes, such as decays of intermediate-scale resonances or showering/hadronization in quantum chromodynamics (QCD). Many new physics scenarios involve jet production, making jets a key target for anomaly detection.

In the soft-collinear limit of QCD, the substructure of a jet factorizes from its kinematics [119–123] (see also [124–127]). Factorization also holds for the decay of an intermediate-scale resonance in the narrow width approximation [128,129], such as for a Lorentz-boosted W/Z boson, Higgs boson, or top quark. Therefore, at leading order in the high-p_T limit, the kinematics of a jet is determined by its transverse momentum p_T and rapidity y. Let $\mathcal{O}$ be a list of jet substructure observables, possibly high dimensional. Then, assuming factorization holds, the distribution of jet kinematics and substructure obey

$$P(p_T, y, \mathcal{O}) \approx \sum_i f_i P_i(p_T, y) P_i(\mathcal{O}|p_T), \tag{1}$$

where P_i refers to the probability density function, $i \in \{q, g, W, t, \cdots\}$ labels the types of initiating particle, and f_i is the fraction of jets initiated by that particle type. Factorization imposes a nontrivial constraint that $P_i(\mathcal{O}|p_T)$ is independent of y for each i *and* that a finite sum over i is sufficient to model the distribution.

If we make an even more restrictive assumption that $\mathcal{O}$ consists of scale- and boost-invariant observables, with no conditional p_T dependence, then we can write

$$P_i(\mathcal{O}|p_T) \approx P_i(\mathcal{O}). \tag{2}$$

Examples of such quasi-invariant observables are N-subjettiness ratios [130,131], D_2 [132], D_3 [133], and N_i [134]. Here, we take the jet substructure to be dominated by the initiating particle's flavor and independent of the remainder of the event, up to subleading corrections. The factorized structure of Eqs. (1) and (2) is what we will exploit for anomaly detection using FORCE.

Consider the case of only two jet types: background (B) QCD jets from high-energy quarks and gluons, and signal (S) jets from the hadronic decay of a new particle. To simplify the algebra, we marginalize over y. Via Eqs. (1) and (2), i.e. assuming p_T and $\mathcal{O}$ are independent in both the signal and background processes, the joint distribution of jet kinematics and substructure is

$$P(p_T, \mathcal{O}) = f_S P_S(p_T) P_S(\mathcal{O}) + f_B P_B(p_T) P_B(\mathcal{O}), \tag{3}$$

where f_S is the fraction of new physics events and f_B is the fraction of QCD events, with $f_S + f_B = 1$. Our goal is to discover and characterize the new physics signal in a data-driven manner.

The key insight behind FORCE is that a machine-learned model trained to predict a jet's p_T from its substructure observable $\mathcal{O}$ yields the optimal S versus B classifier (assuming factorization and with sufficient training and statistics). By "predict," we mean learning the conditional expectation value

$$\hat{p}_T(\mathcal{O}) \equiv \mathbb{E}[p_T|\mathcal{O}], \tag{4}$$

which is a function of $\mathcal{O}$ that can be learned from minimizing the mean-squared error; see the Supplemental Material [135]. With a single factorized process, $\hat{p}_T(\mathcal{O})$ would be independent of $\mathcal{O}$, but the sum of two factorized processes yields nontrivial $\mathcal{O}$ dependence. To see this, recall from the Neyman-Pearson lemma [136] that the signal-to-background likelihood ratio is the optimal new physics classifier derivable from $\mathcal{O}$:

$$L_{S/B}(\mathcal{O}) = \frac{P_S(\mathcal{O})}{P_B(\mathcal{O})}. \tag{5}$$

(A stronger classifier might exist if one includes p_T information, but that requires *a priori* knowledge of $P_S(p_T)/P_B(p_T)$.) From Eq. (3), the conditional distribution can be written as

$$P(p_T|\mathcal{O}) = \frac{(1-f_S)P_B(p_T) + f_S L_{S/B}(\mathcal{O}) P_S(p_T)}{1 - f_S + f_S L_{S/B}(\mathcal{O})}. \tag{6}$$

Taking the expectation value with respect to p_T yields

$$\hat{p}_T(\mathcal{O}) = \langle p_T \rangle_B + f_S \frac{(\langle p_T \rangle_S - \langle p_T \rangle_B) L_{S/B}(\mathcal{O})}{1 - f_S + f_S L_{S/B}(\mathcal{O})}. \tag{7}$$

Remarkably, $\hat{p}_T(\mathcal{O})$ is monotonically related to $L_{S/B}(\mathcal{O})$, so it also defines optimal decision boundaries. A similar observation underpins anomaly detection methods based on classification without labels [4,12,108]. To our knowledge, the first proof that optimal classifiers can be defined through regression (as opposed to classification) appears in Ref. [114]. Note that the factorization assumption is crucial for learning a monotone of $L_{S/B}(\mathcal{O})$ without explicit knowledge of $P_S(\mathcal{O})$ or $P_B(\mathcal{O})$ individually.

Thus, assuming factorized observables, regressing the conditional expectation (FORCE) furnishes a powerful probe of new physics, justifying the FORCE acronym. Interestingly, the same logic holds with more than one type of new particle, such as $pp \to XY$, as long as $\langle p_T \rangle_X = \langle p_T \rangle_Y$ as expected from momentum conservation. If $\langle p_T \rangle_S > \langle p_T \rangle_B$, then Eq. (7) defines an optimal tagger; otherwise, it defines an optimal anti-tagger. In the absence of new physics ($f_S = 0$) or if the signal and background have the same average kinematics ($\langle p_T \rangle_S = \langle p_T \rangle_B$), then Eq. (7) simply returns the expectation value $\langle p_T \rangle$ with no observable dependence. Deviations of the model output from $\langle p_T \rangle$ are therefore a harbinger for a new type of factorized object in the data (or a violation of the factorization assumption).

In summary, FORCE proceeds as follows:

(1) *Define* approximately factorized objects (e.g. jets) with kinematics p_T and scale-/boost-invariant substructure $\mathcal{O}$.

(2) *Train* a machine-learning model $\hat{p}_T(\mathcal{O})$ to predict p_T from $\mathcal{O}$ with the mean-squared error loss.

(3) *Classify* anomalous objects via the model output.

Of course, real collider data is richer than the simple two-category case in Eq. (3). QCD jets themselves are admixtures of quark and gluon jets, each with slightly different kinematics and substructure. Multiple effects can violate the strict version of factorization in Eq. (2), such as partial containment of particle decay products in the jet cone or the logarithmic scale-dependence of QCD due to the running of the strong coupling constant. Further, certain known Standard Model processes, such as jets from hadronically decaying W/Z/Higgs bosons or top quarks, may be considered anomalous beyond the QCD dijet background by our formulation. This behavior may in fact be desirable, and "rediscovering" these particles may be an interesting way to benchmark this technique in data. More broadly, though, the general structure of factorization motivates FORCE as a new physics search strategy.

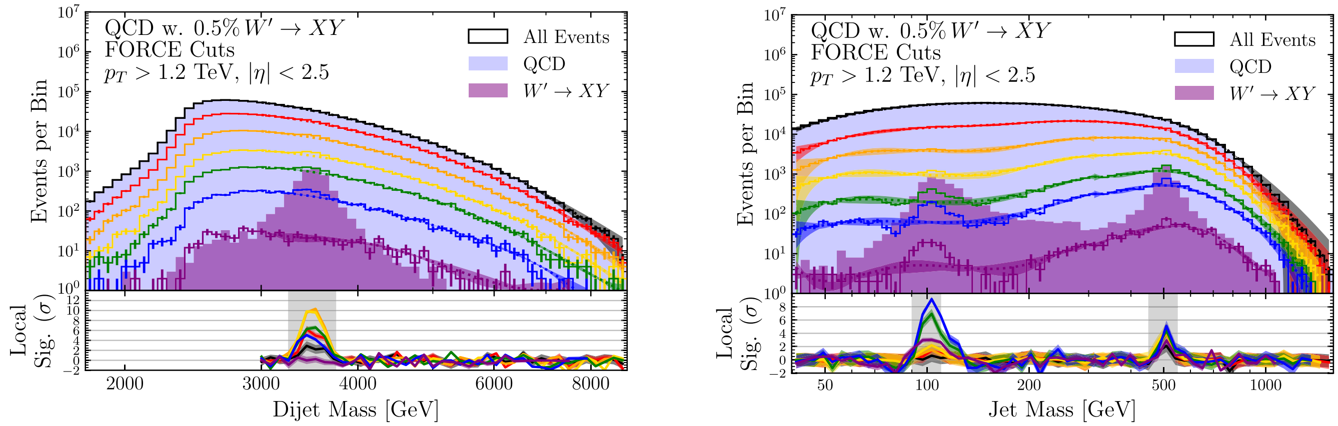
We now showcase FORCE for a new physics search involving dijets. Our case study is based on the development dataset [137] from the LHC Olympics 2020 Anomaly Detection Challenge [39]. This simulated dataset consists of 1 million QCD dijet events and up to 100 thousand $W' \to XY$ events, with the X and Y particles decaying to two quarks. The masses of the three new particles are $m_{W'} = 3.5$ TeV, $m_X = 500$ GeV, and $m_Y = 100$ GeV. The X and Y particles are boosted, giving rise to a dijet resonance with two-pronged jet substructure. While the signal has a W' mass peak, this feature is not used for FORCE training.

The LHC Olympics dataset is generated with PYTHIA 8.219 [138,139] and simulated with DELPHES 3.4.1 [140], excluding pileup or multiple parton interactions. Events are selected to have at least one $R = 1.0$ anti-k_T [141] jet with transverse momentum $p_T > 1.2$ TeV and pseudorapidity $|\eta| < 2.5$. Jets are clustered via the anti-k_T algorithm with a radius of $R = 1.0$ using FastJet 3.3.3 [141,142]. The leading two jets, i.e. those with highest transverse momenta, are recorded in each event as a proxy for the products of the high-energy scattering process. Both jets are used in the analysis, so the anomalies are defined over jets (instead of over events).

For our substructure observables $\mathcal{O}$, we use energy flow polynomials (EFPs) [143,144]. As reviewed in the Supplemental Material [135], EFPs arise from a systematic expansion in energies and angles, and they are sensitive to a broad range of jet features, including the two-prong substructure of the boosted X and Y particles. We compute all 13 EFPs up to and including degree 3 using EnergyFlow 1.0.3 [145], using $z_i = p_{T,i}$ as the energy variable and $\theta_{ij} = (p_i^\mu p_{j\mu}/p_{T,i}p_{T,j})^{1/2}$ as the angular variable. To satisfy Eq. (2), the EFPs need to be made scale- and boost-invariant. Quasi-scale-invariance can be achieved by normalizing the energies to sum to unity. As boosts transverse to the beamline approximately scale energies by γ and angles by $1/\gamma$, the EFPs can be made quasi-boost-invariant by rescaling them via:

$$\mathrm{EFP} \to \frac{\mathrm{EFP}}{(\sum_{i=1}^M p_{T,i})^{N-2d} (\sum_{i=1}^M \sum_{j=1}^M p_{T,i} p_{T,j} \theta_{ij})^d}, \tag{8}$$

where N and d are the energy and angular degrees of the polynomial. This rescaling reduces our basis to seven independent elements. We note that observables desired to be independent of p_T have been employed in prior work



(a) (b)

FIG. 2. The FORCE method applied to a dijet search with a 0.5% new physics signal fraction ($f_S = 0.005$), where the same cut on $\hat{p}_T(\mathcal{O})$ is imposed on both jets. Shown are (a) dijet and (b) jet mass distributions in the top panels and local significance values in the lower panels. The shaded regions "QCD" and "$W' \to XY$" refer to the truth distributions after baseline kinematic selections. The solid lines indicate the data, whereas the dashed lines and shaded areas indicate the background predictions and uncertainties derived from the Legendre fits. The black curve indicates no cuts on the trained model output $\hat{p}_T(\mathcal{O})$, while the sweep from red to purple corresponds to cuts that increase the lower bound on $\hat{p}_T(\mathcal{O})$, with specific cut values chosen manually to highlight the qualitative behavior. Tighter cuts, where the model predictions are further from plain QCD jets, clearly identify the new physics signal with a dijet mass peak at $m_{W'} = 3.5$ TeV and individual jet mass peaks at $m_X = 500$ GeV and $m_Y = 100$ GeV. See the Supplemental Material [135] for different values of f_S.

on anomaly detection [12] and jet-tagging [3]. In the Supplemental Material [135], we show how FORCE performance degrades without this normalization. Interestingly, existing observables for multiprong new physics searches, such as D_2 [132], emerge naturally as elements of this quasi-invariant basis.

The FORCE method works with any machine-learning algorithm whose output $\hat{p}_T(\mathcal{O})$ converges to the conditional expectation $\mathbb{E}[p_T|\mathcal{O}]$. We use a fully-connected neural network consisting of three dense layers with 50 nodes per layer, as well as L2 kernel and bias regularization of 10^{-5} in each layer. Between each dense layer is a dropout [146] layer with $p = 0.1$. Neural networks are implemented and trained with Keras [147] using TensorFlow [148], optimized with Adam [149] with a patience parameter of 10. Since our method is fully unsupervised, seeing no signal/background labels, we utilize the full dataset in training. (In a full analysis, it might be preferable to use statistically independent samples for training and testing.) Our code implementing FORCE is publicly available on GitHub [150].

The dijet and jet mass distributions are shown in Fig. 2 after applying FORCE for a signal fraction of $f_S = 0.005$. Here, we impose a cut on both jets that enforces their model output $\hat{p}_T(\mathcal{O})$ to be above the same threshold. With a strict enough cut, the signal clearly manifests as a peak at $m_{W'} = 3.5$ TeV in the dijet mass distribution, and peaks at $m_X = 500$ GeV and $m_Y = 100$ GeV in the individual jet mass distribution. To estimate the local significance, a background fit is performed using Legendre polynomials outside of the shaded signal region, using fifth order as the central value and between second and seventh orders for the uncertainty band. For the dijet mass background fit, we use data above 3 TeV, and for the jet mass background fits, below 300 GeV for m_Y and above 300 GeV for m_X. We find a boost in significance, where a pre-cut excess of 2σ for the W' and X are increased to $> 5\sigma$, while a precut excess of 1σ for the Y is increased to $> 5\sigma$. Note that although the new physics in this case study appears as a resonance in the jet and dijet mass distributions, a resonance is not a requirement of the FORCE method. (Without a bumplike feature, though, one would have to leverage some other method for background estimation.) Further, by imposing quasi-boost/scale invariance, the model output is largely decorrelated from jet mass (see Fig. 4 in the Supplemental Material [135] and related discussion in Refs. [151–153]).

To test the robustness of FORCE, we apply our method on a range of signal fractions f_S. For stability in this analysis, we train with an equally mixed dataset of 100,000 signal and 100,000 background events, using sample weights in Keras to mimic a signal fraction f_S. We then test the model on the full dataset. To account for variability and obtain error bars, we train an ensemble of 10 different models. The average per-jet classification performance is shown in Fig. 3, where we plot the background rejection factor as a function of signal efficiency. Here, the per-jet performance is evaluated only on the learned $\hat{p}_T(\mathcal{O})$, not including any additional features such as jet mass or assumptions about the $W' \to XY$ event topology. (For these reasons, one cannot directly compare our results to those of previous LHC Olympians.) In the large signal limit, FORCE

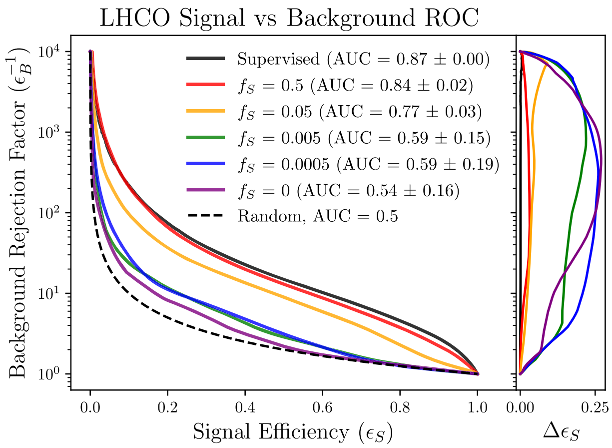


FIG. 3. Per-jet classification performance of FORCE in a dijet search with different signal fractions f_S. The background rejection factor is shown as a function of (left panel) the signal efficiency and (right panel) the standard deviation of the signal efficiency over 10 trainings. See the Supplemental Material [135] for a discussion of the $f_S = 0$ limit, where nontrivial performance can arise from a breakdown of factorization.

approaches the optimal supervised classifier, as predicted by Eq. (7). As the signal fraction decreases, the performance degrades and the variability increases, but there are still substantial gains in sensitivity. Note that the $f_S = 0$ limit still yields reasonable classification performance; this is possibly due to deviations from strict factorization, as discussed further in the Supplemental Material [135]. (Alternatively, since we are applying a no-signal model to a dataset with signal, proper convergence might not be achievable off the data manifold.)

Having established the desired behavior of FORCE on a benchmark collider search, it is worth remarking on several important points. First, our method is based on the inductive bias of factorization, so the performance we saw in the dijet analysis may not translate to other scenarios. This reflects a universal challenge for all approaches to anomaly detection, where the performance of the method depends on the applicability of the assumptions. Nevertheless, limits can be set on the parameters of specific new physics scenarios (even post hoc) by performing pseudoexperiments, injecting various amounts of signal, and repeating the procedure to establish confidence intervals. Second, as the signal fraction decreases, the performance of the learned model becomes highly sensitive to parameter initialization and statistical fluctuations. To ensure robust behavior in this regime, we recommend FORCE be paired with a regularization method like ensemble learning [154]. Third, detector effects can introduce factorization-violating effects, so it may be beneficial to apply FORCE after multidimensional unfolding is applied to the data [155–158]. Jet grooming techniques [159–161] might also improve the factorized behavior of jets at the theoretical level. Finally, we emphasize that no strategy can outperform a targeted search (i.e., hypothesis test) for a specific model, and that the power of data-driven approaches such as FORCE is in broadening the space of new physics scenarios that can be probed.

In summary, we introduced FORCE: an anomaly detection strategy for factorized new physics. By training a machine-learning model to predict the kinematics of factorized objects from their scale- and boost-invariant substructure, we obtain a powerful classifier directly from observed data. We showcased FORCE on a benchmark search for new physics in the dijet final state, where it successfully identified a new physics signal. This work contributes to a growing body of work where powerful computational tools from machine learning are combined with deep theoretical principles to unlock novel collider data analysis strategies. Furthermore, the FORCE method can be easily integrated into these prior methods when viewing the model output as an observable with high discrimination power.

The FORCE framework shifts the discussion of new physics searches from specific models to their general factorized structure, with machine-learning techniques performing detailed observable-level analyses. It would be interesting to generalize FORCE to handle more than one kinematic feature and more than two event categories, which would be important to handle multiple background components. It would also be interesting to combine our reasoning with the factorization of the full event energy flow [125], which may help reframe anomaly detection in the language of “theory space” [162,163]. Though we focused on jets and jet substructure here, this method applies more broadly to any factorized probability distributions, in collider physics and beyond. Data-driven searches hold the potential to fundamentally surprise us, not only by discovering new physics, but by uncovering it in forms that we have failed to imagine.

The authors are grateful to the organizers of the LHC Olympics 2020 Anomaly Detection Challenge for stimulating this research direction and producing excellent public datasets. We are grateful to Samuel Alipour-fard, Rikab Gambhir, Patrick Komiske, Benjamin Nachman, and Nilai Sarda for helpful comments and discussions. J. T. is supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions) and by the Simons Foundation through Investigator Grant 929241. This work was supported by the Office of Nuclear Physics of the U.S. Department of Energy (DOE) under Grant No. DE-SC0011090 and by the DOE Office of High Energy Physics under Grants No. DE-SC0012567 and No. DE-SC0019128.

[1] Georges Aad *et al.* (ATLAS Collaboration), Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B **716**, 1 (2012).
[2] Serguei Chatrchyan *et al.* (CMS Collaboration), Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B **716**, 30 (2012).
[3] J. A. Aguilar-Saavedra, Jack H. Collins, and Rashmish K. Mishra, A generic anti-QCD jet tagger, J. High Energy Phys. 11 (2017) 163.
[4] Jack H. Collins, Kiel Howe, and Benjamin Nachman, Anomaly detection for resonant new physics with machine learning, Phys. Rev. Lett. **121**, 241803 (2018).
[5] Raffaele Tito D'Agnolo and Andrea Wulzer, Learning new physics from a machine, Phys. Rev. D **99**, 015014 (2019).
[6] Andrea De Simone and Thomas Jacques, Guiding new physics searches with unsupervised learning, Eur. Phys. J. C **79**, 289 (2019).
[7] Jan Hajer, Ying-Ying Li, Tao Liu, and He Wang, Novelty detection meets collider physics, Phys. Rev. D **101**, 076015 (2020).
[8] Marco Farina, Yuichiro Nakai, and David Shih, Searching for new physics with deep autoencoders, Phys. Rev. D **101**, 075021 (2020).
[9] Theo Heimel, Gregor Kasieczka, Tilman Plehn, and Jennifer M. Thompson, QCD or what?, SciPost Phys. **6**, 030 (2019).
[10] Alessandro Casa and Giovanna Menardi, Nonparametric semisupervised classification for signal detection in high energy physics, arXiv:1809.02977.
[11] Olmo Cerri, Thong Q. Nguyen, Maurizio Pierini, Maria Spiropulu, and Jean-Roch Vlimant, Variational autoencoders for new physics mining at the large hadron collider, J. High Energy Phys. 05 (2019) 036.
[12] Jack H. Collins, Kiel Howe, and Benjamin Nachman, Extending the search for new resonances with machine learning, Phys. Rev. D **99**, 014038 (2019).
[13] Tuhin S. Roy and Aravind H. Vijay, A robust anomaly finder based on autoencoders, arXiv:1903.02032.
[14] Barry M. Dillon, Darius A. Faroughy, and Jernej F. Kamenik, Uncovering latent jet substructure, Phys. Rev. D **100**, 056002 (2019).
[15] Andrew Blance, Michael Spannowsky, and Philip Waite, Adversarially-trained autoencoders for robust unsupervised new physics searches, J. High Energy Phys. 10 (2019) 047.
[16] M. Romão Crispim, N. F. Castro, R. Pedro, and T. Vale, Transferability of deep learning models in searches for new physics at colliders, Phys. Rev. D **101**, 035042 (2020).
[17] Anna Mullin, Stuart Nicholls, Holly Pacey, Michael Parker, Martin White, and Sarah Williams, Does SUSY have friends? A new approach for LHC event analysis, J. High Energy Phys. 02 (2021) 160.
[18] Raffaele Tito D'Agnolo, Gaia Grosso, Maurizio Pierini, Andrea Wulzer, and Marco Zanetti, Learning multivariate new physics, Eur. Phys. J. C **81**, 89 (2021).
[19] Benjamin Nachman and David Shih, Anomaly detection with density estimation, Phys. Rev. D **101**, 075042 (2020).
[20] Anders Andreassen, Benjamin Nachman, and David Shih, Simulation assisted likelihood-free anomaly detection, Phys. Rev. D **101**, 095004 (2020).
[21] Oz Amram and Cristina Mantilla Suarez, Tag N' Train: A technique to train improved classifiers on unlabeled data, J. High Energy Phys. 01 (2021) 153.
[22] M. Crispim Romão, N. F. Castro, J. G. Milhano, R. Pedro, and T. Vale, Use of a generalized energy Mover's distance in the search for rare phenomena at colliders, Eur. Phys. J. C **81**, 192 (2021).
[23] Oliver Knapp, Olmo Cerri, Guenther Dissertori, Thong Q. Nguyen, Maurizio Pierini, and Jean-Roch Vlimant, Adversarially learned anomaly detection on CMS open data: Re-discovering the top quark, Eur. Phys. J. Plus **136**, 236 (2021).
[24] Georges Aad *et al.* (ATLAS Collaboration), Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector, Phys. Rev. Lett. **125**, 131801 (2020).
[25] B. M. Dillon, D. A. Faroughy, J. F. Kamenik, and M. Szewc, Learning the latent structure of collider events, J. High Energy Phys. 10 (2020) 206.
[26] M. Crispim Romão, N. F. Castro, and R. Pedro, Finding new physics without learning about it: Anomaly detection as a tool for searches at colliders, Eur. Phys. J. C **81**, 27 (2021); **81**, 1020(E) (2021).
[27] Taoli Cheng, Jean-François Arguin, Julien Leissner-Martin, Jacinthe Pilette, and Tobias Golling, Variational autoencoders for anomalous jet tagging, Phys. Rev. D **107**, 016002 (2023).
[28] Charanjit Kaur Khosa and Veronica Sanz, Anomaly awareness, SciPost Phys. **15**, 053 (2023).
[29] Punnathat Thaprasop, Kai Zhou, Jan Steinheimer, and Christoph Herold, Unsupervised outlier detection in heavy-ion collisions, Phys. Scr. **96**, 064003 (2021).
[30] J. A. Aguilar-Saavedra, F. R. Joaquim, and J. F. Seabra, Mass unspecific supervised tagging (MUST) for boosted jets, J. High Energy Phys. 03 (2021) 012; 04 (2021) 133(E).
[31] Stephon Alexander, Sergei Gleyzer, Hanna Parul, Pranath Reddy, Michael W. Toomey, Emanuele Usai, and Ryker Von Klar, Decoding dark matter substructure without supervision, arXiv:2008.12731.
[32] Kees Benkendorfer, Luc Le Pottier, and Benjamin Nachman, Simulation-assisted decorrelation for resonant anomaly detection, Phys. Rev. D **104**, 035003 (2021).
[33] Adrian Alan Pol, Victor Berger, Gianluca Cerminara, Cecile Germain, and Maurizio Pierini, Anomaly detection with conditional variational autoencoders, in *Eighteenth International Conference on Machine Learning and Applications* (2020), arXiv:2010.05531.
[34] Vinicius Mikuni and Florencia Canelli, Unsupervised clustering for collider physics, Phys. Rev. D **103**, 092007 (2021).
[35] Melissa van Beekveld, Sascha Caron, Luc Hendriks, Paul Jackson, Adam Leinweber, Sydney Otten, Riley Patrick, Roberto Ruiz De Austri, Marco Santoni, and Martin White, Combining outlier analysis algorithms to identify new physics at the LHC, J. High Energy Phys. 09 (2021) 024.
[36] Sang Eon Park, Dylan Rankin, Silviu-Marian Udrescu, Mikaeel Yunus, and Philip Harris, Quasi anomalous knowledge: Searching for new physics with embedded knowledge, J. High Energy Phys. 21 (2020) 030.

[37] Darius A. Faroughy, Uncovering hidden new physics patterns in collider events using Bayesian probabilistic models, Proc. Sci. ICHEP2020 (**2021**) 238.
[38] George Stein, Uros Seljak, and Biwei Dai, Unsupervised in-distribution anomaly detection of new physics through conditional density estimation, in *Proceedings of the 34th Conference on Neural Information Processing Systems* (2020), arXiv:2012.11638.
[39] Gregor Kasieczka *et al.*, The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics, Rep. Prog. Phys. **84**, 124201 (2021).
[40] Purvasha Chakravarti, Mikael Kuusela, Jing Lei, and Larry Wasserman, Model-independent detection of new physics signals using interpretable semi-supervised classifier tests, arXiv:2102.07679.
[41] Joshua Batson, C. Grace Haaf, Yonatan Kahn, and Daniel A. Roberts, Topological obstructions to autoencoding, J. High Energy Phys. 04 (2021) 280.
[42] Andrew Blance and Michael Spannowsky, Unsupervised event classification with graphs on classical and photonic quantum computers, J. High Energy Phys. 21 (2020) 170.
[43] Blaž Bortolato, Aleks Smolkovič, Barry M. Dillon, and Jernej F. Kamenik, Bump hunting in latent space, Phys. Rev. D **105**, 115009 (2022).
[44] Jack H. Collins, Pablo Martín-Ramiro, Benjamin Nachman, and David Shih, Comparing weak- and unsupervised methods for resonant anomaly detection, Eur. Phys. J. C **81**, 617 (2021).
[45] Barry M. Dillon, Tilman Plehn, Christof Sauer, and Peter Sorrenson, Better latent spaces for better autoencoders, SciPost Phys. **11**, 061 (2021).
[46] Thorben Finke, Michael Krämer, Alessandro Morandini, Alexander Mück, and Ivan Oleksiyuk, Autoencoders for unsupervised anomaly detection in high energy physics, J. High Energy Phys. 06 (2021) 161.
[47] David Shih, Matthew R. Buckley, Lina Necib, and John Tamanas, Via machinae: Searching for stellar streams using unsupervised machine learning, Mon. Not. R. Astron. Soc. **509**, 5992 (2021).
[48] Oliver Atkinson, Akanksha Bhardwaj, Christoph Englert, Vishal S. Ngairangbam, and Michael Spannowsky, Anomaly detection with convolutional graph neural networks, J. High Energy Phys. 08 (2021) 080.
[49] Alan Kahn, Julia Gonski, Inês Ochoa, Daniel Williams, and Gustaaf Brooijmans, Anomalous jet identification via sequence modeling, J. Instrum. **16**, P08012 (2021).
[50] Thea Aarrestad *et al.*, The dark machines anomaly score challenge: Benchmark data and model independent event classification for the large hadron collider, SciPost Phys. **12**, 043 (2022).
[51] Tommaso Dorigo, Martina Fumanelli, Chiara Maccani, Marija Mojsovska, Giles C. Strong, and Bruno Scarpa, RanBox: Anomaly detection in the copula space, J. High Energy Phys. 01 (2023) 008.
[52] Sascha Caron, Luc Hendriks, and Rob Verheyen, Rare and different: Anomaly scores from a combination of likelihood and out-of-distribution models to detect new physics at the LHC, SciPost Phys. **12**, 077 (2022).
[53] Ekaterina Govorkova, Ema Puljak, Thea Aarrestad, Maurizio Pierini, Kinga Anna Woźniak, and Jennifer Ngadiuba, LHC physics dataset for unsupervised new physics detection at 40 MHz, Sci. Data **9**, 118 (2022).
[54] Gregor Kasieczka, Benjamin Nachman, and David Shih, New methods and datasets for group anomaly detection from fundamental physics, in *Proceedings of the Conference on Knowledge Discovery and Data Mining* (2021), arXiv:2107.02821.
[55] Sergey Volkovich, Federico De Vito Halevy, and Shikma Bressler, A data-directed paradigm for BSM searches: The bump-hunting example, Eur. Phys. J. C **82**, 265 (2022).
[56] Ekaterina Govorkova *et al.*, 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 (2022).
[57] Anna Hallin, Joshua Isaacson, Gregor Kasieczka, Claudius Krause, Benjamin Nachman, Tobias Quadfasel, Matthias Schlaffer, David Shih, and Manuel Sommerhalder, Classifying anomalies through outer density estimation, Phys. Rev. D **106**, 055006 (2022).
[58] Bryan Ostdiek, Deep set auto encoders for anomaly detection in particle physics, SciPost Phys. **12**, 045 (2022).
[59] Katherine Fraser, Samuel Homiller, Rashmish K. Mishra, Bryan Ostdiek, and Matthew D. Schwartz, Challenges for unsupervised anomaly detection in particle physics, J. High Energy Phys. 03 (2022) 066.
[60] Pratik Jawahar, Thea Aarrestad, Nadezda Chernyavskaya, Maurizio Pierini, Kinga A. Wozniak, Jennifer Ngadiuba, Javier Duarte, and Steven Tsan, Improving variational autoencoders for new physics detection at the LHC with normalizing flows, Front. Big Data **5**, 803685 (2022).
[61] Juan Herrero-Garcia, Riley Patrick, and Andre Scaffidi, A semi-supervised approach to dark matter searches in direct detection data with machine learning, J. Cosmol. Astropart. Phys. 02 (2022) 039.
[62] Christopher G. Lester and Rupert Tombs, Using unsupervised learning to detect broken symmetries, with relevance to searches for parity violation in nature. (Previously: "Stressed GANs snag desserts"), arXiv:2111.00616.
[63] J. A. Aguilar-Saavedra, Anomaly detection from mass unspecific jet tagging, Eur. Phys. J. C **82**, 130 (2022).
[64] Rupert Tombs and Christopher G. Lester, A method to challenge symmetries in data with self-supervised learning, J. Instrum. **17**, P08024 (2022).
[65] Vinicius Mikuni, Benjamin Nachman, and David Shih, Online-compatible unsupervised nonresonant anomaly detection, Phys. Rev. D **105**, 055006 (2022).
[66] Sergei Chekanov and Walter Hopkins, Event-based anomaly detection for searches for new physics, Universe **8**, 494 (2022).
[67] Raffaele Tito d'Agnolo, Gaia Grosso, Maurizio Pierini, Andrea Wulzer, and Marco Zanetti, Learning new physics from an imperfect machine, Eur. Phys. J. C **82**, 275 (2022).
[68] Florencia Canelli, Annapaola de Cosa, Luc Le Pottier, Jeremi Niedziela, Kevin Pedro, and Maurizio Pierini, Autoencoders for semivisible jet detection, J. High Energy Phys. 02 (2022) 074.
[69] Vishal S. Ngairangbam, Michael Spannowsky, and Michihisa Takeuchi, Anomaly detection in high-energy physics using a quantum autoencoder, Phys. Rev. D **105**, 095004 (2022).

[70] J. A. Aguilar-Saavedra, Taming modeling uncertainties with mass unspecific supervised tagging, Eur. Phys. J. C **82**, 270 (2022).
[71] Thorsten Buss, Barry M. Dillon, Thorben Finke, Michael Krämer, Alessandro Morandini, Alexander Mück, Ivan Oleksiyuk, and Tilman Plehn, What's anomalous in LHC jets?, SciPost Phys. **15**, 168 (2023).
[72] Layne Bradshaw, Spencer Chang, and Bryan Ostdiek, Creating simple, interpretable anomaly detectors for new physics in jet substructure, Phys. Rev. D **106**, 035014 (2022).
[73] Mattias Birman, Benjamin Nachman, Raphael Sebbah, Gal Sela, Ophir Turetz, and Shikma Bressler, Data-directed search for new physics based on symmetries of the SM, Eur. Phys. J. C **82**, 508 (2022).
[74] John Andrew Raine, Samuel Klein, Debajyoti Sengupta, and Tobias Golling, CURTAINs for your sliding window: Constructing unobserved regions by transforming adjacent intervals, Front. Big Data **6**, 899345 (2023).
[75] Marco Letizia, Gianvito Losapio, Marco Rando, Gaia Grosso, Andrea Wulzer, Maurizio Pierini, Marco Zanetti, and Lorenzo Rosasco, Learning new physics efficiently with nonparametric methods, Eur. Phys. J. C **82**, 879 (2022).
[76] C. Fanelli, J. Giroux, and Z. Papandreou, "Flux+Mutability": A conditional generative approach to one-class classification and anomaly detection, Mach. Learn. Sci. Technol. **3**, 045012 (2022).
[77] Thorben Finke, Michael Krämer, Maximilian Lipp, and Alexander Mück, Boosting mono-jet searches with model-agnostic machine learning, J. High Energy Phys. 08 (2022) 015.
[78] Rob Verheyen, Event generation and density estimation with surjective normalizing flows, SciPost Phys. **13**, 047 (2022).
[79] Barry M. Dillon, Radha Mastandrea, and Benjamin Nachman, Self-supervised anomaly detection for new physics, Phys. Rev. D **106**, 056005 (2022).
[80] Sulaiman Alvi, Christian W. Bauer, and Benjamin Nachman, Quantum anomaly detection for collider physics, J. High Energy Phys. 02 (2023) 220.
[81] Barry M. Dillon, Luigi Favaro, Tilman Plehn, Peter Sorrenson, and Michael Krämer, A normalized autoencoder for LHC triggers, SciPost Phys. Core **6**, 074 (2023).
[82] Sascha Caron, Roberto Ruiz de Austri, and Zhongyi Zhang, Mixture-of-theories training: Can we find new physics and anomalies better by mixing physical theories?, J. High Energy Phys. 03 (2023) 004.
[83] Sang Eon Park, Philip Harris, and Bryan Ostdiek, Neural embedding: Learning the embedding of the manifold of physics data, J. High Energy Phys. 07 (2023) 108.
[84] Gregor Kasieczka, Radha Mastandrea, Vinicius Mikuni, Benjamin Nachman, Mariel Pettee, and David Shih, Anomaly detection under coordinate transformations, Phys. Rev. D **107**, 015009 (2023).
[85] Jernej F. Kamenik and Manuel Szewc, Null hypothesis test for anomaly detection, Phys. Lett. B **840**, 137836 (2023).
[86] Anna Hallin, Gregor Kasieczka, Tobias Quadfasel, David Shih, and Manuel Sommerhalder, Resonant anomaly detection without background sculpting, Phys. Rev. D **107**, 114012 (2023).
[87] Jack Y. Araz and Michael Spannowsky, Quantum-probabilistic Hamiltonian learning for generative modelling and anomaly detection, Phys. Rev. A **108**, 062422 (2023).
[88] Radha Mastandrea and Benjamin Nachman, Efficiently moving instead of reweighting collider events with machine learning, in *Proceedings of the 36th Conference on Neural Information Processing Systems: Workshop on Machine Learning and the Physical Sciences* (2022), arXiv:2212.06155.
[89] Julian Schuhmacher, Laura Boggia, Vasilis Belis, Ema Puljak, Michele Grossi, Maurizio Pierini, Sofia Vallecorsa, Francesco Tacchino, Panagiotis Barkoutsos, and Ivano Tavernelli, Unravelling physics beyond the standard model with classical and quantum anomaly detection, Mach. Learn. Sci. Technol. **4**, 045031 (2023).
[90] Tobias Golling *et al.*, The mass-ive issue: Anomaly detection in jet physics, in *Proceedings of the 34th Conference on Neural Information Processing Systems* (2023), arXiv:2303.14134.
[91] Stephen Roche, Quincy Bayer, Benjamin Carlson, William Ouligian, Pavel Serhiayenka, Joerg Stelzer, and Tae Min Hong, Nanosecond anomaly detection with decision trees for high energy physics and real-time application to exotic Higgs decays, Nat. Commun. **15**, 3527 (2024).
[92] Debajyoti Sengupta, Samuel Klein, John Andrew Raine, and Tobias Golling, CURTAINs flows for flows: Constructing unobserved regions with maximum likelihood estimation, arXiv:2305.04646.
[93] J. A. Aguilar-Saavedra, E. Arganda, F. R. Joaquim, R. M. Sandá Seoane, J. F. Seabra, Gradient Boosting MUST taggers for highly-boosted jets, arXiv:2305.04957.
[94] Louis Vaslin, Vincent Barra, and Julien Donini, GAN-AE: An anomaly detection algorithm for new physics search in LHC data, Eur. Phys. J. C **83**, 1008 (2023).
[95] Georges Aad *et al.* (ATLAS Collaboration), Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states using $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector, Phys. Rev. D **108**, 052009 (2023).
[96] Vinicius Mikuni and Benjamin Nachman, High-dimensional and permutation invariant anomaly detection, SciPost Phys. **16**, 062 (2024).
[97] Tobias Golling, Gregor Kasieczka, Claudius Krause, Radha Mastandrea, Benjamin Nachman, John Andrew Raine, Debajyoti Sengupta, David Shih, and Manuel Sommerhalder, The interplay of machine learning–based resonant anomaly detection methods, Eur. Phys. J. C **84**, 241 (2024).
[98] Sergei V. Chekanov and Rui Zhang, Enhancing the hunt for new phenomena in dijet final-states using anomaly detection filters at the High-Luminosity Large Hadron Collider, Eur. Phys. J. Plus **139**, 237 (2024).
[99] D. Abadjiev *et al.* (CMS ECAL Collaboration), Autoencoder-based anomaly detection system for online data quality monitoring of the CMS electromagnetic calorimeter, Comput. Software Big Sci. **8**, 11 (2024).
[100] Gerrit Bickendorf, Manuel Drees, Gregor Kasieczka, Claudius Krause, and David Shih, Combining resonant and tail-based anomaly detection, Phys. Rev. D **109**, 096031 (2024).

[101] Thorben Finke, Marie Hein, Gregor Kasieczka, Michael Krämer, Alexander Mück, Parada Prangchaikul, Tobias Quadfasel, David Shih, and Manuel Sommerhalder, Back to the roots: Tree-based algorithms for weakly supervised anomaly detection, Phys. Rev. D **109**, 034033 (2024).

[102] Erik Buhmann, Cedric Ewen, Gregor Kasieczka, Vinicius Mikuni, Benjamin Nachman, and David Shih, Full phase space resonant anomaly detection, Phys. Rev. D **109**, 055015 (2024).

[103] Marat Freytsis, Maxim Perelstein, and Yik Chuen San, Anomaly detection in presence of irrelevant features, J. High Energy Phys. 02 (2024) 220.

[104] Timothy Cohen, Martin Jankowiak, Mariangela Lisanti, Hou Keong Lou, and Jay G. Wacker, Jet substructure templates: Data-driven QCD backgrounds for fat jet searches, J. High Energy Phys. 05 (2014) 005.

[105] Joshua Lin, Wahid Bhimji, and Benjamin Nachman, Machine learning templates for QCD factorization in the search for physics beyond the standard model, J. High Energy Phys. 05 (2019) 181.

[106] Lucio Mwinmaarong Dery, Benjamin Nachman, Francesco Rubbo, and Ariel Schwartzman, Weakly supervised classification in high energy physics, J. High Energy Phys. 05 (2017) 145.

[107] Timothy Cohen, Marat Freytsis, and Bryan Ostdiek, (Machine) learning to do more with less, J. High Energy Phys. 02 (2018) 034.

[108] Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler, Classification without labels: Learning from mixed samples in high energy physics, J. High Energy Phys. 10 (2017) 174.

[109] Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Matthew D. Schwartz, Learning to classify from impure samples with high-dimensional data, Phys. Rev. D **98**, 011502 (2018).

[110] Albert M. Sirunyan *et al.* (CMS Collaboration), 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 (2020).

[111] Eric M. Metodiev and Jesse Thaler, Jet topics: Disentangling quarks and gluons at colliders, Phys. Rev. Lett. **120**, 241602 (2018).

[112] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler, An operational definition of quark and gluon jets, J. High Energy Phys. 11 (2018) 059.

[113] Georges Aad *et al.* (ATLAS Collaboration), Properties of jet fragmentation using charged particles measured with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV, Phys. Rev. D **100**, 052011 (2019).

[114] Patrick T. Komiske, Serhii Kryhin, and Jesse Thaler, Disentangling quarks and gluons in CMS open data, Phys. Rev. D **106**, 094021 (2022).

[115] Gavin P. Salam, Towards jetography, Eur. Phys. J. C **67**, 637 (2010).

[116] Andrew J. Larkoski, Ian Moult, and Benjamin Nachman, Jet substructure at the Large Hadron Collider: A review of recent advances in theory and machine learning, Phys. Rep. **841**, 1 (2020).

[117] Roman Kogler *et al.*, Jet substructure at the Large Hadron Collider: Experimental review, Rev. Mod. Phys. **91**, 045003 (2019).

[118] Simone Marzani, Gregory Soyez, and Michael Spannowsky, *Looking Inside Jets: An Introduction to Jet Substructure and Boosted-Object Phenomenology* (Springer, New York, 2019), Vol. 958.

[119] Christian W. Bauer, Sean Fleming, Dan Pirjol, and Iain W. Stewart, An effective field theory for collinear and soft gluons: Heavy to light decays, Phys. Rev. D **63**, 114020 (2001).

[120] Christian W. Bauer and Iain W. Stewart, Invariant operators in collinear effective theory, Phys. Lett. B **516**, 134 (2001).

[121] Christian W. Bauer, Dan Pirjol, and Iain W. Stewart, Soft collinear factorization in effective field theory, Phys. Rev. D **65**, 054022 (2002).

[122] Christian W. Bauer, Sean Fleming, Dan Pirjol, Ira Z. Rothstein, and Iain W. Stewart, Hard scattering factorization from effective field theory, Phys. Rev. D **66**, 014017 (2002).

[123] M. Beneke, A. P. Chapovsky, M. Diehl, and T. Feldmann, Soft collinear effective theory and heavy to light currents beyond leading power, Nucl. Phys. **B643**, 431 (2002).

[124] John C. Collins, Davison E. Soper, and George F. Sterman, Factorization of hard processes in QCD, Adv. Ser. Dir. High Energy Phys. **5**, 1 (1989).

[125] Christian W. Bauer, Andrew Hornig, and Frank J. Tackmann, Factorization for generic jet production, Phys. Rev. D **79**, 114013 (2009).

[126] Ilya Feige and Matthew D. Schwartz, Hard-soft-collinear factorization to all orders, Phys. Rev. D **90**, 105020 (2014).

[127] George Sterman, Comments on collinear factorization, in *Snowmass 2021* (2022), arXiv:2207.06507.

[128] D. Berdine, N. Kauer, and D. Rainwater, Breakdown of the narrow width approximation for new physics, Phys. Rev. Lett. **99**, 111601 (2007).

[129] C. F. Uhlemann and N. Kauer, Narrow-width approximation accuracy, Nucl. Phys. **B814**, 195 (2009).

[130] Jesse Thaler and Ken Van Tilburg, Identifying boosted objects with N-subjettiness, J. High Energy Phys. 03 (2011) 015.

[131] Jesse Thaler and Ken Van Tilburg, Maximizing boosted top identification by minimizing N-subjettiness, J. High Energy Phys. 02 (2012) 093.

[132] Andrew J. Larkoski, Ian Moult, and Duff Neill, Power counting to better jet observables, J. High Energy Phys. 12 (2014) 009.

[133] Andrew J. Larkoski, Ian Moult, and Duff Neill, Building a better boosted top tagger, Phys. Rev. D **91**, 034035 (2015).

[134] Ian Moult, Lina Necib, and Jesse Thaler, New angles on energy correlation functions, J. High Energy Phys. 12 (2016) 153.

[135] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevD.110.055012 for (i) a proof on the conditional expectation minimizing the mean-squared error loss, (ii) an application of the FORCE method to a toy Gaussian dataset, (iii) a review of energy flow polynomials (EFPs), (iv) an analysis on normalization schemes for EFPs, and (v) an analysis on an idealized shuffled dataset.

[136] Jerzy Neyman and Egon S. Pearson, On the problem of the most efficient tests of statistical hypotheses, in *Breakthroughs in Statistics* (Springer, New York, 1992), pp. 73–108.
[137] Gregor Kasieczka, Ben Nachman, and David Shih, R&D dataset for LHC Olympics 2020 anomaly detection challenge (2019), https://zenodo.org/records/2629073.
[138] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands, A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. **178**, 852 (2008).
[139] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands, An introduction to PYTHIA 8.2, Comput. Phys. Commun. **191**, 159 (2015).
[140] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), DELPHES 3, A modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057.
[141] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, The anti-k_t jet clustering algorithm, J. High Energy Phys. 04 (2008) 063.
[142] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, FastJet user manual, Eur. Phys. J. C **72**, 1896 (2012).
[143] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler, Energy flow polynomials: A complete linear basis for jet substructure, J. High Energy Phys. 04 (2018) 013.
[144] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler, Cutting multiparticle correlators down to size, Phys. Rev. D **101**, 036019 (2020).
[145] Patrick T. Komiske and Eric M. Metodiev, EnergyFlow, https://energyflow.network/ (2019).
[146] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. **15**, 1929 (2014).
[147] Francois Chollet, Keras, https://github.com/fchollet/keras (2017).
[148] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard *et al.*, TensorFlow: A system for large-scale machine learning, in *OSDI* (USENIX Association, Berkeley, 2016), Vol. 16, pp. 265–283.
[149] Diederik P. Kingma and Jimmy Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.
[150] Raymond Wynne, FORCE, https://github.com/r-wynne/FORCE (2023).
[151] James Dolen, Philip Harris, Simone Marzani, Salvatore Rappoccio, and Nhan Tran, Thinking outside the ROCs: Designing decorrelated taggers (DDT) for jet substructure, J. High Energy Phys. 05 (2016) 156.
[152] Chase Shimmin, Peter Sadowski, Pierre Baldi, Edison Weik, Daniel Whiteson, Edward Goul, and Andreas Søgaard, Decorrelated jet substructure tagging using adversarial neural networks, Phys. Rev. D **96**, 074034 (2017).
[153] Ian Moult, Benjamin Nachman, and Duff Neill, Convolved substructure: Analytically decorrelating jet substructure observables, J. High Energy Phys. 05 (2018) 002.
[154] Mudasir Ahmad Ganaie, Minghui Hu, Mohammad Tanveer, and Ponnuthurai N. Suganthan, Ensemble deep learning: A review, arXiv:2104.02395.
[155] Anders Andreassen, Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler, OmniFold: A method to simultaneously unfold all observables, Phys. Rev. Lett. **124**, 182001 (2020).
[156] Marco Bellagente, Anja Butter, Gregor Kasieczka, Tilman Plehn, Armand Rousselot, Ramon Winterhalder, Lynton Ardizzone, and Ullrich Köthe, Invertible networks or partons to detector and back again, SciPost Phys. **9**, 074 (2020).
[157] Anders Andreassen, Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, Adi Suresh, and Jesse Thaler, Scaffolding simulations with deep learning for high-dimensional deconvolution, in *Proceedings of the 9th International Conference on Learning Representations* (2021), arXiv:2105.04448.
[158] Miguel Arratia *et al.*, Publishing unbinned differential cross section results, J. Instrum. **17**, P01024 (2022).
[159] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam, Towards an understanding of jet substructure, J. High Energy Phys. 09 (2013) 029.
[160] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler, Soft drop, J. High Energy Phys. 05 (2014) 146.
[161] Christopher Frye, Andrew J. Larkoski, Matthew D. Schwartz, and Kai Yan, Factorization for groomed jet substructure beyond the next-to-leading logarithm, J. High Energy Phys. 07 (2016) 064.
[162] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler, Metric space of collider events, Phys. Rev. Lett. **123**, 041801 (2019).
[163] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler, The hidden geometry of particle collisions, J. High Energy Phys. 07 (2020) 006.