Simulation-Assisted Decorrelation for Resonant Anomaly Detection

Kees Bekendorfer, Physics Department, Reed College
Portland, OR 97202 USA
kebenkend@reed.edu

Luc Le Pottier, Physics Department, University of Michigan
Ann Arbor, MI 48109 USA
luclepot@umich.edu

Benjamin Nachman, Physics Division, Lawrence Berkeley National Laboratory
Berkeley, CA 94720 USA
bpnachman@lbl.gov

Abstract

A growing number of weak- and unsupervised machine learning approaches to anomaly detection are being proposed to significantly extend the search program at the Large Hadron Collider and elsewhere. One of the prototypical examples for these methods is the search for resonant new physics, where a bump hunt can be performed in an invariant mass spectrum. A significant challenge to methods that rely entirely on data is that they are susceptible to sculpting artificial bumps from the dependence of the machine learning classifier on the invariant mass. We explore two solutions to this challenge by minimally incorporating simulation into the learning. In particular, we study the robustness of Simulation Assisted Likelihood-free Anomaly Detection (SALAD), one of our recently developed anomaly detection methods, to correlations between the classifier and the invariant mass. Next, we propose a new approach that only uses the simulation for decorrelation but the Classification without Labels (CWoLa) approach for achieving signal sensitivity.

1 Introduction

Despite compelling experimental (e.g. dark matter) and theoretical (e.g. the hierarchy problem) evidence for new phenomena at the electroweak scale, experiments at the Large Hadron Collider (LHC) have not yet discovered any physics beyond the Standard Model (BSM). There are major search efforts across LHC experiments [9, 12, 10, 20, 21, 19, 42], where most analyses target a particular class of BSM models. While this work is well-motivated and continuing to improve in sensitivity (in part due to machine learning [41, 34, 2, 48]), there is also a growing need for new search strategies capable of discovery in unexpected scenarios.

A variety of automated anomaly detection techniques using innovative machine learning methods are being proposed to cover the unexpected [25, 22, 23, 8, 26, 32, 36, 52, 16, 14, 35, 27, 46, 4, 28, 7, 47, 3, 51, 50, 40, 29, 49, 5, 17, 38, 55]. An important subset of these proposals targets resonant new physics, where sideband methods can be used to estimate the SM background directly from data. A key challenge facing such methods is that the machine learning classifiers must be relatively independent from the resonant feature, for otherwise artificial bumps can be formed. Many automated decorrelation methods have been proposed to ensure that classifiers are decorrelated from particular

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

features by construction [43, 30, 45, 54, 53, 15, 11, 57, 31, 56, 37, 33], but they may not apply in all cases. In particular, weakly supervised approaches that learn directly on the signal region cannot be simply combined with a decorrelation scheme because such an approach could degrade the performance in the presence of a signal. A localized signal would manifest as a dependence between the resonant feature and other features for classification, so forcing independence could eliminate signal sensitivity.

In this paper, two weakly supervised approaches are studied: Classification without Labels (CWoLa) [44, 22, 23, 8] and Simulation Assisted Likelihood-free Anomaly Detection (SALAD) [7]. CWoLa is a method that does not depend on simulation and achieves signal sensitivity by comparing a signal region with nearby sideband regions in the resonance feature. As a result, CWoLa is particularly sensitive to dependencies between the classification features and the resonant feature. SALAD uses a reweighted simulation to achieve signal sensitivity. Since it never directly uses the sideband region, SALAD is expected to be more robust than CWoLa to dependencies. In order to recover the performance of CWoLa in the presence of significant dependence between the classification features and the resonant feature, a new method called simulation augmented CWoLa (SA-CWoLa) is introduced. The SA-CWoLa approach augments the CWoLa loss function to penalize the classifier for learning differences between the signal region and the sideband region in simulation, which is signal-free by construction. All of these methods will be investigated using the correlation test proposed in Ref. [47].

2 Methods

For a set of features $(m, x) \in \mathbb{R}^{n+1}$, let $f : \mathbb{R}^n \to [0, 1]$ be parameterized by a neural network. The observable m is special, for it is the resonance feature that should be relatively independent from $f(x)$. The signal region (SR) is defined by an interval in m and the sidebands (SB) are neighboring intervals.

All neural networks were implemented in KERAS [18] with the TENSORFLOW backend [1] and optimized with ADAM [39]. Each network is composed of three hidden layers with 64 nodes each and use the rectified linear unit (ReLU) activation function. The sigmoid function is used after the last layer. Training proceeds for 10 epochs with a batch size of 200. None of these parameters were optimized; it is likely that improved performance could be achieved with an in-situ optimization based on a validation set.

2.1 Simulation Assisted Likelihood-free Anomaly Detection (SALAD)

The SALAD network [7] is optimized using the following loss:

$$\mathcal{L}_{\text{SALAD}}[f] = -\sum_{i\in\text{SR,data}} \log(f(x_i)) - \sum_{i\in\text{SR,sim.}} w(x_i, m) \log(1 - f(x_i)) \tag{1}$$

where $w(x_i, m) = g(x_i, m)/(1 - g(x_i, m))$ are a set of weights using the Classification for Tuning and Reweighting (DCTR) [6] method. The function g is a parameterized classifier [24, 13] trained to distinguish data and simulation in the sideband:

$$\mathcal{L}[g] = -\sum_{i\in\text{SB,data}} \log(g(x_i, m)) - \sum_{i\in\text{SB,sim.}} \log(1 - g(x_i, m))\,. \tag{2}$$

The above neural networks are optimized with binary cross entropy, but one could use other functions as well, such as the mean-squared error. Intuitively, the idea of SALAD is to train a classifier to distinguish data and simulation in the SR. However, there may be significant differences between the background in data and the background simulation, so a reweighting function is learned in the sidebands that makes the simulation look more like the background in data.

2.2 Simulation Augmented Classification without Labels (CWoLa)

The idea of CWoLa [44] is to construct two mixed samples of data that are each composed of two classes. Using CWoLa for resonant anomaly detection [22, 23], one can construct the mixed samples using the SR and SB. In the absence of signal, the SR and SB should be statistically identical and therefore the CWoLa classifier does not learn anything useful. However, if there is a signal, then it can detect the presence of a difference between the SR and SB. In practice, there are small differences between the SR and SB because there are dependencies between m and x and so CWoLa will only be able to find signals that introduce a bigger difference than already present in the background. The CWoLa anomaly detection strategy was recently used in a low-dimensional application by the ATLAS experiment [8].

We propose a modification of the usual CWoLa loss function in order to construct a simulation-augmented (SA) CWoLa classifier:

$$\begin{aligned}\mathcal{L}_{\text{SA-CWola}}[f] = &-\sum_{i\in\text{SR,data}}\log(f(x_i)) - \sum_{i\in\text{SB,data}}\log(1-f(x_i)) \\ &+\lambda\left(\sum_{i\in\text{SR,sim.}}\log(f(x_i)) + \sum_{i\in\text{SB,sim.}}\log(1-f(x_i))\right)\,, \end{aligned} \tag{3}$$

where $\lambda > 0$ is a hyper-parameter. The limit $\lambda \to 0$ is the usual CWoLa approach and for $\lambda > 0$, the classifier is penalized if it can distinguish the SR from the SB in the (background-only) simulation. In order to help the learning process, the upper and lower sidebands are given the same total weight as each other and together, the same weight as the SR.

3 Results

As a benchmark, 1500 signal events corresponding to a fitted significance of about 2σ is injected into the data for training. For evaluation, the entire signal sample (except for the small number of injected events) is used. Figure 1 shows the performance of various configurations. The fully supervised classifier uses high statistics signal and background samples in the SR with full label information. Since the data are not labeled, this is not achievable in practice. A solid red line labeled 'Optimal CWoLa' corresponds to a classifier trained using two mixed samples, one composed of pure background in the single region and the other composed of mostly background (independent from the first sample) in the SR with the 1500 signal events. This is optimal in the sense that it removes the effect from phase space differences between the SR and SB for the background. The Optimal CWoLa line is far below the fully supervised classifier because the neural network needs to identify a small difference between the mixed samples over the natural statistical fluctuations in both sets. The actual CWoLa method is shown with a dotted red line. By construction, there is a significant difference between the phase space of the SR and SB and so the classifier is unable to identify the signal. At low efficiency, the CWoLa classifier actually anti-tags because the SR-SB differences are such that the signal is more SB-like then SR-like. Despite this drop in performance, the simulation augmenting modification (solid orange) with $\lambda = 0.5$ nearly recovers the full performance of CWoLa.

For comparison, a classifier trained using simulation directly is also presented in Figure 1. The line labeled 'Data vs. Sim.' directly trains a classifier to distinguish the data and simulation in the SR without reweighting. Due to the differences between the background in data and the simulated background, this classifier is not effective. In fact, the signal is more like the background simulation than the data background and so the classifier is worse than random (preferentially removes signal). The performance is significantly improved by adding in the parameterized reweighting, as advocated by Ref. [7]. With this reweighting, the Salad classifier is significantly better than random and is comparable to SA-CWoLa. The Optimal CWoLa line also serves as the upper bound in performance for Salad because it corresponds to the case where the background simulation is statistical identical to the background in data.

4 Conclusions

This paper has investigated the impact of dependencies between m_{jj} and classification features for the resonant anomaly detection methods Salad and CWoLa. A new simulation-augmented approach has been proposed to remedy challenges with the CWoLa method. This modification is

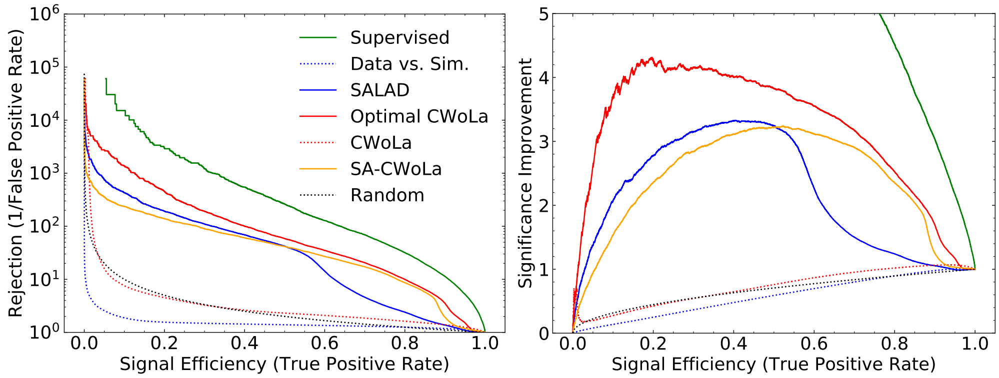

Figure 1: A Receiver Operating Characteristic (ROC) curve (left) and significance improvement curve (right) for various anomaly detection methods described in the text. The significance improvement is defined as the ratio of the signal efficiency to the square root of the background efficiency. A significance improvement of 2 means that the initial significance would be amplified by about a factor of two after employing the anomaly detection strategy. The supervised line is unachievable unless there is no mismodeling and one designed a search for the specific W' signal used in this paper. The curve labeled 'Random' corresponds to equal efficiency for signal and background.

shown to completely recover the performance of CWoLa from the ideal case where dependences are ignored in the training. In both the SALAD and SA-CWoLa methods, background-only simulations provide a critical tool for mitigating the sensitivity of the classifiers on dependences between the resonant feature and the classifier features.

Each of the methods examined here have advantages and weaknesses, and it is likely that multiple approaches will be required to achieve broad sensitivity to BSM physics. Therefore, it is critical to study the sensitivity of each technique to dependencies and propose modifications where possible to build robustness. This paper is an important step in the decorrelation program for automated anomaly detection with machine learning. Tools like the ones proposed here may empower higher-dimensional versions of the existing ATLAS search [8] as well as other related searches by other experiments in the near future.

Broader Impact

The algorithms presented and discussed in this paper could be applied anywhere where a blinded search for outliers is desired. In particular, they might be find applications as a candidate filter for a job or university application; in this case, the algorithm might be able to asses objective quality while explicitly blinded to non-candidate factors. Additionally, this method of semi-supervised detection would be well suited to other physical sciences in which a generalized blind search for outliers is desired.

References

[1] Martín Abadi et al. "TensorFlow: A System for Large-Scale Machine Learning." In: *OSDI*. Vol. 16. 2016, pp. 265–283.

[2] Murat Abdughani et al. "Supervised deep learning in high energy phenomenology: a mini review". In: *Commun. Theor. Phys.* 71 (2019), p. 955. DOI: 10.1088/0253-6102/71/8/955. arXiv: 1905.06047 [hep-ph].

[3] J. A. Aguilar-Saavedra, Jack H. Collins, and Rashmish K. Mishra. "A generic anti-QCD jet tagger". In: *JHEP* 11 (2017), p. 163. DOI: 10.1007/JHEP11(2017)163. arXiv: 1709.01087 [hep-ph].

[4] Giovanna Menardi Alessandro Casa. "Nonparametric semisupervised classification for signal detection in high energy physics". In: (2019). arXiv: 1809.02977 [hep-ex].

[5] Oz Amram and Cristina Mantilla Suarez. "Tag N' Train: A Technique to Train Improved Classifiers on Unlabeled Data". In: (Feb. 2020). arXiv: 2002.12376 [hep-ph].

[6] Anders Andreassen and Benjamin Nachman. “Neural Networks for Full Phase-space Reweighting and Parameter Tuning”. In: *Phys. Rev. D* 101.9 (2020), p. 091901. DOI: 10.1103/PhysRevD.101.091901. arXiv: 1907.08209 [hep-ph].

[7] Anders Andreassen, Benjamin Nachman, and David Shih. “Simulation Assisted Likelihood-free Anomaly Detection”. In: *Phys. Rev. D* 101.9 (2020), p. 095004. DOI: 10.1103/PhysRevD.101.095004. arXiv: 2001.05001 [hep-ph].

[8] ATLAS Collaboration. “Dijet resonance search with weak supervision using 13 TeV pp collisions in the ATLAS detector”. In: (2020). arXiv: 2005.02983 [hep-ex].

[9] ATLAS Collaboration. *Exotic Physics Searches*. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ExoticsPublicResults. 2018.

[10] ATLAS Collaboration. *Higgs and Diboson Searches*. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/HDBSPublicResults. 2019.

[11] ATLAS Collaboration. “Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS”. In: *ATL-PHYS-PUB-2018-014* (2018). URL: http://cds.cern.ch/record/2630973.

[12] ATLAS Collaboration. *Supersymmetry searches*. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults. 2018.

[13] Pierre Baldi et al. “Parameterized neural networks for high-energy physics”. In: *Eur. Phys. J. C* 76.5 (2016), p. 235. DOI: 10.1140/epjc/s10052-016-4099-4. arXiv: 1601.07913 [hep-ex].

[14] Andrew Blance, Michael Spannowsky, and Philip Waite. “Adversarially-trained autoencoders for robust unsupervised new physics searches”. In: *JHEP* 10 (2019), p. 047. DOI: 10.1007/JHEP10(2019)047. arXiv: 1905.10384 [hep-ph].

[15] Layne Bradshaw et al. “Mass Agnostic Jet Taggers”. In: *SciPost Phys.* 8.1 (2020), p. 011. DOI: 10.21468/SciPostPhys.8.1.011. arXiv: 1908.08959 [hep-ph].

[16] Olmo Cerri et al. “Variational Autoencoders for New Physics Mining at the Large Hadron Collider”. In: *JHEP* 05 (2019), p. 036. DOI: 10.1007/JHEP05(2019)036. arXiv: 1811.10276 [hep-ex].

[17] Taoli Cheng et al. “Variational Autoencoders for Anomalous Jet Tagging”. In: (July 2020). arXiv: 2007.01850 [hep-ph].

[18] Fraciois Chollet. *Keras*. https://github.com/fchollet/keras. 2017.

[19] CMS Collaboration. *CMS Beyond-two-generations (B2G) Public Physics Results*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsB2G. 2018.

[20] CMS Collaboration. *CMS Exotica Public Physics Results*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsEXO. 2018.

[21] CMS Collaboration. *CMS Supersymmetry Physics Results*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS. 2018.

[22] Jack H. Collins, Kiel Howe, and Benjamin Nachman. “Anomaly Detection for Resonant New Physics with Machine Learning”. In: *Phys. Rev. Lett.* 121.24 (2018), p. 241803. DOI: 10.1103/PhysRevLett.121.241803. arXiv: 1805.02664 [hep-ph].

[23] Jack H. Collins, Kiel Howe, and Benjamin Nachman. “Extending the search for new resonances with machine learning”. In: *Phys. Rev.* D99.1 (2019), p. 014038. DOI: 10.1103/PhysRevD.99.014038. arXiv: 1902.02634 [hep-ph].

[24] Kyle Cranmer, Juan Pavez, and Gilles Louppe. “Approximating Likelihood Ratios with Calibrated Discriminative Classifiers”. In: (June 2015). arXiv: 1506.02169 [stat.AP].

[25] Raffaele Tito D’Agnolo and Andrea Wulzer. “Learning New Physics from a Machine”. In: *Phys. Rev.* D99.1 (2019), p. 015014. DOI: 10.1103/PhysRevD.99.015014. arXiv: 1806.02350 [hep-ph].

[26] Raffaele Tito D’Agnolo et al. “Learning Multivariate New Physics”. In: (2019). arXiv: 1912.12155 [hep-ph].

[27] Andrea De Simone and Thomas Jacques. “Guiding New Physics Searches with Unsupervised Learning”. In: *Eur. Phys. J.* C79.4 (2019), p. 289. DOI: 10.1140/epjc/s10052-019-6787-3. arXiv: 1807.06038 [hep-ph].

[28] Barry M. Dillon, Darius A. Faroughy, and Jernej F. Kamenik. “Uncovering latent jet substructure”. In: *Phys. Rev.* D100.5 (2019), p. 056002. DOI: 10.1103/PhysRevD.100.056002. arXiv: 1904.04200 [hep-ph].

[29] Barry M. Dillon et al. “Learning the latent structure of collider events”. In: (May 2020). arXiv: 2005.12319 [hep-ph].

[30] James Dolen et al. “Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure”. In: *JHEP* 05 (2016), p. 156. DOI: 10.1007/JHEP05(2016)156. arXiv: 1603.00027 [hep-ph].

[31] Christoph Englert et al. “Machine Learning Uncertainties with Adversarial Neural Networks”. In: *Eur. Phys. J.* C79.1 (2019), p. 4. DOI: 10.1140/epjc/s10052-018-6511-8. arXiv: 1807.08763 [hep-ph].

[32] Marco Farina, Yuichiro Nakai, and David Shih. “Searching for New Physics with Deep Autoencoders”. In: *Phys. Rev. D* 101.7 (2020), p. 075021. DOI: 10.1103/PhysRevD.101.075021. arXiv: 1808.08992 [hep-ph].

[33] M. Schwartz D. Shih G. Kasieczka B. Nachman. “ABCDisCo: Automating the ABCD Method with Machine Learning”. In: (2020). arXiv: 2007.14400 [hep-ph].

[34] Dan Guest, Kyle Cranmer, and Daniel Whiteson. “Deep Learning and its Application to LHC Physics”. In: (2018). arXiv: 1806.11484 [hep-ex].

[35] Jan Hajer et al. “Novelty Detection Meets Collider Physics”. In: *Phys. Rev. D* 101.7 (2020), p. 076015. DOI: 10.1103/PhysRevD.101.076015. arXiv: 1807.10261 [hep-ph].

[36] Theo Heimel et al. “QCD or What?” In: *SciPost Phys.* 6.3 (2019), p. 030. DOI: 10.21468/SciPostPhys.6.3.030. arXiv: 1808.08979 [hep-ph].

[37] Gregor Kasieczka and David Shih. “DisCo Fever: Robust Networks Through Distance Correlation”. In: (Jan. 2020). arXiv: 2001.05310 [hep-ph].

[38] Charanjit K. Khosa and Veronica Sanz. “Anomaly Awareness”. In: (July 2020). arXiv: 2007.14462 [cs.LG].

[39] Diederik Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: (2014). arXiv: 1412.6980 [cs].

[40] Oliver Knapp et al. “Adversarially Learned Anomaly Detection on CMS Open Data: re-discovering the top quark”. In: (2020). arXiv: 2005.01598 [hep-ex].

[41] 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”. In: *Phys. Rept.* 841 (2020), pp. 1–63. DOI: 10.1016/j.physrep.2019.11.001. arXiv: 1709.04464 [hep-ph].

[42] LHCb Collaboration. *Publications of the QCD, Electroweak and Exotica Working Group*. http://lhcbproject.web.cern.ch/lhcbproject/Publications/LHCbProjectPublic/Summary_QEE.html. 2019.

[43] Gilles Louppe, Michael Kagan, and Kyle Cranmer. “Learning to Pivot with Adversarial Networks”. In: *Advances in Neural Information Processing Systems* 30 (2017), p. 981. arXiv: 1611.01046 [stat.ME].

[44] Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler. “Classification without labels: Learning from mixed samples in high energy physics”. In: *JHEP* 10 (2017), p. 174. DOI: 10.1007/JHEP10(2017)174. arXiv: 1708.02949 [hep-ph].

[45] Ian Moult, Benjamin Nachman, and Duff Neill. “Convolved Substructure: Analytically Decorrelating Jet Substructure Observables”. In: *JHEP* 05 (2018), p. 002. DOI: 10.1007/JHEP05(2018)002. arXiv: 1710.06859 [hep-ph].

[46] Anna Mullin et al. “Does SUSY have friends? A new approach for LHC event analysis”. In: (2019). arXiv: 1912.10625 [hep-ph].

[47] Benjamin Nachman and David Shih. “Anomaly Detection with Density Estimation”. In: *Phys. Rev. D* 101 (2020), p. 075042. DOI: 10.1103/PhysRevD.101.075042. arXiv: 2001.04990 [hep-ph].

[48] Alexander Radovic et al. “Machine learning at the energy and intensity frontiers of particle physics”. In: *Nature* 560.7716 (2018), pp. 41–48. DOI: 10.1038/s41586-018-0361-2.

[49] M. Crispim Romao, N.F. Castro, and R. Pedro. “Finding New Physics without learning about it: Anomaly Detection as a tool for Searches at Colliders”. In: (June 2020). arXiv: 2006.05432 [hep-ph].

[50] M. Crispim Romao et al. “Use of a Generalized Energy Mover’s Distance in the Search for Rare Phenomena at Colliders”. In: (Apr. 2020). arXiv: 2004.09360 [hep-ph].

[51] M. Romão Crispim et al. "Transferability of Deep Learning Models in Searches for New Physics at Colliders". In: *Phys. Rev. D* 101.3 (2020), p. 035042. DOI: 10.1103/PhysRevD.101.035042. arXiv: 1912.04220 [hep-ph].

[52] Tuhin S. Roy and Aravind H. Vijay. "A robust anomaly finder based on autoencoder". In: (2019). arXiv: 1903.02032 [hep-ph].

[53] Chase Shimmin et al. "Decorrelated Jet Substructure Tagging using Adversarial Neural Networks". In: *Phys. Rev. D* 96.7 (2017), p. 074034. DOI: 10.1103/PhysRevD.96.074034. arXiv: 1703.03507 [hep-ex].

[54] Justin Stevens and Mike Williams. "uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers". In: *JINST* 8 (2013), P12013. DOI: 10.1088/1748-0221/8/12/P12013. arXiv: 1305.7248 [nucl-ex].

[55] Punnathat Thaprasop et al. "Unsupervised Outlier Detection in Heavy-Ion Collisions". In: (July 2020). arXiv: 2007.15830 [hep-ex].

[56] Stefan Wunsch et al. "Reducing the dependence of the neural network function to systematic uncertainties in the input space". In: *Comput. Softw. Big Sci.* 4.1 (2020), p. 5. DOI: 10.1007/s41781-020-00037-9. arXiv: 1907.11674 [physics.data-an].

[57] Li-Gang Xia. "QBDT, a new boosting decision tree method with systematical uncertainties into training for High Energy Physics". In: *Nucl. Instrum. Meth.* A930 (2019), pp. 15–26. DOI: 10.1016/j.nima.2019.03.088. arXiv: 1810.08387 [physics.data-an].