Simulation-based anomaly detection for multileptons at the LHC

Katarzyna Krzyzanska[a] **and Benjamin Nachman**[b,c]

[a]*Department of Physics, Princeton University,*
Princeton, NJ 08544, U.S.A.

[b]*Physics Division, Lawrence Berkeley National Laboratory,*
Berkeley, CA 94720, U.S.A.

[c]*Berkeley Institute for Data Science, University of California,*
Berkeley, CA 94720, U.S.A.

E-mail: kk19@princeton.edu, bpnachman@lbl.gov

Abstract: Decays of Higgs boson-like particles into multileptons is a well-motivated process for investigating physics beyond the Standard Model (SM). A unique feature of this final state is the precision with which the SM is known. As a result, simulations are used directly to estimate the background. Current searches consider specific models and typically focus on those with a single free parameter to simplify the analysis and interpretation. In this paper, we explore recent proposals for signal model agnostic searches using machine learning in the multilepton final state. These tools can be used to simultaneously search for many models, some of which have no dedicated search at the Large Hadron Collider. We find that the machine learning methods offer broad coverage across parameter space beyond where current searches are sensitive, with a necessary loss of performance compared to dedicated searches by only about one order of magnitude.

Contents

1 Introduction

Many models of new physics contain multiple free parameters that are not predicated by the theory. Typically, searches pick a one- or two-dimensional slice through this space to set limits on model parameters (see refs. [1–7]). This is an effective way to visualize results, but it can also hide potential anomalous regions of the parameter space that are not well-captured by particular slices. In the usual approach, a small number of *signal regions* (SRs) are constructed to achieve broad sensitivity over the low-dimensional slices through parameter space. These SRs are often optimized using benchmark models. In order to improve sensitivity, SR definitions are sometimes parameterized by a subset of the BSM parameters. Scanning through the parameter space in this way will improve the sensitivity to particular models away from the benchmark models at the cost of analysis complexity and a larger trials factor.

An alternative search strategy is *anomaly detection* (AD). AD searches reduce signal or background model dependence by directly training with data. As real data do not have labeled examples (signal versus background) for analysis optimization, these analysis techniques are called *less than supervised*. This paper will focus on signal model-agnostic AD methods. A variety of classical AD searches of this type have been conducted at D0 [8–11], H1 [12, 13], ALEPH [14], CDF [15–17], CMS [18–21], and ATLAS [22–24].

Recently, there have been many proposals for automating AD methods with machine learning [25–63, 63–82] (see refs. [81–84] for overviews of the field). One particularly sensitive class of signal model agnostic searches train classifiers to directly distinguish data in a particular region of phase space (SR) from a prediction of the SM background [25–31, 31, 63, 85]. These searches are called weakly (or semi) supervised because one dataset has known labels (background) while one dataset has noisy labels (background + maybe signal).

These weakly supervised methods are distinguished by the dataset used to estimate the background. If the background is well-understood theoretically, then the reference sample could be simulation [25, 28, 63, 85, 86]. This has the advantage that the background prediction does not need to be learned, but has the disadvantage of being strongly background-model dependent. There are few final states at the LHC for which the background is known precisely enough to be used directly for background estimation. One exception is the final state with four charged leptons. Both ATLAS [87–90] and CMS [91–94] directly use Monte Carlo (MC) simulations to estimate the background and ATLAS even uses machine learning to isolate particular signals [87]. While powerful, this approach is signal model-specific and does not readily extend to models with multidimensional parameters.

In this paper, we explore AD methods applied to the four lepton final state. We train classifiers to distinguish (simulated) data from background predictions. This approach is complementary to direct searches and has broad sensitivity in a signal model-agnostic approach. In particular, we consider a three-dimensional space spanned by the four-lepton invariant mass and the masses of lepton pairs. Even though this is a low-dimensional space, there is currently no search sensitive to $A \to B(\to e^+ e^-)C(\to \mu^+ \mu^-)$ where all three of these masses (m_A, m_B, m_C) are different. One reason that no such search exists, despite plenty of motivation [95], is that it is practically difficult to construct individual search regions for so many mass hypotheses. The method we study in this paper is independent of the signal hypothesis so all signals are essentially tested simultaneously. More importantly, if there was a scan over all hypotheses, there would be a significant look elsewhere effect (trials factor). The AD approach we explore here does not have a trials factor. This means that the local p-value will be worse with our method for a given mass hypothesis, but it may result in a more sensitive global p-value. In addition to proposing a new, specific search that could extend the physics reach of the LHC, we also explore different methods/statistics for performing the AD.

This paper is organized as follows. Section 2 introduces the simulated samples used for the machine learning studies. The machine learning methods are introduced in section 3 and numerical results are presented in section 4. The paper ends with conclusions and outlook in section 5.

2 Simulations

The generation of background and signal events is performed with MAD-GRAPH5_AMC@NLO 2.8.0 [96]. The signal is generated using the Higgs Effective Field Theory (heft) via p p > h with a variable Higgs mass and the background is generated also using heft via p p > e+ e- mu+ mu-. In principle, the same analysis could be applied in the 4-electron and 4-muon final states, but additional complication arises from combinatorical factors. We leave the exploration of such final states to future work. Simulated events are passed to decay (signal only), parton showering, and hadronization using PYTHIA 8.244 [97–99] with its default settings. The detector simulation is parameterized with DELPHES 3.4.2 [100–102] using the default CMS card. In what follows, we will assume that the background is known without any systematic

uncertainties and so the 'data' will be an independent, but statistically identical copy of the SM background. See ref. [85] for a discussion of how this could be extended to include systematic uncertainties. The number of background events is chosen to match[1] the LHC Run 2 dataset of about $150\,\text{fb}^{-1}$.

Focusing only on the leptons,[2] each event is characterized by 12 numbers (four three-momenta). For many models of the form $pp \to A \to B(\to e^+e^-)C(\to \mu^+\mu^-)$, the three masses $m_{e^+e^-\mu^+\mu^-}, m_{e^+e^-}, m_{\mu^+\mu^-}$ are nearly sufficient statistics for characterizing the new physics. In this paper, we consider signals of this form, where A is a variable-mass Higgs boson that decays to two different mass scalars. For this reason, we focus on the three-dimensional problem in this paper. Non-resonant signals and signals with non-trivial spin structures could benefit from using more of the phase space.

The spectra of the three invariant masses for the background and three representative signals are presented in figure 1. As expected, the di-electron and di-muon invariant masses peak near the Z boson mass of $90\,\text{GeV}$ [103] and there are peaks in the four-lepton invariant mass at the Z peak and the Higgs boson mass of about $125\,\text{GeV}$ [103]. Each of the signals is resonant in all three observables with peaks at the masses of the particles. In particular, our three signal models have parent masses of 125, 150, and $250\,\text{GeV}$, respectively. In each case, the parent particle decays to two children, with masses of 25 and $15\,\text{GeV}$ for electrons and muons, respectively. These parameters are chosen to be representative; the resulting less-than-supervised analysis does not rely on them in the machine learning training.

3   Methods

Our machine learning approach is simple, but powerful: we train a classifier to distinguish SM background from 'data'. This 'data' is SM background with some amount of injected signal. If the classifier is able to significantly differentiate these samples, then there is evidence that the 'data' contains BSM events. The significance is determined via bootstrapping [104]. In particular, we create N bootstrap datasets by sampling (with replacement) from the SM background. We then train N classifiers, where each one distinguishes the bootstrap 'data' from the SM background. We compare the classifier performance in the nominal case with the distribution from the bootstraps to compute p-values.

There is no unique choice for which statistic to use to evaluate the performance. Theoretically, the optimal statistics in the absence of systematic uncertainties are those that are monotonically related to the likelihood ratio[3] [105]. In a sense, we are performing a mul-

---

[1]In principle, the simulation could have a larger number of events than in 'data'. Studies in ref. [31] suggest this could be effective in the case of surrogate models (not simulation). We leave a detailed study of the relative size of the simulation dataset in to future studies.

[2]Other event properties could also be useful for discrimination. However, information about the hadronic final state is known with less precision and thus may introduce the need to involve data-driven background estimation.
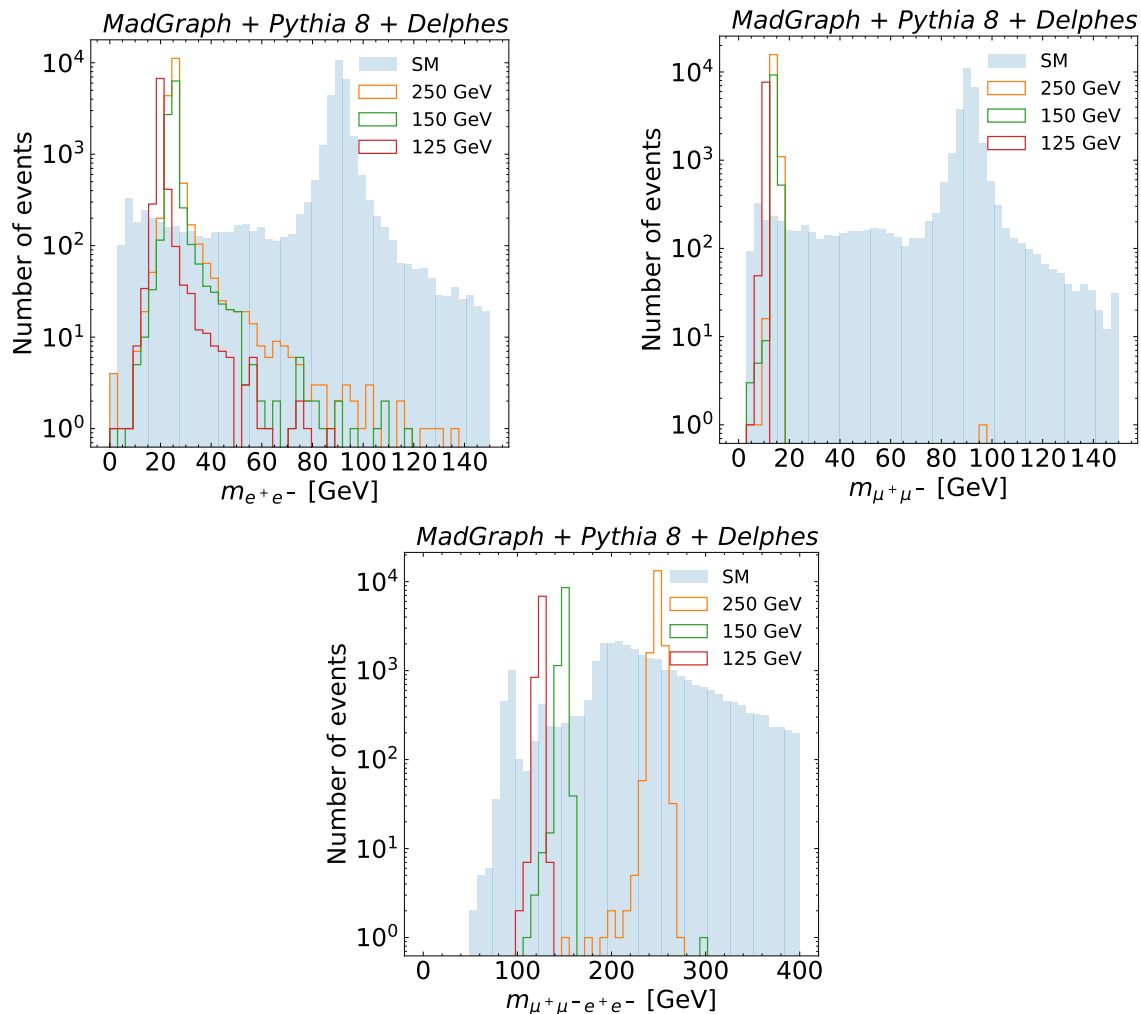
[3]See also appendix A in ref. [30].

Figure 1. The three dimensions used for machine learning: $m_{e^+e^-}$ (left), $m_{\mu^+\mu^-}$ (right) and $m_{4\ell}$ (bottom).

tidimensional, unbinned, and nearly non-parameteric[4] goodness of fit. It may be possible to exploit asymptotic formulae to avoid bootstrapping with additional assumptions [25], but we focus on the numerical approach because it is more general. See ref. [63] for a comparison of bootstrapping to other approaches.

The classifiers are parameterized as fully connected deep neural networks.[5] These networks are implemented in TENSORFLOW [106] 2.2.0 and KERAS [107] and optimized using ADAM [108]. The networks have three hidden layers with 50 nodes per layer. Rectified linear activation functions are used for all intermediate layers. We consider two

---

[4]We will use neural networks for the machine learning; these are by definition parameterized functions, but there are so many parameters that they are effectively non-parametric.

[5]Given the low-dimensionality of our demonstration, a shallower classifier such as a Boosted Decision Tree would likely also be effective. However, neural network training is straight forward and naturally extends to higher dimensions.

loss functions (L):

- Binary Cross Entropy Loss (BCE). This is the most widely-used loss function in machine learning:

$$L[f] = - \sum_{x \in \text{'data'}} \log(f(x)) - \sum_{x \in \text{background}} \log(1 - f(x)), \qquad (3.1)$$

  where the last layer of the neural network is a sigmoid so that $0 \leq f \leq 1$.

- Maximum Likelihood Classifier Loss [25, 109] (MLC). This loss function directly learns the (log) likelihood ratio, which results in useful asymptotic statistical properties:

$$L[f] = - \sum_{x \in \text{'data'}} (f(x) - 1) - \sum_{x \in \text{background}} e^{f(x)}, \qquad (3.2)$$

  where the last layer of the neural network is now linear so $-\infty < f < \infty$.

All of the inputs to the neural networks are standardized to have zero mean and unit variance. In order to improve performance, we train 10 neural networks with different random initialization and take the average value per event. The networks are each trained for a fixed 20 epochs. None of the network or training parameters were highly optimized.

In addition to comparing BCE and MLC, we also explore multiple metrics for performing the hypothesis test. In particular, we consider the average loss, the average score, the maximum score, and the standard deviation over scores. The average loss is the quantity that is optimized. By averaging over the full phase space, it could be that some information about localized anomalies is lost. This is the motivation for examining also the maximum score and the standard deviation across scores. These quantities are more sensitive to localized deviations, but are also more sensitive to background fluctuations. See also ref. [63] for studies with alternative metrics, such as area under the receiver operating characteristic (ROC) curve.

4 Results

The values of the four statistics described in the previous section as a function of the number of injected signal events are presented in figure 2 and 3 for the MLC loss and figure 4 and 5 for the BCE loss. As expected, the loss decreases and the mean/max/standard deviation of the scores increases with more signal events injected. As the number of injected signal events goes to zero, the standard deviation over the scores goes to zero. In the MLC case, the neural network approximates the log likelihood ratio and so the average score approaches 0 as the number of injected signal events goes to zero. The MLC loss is approximately minus two times a χ^2 random variable [25, 109]. In the BCE case, the neural network approximates $p(\text{'data'}|m_{e^+e^-}, m_{\mu^+\mu^-}, m_{4\ell})$, which approaches a δ-function centered at 0.5 as the number of signal events goes to zero. The BCE loss itself approaches $\log(2) \approx 0.69$.

The green (yellow) band corresponds to the 1σ (2σ) region of the background-only hypothesis. The bands for each signal correspond to the standard deviation across 10

independent signal injections. We show zoomed-out (figure 2 and 4) and zoomed-in (figure 3 and 5) versions to emphasize where the bands cross the 2σ line, which approximately corresponds to the 95% exclusion limit. The crossings are listed in tables 1 and 2, respectively. For both loss functions, we find that the standard deviations over the scores is very effective. This could be because the standard deviation is much more sensitive to outliers in the score distribution than the mean. Even in the case that there is a signal, most events are still background, and so the scores are similar to the background-only case. The presence of new physics will therefore be outliers in the score distribution. The loss in the case of the MLC is more useful than the BCE loss and has comparable performance to the score standard deviation. This may be expected, since the MLC loss asymptotically is monotonically related to the likelihood ratio of the two samples (and thus optimal). The max score is must more useful for BCE than for MLC.

The actual limits are comparable for both losses. For example, the limits for the 250 GeV signal are about $0.4^{+0.1}_{-0.3}$ fb (0.4 ± 0.2) for MLC (BCE) and for the other signals are about $0.8^{+0.7}_{-0.2}$ fb (1.2 ± 0.2) for MLC (BCE). To put these limits in context, the lmits in ref. [90] (CMS is similar) for $h \to XX$ has a limit of about 0.04 fb in the $e^+e^-\mu^+\mu^-$ channel (nearly independent of m_X). This limit is about 10 times better than what we find which is the price we pay for being model agnostic.[6] On the other hand, the ATLAS/CMS searches have no sensitivity outside of the fixed m_X, m_h mass windows, while we have broad sensitivity that is nearly independent of all masses.

---

[6]In addition to the impact of the simplifying assumptions we make in the analysis presented here compared to the more detailed ATLAS/CMS searches.
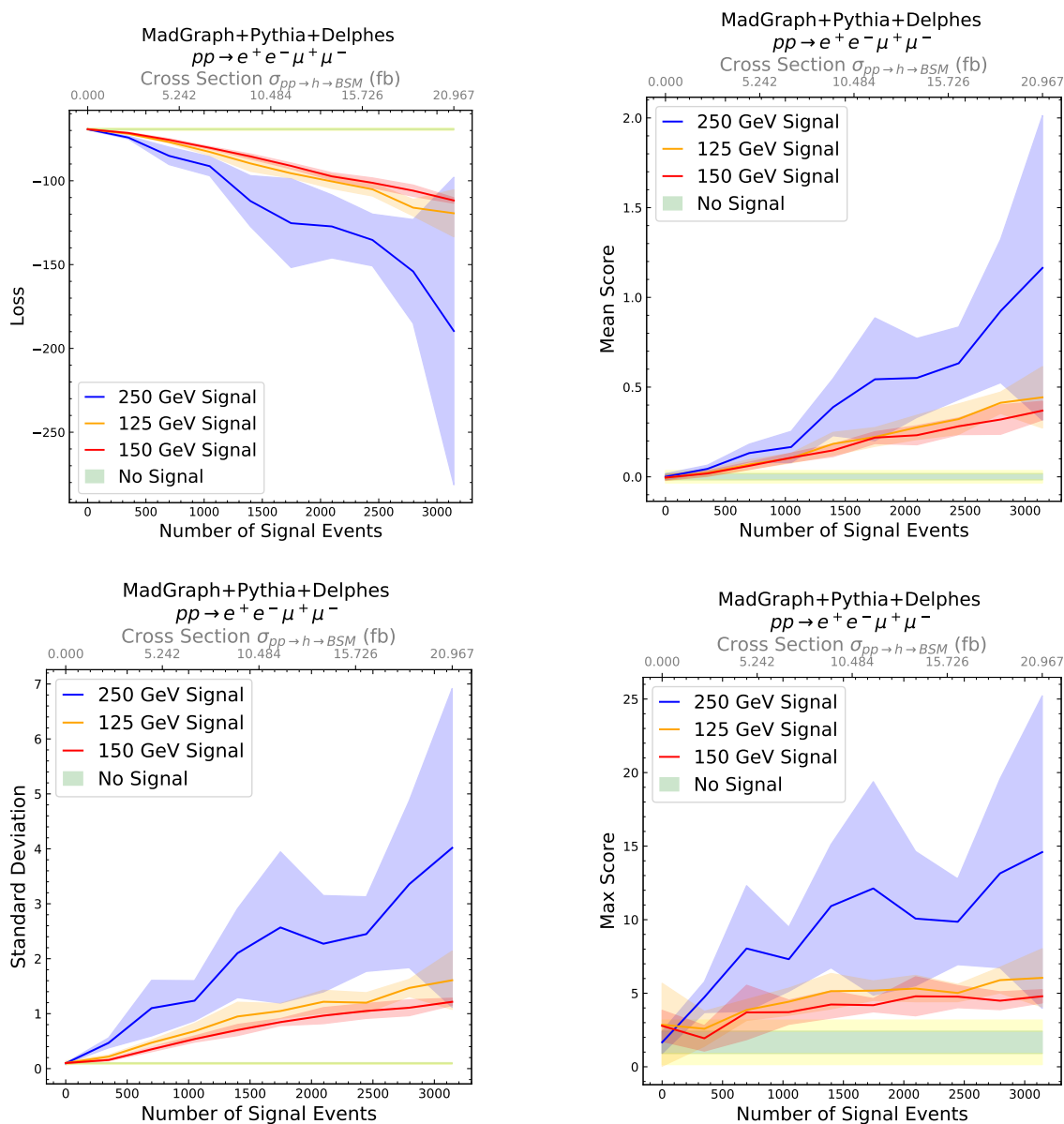
Figure 2. MLC Loss. The average loss (top left), average score (top right), score standard deviation (lower left), and maximum score (lower right) when using the MLC loss as a function of the number of injected signals. The green (yellow) band corresponds to the 1σ (2σ) region of the background-only hypothesis. The bands for each signal correspond to the standard deviation across 10 independent signal injections. The upper axis provides the injected cross section, where the number of background events corresponds to $150\,\text{fb}^{-1}$.
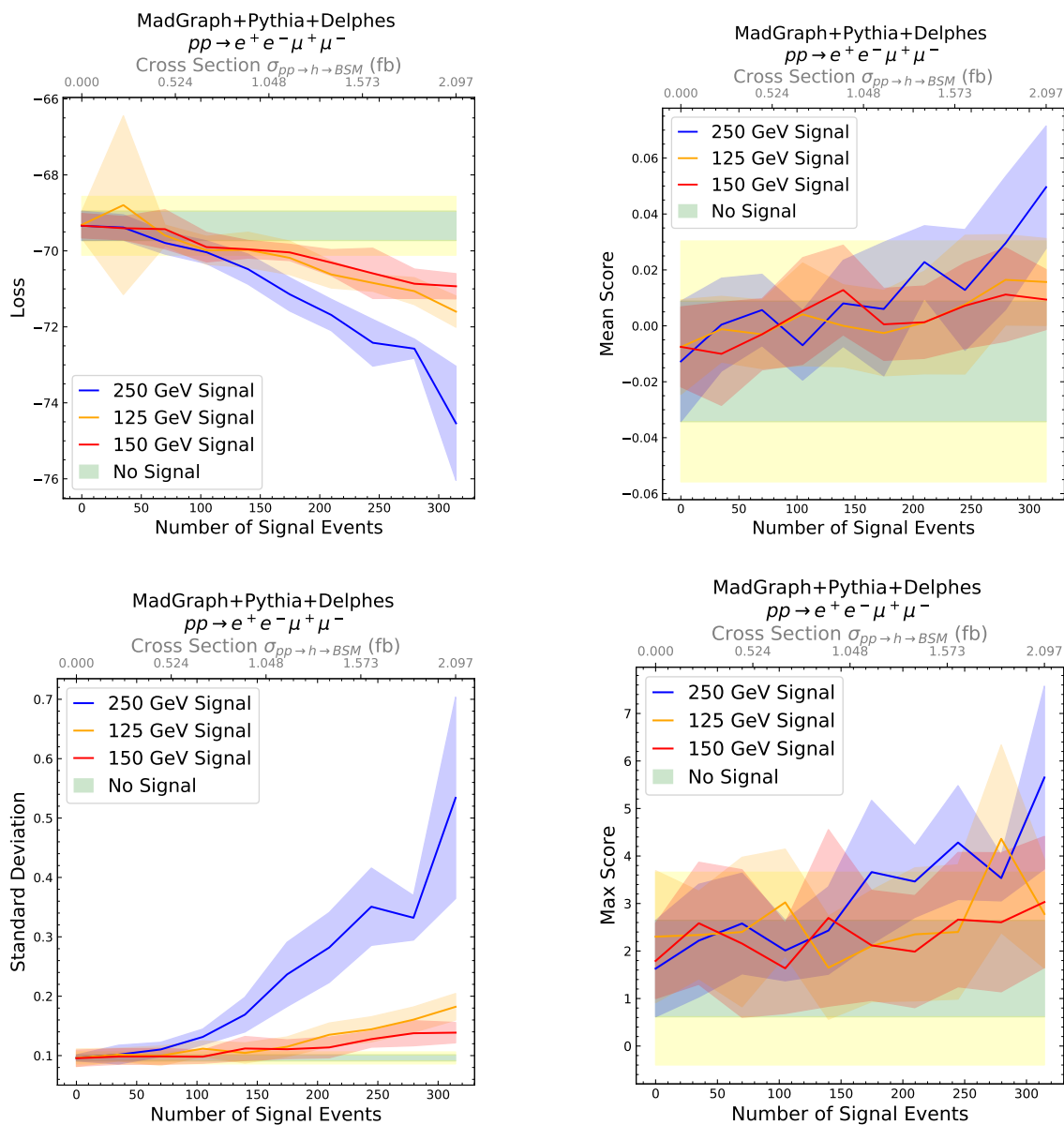
Figure 3. MLC Loss. Same as figure 2, but zoomed in by a factor of 10 on the horizontal axis. The 95% exclusion limits approximately correspond to where the lines for each signal cross the yellow 2σ band for the background-only hypothesis.
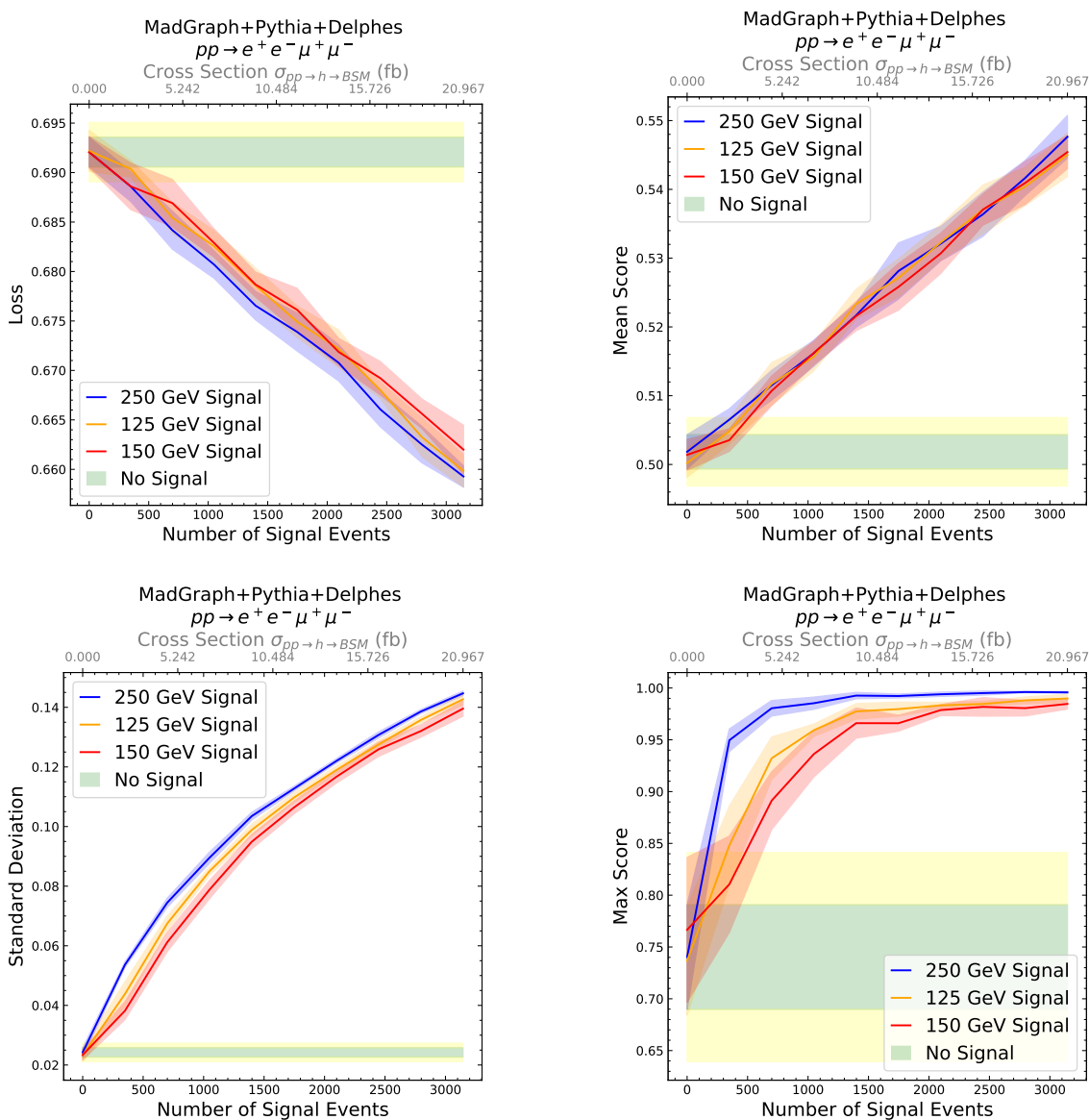
Figure 4. BCE Loss. The average loss (top left), average score (top right), score standard deviation (lower left), and maximum score (lower right) when using the MLC loss as a function of the number of injected signals. The green (yellow) band corresponds to the 1σ (2σ) region of the background-only hypothesis. The bands for each signal correspond to the standard deviation across 10 independent signal injections. The upper axis provides the injected cross section, where the number of background events corresponds to $150\,\mathrm{fb}^{-1}$.
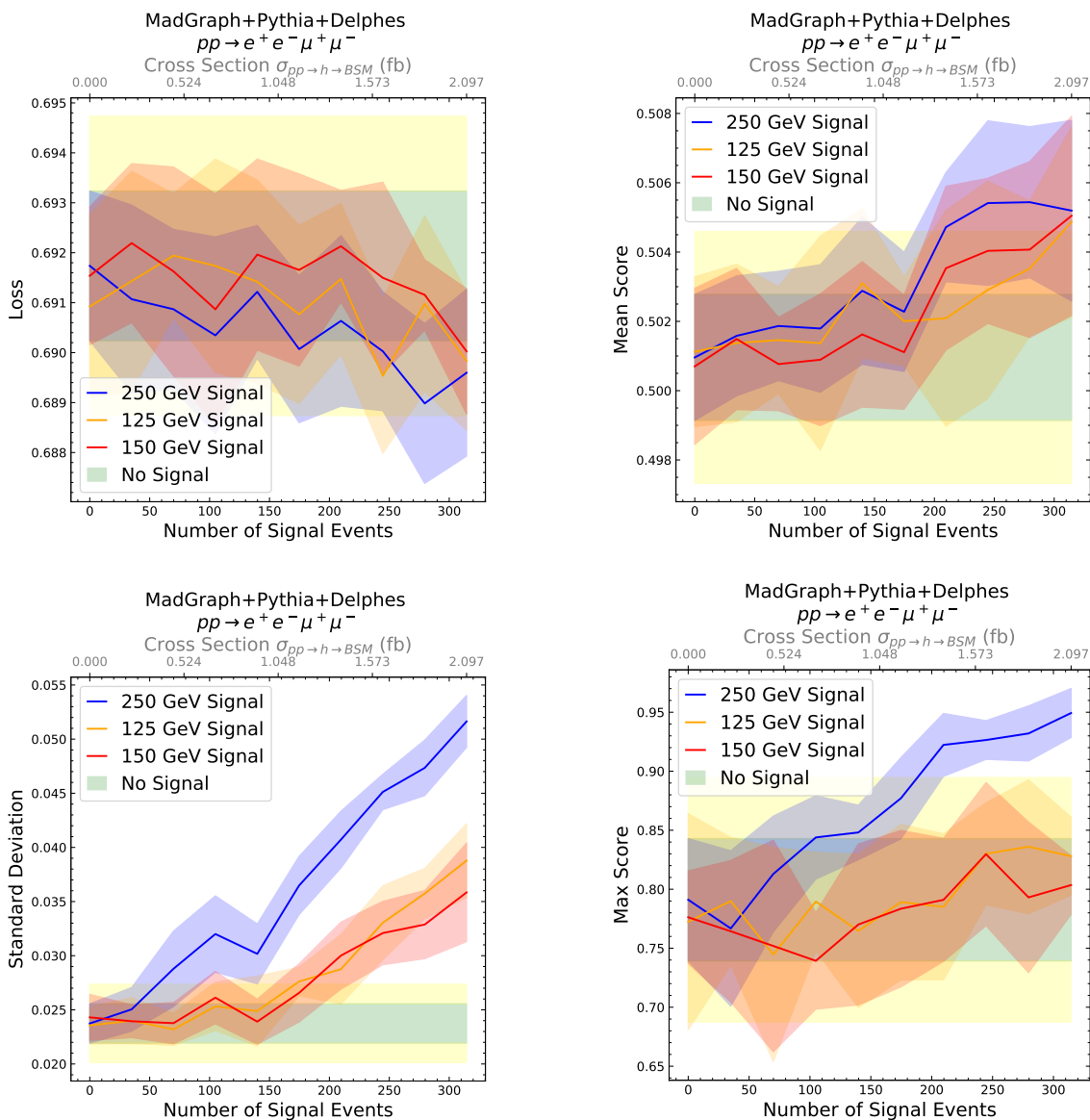
Figure 5. BCE Loss. Same as figure 4, but zoomed in by a factor of 10 on the horizontal axis. The 95% exclusion limits approximately correspond to where the lines for each signal cross the yellow 2σ band for the background-only hypothesis.

	Loss	Mean Score	Standard Deviation	Max Score
250 GeV	0.74	1.9	0.36	1.2
125 GeV	1.1	3.0	0.59	1.8
150 GeV	1.2	3.2	0.83	4.0

Table 1. MLC Loss. Each entry corresponds to the injected cross section limit at which the plots from figure 2 and 3 cross the 2σ band. The point of intersection is calculated as a linear interpolation between data points. These are taken from figure 3 unless the crossing occurs outside of that range.

	Loss	Mean Score	Standard Deviation	Max Score
250 GeV	2.0	1.4	0.4	1.3
125 GeV	2.9	2.0	1.1	2.2
150 GeV	2.0	2.0	1.2	3.2

Table 2. BCE Loss. Each entry corresponds to the injected cross section limit at which the plots from figure 4 and 5 cross the 2σ band. The point of intersection is calculated as a linear interpolation between data points. These are taken from figure 5 unless the crossing occurs outside of that range.

5 Conclusions

In this paper, we have explored the use of less-than-supervised anomaly detection for the four-lepton final state at the LHC. The motivation of this approach is that it is broadly sensitive to BSM models without specifying model parameters and is thus complementary to model-specific analyses. We examine decays of the form $A \to B(\to e^+e^-)C(\to \mu^+\mu^-)$ where the masses m_A, m_B, and m_C are different. As no existing search is sensitive to such decays, we are proposing a new search which could extend the reach of current experiments. The core methodology we use is not new. The idea of performing AD by comparing data with a reference sample was first introduced in refs. [26, 27] and studied in the context of a simulated reference sample in refs. [25, 28, 85]. Our contribution is to note that the four-lepton final state is special in that it is nearly unique at the LHC as a case where simulations are used directly to estimate Standard Model backgrounds. This allows for the data-versus-simulation strategy and we have explicitly demonstrated how the method works using multiple signal models. We have extended previous results by exploring different statistics that can be used to characterize the effective goodness of fit from the machine learning classifiers. While we have focused on resonant anomalies, we note that the methodology studied here will also work in the case of non-resonant modifications to the four-lepton final state. Many parts of the parameter space are statistics limited, but a full investigation of systematic uncertainties will be required for a complete experimental implementation in the future. A variety of uncertainty- and inference-aware methods have been proposed that can be used in this case [85, 110–117].

With no significant evidence for physics beyond the Standard Model at the LHC so far, it is critical to extend our analysis methodologies to include more model agnostic

approaches. It is likely that many techniques will be required to achieve broad sensitivity. We have explored a simulation-driven method in this paper for the four-lepton final state because of the existence of precise background models. Combining simulation-based AD with data-driven background estimation strategies may be able extend the scope of similar methods to a wide variety of final states at the LHC and beyond.

Acknowledgments

References

[1] ATLAS collaboration, *Exotic physics searches*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ExoticsPublicResults, (2022).

[2] ATLAS collaboration, *Supersymmetry searches*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults, (2022).

[3] ATLAS collaboration, *Higgs and diboson searches*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/HDBSPublicResults, (2022).

[4] CMS collaboration, *CMS exotica public physics results*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsEXO, (2022).

[5] CMS collaboration, *CMS supersymmetry physics results*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS, (2022).

[6] CMS collaboration, *CMS beyond-two-generations (B2G) public physics results*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsB2G, (2022).

[7] LHCb collaboration, *Publications of the QCD, electroweak and exotica working group*, http://lhcbproject.web.cern.ch/lhcbproject/Publications/LHCbProjectPublic/Summary_QEE.html, (2022).

[8] B. Abbott et al., *Quasi-model-independent search for new high p_T physics at D0*, [*Phys. Rev. Lett.* **86** (2001) 3712](#) [[hep-ex/0011071](#)].

[9] D0 collaboration, *Search for new physics in eμX data at D0 using SLEUTH: a quasi-model-independent search strategy for new physics*, [*Phys. Rev. D* **62** (2000) 092004](#) [[hep-ex/0006011](#)] [[INSPIRE](#)].

[10] D0 collaboration, *A quasi model independent search for new physics at large transverse momentum*, [*Phys. Rev. D* **64** (2001) 012004](#) [[hep-ex/0011067](#)] [[INSPIRE](#)].

[11] D0 collaboration, *A quasi-model-independent search for new high p_T physics at D0*, *Phys. Rev. Lett.* **86** (2001) 3712 [hep-ex/0011071] [INSPIRE].

[12] H1 collaboration, *A general search for new phenomena at HERA*, *Phys. Lett. B* **674** (2009) 257 [arXiv:0901.0507] [INSPIRE].

[13] H1 collaboration, *A general search for new phenomena in ep scattering at HERA*, *Phys. Lett. B* **602** (2004) 14 [hep-ex/0408044] [INSPIRE].

[14] K.S. Cranmer, *Searching for new physics: contributions to LEP and the LHC*, Ph.D. thesis, University of Wisconsin, Madison, WI, U.S.A. (2005) [INSPIRE].

[15] CDF collaboration, *Model-independent and quasi-model-independent search for new physics at CDF*, *Phys. Rev. D* **78** (2008) 012002 [arXiv:0712.1311] [INSPIRE].

[16] CDF collaboration, *Model-independent global search for new high-p_T physics at CDF*, arXiv:0712.2534 [INSPIRE].

[17] CDF collaboration, *Global search for new physics with* $2.0\,fb^{-1}$ *at CDF*, *Phys. Rev. D* **79** (2009) 011101 [arXiv:0809.3781] [INSPIRE].

[18] CMS collaboration, *MUSiC, a Model Unspecific Search for new physics, in pp Collisions at* $\sqrt{s} = 8\,TeV$, [INSPIRE].

[19] CMS collaboration, *Model unspecific search for new physics in pp collisions at* $\sqrt{s} = 7\,TeV$, [INSPIRE].

[20] CMS collaboration, *MUSiC, a model unspecific search for new physics, in pp collisions at* $\sqrt{s} = 13\,TeV$, CMS-PAS-EXO-19-008, CERN, Geneva, Switzerland (2020).

[21] CMS collaboration, *MUSiC: a model-unspecific search for new physics in proton-proton collisions at* $\sqrt{s} = 13\,TeV$, *Eur. Phys. J. C* **81** (2021) 629 [arXiv:2010.02984] [INSPIRE].

[22] ATLAS collaboration, *A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment*, *Eur. Phys. J. C* **79** (2019) 120 [arXiv:1807.07447] [INSPIRE].

[23] ATLAS collaboration, *A general search for new phenomena with the ATLAS detector in pp collisions at* $\sqrt{s} = 8\,TeV$, ATLAS-CONF-2014-006, CERN, Geneva, Switzerland (2014) [INSPIRE].

[24] ATLAS collaboration, *A general search for new phenomena with the ATLAS detector in pp collisions at* $\sqrt{s} = 7\,TeV$, ATLAS-CONF-2012-107, CERN, Geneva, Switzerland (2012) [INSPIRE].

[25] R.T. D'Agnolo and A. Wulzer, *Learning new physics from a machine*, *Phys. Rev. D* **99** (2019) 015014 [arXiv:1806.02350] [INSPIRE].

[26] J.H. Collins, K. Howe and B. Nachman, *Anomaly detection for resonant new physics with machine learning*, *Phys. Rev. Lett.* **121** (2018) 241803 [arXiv:1805.02664] [INSPIRE].

[27] J.H. Collins, K. Howe and B. Nachman, *Extending the search for new resonances with machine learning*, *Phys. Rev. D* **99** (2019) 014038 [arXiv:1902.02634] [INSPIRE].

[28] R.T. D'Agnolo, G. Grosso, M. Pierini, A. Wulzer and M. Zanetti, *Learning multivariate new physics*, *Eur. Phys. J. C* **81** (2021) 89 [arXiv:1912.12155] [INSPIRE].

[29] A. Andreassen, B. Nachman and D. Shih, *Simulation assisted likelihood-free anomaly detection*, *Phys. Rev. D* **101** (2020) 095004 [arXiv:2001.05001] [INSPIRE].

[30] B. Nachman and D. Shih, *Anomaly detection with density estimation*, *Phys. Rev. D* **101** (2020) 075042 [arXiv:2001.04990] [INSPIRE].

[31] A. Hallin et al., *Classifying anomalies through outer density estimation*, *Phys. Rev. D* **106** (2022) 055006 [arXiv:2109.00546] [INSPIRE].

[32] M. Farina, Y. Nakai and D. Shih, *Searching for new physics with deep autoencoders*, *Phys. Rev. D* **101** (2020) 075021 [arXiv:1808.08992] [INSPIRE].

[33] T. Heimel, G. Kasieczka, T. Plehn and J.M. Thompson, *QCD or what?*, *SciPost Phys.* **6** (2019) 030 [arXiv:1808.08979] [INSPIRE].

[34] T.S. Roy and A.H. Vijay, *A robust anomaly finder based on autoencoders*, arXiv:1903.02032 [INSPIRE].

[35] O. Cerri, T.Q. Nguyen, M. Pierini, M. Spiropulu and J.-R. Vlimant, *Variational autoencoders for new physics mining at the Large Hadron Collider*, *JHEP* **05** (2019) 036 [arXiv:1811.10276] [INSPIRE].

[36] A. Blance, M. Spannowsky and P. Waite, *Adversarially-trained autoencoders for robust unsupervised new physics searches*, *JHEP* **10** (2019) 047 [arXiv:1905.10384] [INSPIRE].

[37] J. Hajer, Y.-Y. Li, T. Liu and H. Wang, *Novelty detection meets collider physics*, *Phys. Rev. D* **101** (2020) 076015 [arXiv:1807.10261] [INSPIRE].

[38] A. De Simone and T. Jacques, *Guiding new physics searches with unsupervised learning*, *Eur. Phys. J. C* **79** (2019) 289 [arXiv:1807.06038] [INSPIRE].

[39] A. Mullin, S. Nicholls, H. Pacey, M. Parker, M. White and S. Williams, *Does SUSY have friends? A new approach for LHC event analysis*, *JHEP* **02** (2021) 160 [arXiv:1912.10625] [INSPIRE].

[40] A. Casa and G. Menardi, *Nonparametric semisupervised classification for signal detection in high energy physics*, arXiv:1809.02977 [INSPIRE].

[41] B.M. Dillon, D.A. Faroughy and J.F. Kamenik, *Uncovering latent jet substructure*, *Phys. Rev. D* **100** (2019) 056002 [arXiv:1904.04200] [INSPIRE].

[42] J.A. Aguilar-Saavedra, J.H. Collins and R.K. Mishra, *A generic anti-QCD jet tagger*, *JHEP* **11** (2017) 163 [arXiv:1709.01087] [INSPIRE].

[43] 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** (2020) 035042 [arXiv:1912.04220] [INSPIRE].

[44] 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** (2021) 192 [arXiv:2004.09360] [INSPIRE].

[45] O. Knapp, O. Cerri, G. Dissertori, T.Q. Nguyen, M. Pierini and J.-R. Vlimant, *Adversarially learned anomaly detection on CMS open data: re-discovering the top quark*, *Eur. Phys. J. Plus* **136** (2021) 236 [arXiv:2005.01598] [INSPIRE].

[46] ATLAS collaboration, *Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector*, *Phys. Rev. Lett.* **125** (2020) 131801 [arXiv:2005.02983] [INSPIRE].

[47] B.M. Dillon, D.A. Faroughy, J.F. Kamenik and M. Szewc, *Learning the latent structure of collider events*, *JHEP* **10** (2020) 206 [arXiv:2005.12319] [INSPIRE].

[48] 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** (2021) 27 [arXiv:2006.05432] [INSPIRE].

[49] O. Amram and C.M. Suarez, *Tag N' Train: a technique to train improved classifiers on unlabeled data*, *JHEP* **01** (2021) 153 [arXiv:2002.12376] [INSPIRE].

[50] T. Cheng, J.-F. Arguin, J. Leissner-Martin, J. Pilette and T. Golling, *Variational autoencoders for anomalous jet tagging*, *Phys. Rev. D* **107** (2023) 016002 [arXiv:2007.01850] [INSPIRE].

[51] C.K. Khosa and V. Sanz, *Anomaly awareness*, arXiv:2007.14462 [INSPIRE].

[52] P. Thaprasop, K. Zhou, J. Steinheimer and C. Herold, *Unsupervised outlier detection in heavy-ion collisions*, *Phys. Scripta* **96** (2021) 064003 [arXiv:2007.15830] [INSPIRE].

[53] S. Alexander et al., *Decoding dark matter substructure without supervision*, arXiv:2008.12731 [INSPIRE].

[54] J.A. Aguilar-Saavedra, F.R. Joaquim and J.F. Seabra, *Mass Unspecific Supervised Tagging (MUST) for boosted jets*, *JHEP* **03** (2021) 012 [arXiv:2008.12792] [INSPIRE].

[55] J.A. Aguilar-Saavedra, *Anomaly detection from mass unspecific jet tagging*, *Eur. Phys. J. C* **82** (2022) 130 [arXiv:2111.02647] [INSPIRE].

[56] K. Benkendorfer, L.L. Pottier and B. Nachman, *Simulation-assisted decorrelation for resonant anomaly detection*, *Phys. Rev. D* **104** (2021) 035003 [arXiv:2009.02205] [INSPIRE].

[57] A.A. Pol, V. Berger, G. Cerminara, C. Germain and M. Pierini, *Anomaly detection with conditional variational autoencoders*, in *Eighteenth international conference on machine learning and applications*, (2020) [arXiv:2010.05531] [INSPIRE].

[58] V. Mikuni and F. Canelli, *Unsupervised clustering for collider physics*, *Phys. Rev. D* **103** (2021) 092007 [arXiv:2010.07106] [INSPIRE].

[59] M. van Beekveld et al., *Combining outlier analysis algorithms to identify new physics at the LHC*, *JHEP* **09** (2021) 024 [arXiv:2010.07940] [INSPIRE].

[60] S.E. Park, D. Rankin, S.-M. Udrescu, M. Yunus and P. Harris, *Quasi anomalous knowledge: searching for new physics with embedded knowledge*, *JHEP* **21** (2020) 030 [arXiv:2011.03550] [INSPIRE].

[61] D.A. Faroughy, *Uncovering hidden new physics patterns in collider events using Bayesian probabilistic models*, *PoS* **ICHEP2020** (2021) 238 [arXiv:2012.08579] [INSPIRE].

[62] G. Stein, U. Seljak and B. Dai, *Unsupervised in-distribution anomaly detection of new physics through conditional density estimation*, in 34th *conference on neural information processing systems*, (2020) [arXiv:2012.11638] [INSPIRE].

[63] P. Chakravarti, M. Kuusela, J. Lei and L. Wasserman, *Model-independent detection of new physics signals using interpretable semi-supervised classifier tests*, arXiv:2102.07679 [INSPIRE].

[64] J. Batson, C.G. Haaf, Y. Kahn and D.A. Roberts, *Topological obstructions to autoencoding*, *JHEP* **04** (2021) 280 [arXiv:2102.08380] [INSPIRE].

[65] A. Blance and M. Spannowsky, *Unsupervised event classification with graphs on classical and photonic quantum computers*, *JHEP* **21** (2020) 170 [arXiv:2103.03897] [INSPIRE].

[66] B. Bortolato, A. Smolkovič, B.M. Dillon and J.F. Kamenik, *Bump hunting in latent space*, *Phys. Rev. D* **105** (2022) 115009 [`arXiv:2103.06595`] [ɪɴSPIRE].

[67] J.H. Collins, P. Martín-Ramiro, B. Nachman and D. Shih, *Comparing weak- and unsupervised methods for resonant anomaly detection*, *Eur. Phys. J. C* **81** (2021) 617 [`arXiv:2104.02092`] [ɪɴSPIRE].

[68] B.M. Dillon, T. Plehn, C. Sauer and P. Sorrenson, *Better latent spaces for better autoencoders*, *SciPost Phys.* **11** (2021) 061 [`arXiv:2104.08291`] [ɪɴSPIRE].

[69] T. Finke, M. Krämer, A. Morandini, A. Mück and I. Oleksiyuk, *Autoencoders for unsupervised anomaly detection in high energy physics*, *JHEP* **06** (2021) 161 [`arXiv:2104.09051`] [ɪɴSPIRE].

[70] D. Shih, M.R. Buckley, L. Necib and J. Tamanas, *Via machinae: searching for stellar streams using unsupervised machine learning*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 5992 [`arXiv:2104.12789`] [ɪɴSPIRE].

[71] O. Atkinson, A. Bhardwaj, C. Englert, V.S. Ngairangbam and M. Spannowsky, *Anomaly detection with convolutional graph neural networks*, *JHEP* **08** (2021) 080 [`arXiv:2105.07988`] [ɪɴSPIRE].

[72] A. Kahn, J. Gonski, I. Ochoa, D. Williams and G. Brooijmans, *Anomalous jet identification via sequence modeling*, *JINST* **16** (2021) P08012 [`arXiv:2105.09274`] [ɪɴSPIRE].

[73] T. Dorigo, M. Fumanelli, C. Maccani, M. Mojsovska, G.C. Strong and B. Scarpa, *RanBox: anomaly detection in the copula space*, *JHEP* **01** (2023) 008 [`arXiv:2106.05747`] [ɪɴSPIRE].

[74] S. Caron, L. Hendriks and R. 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** (2022) 077 [`arXiv:2106.10164`] [ɪɴSPIRE].

[75] E. Govorkova, E. Puljak, T. Aarrestad, M. Pierini, K.A. Woźniak and J. Ngadiuba, *LHC physics dataset for unsupervised new physics detection at 40 MHz*, *Sci. Data* **9** (2022) 118 [`arXiv:2107.02157`] [ɪɴSPIRE].

[76] G. Kasieczka, B. Nachman and D. Shih, *New methods and datasets for group anomaly detection from fundamental physics*, in *Conference on knowledge discovery and data mining*, (2021) [`arXiv:2107.02821`] [ɪɴSPIRE].

[77] S. Volkovich, F. De Vito Halevy and S. Bressler, *A data-directed paradigm for BSM searches: the bump-hunting example*, *Eur. Phys. J. C* **82** (2022) 265 [`arXiv:2107.11573`] [ɪɴSPIRE].

[78] E. Govorkova et al., *Autoencoders on field-programmable gate arrays for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider*, *Nature Mach. Intell.* **4** (2022) 154 [`arXiv:2108.03986`] [ɪɴSPIRE].

[79] B. Ostdiek, *Deep set auto encoders for anomaly detection in particle physics*, *SciPost Phys.* **12** (2022) 045 [`arXiv:2109.01695`] [ɪɴSPIRE].

[80] K. Fraser, S. Homiller, R.K. Mishra, B. Ostdiek and M.D. Schwartz, *Challenges for unsupervised anomaly detection in particle physics*, *JHEP* **03** (2022) 066 [`arXiv:2110.06948`] [ɪɴSPIRE].

[81] G. Kasieczka et al., *The LHC olympics 2020 a community challenge for anomaly detection in high energy physics*, *Rept. Prog. Phys.* **84** (2021) 124201 [`arXiv:2101.08320`] [ɪɴSPIRE].

[82] T. Aarrestad et al., *The dark machines anomaly score challenge: benchmark data and model independent event classification for the Large Hadron Collider*, *SciPost Phys.* **12** (2022) 043 [arXiv:2105.14027] [INSPIRE].

[83] G. Karagiorgi, G. Kasieczka, S. Kravitz, B. Nachman and D. Shih, *Machine learning in the search for new fundamental physics*, arXiv:2112.03769 [INSPIRE].

[84] M. Feickert and B. Nachman, *A living review of machine learning for particle physics*, arXiv:2102.02770 [INSPIRE].

[85] R.T. d'Agnolo, G. Grosso, M. Pierini, A. Wulzer and M. Zanetti, *Learning new physics from an imperfect machine*, *Eur. Phys. J. C* **82** (2022) 275 [arXiv:2111.13633] [INSPIRE].

[86] M. Letizia et al., *Efficient kernel methods for model-independent new physics searches*, in *Proceedings of the deep learning for physical sciences workshop at NeurIPS*, (2021).

[87] ATLAS collaboration, *Search for heavy resonances decaying into a pair of Z bosons in the $\ell^+\ell^-\ell'^+\ell'^-$ and $\ell^+\ell^-\nu\bar{\nu}$ final states using $139\,fb^{-1}$ of proton-proton collisions at $\sqrt{s} = 13\,TeV$ with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) 332 [arXiv:2009.14791] [INSPIRE].

[88] ATLAS collaboration, *Search for Higgs boson decays to beyond-the-Standard-Model light bosons in four-lepton events with the ATLAS detector at $\sqrt{s} = 13\,TeV$*, *JHEP* **06** (2018) 166 [arXiv:1802.03388] [INSPIRE].

[89] ATLAS collaboration, *Measurements of the Higgs boson inclusive and differential fiducial cross sections in the 4ℓ decay channel at $\sqrt{s} = 13\,TeV$*, *Eur. Phys. J. C* **80** (2020) 942 [arXiv:2004.03969] [INSPIRE].

[90] ATLAS collaboration, *Search for Higgs bosons decaying into new spin-0 or spin-1 particles in four-lepton final states with the ATLAS detector with $139\,fb^{-1}$ of pp collision data at $\sqrt{s} = 13\,TeV$*, *JHEP* **03** (2022) 041 [arXiv:2110.13673] [INSPIRE].

[91] CMS collaboration, *Measurements of properties of the Higgs boson and search for an additional resonance in the four-lepton final state at $\sqrt{s} = 13\,TeV$*, CMS-PAS-HIG-16-033, CERN, Geneva, Switzerland (2016).

[92] CMS collaboration, *Search for a low-mass dilepton resonance in Higgs boson decays to four-lepton final states at $\sqrt{s} = 13\,TeV$*, CMS-PAS-HIG-19-007, CERN, Geneva, Switzerland (2020).

[93] CMS collaboration, *Search for low-mass dilepton resonances in Higgs boson decays to four-lepton final states in proton-proton collisions at $\sqrt{s} = 13\,TeV$*, *Eur. Phys. J. C* **82** (2022) 290 [arXiv:2111.01299] [INSPIRE].

[94] CMS collaboration, *Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state*, *Phys. Rev. D* **104** (2021) 052004 [arXiv:2104.12152] [INSPIRE].

[95] T. Robens and T. Stefaniak, *LHC benchmark scenarios for the real Higgs singlet extension of the standard model*, *Eur. Phys. J. C* **76** (2016) 268 [arXiv:1601.07880] [INSPIRE].

[96] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[97] T. Sjostrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

[98] T. Sjostrand, S. Mrenna and P.Z. Skands, *A brief introduction to PYTHIA* 8.1, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820] [InSPIRE].

[99] T. Sjöstrand et al., *An introduction to PYTHIA* 8.2, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [InSPIRE].

[100] DELPHES 3 collaboration, *DELPHES* 3, *a modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [arXiv:1307.6346] [InSPIRE].

[101] A. Mertens, *New features in DELPHES* 3, *J. Phys. Conf. Ser.* **608** (2015) 012045 [InSPIRE].

[102] M. Selvaggi, *DELPHES 3: a modular framework for fast-simulation of generic collider experiments*, *J. Phys. Conf. Ser.* **523** (2014) 012033 [InSPIRE].

[103] PARTICLE DATA GROUP collaboration, *Review of particle physics*, *PTEP* **2020** (2020) 083C01 [InSPIRE].

[104] B. Efron, *Bootstrap methods: another look at the jackknife*, *Annals Statist.* **7** (1979) 1 [InSPIRE].

[105] J. Neyman and E.S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, *Phil. Trans. Roy. Soc. Lond. A* **231** (1933) 289 [InSPIRE].

[106] M. Abadi et al., *TensorFlow: a system for large-scale machine learning*, arXiv:1605.08695.

[107] F. Chollet, *Keras*, https://github.com/fchollet/keras, (2017).

[108] D.P. Kingma and J. Ba, *Adam: a method for stochastic optimization*, arXiv:1412.6980 [InSPIRE].

[109] B. Nachman and J. Thaler, *Learning from many collider events at once*, *Phys. Rev. D* **103** (2021) 116013 [arXiv:2101.07263] [InSPIRE].

[110] A. Ghosh, B. Nachman and D. Whiteson, *Uncertainty-aware machine learning for high energy physics*, *Phys. Rev. D* **104** (2021) 056026 [arXiv:2105.08742] [InSPIRE].

[111] P. De Castro and T. Dorigo, *INFERNO: inference-aware neural optimisation*, *Comput. Phys. Commun.* **244** (2019) 170 [arXiv:1806.04743] [InSPIRE].

[112] S. Wunsch, S. Jörger, R. Wolf and G. Quast, *Optimal statistical inference in the presence of systematic uncertainties using neural network optimization based on binned Poisson likelihoods with nuisance parameters*, *Comput. Softw. Big Sci.* **5** (2021) 4 [arXiv:2003.07186] [InSPIRE].

[113] A. Elwood and D. Krücker, *Direct optimisation of the discovery significance when training neural networks to search for new physics in particle colliders*, arXiv:1806.00322 [InSPIRE].

[114] L.-G. Xia, *QBDT, a new boosting decision tree method with systematical uncertainties into training for high energy physics*, *Nucl. Instrum. Meth. A* **930** (2019) 15 [arXiv:1810.08387] [InSPIRE].

[115] T. Charnock, G. Lavaux and B.D. Wandelt, *Automatic physical inference with information maximizing neural networks*, *Phys. Rev. D* **97** (2018) 083004 [arXiv:1802.03537] [InSPIRE].

[116] J. Alsing and B. Wandelt, *Nuisance hardened data compression for fast likelihood-free inference*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 5093 [arXiv:1903.01473] [InSPIRE].

[117] N. Simpson and L. Heinrich, *neos: end-to-end-optimised summary statistics for high energy physics*, in 20th *international workshop on advanced computing and analysis techniques in physics research: AI decoded — towards sustainable, diverse, performant and effective scientific computing*, (2022) [arXiv:2203.05570] [InSPIRE].