

Boosting mono-jet searches with model-agnostic machine learning

Thorben Finke, Michael Krämer, Maximilian Lipp and Alexander Mück

*Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University,
D-52056 Aachen, Germany*

E-mail: finke@physik.rwth-aachen.de, mkraemer@physik.rwth-aachen.de,
maximilian.lipp@rwth-aachen.de, mueck@physik.rwth-aachen.de

ABSTRACT: We show how weakly supervised machine learning can improve the sensitivity of LHC mono-jet searches to new physics models with anomalous jet dynamics. The Classification Without Labels (CWoLa) method is used to extract all the information available from low-level detector information without any reference to specific new physics models. For the example of a strongly interacting dark matter model, we employ simulated data to show that the discovery potential of an existing generic search can be boosted considerably.

KEYWORDS: Dark Matter at Colliders, Jets and Jet Substructure

ARXIV EPRINT: [2204.11889](https://arxiv.org/abs/2204.11889)

Contents

1	Introduction	1
2	Setup	3
2.1	SM backgrounds	3
2.2	New physics example: strongly interacting dark matter model	4
3	The CWoLa method	5
4	CWoLa for anomalous mono-jets	6
5	Proof of principle	8
6	Reality checks	10
7	Conclusion	12
A	The DGCNN classifier	13
B	Classifier output	14

1 Introduction

The LHC is a designated discovery machine. It has delivered the discovery of a Standard Model-like Higgs particle [1, 2] almost ten years ago and has also provided a wealth of precision measurements. However, new physics beyond the Standard Model (SM) has not been found in direct searches, yet. This is in contrast to our hope that there is a link between LHC physics and dark matter. The LHC experiments will increase the amount of collected data in the coming years, and even more so in the high luminosity phase of the LHC. This vast amount of data has to be scrutinized in all possible ways, in particular also expecting the unexpected.

Model-agnostic Machine Learning (ML) techniques have been shown to provide sensitivity to various new physics signatures. Completely unsupervised methods based on learning a representation of the data to find anomalies in a completely data-driven way have been introduced in refs. [3–9]. Guiding ML methods with simulations while keeping an open mind to what new physics might look like has been used in refs. [10–14]. It is the ultimate vision to have a model-agnostic ML algorithm that uncovers new physics even if it has never been considered by any human physicist. However, there is a long way to go. Being able to tag potential signal events, e.g. with autoencoders or any other anomaly search algorithm, does not mean that one can make a statistically meaningful discovery [15]

if there is no strategy to compare to the SM expectation. Only recently, first steps have been made to develop such a general, complete strategy [16].

It is at the heart of any LHC new physics search to compare the background expectation with a measurement. If the background expectation can be measured in a control region (usually involving some theory assumptions), only data needs to be examined for potential discoveries. ML methods have been proposed to increase the sensitivity of searches employing control regions [17–28] and applied to experiment [29]. In a traditional resonance search or bump hunt, one simply counts events in side bands and signal regions and looks for deviations from a smooth distribution. In such a setup, model-agnostic ML is particularly promising since it can go far beyond counting events. ML can discover patterns in the signal region which are absent in the control regions, e.g. investigating jet substructure in the dijet invariant mass spectrum. To examine the potential of in this context, two community challenges have recently collected many interesting ideas [30, 31].

In this work we employ the Classification Without Labels (CWoLa) method [18] which is designed to uncover any difference between a control and a signal region. A supervised classifier is trained to tag any event as belonging to the signal or the control region. If the control and the signal region have identically distributed features in the absence of new physics this is supposed to be an impossible task. The events being tagged as most signal like will be equally distributed between the control and the signal region. However, if there is a different signature in the signal region due to new physics, the supervised classifier will recognize it and tag it as most signal like. Moreover, the method can provide a statistically meaningful discovery. Labels for new physics events, which are unavailable in real data, are not needed. Only a control region is needed which contains less new physics events than the signal region.

CWoLa has been successfully used for bump hunts [20, 24] and combined with density estimation to improve the sensitivity of the corresponding searches [22]. In the present work, we go beyond bump hunting. We show how to improve mono-jet searches at the LHC using CWoLa. The dominant background in a mono-jet search stems from invisibly decaying weak vector bosons. Hence, events with visibly decaying vector bosons naturally provide a suitable control region. A standard mono-jet search is a cut-and-count search, where features such as the structure of the observed jets do not play a role. Looking for any difference in the jet structure can boost the sensitivity to models with modified jet dynamics. This has been shown for a specific model using supervised ML for example in ref. [32]. Here, we show how the use of CWoLa improves the sensitivity to differences in jet structure in a model unspecific way. We demonstrate the general idea following the most recent ATLAS mono-jet search [33]. A similar search has also been performed by the CMS collaboration [34]. The discovery potential for new physics is illustrated employing a specific dark matter model with a strongly interacting dark sector [32, 35], highlighting the potential as well as the limitations of the method. In particular, we demonstrate that signal regions which are already limited by systematic uncertainties in a standard search, gain sensitivity using model-agnostic ML and will profit in particular from the high luminosity phase at the LHC.

This work is organised as follows: in section 2, we introduce the ATLAS mono-jet search and describe the simulation of all the required data to be used in the following

investigations, i.e. the SM backgrounds and the strongly interacting dark-matter model used as an example. Section 3 summarizes the general CWoLa idea, before the specific CWoLa setup for the mono-jet search is described in section 4. In section 5, we highlight the discovery potential of the method under simplifying assumptions, before we scrutinize and reinforce our findings in section 6. We conclude in section 7 and refer to appendices A and B for more details on the setup and the results, respectively.

2 Setup

The strategy to search for new physics as outlined in this paper is not tied to a specific mono-jet analysis, and it might also be promising for searches with a well-defined control region beyond a mono-jet analysis. However, for concreteness, we follow the ATLAS mono-jet search in ref. [33]. In this search, events with $E_T^{\text{miss}} > 200$ GeV and a leading anti- k_T jet with jet radius 0.4, $p_T^{\text{jet}} > 150$ GeV and $|\eta^{\text{jet}}| < 2.4$ are analysed using an integrated luminosity of 139 fb^{-1} of proton-proton collision data collected between 2015 and 2018 at the LHC in several inclusive and exclusive E_T^{miss} regions. At most three additional jets with $p_T^{\text{jet}} > 30$ GeV and $|\eta^{\text{jet}}| < 2.8$ are allowed in an event. Events with identified leptons are vetoed. We focus on an inclusive signal region with $E_T^{\text{miss}} > 250$ GeV and $\Delta\phi(\mathbf{p}_T^{\text{jet}}, \mathbf{p}_T^{\text{miss}}) > 0.4$ for all jets (denoted as IM1 in ref. [33]). For this region more than 10^6 events have been recorded. The choice of this signal region is motivated by the fact that the CWoLa method benefits from large amounts of training data and will in particular improve searches which are systematics limited. In the high-luminosity phase of the LHC, our machine-learning approach will become more and more applicable also at higher E_T^{miss} thresholds as more and more events are collected. Other machine learning applications for improving the sensitivity of mono-jet searches can be found in refs. [36–38].

2.1 SM backgrounds

The SM background in the signal region consists mainly of Z+jet production where the Z-boson decays invisibly into neutrinos (61%), followed by W+jet production where the W-boson decays invisibly, i.e. it decays leptonically and the charged lepton is not identified (31%). Smaller backgrounds are associated with top-quark production where one top quark decays leptonically (3.5%) and diboson production where one boson decays invisibly (2%).

To reduce the systematic errors of the theory prediction, the background estimate of the ATLAS search [33] uses dedicated control samples where charged leptons from the Z-boson or W-boson decays are identified. The control region is defined by the same cuts as the signal region, but the cut on E_T^{miss} is replaced by a cut on the recoil momentum p_T^{recoil} , where $\mathbf{p}_T^{\text{recoil}} = \mathbf{p}_T^l + \mathbf{p}_T^{\text{miss}}$ and \mathbf{p}_T^l is the transverse momentum of the identified leptons (one for W-boson decays, two for Z-boson decays). Whether an event belongs to the signal or the control region therefore only depends on the decay products of the vector boson. We make use of this control sample in our CWoLa approach as discussed in section 4.

Since we do not have access to LHC data, we investigate our CWoLa based approach using Monte-Carlo simulations. All events are simulated at 13 TeV center-of-mass energy using MadGraph5 [39] for the hard process, Pythia8.2 [40] as a parton shower, Delphes3 [41]

for a basic detector simulation using the ATLAS Card, and FastJet [42] for jet-clustering, with the default settings for all tools.

The CWoLa method is sensitive to differences in jet structure, comparing the jets in the control and the signal region. The leading QCD jets in the Z+jet and W+jet background processes are generated through initial-state radiation (ISR), so their structure is determined by QCD dynamics and independent of the decay channel and the decay dynamics of the vector boson. If the Z boson in the Z+jet process decays into neutrinos or charged leptons, the corresponding ISR QCD jet belongs to the signal or the control region, respectively. Also for W+jet production, the leading jets from ISR can populate the signal or the control region, depending on whether the charged lepton from the W decay is identified or not. As the structure of the ISR jets is the same for all Z+jet and W+jet background processes, we simulate only Z+jet production followed by a Z decay into neutrinos and use the corresponding jets for both the signal and the control region. These events are the basis for the discussion in section 5 where we show the applicability of the CWoLa method in a simplified setup.

As discussed above, the signal region also contains smaller admixtures from top-quark and diboson production. Fat jets (with a jet-radius $R = 0.8$, see section 4) in those processes might have a multi-prong structure since they can contain more than one subjet emerging from the decay products of a top quark or a vector boson. Since the CWoLa method tags those jets as different if they are not present in the control region as well, they have to be treated with some care as discussed in section 6. To keep things simple, we only simulate $t\bar{t}$ production in the semi-leptonic channel, where the transverse momentum of the leptonically decaying W-boson has to fulfill $p_T^{\text{recoil}} > 250 \text{ GeV}$. The leading jets from those $t\bar{t}$ events can populate the signal and the control region, depending on whether the charged lepton is identified or not. Concerning the diboson events, we simulate $W(\rightarrow \text{jets})Z(\rightarrow \nu\bar{\nu})$ events with $E_T^{\text{miss}} > 250 \text{ GeV}$ and $\Delta\phi(\mathbf{p}_T^{\text{jet}}, \mathbf{p}_T^{\text{miss}}) > 0.4$. Again, we assume that a more realistic composition of the diboson events in the signal and control regions would not qualitatively change the jet characteristics of the sample we are interested in. Other backgrounds are strongly suppressed and are thus ignored in this work for the sake of simplicity. The importance of potential additional reducible backgrounds (e.g. electrons being mis-measured as jets) can only be quantified by a proper experimental analysis. Addressing such backgrounds is thus beyond the scope of the present work.

2.2 New physics example: strongly interacting dark matter model

The CWoLa strategy outlined in section 3 and section 4 is completely model agnostic. However, to show the discovery potential, we employ a specific new physics model. Since we are sensitive to modified jet dynamics, we use the strongly interacting dark matter model introduced in ref. [35]. How the modified jet dynamics of this model can be used to improve a mono-jet analysis in a supervised setup has been discussed in ref. [32]. Other machine learning applications to strongly interacting dark matter models can be found in refs. [9, 43, 44]. For our choice of parameters, the model contains a heavy vector boson Z' with mass $m_{Z'} = 2 \text{ TeV}$ interacting with standard model quarks with a coupling $g_d = 0.1$. It couples the Standard Model to a sector of dark quarks q_d with a coupling $e_d = 0.4$.

These quarks are also charged under a dark $SU(3)$ gauge group with a confinement scale $\Lambda_d = 5 \text{ GeV}$. Hence, being pair-produced in a Z' decay they shower and hadronize to form dark pions π_d and dark rho mesons ρ_d with masses $m_{\pi_d} = m_{\rho_d} = 5 \text{ GeV}$. While the neutral rho mesons ρ_d^0 mix with the Z' and decay back to SM quarks promptly, the other mesons are stable dark-matter candidates and escape the detector. Hence, the jets in these models are usually called semi-visible jets. On average the invisible fraction of the jet energy amounts to $r_{\text{inv}} = 0.75$, leading to a specific modified jet structure. Furthermore, the dark jets differ from ordinary QCD jets due to the different running of the dark gauge coupling, the absence of heavy quarks in the shower and the presence of substructure from the decays of dark mesons. Because of the invisible fraction, the Z' decays into dark quarks do not lead to a resonance in the dijet invariant mass. Moreover, a certain fraction of jets turns out to be completely invisible and populates the signal region of our mono-jet search.

We use the UFO [45, 46] implementation of the model to simulate the pair-production of dark quarks starting with Madgraph and using the same tool-chain as discussed in section 2.1. The hidden-valley module of Pythia [47] is used to handle the dark showering, hadronisation and the decay of the ρ_d^0 mesons into SM quarks. The subsequent showering and hadronisation of these quarks as well as the detector simulation and the event selection for our signal region are performed as for the SM backgrounds. The same model parameters and the same tool chain have been also used to produce the Aachen benchmark data set introduced in ref. [48].

3 The CWoLa method

Classification Without Labels (CWoLa) is based on the typical setup in high energy physics experiments: one defines a signal region (SR) and a control region (CR). The signal region is chosen to optimize the fraction $f^{\text{SR}} = N_A^{\text{SR}}/N_B^{\text{SR}}$ for a certain class of models, where N_A^{SR} is the number of new physics or anomalous events/data instances and N_B^{SR} is the number of background events/data instances. The control region should be free of anomalous events or should at least contain a smaller fraction $f^{\text{CR}} = N_A^{\text{CR}}/N_B^{\text{CR}}$ of anomalous events, i.e. $f^{\text{CR}} < f^{\text{SR}}$. The expected number of background events N_B^{SR} measured in the signal region is assumed to be known up to a certain relative error σ which includes statistical as well as systematic uncertainties. The control region is used to minimize the systematic uncertainty using data driven methods. For $f^{\text{SR}} > \sigma$ the measurement is sensitive, e.g. for $f^{\text{SR}} > 5\sigma$ one would expect a discovery.

There are features in each event that are used for the definition of the signal and the background region. CWoLa is based on the crucial assumption that there is an additional set of features such that background events from the signal and the control region are indistinguishable using only those features, i.e. the events in the signal and the control region are drawn from an identical probability distribution concerning this restricted feature space. These features will be called CWoLa features in the following.

In the CWoLa setup, a standard supervised binary classifier is trained to tag each event as belonging to the signal or the control region, using the CWoLa features as input. These labels are available for real experimental data. No reference is made to labeling events as

background or new physics events, which are the labels one would actually like to know, hence the name Classification Without Labels.

In the absence of new physics events and taking the above assumption for granted, the classifier has to fail because the task is impossible. The predicted labels cannot be better than random guessing. However, if there are anomalous events which are drawn from a different probability distribution and $f^{\text{CR}} < f^{\text{SR}}$, the classifier should assign a higher score for those anomalous events to belong to the signal region. Thus, for a given score threshold, the classifier will predominantly select anomalous events.

It is worth stressing the model-agnostic nature of this approach. Searching for a specific new physics model, a difference between the new physics and the background events in some feature will already be used to define the signal region to boost the sensitivity of a simple cut-and-count analysis. However, in this case one has sensitivity to this feature only. CWoLa instead is sensitive to all potential differences in the CWoLa features.

It has been shown that this setup leads to an optimal classifier [18], i.e. it can have the same performance as a supervised classifier working with events labeled as being anomalous or belonging to the background. However, this finding does not imply that the method is guaranteed to work in practice.

CWoLa has been shown to improve the sensitivity of bump-hunt searches [20]. In this context, the control region consists of the side bands of a specified signal region. For concreteness, consider a resonance search using dijet events. An interval in the dijet invariant mass is selected as signal region. It would be populated by new physics events if there was a resonance with a mass in the signal region decaying into a dijet final state. If the structure of new physics jets differs from QCD jets, the CWoLa method can boost the sensitivity of the search by only taking into account events with a classifier score beyond a certain threshold. However, the CWoLa features for the classifier have to be chosen with care since they might be correlated with the dijet mass which defines the signal region. Otherwise the assumption that background jets from the signal and the control region are indistinguishable using the CWoLa features is violated. The corresponding decorrelation of observables has been discussed in ref. [24] and methods to improve on those limitations for bump hunts have also been suggested [22].

4 CWoLa for anomalous mono-jets

To improve an existing or a future mono-jet search, we suggest to use the CWoLa setup in the following way: the signal region consists of events with large missing transverse energy. As discussed in section 2, we search for an excess of events in a signal region defined by $E_T^{\text{miss}} > 250 \text{ GeV}$ and $\Delta\phi(\mathbf{p}_T^{\text{jet}}, \mathbf{p}_T^{\text{miss}}) > 0.4$ for all jets as a concrete example. The control region is defined by events where neutrinos are replaced by charged leptons (see also section 2). Data in the control regions has been recorded in past experimental analyses for most of those backgrounds to control systematics. Details can be also found in ref. [33].

As CWoLa features to the classifier, low level information of the leading fat jet in each event is used. To find the leading fat jet, the constituents of the events in the signal and background region are reclustered into anti- k_T jets with a jet radius $R = 0.8$. In our

simulation based studies, we use the 40 jet constituents with largest transverse momentum in the constituents branch of the Delphes output (see section 2 for the simulation details). The assumption that the CWoLa features are uncorrelated with respect to the definition of signal and background regions, i.e. the leptonic decays of weak bosons, is physically sound: the evolution of the jets from initial-state radiation is driven by QCD dynamics and not influenced by the decay properties of a leptonically decaying weak boson which recoils against the jet. We have verified the independence of the jet structure with respect to the weak-boson decays in our simulation and simplify the simulation accordingly as detailed in section 2. Replacing simulated by experimentally recorded data should be straightforward.

As the binary classifier, we use a Dynamic Graph Convolutional Neural Network (DGCNN) [49] which is based on the ParticleNet architecture [50] and has proven to be an extremely powerful jet tagger [51]. Architecture, preprocessing, and training are discussed in appendix A. Note that instead any powerful supervised classification algorithm could be used, e.g. the recently proposed graph-based LundNet [52] or LorentzNet [53].

The score $s \in [0, 1]$ of the trained classifier for each event can be interpreted as the probability of the event to belong to the signal region (see appendix A for details). For few anomalous events, the output distribution for background events is expected to peak close to $s = 0.5$. As discussed in section 3, anomalous events are expected to be tagged as belonging to the signal region.

Given the classifier score s for each event, we choose a threshold t . For $s > t$, the event is selected as being potentially anomalous. The threshold t is chosen such that one permille of the events in the control region is selected, i.e. $n^{\text{CR}} = \epsilon^{\text{CR}} N^{\text{CR}}$ with $\epsilon^{\text{CR}} = 0.001$. We will comment on this choice below. Here we always assume that there is only background in the control region. Hence, we work with a background rejection $1/\epsilon_{\text{B}}^{\text{CR}} = 1/\epsilon^{\text{CR}} = 1000$. For a given $1/\epsilon_{\text{B}}^{\text{CR}}$, corresponding to a specific threshold t , the classifier selects a number of anomalous events $n_{\text{A}}^{\text{SR}} = \epsilon_{\text{S}}^{\text{SR}} N_{\text{A}}^{\text{SR}}$ in the signal region, where $\epsilon_{\text{S}}^{\text{SR}}$ is the signal efficiency and depends on t . In our weakly supervised setup, $\epsilon_{\text{S}}^{\text{SR}}$ is unknown. In analogy, the classifier selects a certain number of background events from the signal region $n_{\text{B}}^{\text{SR}} = \epsilon_{\text{B}}^{\text{SR}} N_{\text{B}}^{\text{SR}}$. If the CWoLa assumptions hold, we have $\epsilon_{\text{B}}^{\text{SR}} = \epsilon_{\text{B}}^{\text{CR}} = 0.001$.

If there are no anomalies in the signal region, the fraction of selected events from the signal region will be identical to that of the control region, i.e. the expected number of selected events is $n_{\text{exp}}^{\text{SR}} = \epsilon_{\text{B}}^{\text{CR}} N^{\text{SR}}$. This is the Null hypothesis. If $n_{\text{exp}}^{\text{SR}}$ and n^{SR} differ only as expected from statistical fluctuations, no indication of new physics is observed. If there are anomalous events in the signal region and the classifier is successful in identifying them with some $\epsilon_{\text{S}}^{\text{SR}} > \epsilon_{\text{B}}^{\text{SR}}$, the fraction of selected events from the signal region will be larger. If it exceeds the expectation for statistical fluctuations the Null hypothesis can be excluded in the usual way. Due to the model agnostic nature of the method, there are no exclusion limits to be derived for any models. The CWoLa method is a discovery tool.

The choice for the background rejection $1/\epsilon_{\text{B}}^{\text{CR}}$ is to a certain extent arbitrary. Our choice $1/\epsilon_{\text{B}}^{\text{CR}} = 1000$ is driven by the following considerations: although $\epsilon_{\text{S}}^{\text{SR}}$ is unknown, the signal-to-background ratio $n_{\text{A}}^{\text{SR}}/n_{\text{B}}^{\text{SR}}$ is usually a monotonically growing function of the background rejection $1/\epsilon_{\text{B}}^{\text{CR}}$ which favours to choose $1/\epsilon_{\text{B}}^{\text{CR}}$ large. In particular, the signal-to-background ratio for a discovery should not be too small such that one is not too

sensitive to small unknown systematics concerning the validity of the CWoLa assumptions. On the other hand, for increasing $1/\epsilon_B^{\text{CR}}$, a classifier is often not performant enough to not only improve the signal-to-noise ratio but also the significance $n_A^{\text{SR}}/\sqrt{n_{\text{exp}}^{\text{SR}}}$ of an observed excess. Hence, the minimal acceptable signal-to-background ratio for a discovery is a good guide for $1/\epsilon_B^{\text{CR}}$.

As a default, we use one million events in both the signal and the control region which is close to the actual number of observed events in the IM1 signal region of the most recent ATLAS search [33]. For $\epsilon_B^{\text{CR}} = 0.001$, this corresponds to $n^{\text{CR}} = 1000$ selected events. Although n^{CR} is fixed by choosing the threshold t for our data set, n^{CR} for the same t is distributed with the usual statistical uncertainty $\sqrt{n^{\text{CR}}}$ for independent data sets. Therefore the corresponding relative statistical uncertainty due to the limited size of the control region for n^{SR} is roughly $\sigma^{\text{CR}} = 3\%$. Having a signal sample of the same size ($n_{\text{exp}}^{\text{SR}} = n^{\text{CR}}$) and adding the corresponding relative statistical uncertainty σ^{SR} in quadrature, a 5σ discovery needs at least $5\sqrt{(\sigma^{\text{CR}})^2 + (\sigma^{\text{SR}})^2} n_{\text{exp}}^{\text{SR}} = 5\sqrt{2}\sigma^{\text{CR}} n_{\text{exp}}^{\text{SR}} = 5\sqrt{2}n_{\text{exp}}^{\text{SR}} \sim 224$ additional events.

In principle, one could also scan the background rejection. However, here we do not follow this idea in order to avoid discussions about the look-elsewhere effect. The chosen value has not been tuned to the success of finding our example signal. Moreover, as we will see in the following sections, a discovery using our CWoLa setup will most likely be an iterative process which is driven not so much by statistical considerations but by iterated efforts to understand the quality of the control region.

In this setup we are sensitive to modified jet structures which occur in events with large missing transverse energy. A physically well motivated example for such a model is discussed in section 2.2 and used in the following sections to demonstrate the sensitivity of the method.

5 Proof of principle

In this section, we take the CWoLa assumptions for granted. The simulated background events in the signal and the control region are indeed drawn from the same probability distribution since we use the same simulation setup to generate them. We only use Z+jet events which are simulated as discussed in section 2.1. Therefore, we investigate the performance of the CWoLa setup under ideal circumstances. The problems which might arise for more complicated samples of background jets or in defining a suitable control sample are discussed in section 6.

As a default, we use $N^{\text{CR}} = N^{\text{SR}} = 10^6$ events in the control region and in the signal region. We use a fraction $f^{\text{CR}} = 0$ of anomalous events in the control region, i.e. we have $N_B^{\text{CR}} = N^{\text{CR}}$ and only Z+jet events. In the signal region, we have $N_B^{\text{SR}} = (1 - f^{\text{SR}})N^{\text{SR}}$ Z+jet events and $N_A^{\text{SR}} = f^{\text{SR}}N^{\text{SR}}$ new physics events, simulated as detailed in section 2.2. We take the number of new physics events N_A^{SR} as a free parameter. Note that not all new physics events have anomalous leading fat jets since also initial-state radiation QCD jets can be leading. This is making the task to identify new physics even harder.

f^{SR}	$n_{\text{exp}}^{\text{SR}}$	n^{SR}	n_{A}^{SR}	n_{B}^{SR}	$(n^{\text{SR}} - n_{\text{exp}}^{\text{SR}}) / \sqrt{2 n_{\text{exp}}^{\text{SR}}}$
0%	1000	1048	0	1048	1.07
0.2%	1000	1065	47	1018	1.45
0.4%	1000	1107	100	1007	2.39
0.5%	1000	1175	184	991	3.91
0.6%	1000	1306	247	1059	6.84
0.7%	1000	1389	367	1022	8.70
0.8%	1000	1500	419	1081	11.18
1%	1000	1666	625	1041	14.89
2%	1000	2357	1392	965	30.34
4%	1000	4182	3269	913	71.15

Table 1. Number of events n^{SR} selected from the signal region for $N^{\text{CR}} = N^{\text{SR}} = 10^6$ and several signal fractions f^{SR} . We have used the mean score of five classifiers trained on the same data. We also show the number of events expected to be selected in the absence of a new physics signal, $n_{\text{exp}}^{\text{SR}}$, and the number of selected anomalous and background events (n_{A}^{SR} and n_{B}^{SR}). The latter two numbers are not known for real data. The last column shows an estimate of the statistical significance of a possible discovery.

Using $\epsilon_{\text{B}}^{\text{CR}} = 0.001$, as discussed in section 4, the number of events $n^{\text{SR}} = n_{\text{A}}^{\text{SR}} + n_{\text{B}}^{\text{SR}}$ selected by the classifier in the signal region is shown in table 1 for several signal fractions f^{SR} . To be less vulnerable to fluctuations in the training process we train five classifiers on the same data and average their scores. Most importantly, the Null test works fine. If there are no anomalous events in the data ($f^{\text{SR}} = 0$), the selected number is, within statistical fluctuations, in agreement with the expected value $n_{\text{exp}}^{\text{SR}} = 1000$. CWoLa does not provide any false indication for new physics. Hence, overfitting is no issue. Moreover, a signal rate $f^{\text{SR}} = 1\%$, which is still consistent with constraints from the latest ATLAS mono-jet search, leads to $n^{\text{SR}} = 1666$. Without statistical doubts, such a finding would indicate that there is something to be understood about the data. Ideally, a thorough investigation of the selected jets will uncover the unexpected jet structure and hint towards a suitable new physics model.

Around a signal fraction $f^{\text{SR}} = 0.6\%$, the statistical significance rapidly drops below 5σ . Even under the ideal circumstances which are assumed in this section, the classifier is then not able to identify the new physics events as anomalous. Note that only $N_{\text{A}}^{\text{SR}} = 6000$ anomalous events are in the training sample with $N^{\text{SR}} + N^{\text{CR}} = 2 \cdot 10^6$ events. It is extremely challenging for a classifier to efficiently learn the anomalous structures under these circumstances. In a supervised setup, the data instances would be weighted according to their abundance to help the classifier. In our weakly supervised setup, however, this is not possible.

Moreover, our studies show that the absolute number of anomalous events is an essential parameter. If the number of background events is increased for a fixed number of

N^{SR}	$n_{\text{exp}}^{\text{SR}}$	n^{SR}	n_{A}^{SR}	n_{B}^{SR}	$(n^{\text{SR}} - n_{\text{epx}}^{\text{SR}})/\sqrt{2n_{\text{epx}}^{\text{SR}}}$
250k	250	916	686	230	29.78
500k	500	1113	660	453	19.38
1000k	1000	1666	625	1041	14.89

Table 2. Same as table 1 but for a fixed number of anomalous events $N_{\text{A}}^{\text{SR}} = 10\text{k}$ in a signal region with N^{SR} events.

N^{SR}	$n_{\text{exp}}^{\text{SR}}$	n^{SR}	n_{A}^{SR}	n_{B}^{SR}	$(n^{\text{SR}} - n_{\text{epx}}^{\text{SR}})/\sqrt{2n_{\text{epx}}^{\text{SR}}}$
250k	250	362	126	236	5.01
500k	500	680	258	422	5.69
1000k	1000	1666	625	1041	14.89

Table 3. Same as table 1 but for a fixed fraction $f^{\text{SR}} = 0.01$ of anomalous events in a signal region with N^{SR} events.

anomalous events in the signal region, the performance of the CWoLa method is relatively stable, although the signal fraction is decreasing. In table 2, we show the tagged number of events from the signal region for several N^{SR} , with the number of anomalous events fixed at $N_{\text{A}}^{\text{SR}} = 10\text{k}$. Moreover, table 3 shows the improving performance for a fixed signal fraction $f^{\text{SR}} = 0.01$ if the collected data increase. This is good news for the high-luminosity phase of the LHC. Whether improved training strategies or more powerful classifiers can improve the overall performance is left for future research.

We have also investigated the effect of the control region being smaller than the signal region; in the analysis we have considered as an example, the smaller branching ratio of the Z boson into charged leptons leads to a smaller number of events in the control sample. Weighting the events accordingly, we do not observe a significant loss of discovery power beyond the increased statistical error.

6 Reality checks

In section 5, we have shown the sensitivity of the CWoLa method under idealized conditions. In this section, we study how a non-trivial and more realistic composition of the signal and control samples impacts the performance. Here, we assume that the background events in the signal region contain $r_{t\bar{t}}^{\text{SR}} = 3.5\%$ $t\bar{t}$ events and $r_{VV}^{\text{SR}} = 2\%$ diboson events (see section 2.1). In particular, we investigate how well this composition has to be understood and reflected in the control region.

In the absence of any new physics events ($f^{\text{CR}} = f^{\text{SR}} = 0$), we show the results of the CWoLa tagging for several values of $r_{t\bar{t}}^{\text{CR}}$ and r_{VV}^{CR} in table 4. For $r_{t\bar{t}}^{\text{CR}} = r_{VV}^{\text{CR}} = 0$, i.e. the additional backgrounds are absent from the control region, the CWoLa method correctly identifies the different jet structures from the top-quark and weak-boson decays in the signal region. However, this is of course not a sign for new physics but a consequence of our poor modeling of the control region. Although this is a naive example, it highlights

$r_{t\bar{t}}^{\text{CR}}$	r_{VV}^{CR}	$n_{Z+\text{jet}}^{\text{CR}}$	$n_{t\bar{t}}^{\text{CR}}$	n_{VV}^{CR}	$n_{Z+\text{jet}}^{\text{SR}}$	$n_{t\bar{t}}^{\text{SR}}$	n_{VV}^{SR}	n^{SR}
0%	0%	1000	0	0	1053	1127	2043	4223
1.75%	1.0%	541	247	212	581	557	420	1558
2.80%	1.6%	980	3	17	1065	6	17	1088
3.15%	1.8%	957	24	19	962	26	32	1020
3.50%	2.0%	823	160	17	793	175	21	989
5.00%	3.0%	966	18	16	960	22	3	985
3.50%	1.6%	903	89	8	821	93	26	940
2.80%	2.0%	983	5	12	996	7	12	1015

Table 4. Number of events selected by the classifier from the different background classes in the signal and control region for different compositions of the control region. There are no anomalous events ($f^{\text{SR}} = 0$) and we have $r_{t\bar{t}}^{\text{SR}} = 3.5\%$ and $r_{VV}^{\text{SR}} = 2\%$.

the challenge of the CWoLa approach: tagging more signal-region events than expected can always either be due to new physics or due to a mismodeling of the control region. For real data, the leading jets for the control region have to be measured from different event topologies and then combined to form a proper control region that matches the expected rates of the different backgrounds in the signal region, where input from theory and Monte Carlo is required. Therefore, it is a relevant question to which extent the control region needs to be understood. The results in table 4 show, that the understanding and the inclusion of the backgrounds are crucial at the percent level, as it is already the case for the standard searches. However, variations of backgrounds between the control and signal region at the level of a few permille (i.e. a relative understanding at the level of 10%) are tolerable. Hence, one does not need to be perfect. In particular, the CWoLa setup is not too sensitive to overestimating small background components. Other small backgrounds with a different jet structure at the level of a permille or below might be even tolerable in the signal region without including them in the control region at all.

Moreover, in a certain sense, we propose a self-correcting setup. Tagging more events from the signal region than expected in real data would first of all prompt more efforts to better understand the control region. The tagged events might also help in understanding which backgrounds are mismodeled. Only after all those studies one would want to pursue an interpretation of the selected signal events in terms of new physics, also aided by the investigation of the tagged jet’s structure.

In table 5 we again study if the exemplary semivisible jets can be tagged, using a more realistic background sample in the signal as well as in the control region. We see that excesses in n^{SR} over the expected number of observations are dominated by these semivisible jets, even for slightly mismodeled control regions. Completely neglecting additional backgrounds in the control region, however, results in an excess dominated by these backgrounds and only minor enhancement of the fraction of signal jets.

$r_{t\bar{t}}^{\text{CR}}$	r_{VV}^{CR}	$n_{Z+\text{jet}}^{\text{CR}}$	$n_{t\bar{t}}^{\text{CR}}$	n_{VV}^{CR}	$n_{Z+\text{jet}}^{\text{SR}}$	$n_{t\bar{t}}^{\text{SR}}$	n_{VV}^{SR}	n_A^{SR}	n^{SR}
0%	0%	1000	0	0	1089	1245	1826	223	4383
2.80%	1.6%	876	108	16	838	134	37	456	1465
3.50%	2.0%	963	23	14	996	25	32	633	1686
5.00%	3.0%	971	12	17	1034	8	10	575	1627

Table 5. Number of events selected by the classifier from the different classes in the signal and control region for different compositions of the control region. There is a fixed fraction of anomalous events ($f^{\text{SR}} = 0.01$) and we have $r_{t\bar{t}}^{\text{SR}} = 3.5\%$ and $r_{VV}^{\text{SR}} = 2\%$.

Throughout our studies, we have assumed that there are no anomalous new physics events in the control region ($f^{\text{CR}} = 0$). This is a valid assumption if the new physics model does not predict events with isolated leptons. For the strongly interacting dark matter model under consideration, this is the case as the dark mesons have no leptonic decay modes. However, in general new physics models might predict $f^{\text{CR}} \neq 0$. Increasing f^{CR} has the same effect as increasing $r_{t\bar{t}}^{\text{CR}}$ and r_{VV}^{CR} in table 4, where one could view top pair and diboson events as a potential new physics signal. As f^{CR} approaches f^{SR} the method fails. In contrast, small admixtures of new physics events in the control region ($f^{\text{CR}} \ll f^{\text{SR}}$) do not qualitatively change our conclusions.

In appendix B, we further discuss the classifier score using the Monte Carlo labels of the events which are not available for real data.

7 Conclusion

We have demonstrated that the Classification Without Labels (CWoLa) method [18] is a powerful tool to boost the LHC discovery potential for new physics models with anomalous jet dynamics and a mono-jet signature. There are no model-specific assumptions, and the proposed setup can be implemented directly using collected data. The CWoLa method relies on a background-dominated control sample, which is in general available for LHC searches.

Using Monte Carlo simulation and a well-motivated specific new physics model with a strongly interacting hidden sector [35], we show that less than 1% of new physics events in inclusive signal regions would be sufficient to discover signs of new physics. In contrast, the corresponding traditional search [33] is not even sensitive enough to exclude the model at 95% confidence level (the systematics dominated error on the SM prediction for signal region IM1 is 1.2%). We consider the hidden sector model to be a rather challenging test case since the modified jet dynamics of the model is difficult to recognize by unsupervised methods [48] and by dedicated supervised taggers [32].

The CWoLa setup, as a weakly supervised method, is not as sensitive as a dedicated model-specific and Monte Carlo-based approach using supervised methods. Moreover, as a discovery tool, it will not provide exclusion limits for specific model parameters. On the other hand, we emphasise again that the CWoLa method is model agnostic and can be

applied directly to data. We have also demonstrated that the CWoLa method provides a useful data-driven tool to improve the understanding of the signal and control regions.

We have shown that the CWoLa discovery potential is increasing with more data, while traditional searches might already be limited by systematic uncertainties. The method should therefore continue to gain importance with, in particular, the large amounts of data expected in the high-luminosity phase of the LHC. Given that the CWoLa method is model-agnostic and data-driven and that it only requires a background-dominated control sample, we consider CWoLa-assisted searches for new physics and reanalyses of data already collected by the LHC experimental collaborations to be very promising.

Acknowledgments

We would like to thank Elias Bernreuther for discussions and comments on the manuscript. TF is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant 400140256 - GRK 2497: the physics of the heaviest particles at the Large Hadron Collider. The research of MK and AM is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant 396021762 - TRR 257: particle Physics Phenomenology after the Higgs Discovery. The authors gratefully acknowledge the computing time granted by the NHR4CES Resource Allocation Board and provided on the supercomputer CLAIX at RWTH Aachen University as part of the NHR4CES infrastructure. The calculations for this research were conducted with computing resources under the project rwth0934.

A The DGCNN classifier

As input for our Dynamic Graph Convolutional Neural Network (DGCNN) we use the 40 leading p_T constituents of each jet, zero padded when needed. From the 4-momentum of each jet constituent we construct the set of seven input features $\{\Delta\eta, \Delta\phi, \log(p_T), \log(p_T/p_T^{\text{jet}}), \log(E), \log(E/E^{\text{jet}}), \Delta R\}$ with $\Delta\eta = \eta - \eta^{\text{jet}}$, $\Delta\phi = \phi - \phi^{\text{jet}}$ and $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$; η , ϕ , p_T and E refer to the rapidity, the azimuthal angle, the transverse momentum (in GeV), and the energy (in GeV) of the constituent, respectively. The quantities with superscript “jet” refer to the respective characteristics of the jet.

The DGCNN constructs a k -nearest-neighbors (knn) graph with these particles as nodes. We use $k = 16$. The initial graph is constructed using the Euclidean distances of the particles in $\Delta\eta$ and $\Delta\phi$. The network’s architecture is almost identical to the one used in ref. [32] for supervised classification on similar data sets, i.e. we use three EdgeConv blocks with three convolutions each. The number of features is increased successively from the seven input features to 64, 128 and finally 256. The graph is dynamically updated as knn graph after the first and second block using the Euclidean distances between nodes/particles based on all features. The outputs of each block are concatenated with the input resulting in 455 features per particle. After global average pooling the 455 features are fed into a fully connected network with 256, 128 and 2 nodes. We regularize the fully connected network using dropout with a fraction of 0.1 after the first two layers. We apply softmax

activation to the last layer, allowing for a probability interpretation of the output. The two outputs then correspond to the probability of the input to belong to the control region or the signal region. We use the probability for the signal region as score s (see section 4). The only difference to ref. [32] is that we use Leaky ReLU with $\alpha = 0.1$ instead of ReLU as activation function, since we observed that it leads to more stable results.

We implement the network using TensorFlow 2.6.0 [54] and the build-in version of Keras [55]. We use the Adam optimizer [56] with its default settings to minimize the categorical cross entropy. During training we reduce the learning rate by a factor of 0.1 when the training loss does not improve for 8 epochs. If the loss still does not improve for an additional four epochs, we stop the training. We set the maximum number of training epochs to 75, which is hardly ever reached. Note that we evaluate the network on the same data that we use to train. This corresponds to the procedure one would use on experimental data. The results in section 5 show that overfitting is not a problem for such a large dataset, as we see no excess in selected events in the signal region in the absence of signal (see table 1).

B Classifier output

In this appendix we show and analyze the classifier score for the training sample. Here, we make use of the Monte Carlo labels for anomalous/new physics events and background events of the different types. Hence, this information is not available for data collected at the LHC. However, it is nevertheless instructive to investigate it.

In figure 1, we show the signal score s for the simplified setup used in section 5. The scores peak at $s = 0.5$, as this minimizes the loss function if all jets in the control and the signal region are drawn from identical probability distributions. For a well-defined control region there are no jets which are present only in the control region but not in the signal region, such that scores well below $s = 0.5$ would be a sign of overfitting and should not be observed. In the signal region, we distinguish jets from Z+jet background events and semivisible jets from new physics events. For a signal fraction $f^{\text{SR}} = 1\%$ (and $f^{\text{CR}} = 0$), there are enough semivisible jets such that the training is successful in at least identifying the most anomalous jets. Most jets from new physics events are not identified, but this is not a problem for the CWoLa method to work. For $f^{\text{SR}} = 0.5\%$, those anomalous jets are still there, but the training procedure is not able to distinguish them in an efficient way. In particular the range of classifier scores is significantly reduced to $s < 0.55$ ($s < 0.8$ for $f^{\text{SR}} = 1\%$).

In figure 2 and figure 3, we show the signal score s including subdominant backgrounds as discussed in section 6. The signal fraction is always fixed at $f^{\text{SR}} = 1\%$ (and $f^{\text{CR}} = 0$). We use the nominal fractions $r_{t\bar{t}}^{\text{SR}} = 3.5\%$ and $r_{VV}^{\text{SR}} = 2\%$ for the subdominant backgrounds in the signal region and vary $r_{t\bar{t}}^{\text{CR}}$ and r_{VV}^{CR} in the control region. If the mismodelling is too severe (right plot in figure 2), the classifier mainly tags $t\bar{t}$ and diboson events as signal-like, as expected and also shown in table 5. If the mismodelling is reduced (right plot in figure 3) the semi-visible events dominate at large scores as it should be. This is even more so the case for perfect modelling (left plot in figure 2). The left plot in figure 3 shows that slightly overestimating distinct (multi-prong) jets in the control region is less dangerous to

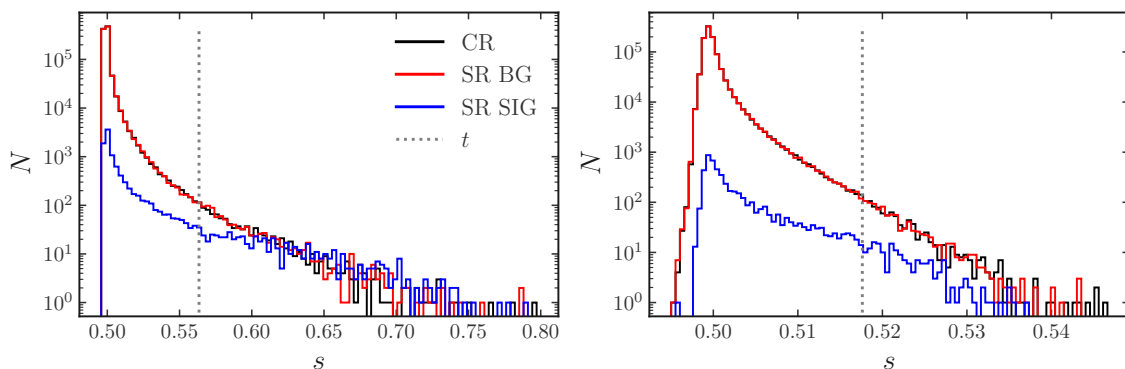


Figure 1. The absolute number of events in the signal region (Z+jet events in red, semivisible jets in blue) and the control region (only Z+jet events in black) is shown as a function of the classifier score $s \in [0, 1]$ for $f^{\text{SR}} = 1\%$ (left) and $f^{\text{SR}} = 0.5\%$ (right). Note the different scale on the x -axis in the two plots. We also show the threshold value t for $\epsilon^{\text{CR}} = 0.001$ as a vertical line.

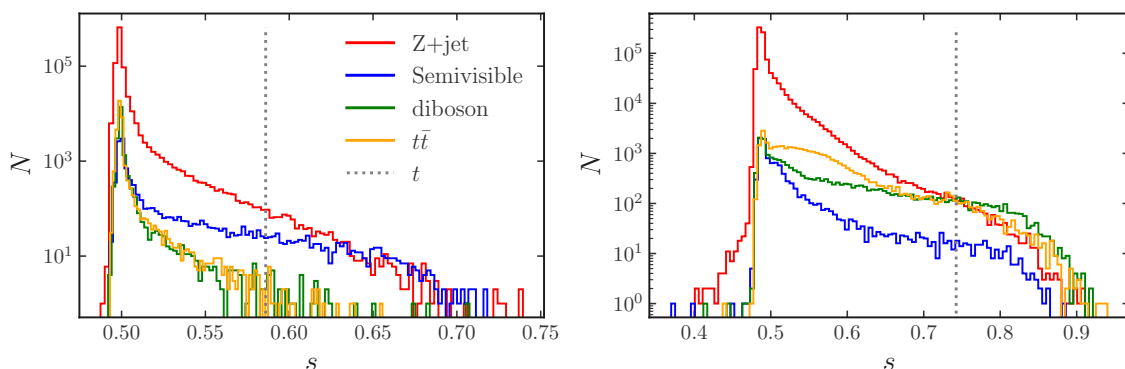


Figure 2. The absolute number of events in the signal region (Z+jet events in red, semivisible jets in blue, $t\bar{t}$ events in orange, diboson events in green) is shown as a function of the classifier score $s \in [0, 1]$. We have $r_{t\bar{t}}^{\text{SR}} = 3.5\%$ and $r_{VV}^{\text{SR}} = 2.0\%$ and use $r_{t\bar{t}}^{\text{CR}} = r_{t\bar{t}}^{\text{SR}}$ and $r_{VV}^{\text{CR}} = r_{VV}^{\text{SR}}$ (left) or $r_{t\bar{t}}^{\text{CR}} = r_{VV}^{\text{CR}} = 0$ (right). We also show the threshold value t for $\epsilon^{\text{CR}} = 0.001$ as a vertical line.

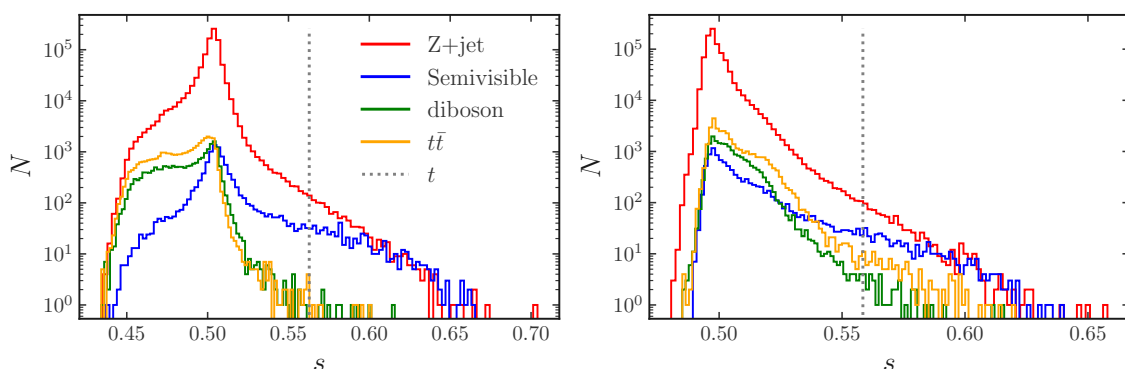


Figure 3. Same as figure 2 but for overestimated (left) additional backgrounds in the control region ($r_{t\bar{t}}^{\text{CR}} = 5.0\%$ and $r_{VV}^{\text{CR}} = 3.0\%$) and underestimated (right) additional backgrounds ($r_{t\bar{t}}^{\text{CR}} = 2.8\%$ and $r_{VV}^{\text{CR}} = 1.6\%$). We also show the threshold value t for $\epsilon^{\text{CR}} = 0.001$ as a vertical line.

the CWoLa method than an underestimation. In this case, these jets are more abundant in the (not perfectly modeled) control region leading to scores $s < 0.5$.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited. SCOAP³ supports the goals of the International Year of Basic Sciences for Sustainable Development.

References

- [1] 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** (2012) 1 [[arXiv:1207.7214](https://arxiv.org/abs/1207.7214)] [[INSPIRE](#)].
- [2] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [[arXiv:1207.7235](https://arxiv.org/abs/1207.7235)] [[INSPIRE](#)].
- [3] A. De Simone and T. Jacques, *Guiding new physics searches with unsupervised learning*, *Eur. Phys. J. C* **79** (2019) 289 [[arXiv:1807.06038](https://arxiv.org/abs/1807.06038)] [[INSPIRE](#)].
- [4] J. Hajer, Y.-Y. Li, T. Liu and H. Wang, *Novelty detection meets collider physics*, *Phys. Rev. D* **101** (2020) 076015 [[arXiv:1807.10261](https://arxiv.org/abs/1807.10261)] [[INSPIRE](#)].
- [5] T. Heimel, G. Kasieczka, T. Plehn and J.M. Thompson, *QCD or what?*, *SciPost Phys.* **6** (2019) 030 [[arXiv:1808.08979](https://arxiv.org/abs/1808.08979)] [[INSPIRE](#)].
- [6] M. Farina, Y. Nakai and D. Shih, *Searching for new physics with deep autoencoders*, *Phys. Rev. D* **101** (2020) 075021 [[arXiv:1808.08992](https://arxiv.org/abs/1808.08992)] [[INSPIRE](#)].
- [7] 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](https://arxiv.org/abs/2005.01598)] [[INSPIRE](#)].
- [8] 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](https://arxiv.org/abs/2105.07988)] [[INSPIRE](#)].
- [9] F. Canelli, A. de Cosa, L.L. Pottier, J. Niedziela, K. Pedro and M. Pierini, *Autoencoders for semivisible jet detection*, *JHEP* **02** (2022) 074 [[arXiv:2112.02864](https://arxiv.org/abs/2112.02864)] [[INSPIRE](#)].
- [10] J.A. Aguilar-Saavedra, J.H. Collins and R.K. Mishra, *A generic anti-QCD jet tagger*, *JHEP* **11** (2017) 163 [[arXiv:1709.01087](https://arxiv.org/abs/1709.01087)] [[INSPIRE](#)].
- [11] R.T. D’Agnolo and A. Wulzer, *Learning new physics from a machine*, *Phys. Rev. D* **99** (2019) 015014 [[arXiv:1806.02350](https://arxiv.org/abs/1806.02350)] [[INSPIRE](#)].
- [12] 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](https://arxiv.org/abs/1912.12155)] [[INSPIRE](#)].
- [13] 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](https://arxiv.org/abs/2002.12376)] [[INSPIRE](#)].
- [14] J.A. Aguilar-Saavedra, F.R. Joaquim and J.F. Seabra, *Mass Unspecific Supervised Tagging (MUST) for boosted jets*, *JHEP* **03** (2021) 012 [Erratum *ibid.* **04** (2021) 133] [[arXiv:2008.12792](https://arxiv.org/abs/2008.12792)] [[INSPIRE](#)].

- [15] C.K. Khosa, V. Sanz and M. Soughton, *A simple guide from machine learning outputs to statistical criteria*, [arXiv:2203.03669](#) [INSPIRE].
- [16] V. Mikuni, B. Nachman and D. Shih, *Online-compatible unsupervised nonresonant anomaly detection*, *Phys. Rev. D* **105** (2022) 055006 [[arXiv:2111.06417](#)] [INSPIRE].
- [17] L.M. Dery, B. Nachman, F. Rubbo and A. Schwartzman, *Weakly supervised classification in high energy physics*, *JHEP* **05** (2017) 145 [[arXiv:1702.00414](#)] [INSPIRE].
- [18] E.M. Metodiev, B. Nachman and J. Thaler, *Classification without labels: learning from mixed samples in high energy physics*, *JHEP* **10** (2017) 174 [[arXiv:1708.02949](#)] [INSPIRE].
- [19] P.T. Komiske, E.M. Metodiev, B. Nachman and M.D. Schwartz, *Learning to classify from impure samples with high-dimensional data*, *Phys. Rev. D* **98** (2018) 011502 [[arXiv:1801.10158](#)] [INSPIRE].
- [20] 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].
- [21] 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].
- [22] B. Nachman and D. Shih, *Anomaly detection with density estimation*, *Phys. Rev. D* **101** (2020) 075042 [[arXiv:2001.04990](#)] [INSPIRE].
- [23] A. Andreassen, B. Nachman and D. Shih, *Simulation assisted likelihood-free anomaly detection*, *Phys. Rev. D* **101** (2020) 095004 [[arXiv:2001.05001](#)] [INSPIRE].
- [24] 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].
- [25] S.-E. Dahbi et al., *Machine learning approach for the search of resonances with topological features at the Large Hadron Collider*, *Int. J. Mod. Phys. A* **37** (2022) 2150241 [[arXiv:2011.09863](#)] [INSPIRE].
- [26] J.S.H. Lee, S.M. Lee, Y. Lee, I. Park, I.J. Watson and S. Yang, *Quark gluon jet discrimination with weakly supervised learning*, *J. Korean Phys. Soc.* **75** (2019) 652 [[arXiv:2012.02540](#)] [INSPIRE].
- [27] A. Hallin et al., *Classifying Anomalies THrough Outer Density Estimation (CATHODE)*, [arXiv:2109.00546](#) [INSPIRE].
- [28] J.A. Raine, S. Klein, D. Sengupta and T. Golling, *CURTAINS for your sliding window: constructing unobserved regions by transforming adjacent intervals*, [arXiv:2203.09470](#) [INSPIRE].
- [29] 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].
- [30] 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].
- [31] 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](#)] [INSPIRE].
- [32] E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer and A. Mück, *Casting a graph net to catch dark showers*, *SciPost Phys.* **10** (2021) 046 [[arXiv:2006.08639](#)] [INSPIRE].

- [33] ATLAS collaboration, *Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **103** (2021) 112006 [[arXiv:2102.10874](#)] [[INSPIRE](#)].
- [34] CMS collaboration, *Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **11** (2021) 153 [[arXiv:2107.13021](#)] [[INSPIRE](#)].
- [35] E. Bernreuther, F. Kahlhoefer, M. Krämer and P. Tunney, *Strongly interacting dark sectors in the early universe and at the LHC through a simplified portal*, *JHEP* **01** (2020) 162 [[arXiv:1907.04346](#)] [[INSPIRE](#)].
- [36] A. Dey, J. Lahiri and B. Mukhopadhyaya, *LHC signals of a heavy doublet Higgs as dark matter portal: cut-based approach and improvement with gradient boosting and neural networks*, *JHEP* **09** (2019) 004 [[arXiv:1905.02242](#)] [[INSPIRE](#)].
- [37] C.K. Khosa, V. Sanz and M. Soughton, *Using machine learning to disentangle LHC signatures of dark matter candidates*, *SciPost Phys.* **10** (2021) 151 [[arXiv:1910.06058](#)] [[INSPIRE](#)].
- [38] E. Arganda, A.D. Medina, A.D. Perez and A. Szykman, *Towards a method to anticipate dark matter signals with deep learning at the LHC*, *SciPost Phys.* **12** (2022) 063 [[arXiv:2105.12018](#)] [[INSPIRE](#)].
- [39] 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](#)].
- [40] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [[arXiv:1410.3012](#)] [[INSPIRE](#)].
- [41] DELPHES 3 collaboration, *DELPHES 3, a modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [[arXiv:1307.6346](#)] [[INSPIRE](#)].
- [42] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J. C* **72** (2012) 1896 [[arXiv:1111.6097](#)] [[INSPIRE](#)].
- [43] M. Park and M. Zhang, *Tagging a jet from a dark sector with jet-substructures at colliders*, *Phys. Rev. D* **100** (2019) 115009 [[arXiv:1712.09279](#)] [[INSPIRE](#)].
- [44] H. Beauchesne and G. Grilli di Cortona, *Event-level variables for semivisible jets using anomalous jet tagging*, in *2022 Snowmass summer study*, (2021) [[arXiv:2111.12156](#)] [[INSPIRE](#)].
- [45] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer and T. Reiter, *UFO — the Universal FeynRules Output*, *Comput. Phys. Commun.* **183** (2012) 1201 [[arXiv:1108.2040](#)] [[INSPIRE](#)].
- [46] N.D. Christensen and C. Duhr, *FeynRules — Feynman rules made easy*, *Comput. Phys. Commun.* **180** (2009) 1614 [[arXiv:0806.4194](#)] [[INSPIRE](#)].
- [47] L. Carloni and T. Sjöstrand, *Visible effects of invisible hidden valley radiation*, *JHEP* **09** (2010) 105 [[arXiv:1006.2911](#)] [[INSPIRE](#)].
- [48] T. Buss et al., *What’s anomalous in LHC jets?*, [[arXiv:2202.00686](#)] [[INSPIRE](#)].

- [49] Y. Wang, Y. Sun, Z. Liu, S.E. Sarma, M.M. Bronstein and J.M. Solomon, *Dynamic graph CNN for learning on point clouds*, *ACM Trans. Graph.* **38** (2019) 146 [[arXiv:1801.07829](#)] [[INSPIRE](#)].
- [50] H. Qu and L. Gouskos, *ParticleNet: jet tagging via particle clouds*, *Phys. Rev. D* **101** (2020) 056019 [[arXiv:1902.08570](#)] [[INSPIRE](#)].
- [51] A. Butter et al., *The machine learning landscape of top taggers*, *SciPost Phys.* **7** (2019) 014 [[arXiv:1902.09914](#)] [[INSPIRE](#)].
- [52] F.A. Dreyer and H. Qu, *Jet tagging in the Lund plane with graph networks*, *JHEP* **03** (2021) 052 [[arXiv:2012.08526](#)] [[INSPIRE](#)].
- [53] S. Gong et al., *An efficient Lorentz equivariant graph neural network for jet tagging*, *JHEP* **07** (2022) 030 [[arXiv:2201.08187](#)] [[INSPIRE](#)].
- [54] M. Abadi et al., *TensorFlow: large-scale machine learning on heterogeneous systems*, <https://www.tensorflow.org/>, (2015).
- [55] F. Chollet et al., *Keras*, <https://github.com/fchollet/keras>, (2015).
- [56] D.P. Kingma and J. Ba, *Adam: a method for stochastic optimization*, [arXiv:1412.6980](#) [[INSPIRE](#)].