Searching for dark jets with displaced vertices using weakly supervised machine learning

Debjyoti Bardhan,[1] **Yevgeny Kats, Noam Wunch**

Department of Physics, Ben-Gurion University, Beer-Sheva 8410501, Israel

E-mail: debjyoti.bardhan@acads.iiserpune.ac.in, katsye@bgu.ac.il, noamwunch1@gmail.com

ABSTRACT: If "dark quarks" from a confining hidden sector are produced at the LHC, they will shower and hadronize to dark sector hadrons, which may decay back to Standard Model particles within the detector, possibly resulting in a collimated spray of particles resembling a QCD jet. In this work we address scenarios in which dark hadrons decay with a measurable small displacement, such that the relevant background is dominated by heavy-flavor jets. Since dark sector parameters are largely unconstrained, and the precise properties of a dark QCD-like theory are difficult to compute or simulate reliably in any case, model-independent, data-based searches for such scenarios are desirable. We explore a search strategy employing weakly supervised machine learning to search for anomalous jets with displaced vertices. The method is tested on several toy signals, demonstrating the feasibility of such a search. Our approach has potential to outperform simple cut-based methods in some cases and has the advantage of being more model-independent.

---

[1]Present address: Department of Physics, Indian Institute of Science Education and Research, Pune, India.

Contents

1 Introduction

The diversity of particles and interactions in the Standard Model (SM), along with the multiple questions that the SM leaves unanswered, make it plausible for additional sectors of particles to exist in nature. One simple possibility is an exotic QCD-like sector [1]. The fermions and gauge bosons of this sector can be charged under a new ("dark") gauge group and neutral under the SM. They are termed dark quarks and dark gluons, in analogy with QCD. If the dark gauge group confines at low energies, the spectrum of this sector will contain composite states neutral under the new gauge group—dark hadrons. One motivation for these models is the possibility that dark hadron species that are stable on cosmological scales may account for dark matter [2–5]. However, such models are interesting also if they do not play this role.

If some portal (for example, a heavy mediator) couples the SM with the hidden sector, dark quarks can potentially be produced at the LHC. If dark quarks are produced, they will undergo parton showering and hadronization in the dark sector, similar to QCD quarks. Species of dark hadrons that are stable on detector scales will escape the detector leaving a trail of missing energy. On the other hand, some species may be unstable, decaying back to the SM within the detector and forming a peculiar jet of SM particles. In this work we aim to obtain better sensitivity to such types of objects, known as *dark jets*.

The collider signature of dark jets (see ref. [6] for a review) is greatly influenced by dark sector specifics. Many typical models would contain light dark pions π', and dark vector mesons ρ' and other hadrons with masses of order of the dark sector confinement scale $\Lambda_{\mathrm{QCD}'}$. In such scenarios, dark jets can be coarsely characterized by three parameters: the average fraction of momentum carried by stable (invisible) dark hadrons r_{inv}, dark pion mass $m_{\pi'}$, and dark pion lifetime $\tau_{\pi'}$.

In cases with sizable r_{inv}, a key signature will be missing energy $\not{E}_T$ aligned with a jet. This scenario of *semivisible jets* was analyzed in refs. [7, 8], where a search program for such models was proposed. References [9–11] suggested also making use of jet substructure variables. A search for resonant production of semivisible jet pairs was conducted by CMS in ref. [12]. It employed the cuts motivated by ref. [8] to probe for models with intermediate r_{inv} and promptly decaying visible dark hadrons. This search also used a boosted decision tree with jet substructure inputs motivated by refs. [9, 10]. ATLAS has performed a search for nonresonant production of semivisible jets in ref. [13]. Potential use of supervised deep neural networks (NN) for the classification of prompt, semivisible jets was studied in refs. [14, 15], weakly supervised learning was considered in ref. [16], and the use of an unsupervised NN, an autoencoder, was considered in ref. [17].

Other types of dark jet scenarios, where missing energy is no longer a dominant signature, are also possible. For example, dark pions with macroscopic flight distances, $c\tau_{\pi'}$ of order $1 - 10$ cm, will manifest as highly displaced objects within the jet. A novel reconstruction object, Emerging Jet, has been proposed for the classification of such jets in ref. [18]. A search for such objects, which are jets with few or no tracks originating from the primary vertex, was later conducted by CMS [19]. This search was sensitive to scenarios with large dark pion flight distance, where QCD background is scarce.

Overall, the large number of unknown dark sector parameters (gauge group, confinement scale, number of dark quark flavors and their masses, additional interactions within the dark sector, type of couplings to the SM and their strength), combined with the difficulty of simulating dark sector showering and hadronization reliably, call for model-independent and simulation-independent searches for anomalous jets. Machine learning (ML), and in particular weakly supervised ML, is a natural tool for such a task.

In this work we propose to employ weakly supervised ML for a largely model-independent, data-based search that would be sensitive to anomalous jets (such as dark jets) containing mildly displaced decays, so that the background is dominated by heavy-flavor jets. We will not assume the anomalous jets to contain missing energy since that case has already been explored a lot in the literature, but will instead rely on the presence of displaced objects. We will assume the anomalous jets to be pair produced in a decay of a heavy resonance.

The rest of the paper is organized as follows. In section 2 we review the relevant ideas of weakly supervised machine learning. In section 3 we describe the proposed search, examine the most important backgrounds and define the features that will be made available to the NN. In section 4 we define the datasets of signals and background that we use to simulate the search. We present the search simulation in section 5. We discuss the results and state our conclusions in section 6. Appendix A describes the event generation. The details of the neural network classifier are provided in appendix B. Jet feature distributions of all benchmark signals compared with the background distributions are presented in appendix C. The fit procedure used for estimating bump significance is described in appendix D.

2 Weakly supervised machine learning

While the most traditional ML approach, that of *fully supervised learning*, can provide very powerful classifiers, using it to search for physics beyond the Standard Model (BSM) requires specifying the exact BSM scenario that is being searched for (and being able to simulate it reliably). This makes fully supervised methods very model specific. In recent years, methods have been developed which lessen signal model dependence for selecting the test statistic. These methods provide different amounts of model independence, with the common trade-off of model independence vs. signal sensitivity.

An example of a completely model-independent test statistic is the output of an autoencoder trained on data (e.g., refs. [20–23]). This *unsupervised learning* method, while being completely model agnostic, lacks sensitivity to many signals [24, 25].

A more moderate approach, in the realm of *weakly supervised learning*, requires knowledge of class proportions. In fully supervised training the true class of each training example, e.g. signal/background, is known and provided to the NN. Knowledge of class proportions means only knowing what fraction of training examples belong to each class. Using class proportions alone, a classifier can learn to distinguish between classes, while training directly on the mixed data. In ML literature this method goes by the name Learning from Label Proportions [26, 27]. It was shown to be effective in quark/gluon discrimination, where calculation of flavor proportions is possible [28, 29].

This was extended to cases where label proportions are unknown, with the sole requirement of two event groups that have different signal proportions—this was termed *Classification Without Labels (CWoLa)* [30, 31]. To implement it in a search, one must separate the data to signal- and background-rich groups, based on some property of the signal model. In the case of a signal resonant in some parameter, the signal-rich sample can be obtained from selecting events near the resonance. This method goes by the name *Extended Bumphunt* [32] and was implemented by ATLAS in ref. [33]. Using CWoLa in a monojet search to enhance its sensitivity to semivisible jets was proposed in ref. [16].

Another approach, *Tag N' Train* [34], suggests using signal dijet topology to obtain the mixed samples. This method uses *cotraining*, with the dataset of dijet events being split into two *views* of each event, one containing first-jet features and the other containing second-jet features. A classifier is trained to discriminate signal-like first jets from background-like first jets. A second classifier is similarly trained on second-jet features. Finally these classifier predictions are combined amounting to an event classifier. Each of the jet classifiers is trained using CWoLa, where signal- and background-rich labels are obtained from some criterion on the other jet in the event. In ref. [34] this criterion was taken to be a cut on the output of an autoencoder trained on the jet. It was shown in [34] that Tag N' Train and Extended Bumphunt can be effectively combined in searches for a dijet resonant signal. In the current work, we adapt this last approach to suggest a new search for dark jets.

3 Proposed search

We propose a largely model-independent, data-based search that would be sensitive to resonantly pair-produced anomalous jets containing mildly displaced decays.

3.1 Strategy

We first select for dijet events with displaced objects, as will be described in section 3.2. Next we define signal and background regions in dijet invariant mass based on a mediator mass and resonance width hypothesis.

An event classifier is obtained according to the following procedure. As in Tag N' Train [34], each of the two leading jets in each event may be assigned a signal- or background-rich *weak label* according to some condition on the "other jet" (among the two) in the event. In Tag N' Train, the "other jet" condition was based on an autoencoder output, which is fully signal model independent. We propose to use some model assumption, namely the fact that dark jets will often have more constituents than SM jets.[1] Therefore, we choose jet constituent count, $n_{\rm obj}$, as our weak classifier.[2] The two jets are ordered by descending p_T and labeled j_1 and j_2. Signal-rich labels are assigned to jets within the signal

---

[1]One could also design an analogous search that would be sensitive to scenarios in which BSM jets have fewer constituents than SM jets.

[2]Other quantities, such as the number or the properties of the displaced vertices in the jet, could serve as alternative weak classifiers. Since our goal is to define a model-independent search, we want the criterion defining the weak labels to be simple and general and not optimized for any particular scenario. The more detailed use of the various features that might distinguish signal from background in each particular scenario is left to the neural network.

region for which the other jet constituent count is greater than some chosen threshold n_{obj}^S. Background-rich labels are assigned to jets coming from the entire mass range (signal region and sidebands) for which the other jet constituent count is smaller than some chosen threshold n_{obj}^B. Using these S/B-rich labels, two classifiers are trained, one on j_1s and the other on j_2s. We define our test statistic for the event as a whole to be the product of the two jet classifier outputs. This quantity will tend to have higher values for signals than for background events. To avoid inference of events used for training, the data should be split into k-folds. The preceding steps should be repeated k times, each time leaving a different fold out of training. The event classifier not trained on a given fold is used to classify the fold events.

The classifier is applied to both signal-region and sideband events, and a cut with efficiency ϵ_D for the data in that entire mass range is applied on the classifier output. The optimal value of ϵ_D, i.e., most sensitive to signal, is model dependent and therefore several values should be used. After applying the cut, the invariant mass distribution in the sidebands is interpolated into the signal region. The expected event count in the signal region, based on the interpolation, is compared to the measured number of events in the signal region. The significance of the excess is estimated based on Poisson statistics and systematic uncertainties of the interpolation.

The search is to be conducted in the form of a bump hunt in dijet invariant mass, i.e. each mediator mass hypothesis, $m_{Z'}$, is considered separately. Resonance width can either be determined from simulation or scanned over (e.g., as in the BUMPHUNTER [35]).

Once a significant excess is identified by the NN, the jets from the events that pass the NN cut can be examined manually to understand their nature.

3.2 Event selection

Event selection for the proposed analysis is performed in two steps: a primary selection for dijet events adhering with trigger limitations, and a more tailored selection for events with displaced objects. Event selection requirements are summarised in table 1.

3.2.1 Primary selection

The main motivation for the primary selection is to adhere with trigger limitations. To ensure this, we follow the cuts of an ATLAS dijet resonance search [36]. First, the two jets are required to have $p_T > 150$ GeV and $|\eta| < 2$. Two more event level cuts are applied. The first is based on half the rapidity separation of the leading jets, $y^* = (y_1 - y_2)/2$. The absolute value of this observable tends to be smaller for s-channel processes, such as our resonant signal. To increase signal purity we therefore require $|y^*| < 0.8$. The second requirement is a minimal azimuthal separation between leading jets, $\Delta\phi(j_1, j_2) = |\phi_1 - \phi_2| > 1$, to prevent excessive overlap between the jets. Finally, a lower bound of 1133 GeV on dijet invariant mass m_{jj} is required to ensure compliance with the trigger [36].

3.2.2 Displaced object selection

We wish to select for dijet events with displaced objects. The criterion we chose for such events is that at least 20% of the jet transverse momentum should be carried by tracks

that are associated with reconstructed displaced vertices. To suppress contributions from long-lived SM hadrons, vertices with two tracks and vertex mass close to the Λ or K_S^0 masses (computed with the appropriate particle identity assumptions for the products) are discarded. A summary of event requirements is given in table 1.

p_T^{jet}	> 150 GeV
$|\eta|^{\text{jet}}$	< 2
m_{jj}	> 1133 GeV
$|y^*|$	< 0.8
$\Delta\phi(jj)$	> 1
$\sum_{\text{disp.vert.}} p_T^{\text{vertex}}/p_T^{\text{jet}}$	> 0.2

Table 1: Event selection summary. Both leading jets (highest p_T) must satisfy p_T, η, and displaced vertex requirements.

3.3 Standard Model background

Displaced vertices are primarily a signature of events containing heavy flavor (b or c) quarks. We therefore expect the leading SM background for our analysis to be dijet events where both leading jets are of a heavy flavor. To estimate the background magnitude and composition we generated events in four groups:

- "bb events" contain a pair of b-flavored jets at the parton level. These events are primarily $b\bar{b}$, with smaller contributions from bb and $\bar{b}\bar{b}$.

- "cc events" contain a pair of c-flavored jets, analogous to the above.

- "bc events" contain one b-flavored jet and one c-flavored jet.

- The fourth group consists of the remaining dijet events (which we will call "other"), which are events that contain only one or no heavy-flavor jets at the parton level.

Event generation, including details about detector simulation and vertexing, is described in appendix A.

The total background cross section after the selection described in section 3.2 is $\sim$ 0.13 pb.[3] The leading contribution ($\sim 50\%$) comes from bb events. The next dominant background ($\sim 37\%$) comes from light or semi-light QCD events, i.e., with less than two final state heavy quarks at the parton level (the "other" category above). This group is dominated by events with gluons splitting to heavy quarks, a process that is significant in

---

[3]To reduce the computational burden, our simulation used leading-order matrix elements, followed by parton showering. However, higher-order QCD corrections, including hard jet radiation, can have some effect on the production cross section, the selection efficiency, and the properties of the two leading jets. While our data-based search methodology is not directly dependent on the simulation details, quantitative claims about the range of scenarios that can be discovered or excluded may be affected.

the hard events under consideration [37]. The remaining groups, bc and cc, account for 10% and 3% of the events, respectively. The selection efficiencies of the different groups are summarized in table 2.

Group	N_{prim}	N_{pass}	ϵ_{DO}	σ_{prim} (pb)	σ (pb)
bb	1066652	100551	0.094	0.71	0.067
jj ("other")	671729	62	$9.2 \cdot 10^{-5}$	530	0.049
cc	2059665	27069	0.013	0.98	0.013
bc	577405	9163	0.016	0.24	0.0038

Table 2: Selection efficiencies and magnitudes of different SM channels. The cross section after the primary selection, σ_{prim}, is derived from the generation level cross section of group events obtained from MadGraph at leading order times the efficiency of the primary selection. Displaced object efficiency is presented with respect to events after primary selection, namely $\epsilon_{\mathrm{DO}} = \frac{N_{\mathrm{pass}}}{N_{\mathrm{prim}}}$ (where N_{prim} and N_{pass} refer to the numbers of our MC events). The final available cross section is determined according to $\sigma = \epsilon_{\mathrm{DO}} \sigma_{\mathrm{prim}}$.

3.4 Jet features used for classification

The cotraining step of the search requires a choice of jet classification model and jet representation. We chose to represent each jet as a list of high-level features as input to a dense NN model. A complete description of the NN model we used is provided in appendix B. More complex representations and architectures, such as a long short-term memory (LSTM) network on lists of vertex features, were also considered. In our testing, these were outperformed by the simple dense architecture and therefore abandoned. This could change as the amount of analysis data grows since more data often favors more complex networks.

Jet features include vertex features chosen to represent the properties of displaced objects within a jet and general jet features that encode complementary jet information. We consider the following vertex features: vertex mass, vertex transverse displacement D_0 divided by the boost factor $\gamma\beta_T$, fraction of jet's transverse momentum carried by the vertex tracks, and vertex track count. For the features above, in the case of more than one reconstructed vertex, the median value across reconstructed vertices is used. The boost factor, $\gamma\beta_T$, is computed according to

$$\gamma\beta_T = \frac{p_T^{\mathrm{vertex}}}{m_{\mathrm{vertex}}} \tag{3.1}$$

where p_T^{vertex} is the magnitude of the vector sum of $\mathbf{p}_T$s of tracks associated to the vertex. Vertex mass is calculated according to

$$m_{\mathrm{vertex}}^2 = \left(\sum_{\mathrm{tracks}} \sqrt{\mathbf{p}_{\mathrm{track}}^2 + m_{\pi^{\pm}}^2} \right)^2 - \left(\sum_{\mathrm{tracks}} \mathbf{p}_{\mathrm{track}} \right)^2 \tag{3.2}$$

i.e. the tracks are assigned the charged pion mass for estimation of their energy, and the sum is over all tracks associated with the vertex. We also supply the total number of

reconstructed vertices in the jet, excluding the primary vertex and the number of particle-flow objects in the jet – n_{obj}.

In our toy dark sector models that will be described in the next section, the discrimination power of each of these features varies with dark sector parameters. The dark pion mass $m_{\pi'}$ directly affects m_{vertex}. Increasing dark pion mass also decreases the number of vertices per jet and increases the number of tracks per vertex. The dark pion lifetime $\tau_{\pi'}$ directly affects $(D_0/\gamma\beta_T)^{\mathrm{vertex}}$ and also indirectly affects the number of vertices. For larger dark pion lifetime, more displaced vertices are distinguished from the primary vertex and therefore the number of displaced vertices increases.

4 Benchmark datasets

While the search is intended to be largely model independent, it is useful to test the strategy on some examples. Since the detailed physics of confining hidden sectors is not known well and is very model dependent, and the simulation tools are limited too, we consider a set of simplistic toy models, defined as follows.

4.1 Benchmark hidden sectors

We base our toy models on the scenario that is obtained in the PYTHIA8 Hidden Valley module [38] for an $SU(3)$ gauge group with a single quark flavour. We consider fully visible jets, i.e. $r_{\mathrm{inv}} = 0$, which manifests as no excessive missing transverse energy. We consider six combinations of the remaining two parameters, with values $(m_{\pi'}, c\tau_{\pi'}) = \{5\,\mathrm{GeV}, 10\,\mathrm{GeV}\} \times \{0.1\,\mathrm{mm}, 0.2\,\mathrm{mm}, 0.3\,\mathrm{mm}\}$. Other mass parameters of the dark sector—confinement scale, constituent dark quark mass, and vector meson mass—were scaled with $m_{\pi'}$, starting at $\Lambda_{\mathrm{QCD'}} = 5\,\mathrm{GeV}$, $m_{q'} = 5\,\mathrm{GeV}$, and $m_{\rho'} = 10.5\,\mathrm{GeV}$ for $m_{\pi'} = 5\,\mathrm{GeV}$. The probability for creating dark vector mesons is kept at its default value of 0.75. We assume the dark vector mesons decay promptly and exclusively to dark pion pairs: $\rho' \to \pi'\pi'$. Our simulated dark pions decay exclusively to SM down quark-antiquark pairs: $\pi' \to d\bar{d}$. Decays to heavier flavor quarks are in principle possible for $m_{\pi'}$ values considered and perhaps even motivated by helicity suppression. However, we found such scenarios less interesting as they produce many additional displaced vertices from the decays of the heavy flavor quarks, amounting to a signal too distinct from QCD background. The benchmark Hidden Valley parameters are summarized in table 3.

We test the sensitivity of our proposed search to resonant dijet events produced via $pp \to Z' \to q'\bar{q}'$ with $m_{Z'} = 2\,\mathrm{TeV}$. We assume an interaction Lagrangian of the following form:

$$\mathcal{L} \supset -g_q Z'_\mu \bar{q}\gamma^\mu q - g_{q'} Z'_\mu \bar{q}'\gamma^\mu q' \,, \tag{4.1}$$

where the first term describes the Z' coupling to SM quarks and the second to dark quarks.

4.2 Benchmark datasets

Our background dataset contains $106k$ events that passed the full event selection in the dijet invariant mass range $1400\,\mathrm{GeV} \leq m_{jj} \leq 2400\,\mathrm{GeV}$. The majority, $89k$, are bb events

Gauge group	SU(3)
$\Lambda_{\mathrm{QCD'}}$	5 / 10 GeV
$n_{q'}$	1
$m_{q'}$	5 / 10 GeV
$m_{\pi'}$	5 / 10 GeV
$m_{\rho'}$	10.5 / 21 GeV
$c\tau_{\pi'}$	0.1 / 0.2 / 0.3 mm
r_{inv}	0

Table 3: Hidden Valley parameters used for the six benchmark signal configurations.

and the rest, $17k$, are cc events. As can be seen in table 2, these two channels combined account for the majority of QCD events that pass selection. We would ideally simulate the entire QCD dijet sample (rather than only bb and cc); however this is too computationally expensive for us due to low selection efficiencies of the displaced objects cut. If we rescale the cross section so that the total background cross section is correct, then based on the analysis of section 3.3 this example corresponds to an integrated luminosity of ~ 800 fb^{-1} available for the analysis.

We will analyze in detail the example of a Z' with mass $m_{Z'} = 2$ TeV and a negligible width relative to the experimental dijet invariant mass resolution. Motivated by the shape of the resulting dijet invariant mass distribution (see, e.g., figure 5a), whose width is not very model dependent since it is dominated by the experimental resolution, we define the signal region to be the invariant mass range $m_{jj} \in [1600, 2000]$ GeV.[4] We define the sidebands as $m_{jj} \in [1400, 1600) \cup (2000, 2400]$ GeV. These boundaries are chosen such that the sidebands and the signal region contain comparable numbers of background events. Approximately 20% of the background events are in the signal region. Signals, one of the six hidden sector configurations described in section 4.1, are injected to this background. Signal size, which we will vary, will be presented in terms of signal fraction $f_S = N_S/(N_B + N_S)$, where N_B and N_S and the background and signal event count in the entire mass range (signal region and sidebands) after event selection.

The feature distributions for the different benchmark signals (and the background) are provided in appendix C. The most discriminating features for this set of benchmarks are the number of objects in the jet, vertex count, and transverse momentum fraction. One can also see that as dark pion displacement increases, from 0.1 mm to 0.3 mm, the number of signal vertices increases because more dark pions decay outside the primary vertex resolution. Vertex mass is a stronger discriminator for the 10 GeV dark pion mass in comparison to the 5 GeV case.

---

[4]This choice of the mass window is based on the region in which the resonant contribution appears in our DELPHES simulation. In an actual experimental analysis, this choice should be reconsidered based on a more accurate simulation of the corresponding detector and jet energy scale corrections that are applied to the reconstructed jets.

5 Example search

In this section we present results of an example search conducted on a simulated benchmark dataset. We provide a detailed account for the case of a dark sector with $m_{\pi'} = 10$ GeV and $c\tau_{\pi'} = 0.2$ mm with a signal fraction $f_S = 0.5\%$, where the number of signal events is $N_S = N_B \cdot \frac{f_S}{1-f_S} = 530$ events. We provide aggregated results for different signal fractions of all other benchmark signals.

5.1 Weak jet classifier

We used the number of particle-flow objects, n_{obj}, of each jet to assign a signal-rich or background-rich (weak) label for the other jet, as described in section 3.1. Background-like threshold, n_{obj}^B, was taken to be the 25% (lower) quantile for the number of particle-flow objects. Signal-like threshold, n_{obj}^S, was taken to be the 75% (upper) quantile for the number of particle-flow objects (amounting to $n_{\mathrm{obj}}^B = n_{\mathrm{obj}}^S$). The softer signal-like threshold is complemented by the invariant mass region selection so that after both cuts the signal- and background-rich labels are approximately balanced. The thresholds were chosen after trying a number of alternatives and finding that results are not very sensitive to this choice. Using tighter signal- and background-rich thresholds amounts to a higher effective signal fraction for training. This comes at the cost of less data available for training. A quantification of this tradeoff is left for future works.

Jet constituent count thresholds corresponding to the chosen quantiles are $n_{\mathrm{obj}}^{\mathrm{thresh}} = 24$ for cuts on j_1 and $n_{\mathrm{obj}}^{\mathrm{thresh}} = 25$ for cuts on j_2. The difference stems from a slightly higher object multiplicity for second jets. These values were unaffected by the small signal fractions considered and are therefore the same for all signals and all signal fractions. From cutting on j_1 and j_2 constituent counts, and requiring that signal-rich jets come from signal-region events, we obtain two background-rich and two signal-rich samples.

In the 0.5% signal fraction case of the $(m_{\pi'}, c\tau_{\pi'}) = (10$ GeV, 0.2 mm$)$ signal, these cuts leave 26292 (23906) background-rich jets and 27824 (28432) signal-rich jets from cuts on j_2 (j_1) constituent counts. From an initial 0.5% signal fraction in the entire dataset, the enriched signal fractions are 1.59% (1.56%) in the signal-rich samples and 0% (0%) in the background-rich samples from cuts on j_2 (j_1) multiplicities.

5.2 Weakly supervised event classifier

After applying the weak cuts to the 0.5% signal fraction case of the signal with $(m_{\pi'}, c\tau_{\pi'}) = (10$ GeV, 0.2 mm$)$, 51% (49%) of jets were assigned weak labels according to $n_{\mathrm{obj}}^{j_1}$ ($n_{\mathrm{obj}}^{j_2}$). Of these events, 10% are put aside for validation to avoid overfitting. A classifier, described in appendix B, is trained to distinguish between the remaining 46% (44%) of events using j_1 (j_2) features and the weak labels. After weak-label assignment and putting aside of the validation set, 48704 jets and 47104 jets are available for training j_1 and j_2 classifiers, respectively. The classifiers were trained for 100 epochs. Learning curves are presented in figure 8. To evaluate the classifiers' performance, a new dataset of $35k$ signal and $35k$ background events was generated. NN outputs and receiver operating characteristic (ROC) curves for the 0.5% signal fraction case of the $(m_{\pi'}, c\tau_{\pi'}) = (10$ GeV, 0.2 mm$)$ signal are

| (a) j_1 classifier | (b) j_2 classifier | (c) combined: $j_1^{\text{pred}} \cdot j_2^{\text{pred}}$ |

Figure 1: NN output distributions for j_1, j_2, and combined classifiers, for the scenario with $(m_{\pi'}, c\tau_{\pi'}) = (10 \text{ GeV}, \ 0.2 \text{ mm})$, $f_S = 0.5\%$.

shown in figures 1 and 2. ROCs comparing discrimination of classifiers trained on different signal fractions are shown in figure 3. As expected, classifier performance deteriorates as f_S is decreased. Still, even at $f_S = 0.1\%$, the classifier is very powerful. However, going to much lower signal fractions is not relevant because they will not be detectable eventually in the bump hunt procedure that is discussed in the next subsection. ROCs comparing the outcomes for the different benchmark signals with 0.5% signal fraction are shown in figure 4.

5.3 Identifying and quantifying an excess

Our null hypothesis, which we will confirm in the following, is that the dijet invariant mass distribution of the background after the NN cut is still well described by a smoothly decreasing function. We construct the following test statistic to probe for deviations from this hypothesis due to a possible signal. We bin the events (with a bin size of 50 GeV in our example) and fit the sidebands to the following three-parameter function

$$\frac{dN}{dm_{jj}} = p_0 \frac{(1 - m_{jj}/\sqrt{s})^{p_1}}{(m_{jj}/\sqrt{s})^{p_2}}, \tag{5.1}$$

also used in ATLAS [39] and CMS [40]. The fit parameters p_i were constrained to positive values. We estimate the number of expected events in the signal region using the fit and compare it to the measured number of events in the signal region. Our test statistic is the excess

$$t = \frac{N_{\text{meas}}^{\text{sig.reg.}} - N_{\text{exp}}^{\text{sig.reg.}}}{\sqrt{\sigma_{\text{meas}}^2 + \sigma_{\text{exp}}^2}} \ . \tag{5.2}$$

The uncertainty in the measured counts is estimated by Poisson statistics as $\sigma_{\text{meas}}^2 = N_{\text{exp}}^{\text{sig.reg.}}$. The uncertainty in the expected counts is obtained by linearly propagating parameter fit uncertainties. Further details of this procedure are provided in appendix D. We obtain test statistic values for different cut efficiencies of the NN. To avoid training many classifiers, as would be done in the k-fold procedure described in section 3.1, we use the

Figure 2: Solid curves are ROCs for NN jet classifiers trained using cotraining, and the event classifier obtained from their product, for the scenario with $(m_{\pi'}, c\tau_{\pi'}) = (10 \text{ GeV}, 0.2 \text{ mm})$, $f_S = 0.5\%$. Dashed curves are ROCs for the weak jet classifiers (constituent count of each jet) and the event classifier obtained from their sum. Shaded areas signify 1σ statistical uncertainty, where for a cut leaving N_B background events out of a total N_B^0 background events, we used $\sigma = N_B^0/N_B^2 \cdot \sqrt{N_B}$. The curves were terminated at $N_B < 10$ ($1/\epsilon_B = 3500$).

entire $106k$ event dataset for semisupervised training and continue with inference on a new (same size) dataset. This is similar to the k-fold procedure for a large enough k.

Let us now exemplify the search with one realization of a background and signal sample. Invariant mass distributions subject to NN cuts of varying efficiency are presented in figure 5 for the $(m_{\pi'}, c\tau_{\pi'}) = (10 \text{ GeV}, 0.2 \text{ mm})$ signal with $f_S = 0.5\%$. The invariant mass spectrum of the entire dataset after the event selection described in section 3.2 is shown in figure 5a. The test statistic significance prior to any further cut is -0.72σ. (The negative sign indicates a downward deviation.) The invariant mass spectra after applying the NN cuts with $\epsilon_D = 2\%$, 1%, 0.6% are shown in figures 5b, 5c, 5d, respectively. Apart from the significances obtained with fit to the sidebands, the information tables at the bottom of these plots show also the significance estimates that would be obtained from the naive calculation of $n_S/\sqrt{n_B}$ in the signal region. As expected, since it does not account for the statistical fluctuations in the sidebands and for the potential contributions in the sidebands due to the signal tails, the naive estimate of the significance is unrealistic, and

Figure 3: ROCs comparing classifiers trained using weakly supervised learning with varying signal fractions f_S of the benchmark signal with $(m_{\pi'}, c\tau_{\pi'}) = (10 \text{ GeV}, \ 0.2 \text{ mm})$.

it is crucial to simulate the fit to the sidebands like we did.

Significance as a function of ϵ_D for different signal fractions is presented in figure 6. The significance peaks at $\epsilon_D \sim f_S$, peaking at higher data efficiencies for greater signal fractions. For $f_S = 0.25\%$ and $f_S = 0.5\%$ the signals pass the discovery threshold of 5σ whereas the $f_S = 0.1\%$ signal falls short. Significance obtained for different benchmark signals with $f_S = 0.5\%$ is presented in figure 7. All benchmark signals are discoverable at this signal fraction.

We also test for the bump significance in a dataset with no signal. This corresponds to the $f_S = 0$ curve in figure 6. The significance fluctuates between ~ 0 and $2\ \sigma$ for the values of ϵ_D considered, which reassures us that no large, spurious bump is carved in the analysis. However, a similar significance trend was observed in a second background realization we tested, suggesting that some $\sim +1\sigma$ bias exists in our significance estimation. A more detailed study, which would involve generating a large number of background realizations, will be needed to quantify the size of this apparent bias more precisely. Additionally, one could explore whether the bias could be reduced by using a different fitting function. More sophisticated methods to reduce sculpting (e.g., along the lines of ref. [41]) could also be explored.

While the weakly supervised machine learning method outlined here is working well, it is interesting to ask how it performs relative to simpler methods. An obvious comparison in

Figure 4: ROCs comparing classifier discrimination for different benchmark signals. All classifiers were trained with $f_S = 0.5\%$.

the case of our benchmark models is to cutting on the jet constituent multiplicity variable. Instead of the loose cut on the multiplicity that was used in the ML approach for producing the weak labels, we apply tight cuts on the sum of the multiplicities of the two jets and use the same sidebands fit procedure. Table 4 summarises the bump significance obtained when cutting on the output of the weakly supervised event classifier and when cutting on the sum of object multiplicities of both jets, for all benchmark signals. The $(m_{\pi'}, c\tau_{\pi'}) = (5 \text{ GeV}, 0.3 \text{ mm})$ signal was discovered at $f_S = 0.25\%$ and all the rest were discovered at $f_S = 0.1\%$. For these signal fractions it was usually the case that cutting on multiplicity slightly outperformed the NN. (Note, however, that significance values like 5σ and higher are somewhat uncertain because they assume the fluctuations due to the fit uncertainties to remain Gaussian far on the tails. Also, the bias discussed in the previous paragraph needs to be quantified for both methods. Therefore, small differences should not be taken too seriously.) The exception is the $(m_{\pi'}, c\tau_{\pi'}) = (5 \text{ GeV}, 0.3 \text{ mm})$ signal that was discovered with higher significance by the NN at $f_S = 0.25\%$.

	$n_S^{\text{sig.reg.}}$	$n_B^{\text{sig.reg.}}$	significance
true	415	35851	2.2σ
sideband fit	-300	36566	-0.72σ

(a) all events

	$n_S^{\text{sig.reg.}}$	$n_B^{\text{sig.reg.}}$	significance
true	413	648	16σ
sideband fit	365	696	5.9σ

(b) $\epsilon_D = 2\%$

	$n_S^{\text{sig.reg.}}$	$n_B^{\text{sig.reg.}}$	significance
true	408	229	27σ
sideband fit	378	259	9.4σ

(c) $\epsilon_D = 1\%$

	$n_S^{\text{sig.reg.}}$	$n_B^{\text{sig.reg.}}$	significance
true	392	57	52σ
sideband fit	366	83	14σ

(d) $\epsilon_D = 0.6\%$

Figure 5: Invariant mass spectrum of events passing NN cut with varying selection efficiency ϵ_D. The sidebands are shaded. The significance in the "true" rows corresponds to true $n_S/\sqrt{n_B}$ within the signal region. The significance in the "sideband fit" rows is obtained from the fit parameters according to eq. (5.2).

Figure 6: Bump significance as a function of selection efficiency at four signal fractions for $(m_{\pi'}, c\tau_{\pi'}) = (10 \text{ GeV}, 0.2 \text{ mm})$ signal.

Figure 7: Bump significance as a function of selection efficiency for the six benchmark signals at $f_S = 0.5\%$.

f_S	$m_{\pi'}$	$c\tau_{\pi'}$	$\max \sigma_{\mathrm{NN}}$	$\max \sigma_{n_{\mathrm{obj}}}$
0.1%	5 GeV	0.1 mm	5.4 σ	7.2 σ
0.1%	5 GeV	0.2 mm	5.2 σ	5.9 σ
0.1%	5 GeV	0.3 mm	4.3 σ	4.8 σ
0.25%	5 GeV	0.3 mm	9.6 σ	8.6 σ
0.1%	10 GeV	0.1 mm	5.4 σ	6.8 σ
0.1%	10 GeV	0.2 mm	4.9 σ	7.2 σ
0.1%	10 GeV	0.3 mm	4 σ	5.6 σ

Table 4: Comparison between object multiplicity cut and weakly supervised NN cut. The columns $\max \sigma_{\mathrm{NN}}$ and $\max \sigma_{n_{\mathrm{obj}}}$ correspond with the maximum bump significance across selection efficiencies ϵ_D for cuts on NN output and cuts on the sum of object multiplicities of both jets, respectively. For this we consider ten values of ϵ_D spaced log uniformly in the range $[0.001, 0.1]$.

6   Summary and conclusions

A hidden ("dark") confining sector may reveal itself at the LHC in the form of anomalous jets, dubbed dark jets, whose properties are very model dependent. In this work we considered dark sectors with dark hadron lifetimes similar to heavy-flavor QCD quarks. A main feature of jets arising from such a sector is displaced vertices from the decays of dark hadrons. We propose using the features of reconstructed vertices to further capture the properties of the displaced objects. The dark sector scenarios we consider are complementary to the ones considered in most of the papers on the subject, which assume the presence of missing energy or very large vertex displacements or do not take advantage of displaced vertices.

The wealth of data collected at the LHC offers an opportunity to harness machine learning to discriminate BSM from SM signatures. A traditional approach to doing so is using MC simulations of signal (or a mix of signals) events and of SM events to train a NN. This paradigm has drawbacks. There are large uncertainties in simulating events, introduced by modeling uncertainties of nonperturbative QCD processes (and in our case also those of the dark confining sector) as well as detector modeling. Another drawback is a lack of generality which translates to reduced sensitivity (if any) to signals not used for training. This is a problem when sensitivity to a wide range of signals is required. Dark sector details are largely unconstrained, allowing for a wide range of dark jet signatures. In this work we propose using the weakly supervised method Tag N' Train in searches for dark jets with displaced vertices. Tag N' Train is a weakly supervised method to obtain a dijet event classifier. The procedure starts with a weak jet classifier. We propose using a cut on jet constituent multiplicity for this stage. This choice makes use of the fact that many dark sector models produce high multiplicity jets. Using the weak labels obtained

from the weak classifier, two classifiers are trained, one for each of the two leading jets in the event. We use dense NN supplied with displaced vertex features, including number of displaced vertices, vertex transverse displacement, vertex mass, number of associated tracks, and the fraction of transverse momentum carried by the vertex out of total jet transverse momentum. Jet constituent multiplicity was also supplied.

We tested this procedure on simulated events with a set of toy dark sector scenarios. We showed that the vertex features can be good discriminators between heavy flavor quark jets and dark jets. We demonstrated a concrete example of a search for resonant dark jet pairs with displaced vertices. The search is conducted in the form of a bump hunt where different mass hypotheses are tested separately. We presented a detailed analysis of the example of a 2 TeV resonance. The resonance mass hypothesis was incorporated in the weak classifier—only jets coming from events within the signal region in invariant mass were candidates to be assigned the signal-rich label. After training the NNs and applying them to simulated data, the significance of the bump was estimated for different NN selection efficiencies. The semisupervised classifier succeeded in learning from auxiliary features specific to the signal that was present in the data for signal fractions as small as 0.1%.

However, at least for the range of examples we examined, the sensitivity of our machine learning method turned out to be comparable (with the details of the comparison depending on the model) to what can be achieved by using the number of objects in the two jets, which by itself is a search that has never been done and is worth pursuing. One cause of the NN not offering a big advantage is the low signal fractions. The discrimination power of CWoLa often deteriorates with decreasing signal fraction while a cut on the number of objects in the jets is unaffected. The effective signal fraction can always be increased by tightening the thresholds of the weak classifier. However, this comes at the cost of fewer events available for training the NN. Therefore, this method might improve as more data is collected and available for analysis.

It could be interesting to extend this method to dark sectors with promptly decaying dark hadrons, where a very different set of features and different backgrounds will be relevant. Another interesting direction would be to consider nonresonant dijet production, where the Tag N' Train method naturally remains applicable.

Acknowledgments

A Event generation

Parton level events at collider energy of 13 TeV were generated using MADGRAPH5 [42] with the NN23LO1 [43] PDF set. A massive Z' mediator with couplings to SM quarks and dark quarks was implemented using the Universal FeynRules Output (UFO) files from [8]. Some parton-level cuts, softer than the eventual selection cuts of section 3.2, were applied in MadGraph to save computation time for the background: jet $p_T > 100$ GeV, $m_{jj} > 1$ TeV, jet $|\eta| < 3$, and $\Delta R(j_1, j_2) > 1$. Showering and hadronization were simulated using PYTHIA8 [44]. Dark-sector showering was done using PYTHIA8's Hidden Valley module [38]. Detector simulation was conducted with DELPHES 3 [45] using the ATLAS detector card with added track smearing according to [46]. Jets were reconstructed from calorimeter deposits using the anti-k_T algorithm [47] with jet radius $R = 0.7$. Particle-Flow[5] constituents were then assigned to jets based on their angular distance (ΔR) from the axes of the reconstructed jets. Vertices were reconstructed with the Adaptive Vertex Fitting algorithm (AVR) [48] using default parameters ($\sigma_{\text{cut,p}} = 2$, $\sigma_{\text{cut,s}} = 6$, and $w_{\min} = 0.5$), implemented in the RAVE toolkit [49]. All event tracks were used for primary vertex reconstruction while only tracks belonging to a given jet were used to find secondary vertices.

B Neural network architecture

We use a dense neural network architecture built and trained using Keras [50] with Tensor-Flow [51] backend. The network has four hidden layers with 32, 16, 16, 4 nodes, respectively. These parameters were coarsely optimized to avoid over/under fitting. The first hidden layer activation is the Leaky Rectified Linear Unit (LeakyReLu). For the remaining three layers Exponential Linear Units (ELU) were used. A sigmoid function was applied to the output. Each hidden layer except the last was followed by a dropout layer with a rate of 0.1. A network summary is provided in table 5. Binary cross-entropy loss function and Adam optimizer were used for training. Each input feature was globally shifted and scaled according to the sample mean and standard deviation of the training data. These scale and shift values are saved. When new data is to be inferred by the classifier it is scaled and shifted by the same values. Examples of learning curves are shown in figure 8.

---

[5]Particle Flow is an algorithm to reconstruct track and calorimeter tower measurements into a list of electrons, muons, charged hadrons, neutral hadrons, and photons.

Layer (type)	Output shape	# Parameters
Layer-1 (Dense)	(None, 32)	224
activation (LeakyReLU)	(None, 32)	0
dropout (Dropout)	(None, 32)	0
Layer-2 (Dense)	(None, 16)	528
activation (ELU)	(None, 16)	0
dropout (Dropout)	(None, 16)	0
Layer-3 (Dense)	(None, 16)	272
activation (ELU)	(None, 16)	0
dropout (Dropout)	(None, 16)	0
Layer-4 (Dense)	(None, 4)	68
activation (ELU)	(None, 4)	0
Output (Dense)	(None, 1)	5
Total parameters: 1,097		

Table 5: NN summary.

(a) j_1 classifier

(b) j_2 classifier

Figure 8: Learning curve of j_1 (left) and j_2 (right) classifiers training on $(m_{\pi'}, c\tau_{\pi'}) = (10 \text{ GeV}, 0.2 \text{ mm})$ signal with $f_S = 0.5\%$. Since the validation set is evaluated without use of the dropout layers it is not surprising that the validation set loss is smaller than the training set loss.

C   Feature distributions

This appendix presents the feature distributions for the benchmark signals and the background based on events in the mass region $m_{jj} \in [1400, 2400]$ GeV after event selection.

C.1 $(m_{\pi'}, c\tau_{\pi'}) = (5$ GeV, 0.1 mm$)$

Figure 9: Distributions of vertex displacement $D_0/\gamma\beta_T$, vertex mass m_{vertex}, vertex transverse momentum fraction $p_T^{\mathrm{vertex}}/p_T^{\mathrm{jet}}$, number of tracks associated to vertex n_{tracks}, total number of jet constituents n_{obj}, and total number of reconstructed displaced vertices n_{vertices} (not including primary vertex). If more than one vertex is reconstructed in a given jet the median value for vertex features is taken. The step-like features in the distribution of $p_T^{\mathrm{vertex}}/p_T^{\mathrm{jet}}$ are an artifact of requiring the sum of this variable over all jet vertices to be greater than 0.2. Jets with only one displaced vertex, which are the majority of background jets, are constrained to values greater than 0.2. Jets with two displaced vertices are constrained to a median greater than 0.1, etc.

C.2 $(m_{\pi'}, c\tau_{\pi'}) = (5 \text{ GeV}, \ 0.2 \text{ mm})$

Figure 10: Distributions of vertex displacement $D_0/\gamma\beta_T$, vertex mass m_{vertex}, vertex transverse momentum fraction $p_T^{\text{vertex}}/p_T^{\text{jet}}$, number of tracks associated to vertex n_{tracks}, total number of jet constituents n_{obj}, and total number of reconstructed displaced vertices n_{vertices} (not including primary vertex). If more than one vertex is reconstructed in a given jet the median value for vertex features is taken. The step-like features in the distribution of $p_T^{\text{vertex}}/p_T^{\text{jet}}$ are an artifact of requiring the sum of this variable over all jet vertices to be greater than 0.2. Jets with only one displaced vertex, which are the majority of background jets, are constrained to values greater than 0.2. Jets with two displaced vertices are constrained to a median greater than 0.1, etc.

C.3 $(m_{\pi'}, c\tau_{\pi'}) = (5\ \text{GeV},\ 0.3\ \text{mm})$

Figure 11: Distributions of vertex displacement $D_0/\gamma\beta_T$, vertex mass m_{vertex}, vertex transverse momentum fraction $p_T^{\text{vertex}}/p_T^{\text{jet}}$, number of tracks associated to vertex n_{tracks}, total number of jet constituents n_{obj}, and total number of reconstructed displaced vertices n_{vertices} (not including primary vertex). If more than one vertex is reconstructed in a given jet the median value for vertex features is taken. The step-like features in the distribution of $p_T^{\text{vertex}}/p_T^{\text{jet}}$ are an artifact of requiring the sum of this variable over all jet vertices to be greater than 0.2. Jets with only one displaced vertex, which are the majority of background jets, are constrained to values greater than 0.2. Jets with two displaced vertices are constrained to a median greater than 0.1, etc.

C.4 $(m_{\pi'}, c\tau_{\pi'}) = (10 \text{ GeV},\ 0.1 \text{ mm})$

Figure 12: Distributions of vertex displacement $D_0/\gamma\beta_T$, vertex mass m_{vertex}, vertex transverse momentum fraction $p_T^{\text{vertex}}/p_T^{\text{jet}}$, number of tracks associated to vertex n_{tracks}, total number of jet constituents n_{obj}, and total number of reconstructed displaced vertices n_{vertices} (not including primary vertex). If more than one vertex is reconstructed in a given jet the median value for vertex features is taken. The step-like features in the distribution of $p_T^{\text{vertex}}/p_T^{\text{jet}}$ are an artifact of requiring the sum of this variable over all jet vertices to be greater than 0.2. Jets with only one displaced vertex, which are the majority of background jets, are constrained to values greater than 0.2. Jets with two displaced vertices are constrained to a median greater than 0.1, etc.

Figure 13: Distributions of vertex displacement $D_0/\gamma\beta_T$, vertex mass m_{vertex}, vertex transverse momentum fraction $p_T^{\text{vertex}}/p_T^{\text{jet}}$, number of tracks associated to vertex n_{tracks}, total number of jet constituents n_{obj}, and total number of reconstructed displaced vertices n_{vertices} (not including primary vertex). If more than one vertex is reconstructed in a given jet the median value for vertex features is taken. The step-like features in the distribution of $p_T^{\text{vertex}}/p_T^{\text{jet}}$ are an artifact of requiring the sum of this variable over all jet vertices to be greater than 0.2. Jets with only one displaced vertex, which are the majority of background jets, are constrained to values greater than 0.2. Jets with two displaced vertices are constrained to a median greater than 0.1, etc.

C.6 $(m_{\pi'}, c\tau_{\pi'}) = (10 \text{ GeV}, \ 0.3 \text{ mm})$

Figure 14: Distributions of vertex displacement $D_0/\gamma\beta_T$, vertex mass m_{vertex}, vertex transverse momentum fraction $p_T^{\text{vertex}}/p_T^{\text{jet}}$, number of tracks associated to vertex n_{tracks}, total number of jet constituents n_{obj}, and total number of reconstructed displaced vertices n_{vertices} (not including primary vertex). If more than one vertex is reconstructed in a given jet the median value for vertex features is taken. The step-like features in the distribution of $p_T^{\text{vertex}}/p_T^{\text{jet}}$ are an artifact of requiring the sum of this variable over all jet vertices to be greater than 0.2. Jets with only one displaced vertex, which are the majority of background jets, are constrained to values greater than 0.2. Jets with two displaced vertices are constrained to a median greater than 0.1, etc.

D Fit procedure

The sidebands were fit using `scipy.curve_fit`, scipy's [52] implementation of nonlinear least squares fit. The fit optimizes the cost function $L = \mathbf{r}^T\mathbf{r}$ where $\mathbf{r}_i$ is the residual in the i'th bin divided by the uncertainty in measured bin counts. The bin count uncertainty in a bin with n counts was taken to be $\sqrt{n}$ according to Poisson statistics. The statistical uncertainty of expected counts in the signal region was estimated according to

$$
\begin{aligned}
\sigma^2_{\text{exp}} = \text{Var}\left(\sum_{x\in\text{sig.reg.}} N(x,\mathbf{p})\right) &\approx \text{Var}\left(\sum_{x\in\text{sig.reg.}} \frac{dN}{d\mathbf{p}}(x,\hat{\mathbf{p}})\cdot(\mathbf{p}-\hat{\mathbf{p}})\right) \\
&= \left(\sum_{x\in\text{sig.reg.}} \frac{dN}{d\mathbf{p}}\right)^T \mathbf{Cov}\left(\sum_{x\in\text{sig.reg.}} \frac{dN}{d\mathbf{p}}\right),
\end{aligned}
\tag{D.1}
$$

where $N(x,\mathbf{p})$ is the fit function from eq. (5.1) multiplied by the bin size, $\mathbf{p}$ is a random variable vector of fit function parameters, and $\hat{\mathbf{p}}$ are the estimated parameters. The covariance matrix for the parameters is estimated by

$$
\mathbf{Cov} = \frac{L}{m-n}(\mathbf{J}^T\mathbf{J})^{-1},
\tag{D.2}
$$

where L is the cost function at $\hat{\mathbf{p}}$, m is the number of points used for the fit, n is the number of parameters $(=3)$, and $\mathbf{J}$ is the Jacobian of $\mathbf{r}$ with respect to the parameters, evaluated at $\hat{\mathbf{p}}$.

References

[1] M. J. Strassler and K. M. Zurek, "Echoes of a hidden valley at hadron colliders," *Phys. Lett. B* **651** (2007) 374–379, `arXiv:hep-ph/0604261`.

[2] Y. Bai and P. Schwaller, "Scale of dark QCD," *Phys. Rev. D* **89** no. 6, (2014) 063522, `arXiv:1306.4676 [hep-ph]`.

[3] 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 [hep-ph]`.

[4] H. Beauchesne, E. Bertuzzo, and G. Grilli Di Cortona, "Dark matter in Hidden Valley models with stable and unstable light dark mesons," *JHEP* **04** (2019) 118, `arXiv:1809.10152 [hep-ph]`.

[5] H. Beauchesne and G. Grilli di Cortona, "Classification of dark pion multiplets as dark matter candidates and collider phenomenology," *JHEP* **02** (2020) 196, `arXiv:1910.10724 [hep-ph]`.

[6] G. Albouy *et al.*, "Theory, phenomenology, and experimental avenues for dark showers: a Snowmass 2021 report," *Eur. Phys. J. C* **82** no. 12, (2022) 1132, `arXiv:2203.09503 [hep-ph]`.

[7] T. Cohen, M. Lisanti, and H. K. Lou, "Semivisible Jets: Dark Matter Undercover at the LHC," *Phys. Rev. Lett.* **115** no. 17, (2015) 171804, `arXiv:1503.00009 [hep-ph]`.

[8] T. Cohen, M. Lisanti, H. K. Lou, and S. Mishra-Sharma, "LHC Searches for Dark Sector Showers," *JHEP* **11** (2017) 196, `arXiv:1707.05326 [hep-ph]`.

[9] M. Park and M. Zhang, "Tagging a jet from a dark sector with jet-substructures at colliders," *Phys. Rev. D* **100** no. 11, (2019) 115009, `arXiv:1712.09279 [hep-ph]`.

[10] T. Cohen, J. Doss, and M. Freytsis, "Jet Substructure from Dark Sector Showers," *JHEP* **09** (2020) 118, `arXiv:2004.00631 [hep-ph]`.

[11] D. Kar and S. Sinha, "Exploring jet substructure in semi-visible jets," *SciPost Phys.* **10** no. 4, (2021) 084, `arXiv:2007.11597 [hep-ph]`.

[12] CMS Collaboration, A. Tumasyan *et al.*, "Search for resonant production of strongly coupled dark matter in proton-proton collisions at 13 TeV," *JHEP* **06** (2022) 156, `arXiv:2112.11125 [hep-ex]`.

[13] ATLAS Collaboration, G. Aad *et al.*, "Search for non-resonant production of semi-visible jets using Run 2 data in ATLAS," `arXiv:2305.18037 [hep-ex]`.

[14] E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer, and A. Mück, "Casting a graph net to catch dark showers," *SciPost Phys.* **10** no. 2, (2021) 046, `arXiv:2006.08639 [hep-ph]`.

[15] C.-T. Lu, H. Lv, W. Shen, L. Wu, and J. Zhang, "Probing Dark QCD Sector through the Higgs Portal with Machine Learning at the LHC," `arXiv:2304.03237 [hep-ph]`.

[16] T. Finke, M. Krämer, M. Lipp, and A. Mück, "Boosting mono-jet searches with model-agnostic machine learning," *JHEP* **08** (2022) 015, `arXiv:2204.11889 [hep-ph]`.

[17] 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 [hep-ph]`.

[18] P. Schwaller, D. Stolarski, and A. Weiler, "Emerging Jets," *JHEP* **05** (2015) 059, `arXiv:1502.05409 [hep-ph]`.

[19] CMS Collaboration, A. M. Sirunyan *et al.*, "Search for new particles decaying to a jet and an emerging jet," *JHEP* **02** (2019) 179, `arXiv:1810.10069 [hep-ex]`.

[20] M. Farina, Y. Nakai, and D. Shih, "Searching for New Physics with Deep Autoencoders," *Phys. Rev. D* **101** no. 7, (2020) 075021, `arXiv:1808.08992 [hep-ph]`.

[21] T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson, "QCD or What?," *SciPost Phys.* **6** no. 3, (2019) 030, `arXiv:1808.08979 [hep-ph]`.

[22] T. Cheng, J.-F. Arguin, J. Leissner-Martin, J. Pilette, and T. Golling, "Variational autoencoders for anomalous jet tagging," *Phys. Rev. D* **107** no. 1, (2023) 016002, `arXiv:2007.01850 [hep-ph]`.

[23] 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 [hep-ph]`.

[24] J. Batson, C. G. Haaf, Y. Kahn, and D. A. Roberts, "Topological Obstructions to Autoencoding," *JHEP* **04** (2021) 280, `arXiv:2102.08380 [hep-ph]`.

[25] 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** no. 7, (2021) 617, `arXiv:2104.02092 [hep-ph]`.

[26] F. X. Yu, K. Choromanski, S. Kumar, T. Jebara, and S.-F. Chang, "On learning from label

proportions," `arXiv:1402.5902 [stat.ML]`.

[27] N. Quadrianto, A. J. Smola, T. S. Caetano, and Q. V. Le, "Estimating labels from label proportions," *Journal of Machine Learning Research* **10** no. 82, (2009) 2349–2374. `http://jmlr.org/papers/v10/quadrianto09a.html`.

[28] L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman, "Weakly Supervised Classification in High Energy Physics," *JHEP* **05** (2017) 145, `arXiv:1702.00414 [hep-ph]`.

[29] T. Cohen, M. Freytsis, and B. Ostdiek, "(Machine) Learning to Do More with Less," *JHEP* **02** (2018) 034, `arXiv:1706.09451 [hep-ph]`.

[30] 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 [hep-ph]`.

[31] 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** no. 1, (2018) 011502(R), `arXiv:1801.10158 [hep-ph]`.

[32] J. H. Collins, K. Howe, and B. Nachman, "Extending the search for new resonances with machine learning," *Phys. Rev. D* **99** no. 1, (2019) 014038, `arXiv:1902.02634 [hep-ph]`.

[33] ATLAS Collaboration, G. Aad *et al.*, "Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector," *Phys. Rev. Lett.* **125** no. 13, (2020) 131801, `arXiv:2005.02983 [hep-ex]`.

[34] 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 [hep-ph]`.

[35] G. Choudalakis, "On hypothesis testing, trials factor, hypertests and the BumpHunter," in *PHYSTAT 2011*. 1, 2011. `arXiv:1101.0390 [physics.data-an]`.

[36] ATLAS Collaboration, G. Aad *et al.*, "Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *JHEP* **03** (2020) 145, `arXiv:1910.08447 [hep-ex]`.

[37] ATLAS Collaboration, G. Aad *et al.*, "Measurement of the flavour composition of dijet events in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," *Eur. Phys. J. C* **73** no. 2, (2013) 2301, `arXiv:1210.0441 [hep-ex]`.

[38] L. Carloni and T. Sjostrand, "Visible Effects of Invisible Hidden Valley Radiation," *JHEP* **09** (2010) 105, `arXiv:1006.2911 [hep-ph]`.

[39] ATLAS Collaboration, M. Aaboud *et al.*, "Search for diboson resonances with boson-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Phys. Lett. B* **777** (2018) 91–113, `arXiv:1708.04445 [hep-ex]`.

[40] CMS Collaboration, A. M. Sirunyan *et al.*, "Search for massive resonances decaying into WW, WZ or ZZ bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV," *JHEP* **03** (2017) 162, `arXiv:1612.09159 [hep-ex]`.

[41] K. Benkendorfer, L. Le Pottier, and B. Nachman, "Simulation-assisted decorrelation for resonant anomaly detection," *Phys. Rev. D* **104** no. 3, (2021) 035003, `arXiv:2009.02205 [hep-ph]`.

[42] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "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 [hep-ph]`.

[43] NNPDF Collaboration, R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo, "Parton distributions with QED corrections," *Nucl. Phys. B* **877** (2013) 290–320, `arXiv:1308.0598 [hep-ph]`.

[44] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.* **191** (2015) 159–177, `arXiv:1410.3012 [hep-ph]`.

[45] DELPHES 3 Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment," *JHEP* **02** (2014) 057, `arXiv:1307.6346 [hep-ex]`.

[46] CMS Collaboration, S. Chatrchyan *et al.*, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *JINST* **9** no. 10, (2014) P10009, `arXiv:1405.6569 [physics.ins-det]`.

[47] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP* **04** (2008) 063, `arXiv:0802.1189 [hep-ph]`.

[48] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting," *J. Phys. G* **34** (2007) N343.

[49] W. Waltenberger, "Rave—a detector-independent toolkit to reconstruct vertices," *IEEE Transactions on Nuclear Science* **58** no. 2, (2011) 434–444. `https://rave.hepforge.org/`.

[50] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

[51] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. `https://www.tensorflow.org/`.

[52] P. Virtanen *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods* **17** (2020) 261–272.