Uncovering latent jet substructure

Barry M. Dillon* and Darius A. Faroughy†
Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

Jernej F. Kamenik‡
Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia and
Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

We apply techniques from Bayesian generative statistical modeling to uncover hidden features in jet substructure observables that discriminate between different a priori unknown underlying short distance physical processes in multi-jet events. In particular, we use a mixed membership model known as Latent Dirichlet Allocation to build a data-driven unsupervised top-quark tagger and $t\bar{t}$ event classifier. We compare our proposal to existing traditional and machine learning approaches to top jet tagging. Finally, employing a toy vector-scalar boson model as a benchmark, we demonstrate the potential for discovering New Physics signatures in multi-jet events in a model independent and unsupervised way.

Introduction. The use of jet substructure techniques in studying large area jets has played an important role in identifying hadronic decays of Higgs and electroweak gauge bosons in runs 1 and 2 of the LHC [1–4]. These techniques have also been used efficiently to tag jets arising from top quarks [5–15]. In the last few years, machine learning (ML) tools have extended the application of jet substructure in tagging jets at the LHC [16–32] through the use of Neural Networks (NNs) to process and ‘learn’ from vast amounts of training data. Since these approaches rely on theoretical predictions for pure signal and background training data sets (typically through Monte Carlo (MC) generators), they (a) are exposed to MC mismodeling of realistic events as reconstructed from real data and detectors; (b) require exact model knowledge of both expected signal and backgrounds. This limits their use in searches for a priori unknown new phenomena in LHC jet events.

There have been recent advances in unsupervised or semi-supervised ML techniques, based on NNs designed to be able to separate signal and background events in mixed samples, and could therefore be run directly on experimental data without the need for pure MC training samples, see e.g. refs. [33–38] and [39–43]. They rely on categorizing and comparing datasets with different expected signal and background admixtures or identifying anomalous events inside large datasets. While these approaches ameliorate the model dependence of fully supervised ML, they are still potentially susceptible to correlated systematics (i.e. detector) effects and/or subject to large look-elsewhere effects. In addition, they generally work best when applied on very large datasets. Consequently their performance may suffer when looking for effects in tails of distributions.

In this Letter, we outline a new technique to classify jets and events in situ within a single mixed event sample, using tools developed in a branch of ML called generative statistical modeling [44, 45]. Developed primarily to identify emergent themes in collections of documents, these models infer the hidden (or latent) structure of a document corpus using posterior Bayesian inference based on word and theme co-occurrence [46]. Translated into the language jet physics, one assumes that observable jet substructure histograms (words) in events (documents) are generated by drawing from pure distributions (themes) of varying proportions. This allows to construct so called statistical mixed membership models of jet substructure.† Furthermore assuming that each event is a mixture of only few pure distributions and that within each of these only few histogram bins have high co-occurrence, such models can be solved using techniques of Latent Dirichlet Allocation (LDA) [48]. Finally, with a trained model at hand, one can define robust parametric jet and event classifiers by inferring on the pure distribution proportions in tested events.

In the following we first present the main ingredients of our proposal in more detail. Then we discuss two proof of principle implementations based on benchmark examples: an unsupervised top quark jet tagger and $t\bar{t}$ event classifier, as well as an unsupervised new physics (NP) search strategy able to identify boosted neutral scalar bosons decaying to pairs of W’s (previously studied in Refs. [37, 38, 49]). We compare them to existing conventional and ML approaches and also outline possible further improvements and future directions.

Generative Bayesian Models of Jet Substructure. We start by considering the formation of a jet stemming from an initial hard seed, as a sequential combination of QCD showering (followed by fragmentation

* Electronic address:barry.dillon@ijs.si
† Electronic address:darius.faroughy@ijs.si
‡ Electronic address:jernej.kamenik@cern.ch

1 Similar techniques have been used recently in a semi-supervised way to reconstruct ‘pure’ quark and gluon jet observable distributions from mixed event samples [17].
and hadronization) and possibly massive particle decays. Next we assume that some relevant information on this intertwined sequence of processes can be recovered by looking at the clustering history of a jet-clustering algorithm. This is in fact the basis for many conventional taggers of massive jets [1, 8, 13].

Within this very simplified picture of jet formation and observation we can draw interesting parallels to so called mixed membership models describing generation of documents in the context of text analysis [48], or genotypes in population studies [50]. In particular, we assume that the observable distribution bins in a clustering profile are populated by drawing from a few ‘pure’ distributions - 

\textit{themes} - corresponding to different contributing physical processes. The likelihood of populating a certain distribution bin \( o \), given a theme \( t \) can then be described by a multinomial distribution \( p(o|t, \beta) \) (a multi-category generalization of the binomial distribution, where the number of categories is given by the number of bins in the distribution and is parametrized by a set of parameters \( \beta \)). In addition, we assume that the likelihood of a given theme contributing to any given event (and thus jet) \( p(t|\omega) \) is also described by some multinomial distribution (parametrized by variables \( \omega \), where the number of categories now corresponds to the number of themes. The \( \omega \)'s themselves are drawn from a probability distribution \( p(\omega|\alpha) \), reflecting the theme proportions in the dataset and parametrized by the hyperparameter \( \alpha \). In this picture the themes \( \beta \) as well as theme proportions \( \omega \) are hidden variables reflecting the thematic structure of the studied event sample. With a given model, the probability that a certain event or jet is generated can be written as a compact expression in terms of the latent variables. For example, the likelihood of generating a jet \( j \) with a clustering history observables \( (o_1, o_2, \ldots, o_n) \) is just

\[
p(j|\alpha, \beta) = \int_\omega p(\omega|\alpha) \prod_{o \in j} \left( \sum_t p(t|\omega)p(o|t, \beta) \right) d\omega. \tag{1}
\]

Statistical models defined in this way are generative in that given the latent variables (themes and theme proportions) the best model will be the one that best reproduces a set of events, i.e. has the best generative power. Therefore, the task of finding the latent variables from a set of training events is specifically to invert the above expression and use the set of events to find the best fit for the latent variables. This can in fact be done using posterior Bayesian inference, i.e.

\[
p(a|x) \propto p(x|a) * p(a), \tag{2}
\]

where \( p(x|a) \) is the likelihood of observing \( x \) given a latent variable \( a \), while \( p(a) \) and \( p(a|x) \) are the prior and posterior distributions of the latent variable itself. The main insight here is that \( p(\omega|\alpha) \) in Eq. (1) is a conjugate prior to the multinomial likelihood \( p(t|\omega) \) and thus forms the multi category generalization of the beta distribution - the Dirichlet distribution. The model is thus called LDA and can be solved approximately \textit{(trained)} in an iterative manner using variational inference [45, 48] or Gibbs sampling [51].

A trained model consists of the latent variables inferred from the training data and the generative model used in constructing Eq. (1). Given a new event or jet we can use the trained model to determine the proportion \( (\omega_t(j)) \) of a theme \( t \) present in the new event or jet \( j \), by maximizing the likelihood function for the document being generated while keeping the theme distributions \( (\beta) \) fixed. In general we can do this for each theme \( t \), but in the case that the model contains only two themes (e.g. \( t = 0, 1 \)) it suffices to choose just one, since \( \sum_t \omega_t(j) = 1 \). In this case we can define a simple classifier e.g. \( h(j) = \omega_1(j) \) based on the proportion of one of the themes in the jet or event. Since the goal is to efficiently discriminate signal from background and not vice versa, one needs to choose the appropriate theme. This can be done by comparing the theme histograms with expectations for pure (e.g. MC generated) signal and background or by directly inferring (running the classifier) on such pure samples. We demonstrate both approaches in the following examples.

\textbf{Unsupervised top tagger.} Our first proof of principle example is a tagger discriminating between boosted hadronically decaying top quarks and QCD jets. Working with a single mixed (tt and QCD) multi-jet event sample we first need to construct the relevant jet substructure observable histograms \( (o) \). We do this by clustering the jets in an event using the Cambridge-Aachen (CA) [52, 53] algorithm with a large radius \( R \). We then proceed to uncluster the jets by reversing each step in the clustering, iteratively separating each (sub)jet into two objects \( j_0 \to j_1 j_2 \). Ordering the subjets by their invariant mass \( m_{j_1} > m_{j_2} \) (and following the standard approach of refs. [1, 3]), we define the relevant clustering observables at each clustering step as

\[
o_{j_0} = \{ m_{j_0}, \frac{m_{j_1}}{m_{j_0}}, \frac{m_{j_2}}{m_{j_0}}, \min(p_{T,1}^2, p_{T,2}^2), \Delta R_{1,2}^2 \}, \tag{3}
\]

where \( p_{T,i} \) is the transverse momentum of a given object \( j_i \) and \( \Delta R_{1,2}^2 = (\phi_1 - \phi_2)^2 + (\eta_1 - \eta_2)^2 \) is the so called planar distance between \( j_1 \) and \( j_2 \) (\( \phi \) and \( \eta \) being the azimuthal angle and pseudorapidity of \( j_i \), respectively). The declustering step is then iteratively repeated on both \( j_1, j_2 \). The procedure is terminated once \( m_{j_0} < m_{\text{min}} \), where \( m_{\text{min}} \) is an algorithm parameter, which we choose to lie below the lowest massive resonance of interest. In the case of the top tagger, we fix \( m_{\text{min}} = 30 \text{ GeV} \ll m_W \), but have checked that lowering this threshold by a factor of a few does not significantly affect the results. The output of such a procedure is a (typically a rather sparse) four-dimensional histogram of \( o_{j} \) which can be defined either \textit{per jet} or \textit{even per event}. After mapping individual histogram bins into \textit{words}, we feed individual jets or events as \textit{documents} into an LDA implementation using the software
package Gensim [54, 55], fixing the number of themes to two (\(\omega_{0,1}\)). Further technical details of the required binning and mapping of data onto (one-dimensional) text vocabularies compatible with Gensim, as well as a detailed analysis of the convergence of the algorithm when applied on sparse jet substructure data will be presented elsewhere [56]. Here we only focus on the consistency and stability of the resulting trained models. For this purpose we use the \(k\)-folding method with \(k = 10\). This involves splitting the training data into \(k\) different mutually exclusive blocks and then running the training \(k\) times on event samples built from \(k - 1\) blocks, with the combination changing on each training run. The performance of the tagger is tested on events or jets from the remaining block.

In order to evaluate the performance of the tagger and compare it to existing methods, we construct a receiver operating characteristic (ROC) curve for our tagger. This is the only step where one needs to rely on access to pure samples (either MC generated or pre-tagged in some other way using observables orthogonal to \(o_j\)). In particular, we construct the ROC curve by performing the classification on such pure samples while continuously varying the threshold of the theme proportion defining the classifier \(h(j)\). This is done for all \(k\) sets of results and we calculate the median mis-tag rate (\(\varepsilon_b\)) for each signal efficiency (\(\varepsilon_s\)), as well as the mean absolute deviation of the mis-tag rate to evaluate the stability and consistency of the tagger.

Our training samples for the QCD di-jet background and the (hadronic) \(tt\) signal both consist of \(\approx 84,000\) 13 TeV \(pp\) collision events, where the final state particles are clustered into \(R = 1.5\) CA jets with \(p_T\) in the range \([350, 450]\) GeV. The samples are generated using aMC@NLO 2.6.1 [57] interfaced with Pythia 8.2 [58] for showering and hadronization, while jet clustering is performed using FastJet 3.2.0 [59]. Note that no grooming is performed on the jets. We have also checked explicitly that applying jet (sub)cluster energy smearing consistent with the parametric fast detector simulation of ATLAS implemented in Delphes 3.4.1 [60] has no significant effect on our results.

We train the top tagger on four test cases: supervised, and unsupervised mixed samples with \(S/B = 1, 1/9, 1/99\). In the supervised case we collapse the pure samples into single documents such that they are processed by the algorithm in a single block, essentially providing the labelling of the data required in supervised algorithms. For the different \(S/B\) ratios each jet or event is represented by a single document. However, we inform the tagger to search for certain \(S/B\) ratios by setting the hyperparameters of the Dirichlet distribution accordingly, i.e. \(\alpha = [0.5, 0.5], [0.9, 0.1],\) and \([0.99, 0.01]\). Note that these may not be the optimal choices, but they are based on the intuition from the values of \(S/B\) and give a useful parameterization to demonstrate the performance of the algorithm. We also stress that \(O(1)\) variations in \(\alpha\) have only a small effect on the performance of the algorithm provided that the hierarchy in the elements of \(\alpha\) approximately reflect the \(S/B\) ratio, and that the elements are smaller than one. More details on the dependence of the algorithm on these hyperparameters, and how to determine their optimal values without prior knowledge of the \(S/B\) ratios, will be presented elsewhere [56].

In Fig. 1 (upper panel) we plot the ROC curves for our top jet taggers, where separate documents are represented by individual jets, and compare these to various supervised taggers in the literature [8, 22, 23]. We see that the taggers perform well and with relatively small variance, with the supervised tagger performing the best. An interesting observation is that at high background rejection rates (\(1/\varepsilon_b \gg O(\text{few})\)) the taggers trained on smaller \(S/B\) perform slightly better than the
tagger trained on the $S/B = 1$ sample, although the differences are comparable to the estimated uncertainties. This is essentially because the algorithm is designed to discern features in the jet substructure, which are subsequently used to tag jets and events. In the supervised and $S/B = 1$ case, the algorithm discovers features in fat jets where fat jets both near $m_{j0} \sim m_t$ and $m_{j0} \sim m_W$ (see the right plot in Fig. 2), while in the lower $S/B$ cases the algorithm is only able to identify $m_{j0} \sim m_t$ as relevant. On the other hand, lower $m_{j0}$ regions generically feature more prominently in QCD jets (see left plot in Fig. 2). Thus, while a very accurate determination of the features near $m_{j0} \sim m_W$ in the supervised case helps the performance of the tagging algorithm, the worse resolution in the unsupervised $S/B = 1$ case leads to worse tagging performance compared to lower $S/B$ examples. We see that the performance of the unsupervised taggers is comparable to the original JH top tagger [8], although it falls short in comparison to the others. We note that the observables we use mostly match those used in the JH top tagger, hence the similar performance is indeed encouraging.

In Fig. 1 (lower panel) we plot the ROC curves for our $tt$ event classifiers, where a single document now contains all jets within the selected $p_T$ region in an event, and again compare these to the top jet taggers in the literature. To make the comparison with other taggers fair, we re-scale those results by defining an event tagging efficiency ($\epsilon_\varepsilon$) in terms of the jet tagging efficiency ($\epsilon_j$) and the fraction of events in our pure samples with one ($f_1$) and two ($f_2$) jets passing the selection cuts. This means in practice that tagging an event as $tt$ requires at least one jet in the event to be tagged as a top jet. The ROC curves do not change significantly under this re-scaling, instead the points move along a trajectory towards higher efficiencies approximately equal to that of the ROC curve for jet tagging. We see again that the classifier performs very well in all cases, performing as well as the JH top tagger even for low $S/B$.

We observe that the LDA algorithm performs relatively better when characterizing and tagging events than jets, mainly due to the larger amount of substructure (words) in each document. With more data per document it is easier for the algorithm to identify co-occurrences between the different features shared by jets in the same event. For this reason it is also easier for the trained model to infer the correct thematic structure from events, than from jets.

The themes discovered by the unsupervised training algorithm contain valuable information about the substructure of the events or jets. In Fig. 2 we plot the substructure probability distributions of the two themes discovered by the top tagger (with $S/B = 1$) projected onto the plane of $m_{j0}$ and $m_{j1}/m_{j0}$. We observe that while the distribution on the left-hand side plot (the “QCD” theme) is fairly unremarkable (mostly monotonic and smooth) and peaks towards $(m_j \to 0, m_{j1}/m_{j0} \to 1)$, the theme on the right-hand side plot (the “tt” theme) clearly exhibits a heavily weighted feature at both $m_{j0} \sim m_t$ and $m_{j0} \sim m_W$, even identifying the $W$ subjet arising from the decay of the top quark within the jet resulting in a mass drop of $m_{j1}/m_{j0} \sim m_W/m_t \simeq 0.45$. On the other hand, the broad $m_{j1}/m_{j0} \sim 0.2 \gtrsim 0$ feature at $m_{j0} \sim m_W$ is expected due to the fact that the mass drop is defined with the heaviest daughter subjet in the numerator thus skewing the $m_{j1}/m_{j0}$ distribution away from zero.

**Unsupervised NP search.** As a second example, we consider a NP model [61, 62] containing a heavy $W'$ boson plus a heavy scalar $\phi$. Signal events thus consist of resonant $W'$ production (at $m_{W'} = 3$ TeV), followed by $W' \to W\phi$ decays (where we choose $m_\phi = 400$ GeV $\ll m_{W'}$ such that both the $W$ and the $\phi$ coming from $W'$ decays are boosted). Finally, the scalar further decays as $\phi \to W^+W^-$. Using the same event generation, jet clustering/de-clustering procedure, observable basis $o_j$, and the same LDA tagging algorithm as before, we apply our procedure to the all-hadronic final state of this NP process in a region dominated by QCD background. The same model has been previously studied using the unsupervised ML approach called *classification without labels* (CWoLa) [37, 38]. It is based on mixed sample classification using phase space regions with vastly different $S/B$ ratios processed by deep NNs. In order to

---

2 We have checked that the fractions of events with zero or more

---

than two jets passing the selection cuts are negligible.
quantitatively compare our results to CWoLa, our signal and background event samples mirror directly those in Ref. [38]. In particular, we consider just the signal region, $2730 \leq m_{jj} \leq 3189$ GeV, and cut jets with $p_T$ below 400 GeV. The 30 GeV cut on the subjet invariant mass is also applied, just as in the top tagger case. After the selection cuts we work with $\sim 60,000$ events in both the signal and background samples. We train three different taggers; a supervised tagger, and two taggers with $S/B = 1.1 \times 10^{-2}$ and $5.8 \times 10^{-3}$. The Dirichlet hyperparameters $\alpha$ are chosen in the same way as in the previous section, i.e. $\alpha = [0.5, 0.5], [0.989, 0.011]$, and $[0.942, 0.058]$. To evaluate the robustness of the taggers we again employ the $k$-folding procedure with $k=10$.

In the upper plot of Fig. 3 we show the ROC curves for our taggers and compare the results to those from CWoLa [38]. We see that in most of the parameter space the LDA-based tagger outperforms the CWoLa tagger, most notably at high signal efficiencies.

In the lower plot of Fig. 3 we also show the probability distributions of the discovered themes in the plane of $m_{j_0}$ and $m_{j_1}/m_{j_0}$ for the LDA model trained on event samples with $S/B = 1.1 \times 10^{-2}$. Features in the subjet mass at $m_{j_0} \sim m_W$ and at $m_{j_0} \sim m_\phi$ are clearly discernible in one of the themes (the “$\phi W$” theme), as well as mass drops related to the decays of the heavy scalar and the $W$ bosons.

**Conclusions.** We have demonstrated a new unsupervised ML technique for disentangling signal and background events in mixed samples by identifying features in jet substructure observables that differentiate between the two. To do so we have mapped jet substructure distributions onto a LDA model, a generative probabilistic model (mixed membership model) widely used in Bayesian statistics approaches to unsupervised ML. Assuming that the kinematic observable distributions within jets or events are sampled from a fixed set of (latent) themes, LDA can learn the thematic structure that most likely generated the observed data (the later being either in the form of reconstructed real LHC events or un-labeled MC-generated samples). Furthermore, we have shown that the learned structure from a two-theme LDA model can be used to build unsupervised jet taggers or event classifiers that efficiently discriminate between signal and background in previously unseen data.

As a first example we have trained a two-theme LDA model on MC-generated event samples consisting of different mixtures of $pp \rightarrow t\bar{t}$ and QCD di-jet events. Our results show that the top-jet taggers and $t\bar{t}$ event classifiers constructed from the discovered themes have a very good discrimination power when applied to previously unseen pure samples, even if trained on data with $S/B$ ratios as low as 1%. Our results are in some cases comparable even with fully supervised taggers in the literature. In addition we have explored the viability of LDA discovering NP phenomena in multi-jet events. Using a benchmark NP (vector $W'$ - scalar $\phi$) model we have studied $pp \rightarrow W' \rightarrow \phi W \rightarrow W W W'$ with hadronically decaying $W$ bosons and a (boosted) new scalar $\phi$ with mass $m_\phi \ll m_{W'}$. The resulting LDA event classifiers from training samples with $S/B$ as low as a few per-mille, when applied to pure samples, produce excellent signal efficiencies and QCD rejection rates that can outperform other existing approaches.

Besides being a fully unsupervised ML technique, one advantage of performing LDA on jet clustering history observables, is the possibility of interpreting the thematic structure discovered by the model from the data. In both examples presented here, the features in the probability distributions over the kinematical observables of the two uncovered themes match to a high degree the expected features of the underlying hard processes - hadronic decays of top-quarks (or $\phi \rightarrow W^+ W^-$) and the QCD back-
aim of performing unsupervised data-driven searches for

multiple backgrounds together with the signal. In future work we will also detail how these techniques can be employed as part of a broad search strategy for new phenomena in multi-jet invariant mass spectra with the aim of performing unsupervised data-driven searches for NP at high $p_T$.

ACKNOWLEDGMENTS

We thank Jasna Urbančič, Erik Novak and Klemen Kenda for initial involvement in the project as well as Jack Collins for generously providing the $W' - \phi$ NP model implementation for use in aMC@NLO. We also thank Česár A. Ojeda and Bryan Zaldivar for useful discussions. DAF is supported by the Young Researchers Programme of the Slovenian Research Agency under the grant No. 37468. JFK and BMD acknowledge the financial support from the Slovenian Research Agency (research core funding No. P1-0035 and J1-8137).

[1] J. M. Butterworth, A. R. Davison, M. Rubin, G. P. Salam, Jet substructure as a new Higgs search channel at the LHC, Phys. Rev. Lett. 100 (2008) 242001. arXiv:0802.2470, doi:10.1103/PhysRevLett.100.242001.

[2] J. M. Butterworth, B. E. Cox, J. R. Forshaw, $W W$ scattering at the CERN LHC, Phys. Rev. D65 (2002) 096014. arXiv:hep-ph/0201098, doi:10.1103/PhysRevD.65.096014.

[3] J. M. Butterworth, J. R. Ellis, A. R. Raklev, Reconstruction of particle mass spectra using hadronic decays, JHEP 05 (2007) 033. arXiv:hep-ph/0702150, doi:10.1088/1126-6708/2007/05/033.

[4] Y. Cui, Z. Han, M. D. Schwartz, W-jet Tagging: Optimizing the Identification of Boosted Hadronically-Decaying W Bosons, Phys. Rev. D83 (2011) 074023. arXiv:1012.2077, doi:10.1103/PhysRevD.83.074023.

[5] W. Skiba, D. Tucker-Smith, Using jet mass to discover vector quarks at the LHC, Phys. Rev. D75 (2007) 115010. arXiv:hep-ph/0701247, doi:10.1103/PhysRevD.75.115010.

[6] B. Holdom, t-prime at the LHC: The Physics of discovery, JHEP 03 (2007) 063. arXiv:hep-ph/0702037, doi:10.1088/1126-6708/2007/03/063.

[7] M. Gerbush, T. J. Khoo, D. J. Phalen, A. Pierce, D. Tucker-Smith, Color-octet scalars at the CERN LHC, Phys. Rev. D77 (2008) 095003. arXiv:0710.3133, doi:10.1103/PhysRevD.77.095003.

[8] D. E. Kaplan, K. Rehermann, M. D. Schwartz, B. Tweedie, Top Tagger: A Method for Identifying Boosted Hadronically Decaying Top Quarks, Phys. Rev. Lett. 101 (2008) 142001. arXiv:0806.0848, doi:10.1103/PhysRevLett.101.142001.

[9] L. G. Almeida, S. J. Lee, G. Perez, I. Sung, J. Virzi, Top Jets at the LHC, Phys. Rev. D79 (2009) 074012. arXiv:0810.0934, doi:10.1103/PhysRevD.79.074012.

[10] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung, J. Virzi, Substructure of high-$p_T$ Jets at the LHC, Phys. Rev. D79 (2009) 074017. arXiv:0807.0234, doi:10.1103/PhysRevD.79.074017.

[11] L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, I. Sung, Template Overlap Method for Massive Jets, Phys. Rev. D82 (2010) 054034. arXiv:1006.2035, doi:10.1103/PhysRevD.82.054034.

[12] M. Backović, J. Juknevich, TemplateTagger v1.0.0: A Template Matching Tool for Jet Substructure, Comput. Phys. Commun. 185 (2014) 1322–1338. arXiv:1212.2978, doi:10.1016/j.cpc.2013.12.018.

[13] T. Plehn, G. P. Salam, M. Spannowsky, Fat Jets for a Light Higgs, Phys. Rev. Lett. 104 (2010) 111801. arXiv:0910.5472, doi:10.1103/PhysRevLett.104.111801.

[14] T. Plehn, M. Spannowsky, M. Takeuchi, D. Zerwas, Stop Reconstruction with Tagged Tops, JHEP 10 (2010) 078. arXiv:1006.2833, doi:10.1007/JHEP10(2010)078.

[15] D. E. Soper, M. Spannowsky, Finding top quarks with shower deconstruction, Phys. Rev. D87 (2013) 054012. arXiv:1211.3140, doi:10.1103/PhysRevD.87.054012.

[16] A. J. Larkoski, I. Moul, B. Nachman, Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning arXiv:1709.04464.

[17] J. Cogan, M. Kagan, E. Strauss, A. Schwartzman, Jet-Images: Computer Vision Inspired Techniques for Jet Tagging, JHEP 02 (2015) 118. arXiv:1407.5675, doi:10.1007/JHEP02(2015)118.

[18] L. G. Almeida, M. Backović, M. Cliche, S. J. Lee, J. M. Butterworth, A. Schwartzman, Jet-images ? deep learning edition, arXiv:1511.05190.

[19] J. Cogan, M. Kagan, E. Strauss, A. Schwartzman, Jet-Images: Computer Vision Inspired Techniques for Jet Tagging, JHEP 02 (2015) 118. arXiv:1407.5675, doi:10.1007/JHEP02(2015)118.

[20] L. G. Almeida, M. Backović, M. Cliche, S. J. Lee, M. Perelstein, Playing Tag with ANN: Boosted Top Identification with Pattern Recognition, JHEP 07 (2015) 086. arXiv:1501.05968, doi:10.1007/JHEP07(2015)086.

[21] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, A. Schwartzman, Jet-images ? deep learning edition, JHEP 07 (2016) 069. arXiv:1511.05190, doi:10.1007/JHEP07(2016)069.

[22] P. Baldi, K. Bauer, C. Eng, P. Sadowski, D. Whiteson, Jet Substructure Classification in High-Energy Physics with Deep Neural Networks, Phys. Rev. D93 (9) (2016) 094034. arXiv:1603.09349, doi:10.1103/PhysRevD.93.094034.

[23] J. Barnard, E. N. Dawe, M. J. Dolan, N. Rajcic, Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks, Phys. Rev. D95 (1) (2017) 014018. arXiv:1609.00607, doi:10.1103/PhysRevD.95.014018.

[24] G. Kasiezcka, T. Plehn, M. Russell, T. Schell, Deep-learning Top Taggers or The End of QCD?, JHEP
[54] K. Datta, A. Larkoski, How Much Information is in a Jet?, JHEP 06 (2017) 073. arXiv:1704.02124.
[27] K. Datta, A. Larkoski, Novel Jet Observables from Machine Learning, JHEP 03 (2018) 086. arXiv:1710.01035.
[28] A. Andreassen, I. Feige, C. Frye, M. D. Schwartz, JUNIPR: a Framework for Unsupervised Machine Learning in Particle Physics, Eur. Phys. J. C 79 (2) (2019) 102. arXiv:1804.09720, doi:10.1140/epjc/s10052-019-6607-9.
[29] S. Macaluso, D. Shih, Pulling Out All the Tops with Computer Vision and Deep Learning, JHEP 10 (2018) 121. arXiv:1803.00107, doi:10.1007/JHEP10(2018)121.
[30] K. Datta, A. Larkoski, B. Nachman, Automating the Construction of Jet Observables with Machine Learning, arXiv:1902.07180.
[31] L. M. Dery, B. Nachman, F. Rubbo, A. Schwartzman, Weakly Supervised Classification in High Energy Physics, JHEP 05 (2017) 145. arXiv:1702.00414, doi:10.1007/JHEP05(2017)145.
[32] E. M. Metodiev, B. Nachman, J. Thaler, Classification without labels: Learning from mixed samples in high energy physics, JHEP 10 (2017) 174. arXiv:1708.02949, doi:10.1007/JHEP10(2017)174.
[33] P. T. Komiske, E. M. Metodiev, B. Nachman, M. D. Schwartz, Learning to classify from impure samples with high-dimensional data, Phys. Rev. D98 (1) (2018) 011502. arXiv:1801.10158, doi:10.1103/PhysRevD.98.011502.
[34] T. Cohen, M. Freytsis, B. Ostdiek, (Machine) Learning to Do More with Less, JHEP 02 (2018) 034. arXiv:1706.09451, doi:10.1007/JHEP02(2018)034.
[35] J. H. Collins, K. Howe, B. Nachman, Anomaly Detection for Resonant New Physics with Machine Learning, Phys. Rev. Lett. 121 (24) (2018) 241803. arXiv:1805.02664, doi:10.1103/PhysRevLett.121.241803.
[36] J. H. Collins, K. Howe, B. Nachman, Extending the search for new resonances with machine learning, Phys. Rev. D99 (1) (2019) 014038. arXiv:1902.02634, doi:10.1103/PhysRevD.99.014038.
[37] J. A. Aguilar-Saavedra, J. H. Collins, R. K. Mishra, A generic anti-QCD jet tagger, JHEP 11 (2017) 163. arXiv:1709.01087, doi:10.1007/JHEP11(2017)163.
[38] J. Hajer, Y.-Y. Li, T. Liu, H. Wang, Novelty Detection Meets Collider Physics, arXiv:1807.10261.
[39] T. Heimel, G. Kasieczka, T. Plehn, J. M. Thompson, QCD or What?, SciPost Phys. 6 (2019) 030. arXiv:1808.08979, doi:10.21468/SciPostPhys.6.3.030.
[40] M. Farina, Y. Nakai, D. Shih, Searching for New Physics with Deep Autoencoders, arXiv:1808.08992.
[41] O. Cerri, T. Q. Nguyen, M. Pierini, M. Spiropulu, J.-R. Vlimant, Variational Autoencoders for New Physics Mining at the Large Hadron Collider, arXiv:1811.10276.
[42] S. Dearwerther, S. T. Dumas, G. W. Furnas, T. K. Landauer, R. Harshman, Indexing by latent semantic analysis, JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE 41 (6) (1990) 391–407.
[43] T. Hofmann, Probabilistic latent semantic analysis (1999) 289–296.
[44] K. Nigam, A. K. McCallum, S. Thrun, T. Mitchell, Text classification from labeled and unlabeled documents using em (1999) 103–134.
[45] E. M. Metodiev, J. Thaler, Jet Topics: Disentangling Quarks and Gluons at Colliders, Phys. Rev. Lett. 120 (24) (2018) 241602. arXiv:1802.00008, doi:10.1103/PhysRevLett.120.241602.
[46] D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty, Latent dirichlet allocation, Journal of Machine Learning Research 3 (2003) 2003.
[47] K. Agashe, J. H. Collins, P. Du, S. Hong, D. Kim, R. K. Mishra, Detecting a Boosted Diboson Resonance, JHEP 11 (2018) 027. arXiv:1809.07334, doi:10.1007/JHEP11(2018)027.
[48] J. K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2) (2000) 945–959. arXiv:http://www.genetics.org/content/155/2/945.full.pdf.
URL http://www.genetics.org/content/155/2/945.
[49] T. L. Griffiths, M. Steyvers, Finding scientific topics, Proceedings of the National Academy of Sciences 101 (suppl 1) (2004) 5228–5235. arXiv:https://www.pnas.org/content/101/suppl_1/5228.full.pdf, doi:10.1073/pnas.0307752101.
URL https://www.pnas.org/content/101/suppl_1/5228.
[50] Y. L. Dokshitzer, G. D. Leder, S. Moretti, B. R. Webber, Better jet clustering algorithms, JHEP 08 (1997) 001. arXiv:hep-ph/9707323, doi:10.1088/1126-6708/1997/08/001.
[51] M. Wobisch, T. Wengler, Hadronization corrections to jet cross-sections in deep inelastic scattering, in: Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998–1999, 1998, pp. 270–279. arXiv:hep-ph/9907280.
[52] R. Rehãlek, P. Sojka, Software Framework for Topic Modelling with Large Corpora (2010) 45–50http://is.muni.cz/publication/88493/en.
[53] M. D. Hoffman, D. M. Blei, F. Bach, Online learning for latent dirichlet allocation (2010) 856–864. URL http://dl.acm.org/citation.cfm?id=2997189.2997285.
[54] B. Dillon, D. A. Faroughy, J. F. Kamenik, in preparation.
[55] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, 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, doi:10.1007/JHEP07(2014)079.

[58] T. Sjostrand, S. Mrenna, P. Z. Skands, A Brief Introduction to PYTHIA 8.1, Comput. Phys. Commun. 178 (2008) 852–867. arXiv:0710.3820, doi:10.1016/j.cpc.2008.01.036.

[59] M. Cacciari, G. P. Salam, G. Soyez, FastJet User Manual, Eur. Phys. J. C72 (2012) 1896. arXiv:1111.6097, doi:10.1140/epjc/s10052-012-1896-2.

[60] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, M. Selvaggi, DELPHES 3, A modular framework for fast simulation of a generic collider experiment, JHEP 02 (2014) 057. arXiv:1307.6346, doi:10.1007/JHEP02(2014)057.

[61] K. Agashe, P. Du, S. Hong, R. Sundrum, Flavor Universal Resonances and Warped Gravity, JHEP 01 (2017) 016. arXiv:1608.00526, doi:10.1007/JHEP01(2017)016.

[62] K. Agashe, J. H. Collins, P. Du, S. Hong, D. Kim, R. K. Mishra, Dedicated Strategies for Triboson Signals from Cascade Decays of Vector Resonances arXiv:1711.09920.