Machine learning-based jet and event classification at the Electron-Ion Collider with applications to hadron structure and spin physics

Kyle Lee,[a,b] James Mulligan,[a,c] Mateusz Płoskoń,[a] Felix Ringer[d,e,f,g] and Feng Yuan[a]

[a]*Nuclear Science Division, Lawrence Berkeley National Laboratory,*
Berkeley, California 94720, U.S.A.

[b]*Center for Theoretical Physics, Massachusetts Institute of Technology,*
Cambridge, MA 02139, U.S.A.

[c]*Physics Department, University of California,*
Berkeley, CA 94720, U.S.A.

[d]*C.N. Yang Institute for Theoretical Physics, Stony Brook University,*
Stony Brook, NY 11794, U.S.A.

[e]*Department of Physics and Astronomy, Stony Brook University,*
Stony Brook, NY 11794, U.S.A.

[f]*Department of Physics, Old Dominion University,*
Norfolk, VA 23529, U.S.A.

[g]*Thomas Jefferson National Accelerator Facility,*
Newport News, VA 23606, U.S.A.

E-mail: kylel@mit.edu, james.mulligan@berkeley.edu, mploskon@lbl.gov, fmringer@jlab.org, fyuan@lbl.gov

ABSTRACT: We explore machine learning-based jet and event identification at the future Electron-Ion Collider (EIC). We study the effectiveness of machine learning-based classifiers at relatively low EIC energies, focusing on (i) identifying the flavor of the jet and (ii) identifying the underlying hard process of the event. We propose applications of our machine learning-based jet identification in the key research areas at the future EIC and current Relativistic Heavy Ion Collider program, including enhancing constraints on (transverse momentum dependent) parton distribution functions, improving experimental access to transverse spin asymmetries, studying photon structure, and quantifying the modification of hadrons and jets in the cold nuclear matter environment in electron-nucleus collisions. We establish first benchmarks and contrast the estimated performance of flavor tagging at the EIC with that at the Large Hadron Collider. We perform studies relevant to aspects of detector design including particle identification, charge information, and minimum transverse momentum capabilities. Additionally, we study the impact of using full event information instead of using only information associated with the identified jet. These methods can be deployed either on suitably accurate Monte Carlo event generators, or, for several applications, directly on experimental data. We provide an outlook for ultimately connecting these machine learning-based methods with first principles calculations in quantum chromodynamics.

Contents

1   Introduction

The future Electron-Ion Collider (EIC) will map out the structure of nucleons and nuclei with unprecedented precision and allow for novel studies of quantum chromodynamics (QCD) including explorations of the mechanism of hadronization and the properties of cold nuclear matter [1]. The EIC center-of-mass energy of up to $\sqrt{s} = 141$ GeV will allow for jets to play a major role in the EIC science program. Jets are collimated sprays of particles observed in detectors that arise from the multiple soft and collinear emissions of highly energetic scattered quarks and gluons. The measured energy and the direction of a jet represent good proxies of the corresponding quantities at the level of quarks and gluons that initiate the observed jet. This close correspondence between partons and jets is expected to hold at the EIC [2, 3]. In this work, we explore the use of machine learning to classify different flavors of jets as well as the underlying hard processes in events using machine learning. Because of the close connection between jets and partons, accurate classification of jets as well as the underlying hard processes has important implications for the major scientific goals of the EIC. We will outline specific applications of these methods that we envision at the future EIC. Moreover, we expect that the tools and

applications discussed in this work are also relevant to the ongoing experimental program at the Relativistic Heavy Ion Collider (RHIC).

The rapid progress in artificial intelligence and machine learning over the last decade has led to various applications in nuclear and high energy physics, including in the areas of classification, generative modeling, regression, and inference [4–8]. In the context of the EIC, machine learning techniques have been proposed to determine kinematic variables in Deep Inelastic Scattering (DIS) [9, 10] and to extract quantum correlation functions such as parton distribution functions (PDFs), transverse momentum dependent distribution functions (TMDs), generalized parton distributions (GPDs), and fragmentation functions [11–15].

In high-energy collider physics, the classification of jets (quark vs. gluon, QCD vs. W/Z, proton-proton vs. heavy-ion) has been studied with increasing sophistication over the past years [16–32]. An important goal of these studies is to increase the sensitivity to potential signals of physics beyond the Standard Model. At the Large Hadron Collider (LHC), machine learning based jet taggers have succeeded in significantly outperforming traditional jet taggers [33], including architectures based on Convolutional Neural Networks [19, 34], deep sets [35], and transformer models [36]. Compared to tagging techniques based on traditional observables, algorithms based on supervised machine learning algorithms have both benefits and drawbacks. The principal benefit is the ability of the machine learning algorithm to take advantage of the full information at hand and thereby significantly outperform algorithms based on simple closed-form observables. The drawbacks in doing so are that the results can be difficult to interpret and connect to first principles since it is unclear what the machine is learning and that the (simulated or experimental) data used in the training process may contain biases. Numerous efforts have been taken to address the limitations of machine learning based algorithms, both to improve the interpretability of machine learned results and bring under control the biases of the training data set [33]. For example, in refs. [37–40] complete bases of jet substructure observables were introduced which span the phase space of emissions inside jets. This large set of perturbatively calculable observables can rival the performance of successful machine learning algorithms, which has allowed for insights into how machines learn and an improved understanding of their good performance. In this way, machine learning based algorithms have driven progress both in the performance of the classification tasks but also in pushing analytical approaches forward. A key question at the EIC is understanding how large an improvement in performance machine learning based tagging algorithms can provide, and to identify applications of machine learning based algorithms that can drive forward the physics goals of the EIC, as it has at the LHC.

In this work, we explore the application of machine learning-based classifiers at the comparatively low EIC energies. Typically, jet classification studies are carried out at LHC energies where the jet transverse momentum p_T^{jet} is $\mathcal{O}(100-1000\,\text{GeV})$. Instead, at the EIC, jets will be produced predominantly with p_T^{jet} of 10-30 GeV. Moreover, an important aspect of machine learning for jets is the sparsity of the data compared to typical tasks in computer vision. Due to the relatively low number of hadrons that make up EIC jets, we expect an increased level of sparsity compared to LHC jets. We will explore if the reduced energy

and increased sparsity of jets at the EIC affect the performance of machine learning based classifiers. Additionally, at the low EIC energies Monte Carlo event generators must rely on a larger amount of non-perturbative modeling, which can cause biases in the simulated data used to train machine learning classifiers. For the study of spin physics and cold nuclear matter effects, we will propose an approach that can mitigate this issue by training directly on experimental data.

We will explore both quark flavor jet tagging and quark vs. gluon jet tagging, as well as tagging the underlying hard process in the event. For jet flavor tagging, as examples we consider the binary classification tasks: u vs. d quark [41], ud vs. s and uds vs. c jet tagging and we compare to the jet charge [42, 43] as a reference. We expect that the flavor tagging of jets will be an important component to constrain collinear and transverse momentum dependent PDFs, which we discuss more below. For quark vs. gluon tagging and tagging the underlying hard process in events, we consider the classification of $qq, q\bar{q}$ vs. gg di-jet topologies as well as the classification of direct vs. resolved photon contributions to the photoproduction di-jet cross section. The latter provides a new opportunity to enhance constraints on the parton-in-photon PDF.

In addition to using the in-jet information to train the classifiers, we will investigate how the performance can be improved by not only making use of the particles inside the jet but also out-of-jet particles in the event. By making use of this additional information, we extend the jet classification task to the classification of the underlying hard process in the event. For typical applications that we foresee at the future EIC, the classification of the underlying hard process can improve measurements that are not necessarily limited to in-jet dynamics. Event-wide classification algorithms can help to improve the measurement of spin asymmetries and studies of cold nuclear matter. The machine-learned event-wide information can also be mapped to traditional observables like N-jettiness ratios [44] or the jet pull [45, 46].

In addition to baseline studies of the jet flavor tagging performance, we will gauge the importance of several aspects of detector design within the future detector setup at the EIC [1]. We will consider the role of particle identification (PID) information, charge information, and minimum transverse momentum capabilities on the jet flavor tagging performance. In these studies, which serve as a first step, we retain the ideal efficiency and PID capabilities for the detected particle species, leaving the implementation of a simulated detector response and the impact on the jet energy scale and resolution for future work. Nonetheless, by varying the minimum particle p_T of the particles input to the classifier as well as comparing the performance when PID or charge information are included, we will elucidate baseline considerations on the importance of reconstructing low p_T particles and reconstructing PID vs. charge information. For strange quark initiated jets, we additionally study the impact of identifying the weakly decaying strange hadrons from their decay products by comparing the jet flavor tagging performance when the classifier is supplied with the undecayed strange hadrons vs. their decay products.

We foresee several specific applications of machine learning based jet and event classification to the major physics goals of the EIC, several aspects of which we will discuss in further detail in section 2:

(i) Strengthening constraints on (transverse momentum dependent) PDFs. The flavor tagging of jets will be an important component to constrain collinear and transverse momentum dependent PDFs. For example, charm-tagged jets can increase the sensitivity to the (collinear) strange quark PDF in charged current events [47]. In addition, di-jet events with charm and anti-charm tagging can also help to constrain the gluon TMD, including the gluon Sivers function at the EIC [48]. Jet substructure observables have also been proposed to constrain the gluon PDF at the LHC [49]. New opportunities relevant to RHIC and the EIC include the gluon helicity distribution, the parton-in-photon PDF, spin-dependent TMD PDFs and fragmentation functions some of which we explore quantitatively in this work. In our studies, we find that machine learning-based classifiers outperform traditional observables like the jet charge and therefore we expect that machine learning can significantly enhance the constraints on PDFs. Machine learned classifiers provide an upper bound on the information content contained in the jet or event [38] and can be used to design closed-form observables using symbolic regression techniques that are calculable in perturbative QCD [36, 50, 51]. Additionally, machine-learned event-by-event classifiers may eventually be directly included in global analyses of quantum correlation functions like PDFs.

(ii) Enhancing the sensitivity to transverse single spin asymmetries. Transverse Single Spin Asymmetries (TSSAs) constitute some of the hallmark measurements at RHIC and the future EIC and they provide constraints on the spin structure of the proton. TSSAs are defined as the difference of cross sections where the incoming protons have different transverse spin ($\uparrow\downarrow$) orientations

$$A_{UT} = \frac{\mathrm{d}\sigma^\uparrow - \mathrm{d}\sigma^\downarrow}{\mathrm{d}\sigma^\uparrow + \mathrm{d}\sigma^\downarrow}. \tag{1.1}$$

However, it has generally been challenging to measure non-zero TSSAs, in particular, for those associated with jets [52]. Recently, the STAR collaboration used the jet charge as an additional measurement to increase the size of the asymmetry [53], which was also proposed in ref. [54]. In this paper, we propose that an enhancement of the TSSA signal

$$\max_\theta |A_{UT}(\theta)|, \tag{1.2}$$

can be achieved by including an additional machine-learned measurement, which is here given by the parameters θ. This can be achieved by formulating the regression task in eq. (1.2) as a classification problem of jets (or events) that are obtained in scattering processes with differently polarized protons ($\uparrow\downarrow$) in the initial state. By applying a classifier trained to distinguish jets in events with different initial spin orientations as an additional measurement, similar to the jet charge, larger spin asymmetries may be obtained, which can provide better constraints on the corresponding quantum correlation functions in global analyses.

(iii) Elucidating cold nuclear matter effects. One of the goals of the EIC is to achieve an understanding of the transport properties of nuclear matter such as the jet transport coefficient $\hat{q}$, which denotes the mean square of the momentum transfer between a propagating hard jet and the nuclear medium [55–59]. This can be achieved by comparing jet

observables in eA collisions to those in ep collisions, similar to the jet quenching program comparing AA and pp collisions at RHIC and the LHC. The entire basis for extracting such properties of nuclear matter is the difference between eA and ep observables. By training machine learning methods to distinguish these two classes of events, one can use interpretable machine learning methods to gain insight into the type of information responsible for these differences, and thereby make connections to calculable observables in perturbative QCD [50]. Additionally, by tagging quark and gluon jets separately, one can achieve a more detailed understanding of the jet quenching interaction. This has remained challenging in the AA jet quenching program [30, 60], and the eA program will offer a cleaner environment where such techniques may be more likely to succeed.

The remainder of this work is organized as follows. In section 2, we propose several applications of machine learning based jet identification we carry out in this paper to the scientific program of the EIC. In section 3, we discuss the event generation setup of our studies and present the different machine learning algorithms used in this work. In section 4 we present results for jet flavor classification at the EIC, and in section 5 we extend the classification to identify underlying hard processes in full events. In section 6, we draw conclusions and provide an outlook.

2 Machine learning applications to hadron structure and spin physics

In this section, we propose several applications of machine learning based jet classification algorithms to the scientific program of the EIC and the ongoing RHIC program. In section 2.1 we provide a high-level description of applications of jet tagging that can gain stronger constraints with machine learning methods. In section 2.2, we outline a proposal to enhance spin asymmetries with machine learning that is largely independent of model biases.

2.1 Maximizing jet flavor tagging performance

The measurement of longitudinal and transverse spin asymmetries provides constraints on the spin decomposition of the proton. Using QCD factorization, initial and final state spin effects can be disentangled within global analyses of the available data. As a representative example, we consider TSSAs where one of the incoming protons is transversely polarized, see eq. (1.1). These asymmetries are generally small due to the cancellation of different parton contributions with opposite sign. In particular, measurements of Sivers [61] and Collins [62] asymmetries are often close to zero due to large cancellations between different PDFs and fragmentation functions.

As discussed also in section 1, since these asymmetries are small, experimental measurements at RHIC and the future EIC are challenging. Due to the relatively large experimental uncertainties, the measured asymmetries are often small or consistent with zero. Measurements have been performed by STAR using di-jet correlations $p^\uparrow + p \rightarrow \text{dijets} + X$ [52, 53] as well as with single-inclusive measurements by PHENIX and STAR using pions $p^\uparrow + p \rightarrow \pi + X$, jets, open heavy-flavor mesons and photons [63–67]. Theory calculations corresponding to the di-jet measurements have also recently been performed [68–70].

The reason for the small size of the asymmetries is due to approximate cancellations which can be understood from momentum sum rules. In the following, we will consider the Schäfer-Teryaev sum rule [71, 72] and the Burkardt [73–75] sum rule that are satisfied by the Collins and Sivers functions, respectively. Both of these sum rules state that average transverse momentum should sum to zero when summed over either the outgoing hadron flavors (Collins) or incoming quark flavors (Sivers). We note that the derivation of these sum rules involves bare quark and gluon operators and it is therefore unknown how much the sum rules are violated due to renormalization. Nevertheless, they provide an intuitive understanding of the large cancellations between different quark flavors to first order.

First, we consider the Sivers function $f_{1T}^{\perp a}(x, \vec{k}_T^2)$, which describes the longitudinal x and transverse momentum k_T anisotropy of partons inside a transversely polarized proton. Here the superscript $a = q, \bar{q}, g$ denotes the parton inside the proton. Including appropriate prefactors and formally integrating over the transverse momentum dependence, we find

$$f_{1T}^{\perp(1)a}(x) = \int \mathrm{d}^2\vec{k}_T \frac{\vec{k}_T^2}{2M^2} f_{1T}^{\perp a}(x, \vec{k}_T^2) \,, \tag{2.1}$$

where M is the proton mass. The Burkardt sum rule for the Sivers function states that the following integral vanishes [74]

$$\sum_{a=q,\bar{q},g} \int_0^1 \mathrm{d}x f_{1T}^{\perp(1)a}(x) = 0 \,. \tag{2.2}$$

Under the assumption that the valence quark distributions dominate, the Burkardt sum rule leads to u and d-quark Sivers functions that have opposite signs and similar magnitudes. This expected behavior of the u and d quark Sivers functions has been confirmed by recent global analyses [76–78]. At the EIC, the Sivers function can be measured for example in lepton-jet correlations [79, 80]. In order to obtain larger spin asymmetries, and hence a greater sensitivity to the underlying Sivers function, we propose that machine learning-based classifiers can be included to isolate the contribution of different quark flavors.

Second, we consider the Collins fragmentation function $H_{1,h/q}^{\perp}(z, \vec{P}_\perp^2)$ as an example of spin-dependent dynamics in the final state where large flavor cancellations are expected. It describes the longitudinal z and transverse momentum $P_\perp$ distribution of a final-state hadron that originates from a transversely polarized parton. After integrating out the transverse momentum dependence, we find

$$H_{1,h/q}^{\perp(3)}(z) = \int \mathrm{d}^2\vec{P}_\perp \frac{\vec{P}_\perp^2}{M_h} H_{1,h/q}^{\perp}(z, \vec{P}_\perp^2) \,, \tag{2.3}$$

where M_h is the mass of the observed hadron. The Schäfer-Teryaev sum rule for the Collins function states that the integral over the longitudinal momentum fraction vanishes after we sum over all hadron species [71, 72]

$$\sum_h \int_0^1 \mathrm{d}z \, H_{1,h/q}^{\perp(3)}(z) = 0 \,. \tag{2.4}$$

The cancellation that results here from summing over all hadrons is typically avoided by measuring identified hadrons in the final states e.g. by measuring π^+ and π^- cross sections separately [66]. Nevertheless, there can be further cancellations, which can be seen as follows. For simplicity, we now assume that isospin symmetry holds and we assume that the light parton-to-pion fragmentation process dominates. In this case, only two fragmentation channels remain. The *favored* fragmentation functions are pion fragmentation functions for a valence parton u or d, i.e. $H_{1,\pi^+/u}^\perp$ or $H_{1,\pi^-/d}^\perp$, respectively, and the *unfavored* fragmentation functions are pion fragmentation functions of u or d that are not a valence parton, i.e. $H_{1,\pi^-/u}^\perp$ or $H_{1,\pi^+/d}^\perp$. Assuming that the fragmentation of light partons to pions dominates, this then implies that the favored and unfavored contributions cancel according to the Schäfer-Teryaev sum rule in eq. (2.4) for a given parton. If we now choose to measure an identified hadron, say π^+, we expect

$$\int_0^1 \mathrm{d}z \left(H_{1,\pi^+/u}^{\perp(3)}(z) + H_{1,\pi^+/d}^{\perp(3)}(z) \right) \approx 0 \,. \tag{2.5}$$

As the flavor of the jet corresponds to the flavor of the fragmenting parton, up to higher order corrections in QCD, this cancellation is relevant when we consider for example the distribution of identified hadrons inside jets [81–85]. In particular, one can study azimuthal asymmetries that involve the correlation of the transversity PDF and the Collins fragmentation function of the hadron inside the jet [86, 87]. Therefore, in order to avoid the cancellation in eq. (2.5), we propose that a machine learned classifier can be used to tag the flavor of the observed jet. Moreover, recently refs. [88, 89] proposed to measure spin asymmetries similar to the Collins asymmetry directly using jets instead of identified hadrons. Analogous sum rules as in eq. (2.4) apply that are expected to lead to small spin asymmetries. In order to address this problem, refs. [54, 88, 89] proposed the use of the jet charge as an additional measurement, which avoids large cancellations between different quark flavors. Here we also propose that the use of machine learning-based classification of the jet flavor can enhance the size of the asymmetry compared to more traditional observables.

We expect that various other measurements and science goals of the EIC and RHIC will greatly benefit from machine learned classifiers that can identify the jet flavor or the hard-scattering event. While some of them will be discussed in this work, we leave more detailed quantitative studies of the following topics for future work:

- Quark flavor and quark vs. gluon jet identification can help to improve the sensitivity to the longitudinally polarized gluon distribution Δg. In particular, it may be possible to distinguish the positive and negative solutions for Δg that were found in recent global analyses [90]. See also refs. [91–95] for recent discussions and experimental results.

- Quark vs. gluon jet classification may help to improve measurements of the gluon Sivers function at RHIC and the future EIC [79, 96].

- The techniques discussed here may also improve searches of physics beyond the Standard Model at the EIC [97–99]. For example, in ref. [100] jet charge-weighted TSSAs were proposed in this context.

- Exclusive / diffractive processes involving jets can provide constraints on GPDs and Wigner functions [101–103]. We expect that machine learning based classifiers may help to better pin down these higher-dimensional parton distribution functions along with knowledge about the exact kinematics of the di-jet events [104, 105].

2.2 Maximizing the size of spin asymmetries

In the previous section, we discussed several areas where machine learning-based jet and event flavor tagging can play an important role to support the EIC and RHIC science programs. We implicitly adopted a "UV definition" of the jet flavor. In this case, the flavor of a jet is defined as the hard parton that initiates the jet and it can directly be used in Monte Carlo event generators. The machine learning algorithm is then trained to recover the assigned flavor label from the IR physics, i.e. the hadrons that make up the jet [106]. There are theoretical ambiguities associated with this approach [107] and since the UV label of the jet flavor is not accessible experimentally, machine learning algorithms have to be trained on simulated data. This definition has been widely used for machine learning studies of jet classification at the LHC and various approaches have been developed to minimize the biases of this approach. For example, data-driven methods [108] and weakly supervised learning [109] have been introduced, which are tailored toward the physics goals at the LHC.

In the spin physics context, we propose an alternative approach to directly train the machine learning algorithms at the observable level. We expect that several of the science goals discussed in the previous section can be achieved without explicitly relying on simulated data. Spin asymmetries such as TSSAs in eq. (1.1) are measured as the difference between cross sections with different (longitudinal or transverse) spin orientations of the particles in the initial state. Therefore, we can train a machine learning algorithm that directly maximizes the size (positive or negative sign) of the spin asymmetry as repeated in eq. (1.2):

$$\max_{\theta} |A_{UT}(\theta)| \,. \tag{2.6}$$

The machine learning algorithm is given here in terms of the set of parameters θ. The optimization of a machine learning algorithm using eq. (2.6) only requires experimentally accessible / hadron-level information. The training does not explicitly require a UV definition of the jet flavor and it does not rely on simulated data. As discussed above, spin asymmetries are often small due to cancellations between different flavor combinations. By including a machine learning algorithm that is using the size of the spin asymmetry as an optimization metric or loss function, we can effectively achieve a flavor separation. Within QCD factorization, the sum rules in eqs. (2.2) and (2.4) provide the direct connection of the machine learning algorithm that is optimized using the objective function in eq. (2.6) to parton level quantities.

Instead of solving the regression problem in eq. (2.6) directly, we can formulate the task as a classification problem where the machine learning techniques discussed in the following sections can be applied. By training a classifier that distinguishes jets produced in events where the incoming proton has opposite transverse or longitudinal spin orientation, we can find a classifier that maximizes the corresponding spin asymmetry. We note that this approach is similar to other discrimination tasks where the training labels are known such as between jets in proton-proton and heavy-ion collisions where the trained classifier can be used to maximize the deviation of the nuclear modification factor from unity:

$$
\max_\theta |R_{AA}(\theta) - 1| = \max_\theta \left| \frac{\mathrm{d}\sigma_{AA}}{\mathrm{d}\sigma_{pp}}(\theta) - 1 \right| . \tag{2.7}
$$

See ref. [50] for more details. The identification of a machine learned-classifier can be performed directly on data before unfolding [50] or using corrected full events [110]. Subsequently, an observable can be identified that is calculable in perturbative QCD and that approximates the performance of the machine learned-classifier. For example, this can be achieved using complete sets of observables such as the N-subjettiness basis or Energy Flow Polynomials (EFPs) that will be discussed below. Note that in the case where the classifier is trained on uncorrected detector-level inputs, one must carefully account for systematic differences in detector conditions between the two data samples. In this case, the designed observables will not be optimal but rather *approximate* maximally discriminating observables — where the deviation of the approximate observable from the optimal observable is driven by the extent to which detector conditions are understood. These identified observables can then be measured with traditional techniques that correct for detector and background effects. In this way, one can verify — without any reliance on ML — whether the designed observables maximize spin asymmetries. Ultimately, the observables can then be included in a global analyses of quantum correlation functions.

3 Simulation and training setup

To perform the studies in the remainder of this article, we generate simulated events using the Monte Carlo event generator PYTHIA6 [111], which serves as the training data for the (supervised) machine learning based classification algorithms. In the following, we describe the simulated event sample and the machine learning architecture.

3.1 Event generation

We generate two data sets for the following studies, both using PYTHIA6 [111] with the eRHIC tune [112]. We use CTEQ6.1 [113] and SAS 1D-LO [114] proton and photon PDFs, respectively. See figure 1 for an illustration of the two processes in the respective data sets.

First, we generate jet samples using leading-order (LO) DIS as the hard-scattering process for the jet flavor tagging studies discussed in section 4. At LO, the final state consists of the scattered electron and a single jet originating from different quark flavors. The LO DIS process is given the process number 99 according to PYTHIA6. We then identify the jet flavor with the flavor of the underlying quark in the LO DIS process
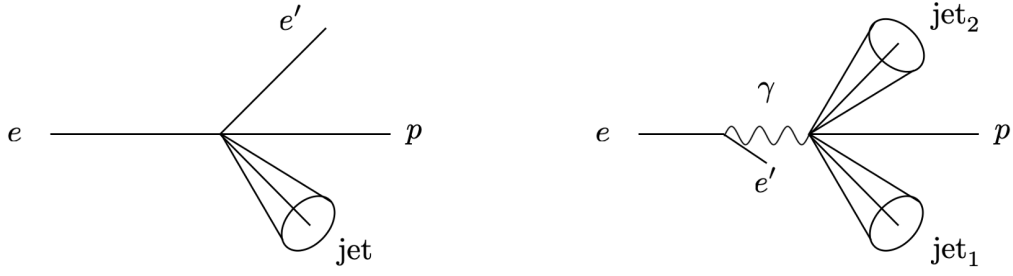
Figure 1. Illustration of the jet production processes considered in this work. Left: High-Q^2 electron-proton scattering. At leading order, the final state consists of the scattered electron and a single jet originating from different quark flavors. Right: Low-Q^2 photoproduction, where we include both the direct and the resolved contribution. At leading order, the final state consists of the scattered electron in the forward direction close to the beam axis and a di-jet pair, which can be initiated by both quarks and gluons. In both cases, the transverse momentum of the jets is measured relative to the beam axis in the laboratory frame.

($\gamma^* q \to q$). We require the photon virtuality and inelasticity to be in the range $25 < Q^2 < 1000\,\mathrm{GeV}^2$ and $0.1 < y < 0.85$, respectively.

Since gluons do not contribute at LO in DIS, we generate a second data set for our studies in section 5 of quark vs. gluon jet tagging using di-jet events in low-Q^2 photoproduction events, including both the direct and resolved photon contributions. At LO, the final state consists of the scattered electron in the forward direction close to the beam axis and a di-jet pair, which can be initiated by both quarks and gluons. We require low $10^{-5} < Q^2 < 1\,\mathrm{GeV}^2$, while maintaining the same cut on the inelasticity $0.1 < y < 0.85$. We identify quark and gluon jets in the photoproduction events using the PYTHIA6 resolved processes $11\,(qq \to qq), 12\,(q\bar{q} \to q\bar{q}), 53\,(gg \to q\bar{q})$ and $13\,(q\bar{q} \to gg), 68\,(gg \to gg)$ and the direct photon-gluon fusion processes $135\,(\gamma_T^* g \to q\bar{q}), 136\,(\gamma_L^* g \to q\bar{q})$, where the subscripts T and L denote the transverse and longitudinal polarization contributions, respectively. For the quark vs. gluon jet studies in section 5.1, we neglect the resolved process $28\,(qg \to qg)$ and the direct QCD Compton processes $131\,(\gamma_T^* q \to qg)$ and $132\,(\gamma_L^* q \to qg)$ in order to avoid ambiguity in labeling processes with qg final states, whereas in section 5.2 we include them. We note that jet cross sections and jet substructure observables in EIC photoproduction events were considered in refs. [115–117]. In addition, a comparison of PYTHIA6 results to jet substructure data from HERA was performed in ref. [117].

In accordance with experimental particle detection capabilities, we include all particles in the event and in the jet reconstruction with a lifetime of $c\tau > 1\,\mathrm{cm}$. This includes

$$\gamma,\, e^-,\, \mu^-,\, \pi^-,\, p,\, n,\, K_L^0,\, K_S^0,\, K^-,$$
$$\Lambda^0,\, \Xi^0,\, \Xi^-,\, \Sigma^\pm,\, \Omega^-,\qquad\qquad (3.1)$$

and the corresponding anti-particles. Particles with $c\tau < 1\,\mathrm{cm}$, such as neutral pions π^0, are decayed until daughters with $c\tau > 1\,\mathrm{cm}$ are produced. The scattered electron is identified as the leading electron in the event and removed before we run the jet clustering algorithm. From the scattered electron we determine the virtuality Q^2 of the exchanged

photon. We leave the implementation of a simulated detector response for future work, and as a first step we will instead examine the impact of PID, charge information, and minimum transverse momentum requirements on the performance of these taggers in section 4.

Jets are reconstructed in the laboratory frame with the anti-k_T algorithm [118]. We choose a jet radius parameter of $R = 1.0$ and we consider the rapidity range $|\eta_{\text{lab}}| < 4$. For the single jets identified in the DIS events, we require the transverse momentum of the identified jets to be $p_T^{\text{jet}} > 10\,\text{GeV}$. We consider a sample of 14M events satisfying these criteria. For the di-jets identified in the low-Q^2 photoproduction events, we require the transverse momentum of the leading jet to be $p_{T1}^{\text{jet}} > 8\,\text{GeV}$ and the subleading jet to be $p_{T2}^{\text{jet}} > 5\,\text{GeV}$, as well as the third leading jet to have $p_{T3}^{\text{jet}} < 4\,\text{GeV}$. We consider a sample of 1.4M events satisfying these criteria, corresponding to an integrated luminosity of approximately $1\,\text{fb}^{-1}$ [117], corresponding to approximately one month of EIC runtime (estimated at approximately $10\,\text{fb}^{-1}$ per year [1]). The size of the sample is determined by studying when the classification performance approximately saturates as the statistics are increased.

We use supervised training throughout this work. In order to deploy supervised models on real data, one can either perform training on simulations or in some cases on experimental data itself. Training models on Monte Carlo simulations may be viable at the EIC if event simulations based on parton showers are sufficiently reliable in the sense that the machine learning algorithms can be trained on a suite of simulations and deployed on data. In addition, some of the studies proposed here can be performed on experimental data before unfolding of detector effects. For example, classifying different spin orientations or classifying eA vs. ep observables are cases where the training labels are known experimentally, unlike the case in quark flavor or quark vs. gluon classification. This will allow for the identification of suitable observable for which unfolding can be performed afterwards [50]. Furthermore, recent advances in unfolding methods may allow one to unfold entire events [110]. We leave a more detailed exploration of these aspects for future work.

3.2 Machine learning algorithms

In the next sections, we will study the binary classification of jets with different quark flavor and quark vs. gluon jets using machine learning algorithms. For this task we choose deep sets [119–121] as our default classifier, which were introduced as Particle Flow Networks (PFNs) [35] for data obtained in high-energy collisions in particle and nuclear physics. The information about the particles in a jet or collider event can be considered as a set of four vectors with variable length event-by-event. A PFN is a neural network, which is invariant with respect to permutations of the input variables and it can naturally handle input with different length. This choice of machine learning architecture appears to be natural for data in collider physics since the number of particles varies event-by-event and there is no inherent ordering of the particles inside a jet or of the particles in the entire event. We note that other machine learning based classifiers were found to perform similarly or worse for analogous tasks at LHC energies [35, 36]. The PFNs take as input the information of all the particles inside a reconstructed jet. We represent the per-particle input variables as

$$p_i = (z_i, \eta_i, \phi_i, \text{PID}_i), \tag{3.2}$$

where $z_i = p_{Ti}/p_T^{\text{jet}}$ is the normalized transverse momentum of particle i with respect to the beam axis, and η_i, ϕ_i are its rapidity and azimuthal angle. Following ref. [35], we take η_i, ϕ_i relative to the (E-scheme [122]) jet axis. Lastly, PID_i in eq. (3.2) denotes the particle identification number. See eq. (3.1) for the different particles in our data set. We map the PIDs to numerical values in an interval centered around zero. The numerical values are separated by 0.1 and particles (anti particles) are assigned positive (negative) values.

A PFN denoted $f(p_1, \ldots, p_N)$ takes as input the kinematics of the N particles in the event. It is constructed such that it satisfies $f(p_{\pi(1)}, \ldots, p_{\pi(N)})$, where π denotes the permutation operator. The required permutation invariance can be achieved by expressing f as [119]

$$f(p_1, \ldots, p_N) = F\left(\sum_{i=1}^{N} \Phi(p_i) \right). \tag{3.3}$$

Here Φ, F denote fully connected feed-forward neural networks with a certain number of hidden layers. The connections between hidden layers are parametrized in terms of weights and each node has a bias term. The per-particle neural network $\Phi : \mathbb{R}^4 \to \mathbb{R}^d$ maps the input to a d-dimensional latent space. The summation operation in latent space leads to the permutation invariance of f. The second neural network is a map between the latent space and the final output of the binary classifier $F : \mathbb{R}^d \to \mathbb{R}$.

For quark vs. gluon jet classification, we additionally consider Energy Flow Networks (EFNs), which are Infrared-Collinear (IRC) safe architectures closely related to PFNs [35]. IRC safety is built into the permutation invariant neural network in eq. (3.3) by constructing an EFN denoted

$$\tilde{f}(p_1, \ldots, p_M) = F\left(\sum_{i=1}^{M} z_i \Phi(\hat{p}_i) \right), \tag{3.4}$$

where every particle inside a jet is written in terms of its transverse momentum momentum fractions z_i and a 2-component vector which contains the angular variables $\hat{p}_i = (y_i, \phi_i)$. Due to the weighting of Φ with the momentum fraction z_i, the resulting expression is IRC safe [35].

We parametrize the functions Φ and F in eqs. (3.3) and (3.4) in terms of DNNs, using the `EnergyFlow` package [35] with `Keras` [123]/`TensorFlow` [124]. For Φ we use two hidden layers with 100 nodes each and a latent space dimension of $d = 256$. For F we include three layers with 100 nodes each. For each dense layer we use the ReLU activation function [125] and we use the softmax activation function for the final output layer of the classifier. We train the neural networks using the Adam optimizer [126] and the binary cross entropy loss function [127], and train for 10 epochs with a batch size of 500. We do not perform additional hyperparameter optimization for the PFNs or EFNs. We reserve 20% of the training sample as a validation set, and an additional 20% as a test set on which all metrics are reported. We train the models using an NVIDIA A100 GPU on the Perlmutter supercomputer.

For quark vs. gluon tagging, we will also consider dense neural networks (DNNs) that take as input a list of observables that are IRC safe and generally calculable within perturbative QCD. The resulting classifier is generally Sudakov safe [38]. The observables

that are taken as input to the DNN form a complete basis of observables. As an example we consider Energy Flow Polynomials (EFPs) [37]. Alternately, one could consider the N-subjettiness basis [38–40]. The EFPs constitute a linear basis of jet substructure observables and they are defined as

$$\mathrm{EFP}_G = \sum_{i_1} \cdots \sum_{i_V} z_{i_1} \cdots z_{i_V} \prod_{(k,l) \in E} \theta_{i_k i_l} \tag{3.5}$$

where we sum over all particles inside the jet and z_i, θ_{ij} denote the longitudinal momentum fraction of particle i and the relative angle between particles i and j, respectively. The subscript $G = (V, E)$ indicates that EFPs are defined in terms of a graph that specifies which terms are included on the right hand side of eq. (3.5). See ref. [37] for more details. We note that this basis is insensitive to quark flavor differences but provides a powerful discriminant for quark vs. gluon jet tagging. In addition, they provide an increased degree of interpretability compared to PFNs. For the EFP DNNs, we use 3 hidden layers containing between 32-512 nodes, each with a ReLU activation function [125], followed by a sigmoid activation for the final output layer. We train the neural network with the Adam optimizer [126] and a learning rate ranging from 0.01 to 0.001 and batch size 1000, with the binary cross entropy loss function of ref. [127]. We use `Keras` [123]/`TensorFlow` [124] for the implementation, and determine the number of nodes in each hidden layer and the learning rate using a hyperparameter optimization with the Hyperband algorithm [128] implemented in Keras Tuner [129]. As in the case of the PFNs, we reserve 20% of the training sample as a validation set (on which the hyperparameter tuning is performed), and an additional 20% as a test set on which all metrics are reported.

The performance of a classifier can be assessed by various metrics quantifying the rates of correct and incorrect identification of the two classes. There are four possible outcomes of a classifier's prediction, "True/False Positive/Negative", where "True/False" denotes whether the classifier prediction was correct, and "Positive/Negative" refers to the predicted class label. In this work, we will use the following conventions for the positive class:

- u vs. d classification (section 4.1): d
- ud vs. s classification (section 4.3): s
- uds vs. c classification (section 4.3): c
- $qq/q\bar{q}$ vs. gg classification (section 5.1): $qq/q\bar{q}$
- *direct* vs. *resolved* classification (section 5.2): *direct*

We will consider two metrics in this work, the Receiver Operating Characteristic (ROC) curve and the Precision-Recall (PR) curve. These metrics are defined in terms of the following quantities each of which are cumulative distribution functions of corresponding probability distributions:

- True Positive Rate (TPR, also known as Recall):

$$\frac{\text{True Positives}}{\text{Total Positives}}$$

- False Positive Rate (FPR):
$$\frac{\text{False Positives}}{\text{Total Negatives}}$$

- Precision:
$$\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

The ROC curve shows the TPR vs. the FPR as the decision threshold is varied. A random classifier follows a diagonal line with an area under the curve (AUC) of 0.5 and the better a classifier is, the closer the curve is to the upper left edge of the plot, with a perfect classifier having AUC = 1. The ROC curve does not depend on the relative proportions of the two classes, and we will use it for classification tasks where there is not a large imbalance in the proportions of the two classes, such as u vs. d and q vs. g classification.

The PR curve shows the precision vs. recall as the decision threshold is varied. The PR curve explicitly depends on the relative proportions of the two classes, since the precision is a measure of the purity of the predicted positive class. A random classifier based solely on the relative proportions of the two classes follows a line of constant precision. The larger the precision and recall, the better the classifier is. A classifier with high precision but low recall returns only a small fraction of positive cases (low efficiency) but most of them being identified correctly (high purity), whereas a classifier with low precision but high recall returns a large fraction of positive cases (high efficiency) but with many of them being identified incorrectly (low purity). We will use the PR for classification tasks where there is a large imbalance in the proportions of the two classes, such as strange and charm jet classification.

4   Jet flavor tagging

Using the LO DIS events described in section 3.1, we now study various binary classifications of quark-jet flavors. We consider several different classification groupings: u vs. d, ud vs. s, and uds vs. c quark jets. We will study the role of PID information, charge information, and minimum particle transverse momentum thresholds on the performance of the classifiers, as well as the role of both in-jet and out-of-jet particles.

We will benchmark our machine learning-based algorithms against the energy-weighted jet charge [130]

$$Q_\kappa = \sum_{i \in \text{jet}} z_i^\kappa Q_i \,, \tag{4.1}$$

where $z_i = p_{Ti}/p_T^{\text{jet}}$ denotes the longitudinal momentum fraction of the hadrons i inside the jet and Q_i is their electric charge. The weighting factor z_i^κ reduces the sensitivity to experimental uncertainties and κ is a free parameter that we will vary in our numerical studies below. The jet charge is soft safe but collinear unsafe, which means that theoretical calculations require a nonperturbative input that needs to be determined from experiment. Theoretical calculations of the jet charge were performed in refs. [42, 43]. Extensions of the jet charge definition in eq. (4.1) were proposed in refs. [41, 131]. Theoretical work on defining the flavor of jets can be found in refs. [132–136]. Experimental measurements at
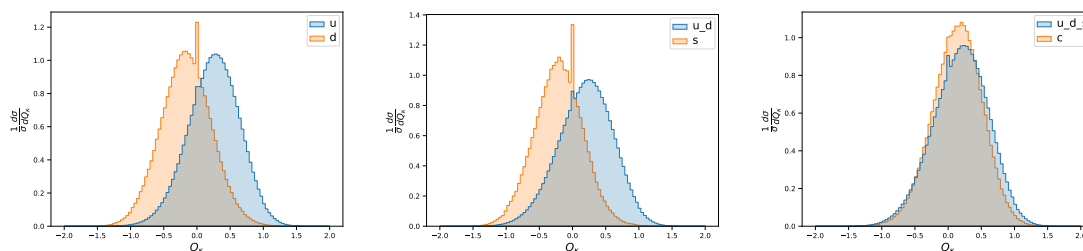
Figure 2. The jet charge distribution for EIC jets with $p_T^{\text{jet}} > 10\,\text{GeV}$ produced in high-Q^2 events as shown on the left side of figure 1. The three panels show the results for different flavor discrimination: u vs. d (left), ud vs. s (middle), and uds vs. c (right) for a jet charge parameter of $\kappa = 0.5$, see eq. (4.1). The jet charge is able to distinguish u from d, s reasonably well, whereas it is a relatively poor discriminator for u vs. c or q vs. g. Note that a peak at $Q_\kappa = 0$ arises from jets that contain only neutral particles, which happens more frequently compared to its counterpart at the LHC due to lower particle multiplicity at the EIC.

the LHC can be found in refs. [137–139]. In figure 2, we show the jet charge distributions for the LO DIS jets considered in this section. The jet charge is able to distinguish jets initiated by quarks of different electric charge reasonably well, such as u from d, s, whereas it is a relatively poor discriminator for u vs. c since they have the same electric charge, and similarly for q vs. g (not shown here). The jet charge thereby serves as a reference to which the performance of our machine learning-based algorithms can be compared.

In order to study the role of PID information and charge information, we consider three variations of the information input to the PFN training:

- "PFN w/ PID": $p_i = (z_i, \eta_i, \phi_i, \text{PID}_i)$,

- "PFN w/ charge": $p_i = (z_i, \eta_i, \phi_i, Q_i)$,

- "PFN w/o PID, charge": $p_i = (z_i, \eta_i, \phi_i)$.

We note that the "PFN w/ charge" classifier uses the same experimental information as the jet charge, whereas the "PFN w/ PID" uses full PID information, which is not used by the jet charge. Similarly, we consider varying the minimum transverse momentum of jet constituents input to the PFN training, varying between $p_{T,\text{particle}} = 0.1-0.4\,\text{GeV}$. While we do not consider the exact PID capabilities or single-particle efficiencies of the proposed EIC detectors, these variations provide a first-order estimate of the importance of PID and minimum particle transverse momentum detection capabilities and serve as an initial quantification of the value that may be gained in jet tagging performance by investing in improved PID or minimum particle transverse momentum capabilities.

4.1 u vs. d quark jets

To begin, we consider the classification of u vs. d initiated jets. Our results are shown in figure 3. We find that while the jet charge is a fairly good discriminator of u vs. d jets, the PFN (which uses the full four-vector information of the final-state particles) improves the performance when either charge information is included or even more so
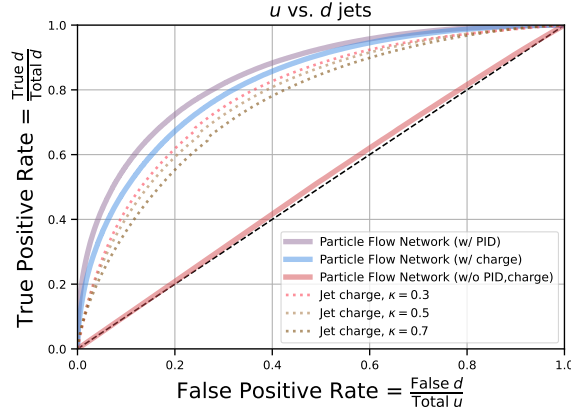
Figure 3. ROC curve for u vs. d jet flavor tagging using the jet charge and PFNs for jets with $p_T^{\text{jet}} > 10\,\text{GeV}$ and $p_{T,\text{particle}} > 0.1\,\text{GeV}$. We consider three variations of the input to the PFN, providing either PID information for all particles, charge information for all particles, or neither.
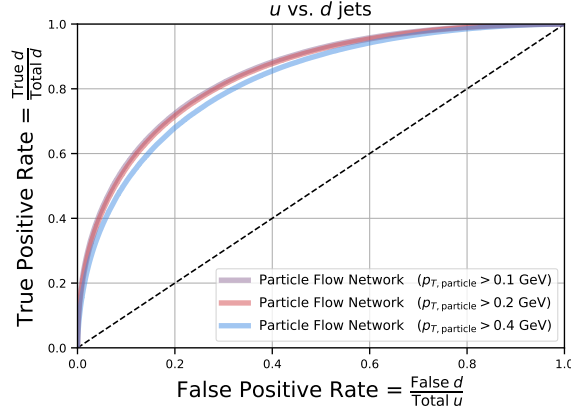

Figure 4. ROC curves for u vs. d jet flavor tagging using using PFNs with PID information for jets with $p_T^{\text{jet}} > 10\,\text{GeV}$ and different cuts on the minimum $p_{T,\text{particle}}$ required of jet constituents.

when PID information is included. When neither PID nor charge information is included, the classifier cannot significantly distinguish u jets from d jets in PYTHIA6. The increase in performance when adding PID information rather than charge information is fairly small, especially noting that experimental PID capabilities are not perfectly efficient as assumed in our studies. We will see in section 4.3, however, that for strange and charm quark jet identification, PID information provides a substantial improvement in performance.

Next, we consider the role of the minimum transverse momentum of jet constituents input to the PFN training. Figure 4 shows the results when varying the minimum threshold between $p_{T,\text{particle}} > 0.1 - 0.4\,\text{GeV}$. We find only a minor difference in the classifier performance when varying the minimum $p_{T,\text{particle}}$ between $0.1\,\text{GeV}$ and $0.4\,\text{GeV}$, suggesting that the minimum $p_{T,\text{particle}}$ detector requirements are not essential for classifying jet flavor using the in-jet information. We will see, however, in the next section that this has a stronger impact when considering the out-of-jet particles.
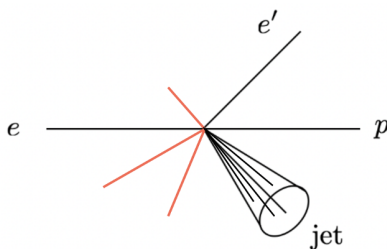
Figure 5. Illustration of particles inside the jet (black) and out-of-jet radiation (red), which we also take into account to classify both the flavor of the jet as discussed in section 4.2 and the underlying hard process in the event as discussed in section 5. As an example, we show jet production in a high-Q^2 DIS scattering process.

4.2 Out-of-jet information

The motivation of machine learned-jet classification at the EIC and RHIC is quite different compared to the LHC. For example, at the LHC di-jet reference processes can be used as calibration and the resulting classifier can be applied to identify jets in multi-jet events to search for physics beyond the Standard Model. Instead, at RHIC and the EIC the focus will be on improving for example measurements of spin asymmetries as discussed above or to improve constraints on cold nuclear matter effects. Therefore, at RHIC and the EIC, the classifier does not need to be limited to the particles inside the identified jet. We note that event-wide information was also used in classification studies at the LHC, see for example refs. [140, 141]. In this section, we investigate how the performance can be improved by not only making use of the particles inside the jet but also out-of-jet particles to classify the jet flavor, as shown in figure 5. While we have used a relatively large jet radius $R = 1.0$ in these studies, this choice is somewhat arbitrary and neglects the role of large-angle radiation and correlations across the entire event. We therefore compare the performance of a PFN supplied with only in-jet particles to that of a PFN supplied with both in-jet and out-of-jet particles.

Figure 6 shows the results of this comparison. We show the comparison for two different minimum $p_{T,\text{particle}}$ thresholds, $0.1\,\text{GeV}$ and $0.4\,\text{GeV}$. We find that the difference between the in-jet classifier and the in-jet + out-of-jet classifier is significant for the case $p_{T,\text{particle}} > 0.1\,\text{GeV}$, whereas the difference is almost negligible for $p_{T,\text{particle}} > 0.4\,\text{GeV}$. This suggests that the *soft* out-of-jet particles play a significant role in boosting the classification performance — despite that the soft in-jet particles had little impact (see figure 4). This motivates further study of the origin and role of out-of-jet radiation, since our results suggest it can provide a significant boost in jet (or event) flavor tagging performance. In section 5 we will revisit the role of out-of-jet particles in order to classify the underlying hard process of the event.

4.3 Strange and charm

We now turn to the identification of strange- and charm-quark initiated jets. Since strange- and charm-initiated jets are considerably more rare than up- or down-initiated jets (for our
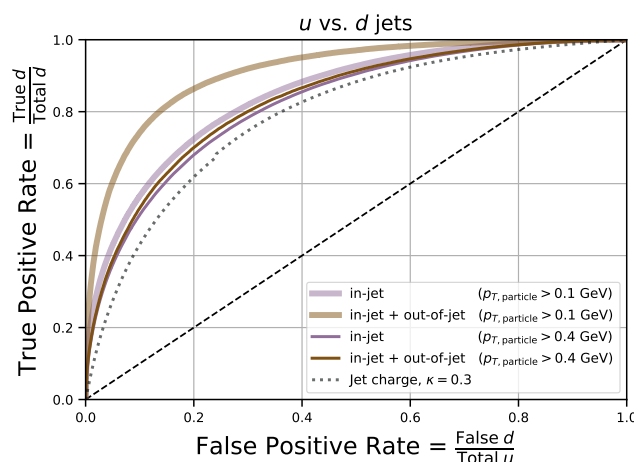
Figure 6. ROC curves for u vs. d jet flavor tagging using PFNs with PID information for jets with $p_T^{\rm jet} > 10\,{\rm GeV}$, using either in-jet information as input or using both in-jet and out-of-jet information as input. We consider two different cuts on the minimum $p_{T,{\rm particle}}$ required of both the in-jet and out-of-jet particles, which illustrate that soft out-of-jet particles play a significant role in boosting the classification performance.

kinematics, the relative $u:d:s:c$ ratios are approximately $33:5:1:2$), we quantify the classification performance using both the ROC curve and the precision-recall curve. In fact, strange jets are even more rare than charm jets, since despite that the proton PDF contains a larger quantity of strange than charm, the overall cross section for charm is larger due to its larger electric charge.

Strange and charm jets also differ from up and down jets in that strange and charm hadrons have limited decay lifetimes. In the case of strange quarks, there are a variety of weakly decaying strange hadrons with lifetimes $1\,{\rm cm} < c\tau < 10\,{\rm cm}$ (namely K_S^0, Λ^0, Ξ^0, Ξ^-, $\Sigma^\pm$, Ω^- and their associated antiparticles) which therefore decay on a length scale comparable to the size of the innermost tracking layers of collider experiments [142]. We therefore will contrast the classification performance depending on whether the PFN is provided the undecayed strange hadrons or only the decay products of these hadrons. In the case of charm quarks, on the other hand, all charm hadrons decay with lifetimes much shorter than $c\tau = 1\,{\rm cm}$, and cannot be directly detected by experiments but rather must be reconstructed using the invariant mass of decay products of exclusive charm hadrons or by tagging displaced vertices. A large literature exists on charm-jet tagging algorithms, but we will not pursue performance comparisons here [143–146].

Figure 7 shows the results for ud vs. s jet classification with final-state particle decay lifetimes of $c\tau > 1\,{\rm cm}$ and $c\tau > 10\,{\rm cm}$, respectively. We find several notable differences compared to the u vs. d classification. First, the PFN with PID dramatically outperforms the jet charge. We also provide as a reference a simple "Leading strange tagger" which classifies the jet flavor purely based on whether the highest p_T particle in the jet is a strange hadron. The PFN also dramatically outperforms this. This provides a clear illustration
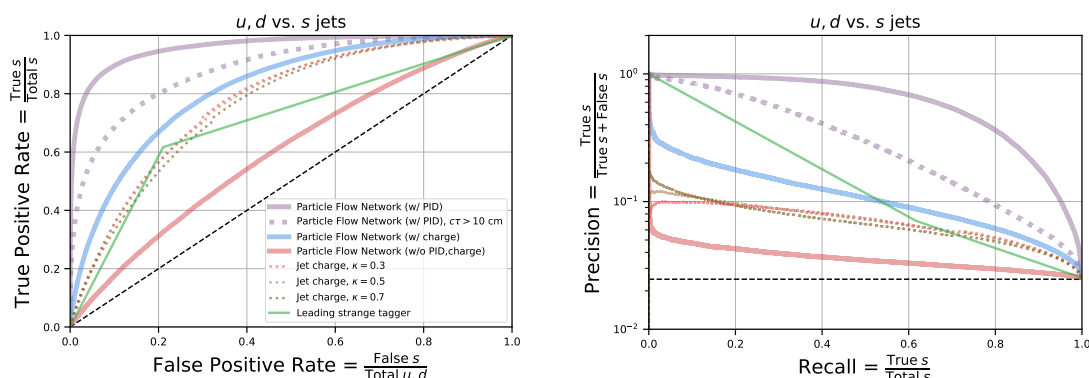
Figure 7. ROC curve (left) and PR curve (right) for ud vs s jet flavor tagging using the jet charge and PFNs for jets with $p_T^{\text{jet}} > 10\,\text{GeV}$ and $p_{T,\text{particle}} > 0.1\,\text{GeV}$. We consider several variations of the input to the PFN, providing either PID information for all particles, charge information for all particles, or neither. All curves are constructed from particles with a decay length $c\tau > 1\,\text{cm}$ (in which the weakly-decaying strange hadrons K_S^0, Λ^0, Ξ^0, Ξ^-, $\Sigma^\pm$, Ω^- and their associated antiparticles are undecayed), except the curve labeled $c\tau > 10\,\text{cm}$, which is constructed from particles with a decay length $c\tau > 10\,\text{cm}$ (in which the above weakly-decaying strange hadrons are decayed). The dashed black lines correspond to a random classifier.

of the value of machine learning-based jet flavor identification. Second, the overall performance of ud vs. s tagging is significantly improved when PID information is provided relative to charge information, especially when the weakly decaying strange hadrons with $c\tau > 1\,\text{cm}$ are included as input to the PFN. If only charge information is supplied, the performance decreases substantially. This provides a clear illustration that PID information is highly valuable to obtaining the best possible strange-jet tagging performance. We leave further study, such as whether providing PID information of the leading particle rather than all particles, which could substantially lessen the experimental efforts, to future work. Third, if neither PID nor charge information is provided, the performance is yet again substantially worse — however it is still notably better than in the u vs. d case. This illustrates the relative importance of particle identification vs. fragmentation in determining the jet flavor — since when neither PID nor charge information is provided the machine learning algorithm can only learn from the differences in fragmentation between ud and s jets.

Figure 8 shows the results for uds vs. c jet classification. In this case, the jet charge is not expected to be a good discriminator, since u (which dominates the uds sample) and c jets have the same electric charge. We find similarly strong performance of the PFN classifier when PID information is included, with an even larger benefit of providing PID information relative to charge information. Additionally, we note that the PFN that is supplied with neither PID nor charge information performs better than the previous cases, illustrating that the amount of information in the fragmentation pattern unrelated to particle PID or charge plays an increasing role for heavier quarks, as expected [147, 148].
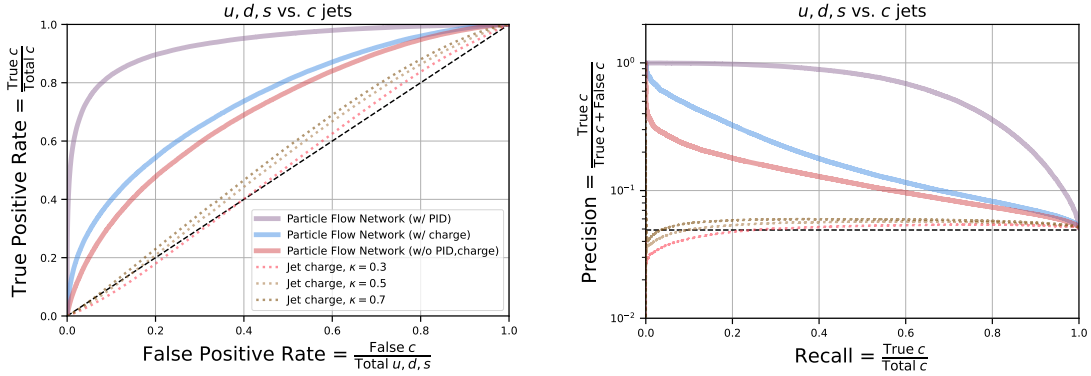
Figure 8. ROC curve (left) and PR curve (right) for *uds* vs *c* jet flavor tagging using the jet charge and PFNs for jets with $p_T^{\text{jet}} > 10\,\text{GeV}$ and $p_{T,\text{particle}} > 0.1\,\text{GeV}$. We consider several variations of the input to the PFN, providing either PID information for all particles, charge information for all particles, or neither. The dashed black lines correspond to a random classifier.

5 Hard process tagging

The classification of the underlying hard process is often of primary interest instead of the classification of a single jet. To do this, we propose to not only utilize the particles inside the reconstructed jet but to also take as input particles outside the jet, similar to the studies done in section 4.2 and shown in figure 5. Note that we still require a jet with a given transverse momentum to identify the entire event to ensure the presence of a hard-scale, which allows for the interpretation or applicability of perturbative techniques in QCD. The additional information contained in the dynamics of particles outside the reconstructed jet can generally increase the performance of the machine learning algorithm. We note that event type classification using triggers and machine learning was discussed in ref. [4] and references therein. Different than ref. [4], we aim here at identifying the underlying hard process in the event at parton level. As discussed in section 3.2 above, the in-jet information that is used to train machine learned classifiers can be captured by complete sets of observables like N-subjettiness and EFPs. Similar observable bases can be constructed for out-of-jet information and correlations between jets (such as in photoproduction events, see figure 1) can be captured by observables like the jet pull [45, 46].

We consider two examples of event classification in this section. In both cases, we use low-Q^2 photoproduction events that contain a di-jet signal with the transverse momentum of the leading jet required to be $p_{T1}^{\text{jet}} > 8\,\text{GeV}$ and the subleading jet to be $p_{T2}^{\text{jet}} > 5\,\text{GeV}$, as described in section 3.1. First, we consider the classification quark vs. gluon jet topologies by discriminating qq or $q\bar{q}$ di-jet topologies from gg topologies. Second, we consider the classification of direct vs. resolved photoproduction processes. Similar to the in-jet particles, we normalize the transverse momentum of out-of-jet particles relative to the leading jet transverse momentum $z_i = p_{Ti}/p_{T1}^{\text{jet}}$. Since we divide by the transverse momentum of the leading jet in the event, we have $z_i < 1$ for both in-jet and out-of-jet particles. Moreover, we count the values (η_i, ϕ_i) of out-of-jet particles relative to the leading jet axis.
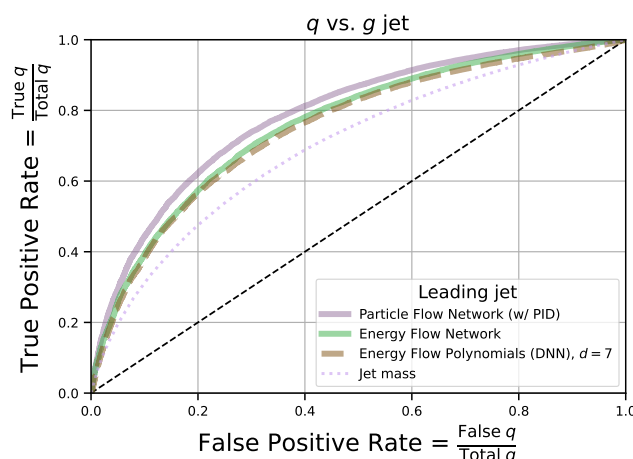
Figure 9. ROC curves for quark vs. gluon jet tagging at the EIC using the leading jet information from quark and gluon di-jets in low-Q^2 protoproduction events containing $qq, q\bar{q}, gg$ topologies. The leading jet is required to be $p_{T1}^{\text{jet}} > 8\,\text{GeV}$, the subleading jet to be $p_{T2}^{\text{jet}} > 5\,\text{GeV}$, and the third leading jet to be $p_{T3}^{\text{jet}} < 4\,\text{GeV}$, see figure 1. We consider several models: (i) PFN including PID information, (ii) EFN, (iii) DNN with EFPs for two different dimensions d, and (iv) jet mass.

5.1 Quark vs. gluon jet tagging

We consider events with quark and gluon di-jet topologies by considering both direct and resolved processes that result in qq-, $q\bar{q}$-, or gg-initiated di-jets, as described in section 3.1. We then train PFNs using either (i) the particles in the leading jet, (ii) the particles in both the leading and subleading jet, or (iii) all particles in the event with $p_{T,\text{particle}} > 0.1\,\text{GeV}$.

Figure 9 shows the classification performance of quark vs. gluon jet event topologies when trained with only the leading jet particles. Models trained with the leading jet particles correspond most closely to previous studies of the classification of quark vs. gluon single jets [27, 34]. We consider a PFN trained with PID information, as well as an IRC-safe EFN, which performs slightly worse than the PFN, which is typical in quark vs. gluon jet classification at the LHC. While the performance at the low EIC jet energies considered here is lower than quark vs. gluon classification with high-p_T^{jet} jets at the LHC, the PFN and EFN still are able to achieve substantial classification performance and large improvements compared to single observables such as the jet mass. Additionally, we compare the a DNN that uses IRC-safe EFPs with dimension $d = 7$ as input, which gives a performance that approaches that of the deep set models.

Figure 10 shows the classification performance of quark vs. gluon jet event topologies as the subleading jet particles and out-of-di-jet particles are added to the PFN training input. As the subleading jet particles and out-of-di-jet particles are added to the PFN training input, the performance significantly increases.
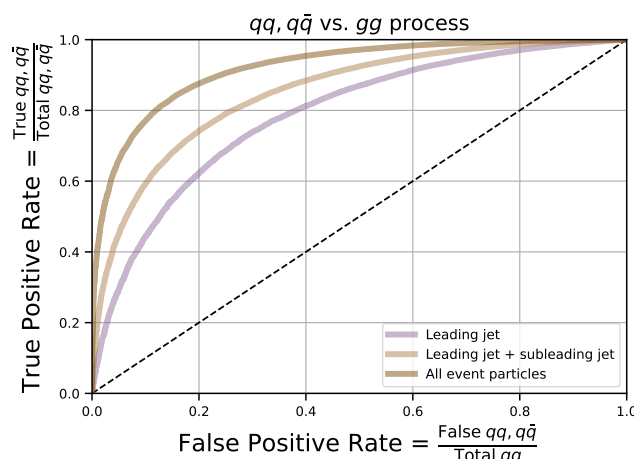
Figure 10. ROC curves for quark vs. gluon event tagging at the EIC with PFNs including PID information and considering as input to the PFN either the leading jet particles, the leading and subleading jet particles, or all particles in the event with $p_{T,\text{particle}} > 0.1\,\text{GeV}$. Here we consider quark and gluon di-jets in low-Q^2 protoproduction events containing $qq, q\bar{q}, gg$ topologies. The leading jet is required to be $p_{T1}^{\text{jet}} > 8\,\text{GeV}$, the subleading jet to be $p_{T2}^{\text{jet}} > 5\,\text{GeV}$, and the third leading jet to be $p_{T3}^{\text{jet}} < 4\,\text{GeV}$, see figure 1.

5.2 Direct vs. resolved processes and improved constraints on photon structure

Next, we consider discriminating events that arise from direct vs. resolved photoproduction processes. The direct processes correspond to those where the low-Q^2 quasi-real photon directly participates in the hard-scattering process. Instead, the resolved process corresponds to the case where the parton content of the photon is resolved and only quarks and gluons are involved in the hard-scattering process that produces the di-jet pair in the final state. The resolved process provides access to the nonperturbative parton-in-photon PDF, which is of particular interest at the future EIC. While the unpolarized di-jet photoproduction cross section has been measured at HERA by the H1 Collaboration [149], the polarized cross section will be measured for the first time at the EIC and the helicity parton-in-photon PDFs are currently unconstrained from experimental data. See refs. [116, 117, 150–152] for theoretical work on the partonic structure of photons in the context of the future EIC.

An approximate separation of the direct and resolved contribution can significantly enhance the sensitivity to the nonperturbative parton-in-photon PDF. Traditionally, this separation has been achieved by measuring a multi-differential cross section including the kinematic variable x_γ, which is defined in terms of the electron energy and the di-jet transverse momenta and rapidities [149, 150]. At LO in QCD, x_γ corresponds to the momentum fraction of the parton inside the photon such that for $x_\gamma \to 1(0)$, the direct (resolved) process dominates. Instead, here we propose that the performance can be augmented using a machine learning-based binary classifier that can make use of the full event information.
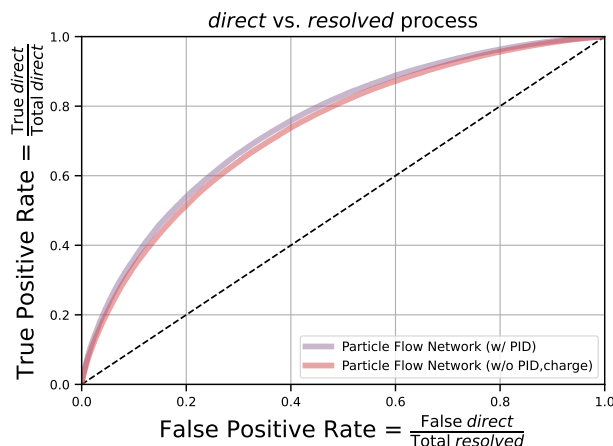
Figure 11. ROC curves for direct vs. resolved process tagging at the EIC. Here we consider quark and gluon jets produced in low-Q^2 photoproduction events, see figure 1.

In order to explore this approach, we train a PFN to distinguish the direct vs. resolved photoproduction processes using the leading and subleading jet information. We consider events with the same quark and gluon di-jet topologies described in the previous section except additionally including *qg*-initiated di-jets in addition to *qq*-, *q$\bar{q}$*-, and *gg*-initiated di-jets. Figure 11 shows the classification performance of direct vs. resolved photoproduction processes. We find that the performance is worse than the quark vs. gluon di-jet topology classification shown in the previous section, which is unsurprising given that the direct and resolved contributions contain both quark and gluon jets. We furthermore find that the impact of supplying PID information to the PFN is almost negligible in this case. It would be instructive to combine the information of the "QCD-inspired" variable x_γ with the machine-learned classifier described here. Note that in figure 11 we have not included information from the electron. In addition, it would be interesting to combine the tagging of the direct vs. resolved process with jet flavor identification as discussed in previous sections. This would allow for a flavor separation of the parton-in-photon PDFs. We note that a flavor separation based on identified hadrons inside the di-jets was employed in ref. [150]. We leave the exploration of these topics as well as quantitative impact studies at the EIC for future work.

6 Conclusions and outlook

We have presented first studies of machine learning based jet and event classification using simulated events at the EIC. While the performance of jet flavor classification is more challenging than with high transverse momentum jets at the LHC, machine learning-based classification algorithms offer important advantages in performance and many prospects for interpretability. We found that machine learning algorithms outperform traditional observables used to identify jet flavor, such as the jet charge. In order to provide input to the detector specifications at the EIC, we investigated the impact of PID information, charge information, and minimum particle transverse momentum requirements. We found

that providing charge information is sufficient for u vs. d jet classification, but that PID information gives a large improvement to strange and charm jet tagging capabilities. We found that soft particles with $0.1 < p_T < 0.4\,\text{GeV}$ have only minor impact when jet flavor classification is performed using in-jet particles, but that out-of-jet soft particles give substantial improvement to the classification performance. The studies performed here can be extended to a full detector simulation and additional kinematics. Future detailed studies on the impact of PID for strange and charm quark-initiated jets may be useful, such as investigating whether it is important for the all particles to be identified or whether a small number of leading particles is responsible for most of the flavor tagging performance. Another future direction is the exploration of different machine learning architectures. In this work, we limited ourselves to algorithms that have been known to perform well in the LHC environment. The jet quark flavor classification could also be extended from binary classification to multi-label classification for all quark flavors simultaneously.

We have proposed several applications in which these methods will play an important role in the EIC science program. Jet flavor tagging can lead to improved constraints of collinear and transverse momentum dependent parton distribution functions, such as the strange quark PDF and the gluon TMD. Eventually, machine-learned event-by-event classifiers may be directly included in global analyses of quantum correlation functions like PDFs. Flavor tagging also provides opportunities to increase experimental access to transverse single spin asymmetries, since one can use the magnitude of the asymmetry as an objective function for the machine algorithm to maximize. Tagging the underlying hard process in events can enhance constraints on the parton-in-photon PDF when combined with previous work [150]. Additionally, the methods outlined here can be applied to classify ep vs. eA collisions and provide new insight to cold nuclear matter effects, and can be extended to produce increasingly interpretable results [50]. The supervised machine learning methods presented here to classify jet flavor require suitably accurate Monte Carlo event generators to produce training data. This will require repeating these studies with additional and increasingly sophisticated Monte Carlo event generators. On the other hand, the applications we proposed for (i) maximizing spin asymmetries and (ii) classifying ep vs. eA can be performed directly on experimental data since the training labels are known. While the latter can currently only be explored using simulations, the former can be deployed immediately at RHIC.

Data availability. The data sets used in this work can be found at:

https://zenodo.org/record/7538810#.Y8RcaS-B2gQ

We provide the particle four vectors for both the LO DIS sample and photoproduction sample described above.

Acknowledgments

References

[1] R. Abdul Khalek et al., *Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report*, *Nucl. Phys. A* **1026** (2022) 122447 [arXiv:2103.05419] [INSPIRE].

[2] M. Arratia, Y. Song, F. Ringer and B.V. Jacak, *Jets as precision probes in electron-nucleus collisions at the future Electron-Ion Collider*, *Phys. Rev. C* **101** (2020) 065204 [arXiv:1912.05931] [INSPIRE].

[3] B.S. Page, X. Chu and E.C. Aschenauer, *Experimental Aspects of Jet Physics at a Future EIC*, *Phys. Rev. D* **101** (2020) 072003 [arXiv:1911.00657] [INSPIRE].

[4] A. Boehnlein et al., *Colloquium: Machine learning in nuclear physics*, *Rev. Mod. Phys.* **94** (2022) 031003 [arXiv:2112.02309] [INSPIRE].

[5] M. Feickert and B. Nachman, *A Living Review of Machine Learning for Particle Physics*, arXiv:2102.02770 [INSPIRE].

[6] Y. Alanazi et al., *Simulation of electron-proton scattering events by a Feature-Augmented and Transformed Generative Adversarial Network (FAT-GAN)*, in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, on-line (2020), pg. 2126, https://doi.org/10.24963/ijcai.2021/293 [arXiv:2001.11103] [INSPIRE].

[7] Y.S. Lai, D. Neill, M. Płoskoń and F. Ringer, *Explainable machine learning of the underlying physics of high-energy particle collisions*, *Phys. Lett. B* **829** (2022) 137055 [arXiv:2012.06582] [INSPIRE].

[8] A. Butter et al., *Generative Networks for Precision Enthusiasts*, arXiv:2110.13632 [INSPIRE].

[9] M. Arratia, D. Britzger, O. Long and B. Nachman, *Reconstructing the kinematics of deep inelastic scattering with deep learning*, *Nucl. Instrum. Meth. A* **1025** (2022) 166164 [arXiv:2110.05505] [INSPIRE].

[10] M. Diefenthaler, A. Farhat, A. Verbytskyi and Y. Xu, *Deeply learning deep inelastic scattering kinematics*, *Eur. Phys. J. C* **82** (2022) 1064 [arXiv:2108.11638] [INSPIRE].

[11] NNPDF collaboration, *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [arXiv:1410.8849] [InSPIRE].

[12] S. Alekhin, J. Blümlein and S. Moch, *NLO PDFs from the ABMP16 fit*, *Eur. Phys. J. C* **78** (2018) 477 [arXiv:1803.07537] [InSPIRE].

[13] J. Bringewatt, N. Sato, W. Melnitchouk, J.-W. Qiu, F. Steffens and M. Constantinou, *Confronting lattice parton distributions with global QCD analysis*, *Phys. Rev. D* **103** (2021) 016003 [arXiv:2010.00548] [InSPIRE].

[14] T.-J. Hou et al., *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, *Phys. Rev. D* **103** (2021) 014013 [arXiv:1912.10053] [InSPIRE].

[15] I. Borsa, R. Sassot, D. de Florian, M. Stratmann and W. Vogelsang, *Towards a Global QCD Analysis of Fragmentation Functions at Next-to-Next-to-Leading Order Accuracy*, *Phys. Rev. Lett.* **129** (2022) 012002 [arXiv:2202.05060] [InSPIRE].

[16] J. Gallicchio and M.D. Schwartz, *Quark and Gluon Tagging at the LHC*, *Phys. Rev. Lett.* **107** (2011) 172001 [arXiv:1106.3076] [InSPIRE].

[17] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [arXiv:1305.0007] [InSPIRE].

[18] ATLAS collaboration, *Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **74** (2014) 3023 [arXiv:1405.6583] [InSPIRE].

[19] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, *JHEP* **07** (2016) 069 [arXiv:1511.05190] [InSPIRE].

[20] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet Substructure Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev. D* **93** (2016) 094034 [arXiv:1603.09349] [InSPIRE].

[21] G. Louppe, K. Cho, C. Becot and K. Cranmer, *QCD-Aware Recursive Neural Networks for Jet Physics*, *JHEP* **01** (2019) 057 [arXiv:1702.00748] [InSPIRE].

[22] A. Andreassen, I. Feige, C. Frye and M.D. Schwartz, *JUNIPR: a Framework for Unsupervised Machine Learning in Particle Physics*, *Eur. Phys. J. C* **79** (2019) 102 [arXiv:1804.09720] [InSPIRE].

[23] A.J. Larkoski and E.M. Metodiev, *A Theory of Quark vs. Gluon Discrimination*, *JHEP* **10** (2019) 014 [arXiv:1906.01639] [InSPIRE].

[24] S. Marzani, G. Soyez and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, *Lect. Notes Phys.* **958** (2019) 1 [arXiv:1901.10342] [InSPIRE].

[25] Y.-C.J. Chen, C.-W. Chiang, G. Cottin and D. Shih, *Boosted W and Z tagging with jet charge and deep learning*, *Phys. Rev. D* **101** (2020) 053001 [arXiv:1908.08256] [InSPIRE].

[26] CMS collaboration, *Measurement of quark- and gluon-like jet fractions using jet charge in PbPb and pp collisions at 5.02 TeV*, *JHEP* **07** (2020) 115 [arXiv:2004.00602] [InSPIRE].

[27] F.A. Dreyer and H. Qu, *Jet tagging in the Lund plane with graph networks*, *JHEP* **03** (2021) 052 [arXiv:2012.08526] [InSPIRE].

[28] I.W. Stewart and X. Yao, *Pure quark and gluon observables in collinear drop*, *JHEP* **09** (2022) 120 [arXiv:2203.14980] [InSPIRE].

[29] S. Bright-Thonney, I. Moult, B. Nachman and S. Prestel, *Systematic quark/gluon identification with ratios of likelihoods*, *JHEP* **12** (2022) 021 [arXiv:2207.12411] [InSPIRE].

[30] Y.-T. Chien and R. Kunnawalkam Elayavalli, *Probing heavy ion collisions using quark and gluon jet substructure*, arXiv:1803.03589 [InSPIRE].

[31] L. Apolinário, N.F. Castro, M. Crispim Romão, J.G. Milhano, R. Pedro and F.C.R. Peres, *Deep Learning for the classification of quenched jets*, *JHEP* **11** (2021) 219 [arXiv:2106.08869] [InSPIRE].

[32] L. Liu, J. Velkovska and M. Verweij, *Identifying quenched jets in heavy ion collisions with machine learning*, arXiv:2206.01628 [InSPIRE].

[33] A.J. Larkoski, I. Moult and B. Nachman, *Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning*, *Phys. Rept.* **841** (2020) 1 [arXiv:1709.04464] [InSPIRE].

[34] P.T. Komiske, E.M. Metodiev and M.D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, *JHEP* **01** (2017) 110 [arXiv:1612.01551] [InSPIRE].

[35] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121 [arXiv:1810.05165] [InSPIRE].

[36] Y. Lu, A. Romero, M.J. Fenton, D. Whiteson and P. Baldi, *Resolving extreme jet substructure*, *JHEP* **08** (2022) 046 [arXiv:2202.00723] [InSPIRE].

[37] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy flow polynomials: A complete linear basis for jet substructure*, *JHEP* **04** (2018) 013 [arXiv:1712.07124] [InSPIRE].

[38] K. Datta and A. Larkoski, *How Much Information is in a Jet?*, *JHEP* **06** (2017) 073 [arXiv:1704.08249] [InSPIRE].

[39] K. Datta and A.J. Larkoski, *Novel Jet Observables from Machine Learning*, *JHEP* **03** (2018) 086 [arXiv:1710.01305] [InSPIRE].

[40] K. Datta, A. Larkoski and B. Nachman, *Automating the Construction of Jet Observables with Machine Learning*, *Phys. Rev. D* **100** (2019) 095016 [arXiv:1902.07180] [InSPIRE].

[41] K. Fraser and M.D. Schwartz, *Jet Charge and Machine Learning*, *JHEP* **10** (2018) 093 [arXiv:1803.08066] [InSPIRE].

[42] W.J. Waalewijn, *Calculating the Charge of a Jet*, *Phys. Rev. D* **86** (2012) 094030 [arXiv:1209.3019] [InSPIRE].

[43] D. Krohn, M.D. Schwartz, T. Lin and W.J. Waalewijn, *Jet Charge at the LHC*, *Phys. Rev. Lett.* **110** (2013) 212001 [arXiv:1209.2421] [InSPIRE].

[44] I.W. Stewart, F.J. Tackmann and W.J. Waalewijn, *N-Jettiness: An Inclusive Event Shape to Veto Jets*, *Phys. Rev. Lett.* **105** (2010) 092002 [arXiv:1004.2489] [InSPIRE].

[45] J. Gallicchio and M.D. Schwartz, *Seeing in Color: Jet Superstructure*, *Phys. Rev. Lett.* **105** (2010) 022001 [arXiv:1001.5027] [InSPIRE].

[46] A.J. Larkoski, S. Marzani and C. Wu, *Theory Predictions for the Pull Angle*, *Phys. Rev. D* **99** (2019) 091502 [arXiv:1903.02275] [InSPIRE].

[47] M. Arratia, Y. Furletova, T.J. Hobbs, F. Olness and S.J. Sekula, *Charm jets as a probe for strangeness at the future Electron-Ion Collider*, *Phys. Rev. D* **103** (2021) 074023 [arXiv:2006.12520] [InSPIRE].

[48] J. Arrington et al., *EIC Physics from An All-Silicon Tracking Detector*, `arXiv:2102.08337` [inSPIRE].

[49] S. Caletti, O. Fedkevych, S. Marzani and D. Reichelt, *Tagging the initial-state gluon*, *Eur. Phys. J. C* **81** (2021) 844 [`arXiv:2108.10024`] [inSPIRE].

[50] Y.S. Lai, J. Mulligan, M. Płoskoń and F. Ringer, *The information content of jet quenching and machine learning assisted observable design*, *JHEP* **10** (2022) 011 [`arXiv:2111.14589`] [inSPIRE].

[51] M. Cranmer et al., *Discovering Symbolic Models from Deep Learning with Inductive Biases*, `arXiv:2006.11287` [inSPIRE].

[52] STAR collaboration, *Measurement of transverse single-spin asymmetries for di-jet production in proton-proton collisions at $\sqrt{s} = 200$ GeV*, *Phys. Rev. Lett.* **99** (2007) 142003 [`arXiv:0705.4629`] [inSPIRE].

[53] R. Fatemi, *News from RHIC spin*, EINN Workshop, Paphos, Cyprus (2019).

[54] Z.-B. Kang, X. Liu, S. Mantry and D.Y. Shao, *Jet Charge: A Flavor Prism for Spin Asymmetries at the EIC*, *Phys. Rev. Lett.* **125** (2020) 242003 [`arXiv:2008.00655`] [inSPIRE].

[55] P. Ru, Z.-B. Kang, E. Wang, H. Xing and B.-W. Zhang, *Global extraction of the jet transport coefficient in cold nuclear matter*, *Phys. Rev. D* **103** (2021) L031901 [`arXiv:1907.11808`] [inSPIRE].

[56] H.T. Li, Z.L. Liu and I. Vitev, *Heavy meson tomography of cold nuclear matter at the electron-ion collider*, *Phys. Lett. B* **816** (2021) 136261 [`arXiv:2007.10994`] [inSPIRE].

[57] M. Xie, X.-N. Wang and H.-Z. Zhang, *γ-hadron spectra in $p+Pb$ collisions at $\sqrt{s_{NN}} = 5.02$ TeV*, *Phys. Rev. C* **103** (2021) 034911 [`arXiv:2003.02441`] [inSPIRE].

[58] Y.-Y. Zhang and X.-N. Wang, *Parton rescattering and gluon saturation in dijet production at EIC*, *Phys. Rev. D* **105** (2022) 034015 [`arXiv:2104.04520`] [inSPIRE].

[59] H. Clayton, M.D. Sievert and W.A. Horowitz, *Jet broadening in the opacity and twist expansions*, *Eur. Phys. J. C* **82** (2022) 437 [`arXiv:2110.14737`] [inSPIRE].

[60] Y. Ying, J. Brewer, Y. Chen and Y.-J. Lee, *Data-driven extraction of the substructure of quark and gluon jets in proton-proton and heavy-ion collisions*, `arXiv:2204.00641` [inSPIRE].

[61] D.W. Sivers, *Single Spin Production Asymmetries from the Hard Scattering of Point-Like Constituents*, *Phys. Rev. D* **41** (1990) 83 [inSPIRE].

[62] J.C. Collins, *Fragmentation of transversely polarized quarks probed in transverse momentum distributions*, *Nucl. Phys. B* **396** (1993) 161 [`hep-ph/9208213`] [inSPIRE].

[63] STAR collaboration, *Measurement of transverse single-spin asymmetries of π^0 and electromagnetic jets at forward rapidity in 200 and 500 GeV transversely polarized proton-proton collisions*, *Phys. Rev. D* **103** (2021) 092009 [`arXiv:2012.11428`] [inSPIRE].

[64] PHENIX collaboration, *Transverse-single-spin asymmetries of charged pions at midrapidity in transversely polarized $p+p$ collisions at $\sqrt{s} = 200$ GeV*, *Phys. Rev. D* **105** (2022) 032003 [`arXiv:2112.05680`] [inSPIRE].

[65] PHENIX collaboration, *Probing Gluon Spin-Momentum Correlations in Transversely Polarized Protons through Midrapidity Isolated Direct Photons in $p^{\uparrow} + p$ Collisions at $\sqrt{s} = 200\,GeV$*, *Phys. Rev. Lett.* **127** (2021) 162001 [arXiv:2102.13585] [INSPIRE].

[66] STAR collaboration, *Azimuthal transverse single-spin asymmetries of inclusive jets and identified hadrons within jets from polarized pp collisions at $\sqrt{s} = 200\,GeV$*, *Phys. Rev. D* **106** (2022) 072010 [arXiv:2205.11800] [INSPIRE].

[67] PHENIX collaboration, *Improving constraints on gluon spin-momentum correlations in transversely polarized protons via midrapidity open-heavy-flavor electrons in $p^{\uparrow} + p$ collisions at $\sqrt{s} = 200\,GeV$*, arXiv:2204.12899 [INSPIRE].

[68] D. Boer and W. Vogelsang, *Asymmetric jet correlations in p p uparrow scattering*, *Phys. Rev. D* **69** (2004) 094025 [hep-ph/0312320] [INSPIRE].

[69] X. Liu, F. Ringer, W. Vogelsang and F. Yuan, *Factorization and its Breaking in Dijet Single Transverse Spin Asymmetries in pp Collisions*, *Phys. Rev. D* **102** (2020) 114012 [arXiv:2008.03666] [INSPIRE].

[70] Z.-B. Kang, K. Lee, D.Y. Shao and J. Terry, *The Sivers Asymmetry in Hadronic Dijet Production*, *JHEP* **02** (2021) 066 [arXiv:2008.05470] [INSPIRE].

[71] A. Schafer and O.V. Teryaev, *Sum rules for the T-odd fragmentation functions*, *Phys. Rev. D* **61** (2000) 077903 [hep-ph/9908412] [INSPIRE].

[72] S. Meissner, A. Metz and D. Pitonyak, *Momentum sum rules for fragmentation functions*, *Phys. Lett. B* **690** (2010) 296 [arXiv:1002.4393] [INSPIRE].

[73] K. Goeke, S. Meissner, A. Metz and M. Schlegel, *Checking the Burkardt sum rule for the Sivers function by model calculations*, *Phys. Lett. B* **637** (2006) 241 [hep-ph/0601133] [INSPIRE].

[74] M. Burkardt, *Sivers mechanism for gluons*, *Phys. Rev. D* **69** (2004) 091501 [hep-ph/0402014] [INSPIRE].

[75] M. Burkardt, *Quark correlations and single spin asymmetries*, *Phys. Rev. D* **69** (2004) 057501 [hep-ph/0311013] [INSPIRE].

[76] JEFFERSON LAB ANGULAR MOMENTUM collaboration, *Origin of single transverse-spin asymmetries in high-energy collisions*, *Phys. Rev. D* **102** (2020) 054002 [arXiv:2002.08384] [INSPIRE].

[77] M. Bury, A. Prokudin and A. Vladimirov, *Extraction of the Sivers Function from SIDIS, Drell-Yan, and $W^{\pm}/Z$ Data at Next-to-Next-to-Next-to Leading Order*, *Phys. Rev. Lett.* **126** (2021) 112002 [arXiv:2012.05135] [INSPIRE].

[78] M. Bury, A. Prokudin and A. Vladimirov, *Extraction of the Sivers function from SIDIS, Drell-Yan, and $W^{\pm}/Z$ boson production data with TMD evolution*, *JHEP* **05** (2021) 151 [arXiv:2103.03270] [INSPIRE].

[79] X. Liu, F. Ringer, W. Vogelsang and F. Yuan, *Lepton-jet Correlations in Deep Inelastic Scattering at the Electron-Ion Collider*, *Phys. Rev. Lett.* **122** (2019) 192003 [arXiv:1812.08077] [INSPIRE].

[80] Z.-B. Kang, K. Lee, D.Y. Shao and F. Zhao, *Spin asymmetries in electron-jet production at the future electron ion collider*, *JHEP* **11** (2021) 005 [arXiv:2106.15624] [INSPIRE].

[81] M. Procura and I.W. Stewart, *Quark Fragmentation within an Identified Jet*, *Phys. Rev. D* **81** (2010) 074009 [arXiv:0911.4980] [*Erratum ibid.* **83** (2011) 039902] [INSPIRE].

[82] A. Jain, M. Procura and W.J. Waalewijn, *Parton Fragmentation within an Identified Jet at NNLL*, *JHEP* **05** (2011) 035 [arXiv:1101.4953] [INSPIRE].

[83] M. Procura and W.J. Waalewijn, *Fragmentation in Jets: Cone and Threshold Effects*, *Phys. Rev. D* **85** (2012) 114041 [arXiv:1111.6605] [INSPIRE].

[84] Z.-B. Kang, F. Ringer and I. Vitev, *Jet substructure using semi-inclusive jet functions in SCET*, *JHEP* **11** (2016) 155 [arXiv:1606.07063] [INSPIRE].

[85] Z.-B. Kang, K. Lee and F. Zhao, *Polarized jet fragmentation functions*, *Phys. Lett. B* **809** (2020) 135756 [arXiv:2005.02398] [INSPIRE].

[86] F. Yuan, *Azimuthal asymmetric distribution of hadrons inside a jet at hadron collider*, *Phys. Rev. Lett.* **100** (2008) 032003 [arXiv:0709.3272] [INSPIRE].

[87] M. Arratia, Z.-B. Kang, A. Prokudin and F. Ringer, *Jet-based measurements of Sivers and Collins asymmetries at the future electron-ion collider*, *Phys. Rev. D* **102** (2020) 074015 [arXiv:2007.07281] [INSPIRE].

[88] X. Liu and H. Xing, *The Time-reversal Odd Side of a Jet*, arXiv:2104.03328 [INSPIRE].

[89] W.K. Lai, X. Liu, M. Wang and H. Xing, *Unveiling Nucleon 3D Chiral-Odd Structure with Jet Axes*, arXiv:2205.04570 [INSPIRE].

[90] JEFFERSON LAB ANGULAR MOMENTUM (JAM) collaboration, *How well do we know the gluon polarization in the proton?*, *Phys. Rev. D* **105** (2022) 074022 [arXiv:2201.02075] [INSPIRE].

[91] D.P. Anderle et al., *Probing gluon helicity with heavy flavor at the Electron-Ion Collider*, *Phys. Rev. D* **104** (2021) 114039 [arXiv:2110.04489] [INSPIRE].

[92] D. de Florian, R. Sassot, M. Stratmann and W. Vogelsang, *Evidence for polarization of gluons in the proton*, *Phys. Rev. Lett.* **113** (2014) 012001 [arXiv:1404.4293] [INSPIRE].

[93] NNPDF collaboration, *A first unbiased global determination of polarized PDFs and their uncertainties*, *Nucl. Phys. B* **887** (2014) 276 [arXiv:1406.5539] [INSPIRE].

[94] PHENIX collaboration, *Measurement of charged pion double spin asymmetries at midrapidity in longitudinally polarized $p + p$ collisions at $\sqrt{s} = 510\,GeV$*, *Phys. Rev. D* **102** (2020) 032001 [arXiv:2004.02681] [INSPIRE].

[95] STAR collaboration, *Longitudinal double-spin asymmetry for inclusive jet and dijet production in polarized proton collisions at $\sqrt{s} = 510\,GeV$*, *Phys. Rev. D* **105** (2022) 092011 [arXiv:2110.11020] [INSPIRE].

[96] L. Zheng, E.C. Aschenauer, J.H. Lee, B.-W. Xiao and Z.-B. Yin, *Accessing the gluon Sivers function at a future electron-ion collider*, *Phys. Rev. D* **98** (2018) 034011 [arXiv:1805.05290] [INSPIRE].

[97] V. Cirigliano, K. Fuyuto, C. Lee, E. Mereghetti and B. Yan, *Charged Lepton Flavor Violation at the EIC*, *JHEP* **03** (2021) 256 [arXiv:2102.06176] [INSPIRE].

[98] R. Boughezal et al., *Neutral-current electroweak physics and SMEFT studies at the EIC*, *Phys. Rev. D* **106** (2022) 016006 [arXiv:2204.07557] [INSPIRE].

[99] J.L. Zhang et al., *Search for $e \to \tau$ Charged Lepton Flavor Violation at the EIC with the ECCE Detector*, arXiv:2207.10261 [INSPIRE].

[100] H.T. Li, B. Yan and C.P. Yuan, *Jet charge: A new tool to probe the anomalous* $Zb\bar{b}$ *couplings at the EIC*, *Phys. Lett. B* **833** (2022) 137300 [arXiv:2112.07747] [INSPIRE].

[101] Y. Hatta, N. Mueller, T. Ueda and F. Yuan, *QCD Resummation in Hard Diffractive Dijet Production at the Electron-Ion Collider*, *Phys. Lett. B* **802** (2020) 135211 [arXiv:1907.09491] [INSPIRE].

[102] Y. Hatta, B.-W. Xiao, F. Yuan and J. Zhou, *Anisotropy in Dijet Production in Exclusive and Inclusive Processes*, *Phys. Rev. Lett.* **126** (2021) 142001 [arXiv:2010.10774] [INSPIRE].

[103] Y. Hatta, B.-W. Xiao, F. Yuan and J. Zhou, *Azimuthal angular asymmetry of soft gluon radiation in jet production*, *Phys. Rev. D* **104** (2021) 054037 [arXiv:2106.05307] [INSPIRE].

[104] Y. Hatta, B.-W. Xiao and F. Yuan, *Probing the Small-x Gluon Tomography in Correlated Hard Diffractive Dijet Production in Deep Inelastic Scattering*, *Phys. Rev. Lett.* **116** (2016) 202301 [arXiv:1601.01585] [INSPIRE].

[105] E. Iancu, A.H. Mueller and D.N. Triantafyllopoulos, *Probing Parton Saturation and the Gluon Dipole via Diffractive Jet Production at the Electron-Ion Collider*, *Phys. Rev. Lett.* **128** (2022) 202001 [arXiv:2112.06353] [INSPIRE].

[106] D. Neill and W.J. Waalewijn, *Entropy of a Jet*, *Phys. Rev. Lett.* **123** (2019) 142001 [arXiv:1811.01021] [INSPIRE].

[107] P. Gras et al., *Systematics of quark/gluon tagging*, *JHEP* **07** (2017) 091 [arXiv:1704.03878] [INSPIRE].

[108] E.M. Metodiev and J. Thaler, *Jet Topics: Disentangling Quarks and Gluons at Colliders*, *Phys. Rev. Lett.* **120** (2018) 241602 [arXiv:1802.00008] [INSPIRE].

[109] L.M. Dery, B. Nachman, F. Rubbo and A. Schwartzman, *Weakly Supervised Classification in High Energy Physics*, *JHEP* **05** (2017) 145 [arXiv:1702.00414] [INSPIRE].

[110] A. Andreassen, P.T. Komiske, E.M. Metodiev, B. Nachman and J. Thaler, *OmniFold: A Method to Simultaneously Unfold All Observables*, *Phys. Rev. Lett.* **124** (2020) 182001 [arXiv:1911.09107] [INSPIRE].

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

[112] PYTHIA6 with Radiative Corrections, https://eic.github.io/software/pythia6.html.

[113] D. Stump et al., *Inclusive jet production, parton distributions, and the search for new physics*, *JHEP* **10** (2003) 046 [hep-ph/0303013] [INSPIRE].

[114] G.A. Schuler and T. Sjostrand, *Low and high mass components of the photon distribution functions*, *Z. Phys. C* **68** (1995) 607 [hep-ph/9503384] [INSPIRE].

[115] B. Jager, M. Stratmann and W. Vogelsang, *Longitudinally polarized photoproduction of inclusive hadrons at fixed-target experiments*, *Eur. Phys. J. C* **44** (2005) 533 [hep-ph/0505157] [INSPIRE].

[116] B. Jager, *Photoproduction of single inclusive jets at future ep colliders in next-to-leading order QCD*, *Phys. Rev. D* **78** (2008) 034017 [arXiv:0807.0066] [INSPIRE].

[117] E.-C. Aschenauer, K. Lee, B.S. Page and F. Ringer, *Jet angularities in photoproduction at the Electron-Ion Collider*, *Phys. Rev. D* **101** (2020) 054028 [arXiv:1910.11460] [INSPIRE].

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

[119] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov and A. Smola, *Deep Sets*, *arXiv e-prints* (2017) arXiv:1703.06114 [`arXiv:1703.06114`].

[120] E. Wagstaff, F.B. Fuchs, M. Engelcke, I. Posner and M. Osborne, *On the Limitations of Representing Functions on Sets*, *arXiv e-prints* (2019) arXiv:1901.09006 [`arXiv:1901.09006`].

[121] B. Bloem-Reddy and Y. Whye Teh, *Probabilistic symmetries and invariant neural networks*, *J. Mach. Learn. Res.* **21** (2020) 3535, http://jmlr.org/papers/v21/19-322.html [`arXiv:1901.06082`].

[122] G.C. Blazey et al., *Run II jet physics*, in *Physics at Run II: QCD and Weak Boson Physics Workshop: Final General Meeting*, Batavia, U.S.A. (2000), pg. 47 [`hep-ex/0005012`] [INSPIRE].

[123] F. Chollet et al., *Keras*, https://github.com/fchollet/keras (2015).

[124] M. Abadi et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, http://tensorflow.org/ (2015).

[125] V. Nair and G.E. Hinton, *Rectified linear units improve restricted Boltzmann machines*, in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, Haifa, Israel (2010), pg. 807.

[126] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, `arXiv:1412.6980` [INSPIRE].

[127] D.R. Cox, *The regression analysis of binary sequences*, *J. Roy. Stat. Soc. B* **20** (1958) 215.

[128] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh and A. Talwalkar, *Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization*, *J. Mach. Learn. Res.* **18** (2017) 6765 [`arXiv:1603.06560`].

[129] T. O'Malley et al., *Kerastuner*, https://github.com/keras-team/keras-tuner (2019).

[130] R.D. Field and R.P. Feynman, *A Parametrization of the Properties of Quark Jets*, *Nucl. Phys. B* **136** (1978) 1 [INSPIRE].

[131] Z.-B. Kang, X. Liu, S. Mantry, M.C. Spraker and T. Wilson, *Dynamic Jet Charge*, *Phys. Rev. D* **103** (2021) 074028 [`arXiv:2101.04304`] [INSPIRE].

[132] A. Banfi, G.P. Salam and G. Zanderighi, *Infrared safe definition of jet flavor*, *Eur. Phys. J. C* **47** (2006) 113 [`hep-ph/0601139`] [INSPIRE].

[133] A. Banfi, G.P. Salam and G. Zanderighi, *Accurate QCD predictions for heavy-quark jets at the Tevatron and LHC*, *JHEP* **07** (2007) 026 [`arXiv:0704.2999`] [INSPIRE].

[134] S. Caletti, A.J. Larkoski, S. Marzani and D. Reichelt, *Practical jet flavour through NNLO*, *Eur. Phys. J. C* **82** (2022) 632 [`arXiv:2205.01109`] [INSPIRE].

[135] S. Caletti, A.J. Larkoski, S. Marzani and D. Reichelt, *A fragmentation approach to jet flavor*, *JHEP* **10** (2022) 158 [`arXiv:2205.01117`] [INSPIRE].

[136] R. Gauld, A. Huss and G. Stagnitto, *A dress of flavour to suit any jet*, Tech. Rep. BONN-TH-2022-17 (8, 2022) [INSPIRE].

[137] ATLAS collaboration, *Measurement of jet charge in dijet events from $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, *Phys. Rev. D* **93** (2016) 052003 [`arXiv:1509.05190`] [INSPIRE].

[138] CMS collaboration, *Measurements of jet charge with dijet events in pp collisions at* $\sqrt{s} = 8$ *TeV*, *JHEP* **10** (2017) 131 [arXiv:1706.05868] [INSPIRE].

[139] CMS collaboration, *Studies of Quark and Gluon Contributions to Jets using Jet Charge Measurements in pp and PbPb Collisions*, *Nucl. Phys. A* **1005** (2021) 121803 [arXiv:2004.14600] [INSPIRE].

[140] J. Lin, M. Freytsis, I. Moult and B. Nachman, *Boosting* $H \to b\bar{b}$ *with Machine Learning*, *JHEP* **10** (2018) 101 [arXiv:1807.10768] [INSPIRE].

[141] C.-W. Chiang, D. Shih and S.-F. Wei, *VBF vs. GGF Higgs with Full-Event Deep Learning: Towards a Decay-Agnostic Tagger*, *Phys. Rev. D* **107** (2023) 016014 [arXiv:2209.05518] [INSPIRE].

[142] ALICE collaboration, *The ALICE definition of primary particles*, ALICE-PUBLIC-2017-005 (2017) [INSPIRE].

[143] LHCb collaboration, *Identification of charm jets at LHCb*, 2022 *JINST* **17** P02028 [arXiv:2112.08435] [INSPIRE].

[144] CMS collaboration, *A new calibration method for charm jet identification validated with proton-proton collision events at* $\sqrt{s} = 13$ *TeV*, 2022 *JINST* **17** P03014 [arXiv:2111.03027] [INSPIRE].

[145] CMS collaboration, *Search for Higgs boson decay to a charm quark-antiquark pair in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, arXiv:2205.05550 [INSPIRE].

[146] CMS collaboration, *A search for the standard model Higgs boson decaying to charm quarks*, *JHEP* **03** (2020) 131 [arXiv:1912.01662] [INSPIRE].

[147] Y.L. Dokshitzer, V.A. Khoze and S.I. Troian, *On specific QCD properties of heavy quark fragmentation ('dead cone')*, *J. Phys. G* **17** (1991) 1602 [INSPIRE].

[148] ALICE collaboration, *Direct observation of the dead-cone effect in quantum chromodynamics*, *Nature* **605** (2022) 440 [arXiv:2106.05713] [*Erratum ibid.* **607** (2022) E22] [INSPIRE].

[149] H1 collaboration, *Measurement of the inclusive dijet cross-section in photoproduction and determination of an effective parton distribution in the photon*, *Eur. Phys. J. C* **1** (1998) 97 [hep-ex/9709004] [INSPIRE].

[150] X. Chu, E.-C. Aschenauer, J.-H. Lee and L. Zheng, *Photon structure studied at an Electron Ion Collider*, *Phys. Rev. D* **96** (2017) 074035 [arXiv:1705.08831] [INSPIRE].

[151] C. Uebler, A. Schäfer and W. Vogelsang, *Direct- and Resolved-Photon Threshold Resummation for Polarized High-p_T Hadron Production at COMPASS*, *Phys. Rev. D* **96** (2017) 074026 [arXiv:1708.08284] [INSPIRE].

[152] R. Boughezal, F. Petriello and H. Xing, *Inclusive jet production as a probe of polarized parton distribution functions at a future EIC*, *Phys. Rev. D* **98** (2018) 054031 [arXiv:1806.07311] [INSPIRE].