Precision studies for the partonic kinematics calculation through machine learning

D. F. Renteria-Estrada[a,b,*], R. J. Hernandez-Pinto[b], G. F. R. Sborlini[c,d], and Pia Zurita[e,f]

[a]*Instituto de Física Corpuscular, Universitat de València – Consejo Superior de Investigaciones Científicas,*
Parc Científic, E-46980 Paterna, Valencia, Spain.
*Speaker
[b]*Facultad de Ciencias Físico-Matemáticas, Universidad Autónoma de Sinaloa,*
Ciudad Universitaria, 80000 Culiacán, México.
[c]*Departamento de Física Fundamental e IUFFyM, Universidad de Salamanca,*
37008 Salamanca, Spain.
[d]*Escuela de Ciencias, Ingeniería y Diseño, Universidad Europea de Valencia,*
Paseo de la Alameda 7, 46010 Valencia, Spain.
[e]*Departamento de Física Teórica & IPARCOS,*
Universidad Complutense de Madrid, E-28040 Madrid, Spain.
[f]*Institut für Theoretische Physik, Universität Regensburg,*
93040 Regensburg, Germany, Universität Regensburg, Germany.

High Energy collider experiments are moving to the highest precision frontier quickly. The predictions of observables are based on the factorization formula, which helps to connect small to large distances. These predictions can be contrasted with experimental measurements and the success of this phenomenological approach is based on the correct description of nature. The application of the method to proton-proton colliders brings new challenges due to the proton structure and the detectors efficiency on reconstructing hadrons. Furthermore, since the non-perturbative distribution functions takes an important role to describe the experimental distributions, the presence of them makes the information of the partons diluted. At Leading Order (LO) in perturbative calculations, the momentum fractions involved in hard scattering processes are known exactly in terms of kinematical variables of initial and final states hadrons. However, at Next-to-Leading Order (NLO) and beyond, a closed analytical formula is not available. Furthermore, from the pure theoretical calculation, the exact definition of the momentum fraction is very challenging. In this work, we report a methodology based on Machine Learning techniques for the extraction of momentum fractions for $p + p \rightarrow \pi^+ + \gamma$ using a Monte Carlo simulation including quantum corrections up to Next-to-Leading Order in Quantum Chromodynamics and Leading Order in Quantum Electrodynamics. Our findings point towards a methodology to find the fundamental properties of the internal structure of hadrons because the reconstructed momentum fractions deeply relate our perturbative models with experimental measurements.

Keywords: High energy; proton-proton; quantum electrodynamics.

DOI: https://doi.org/10.31349/SuplRevMexFis.4.021134

1. Introduction

Technological advances have enabled important break-throughs in science, and in the case of high energy physics have allowed us to discover new particles and test models. To mention some specific examples, technology has favored perturbative Quantum Chromodynamics (pQCD) to achieve greater precision in computational calculations and to improve the efficiency of such calculations. The trend right now points towards Artificial Intelligence (AI) and Machine Learning (ML) techniques applied to high energy experiments in parton shower Monte Carlo (MC) reconstruction applied to jets physics [1], reconstruction of deep inelastic dispersion kinematics [2,3], or more phenomenological analysis, including the determination of partonic distribution densities by the NNPDF collaboration [4-11].

As is well known, the Quantum Chromodynamics (QCD) is a theory that studies the strong interactions between partons, but in hadronic collisions it is necessary to resort to pQCD theory. Calculating the scattering cross-section is a useful measure to study these interactions, so it is important to know the accessible kinematic of the detector how to reconstruct the underlying partonic momentum fractions. In a perturbation theory, these quantities are only accurately determined when analytical calculations are computed up to the Leading Order (LO) precision. However, obtaining the scattering cross-section at LO accuracy is not sufficient to describe precise experimental measurements. To achieve a more reliable the phenomenological description, higher-order calculations are required. In pQCD, this feature happens even for the first or second order, the so called Next-to-Leading Order (NLO) and Next-to-Next-to-Leading Order (NNLO) respectively. Furthermore, the problem grows when we start

combining the effects of the interaction of photons with partons and the QCD interactions, since non-perturbative effects must be considered, hadronization.

In this work, we propose a methodology based on Neural Networks using Machine Learning (ML) to reconstruct the partonic momentum fractions in photon-hadron production at colliders, including up to NLO QCD and LO QED correction. We base our work on Refs. [12, 13], which we briefly explain in Sec. 2. In Sec. 3 we report the phenomenological results of studying the hadronic differential cross-section distribution as a function of the phase-space of the system. Finally, in Sec. 3.2 we show the Neural Network (NN) results for predicting the partonic momentum fractions of the scattering process in the initial and final state at LO QED + NLO QCD accuracy.

2. Computational details

The efforts in the scientific community to obtain greater accuracy into quarks and gluons distributions, had contributed to solve the problem of the proton spin crisis. Still, there is a long way to pursue and one direction is the precise determination of the partonic moments in the pQCD theory. In this context, studying the parton momentum fractions in hadronic collisions constitutes an advance towards the search of greater precision since they carry an imprint of the parton level kinematics. In our study, we are interested to perform the MC simulation of the cross-section for the reaction,

$$ p\, p \to \pi^+ + \gamma \,, \tag{1} $$

where the photon shall be produced by the direct interaction of the original partons in the collision, which we will refer to as the hard photon. Fortunately, the dispersion between hadrons can be extracted by means of hard process factorization properties in the QCD [14]. Therefore, the differential cross-section of the proton-proton process is obtained through the convolution of the Parton Distribution Function (PDF), Fragmentation Function (FF) and the differential partonic cross-section. The non-perturbative part corresponds to the PDF and the FF, which provide us the information on the probabilities of finding the partons inside the colliding hadrons and the probability of generating a given meson/hadron from a certain parton, respectively. Explicitly, the differential cross-section of producing a positively charged pion in addition with a photon in the proton-proton collisions is given by,

$$ d\sigma_{p_1\,p_2\to\pi\,\gamma} = \sum_{a_1 a_2 a_3 a_4} \int dx_1 dx_2 dz dz'\, f_{a_1}^{(p_1)}(x_1,\mu_I) $$
$$ \times f_{a_2}^{(p_2)}(x_2,\mu_I)\, d_{a_3}^{(\pi)}(z,\mu_F) d_{a_4}^{(\gamma)}(z',\mu_F) $$
$$ \times d\hat{\sigma}_{a_1\,a_2\to a_3\,a_4}(x_1,x_2,z,z',\mu_R) \,, \tag{2} $$

where we sum over all flavors of quarks and gluons a_1, a_2 and a_3, $f_a^{(h)}(x,\mu_I)$ represents the PDF which indicates the

probability of finding the parton a inside the hadron h with momentum fraction x and initial energy scale μ_I. Analogously, $d_b^{(h)}(z,\mu_F)$ stands for the probability that the parton b produced during the hard scattering, hadronizes in the hadron h with momentum fraction z at a typical energy scale μ_F. Finally, we cast in $d\hat{\sigma}_{a_1\,a_2\to a_3\,a_4}$ the differential partonic cross-section of the interaction between the partons a_1 and a_2 producing the partons a_3 and a_4. In practice, the nature of the detected photons is indistinguishable, however, there are reconstructions algorithms that permit to retain a high proportion of photons expected to come directly from the parton-level interaction (*i.e.* from the hard-scattering). For this purpose, we impose a criterion to isolate photons [15] which allows us to obtain independence of the parton-to-photon fragmentation function[i] and also to solve the collinear divergences. Therefore, the differential cross-section of producing a positively charged pion plus a hard photon into proton-proton collision can be rewritten as follows,

$$ d\sigma_{p_1\,p_2\to\pi\,\gamma} = \sum_{a_1 a_2 a_3} \int dx_1 dx_2 dz\, f_{a_1}^{(p_1)}(x_1,\mu_I) $$
$$ \times f_{a_2}^{(p_2)}(x_2,\mu_I)\, d_{a_3}^{(\pi)}(z,\mu_F) $$
$$ \times d\hat{\sigma}_{a_1\,a_2\to a_3\,\gamma}^{\mathrm{ISO}}(x_1,x_2,z,\mu_R), \tag{3} $$

where $d\hat{\sigma}_{a_1\,a_2\to a_3\,\gamma}^{\mathrm{ISO}}$ is the differential cross-section to generate a parton a_3 and an isolated photon from the interaction between partons a_1 and a_2 taking into account the isolation algorithm. In our case, we perform the calculation considering including up to LO QED + NLO QCD. In this way, the partonic cross-section is the sum of the QED plus the QCD contributions, namely,

$$ d\hat{\sigma}_{a_1\,a_2\to a_3\,\gamma}^{\mathrm{ISO}} = d\hat{\sigma}_{a_1\,a_2\to a_3\,\gamma}^{\mathrm{ISO,\,QED}} + d\hat{\sigma}_{a_1\,a_2\to a_3\,\gamma}^{\mathrm{ISO,\,QCD}} \,. \tag{4} $$

Once we have applied the isolation criteria, the partonic channels for the $2 \to 2$ processes in QCD are,

$$ q\bar{q} \to \gamma g, \quad qg \to \gamma q \,, \tag{5} $$

meanwhile, the channels in the QED interaction are,

$$ q\bar{q} \to \gamma\gamma, \quad q\gamma \to \gamma q \,, \tag{6} $$

and, finally, the processes $2 \to 3$ considered in the NLO contribution of QCD are those as,

$$ q\bar{q} \to \gamma gg \,, \quad qg \to \gamma gq \,, \quad gg \to \gamma q\bar{q} \,, $$
$$ q\bar{q} \to \gamma Q\bar{Q} \,, \quad qQ \to \gamma qQ \,. \tag{7} $$

3. Phenomenological results

The calculations described in Sec. 2 were implemented in the MC integrator based in Ref. [12] and considered two different scenarios for the scattering kinematics. In the first scenario, we seek to study the kinematic configuration of PHENIX in the RHIC experiment with a center-of-mass (c.m.) energy of

$\sqrt{S_{\mathrm{CM}}} = 500$ GeV, and considering the phenomenological cuts,

$$\{|\eta^\pi|, |\eta^\gamma|\} \leq 0.35 \,,$$

$$p_T^\pi \geq 2 \ \mathrm{GeV} \,,$$

$$5 \, \mathrm{GeV} \leq p_T^\gamma \leq 15 \ \mathrm{GeV}, \tag{8}$$

where η and p_T are the pseudorapidity and transverse momentum, respectively. On the other hand, in the second scenario we explore a c.m. energy of $\sqrt{S_{\mathrm{CM}}} = 13$ TeV corresponding to LHC RUN II configuration, keeping the same cuts for the pion kinematics. However, motivated by studies about the photon detection efficiency in ATLAS and CMS in different energies [17–19], we impose the following cuts on the photon kinematics,

$$|\eta^\gamma| \leq 2.5 \,,$$

$$p_T^\gamma \geq 30 \ \mathrm{GeV}. \tag{9}$$

In addition, it is important to mention that we implement the restriction $|\phi^\pi - \phi^\gamma| \geq 2$ on the azimuthal angles of the pion and photon, in order to keep a *back-to-back* configuration. Finally, we set the renormalization and factorization scales as the average of the pion and photon transverse momenta,

$$\mu_R = \mu_I = \mu_F = \frac{p_T^\pi + p_T^\gamma}{2} \,. \tag{10}$$

3.1. Partonic kinematics analysis

Now, we study the momentum fraction distributions with the purpose of imposing constraints on the NN training. As it is well known, the momentum fractions are exactly determined at LO accuracy, explicitly as follows,

$$x_{1,2} = \frac{p_T^\gamma \left(\exp\left(\eta^{\pm\pi} \right) + \exp\left(\eta^{\pm\gamma} \right) \right)}{\sqrt{S_{CM}}} \,, \tag{11}$$

$$z = \frac{p_T^\pi}{p_T^\gamma} \,. \tag{12}$$

To compute Eq. (3), we use in the MC simulation the PDF set `NNPDF4.0NLO` [10] for the NLO QCD prediction, while for LO QED + NLO QCD quantum corrections we used the set `NNPDF3.1luxQEDNLO` [20–23]. Likewise, we selected the most up-to-date FF of pions at NLO given by collaboration `DSS2014` [24, 25]. Additionally, by simplicity we define the same partonic momentum fraction for both protons as $x = \{x_1, x_2\}$.

In Fig. 1 we display the differential cross-section per bin of the $pp \rightarrow \pi^+ + \gamma$ scattering process as a function of the partonic momentum fraction x, both for energies of RHIC (right) and LHC RUN II (left) experiments. A significant contribution to higher-order corrections is observed with respect to the LO contribution in QCD (red dots). The effects of QED and QCD (blue dashed line) are smaller with respect to only NLO QCD contribution (black line). Furthermore, as
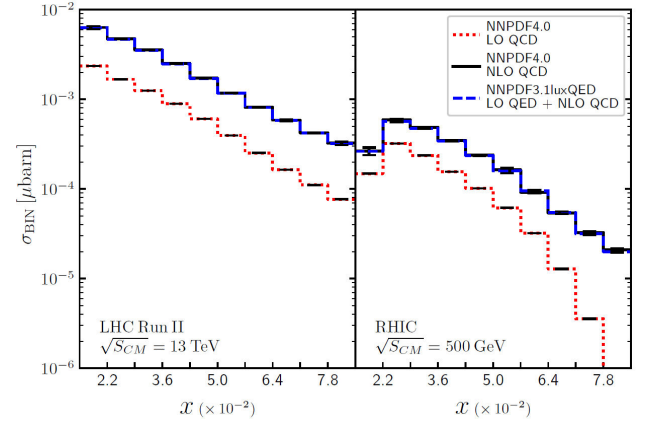

FIGURE 1. Cross-section as a function of the partonic momentum fractions, for RHIC and LHC Run II .
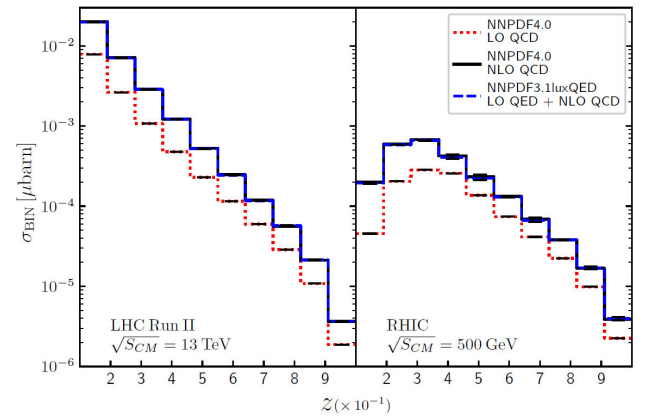

FIGURE 2. Cross-section as a function of the partonic momentum fractions, for RHIC and LHC Run II.

expected from LO, the maximum limit imposed on the photon momentum $p_T^\gamma \leq 15$ GeV at RHIC energies, manifests itself in a peak of the $d\sigma_{pp \rightarrow \pi+\gamma}$ distribution at $x \approx 2.5 \times 10^{-2}$, both for the Born level and for higher-orders. Similarly, in Eq. (11) the p_T and η restrictions in RHIC, limit the range that x can be computed to $x_{\max} \approx 0.01$. For this reason, we limited our analysis to observing the LHC kinematics in the same RHIC range. As a matter of fact, in Fig. 2 we show an analysis of the differential cross-section distribution analogous to Fig. 1 but as a function of z. Here, we can see that the LHC kinematics implies that cross-section decreases faster than the calculated with RHIC kinematics. The distributions show a peak at $z_{\mathrm{peak}} \approx 0.35$ in RHIC whereas $z_{\mathrm{peak}} \approx 0.25$ for the LHC RUN II. It is important to mention that those peaks correspond to a greater extent to the constraints implemented in the FF.

3.2. NN prediction

Taking into account the effects of kinematical cuts on the cross-section distribution, we will focus our efforts on reconstructing the partonic momentum fractions x and z. Since we now know that in the LO case our phase-space is well-defined, we will use this information to check the reliability

TABLE I. Architecture for the MLP best fit parameters for the reconstruction of the momentum fractions at LO in QCD: X_{REC}(LO) and Z_{REC}(LO) (second and third columns), and for the momentum fractions at NLO QCD + LO QED: X_{REC}(NLO) and Z_{REC}(NLO) (fourth and fifth columns).

	X_{REC} (LO)	Z_{REC} (LO)	X_{REC} (NLO)	Z_{REC} (NLO)
# of hidden layers	2	1	5	5
# of neurons/layer	200	100	300	300
activation function	ReLU	ReLU	ReLU	ReLU
# iterations	1×10^5	1×10^5	1×10^{12}	1×10^{12}
learning rate	1×10^{-3}	1×10^{-3}	1×10^{-4}	1×10^{-4}

of our method. In this task, we implement a NN with Python 3 library `scikit-learn` [26], specifically, we train a Multilayer Perceptron (MLP) whose architecture is shown in Table I. For this purpose, we define X_{REC} and Z_{REC} as target variables of the NN. For the training to predict the x variable in the LO case, we set the MLP with two hidden layers and 200 neurons per layer, using 10^5 interactions per event number. In addition, we find some parameters sensitive to the output variables, such as the maximum number of interactions and the learning rate. Finally, for all architectures, we use the Rectified Linear Unit (ReLU) activation function.

The methodology to train the NN consists of generating random points (x, z), assigning them as the target variables. In this way, we use 20% of the data to test the results of the NN and 80% for training the NN. In the LO case, the inputs values are the kinematic variables accessible in the experimental environment, this mean,

$$\mathcal{V}_{\text{Exp}} = \{p_T^\gamma, p_T^\pi, \eta^\gamma, \eta^\pi, \cos(\phi^\pi - \phi^\gamma)\}. \qquad (13)$$

Conversely, due to the discussion in Sec. 3.1, the kinematic cutoffs imposed restrict the range in which we can measure x and z, therefore, we reconstruct the partonic momentum fractions in the range $x \in (0.015, 0.075)$ and $z \in (0.1, 1)$.

The results at LO are displayed in Fig. 3, where we have defined the variables X_{REAL} and Z_{REAL} as the variables randomly generated by our MC simulation. The color scale indicates the probability that the Neural Network predicts $\{X_{\text{REC}}, Z_{\text{REC}}\}$ in the same bin as $\{X_{\text{REAL}}, Z_{\text{REAL}}\}$. With this architecture, the NN can access the momentum fractions with probability greater than 90% for $x < 6.9 \times 10^{-2}$, and a probability around 80% for $x > 6.9 \times 10^{-2}$. Meanwhile, the NN trained to predict z fraction, have simpler architecture and reconstructs the momentum fraction with success rate close to 100% in $z > 0.19$. At the Born level, x is proportional to the photon transverse momentum, while z is inversely proportional, so the imposed upper bound of $P_T^\gamma \leq 15$ GeV has repercussions on the reconstruction of x when large values of X_{REAL} and low values of Z_{REAL} are explored.

Once we have checked the advantage of the NN, we train it to perform predictions at NLO level. However, since having a better fit becomes more complex, we opt for a more robust architecture. By the nature of the process, the experiment is not able to distinguish between a LO and NLO contribution, therefore, for the NLO QCD + LO QED correction it is necessary to discretize the kinematic variables used as input values namely,
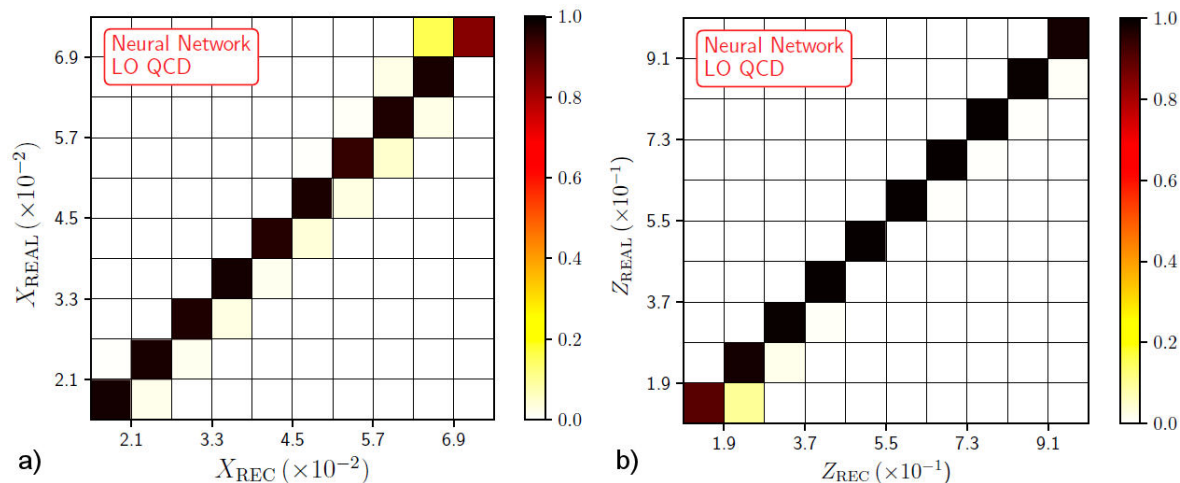

a)

b)

FIGURE 3. Comparison of the momentum fractions at LO accuracy obtained with MLP neural networks with the parameters given in Table I.
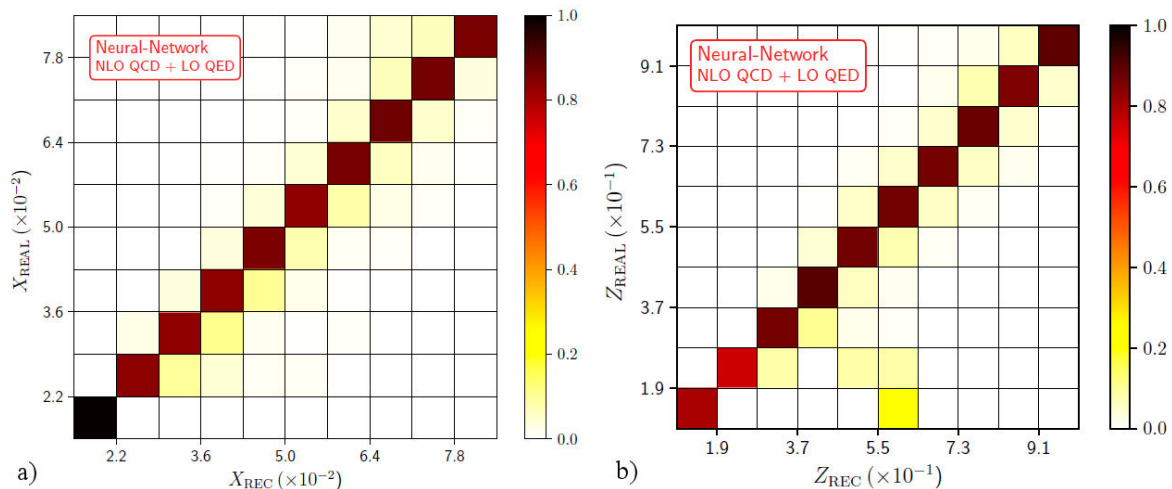
FIGURE 4. Comparison of the momentum fractions at LO QED + NLO QCD accuracy obtained with MLP neural networks with the parameters given in Table I.

$$p = \{\bar{p}_T^\gamma, \bar{p}_T^\pi, \bar{\eta}^\gamma, \bar{\eta}^\pi, \overline{\cos}(\phi^\pi - \phi^\gamma)\} \in \bar{\mathcal{V}}_{\text{Exp}}, \qquad (14)$$

where $\bar{a}$ denotes the mean value of each variable of $\mathcal{V}_{\text{Exp}}$ per bin. Now we discretize the space with ten bins for p_T, five bins for η and six bins for $\cos(\phi^\pi - \phi^\gamma)$. This leads us to weigh the partonic momentum fractions with the average cross-section corresponding to the bin of the event $\bar{\mathcal{V}}_{\text{Exp}}$. Explicitly, the target variables are now,

$$X_{1,\text{REC}} = \sum_i (x_1)_i \frac{d\sigma_j}{dx_1}(p_j; (x_1)_i), \qquad (15)$$

$$X_{2,\text{REC}} = \sum_i (x_2)_i \frac{d\sigma_j}{dx_2}(p_j; (x_2)_i), \qquad (16)$$

$$Z_{\text{REC}} = \sum_i z_i \frac{d\sigma_j}{dz}(p_j; z_i), \qquad (17)$$

The results at NLO QCD + LO QED accuracy are presented in Fig. 4, for which we use a MLP whose parameters are those of Table I (fourth and fifth columns). We generated 10^9 events for LO case, from which only 30% passed the kinematic cuts. In addition, the discretization of the input and output values at NLO led to generate 15000 variables in our phase-space. For both x and z, the NN showed favorable results and a remarked improvement over the work made in Ref. [12]. The prediction for x is excellent in the lowest rank bin, while in the rest bins the probability of reconstruction oscillates around 80%. On the other hand, the fraction z achieves the best prediction in the range $z \in (0.28, 1)$; on the other hand, we find that for low values of z, our result continues to show a strong correlation.

4. Conclusions and outlook

In the context of hadron collisions, it is difficult to obtain closed analytical formulae to related parton momenta fraction and experimentally-accessible variables. This fact leads

us to use only approximate formulas in the calculation of the scattering cross-section. In this work, we seek to access the partonic momentum fractions for the production of a hadron plus a direct photon in proton-proton collisions using ML tools. First, we calculated the cross-section differential distribution of the scattering process $pp \rightarrow \pi^+ + \gamma$ with RHIC and LHC RUN II kinematics. The results indicated that the kinematical cuts imposed by both experiments limit the range of accessibility of x. On the other hand, we noted that RHIC experiment has a peak in both x and z, which was observed due to restrictions of p_T^γ and pion fragmentation. Likewise, we have built a MLP NN to access the partonic momentum fractions at NLO QCD + LO QED accuracy. In the first instance, we obtained predictions with LO QCD accuracy to compare the fitting efficiency of the NN, and we obtained nearly perfect results. Subsequently, with the help of Deep Learning, we accessed the partonic momentum fractions with NLO QCD + LO QED precision getting results with higher accuracy. Finally, this work not only lays the foundation for higher accuracy predictions of High Energy Physics, but also, as mentioned in Ref. [27], it can help to find constraints on the final states of heavy hadrons.

Acknowledgements

i. The smooth isolation prescription algorithm can be found in Refs. [15,16].

1. K. Cranmer, *et al.*, Reframing Jet Physics with New Computational Methods, *EPJ Web Conf.* **251** (2021) 03059, https://doi.org/10.1051/epjconf/202125103059.

2. M. Diefenthaler *et al.*, Deeply learning deep inelastic scattering kinematics, *Eur. Phys. J. C* **82** (2022) 1064, https://doi.org/10.1140/epjc/s10052-022-10964-z.

3. M. Arratia *et al.*, Reconstructing the kinematics of deep inelastic scattering with deep learning, *Nucl. Instrum. Meth. A* **1025** (2022) 166164, https://doi.org/10.1016/j.nima.2021.166164.

4. R. D. Ball *et al.*, A determination of parton distributions with faithful uncertainty estimation, *Nucl. Phys. B* **809** (2009) 1, https://doi.org/10.1016/j.nuclphysb.2008.09.037.

5. R. D. Ball *et al.*, Parton distributions with LHC data, *Nucl. Phys. B* **867** (2013) 244, https://doi.org/10.1016/j.nuclphysb.2012.10.003.

6. T. N. Collaboration *et al.*, Neural network determination of parton distributions: the nonsinglet case, *JHEP* **2007** (2007) 039, https://dx.doi.org/10.1088/1126-6708/2007/03/039.

7. S. Forte *et al.*, Neural network parametrization of deep-inelastic structure functions, *JHEP* **2002** (2002) 062, https://dx.doi.org/10.1088/1126-6708/2002/05/062.

8. R. D. Ball *et al.*, Parton distributions for the LHC run II, *JHEP* **2015** (2015) 40, https://doi.org/10.1007/JHEP04(2015)040.

9. R. Abdul Khalek *et al.*, A first determination of parton distributions with theoretical uncertainties, *Eur. Phys. J. C* **79** (2019) 838, https://doi.org/10.1140/epjc/s10052-019-7364-5.

10. R. D. Ball *et al.*, The path to proton structure at 1% accuracy, *Eur. Phys. J. C* **82** (2022) 428, https://doi.org/10.1140/epjc/s10052-022-10328-7.

11. J. Rojo and J. I. Latorre, Neural network parametrization of spectral functions from hadronic tau decays and determination of QCD vacuum condensates, *JHEP* **2004** (2004) 055, https://dx.doi.org/10.1088/1126-6708/2004/01/055.

12. D. de Florian and G. F. R. Sborlini, Hadron plus photon production in polarized hadronic collisions at next-to-leading order accuracy, *Phys. Rev. D* **83** (2011) 074022, https://doi.org/10.1103/PhysRevD.83.074022.

13. D. F. Rentería Estrada *et al.*, Reconstructing partonic kinematics at colliders with machine learning, *SciPost Phys. Core* **5** (2022) 049, https://doi.org/10.21468/SciPostPhysCore.5.4.049.

14. J. C. Collins, D. E. Soper, and G. Sterman, Factorization of Hard Processes in QCD, *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1, https://doi.org/10.1142/9789814503266_0001.

15. S. Frixione, Isolated photons in perturbative QCD, *Phys. Lett. B* **429** (1998) 369, https://doi.org/10.1016/S0370-2693(98)00454-7.

16. D. F. Rentería-Estrada, R. J. Hernández-Pinto, and G. Sborlini, Analysis of the Internal Structure of Hadrons Using Direct Photon Production, *Symmetry* **13** (2021) 942, https://doi.org/10.3390/sym13060942.

17. P. Kontaxakis, The Level-1 CMS electron and photon trigger for the LHC Run II, *PoS LHCP* **2018** (2018) 073, https://doi.org/10.22323/1.321.0073.

18. G. Aad *et al.*, Performance of electron and photon triggers in ATLAS during LHC Run 2, *Eur. Phys. J. C* **80** (2020) 47, https://doi.org/10.1140/epjc/s10052-019-7500-2.

19. G. Aad *et al.*, Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data, *JINST* **14** (2019) P12006, https://dx.doi.org/10.1088/1748-0221/14/12/P12006.

20. J. M. Campbell *et al.*, Direct photon production and PDF fits reloaded, *Eur. Phys. J. C* **78** (2018) 470, https://doi.org/10.1140/epjc/s10052-018-5944-4.

21. V. Berton *et al.*, Illuminating the photon content of the proton within a global PDF analysis, *SciPost Phys.* **5** (2018) 008, https://doi.org/10.21468/SciPostPhys.5.1.008.

22. A. Manohar *et al.*, How Bright is the Proton? A Precise Determination of the Photon Parton Distribution Function, *Phys. Rev. Lett.* **117** (2016) 242002, https://doi.org/10.1103/PhysRevLett.117.242002.

23. A. V. Manohar *et al.*, The photon content of the proton, *JHEP* **2017** (2017) 46, https://doi.org/10.1007/JHEP12(2017)046.

24. D. de Florian *et al.*, Parton-to-pion fragmentation reloaded, *Phys. Rev. D* **91** (2015) 014035, https://doi.org/10.1103/PhysRevD.91.014035.

25. D. de Florian, R. Sassot, and M. Stratmann, Global analysis of fragmentation functions for protons and charged hadrons, *Phys. Rev. D* **76** (2007) 074033, https://doi.org/10.1103/PhysRevD.76.074033.

26. F. Pedregosa *et al.*, Scikit-learn: Machine Learning in Python, *J. Machine Learning Res.* **12** (2011) 2825-2830,

27. S. A. Ochoa-Oregon *et al.*, Constraining fragmentation functions through hadronphoton production at higher-orders, *Phys. Rev. D* **107** (2023) 096002, https://doi.org/10.1103/PhysRevD.107.096002.