

AI4EIC Hackathon

AI4EIC Hackathon Team

October 14, 2022

Abstract

For this hackathon we propose problems of different difficulty, starting from one that is accessible to everyone, that are deemed to be solvable in a one-day event. We focus on the dual-radiator Ring Imaging Cherenkov (dRICH) detector under development as part of the particle-identification (PID) system at the future Electron-Ion Collider (EIC). Data are produced using the ePIC software stack. We choose as a working point a range in momentum around 15 GeV, which corresponds to a region where potentially both radiators can contribute to the π/K separation (and certainly at larger momenta the gas radiator can contribute to PID). Despite the inherent ‘simplicity’ of the problems, given the approximations made as explained in this document, this event can potentially become a first step towards machine learning/deep learning application for PID with the dRICH.

1 Introduction to dRICH

The dRICH (dual-radiator Ring Imaging Cherenkov detector) [1] is a PID detector that will provide full hadronic ($\pi/K/p$ better than 3σ apart) from a few GeV/c up to $\approx 50GeV/c$ in the hadron going direction at ePIC. The dRICH consists of 6 identical, transversely open sectors. Each contains two radiators (aerogel and C_2F_6 gas), sharing the same outward focusing mirror and readout planes, which are instrumented with highly segmented photosensors ($3\text{ mm} \times 3\text{ mm}$ pixels), located outside of charged particle acceptance. The photo sensor tiles are arranged on a curved surface to compensate aberrations. Photons from a Cherenkov cone may split over two or more sectors thanks to the open geometry of the dRICH sectors.

Sketch of the dRICH is shown in Fig. 1. Also, an interactive visualization of the dRICH detector can be found [here](#)

2 Details on the data

2.1 Format of the data

Training and test data is given in `.csv` format. For all the questions in the hackathon the format of data remains the same. Structure of data is as follows:

1. Total number of columns in all datasets are 185.
2. Each row corresponds to a single event with a single type of particle (π^+ or K^+).
3. First 5 columns include `eventID`, `PID`, `momentum [GeV]`, `theta [θ in degree ($^\circ$)]`, and `phi [ϕ in degree ($^\circ$)]`. Note that, the θ , ϕ and momentum (p) correspond to their values at the interaction point where the collision of the beams occur.
4. The rest of 180 columns are named as
 - (a) (`X0`, `X1`, `.....`, `X59`), (`Z0`, `Z1`, `.....`, `Z59`), (`Y0`, `Y1`, `.....`, `Y59`), corresponds to the positions on the photo-sensor of the detected optical photons produced by the radiators (gas and aerogel), due to the tracks.

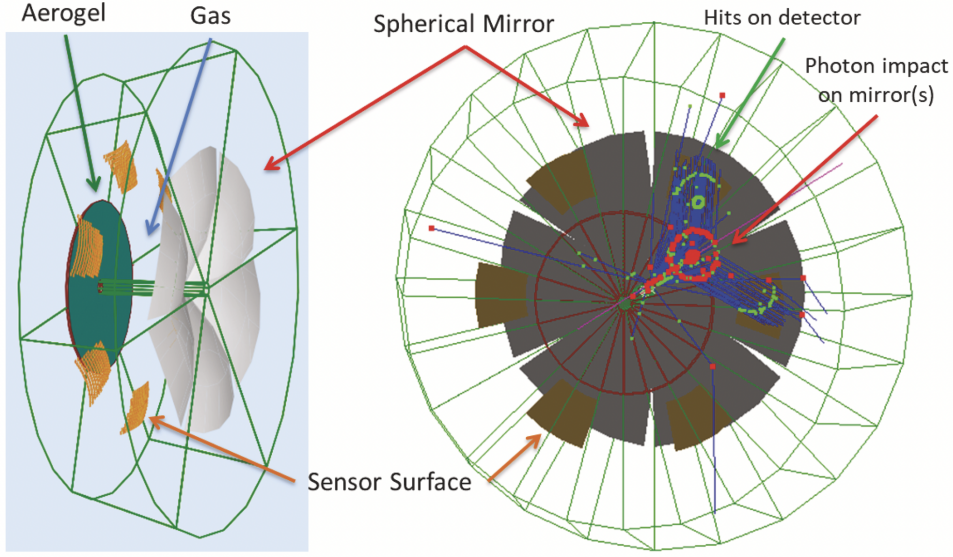


Figure 1: Figure from [2]. dRICH 3D model: downstream view (left) and upstream view (right) with example of a 10 GeV/c simulated pion: red and green points represent reflection and absorption respectively. In the represented event, the photons reflect on a single mirror and then split on two sectors; the aerogel photons end in two different sensors (two large ring arcs) while the gas photons concentrate in a single sensor (small complete ring); few photons scatter around. The open sectors geometry facilitates the collection of the largest amount of informative photons, keeping relatively small the overall photon detector surface.

- (b) The point (X_N, Y_N, Z_N) corresponds to the location of N^{th} detected optical photon. For ease of data formatting, the maximum number of detected optical photons are restricted to 60 (not so likely to have more than 60 detected optical photons in the kinematic settings in the following problems). When the number of detected optical photons (N) is less than 60, the corresponding $(X_N - X_{59}, Y_N - Y_{59}, Z_N - Z_{59})$ are padded with $(0., 0., 0.)$.
- (c) The points (X_N, Y_N, Z_N) are in [mm].

The format of data is shown in the table 1

eventID	PID	momentum	theta	phi	X0	...	X59	Y0	...	Y59	Z0	...	Z59
0	211 or 321	p [GeV/c]	θ (deg)	ϕ (deg)	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]
1	211 or 321	p [GeV/c]	θ (deg)	ϕ (deg)	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]
..	211 or 321	p [GeV/c]	θ (deg)	ϕ (deg)	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]
..	211 or 321	p [GeV/c]	θ (deg)	ϕ (deg)	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]
N	211 or 321	p [GeV/c]	θ (deg)	ϕ (deg)	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]

Table 1: Table summarizing the data format for the hackathon problems

2.2 Details on simulation

A sample event in an XY plane for a π^+ and K^+ is shown in Fig 2.

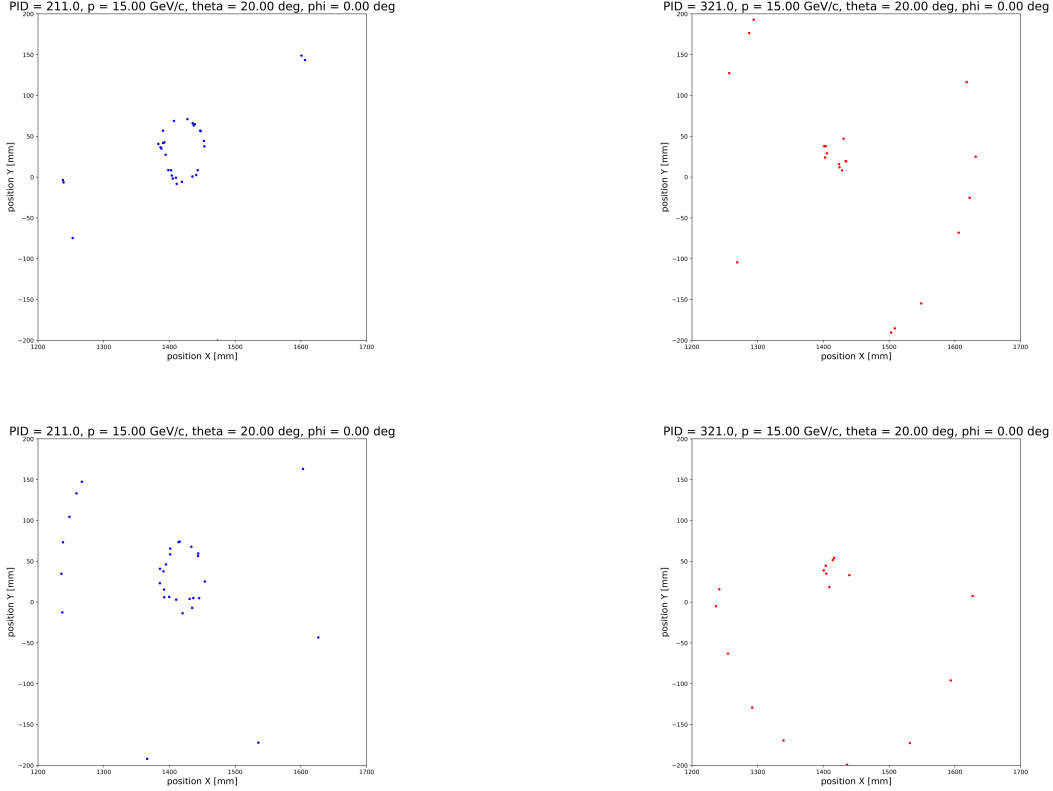


Figure 2: XY projection of an event corresponding to pion and kaon. The smaller ring is caused by the gaseous radiator (smaller refractive index) while the larger ring is caused by the aerogel radiator (larger refractive index). The figures on the left correspond to π^+ while the figures on the right correspond to K^+ .

3 Question 1

3.1 Scenario (Statement)

Training Events	1.5 Million Events	With Magnetic Field ($\sim 1.5T$)
Momentum	15 GeV/c	at Interaction Point (0, 0, 0)
Theta θ	20°	at Interaction Point (0, 0, 0)
Phi ϕ	0°	at Interaction Point (0, 0, 0)

Table 2: Kinematics for data corresponding to Question 1.

3.2 The challenge...

The challenge here is to classify between pions (π^+ PID = 211) and kaons (K^+ PID = 321) given the positions of the optical photon hits in the dRICH detector. A test data set of size 100k events is given to the participants

for which predictions will be accepted at the [leaderboard](#).

Please refer the section 7 on how the results have to be formatted for judging.

4 Question 2

4.1 Scenario (Statement)

Training Events	3 Million Events	With Magnetic Field ($\sim 1.5T$)
Momentum	$15 - 20 GeV/c$	at Interaction Point (0, 0, 0)
Theta θ	$15 - 16^\circ$	at Interaction Point (0, 0, 0)
Phi ϕ	$0 - 5^\circ$	at Interaction Point (0, 0, 0)

Table 3: Kinematics for data corresponding to Question 2. The kinematics are produced in a continuous range instead of a fixed kinematics.

4.2 The challenge...

The challenge here is to classify between pions (π^+ PID = 211) and kaons (K^+ PID = 321) given the positions of the optical photon hits in the dRICH detector. A test data set of size $100k$ events is given to the participants for which predictions will be accepted at [leaderboard](#).

5 Question 3

5.1 Scenario (Statement)

Training Events	3 Million Events	With Magnetic Field ($\sim 1.5T$)
Momentum	$15 - 20 GeV/c$	at Interaction Point (0, 0, 0)
Theta θ	$15 - 16^\circ$	at Interaction Point (0, 0, 0)
Phi ϕ	$0 - 5^\circ$	at Interaction Point (0, 0, 0)

Table 4: Kinematics for data corresponding to Question 3. The kinematics are produced in a continuous range instead of a fixed kinematics. There are random noise hits (Average of 1/10th the expected number optical photons) added on top of each event.

5.2 The challenge...

Can one come up with an intelligent solution to classify π^+ and K^+ even in a noisy environment. To each event, random noisy hits (Average 1/10th of photon yield) are added on top of the detected optical photons due to the charged particle interaction in dRICH. The challenge here is to classify between pions (π^+ PID = 211) and kaons (K^+ PID = 321) given the positions of the detected optical photon hits in the dRICH detector. A test data set of size $100k$ events is given to the participants for which predictions will be accepted at [leaderboard](#).

6 Resources provided

Participants on request, are provided with `aws` instances to use during the hackathon event. To facilitate participants to efficiently train, a modular template has been given to the participants as well. The modular template

loads data (training) and one can plug and play different models during the hackathon event. Participants can choose between the PyTorch or TensorFlow templates that has been provided. We highly recommend the use of dataloader and the template for all participants and in particular the participants who use `aws` instances.

7 Result Submission format

Participants are given the `test` dataset with `eventID` in their columns. Make your predictions should be submitted in its PDG code only for π^+ (211) and K^+ (321). Make sure to submit the prediction files in the form of `.csv` files. The submission portal accepts only `.csv` format. **NOTE: PLEASE DO NOT HAVE ANY HEADER IN THE PREDICTION FILE.**

A sample format for the prediction file can be formatted like the table below (make sure the column separates are `;` (`.csv`). The prediction file is usually not bigger than `1MB` and the submission portal has a file size limit of `3MB`.

0	Predicted label (211/321)
1	Predicted label (211/321)
...	
...	
99999	Predicted label (211/321)

8 Judging mechanism and time limit

8.1 Judging mechanism

Participants can make multiple submissions to the problems throughout the hackathon event. There is no constraint on the number of times a solution is submitted to the given problem. The solutions to problems that are submitted by the participants are evaluated on the basis of accuracy, The accuracy is calculated as:

$$\text{ACCURACY} = \frac{\sum_{N_{test}} |y_{pred} == y_{true}|}{N_{test}}$$

The problems are curated in increasing order of difficulty. Therefore, scoring is done accordingly depending on the problem. Each problem has a threshold on its accuracy which is based on a "custom" benchmark study that was done ahead in time.

Scoring is done in the following manner: If a solution submitted to a problem has its accuracy less than its threshold, then a score of 0.0 is returned. (Do not give up keep trying...)

$$\text{SCORE} = 0.0 \quad \{\text{IF ACCURACY} < \text{THRESHOLD}\}$$

If the solution submitted to the problem is above its threshold, then the accuracy is scaled such that the score is between 50 and 100.

$$\text{SCORE} = 50.0 + 50.0 \times \frac{(\text{ACCURACY} - \text{THRESHOLD})}{100.0 - \text{THRESHOLD}}$$

The threshold for the problems are given below.

Problem Number	Threshold Accuracy
Problem 1	94%
Problem 2	86%
Problem 3	80%

8.2 Time limit

The hackathon starts at 10:00AM E.T on 14th October 2022, and submissions will be accepted until 5:00PM E.T on 14th October 2022. Submissions beyond 5:00PM E.T on 14th October 2022, will not be considered for the prize for the hackathon. However, for participants who are interested in getting continuing to work on the problems post hackathon, can continue to submit solutions until 5:00PM E.T 17th October 2022. This is just to encourage interested participants, This submissions will not be a part of hackathon evaluation.

References

- [1] E. Cisbani, A. Del Dotto, C. Fanelli, M. Williams, M. Alfred, F. Barbosa, L. Barion, V. Berdnikov, W. Brooks, T. Cao, et al. AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case. *Journal of Instrumentation*, 15(05):P05009, 2020.
- [2] A. Del Dotto, C.-P. Wong, and et. al. Design and r&d of rich detectors for eic experiments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 876:237–240, 2017. ISSN 0168-9002. doi: <https://doi.org/10.1016/j.nima.2017.03.032>. URL <https://www.sciencedirect.com/science/article/pii/S0168900217303728>. The 9th international workshop on Ring Imaging Cherenkov Detectors (RICH2016).