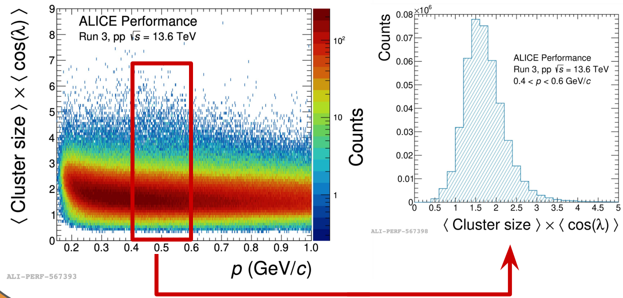
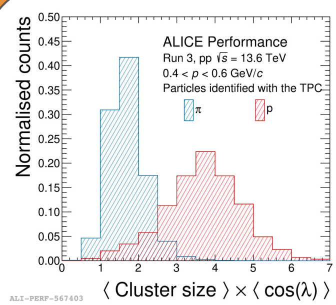
Particle Identification with Machine Learning for ALICE ITS2

Giorgio Alberto Lucia (giorgio.lucia@edu.unito.it) on behalf of the ALICE Collaboration
6th Inter-experiment Machine Learning Workshop, CERN (Genève), 29/01-02/02/2024

- Particle identification (PID) information is provided by various ALICE detectors through complementary experimental techniques
- The Run 2 ALICE Inner Tracking System (ITS1) provided PID information for low-momentum particles by measuring their specific energy loss
- Upgraded ITS for the LHC Run 3 (ITS2)[1]:
  - higher granularity (pixel size reduced from 50 µm × 425 µm to 29.24 µm × 26.88 µm) → better spatial resolution (5 μm)
  - lower material budget (0.36 X_0 %)
  - it can cope with higher interaction rates (50 kHz in Pb—Pb collisions)
  - digital readout → ITS2 cannot measure the energy loss directly

- Cluster size, i.e. the number of pixels turned on by a charged particle traversing a layer of the ITS2, used as a proxy of the particle energy loss
  - cluster size corrected by particle's incident angle on the ITS layer (λ)
  - cluster size distribution for different species is not separated





- Particle distributions normalised per species
  - π and p tracks tagged with the ALICE Time-Projection Chamber (TPC)
- Sensitivity to the different energy loss in the detector material

Particle identification with Machine Learning

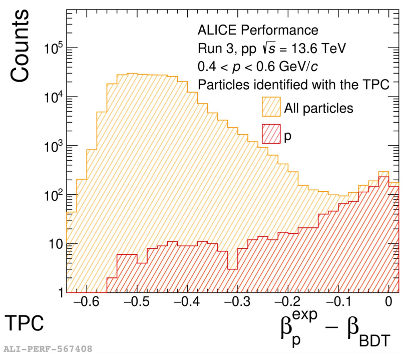
- Separation can be improved with Machine Learning (ML)
- Boosted Decision Tree (BDT) employed for regression
  - inference of the particle $\beta \rightarrow \beta_{BDT}$

$$\beta = \frac{p}{\sqrt{p^2 + m^2}}$$

- Training sample:
  - ITS-TPC matched tracks, particles tagged with the TPC information
  - full data-driven approach
  - training variables: cluster size (average and on different ITS2 layers), momentum (p), sin(φ) (φ is the particle angle on the transverse plane), tan(λ)
- Regressor implemented by XGBoost[1]
  - hyperparameters of the BDT optimised with Optuna[2]

Results and outlook

- Regressor trained and applied on Run 3 data
  - pp collisions at $\sqrt{s}$ = 13.6 TeV
- BDT performance evaluated for protons with 0.4 < p < 0.6 GeV/c through the discrepancy between β_{BDT} and β^{exp}
- Proton distribution peaked around zero
  - **well separated from the rest of the particles**
  - **PID at low momentum feasible with ML**



- ITS PID will be used for:
  - identifying low-momentum particles
  - improving the matching with TPC
  - analysis of (hyper)nuclei decayed or absorbed before reaching the TPC

[1] ALICE Collaboration, *J.Phys. G* 41 (2014) 087002
[2] Chen et al., 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, 2016
[3] Takuya A. et al, KDD (2019)