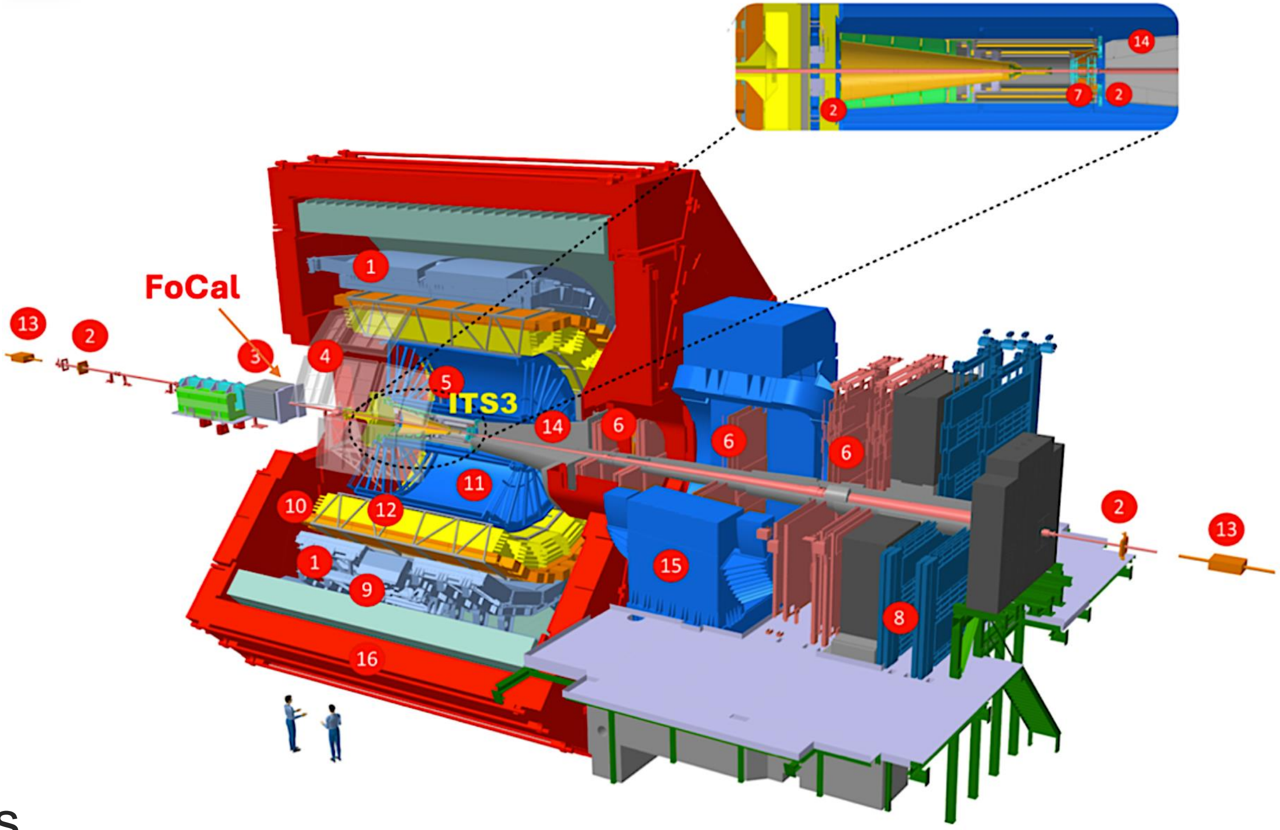
Machine Learning Particle Identification for Run 3 Pb-Pb Collisions in ALICE

Robert Forynski (University of Derby) on behalf of the ALICE Collaboration
IOP Nuclear Physics Conference 13-15 April 2026, Brighton, UK

Introduction to ALICE (A Large Ion Collider Experiment)

- ALICE studies the **quark-gluon plasma** created in high-energy heavy-ion collisions at the LHC.
- Runs 3 and 4: much **higher data rates** and more Pb–Pb collisions **than ever before**.
- Analyses rely on identifying pions, kaons, protons, and electrons over a wide momentum range.

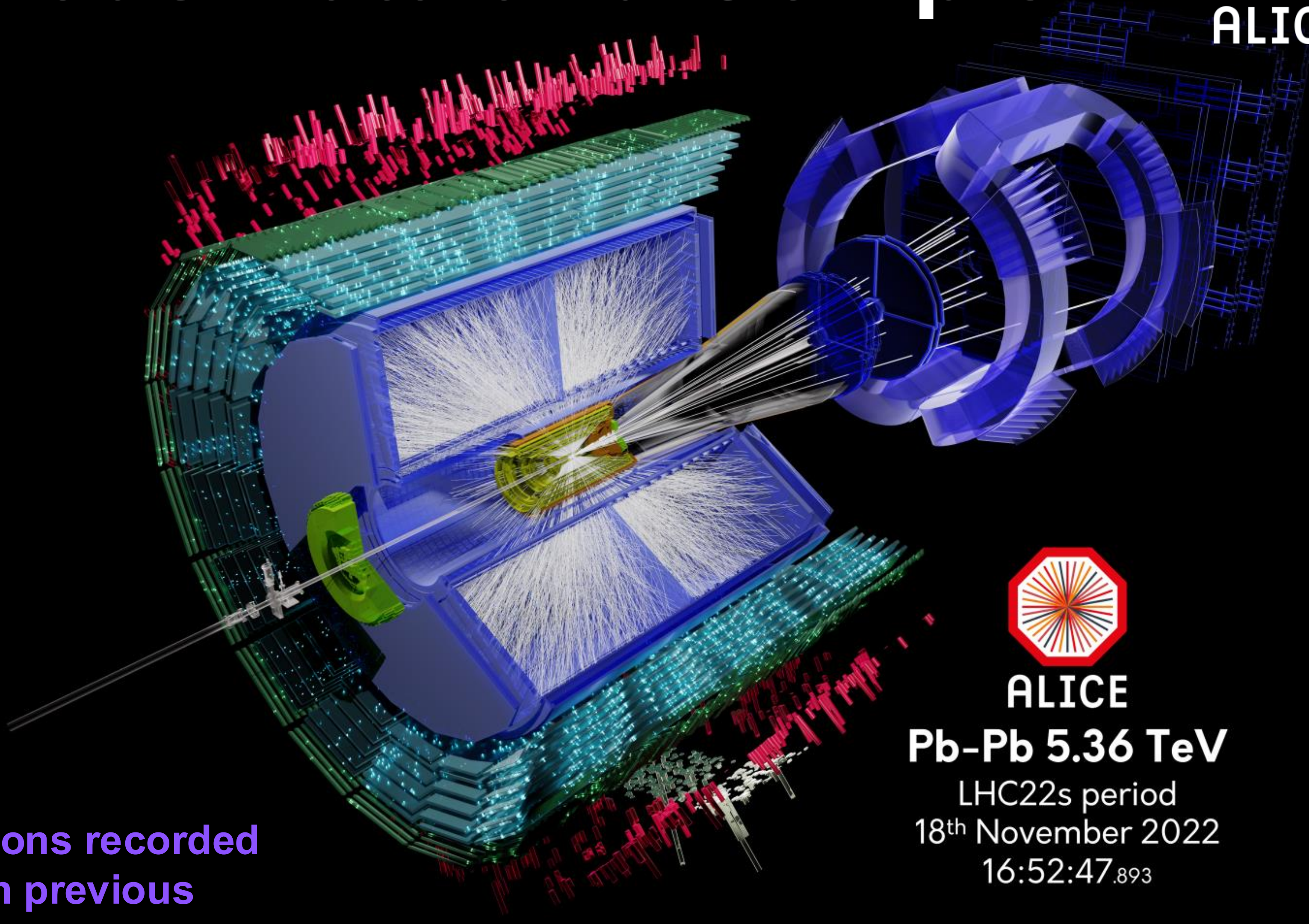


1. **EMCAL |** Electromagnetic Calorimeter
2. **FIT |** Fast Interaction Trigger
3. **FoCal |** Forward Calorimeter (in front of compensator magnet)
4. **HMPID|** High Momentum Particle Identification Detector
5. **ITS|** Inner Tracking System
6. **MCH|** Muon Tracking Chambers
7. **MFT|** Muon Forward Tracker
8. **MID|** Muon Identifier
9. **PHOS/CPV|** Photon Spectrometer
10. **TOF|** Time Of Flight
11. **TPC|** Time Projection Chamber
12. **TRD|** Transition Radiation Detector
13. **ZDC|** Zero Degree Calorimeter
14. **Absorber**
15. **Dipole Magnet**
16. **L3 Magnet**

Run 3: Enormous Data and Complex Events

- Central Pb–Pb collisions produce thousands of tracks per event.
- We need fast, robust methods to tell which tracks are pions, kaons, protons, electrons.

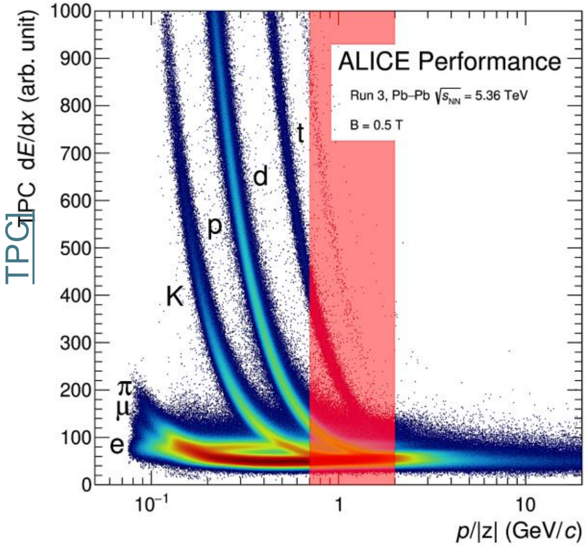
10 billion heavy-ion collisions recorded
10× higher luminosity than previous runs



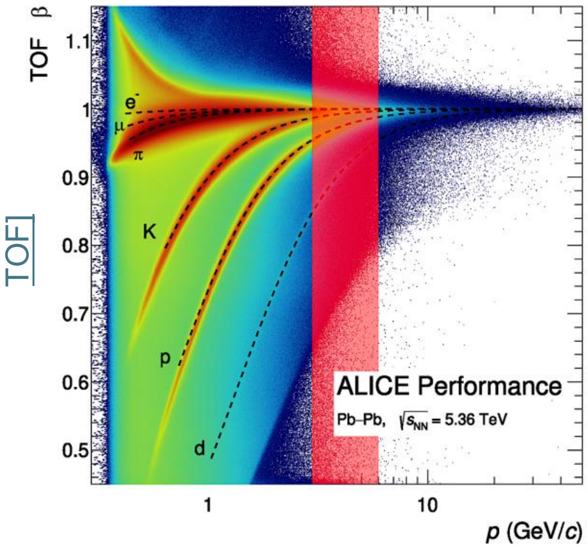
[ALICE event displays and geometry rendering of lead-lead ion collisions during Run 3]

How ALICE identifies particles today

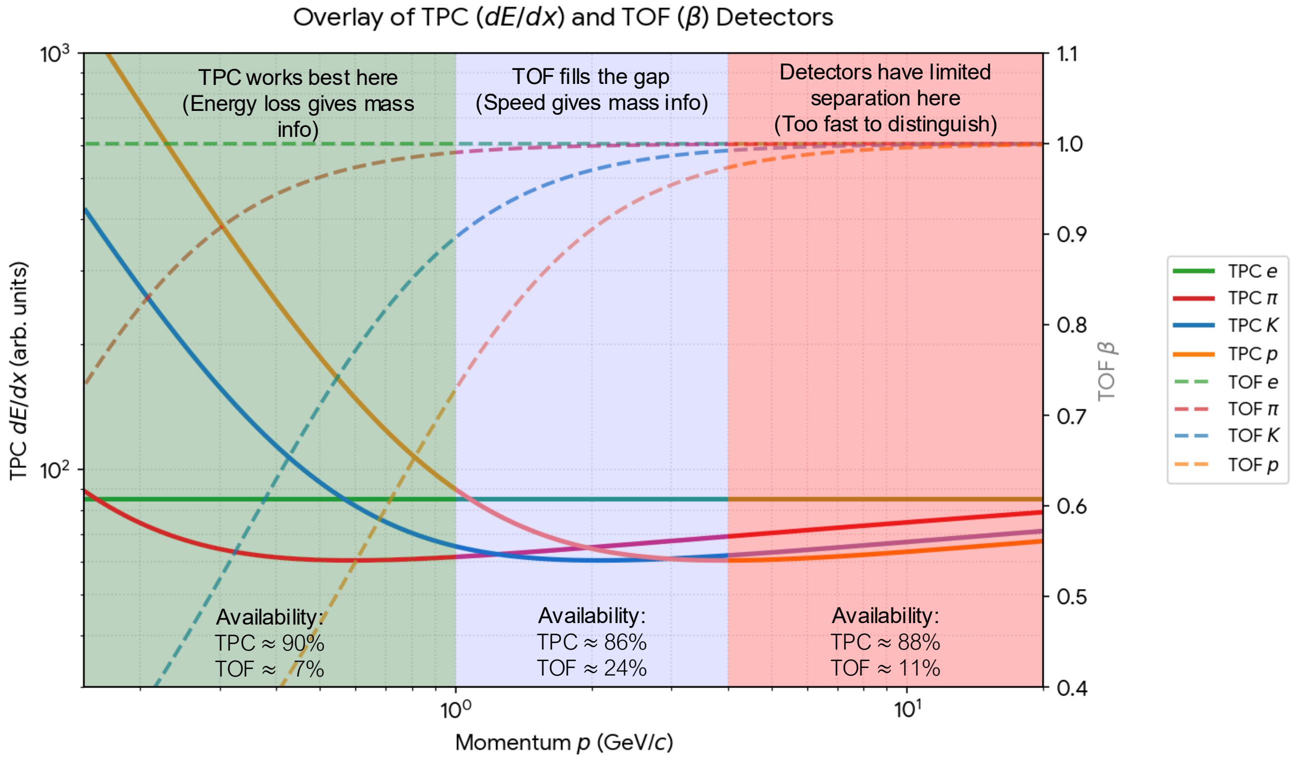
[The particle identification with TPC]



[The particle identification with TOF]



ALI-PERF-577978



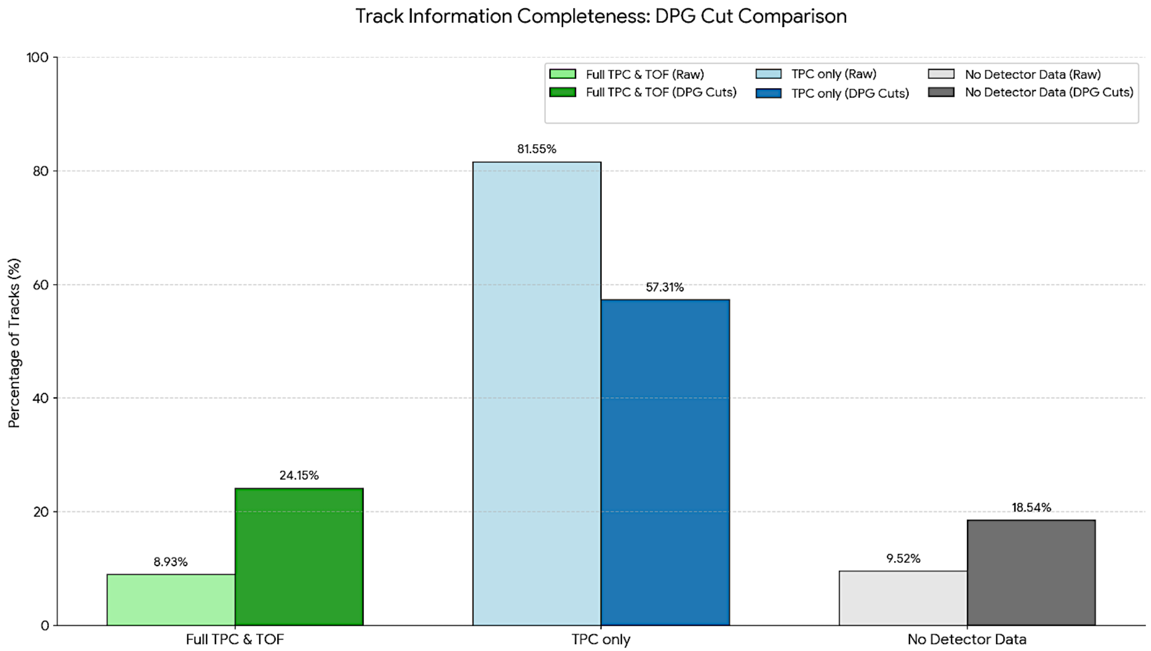
- **TPC measures how much energy a track loses in the gas.** Lighter and heavier particles lose different amounts.
- **TOF measures how long tracks take to reach the outer detector**. Heavier particles arrive slightly later.
- **Combining these gives us good particle identification** where both measurements are available.

Limitations of current methods

- In general, only **about 8%** of tracks have complete information from both TPC and TOF. With the **Data Preparation Group** (DPG) **analysis cuts,** this increases to **~24%.**

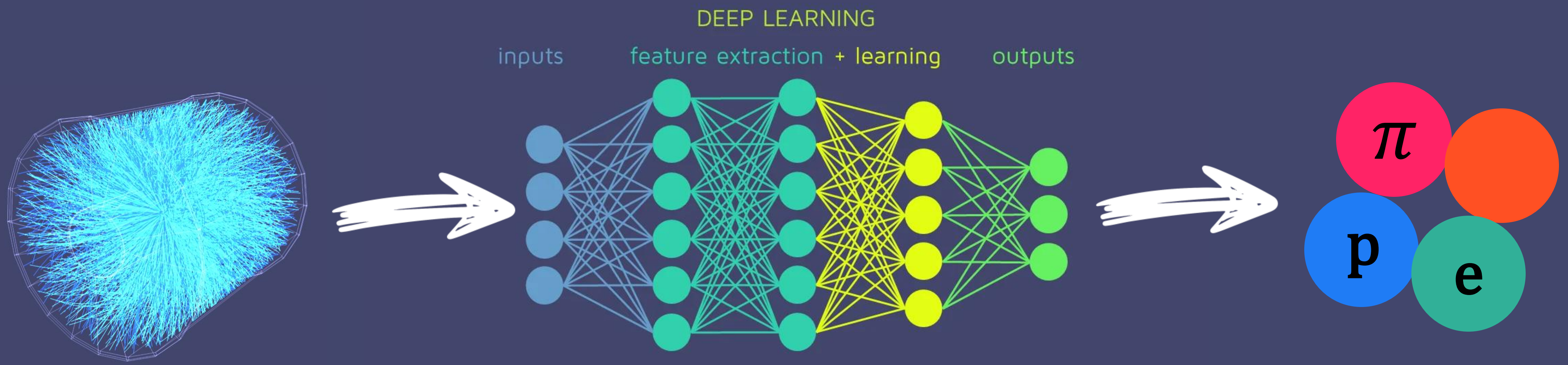
- **Standard methods** (cut-based / Bayesian combination) work very well where everything is measured but **cannot be applied to the majority of tracks.**

- In some momentum regions (around 1 GeV/*c*) separation between particle types is particularly difficult.



Dataset: Monte Carlo-reconstructed 2023 Pb-Pb data

Goal of this work



- Use modern classification techniques to identify more particles over a wide momentum range.
- Maintain high efficiency (capture the maximum number of true tracks) and high purity (to mitigate misidentification background).
- Do this in a way that can cope with the large Run 3 datasets.

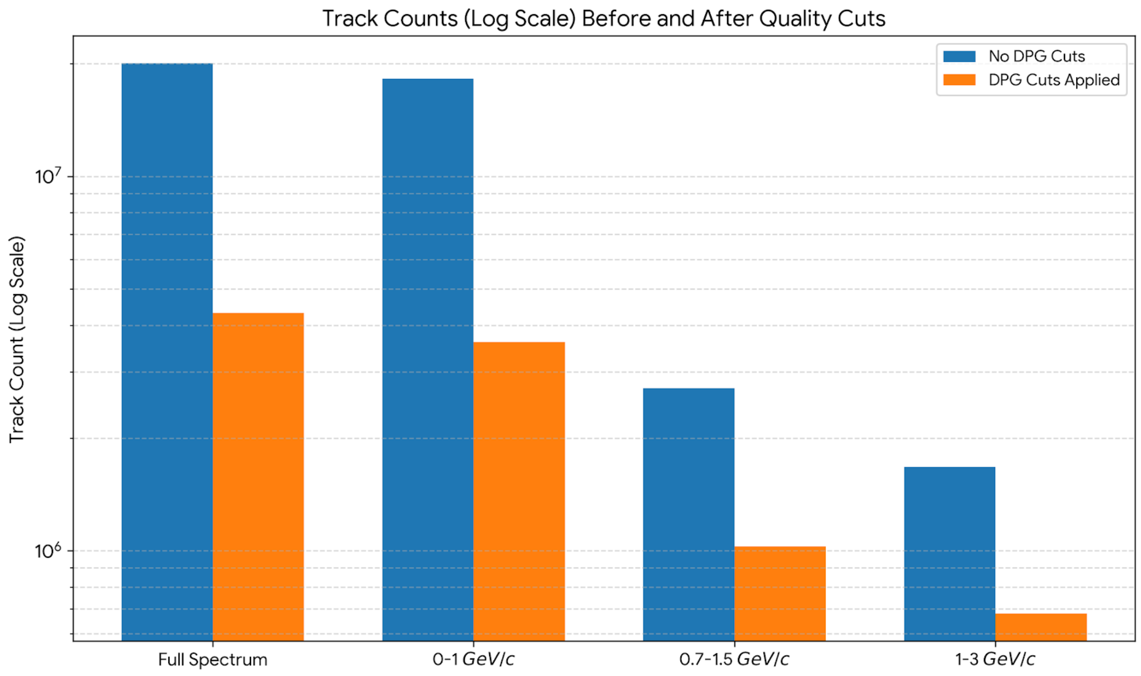
Dataset and selections

- **Simulated Run 3 Pb–Pb** collisions with detector effects, about **20 million tracks**.

- **Four species**: pion, kaon, proton, electron.

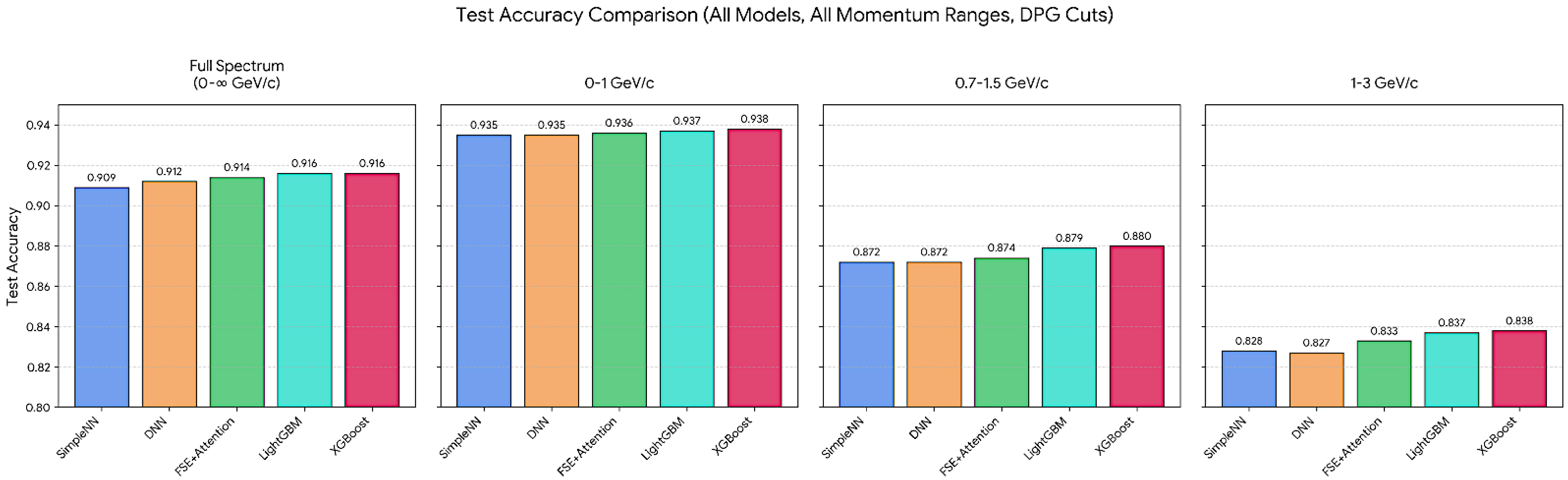
- We look at **different momentum ranges**: 0-∞, 0–1, 0.7–1.5, 1–3 GeV/c.

- Standard **ALICE track selections applied** (see backup slide).



Dataset: Monte Carlo-reconstructed 2023 Pb-Pb data

Classification methods

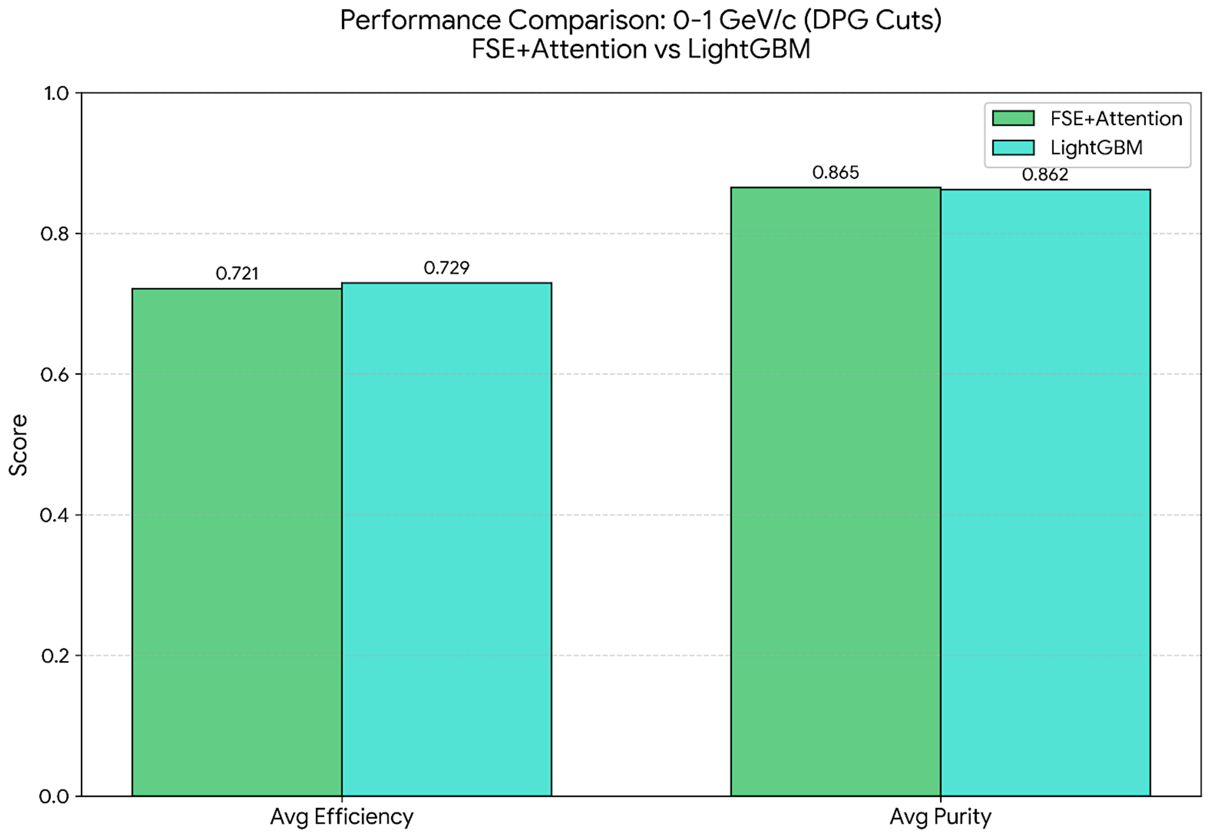


- We **train several types of classifiers**: deep neural networks and boosted decision trees.
- They take **as input** the track **kinematics**, **TPC** energy loss, **TOF** time, and whenever available, the **standard** ALICE **probability** estimates. [Particle identification in ALICE: a Bayesian approach]
- All methods are trained on the same simulated data and tested on independent samples.

Main results: efficiency & purity

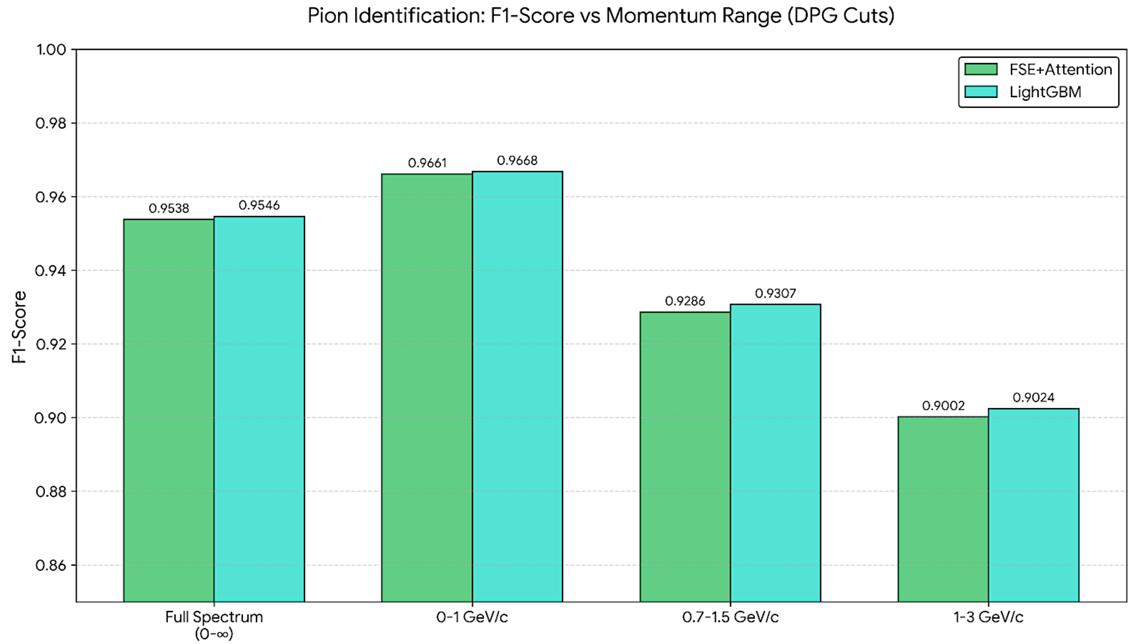
- For **pions and protons**, we reach around **90% efficiency and 90% purity** in the 0–1 GeV/*c* range.
- **Kaons** and **electrons** are more **challenging**, but we still achieve useful separation.

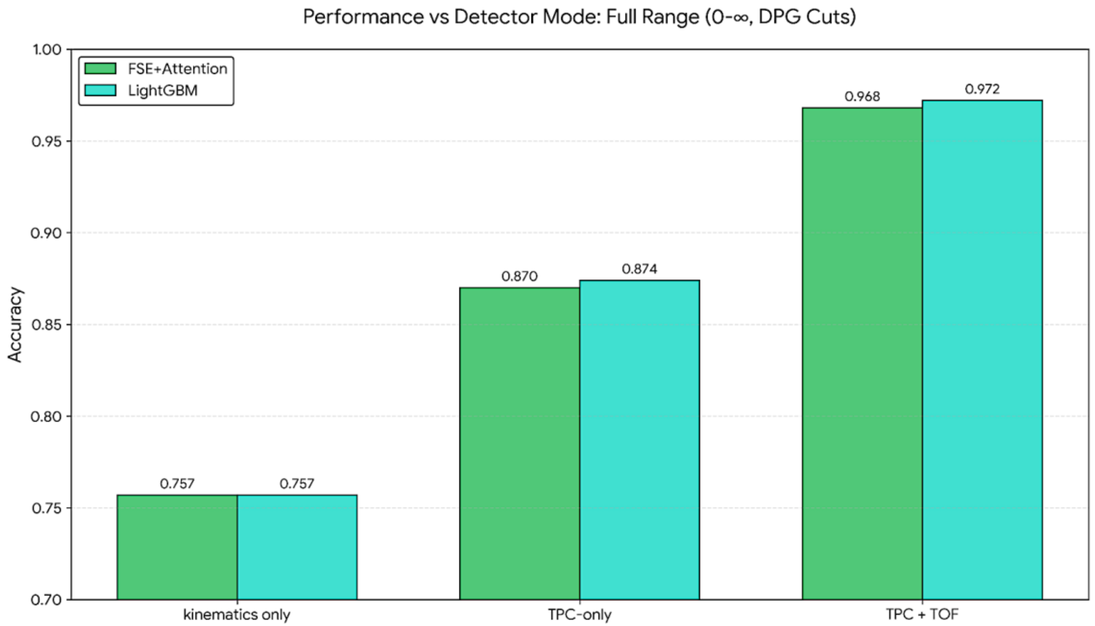
FSE+Attention				LightGBM			
Particle	Efficiency	Purity	F1-Score	Particle	Efficiency	Purity	F1-Score
Pion	0.9911	0.9423	0.9661	Pion	0.9914	0.9434	0.9668
Kaon	0.4824	0.8360	0.6118	Kaon	0.4825	0.8538	0.6166
Proton	0.7345	0.9535	0.8298	Proton	0.7350	0.9571	0.8315
Electron	0.3561	0.7228	0.4771	Electron	0.4043	0.7211	0.5182



Dataset: Monte Carlo-reconstructed 2023 Pb-Pb data

Dependence on momentum and detectors



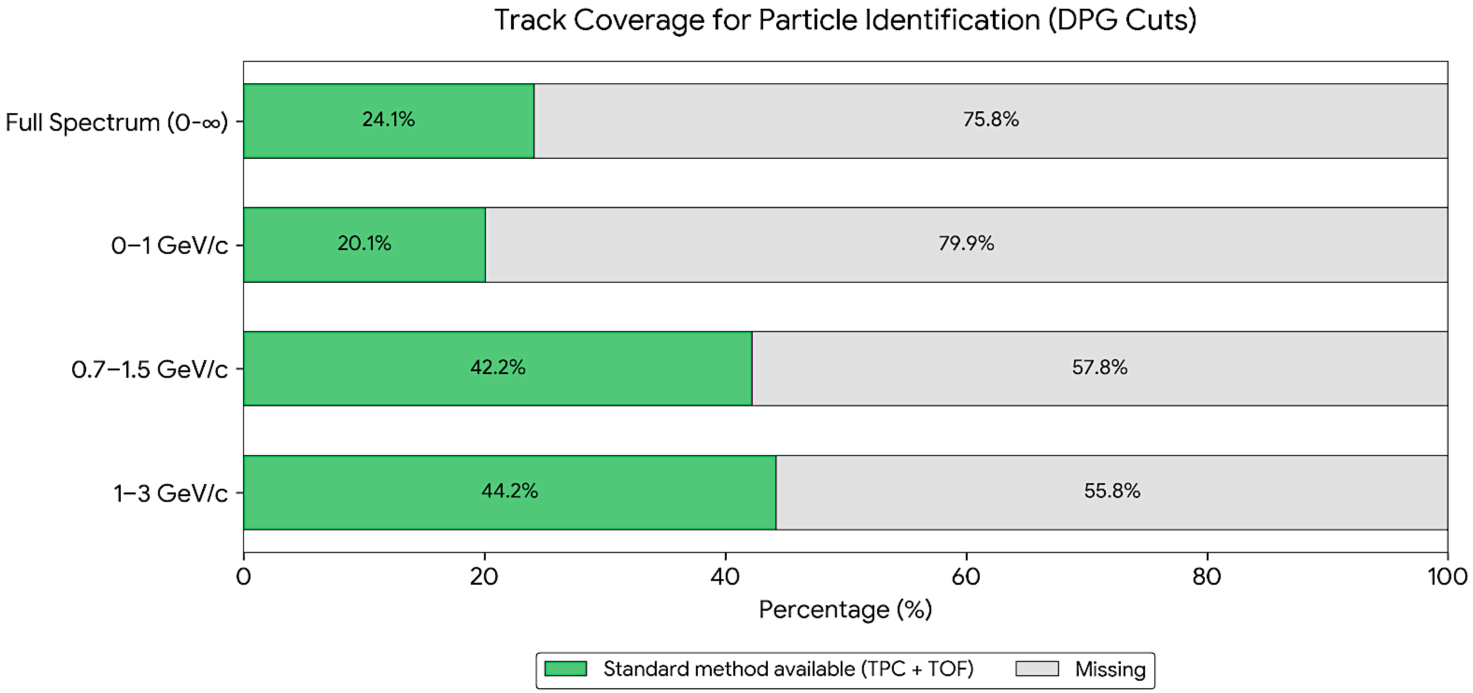


- **Performance** is best in 0–1 GeV/*c*, and **decreases at higher momentum**, where separation in TPC and TOF is harder.

- Having both **TPC and TOF** information is **crucial**.
- **Quality cuts** also improve performance.
- The **classifier** choice impacts results by a few %.
- The detector information is the main driver.

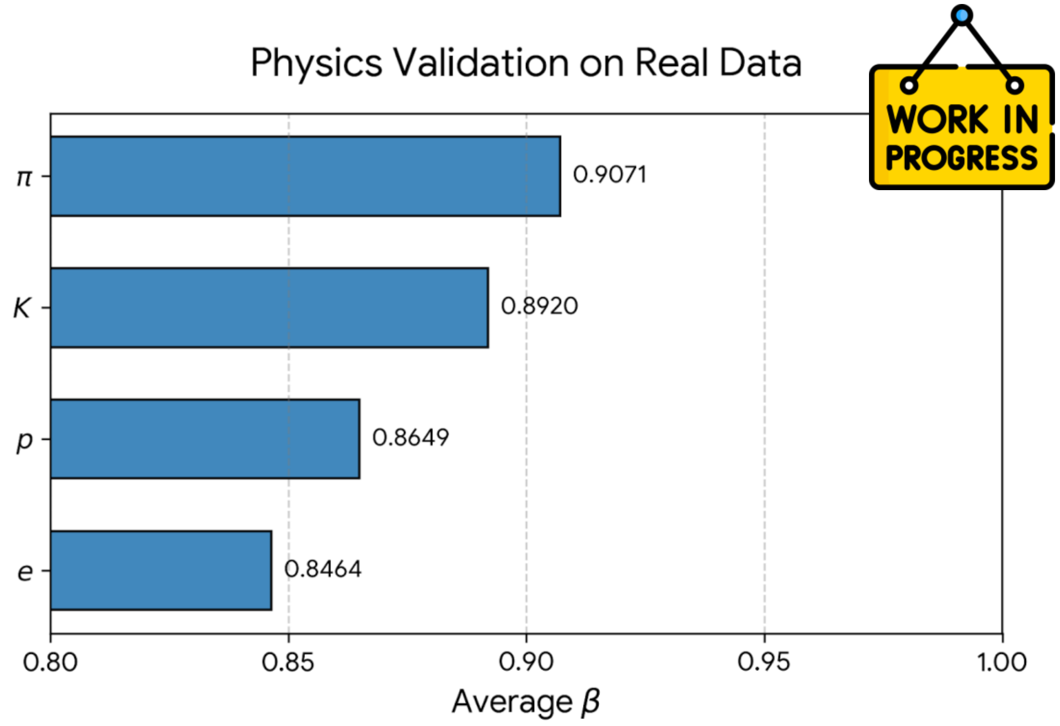
Relation to current ALICE method

- **Where** the **standard** ALICE probability-based **method is defined** and all detectors are present, our **classifier gives similar or better** identification **performance** for the same tracks.
- More importantly, it **can provide particle probabilities** for the majority of tracks **where the standard method is not defined**, extending the useful phase-space.



First checks on real data

- We applied one of the trained classifiers to about **8 million real Run 3 Pb–Pb tracks**.
- The predicted species fractions and the ordering of **TOF times look physically reasonable**.
- We see some **tension in the TPC energy-loss distributions**, which likely reflects differences between simulation and current calibrations; this is **under investigation**.



Predicted species show the expected TOF ordering in real data.

Dataset: Raw 2023 Pb-Pb data

Conclusions

- **Machine-learning classifiers** can provide **high-quality particle identification** across 0–3 GeV/*c* in Pb–Pb simulations.

- Using both TPC and TOF information and applying standard track-quality cuts are the main components. The choice of classifier gives additional but smaller gains.

- **First tests on real data are encouraging** but highlight the need for careful detector calibration and further validation.

- **Future work** includes improving kaon and electron performance, incorporating additional detectors, and integrating these methods into ALICE analysis workflows.

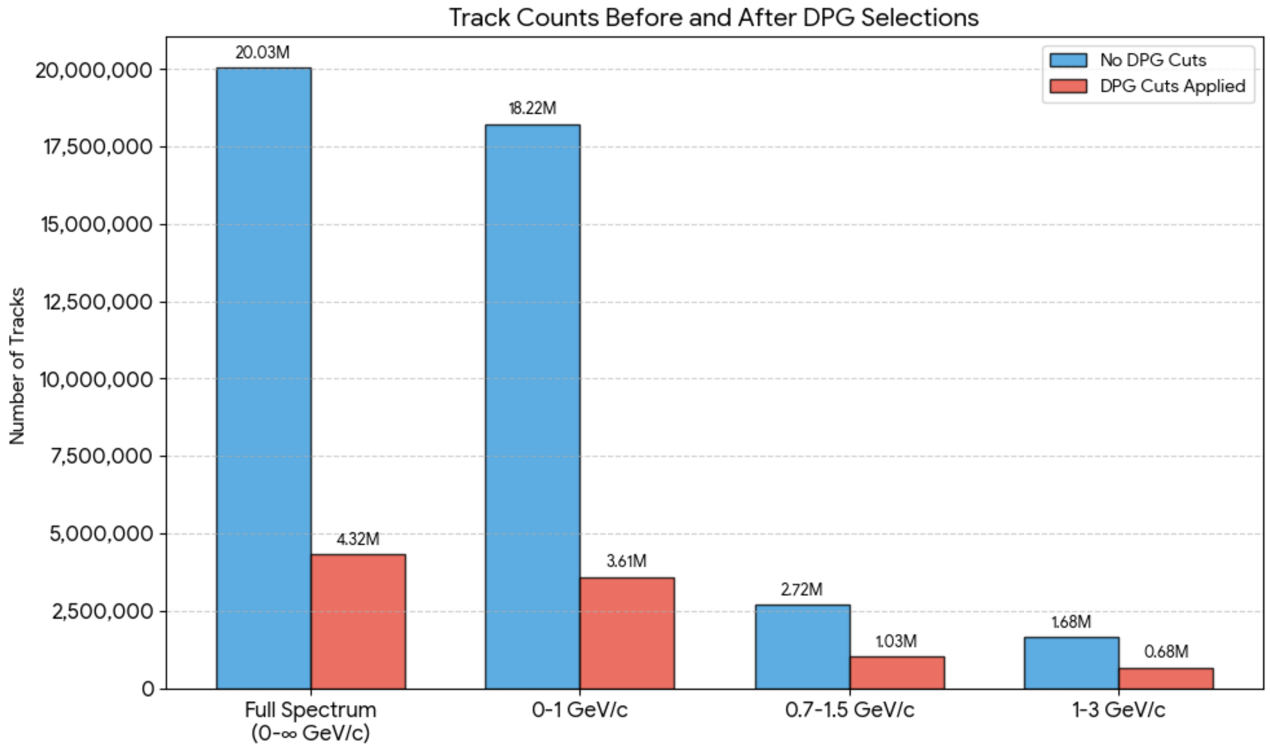
Thank you for your attention

Robert Forynski
robert.forynski@cern.ch

Backup

Standard ALICE track selections

- $|\eta| < 0.8$
- DCA_xy < 0.105 cm, DCA_z < 0.12 cm
- TPC clusters ≥ 70
- $p_T > 0.1$ GeV/c
- **DPG recommended cuts applied for main results.**



Classifier architectures

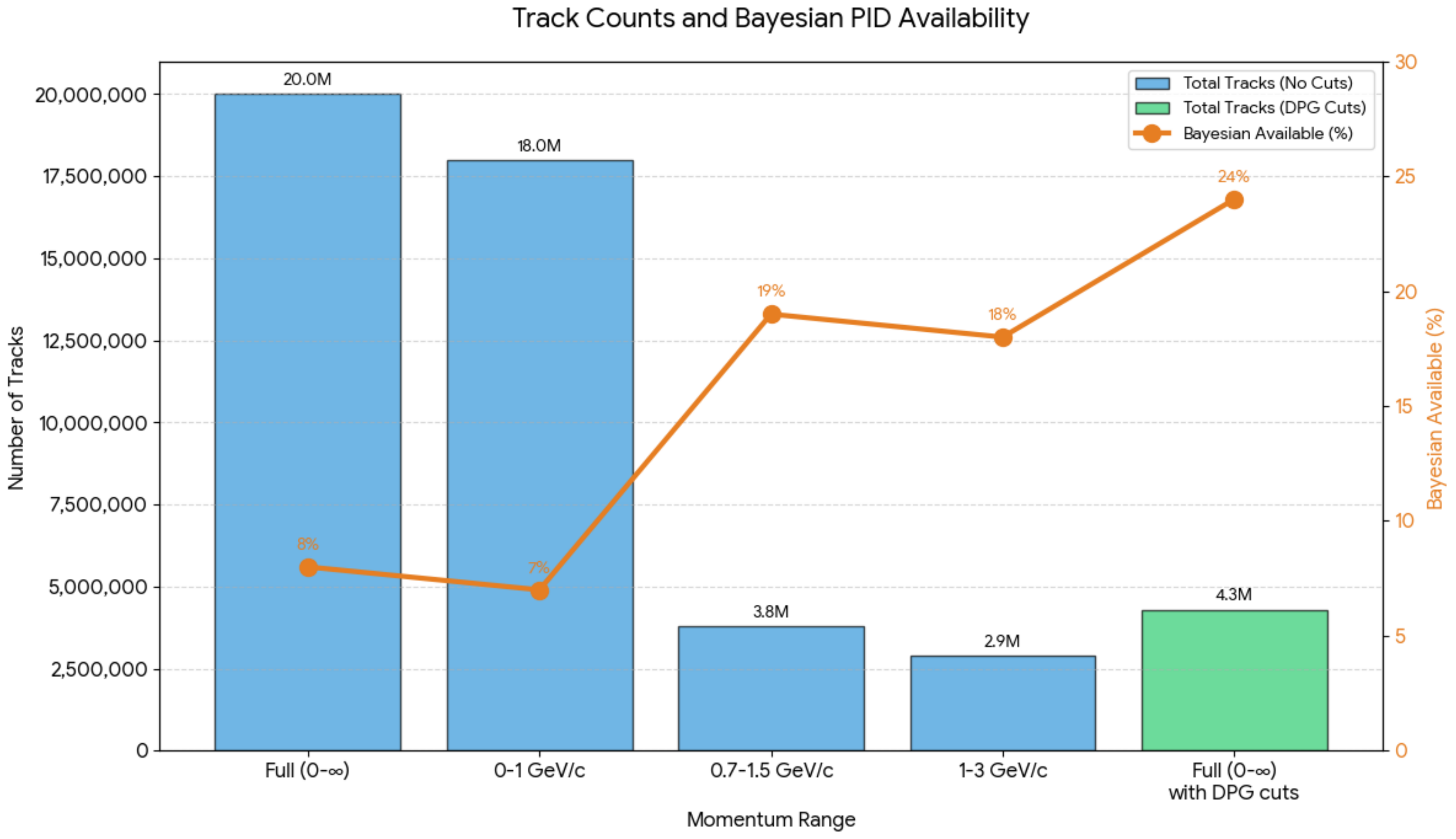
- **SimpleNN and DNN** are standard feed-forward neural networks.

- **FSE + Attention** is a feature-based neural network that learns which detector inputs are most useful.

- **LightGBM and XGBoost** are boosted decision-tree methods, included as strong non-neural baselines

Bayesian particle identification (PID) availability

- **Machine-learning accuracy 87%-96%** vs nσ Bayesian traditional method accuracy ~65–85% (where Bayesian is defined).
- **Extends to 75–85% on remaining tracks** where Bayesian traditional method is not defined.



Detector availability across momentum ranges

- DPG cuts increase full TPC+TOF availability, especially at higher momentum.
- TPC remains available for most tracks. TOF coverage is the main limiting factor.