Machine learning applications for particle identification in MPD

A. Ayriyan, H. Grigorian, A. Mudrokh, V. Papoyan

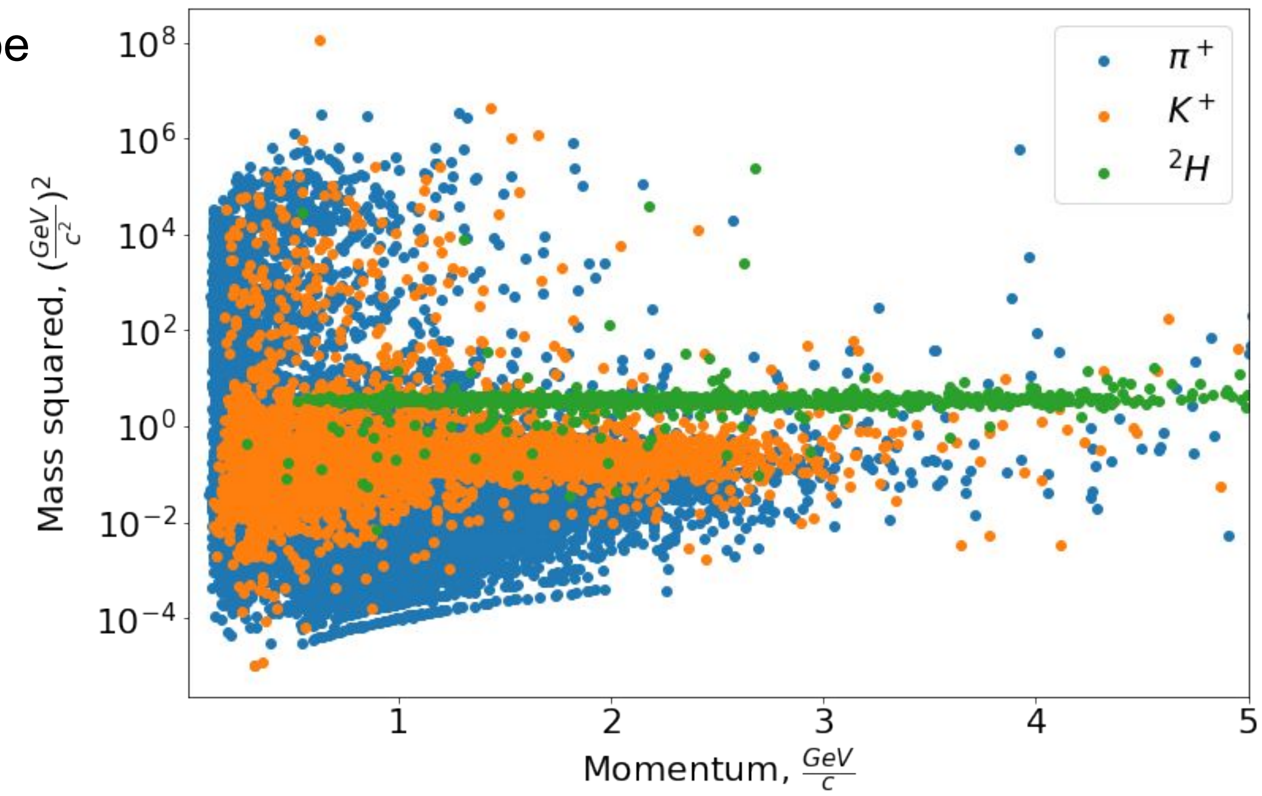
12.07.2022

Particle Identification

Particle IDentification (**PID**) is the task of identifying the particle type associated with a given track.

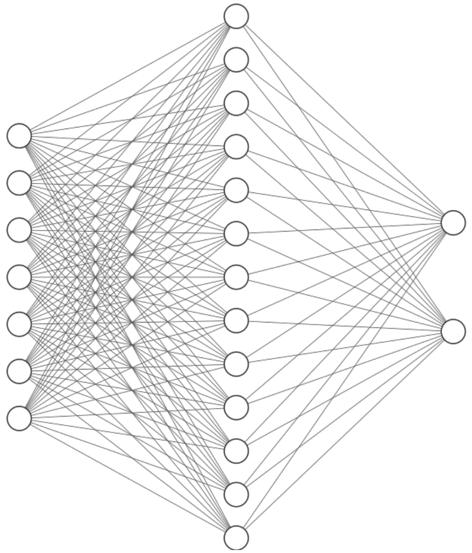
In Machine Learning terms, PID can be considered as:

1. <u>multiclass classification problem</u>;
2. binary classification problem
   a. one-vs-rest;
   b. one-vs-one.



Machine Learning in PID

Present time ML methods for PID are widely used.

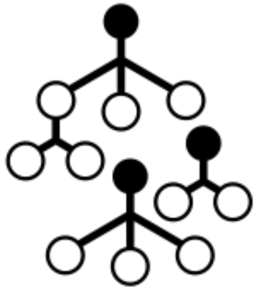


ProbNN (Shallow Neural Networks):

one-particle-vs-rest strategy; One shallow neural network for the each particle type

DNN (Deep Neural Network):

multiclass strategy; Deep NN with three hidden layers



XGboost & **CatBoost** (Boosted Decision Trees):

multiclass strategy; CatBoost uses **oblivious** trees (robust to noise) [1]

The preliminary results were obtained by application of **Decision tree**.

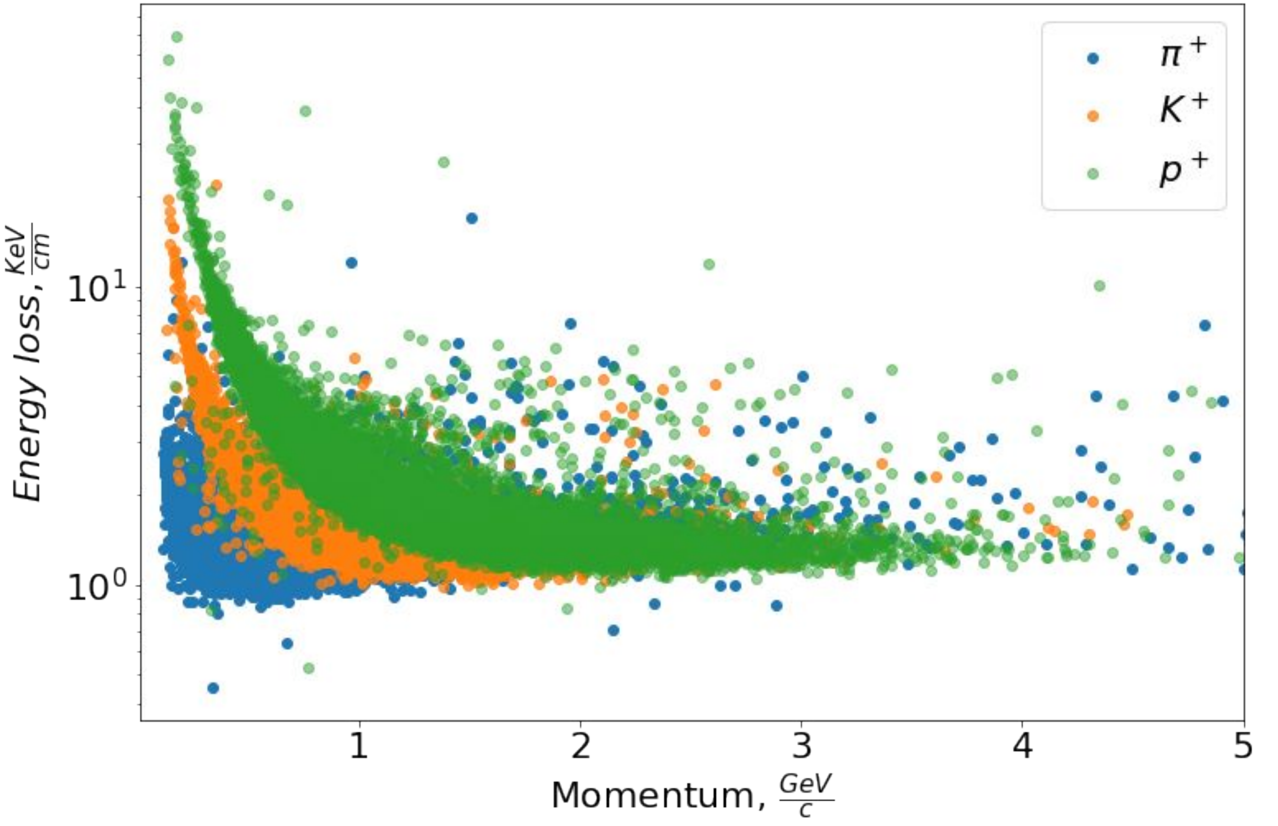
[1] Derkach D. et al. Machine-Learning-based global particle-identification algorithms at the LHCb experiment //Journal of Physics: Conference Series. – IOP Publishing, 2018. – T. 1085. – №. 4. – C. 042038.

Data set

The **Decision tree** model is trained on Monte-Carlo data (24M tracks in total).

PHQMD; minimum bias

Bi+Bi @ 9.2 GeV



Feature vector:

- momentum;
- charge;
- energy loss;
- mass squared;
- number of hits in TPC;
- pseudorapidity;
- dca.

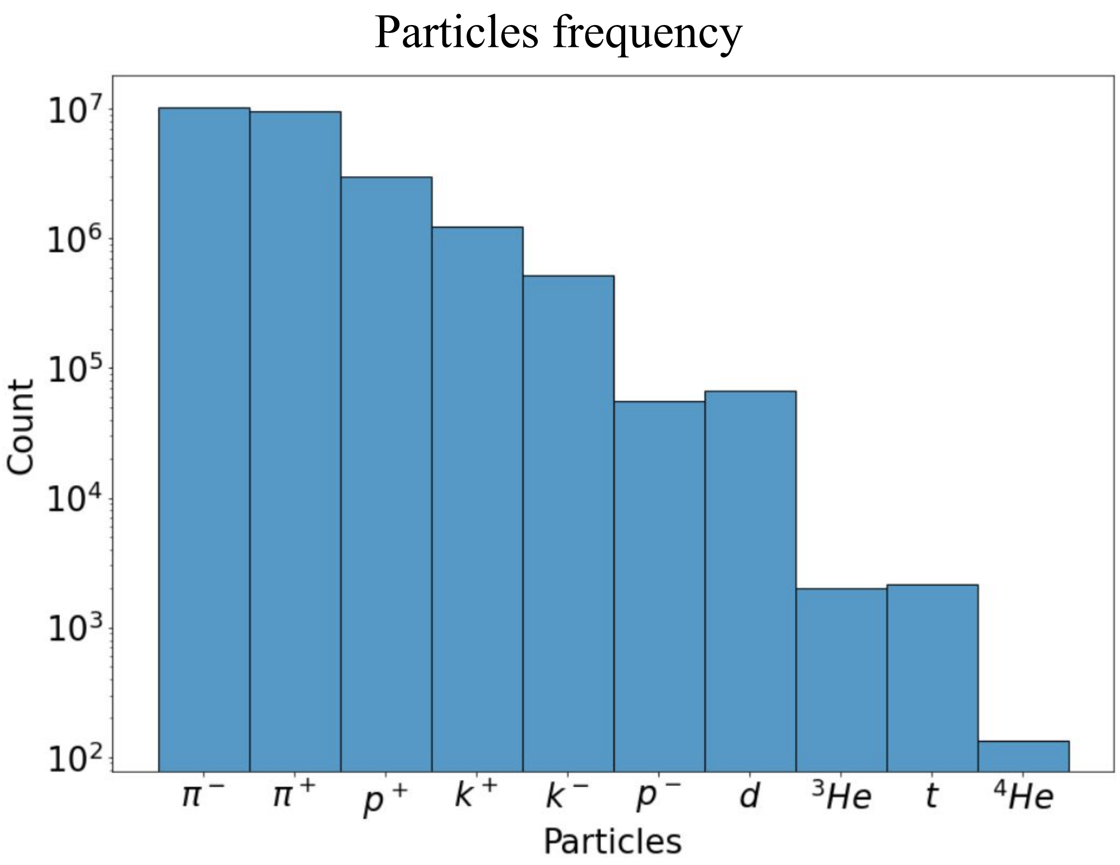
Train and Test Samples

Train sample: random 70% Monte-Carlo tracks.

Test sample: remaining 30% tracks.

Classes are **imbalanced** - not having enough tracks for the minority classes (^{4}He, t, ^{3}He).

PID efficiency reduction for minority classes. Balanced data are better for training.



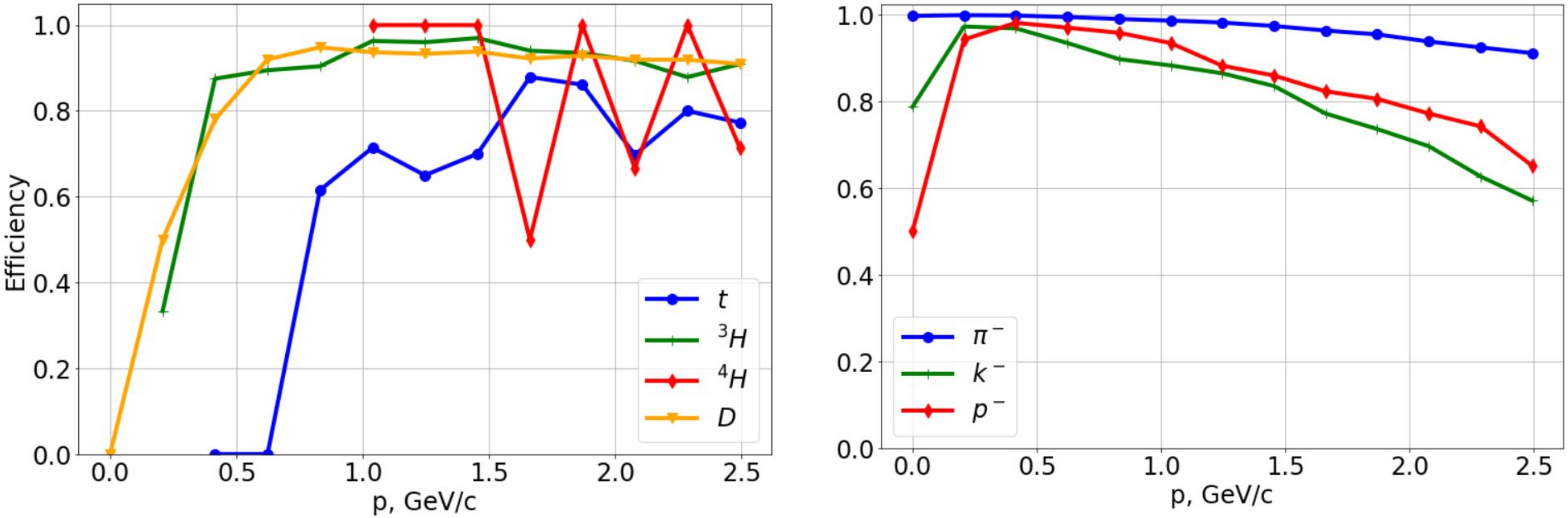
The preliminary results

$$Efficiency = \frac{right\ identified\ tracks}{all\ tracks}$$

Decision tree parameters:

- **criterion** : gini;
- **depth** : 7.

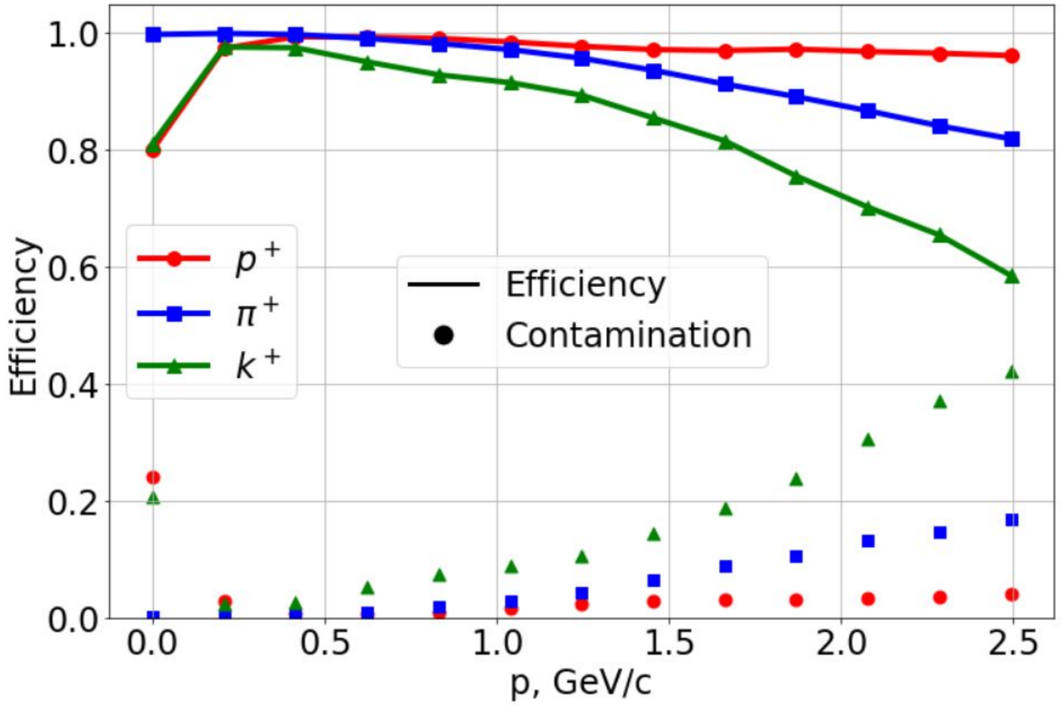
Combined PID efficiency:



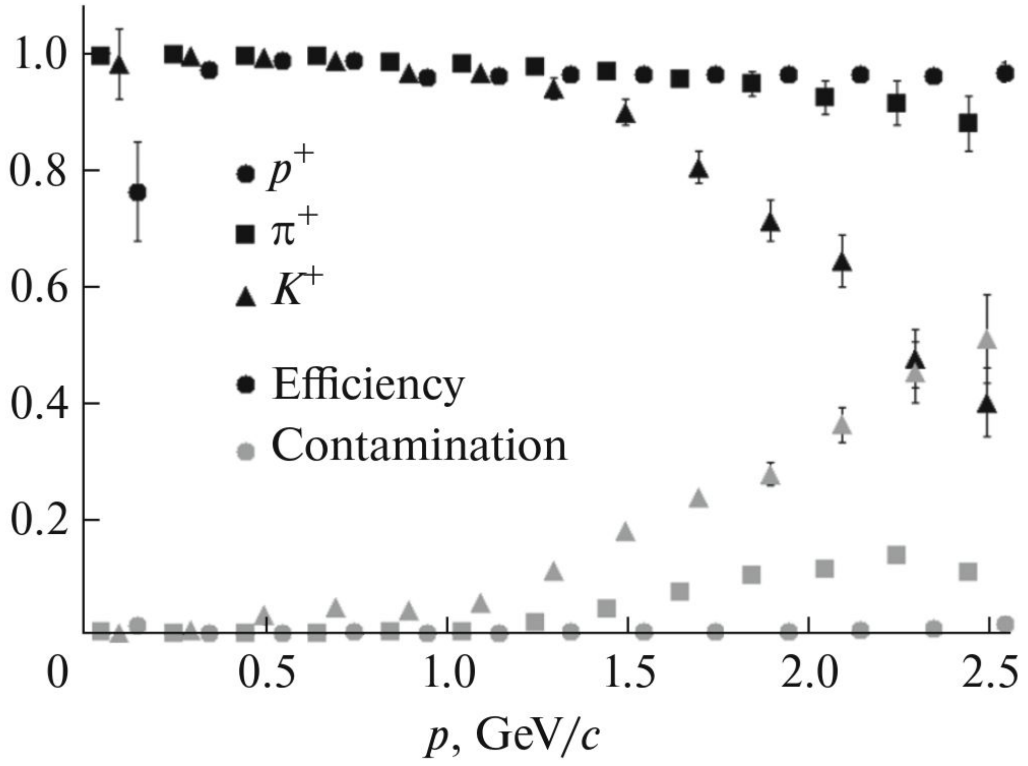
The preliminary results

$$Contamination = \frac{wrong\ identified\ tracks}{identified\ tracks}$$

Combined PID efficiency (Decision Tree):



Combined PID efficiency (MPD) [1]:



[1] Kolesnikov V. et al. Towards a realistic Monte Carlo simulation of the MPD detector at NICA //Physics of Particles and Nuclei Letters. – 2019. – T. 16. – №. 1. – C. 6-15.

Conclusion and Outlook

1. Application of simple Decision Tree approach allowed **to reproduce** the properties of the PID MPD results. For some of particles the efficiency becomes **even better**.
2. A new **balanced training data set** will be generated for all particle classes and all momentum range. Such dataset is expected to **increase** the PID efficiency.
3. Decision Tree approach will be **replaced** to Boosting Decision Tree and Random Forest algorithms.