



# Particle-identification techniques and performance at LHCb in Run 2

M. Hushchyn\*, V. Chekalina, on behalf of the LHCb collaboration

National Research University Higher School of Economics, Moscow, Russia  
Yandex School of Data Analysis, Moscow, Russia

## ARTICLE INFO

### Keywords:

Photo detectors and PID

## ABSTRACT

One of the most challenging data analysis tasks of modern High Energy Physics experiments is the identification of particles. In this proceedings we review the new approaches used for particle identification at the LHCb experiment. Machine-Learning based techniques are used to identify the species of charged and neutral particles using several observables obtained by the LHCb sub-detectors. We show the performances of various solutions based on Neural Network and Boosted Decision Tree models.

## Contents

1. Introduction .....	568
2. Charged particles .....	568
3. Neutral particle.....	569
4. Conclusion .....	569
Acknowledgement.....	569
References.....	569

## 1. Introduction

Particle identification (PID) plays a crucial role in LHCb analyses. The LHCb PID system is composed of Cherenkov detectors, muon chambers and a calorimeter system. Combining information from these subdetectors allows to recognize long-lived charged and neutral particles [1]. Advanced multivariate techniques can be used to improve the PID performance and control systematic uncertainties in a data-driven way. New models were developed using environment of the LHCb Run 2. These models show in general good performance and in addition a better efficiency flatness.

## 2. Charged particles

There are five types of charged particles that leave tracks at the LHCb detector, namely, electron, muon, pion, kaon, proton. This set of real tracks is complemented by ghost tracks, which do not correspond to a real particle passed through the detector. The PID algorithm has to provide one of these six possible outcomes thus forming a multi-class problem. The current solution was developed during Run 1 of data taking and consists in a set of 5 neural networks with one hidden layer trained to distinguish one particle species from the rest [1].

To improve the baseline PID solution performance, several new classification algorithms are implemented. Deep NN classifier combines subdetectors responses with deep neural network trained using Keras library [2]. The gradient boosting solution trained using CatBoost library [3] has good performance for some of the signal-background combinations (see Fig. 1). Understanding if the new approach introduces systematic biases is the goal of an ongoing study. In addition, Flat 4d classifier is a boosted decision trees classifier with flat dependency of signal efficiency on particle kinematics (total momentum, transverse momentum and pseudo-rapidity) and track multiplicity observables [4].

The classifiers are trained using a Monte Carlo (MC) sample containing all charged particle types and ghost track candidates. Additional MC samples are used to estimate the classifier performance on real data [1]. The classifier is also tested on data from the 2016 data taking and contain particles that can be identified purely based on only kinematic properties. The samples used in this work were reconstructed in several decay modes:  $J/\Psi \rightarrow \mu^+\mu^-$ ,  $B^+ \rightarrow J/\Psi(e^+e^-)K^+$ ,  $D^{*+} \rightarrow D^0(K^-\pi^+)\pi^+$ ,  $\Lambda^0 \rightarrow p\pi^-$ .

The PID performance of each classifier is shown in Fig. 1.

The PID response strongly depends on the kinematic of the charged particle. This correlation leads to a strong dependency of the PID efficiency on the kinematic variables (Fig. 2). Relative to the baseline model, the Flat 4d model has a flatter PID efficiency, which can help to

\* Corresponding author.

E-mail address: [mikhail.hushchyn@cern.ch](mailto:mikhail.hushchyn@cern.ch) (M. Hushchyn).

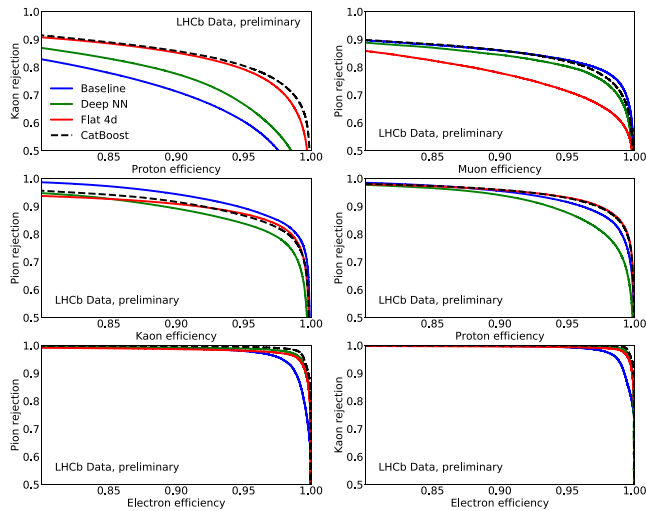


Fig. 1. Dependences between background rejection and signal efficiency for six particle pairs.

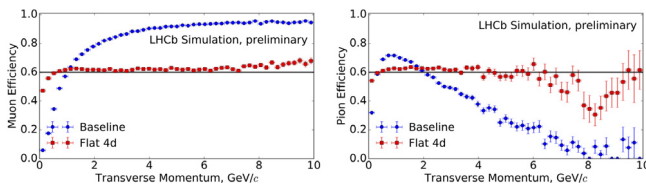


Fig. 2. Dependence between Flat 4d model efficiencies and particle transverse momentum for each particle type. The curves correspond to the same global signal efficiency of 60%.

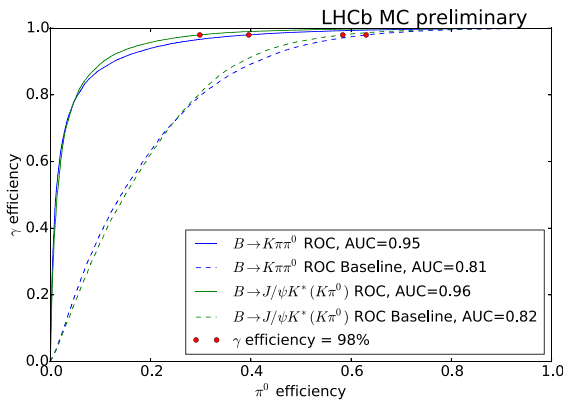


Fig. 3. Photon efficiency vs mis-identification curves for the baseline (dashed line) and the new XGBoost model (solid line). Different colours refer to different test samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reduce the systematic uncertainties in the physics analysis. The classifier achieves flat dependency using a modified loss function [4].

### 3. Neutral particle

The identification of neutral particles is based on the response of the calorimeter system. One important classification problem is to separate photons produced promptly in the proton–proton collision from photons produced in  $\pi^0$  decays. The most challenging case is when two photons from the  $\pi^0$  decay are highly collinear, thus producing a calorimeter cluster that looks similar to a cluster produced by a prompt photon.

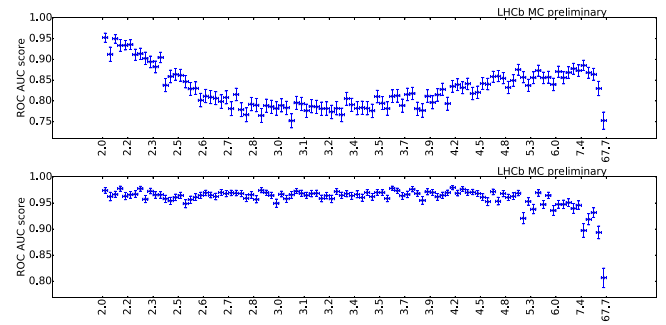


Fig. 4. Baseline (top) and XGBoost (bottom) models quality as a function of transverse energy.

The baseline model used to separate these responses is based on a 2-layer neural network from the TMVA library [5]. The neural network inputs are several aggregated features describing the cluster shape in a  $3 \times 3$  electromagnetic calorimeter (ECAL) cells area around the cluster seed. The feature set includes the variables depended on the energy distribution in the cluster and the second moments of the cluster shape  $S_{xx}$ ,  $S_{yy}$ ,  $S_{xy}$ . To improve the efficiency we use a XGBoost model which is a gradient boosting over decision trees classifier. In this case, inputs are raw energy values in  $5 \times 5$  ECAL and Preshower Detector (PS) cells around the cell seed.

The model is trained using MC samples. Photons and  $\pi^0$  are obtained from  $B^0 \rightarrow K\pi\gamma$  and  $B^0 \rightarrow K\pi\pi^0$  decays respectively. The decay channel  $B^0 \rightarrow J/\psi K^*$  with  $K^* \rightarrow K\pi^0$  is used as an extra source of  $\pi^0$  to check the models stability. The performance evaluated on the simulated events shows promising results. The obtained quality of  $\gamma$ - $\pi^0$  separation is demonstrated in Fig. 3.

The uniform dependence of the quality of the model from particles properties is very important for optimal performance. Fig. 4 demonstrates that the new approach provides flat dependency of quality from the transverse energy  $E_T$ .

### 4. Conclusion

Advanced machine learning techniques allow to increase particle identification performance both for charged and neutral particles. Combining information from the LHCb tracking system, ring-imaging Cherenkov detectors, electromagnetic and hadron calorimeters, and muon chambers allows to achieve high background rejection for charged particle identification. Using machine learning approaches to analyse the raw energy deposits in the calorimeter provides good prospects for high quality neutral particle identification.

### Acknowledgement

The research leading to these results has received funding from Russian Science Foundation under grant agreement no. 17-72-20127.

### References

- [1] LHCb collaboration, LHCb detector performance, *Internat. J. Modern Phys. B* **30** (2015) 1530022.
- [2] F. Chollet, Keras, 2015. <https://github.com/fchollet/keras>.
- [3] A.-V. Dorogush, A. Gulin, G. Gusev, N. Kazeev, L. Ostroumova Prokhorenkova, A. Vorobei, Fighting biases with dynamic boosting, 2017. [arXiv:1706.09516](https://arxiv.org/abs/1706.09516).
- [4] A. Rogozhnikov, A. Bukva, V. Gligorov, A. Ustyuzhanin, M. WilliamsNew, New approaches for boosting to uniformity, *J. Instrum.* **10** (2015) 03002.
- [5] M. Calvo, E. Cogneras, O. Deschamps, M. Hoballah, A tool for  $\gamma$   $\pi^0$  separation at high energies LHCb-PUB-2015-016, 2015.