Boosting Neutral Particles Identification by Boosting Trees: LHCb case

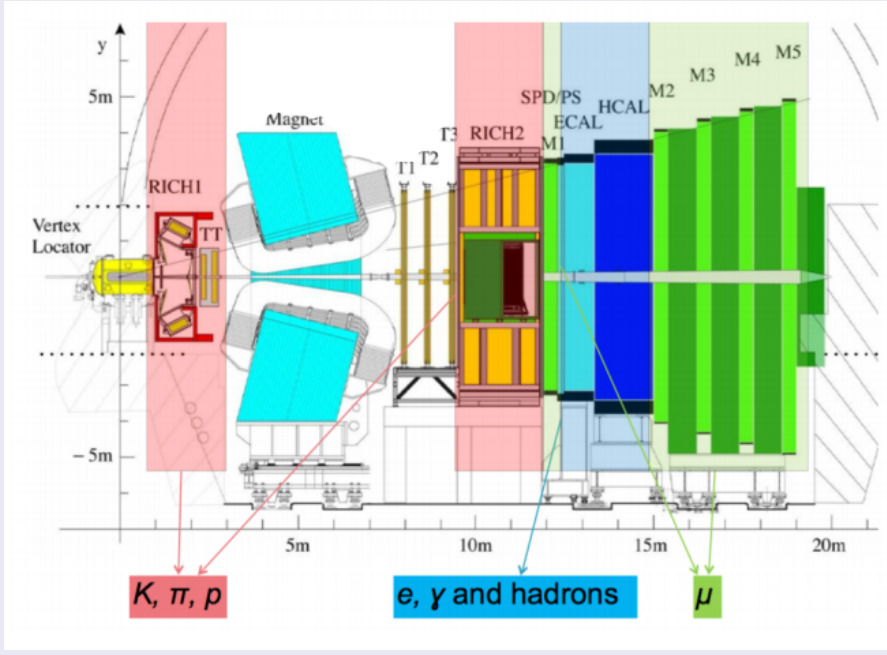
Chekalina Viktoria on behalf of the LHCb Collaboration

NRU Higher School of Economics, Yandex Data School

11 July 2018

LHCb PID sub-detectors

- Ring Imaging Cherenkov detectors (RICH)
- Tracking system
- Hadronic and Electromagnetic calorimeters (HCAL and ECAL)
- Muon chambers

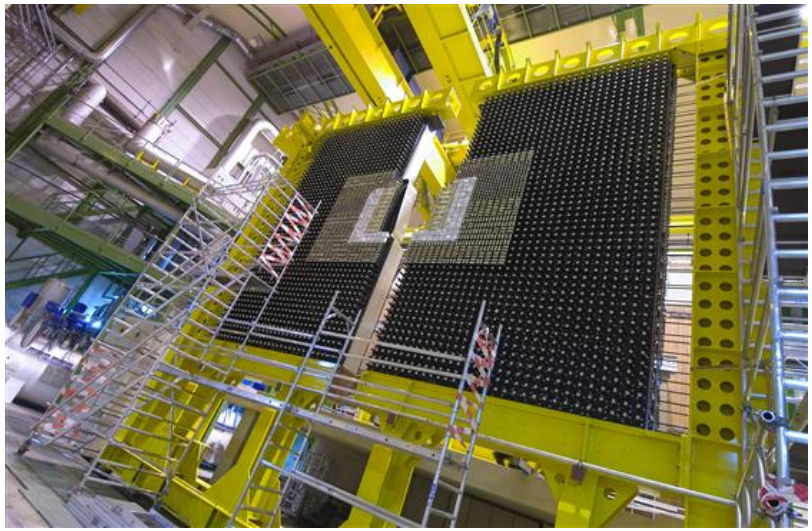


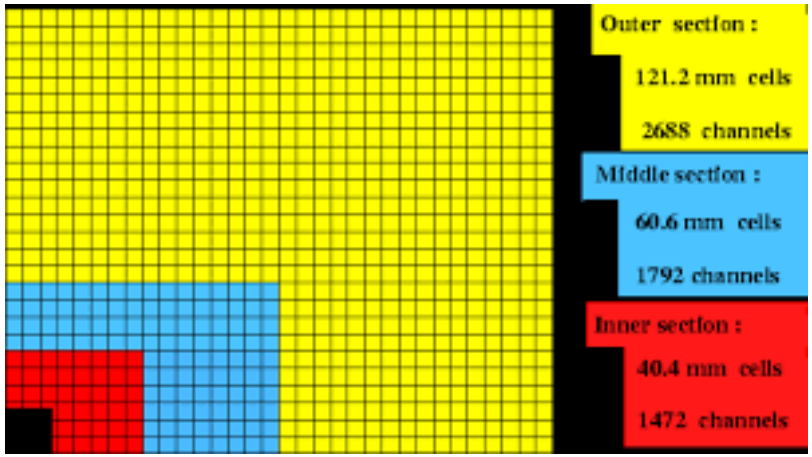
LHCb tracking system

Particles

- Charged particles: π, e, μ, K, p
- Neutral particles: π^0, γ

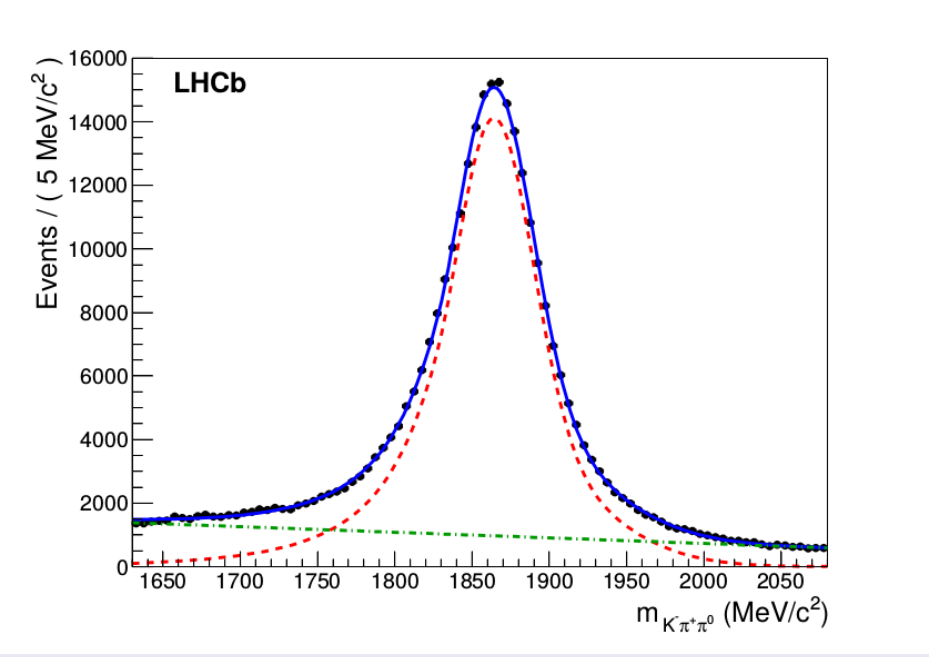
LHCb calorimeter





- Shashlik technology
- 1*1, 2*2, 3*3 module granularities

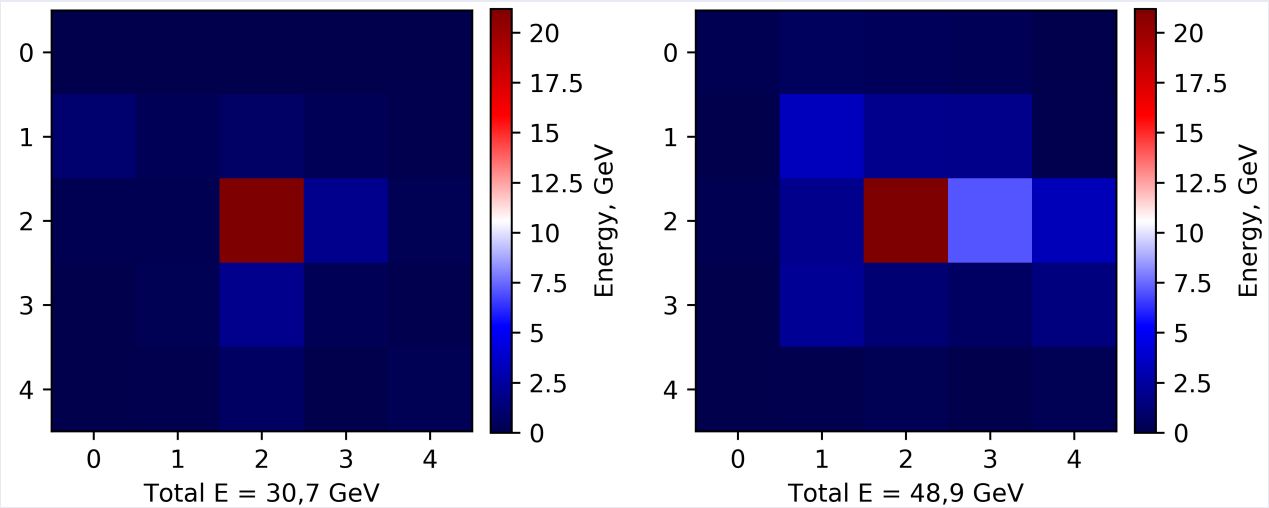
- Some π^0 decay into photons before the calorimeter
- We want to distinguish high energy photons from photons coming from π^0 decays



Suppression of the π^0's background

- Cluster - the group of cells, which show energy deposits where a particle hits the calorimeter

Clusters for photon and π^0photon



Responses from single photon(left) and merged π^0(right)

Quality metrics

- Efficiency (recall) on photons - number of true recognized photons to all real photons
- Fake rate (at a certain efficiency) on π^0's
- Receiver operating characteristic (ROC) curve as an independent characteristic of the model, ROC score - area under ROC curve
- Flat dependence on energy

Training and validation

- use $B^0 \to K\pi\gamma$ to obtain γ to training
- To prevent classifier from separating particle by energy, we use kinematically similar π^0 from $B^0 \to K\pi\pi^0$
- For stability check, we use $B^0 \to J/\psi K^*$ with $K^* \to K\pi^0$ as an extra π^0 source

Definitions

- (x_c, y_c) - coordinates of the cluster center of gravity, e_i - energy of the ith cell, (x_i, y_i) - the cell's coordinates.
- $S_{xx} = \frac{\sum_{i=1}^{N} e_i (x_i - x_c)^2}{\sum_{i=1}^{N} e_i}$, $S_{yy} = \frac{\sum_{i=1}^{N} e_i (x_i - x_c)^2}{\sum_{i=1}^{N} e_i}$, $S_{xy} = S_{yx} = \frac{\sum_{i=1}^{N} e_i (x_i - x_c)(y_i - y_c)}{\sum_{i=1}^{N} e_i}$
- E_{seed} - energy of the center seed, E_{cl} - energy in full cluster, E_{snd} - the second largest energy in cells

Baseline approach

LHCb-PUB-2015-016

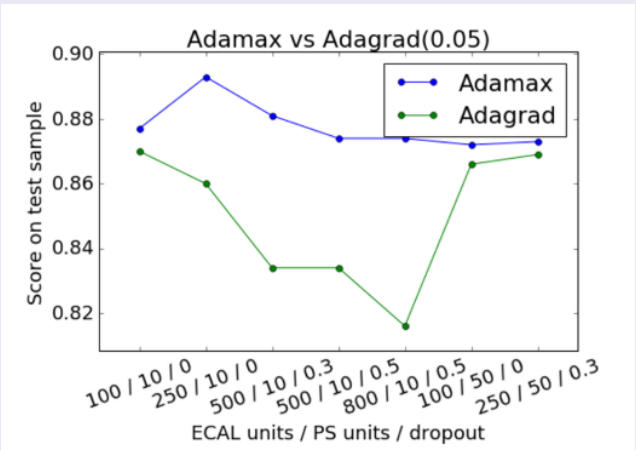
- Consider 3*3 cluster
- Use "shape" and "asymmetry" properties of the clusters as a features:
  $\frac{E_{seed}}{E_{cl}}$, $\frac{E_{seed}+E_{snd}}{E_{cl}}$,
  $k = \sqrt{(1 - 4\frac{S_{xx}S_{yy} - S_{xy}^2}{(S_{xx}+S_{yy})^2})}$,
  $asym = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$, $r_2 = \langle r \rangle = S_{xx} + S_{yy}$
- Use 2-layer TMVA MLP classifier

New approach

- Consider 5*5 cluster
- Use energy in each cell as a feature
- Use several models and look for the best one

New approach: Neural Network (NN) classifiers

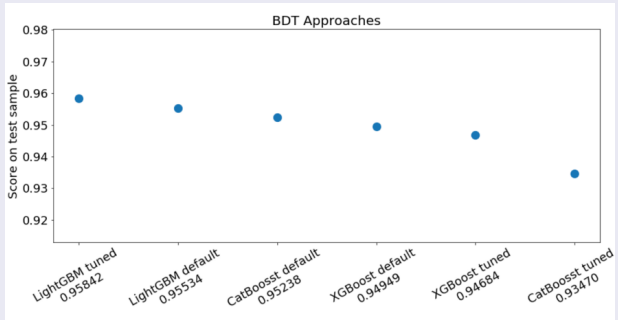
- We vary the number of layers, the number of units and the way of optimizing
- Neural network classifiers give at most 0.89 as a ROC-score



NN with different architecture based on Adamax (violet) and Adagrad (green) optimizer

New approach: Boosted Tree (BDT) classifiers

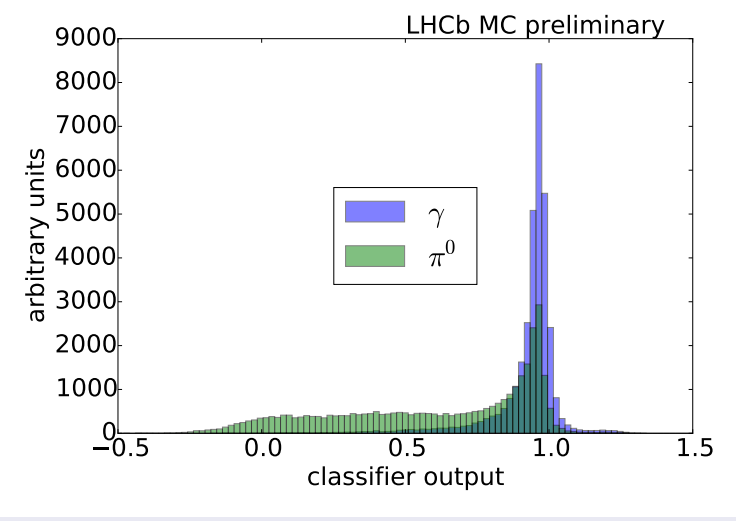
- We use LightGBM, XGBoost and CatBoost models
- Different models over the boosted decision tree give the similar results



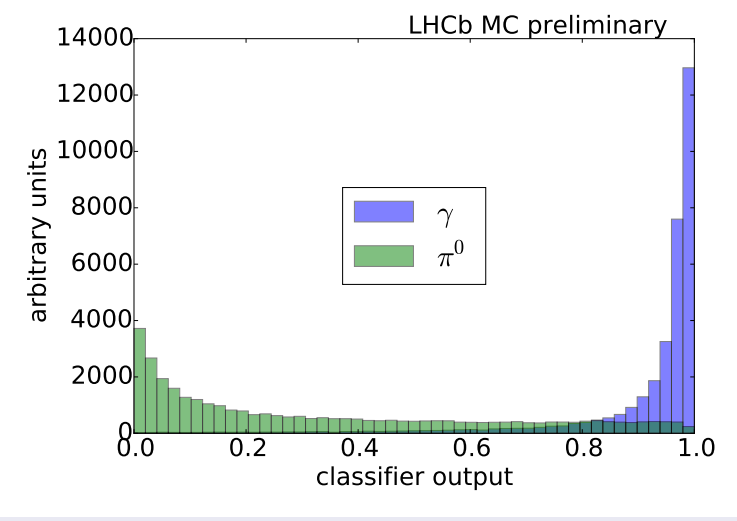
Scores of tuned and untuned BDT models

Results: Monte-Carlo (MC)

- Classifiers' responses on different particle types



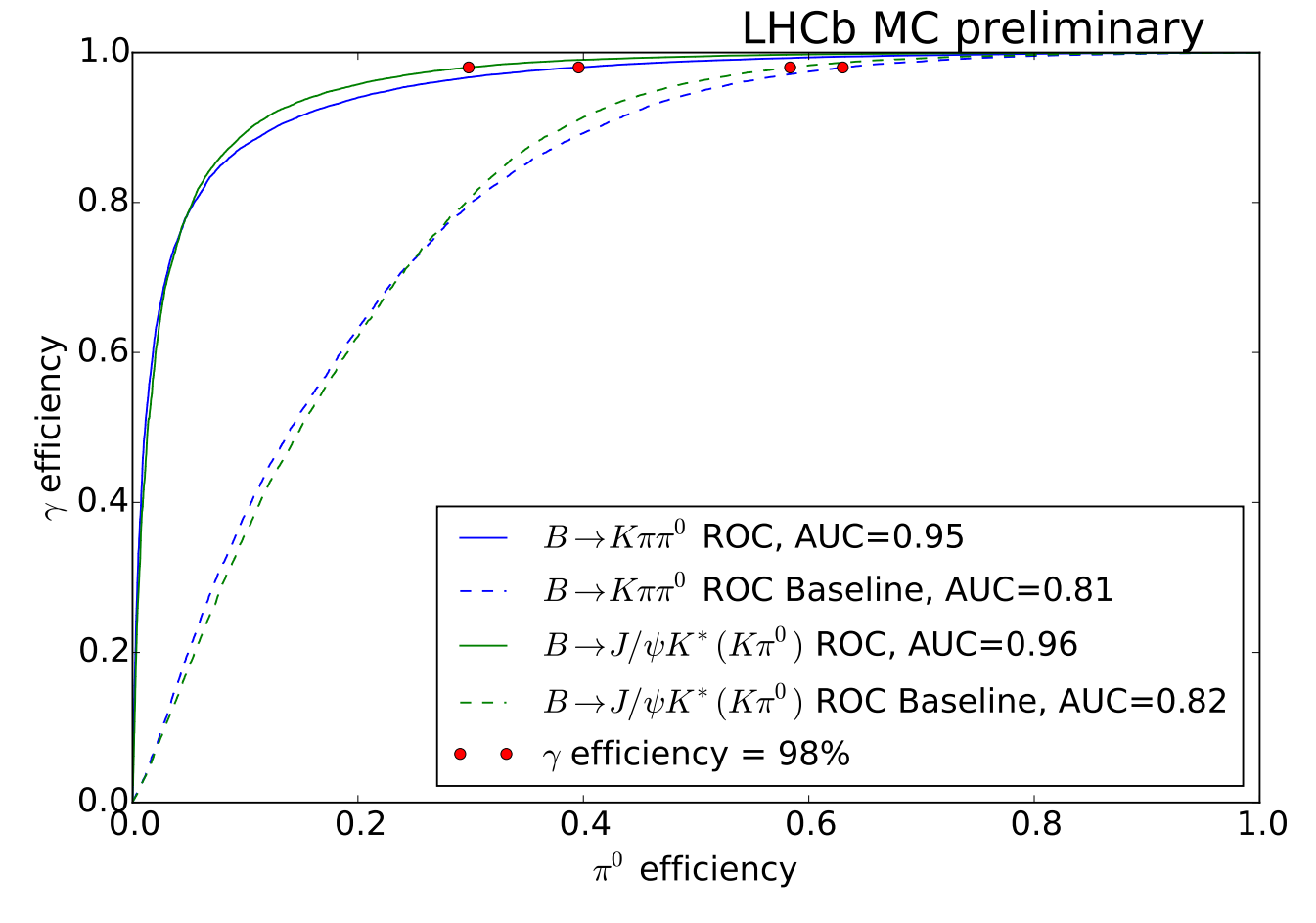
Baseline output



XGBoost approach output

Results: Monte-Carlo (MC)

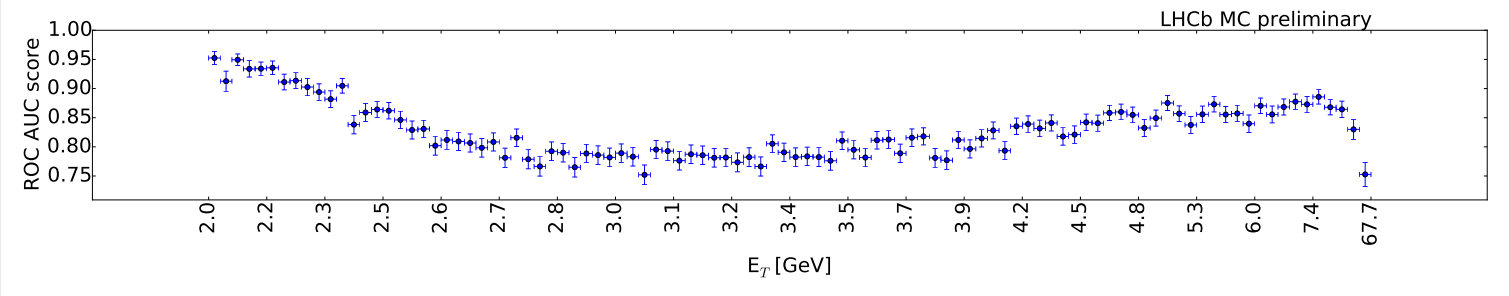
- Performance



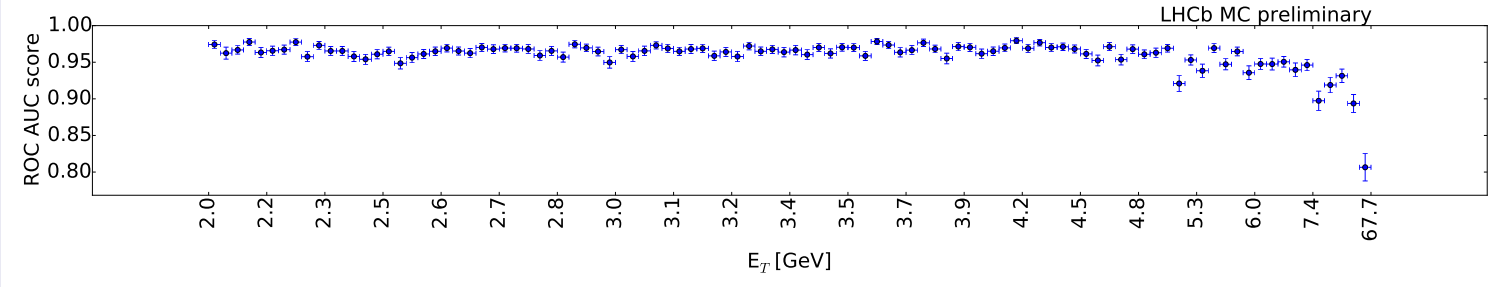
ROC curves for the baseline (dashed line) and new method (solid line). Different colors refer to different test samples

Dependency on transverse energy

- The flat dependency on energy can help to reduce the systematic uncertainties in the physics analysis



Baseline model quality as a function of transverse energy.



BDT model quality as a function of transverse energy.

Conclusions

- We developed a new procedure to separate photons from merged π^0.
- New approach shows good performance on simulated data. Classifier's quality does not depend on energy.
- Validation on real data requires a thoughtful approach.

Next step: Validating on real data

- It is not trivial to select calibration samples from real data.
- To train and validate we use π^0's from rare decays. The π^0 from real data have varied energy distributions, which can affect to classifier response.