Muon/Pion Identification Based on Machine Learning Algorithm at BESIII

Yuncong Zhai, Teng Li, Xingtao Huang

Shandong University, Qingdao, China

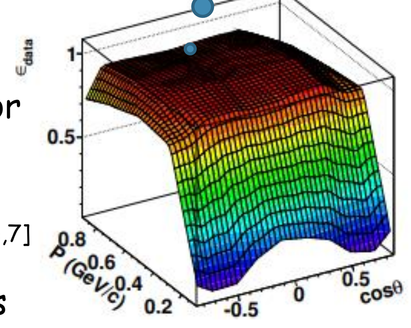
前沿交叉科学青岛研究院 INSTITUTE OF FRONTIER AND INTERDISCIPLINARY SCIENCE

INTRODUCTION

- Particle identification (PID) is one of the most important and commonly used tools for the physics analysis in collider physics experiments.
- For BESIII experiment, traditional methods like the maximum likelihood method are difficult to improve due to the intrinsic correlations between input variables.
  - Especially for very challenging problem: muon/pion separation
- In recent decades, Machine learning (ML) has provided a powerful toolbox.
  - ML based techniques have been rapidly developed and have shown successful applications in HEP experiments[1,2]
  - One of the obvious advantages of applying ML to PID is its capability of combing many correlated variables to solve the most difficult problems for traditional methods[3-5]
  - Preliminary results show that the gradient boosting decision tree (BDT) [6,7] model provides obviously higher discrimination power than traditional ones
- Targeting at the muon/pion identification problem at the BESIII experiment, we have developed a new PID algorithm based on the BDT algorithm.
  - Further improving the performance of traditional PID algorithms and exploring its physical potential

Great room for improvement at certain regions

□ The muon discrimination efficiency w.r.t. momentum and cos θ by traditional PID software.

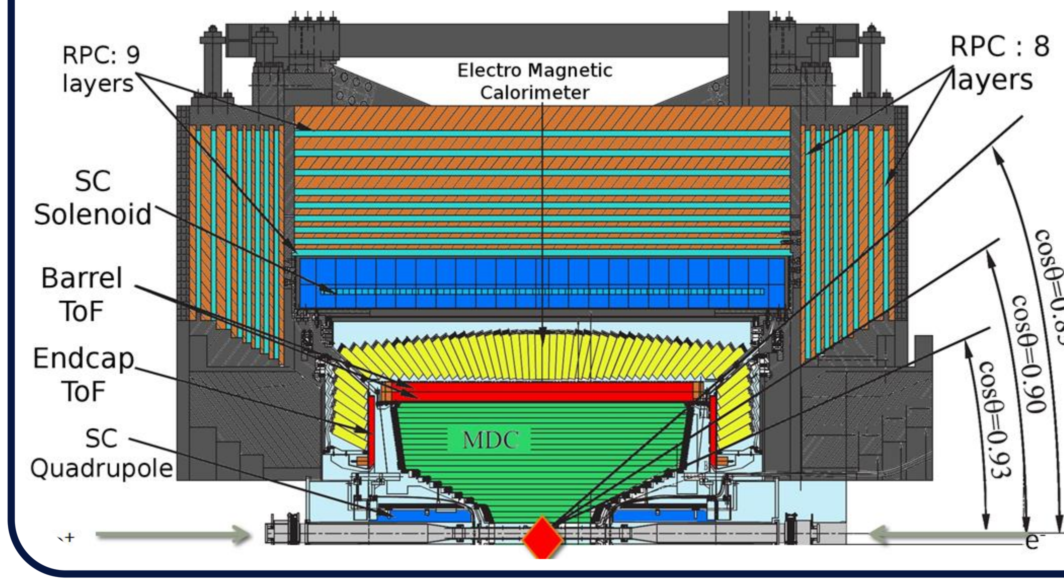


BESIII EXPERIMENT

- The Beijing Spectrometer III (BESIII) is a collider physics experiment running on the Beijing Electron-Positron Collider II (BEPCII) [8].

The BEPCII description

- Center-of-mass energy: 2.0 to 4.95 GeV
- Peak luminosity: 1 x $10^{33}cm^{-2}s^{-1}$ ($\sqrt{s}$ = 3.77 GeV)



The BESIII detector

- Helium-based multilayer drift chamber (MDC)
- Plastic scintillator time-of-flight system (TOF)
- CsI(Tl) electromagnetic calorimeter (EMC)
- Muon counter(MUC)
- Momentum resolution: 0.5% at 1 GeV/c
- dE/dx resolution: 6%
- Energy resolution: 2.5% (5%) at 1 GeV in the barrel (end cap)
- Time resolution : 68 ps (barrel), 60 ps (end cap)

METHODOLOGY

In order to fully explore the PID performance of the detector. Using advanced BDT (XGBoost[9]) , develop a novel muon/pion PID algorithm.

01 Configuration

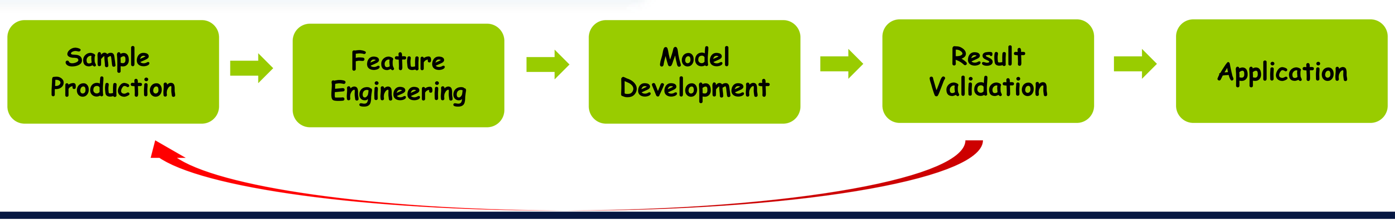
- Based on a data-driven approach, BDT is used as a key technical approach.
- Selected hyper-parameters:
  - max_depth: 8
  - n_estimators: 300

02 Systematic errors

- Systematic error:

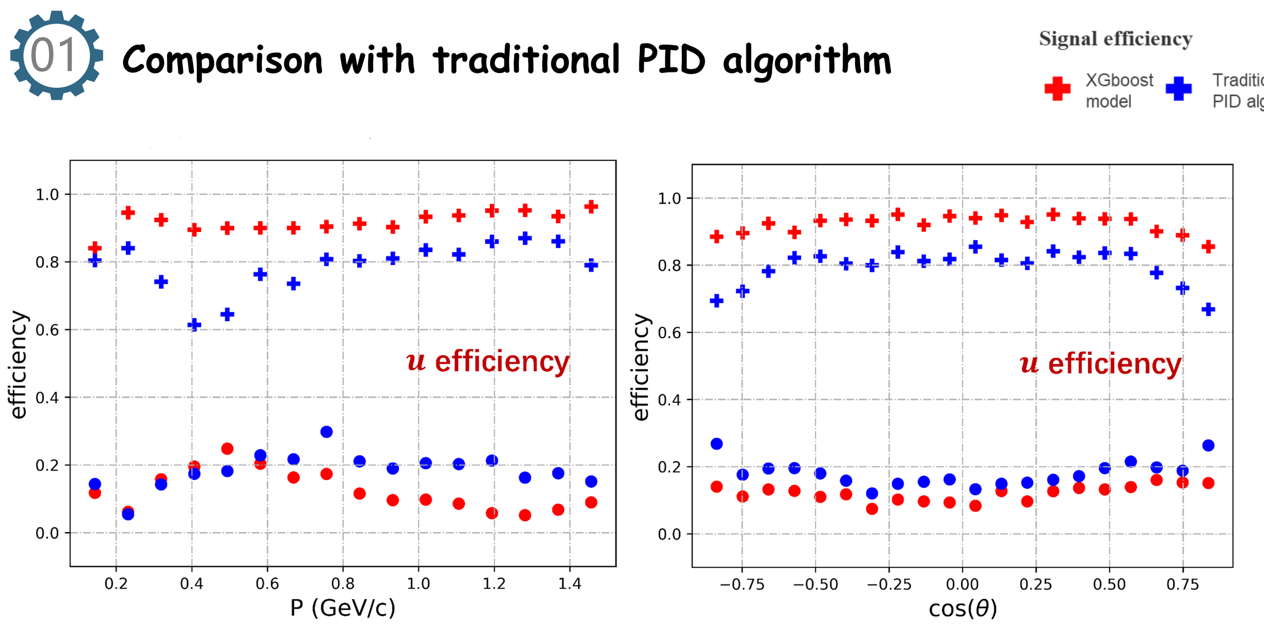
$$\Delta\varepsilon = \frac{\varepsilon(Data)-\varepsilon(MC)}{\varepsilon(MC)} \quad (\varepsilon:\ \text{PID efficiency})$$

- Through detailed cross-validation to evaluate deviations :
  - Different decay processes
  - MC/data



PERFORMANCE

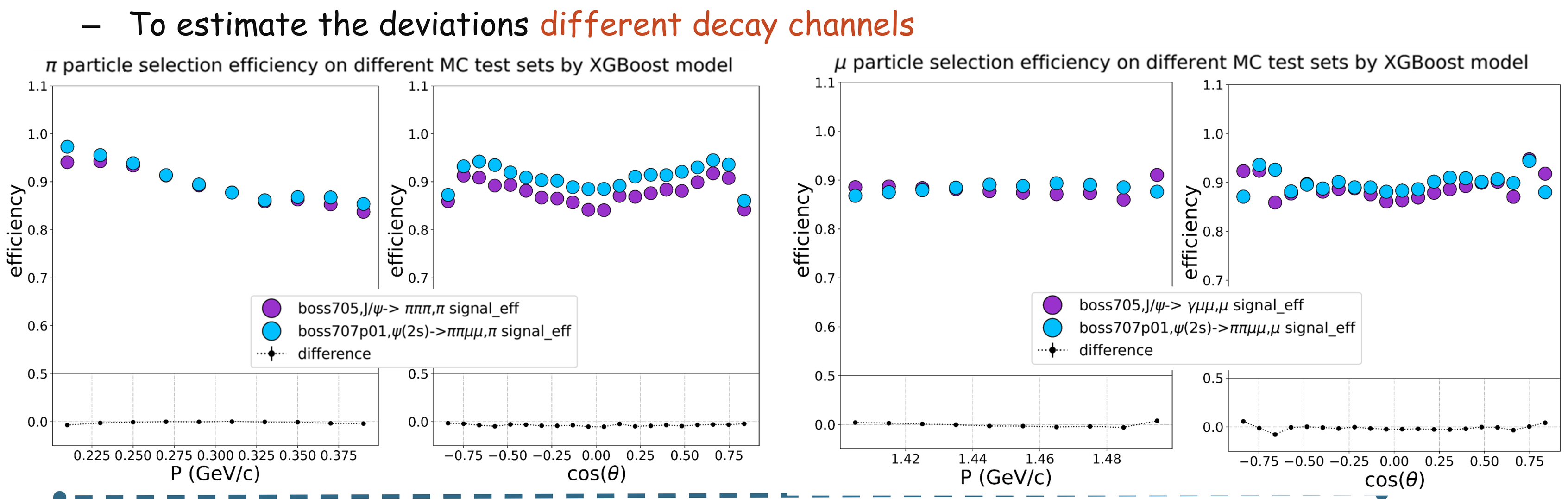
01 Comparison with traditional PID algorithm



- Signal efficiency : $\frac{\text{The number of signal selected correctly}}{\text{The total number of signal}}$
- Background efficiency : $\frac{\text{The number of background misidentified as signal}}{\text{The total number of background}}$

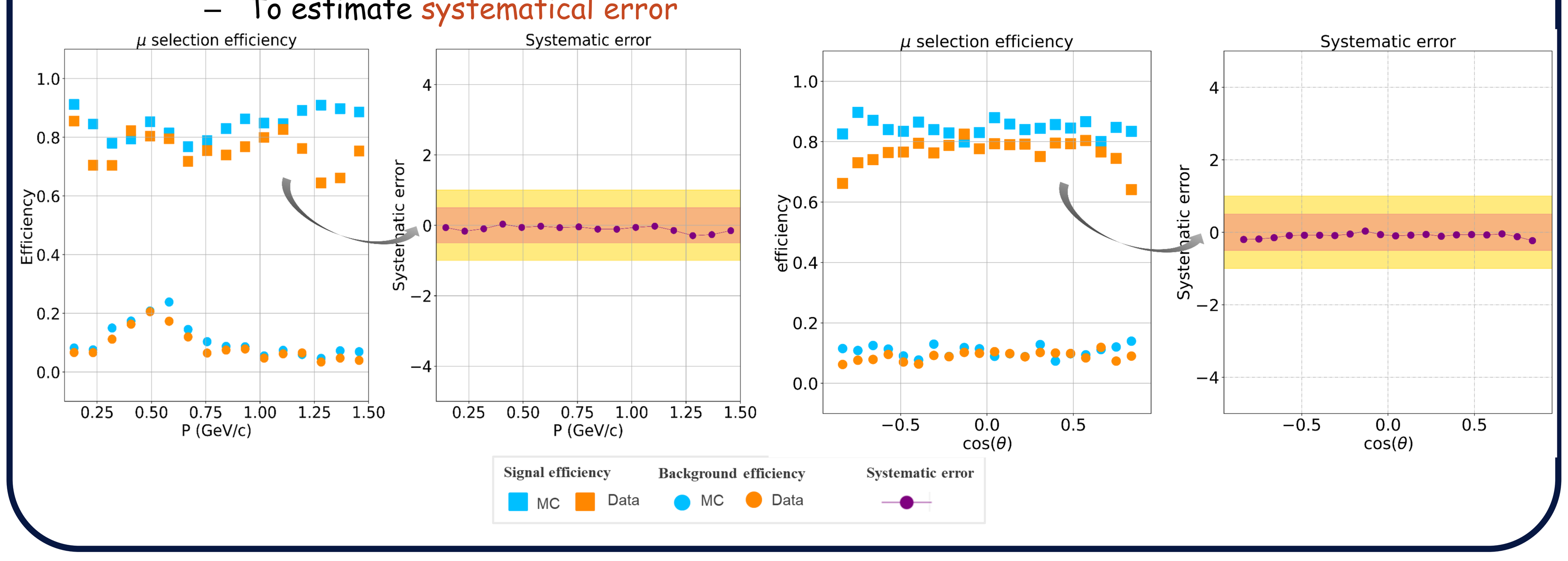
02 Cross validation between different decay processes

- To check generalization ability
- To estimate the deviations different decay channels



03 Cross validation between MC and Data

- To estimate systematical error

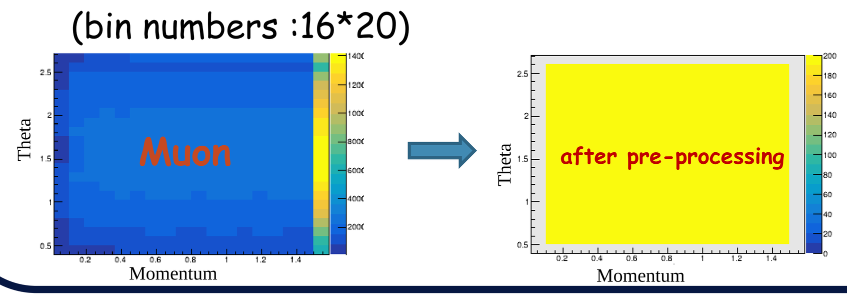


DATA SAMPLE

Based on the substantial amount of high-quality Monte Carlo simulation (MC)/real data samples from BESIII, relying on its mature offline software system (BOSS).

Train sample

- Single muon/pion MC samples
- High purity and well distribution (Pre-processing)
  - Make sure the distribution of p and cos θ is flattened to avoid bias
  - 0.1 GeV/c < p < 1.5 GeV/c, −0.88 < cos θ < 0.88 (bin numbers :16*20)



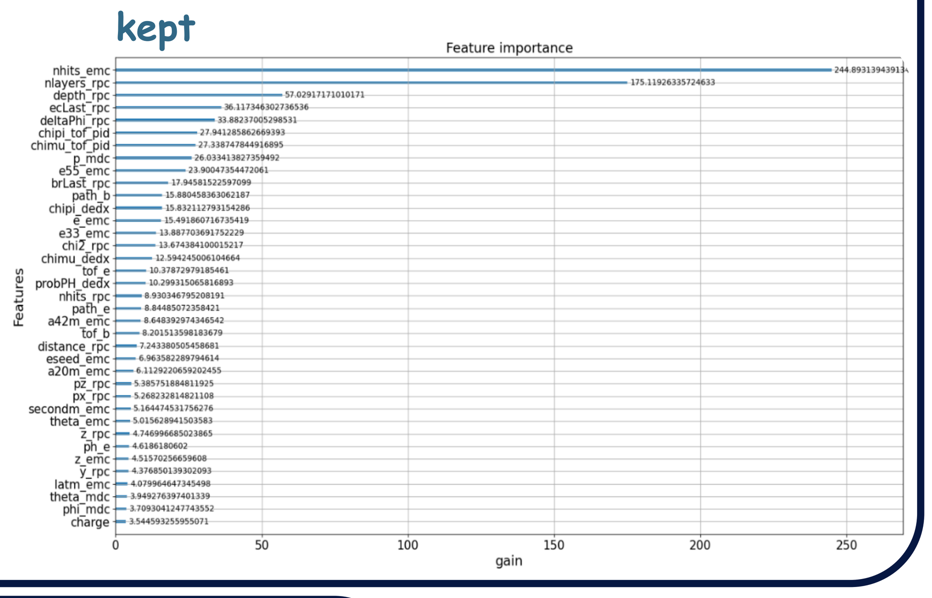
Cross-validation sample

The purity (P) of the μ/π samples : $\frac{N_{sample\ ture}}{N_{sample}}$

- MC/data:
  - $J/\psi \to \pi^+\pi^-\pi^0 \to \pi^+\pi^-\gamma\gamma$ (P = 99.37%)
  - $J/\psi \to \gamma\mu^+\mu^-$ (P = 97.97%)
- Different decay processes:
  - $\psi(2s) \to \pi^+\pi^- J/\psi \to \pi^+\pi^-\mu^+\mu^-$ (P = 99.13%)

FEATURE SELECTION

- To extract effective features from a large amount of interrelated sub-detectors information.
- First model trained with all 108 features.
  - Contain MDC, dE/dX, TOF, EMC, MUC information
  - Based on XGBoost
- Features are then selected according to feature importance.
- Eliminate redundant and strongly-correlated features, 37 features are kept



CONCLUSION

- ✓ A muon/pion identification algorithm based on machine learning model (XGBoost) is developed based on the high quality data samples and has been integrated into the BOSS.
- ✓ Performance analysis shows XGBoost model provides obviously higher discrimination power than traditional methods.
- ✓ Detailed cross-validation was conducted and an evaluation method for the systematic error of the machine learning model was provided, which can be used by BESIII physics analysts.

REFERENCE

[1]Dimitri Bourilkov. Machine and deep learning applications in particle physics. International Journal of Modern Physics A, 34(35):1930019, Dec 2019.

[2] Karagiorgi G, Kasieczka G, Kravitz S, et al. Machine learning in the search for new fundamental physics[J]. Nature Reviews Physics, 2022, 4(6): 399-412.

[3] Graczykowski Ł K, Kabus M, Jakubowska M, et al. arXiv: Using Machine Learning for Particle Identification in ALICE[R]. 2022.

[4] Abratenko P, Alrashed M, An R, et al. Convolutional neural network for multiple particle identification in the MicroBooNE liquid argon time projection chamber[J]. Physical Review D, 2021, 103(9): 092003.

[5] M. Hushchyn, V. Chekalina,Particle-identification techniques and performance at LHCb in Run 2,Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,Volume 936,2019,Pages 568-569,ISSN 0168-9002,

[6] Robert E Schapire. A brief introduction to boosting. In Ijcai, volume 99, pages 1401-1406. Citeseer, 1999.

[7] Laviniu Aurelian Badulescu. Attribute selection measure in decision tree growing. 2007.

[8] Guo, YX., Sun, SS., An, FF. et al. The study of time calibration for upgraded end cap TOF of BESIII. Radiat Detect Technol Methods 1, 15 (2017). https://doi.org/10.1007/s41605-017-0012-4

[9]Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pages 785–794, 2016.

CHEP 2023

Author Email : zhaiyc@mail.sdu.edu.cn