Institute of High Energy Physics Chinese Academy of Sciences

University of Chinese Academy of Sciences

ML-based PID at BESIII

Hao Yuan, Zhenyuan Chen, Shengseng Sun

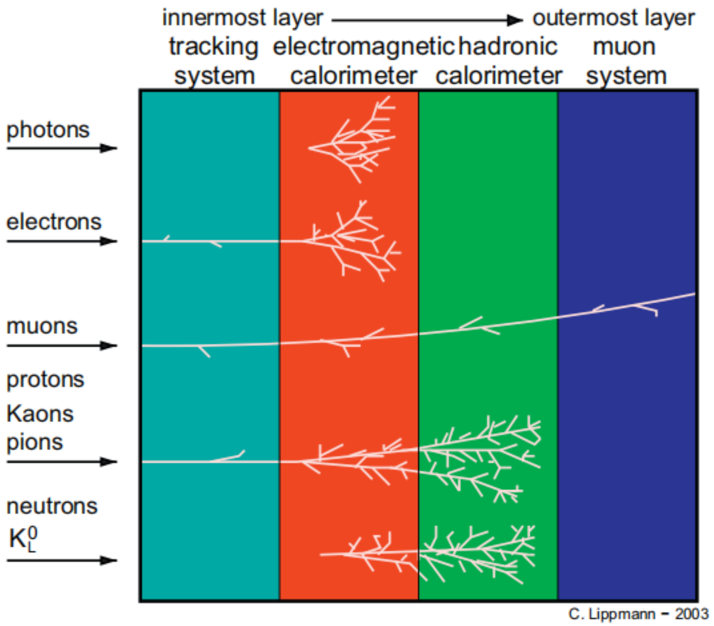
STCF 2026.7.2

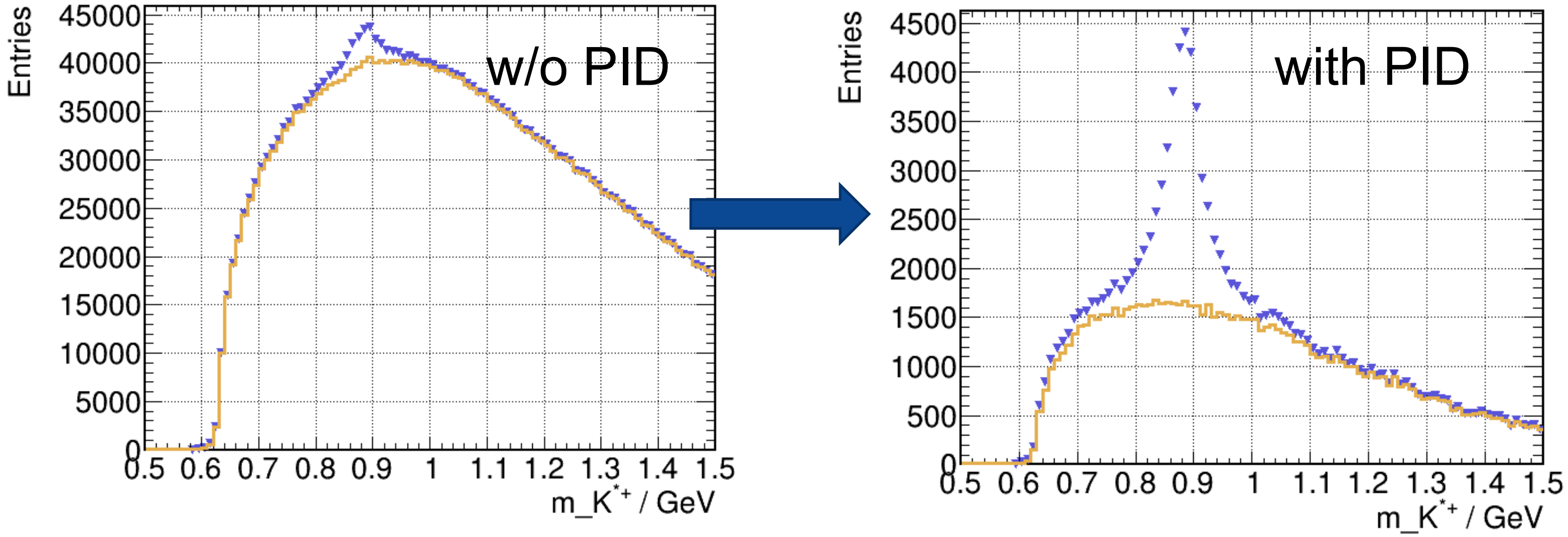
Xi an

Content

- PID
- BESIII
- ML
- Performance
- Application
- Summary

Why Identify Particles ?



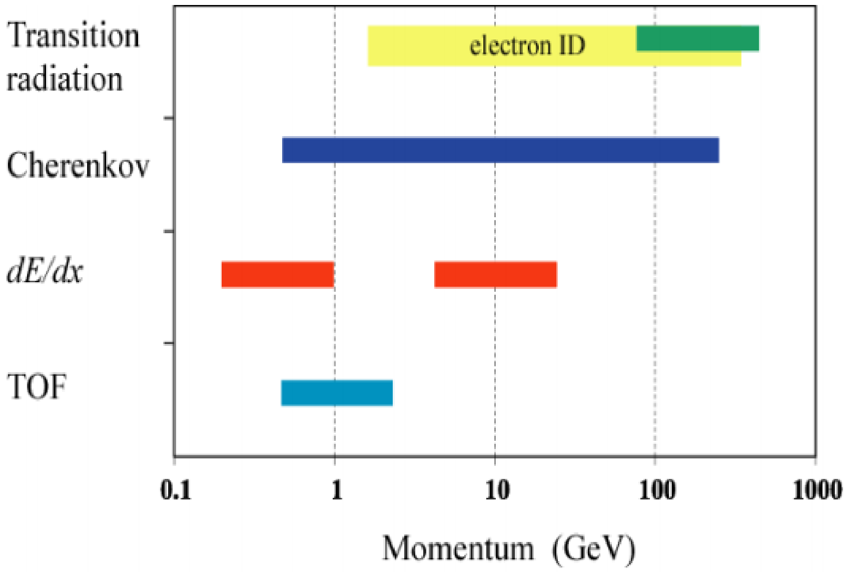


incMC J/ψ -> $K^{-}K^{*+}$

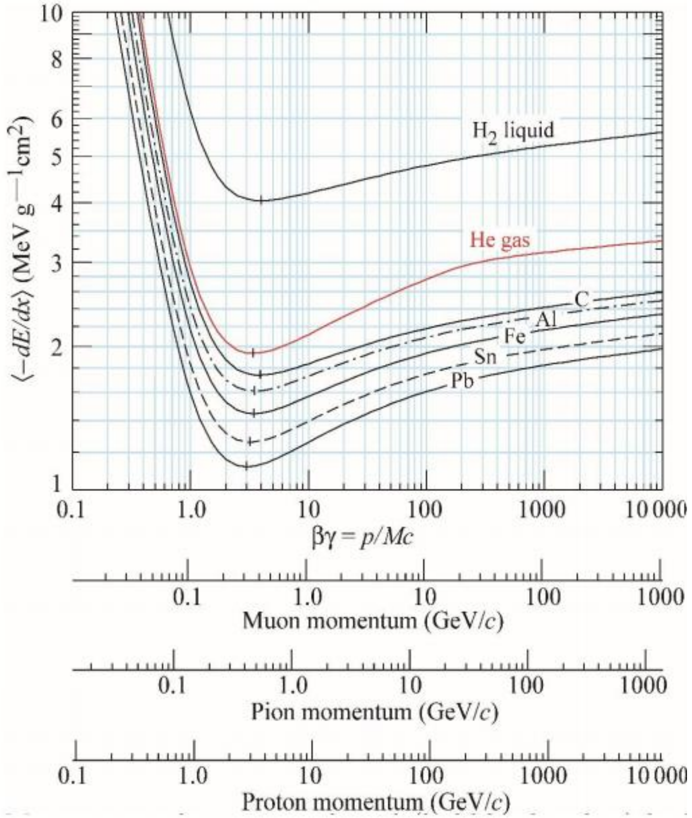
- Particle Identification (PID) is a crucial aspect of most particle physics experiments
- The effective identification of final state particles in physical processes significantly reduces background and improves the accuracy of physical analysis

How to Perform Particle Identification ?

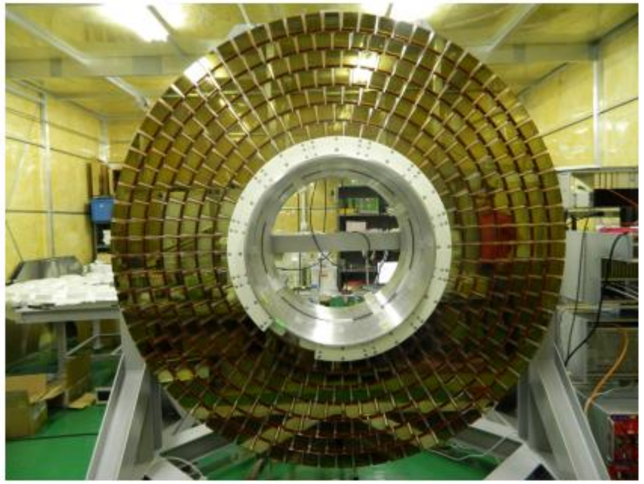
- Energy Loss Measurement (dE/dx)
- Time of Flight (TOF)
- Cherenkov Radiation
- Transition Radiation



dE/dx

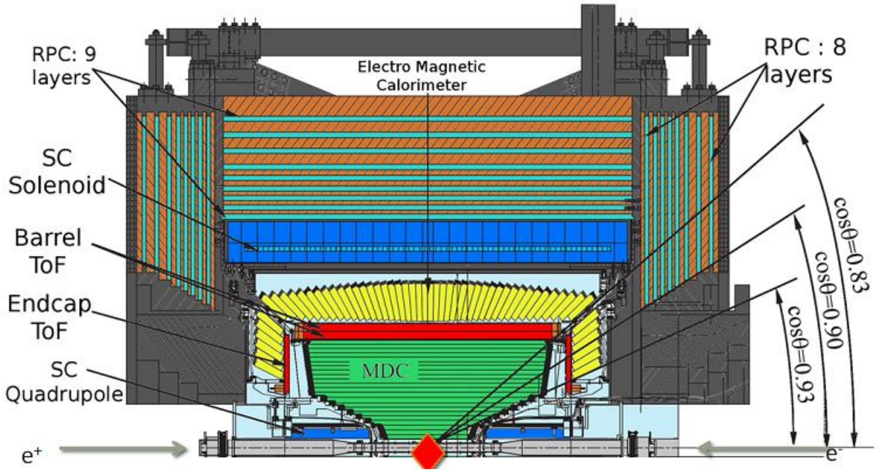


Belle II ARICH



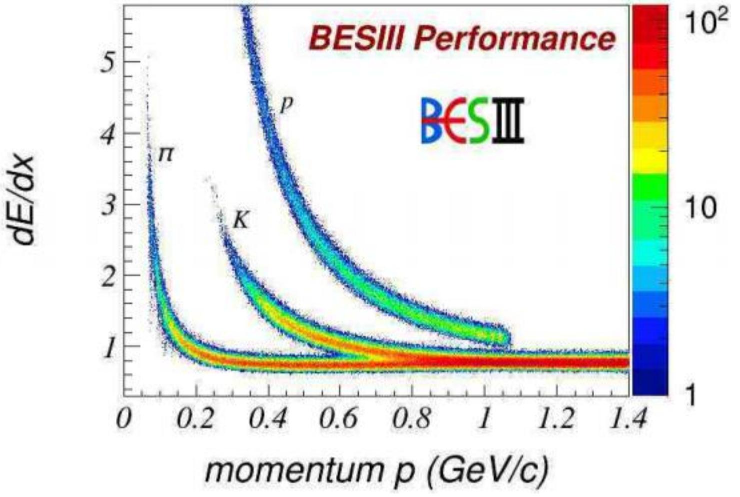
$$v_{th} \geq \frac{c}{n} \; or \; \beta \geq \frac{1}{n} \Rightarrow \cos\theta_c = \frac{1}{n\beta}$$

BESIII PID

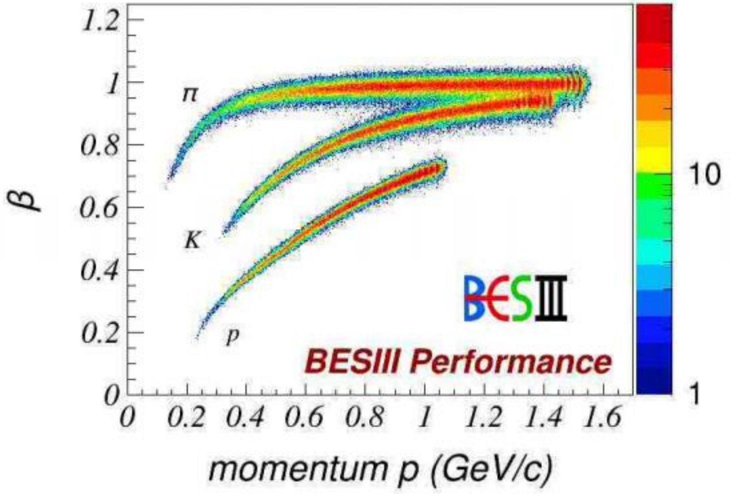


- The BESIII experiment operates in the τ-charm energy region
- Final states of physical processes predominantly consist of hadronic particles (π, K, p)
- Conventional PID methods: dE/dx + TOF
- The PID efficiency in the high momentum region does not fully meet the physical requirements

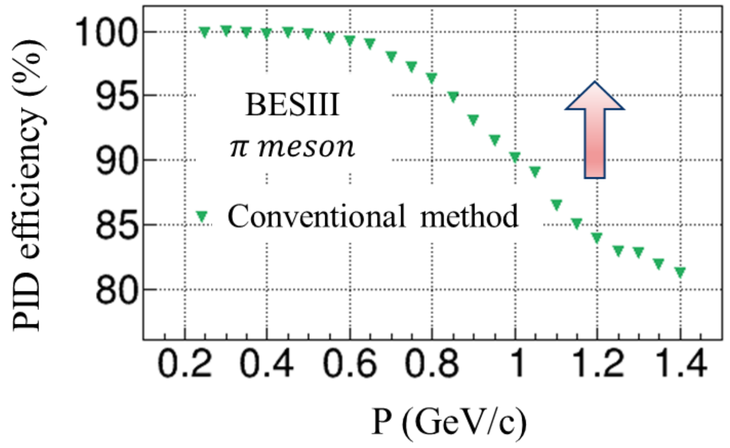
dE/dx



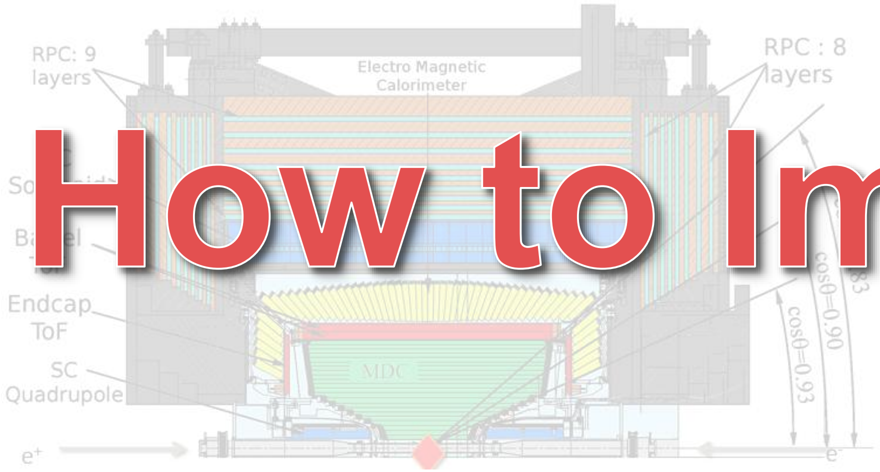
TOF



π PID efficiency



BESIII PID

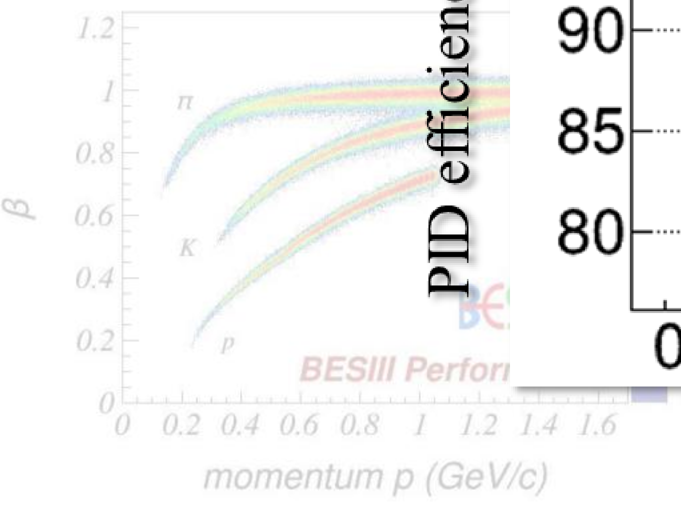


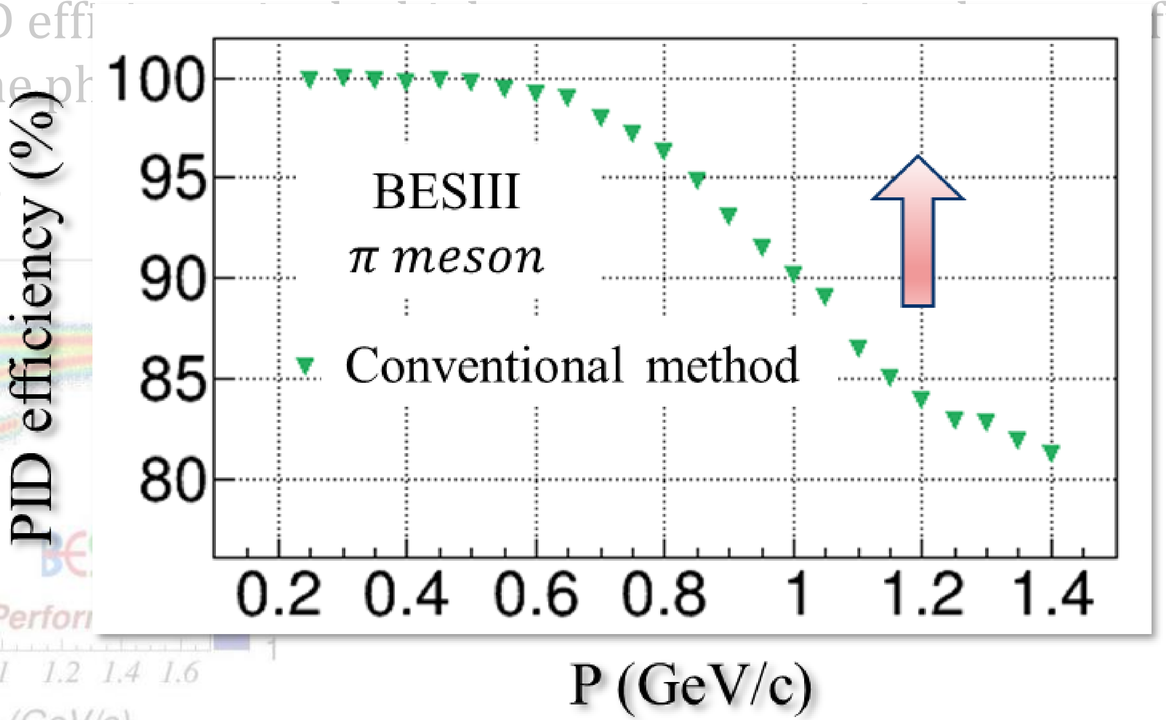
- The BESIII experiment operates in the τ-charm energy region
- Final states of physical processes predominantly consist of hadronic particles (π, K, p)
- Conventional PID methods: dE/dx + TOF
- The PID effi… fully meet the ph…

How to Improve

dE/dx

TOF

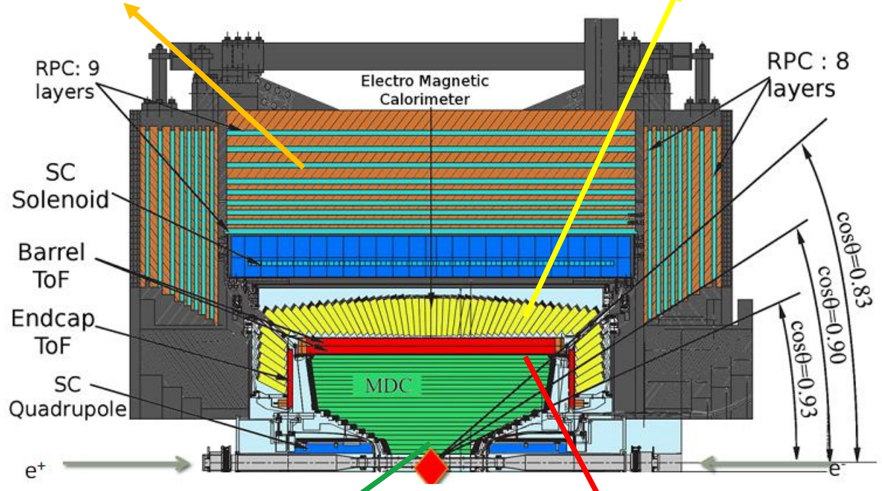




More Detector Information + ML

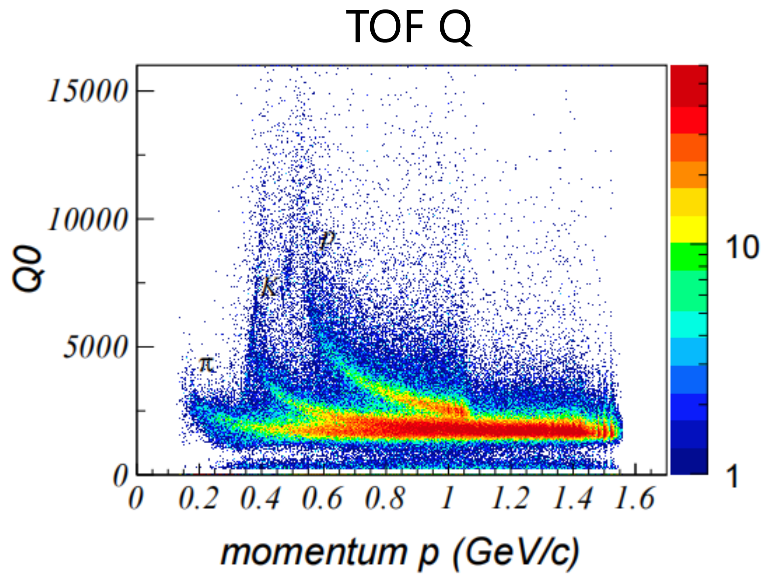
MUC:
$depth$, mucnumHits, mucnumLayers…

EMC:
E/p, $e3\times3/e5\times5$, secondMoment…



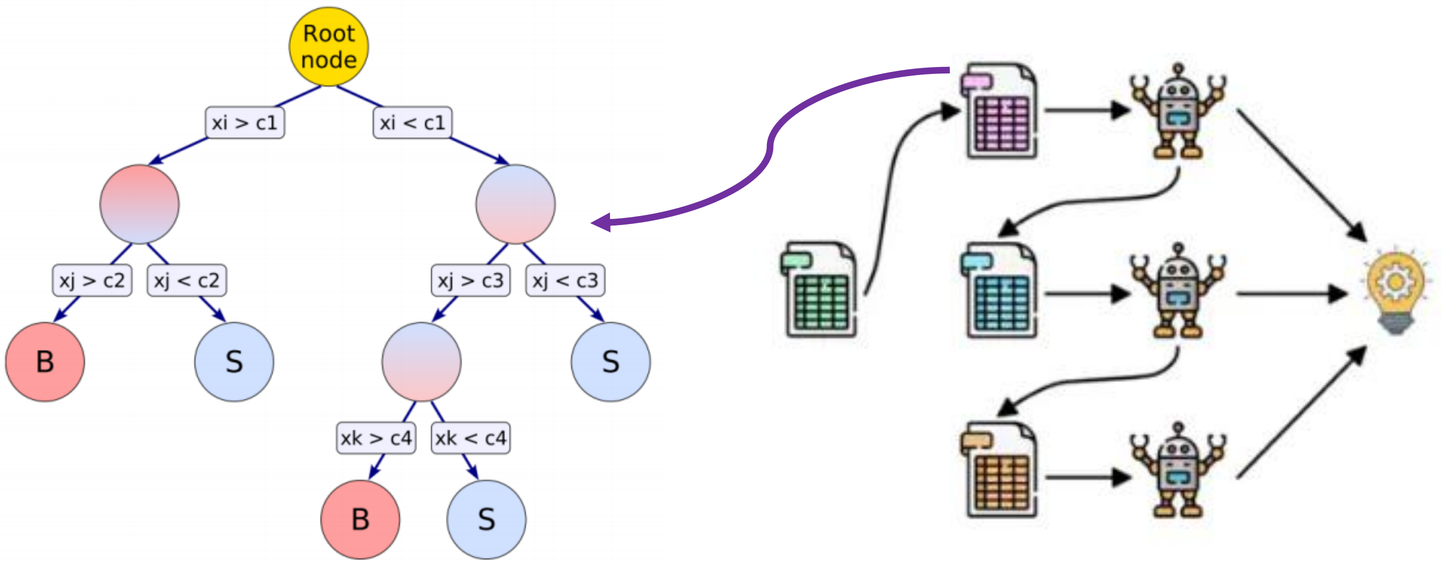
MDC:
$cos\theta$, momentum, charge …

TOF:
Q_0, t_{11}, t_{12}, β_{11}
β_{12} …

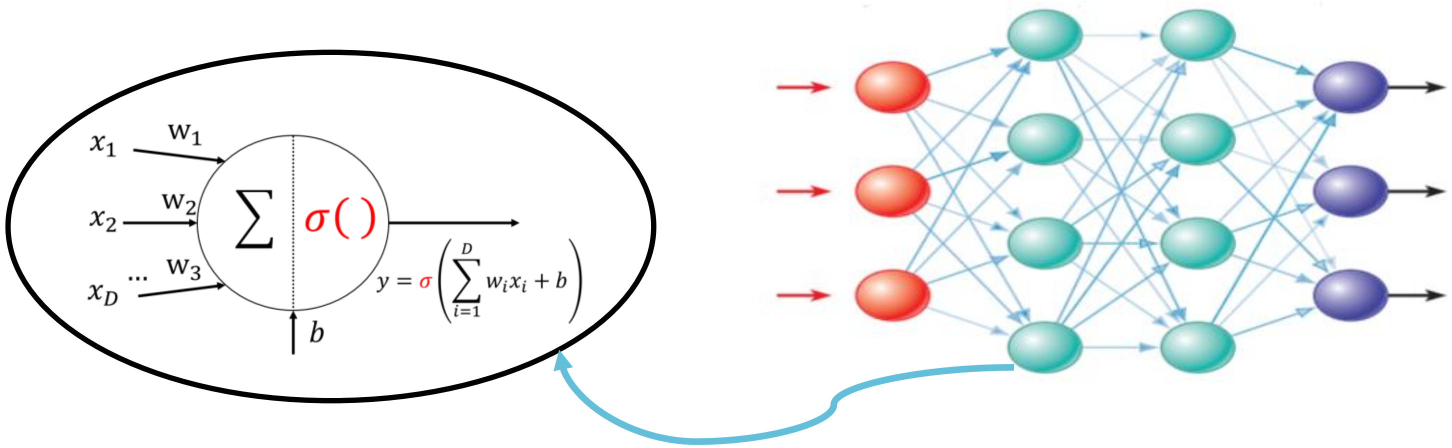


- ➢ Some other detector information also has particle identification capabilities
- ➢ Utilizing machine learning to integrate information from four sub-detectors enhances PID efficiency

ML Models



BDTG

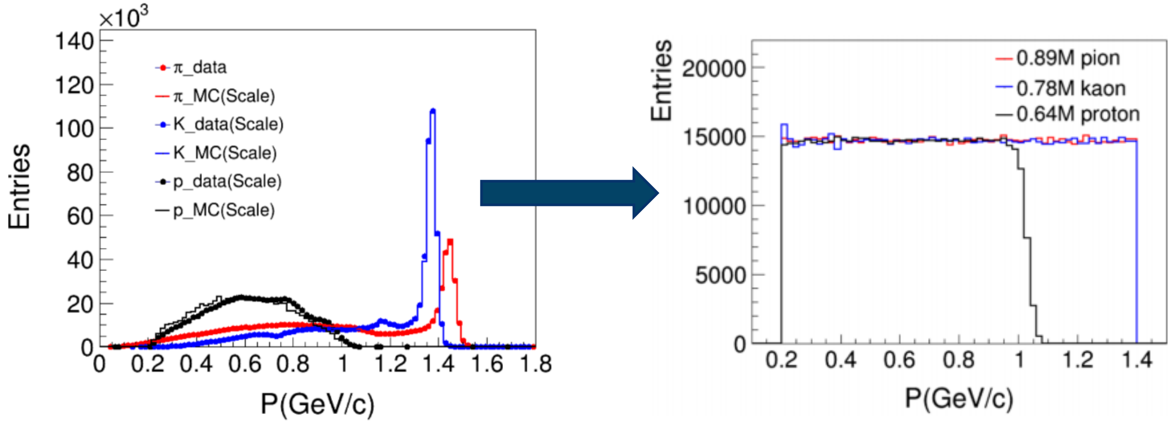


DNN

ML Models

Data Acquisition And Processing → Feature Selection → Hyperparameter Tuning → Model Evaluation

- **Event selection** from 10 billion J/ψ (MC/real data) events without particle identification:
    1. $J/\psi \rightarrow \pi^+ \pi^- \pi^0, \pi^0 \rightarrow \gamma\gamma$
    2. $J/\psi \rightarrow K_s{}^0 K^{\pm} \pi^{\mp}, K_s{}^0 \rightarrow \pi^+ \pi^-$
    3. $J/\psi \rightarrow p\bar{p}\pi^+\pi^-$
- **Data cleaning:**
    - ✓ Remove outliers
    - ✓ TOF, EMC, and MUC measurements are set to 0 for particles with momentum below 0.4 GeV/c
- **Data Transformation**:



MC/real data	Train sample	Validation sample	Test sample
pion	0.57M(0.56M)	0.14M(0.14M)	0.18M(0.17M)
kaon	0.50M(0.51M)	0.12M(0.13M)	0.16M(0.16M)
proton	0.41M(0.39M)	0.10M(0.10M)	0.13M(0.12M)

ML Models

Data Acquisition And Processing → **Feature Selection** → Hyperparameter Tuning → Model Evaluation

100+ Feature

MUC: $depth$, mucnumHits, mucnumLayers…	EMC: $E/p, e3\times3/e5\times5$, secondMoment…
MDC: $dE/dx,\ cos\theta$, momentum, charge …	TOF: $Q_0, t_{11}, t_{12}, \beta_{11}$ β_{12} …

- Filter methods
- Wrapper Methods
- Embedded Methods

→

16 Feature

MUC: 1 $depth$	EMC: 6 $E/p, eSeed/e3\times3$, secondMoment,latMoment, $NHits_{EMC}, \Delta\theta$
MDC: 4 $dE/dx,\ cos\theta$, momentum, charge	TOF: 5 $Q_0, t_{11}, t_{12}, t_{21}\ t_{22}$

Is more always better?

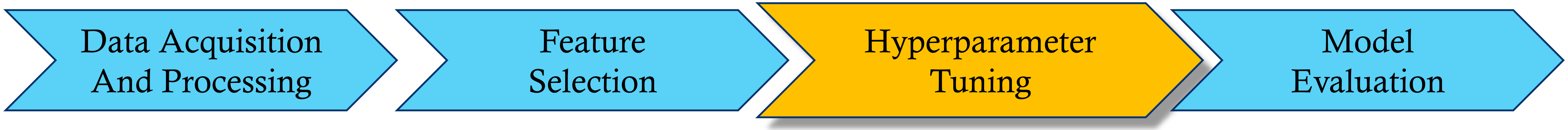
- Larger Models
- Increased Time Consumption
- Risk of Overfitting

✓ higher accuracy and better generalization of the model

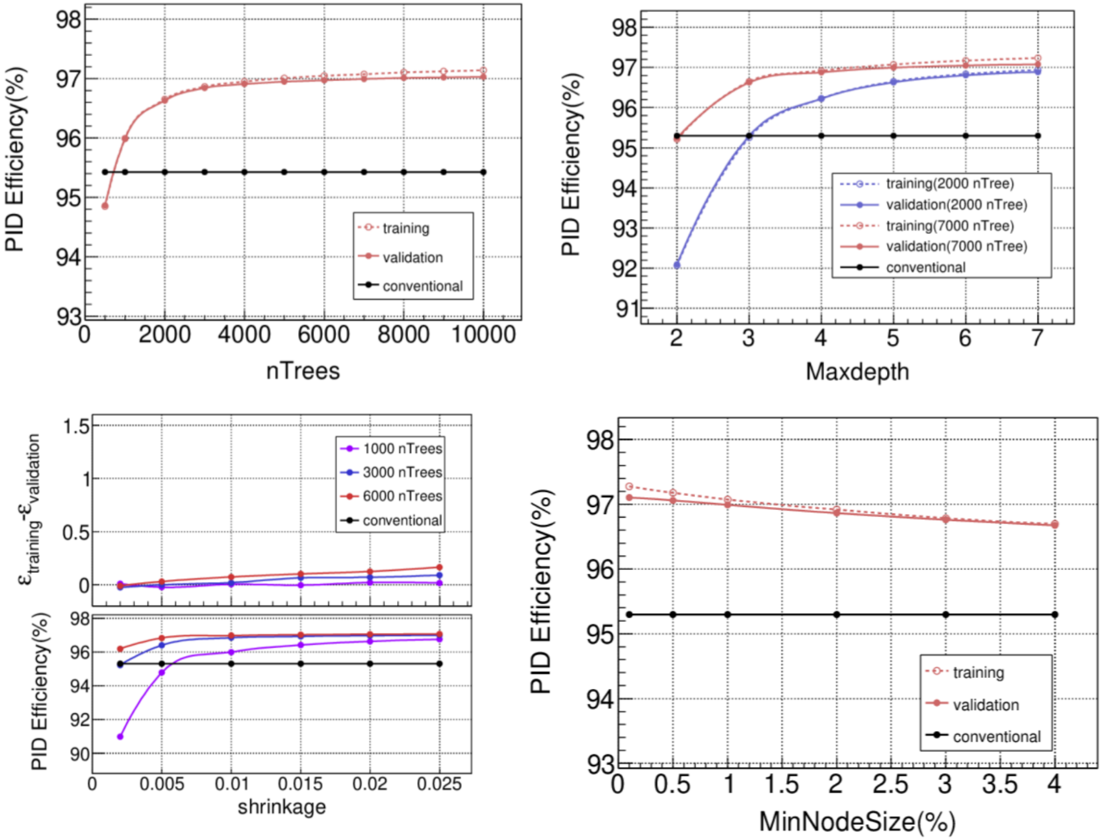
✓ computationally less expensive

✓ lower the risk of overfitting

ML Models



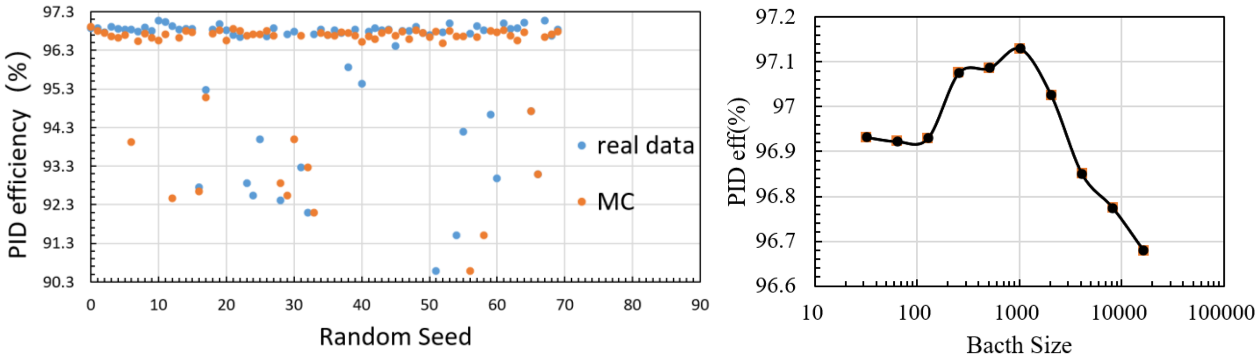
BDTG



DNN

Network Architecture

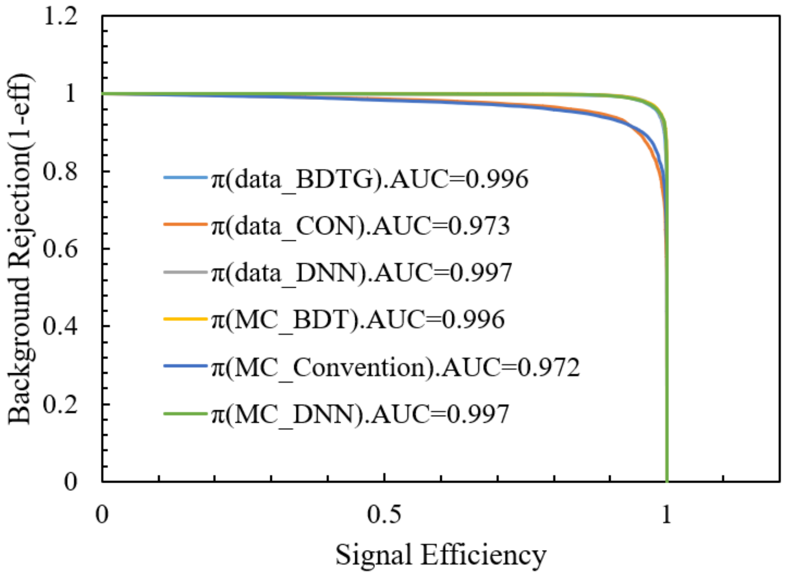
- **Input Layer:** 16 features
- **Hidden Layers:** TANH|20,TANH|40,TANH|20
- **Output Layers**: 3 (π, K, p) $\left(\mathrm{Softmax}(\mathrm{y_i}) = {e^{y_i}}/{\sum_j e^{y_j}}\right)$



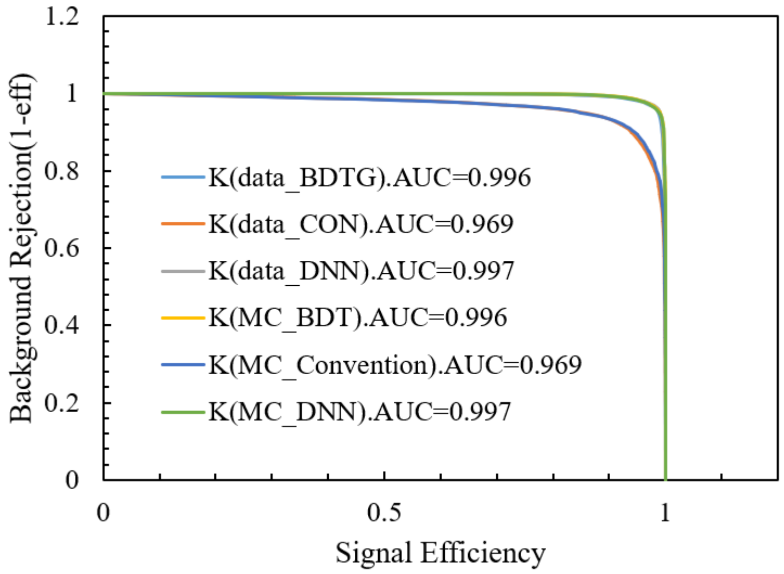
ML Models

Data Acquisition And Processing → Feature Selection → Hyperparameter Tuning → **Model Evaluation**

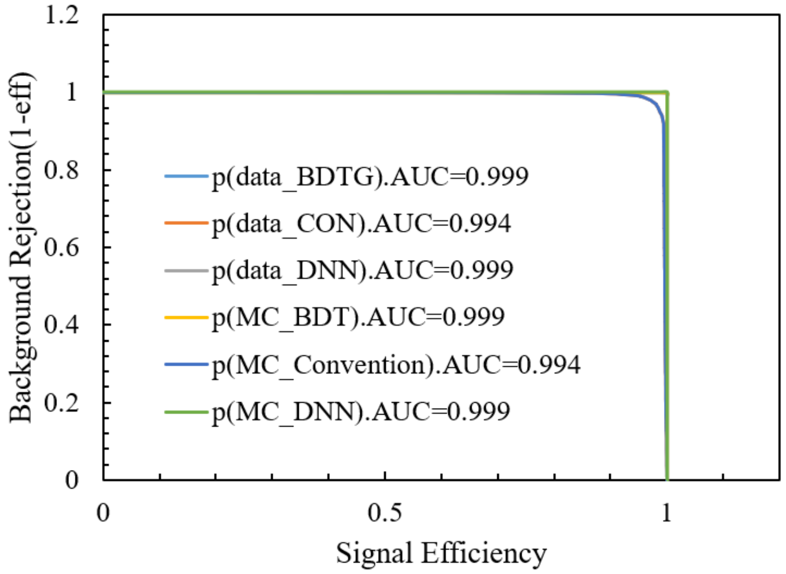
π



K

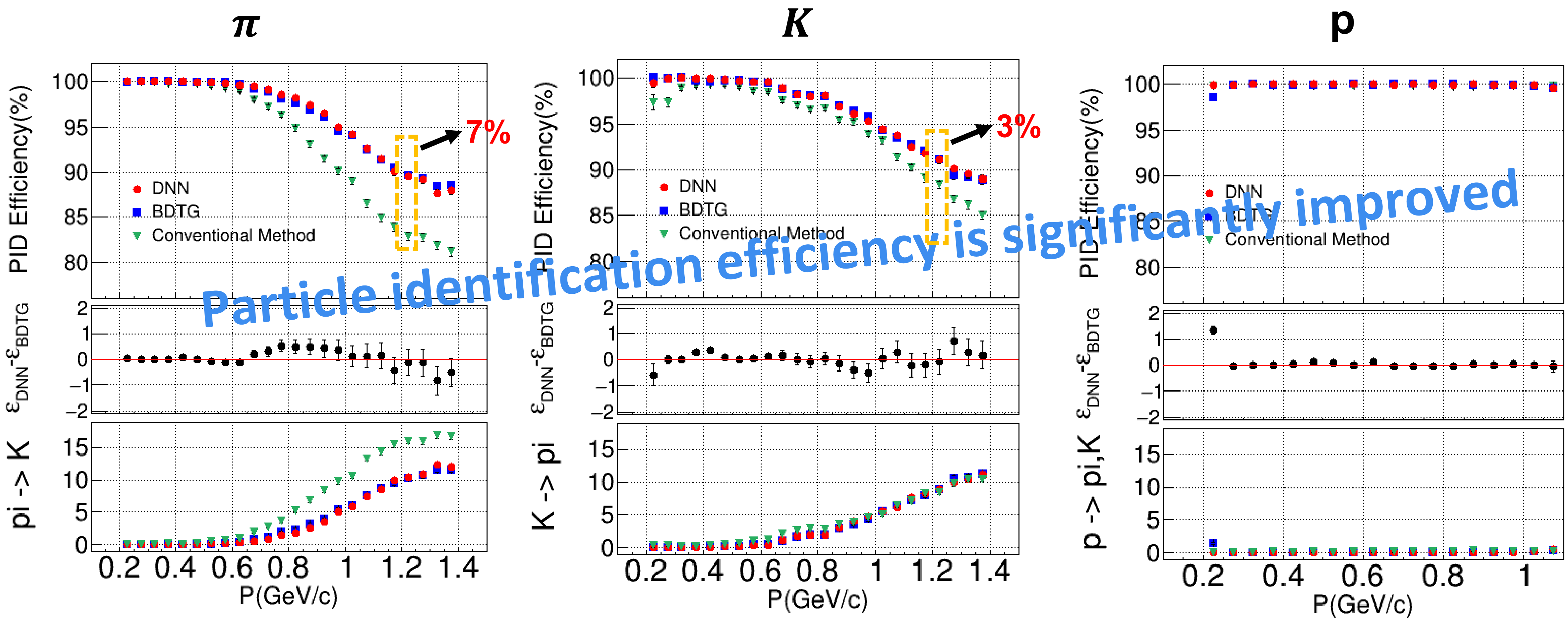


p



Performance —— PID eff vs momentum

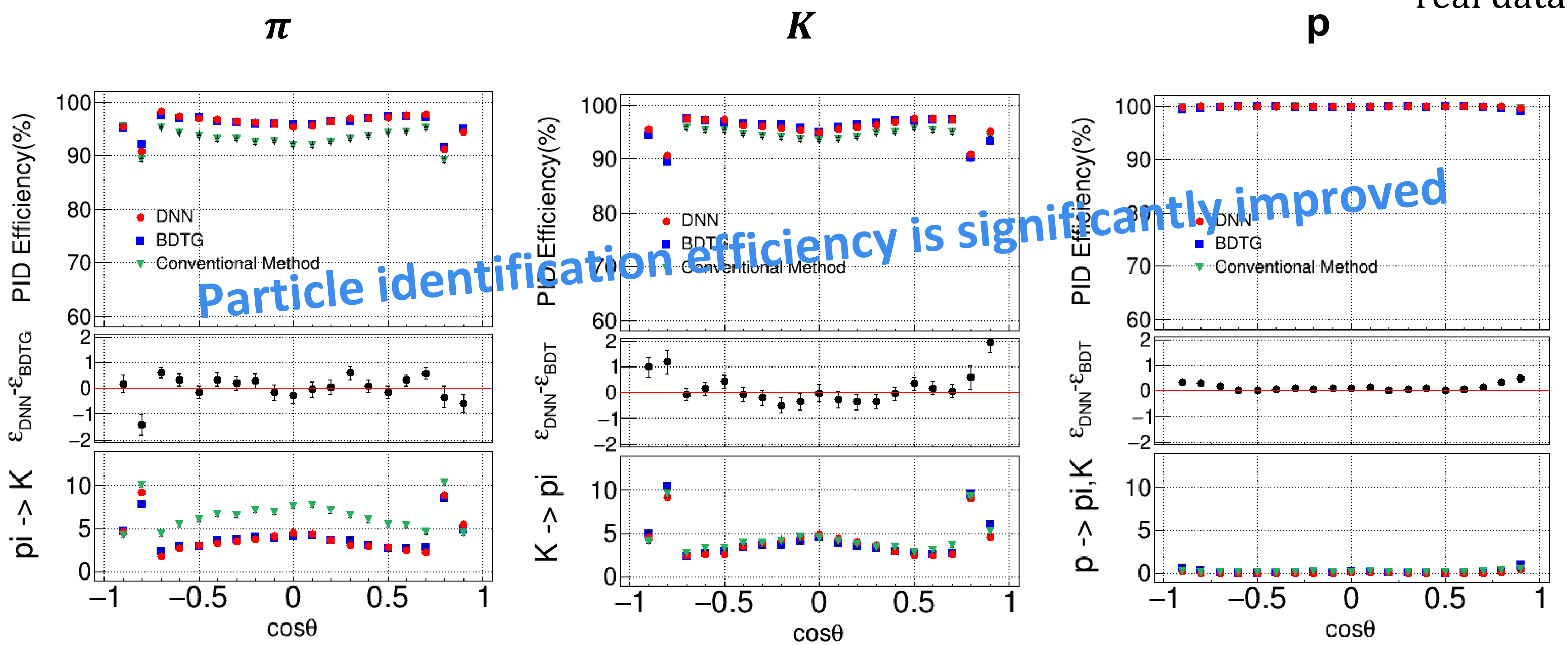
real data



Our initial goal has been achieved——higher PID eff

Performance —— PID eff vs $cos\theta$

real data

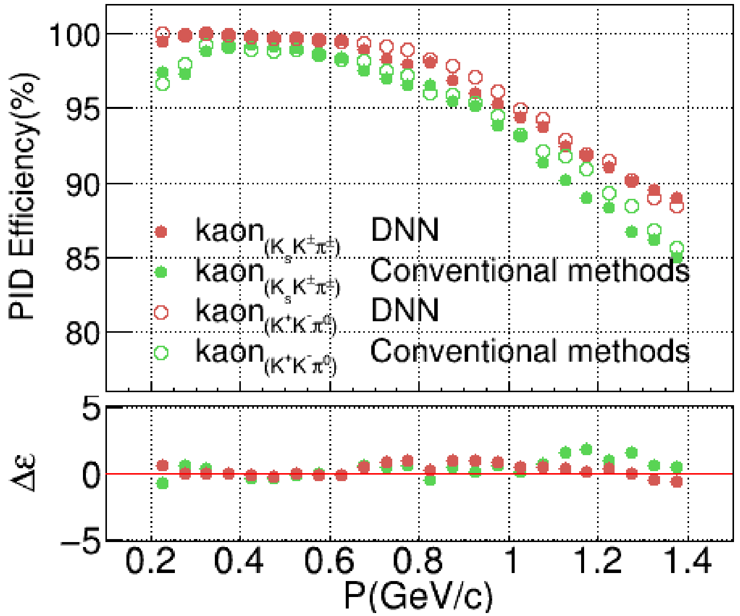


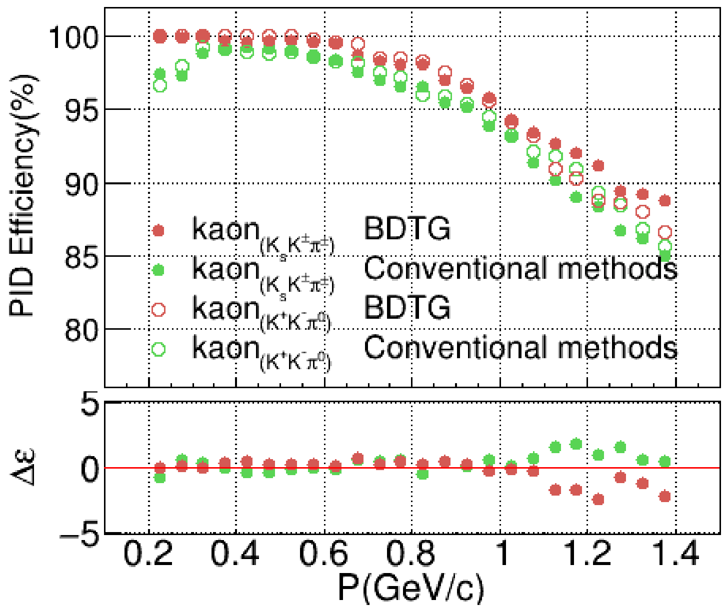
More Check —— cross-validation

real data

$J/\psi \rightarrow K^+K^-\pi^0$ 0.82M

Same data processing steps as before





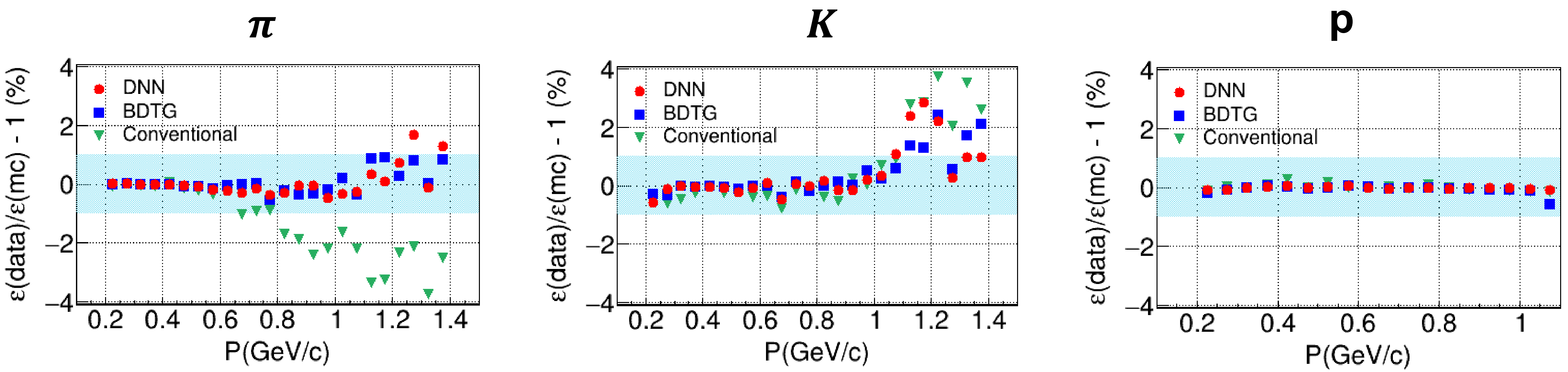
The performance of the model is independent of the physical process

More Check —— Systematic error vs momentum

- If systematic error increase, the gain does not outweigh the loss
- Fortunately, systematic error of ML compared to Conventional Method is reduced

Systematic error： $\Delta\varepsilon = \frac{\varepsilon(data)-\varepsilon(MC)}{\varepsilon(MC)}$

- ε： PID eff



Now we can say: The DNN and BDTG models hold application value

Deployment

In BESIII Offline Software System (BOSS)

DnnPID

\# GetDetectorInfo()
\# TMVACalculation()
\# Prob(5)

Read Infroamtion

Calculate probability of π, K, p

return

Calling method

Rhopi.cxx

```
ParticleID *pid = ParticleID::instance();
for(int i = 0; i < nGood; i++) {
        EvtRecTrackIterator itTrk = evtRecTrkCol->begin() + iGood[i];
        pid->init();
         …
        pid->usePidSys(pid->useDedx() | pid->useTofCorr()); // use PID sub-system
        // pid->usePidSys(pid->useDnn()); // use Deep Neural Networks model
        // pid->usePidSys(pid->useBdt());  // use Gradient Boosting Decision Treemodel
        pid->identify(pid->onlyPionKaonProton()); // seperater Pion/Kaon/Proton
        …
        pid->calculate();
        if((pid->probPion() < pid->probProton()) || (pid->probPion() < pid->probKaon())) continue;
        …
}
```

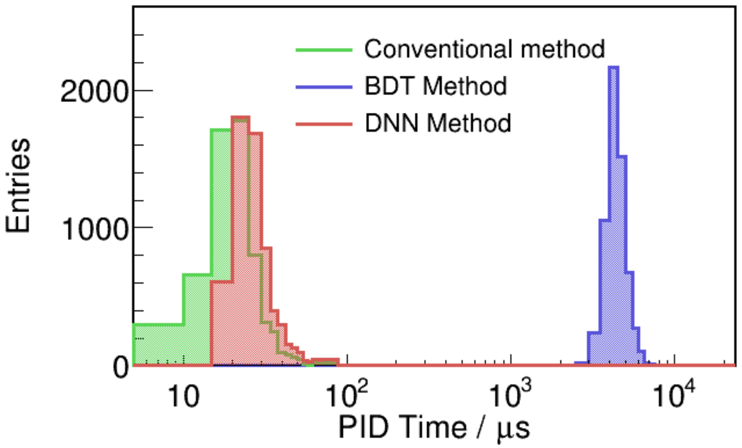
Time consumption

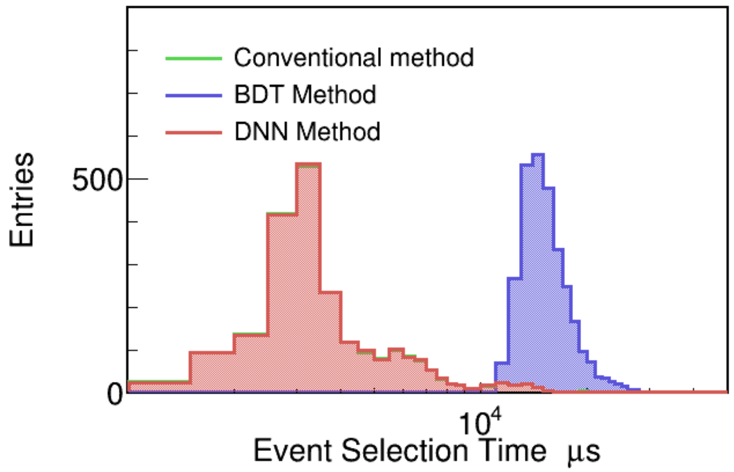
In practical applications, time consumption is a critical concern that determines the feasibility of implementation, impact on Resource Allocation and User Experience

$J/\psi \to \pi^+\pi^-\pi^0$

Event Selection:

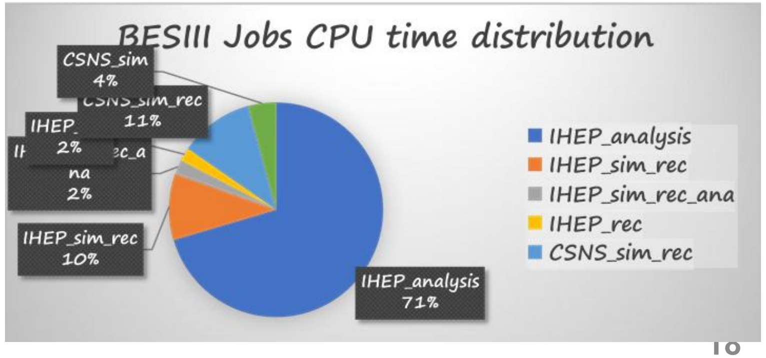
- Charged track
- Photon Selection
- PID
- 1C / 4C kinematic fit





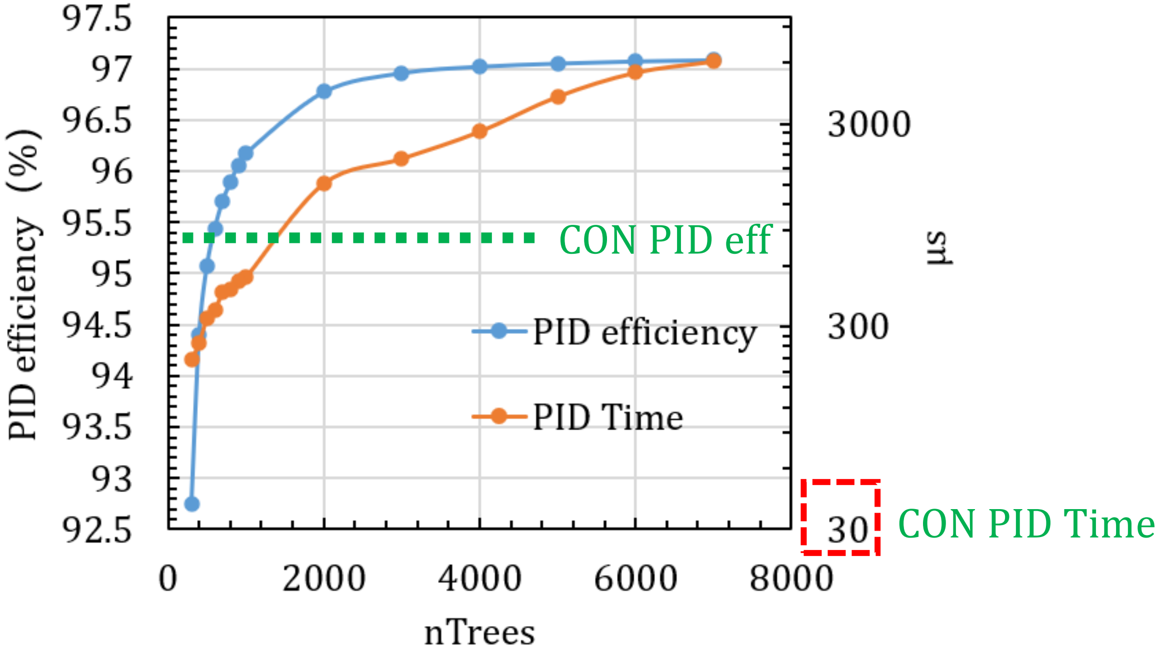
Time consumption (CPU)

- DNN vs CON ~ 1 : 1
- BDTG vs CON： Significant Computational Resource Demands
  - several hundred times more time for PID
  - five times more time for event selection

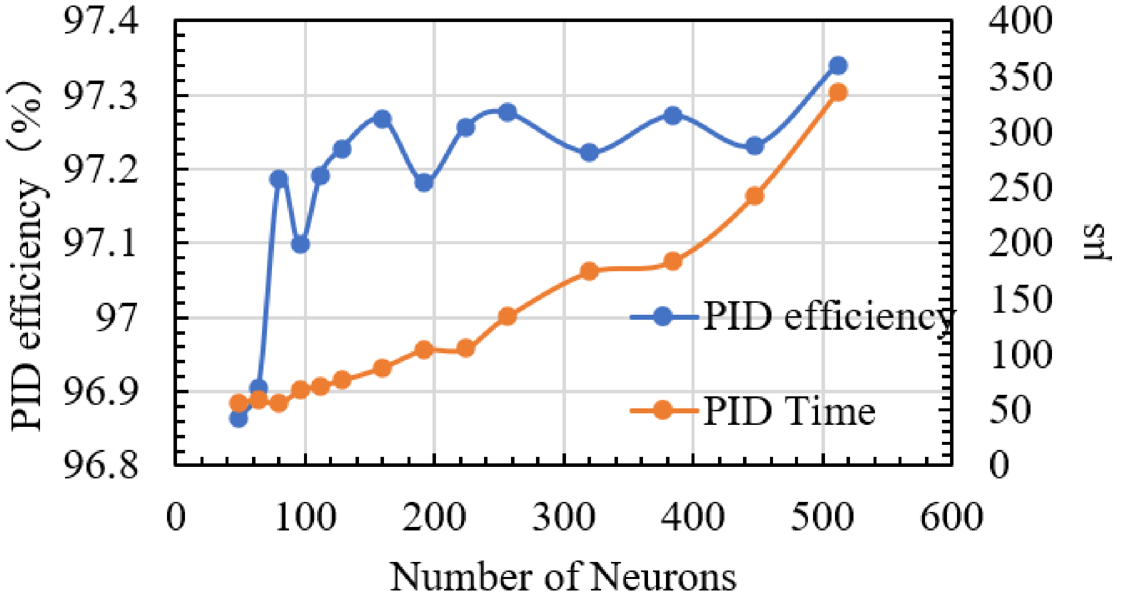


Time consumption

BDTG



DNN



As nTrees decreases,
PID efficiency sharply declines,
but time consumption remains high

Increasing the number of neurons did not significantly enhance PID efficiency.
The optimal number of neurons was selected by balancing time consumption and PID performance

Application

J/ψ incMC

Signal: K^-K^{*+}

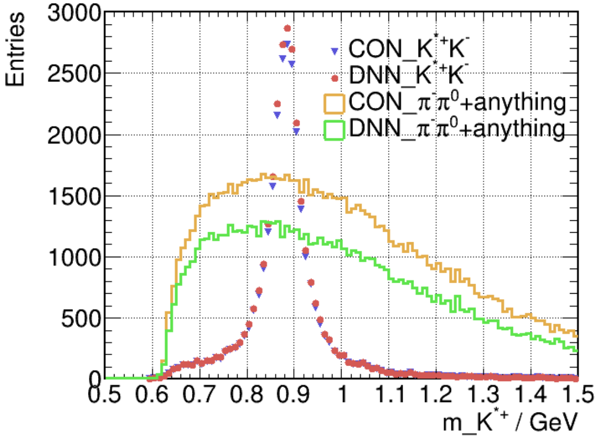
Background: $\pi^-\pi^0 + anything$

PID: $prob_{K-} > prob_{\pi,P}$

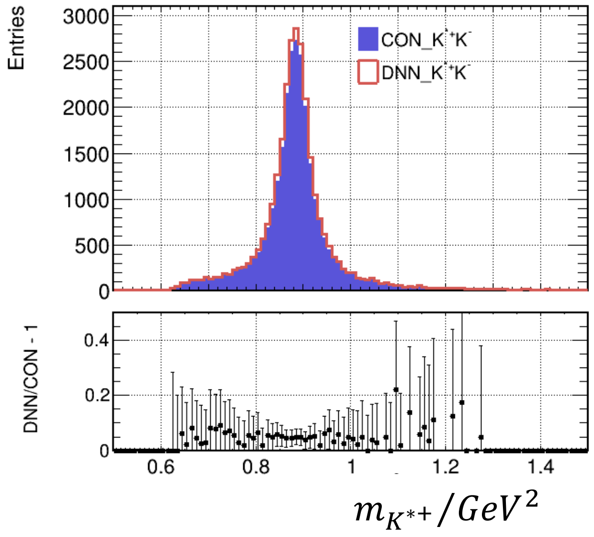
method	Events before PID	PID (0.5<p<1.5) π_0 Cut	$s/\sqrt{s+b}$ (0.5<$m_{K^{*+}}$<1.2)
CON	$K^{*+}K^-$: 37361	27429 (73.42%)	77.98
	$\pi^-\pi^0 + anything$: 3458150	96284 (2.78%)	
DNN	$K^{*+}K^-$: 37361	28704 (76.83%)	90.54
	$\pi^-\pi^0 + anything$: 3458150	71802 (2.08%)	

~16%

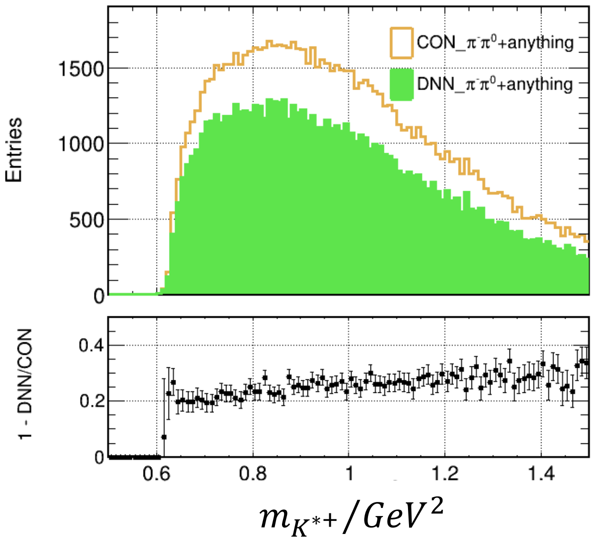
$m_{K^{*+}}$ after PID



Signal after PID



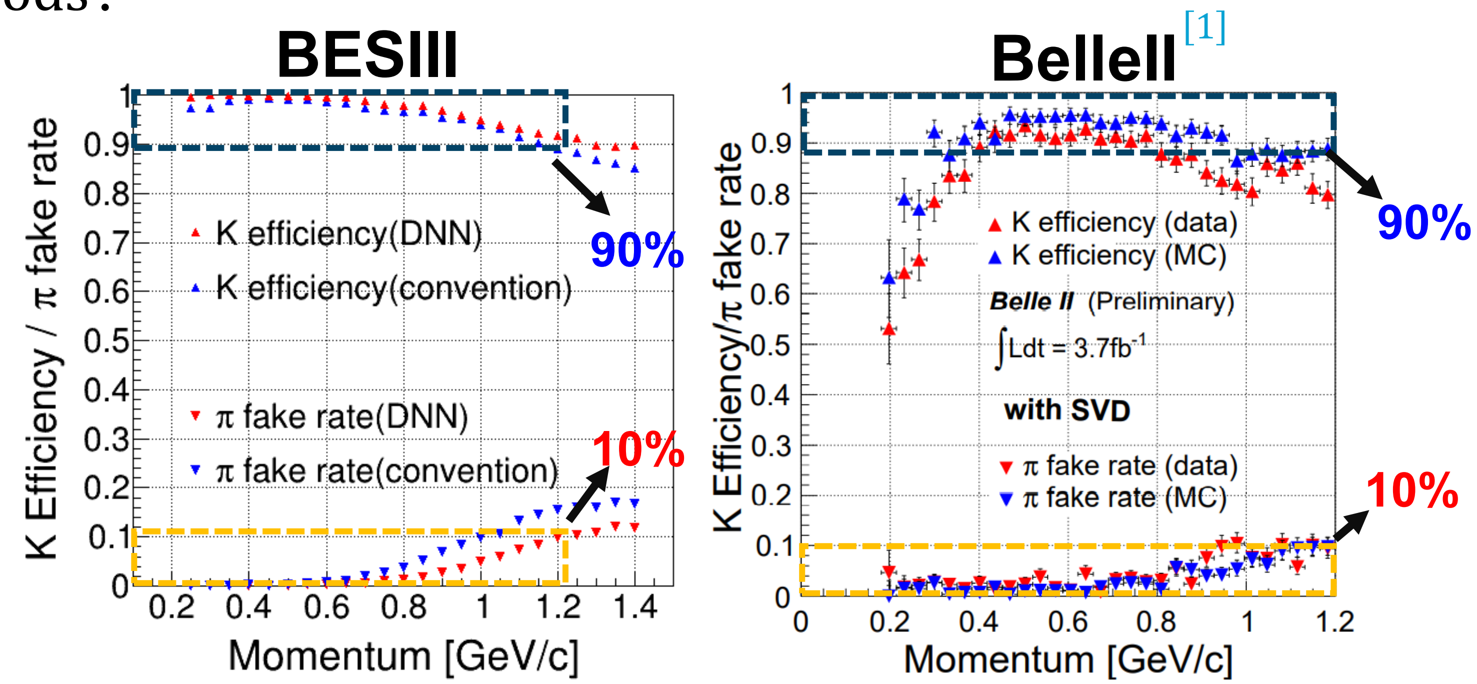
background after PID



Signal-to-background ratios is improved

Summary

- By leveraging machine learning techniques to efficiently integrate hadron identification informations from the four sub-detectors of the BESIII experiment, we have developed the new fast and precise particle identification algorithm that effectively enhances signal significance and physical precision
- PID based on ML vs Conventional methods：
  - PID eff: significant Improvements
    - π~7%@1.2GeV/c
    - K~3%@1.2GeV/c
  - Sys err: Noticeable Improvements
    - π , K: overall Less Than 1%
    - p: Keep it small
  - Time consumption:
    - DNN / CON ~ 1
    - BDTG / CON : unacceptable

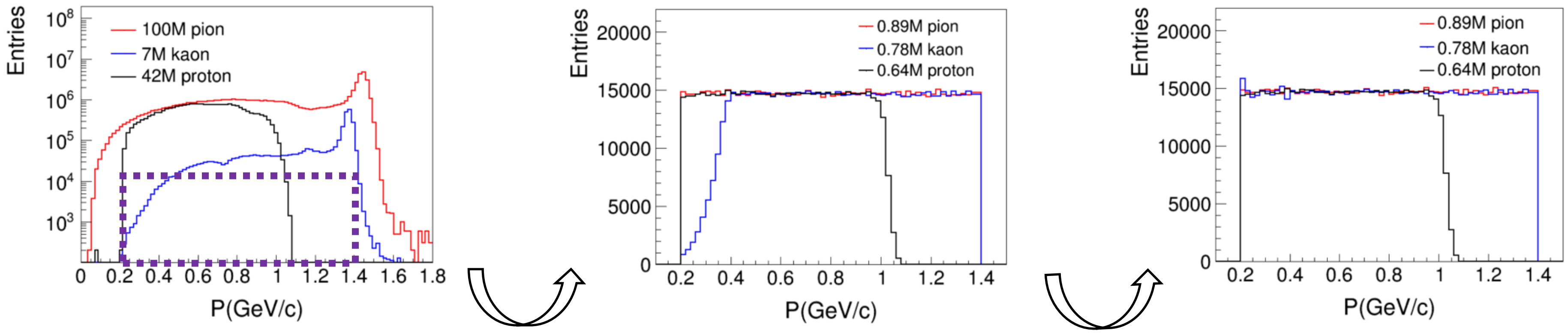


Thank you for your attention!

References

［1］ S. Hazra, et al., Particle Identification in Belle II Silicon Vertex Detector, JPS Conf. Proc. 34 (2021) 01001

Data Processing



- K meson