Particle Identification Algorithms Based on Machine Learning for STCF

Yuncong Zhai[1], Zhipeng Yao[1], Xiaoshuai Qin[1], Teng Li[1], Xingtao Huang[1]

山东大学(青岛) SHANDONG UNIVERSITY QINGDAO

1. Shandong University, Qingdao, China.

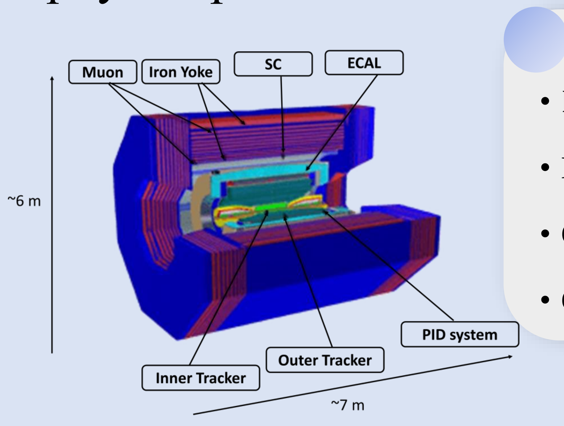
Author Email： zhaiyc@mail.sdu.edu.cn

前沿交叉科学青岛研究院

Introduction

*** The Super Tau Charm Facility (STCF):**

- **STCF** is the next generation positron-electron collider in China, designed specifically to explore various physics phenomena in the τ-charm energy region.



Schematic layout of the STCF detector concept.

Parameters of STCF

- E_{cm} = 2-7 GeV
- L = 0.5×10^{35} cm^{-2} s^{-1}
- Circumference: Double-ring, 600-800 m
- Crossing angle: 2×30mrad

Physics Objectives

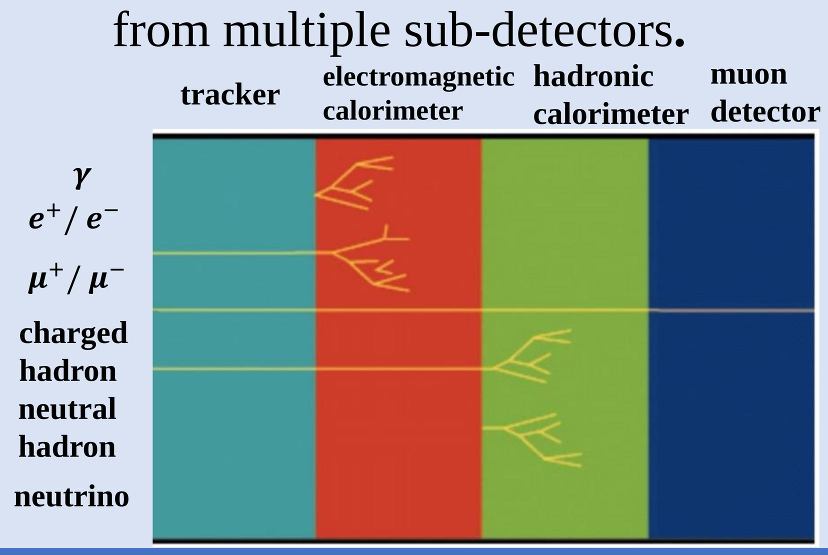
- Rich physics with c quark and τ leptons
- Studying flavor physics and CP violation physics
- Non-perturbed strong interaction and new exotic hadronic states
- Searching for new physics

*** Particle identification (PID):**

•**Particle identification (PID)** is one of the most important and commonly used tools for the physics analysis in STCF.

•The PID algorithm performance is crutial for exploiting the potential of STCF detectors.

•Better particle identification usually requires the combination of information from multiple sub-detectors.



- *π/K (K/p) 3-4σ separation up to 2Gev/c*
- *μ/π up to 2Gev/c, π suppression ~ 3%*
- *Good discrimination power for γ/n/K_L^0*

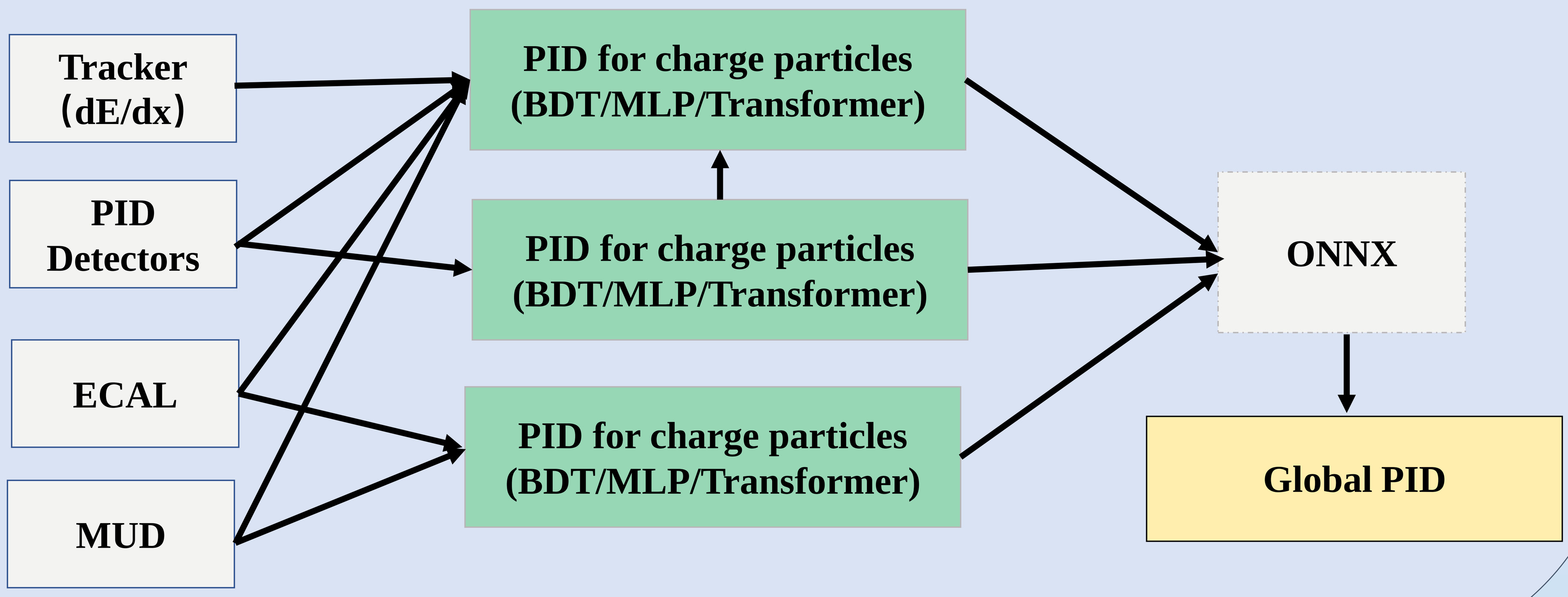
*** Particle Identification Algorithms Based on Machine Learning :**

•The data-driven **machine learning (ML)** approach has opened up new avenues for improving PID performance, with strong modeling and generalization capabilities.

•ML excels at integrating information from all sub-detectors and performing "intelligent fusion" of trajectory information.

•Innovated and developed a Global Particle Identification (GlobalPID) software algorithm based on the ML techniques.

- Identification for charged particles (e/μ/π/K/p) :Taking BDT (based on XGBoost) as a baseline model
- Particle identification of hadrons based on Convolutional Neural Networks (CNN).
- Neutral particle(γ/K_L^0/n) identification based on CNN.



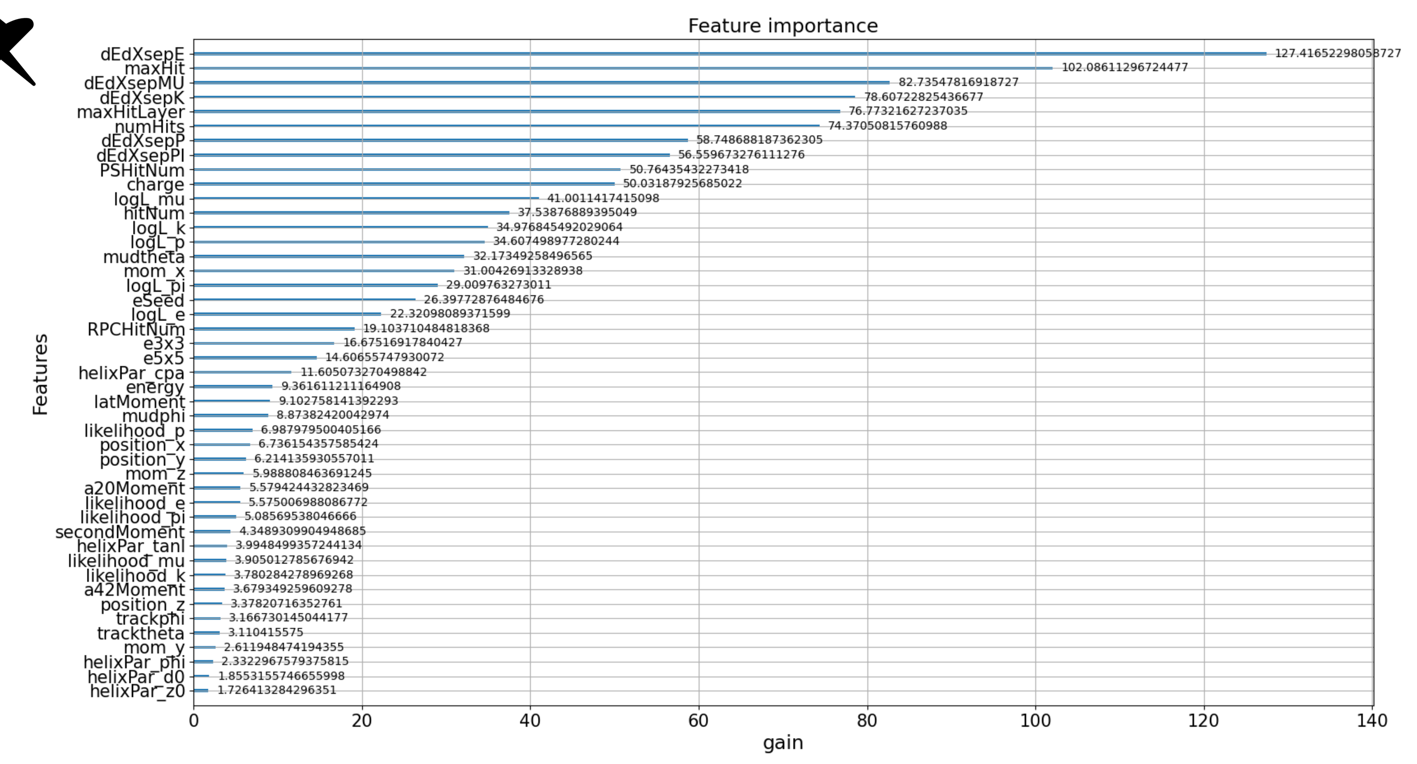
Identification for charged particles

*** Data Sample:**

- Selecting a subset of the most informative features from large amount of interrelated sub-detectors information can help stabilize the model training process
  - MC single charged track using ParticleGun
  - 50000 tracks for each type (e±, μ±, π±, K±, p±)
  - p∈(0.2, 2.4)Gev/c, θ∈(0°, 180°), phi ∈(0°, 360°)
  - Pre-processing: Flatten momentum and θ spectrum to avoid bias due to p/θ distribution
  - 45 features are kept

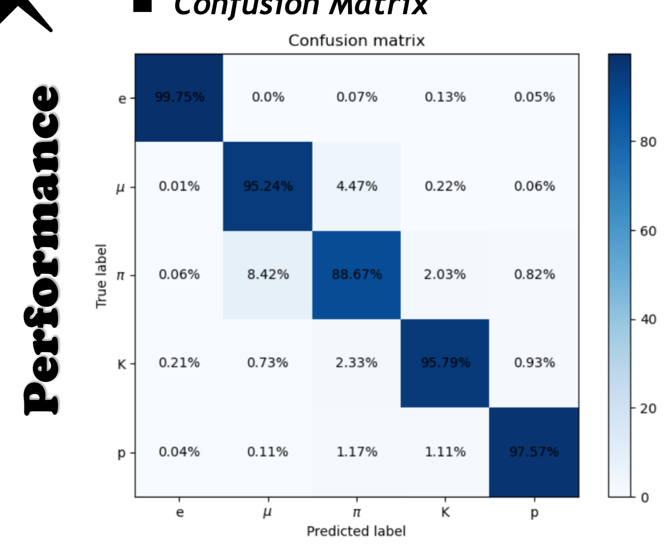
*** Training and Tuning:**

- Target: automated optimization of BDT hyperparameters
  - Reduce manual intervention and time costs
  - Improve model efficiency and reliability
- Selected hyperparameters
  - max_depth: 7
  - n_estimators: 800
- Optimal hyperparameters are obtained based on GridSearchCV



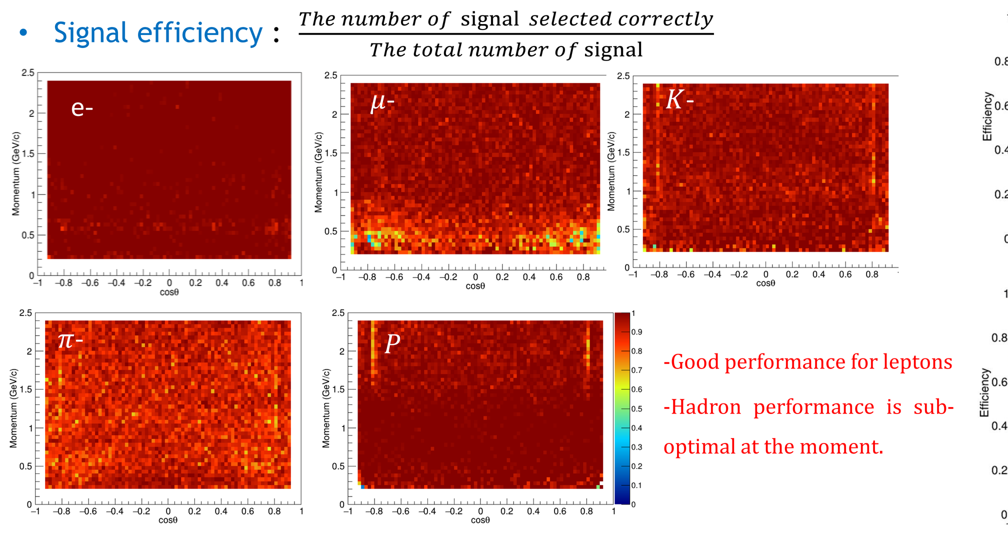
Performance

■ *Confusion Matrix*

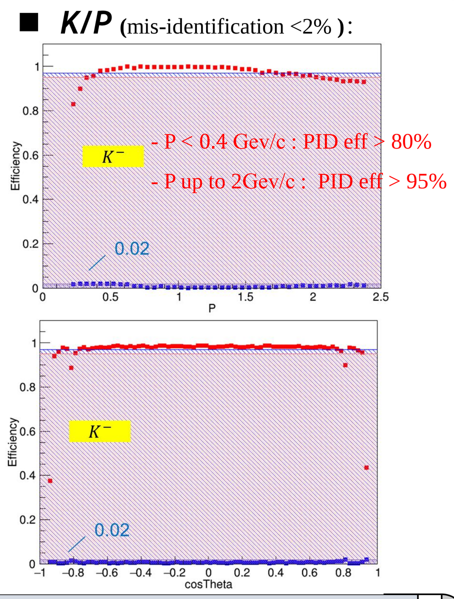


■ *Charged Particle Identification Performance In GlobalPID*

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



■ *K/P* (mis-identification <2%):



Identification of charged hadrons

*** Data Sample:**

- 2D pixel map (200*216)

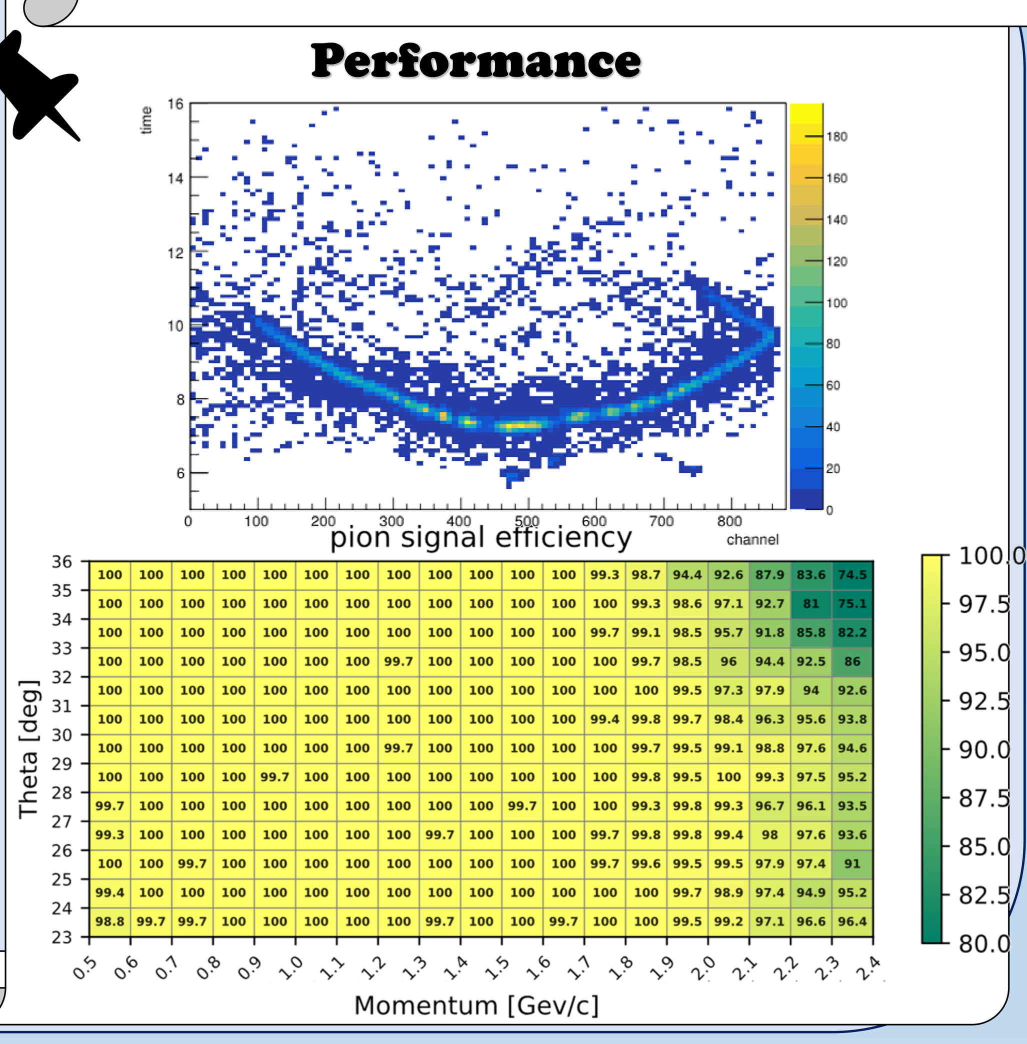
- X-label: the hit position of Cherenkov photons
- Y-label: the arrival time of Cherenkov photons
- Value: the number of photons within this bin
- momentum and position as additional features

- Charged hadrons Data Sample：

- Pion/Kaon
- Generated by ParticleGun
- p ∈ (0.5, 2.4) Gev/c, θ ∈ (23°, 36°), φ ∈ (0°, 180°)

- Using EfficientNetV2-S as the baseline model

Performance



Neutral particle identification

*** Neutral particle identification needs to consider the energy deposition, time response and MUD hitting information of ECAL**

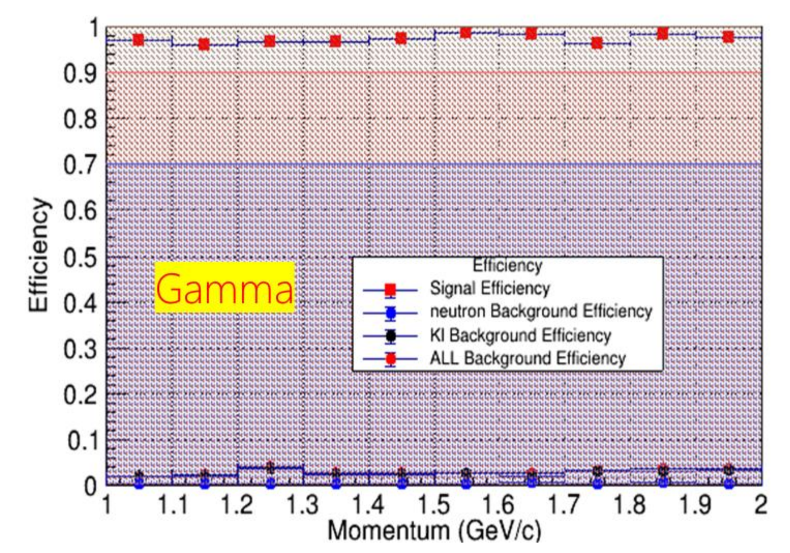
- The energy deposition information of the electromagnetic calorimeter is projected onto a two-dimensional plane
- The MUD hitting information is used as additional input
- CNN is used for image classification
- Neutral Particle Data Sample：
  - γ/K_L/n
  - Generated by ParticleGun
  - 100,000(Each type)
  - P ∈ （0，2.0）Gev/c, θ = 90°, φ = 0°

***The initial implementation of a global neutral particle discriminator based on CNN**

Performance

■ *Gamma*:

- *Good photon discrimination performance*
- *Signal efficiency > 90%*



Reference

1. M. Achasov, V. Blinov, A. V. Bobrov, D. A. Bodrov, A. E. Bondar, S. D. Vorobiev, D. S. Gorbunov,V. P. Druzhinin, D. A. Epifanov, A. S. Kuzmin et al., "Experiments at the super charm-tau factory," Uspekhi Fizicheskikh Nauk, vol. 194, no. 1, pp. 60–76, 2024.
2. L. A. Badulescu, "Attribute selection measure in decision tree growing," Proceedings of Annals of University of Craiova, 2007.
3. A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," Journal of Chemometrics: A Journal of the Chemometrics Society, vol. 18, no. 6, pp. 275–285, 2004.
4. J. S. Kushwah, A. Kumar, S. Patel, R. Soni, A. Gawande, and S. Gupta, "Comparative study of regressor and classifier with decision tree using modern tools," Materials Today: Proceedings, vol. 56, pp. 3571–3576, 2022.
5. T.-H. Lee, A. Ullah, and R. Wang, "Bootstrap aggregating and random forest," Macroeconomic forecasting in the era of big data: Theory and practice, pp. 389–429, 2020.
6. I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: a review," Procedia Manufacturing, vol. 35, pp. 698–703, 2019.
7. L. Breiman, Classification and regression trees. Routledge, 2017.
8. R. E. Schapire et al., "A brief introduction to boosting," in Ijcai, vol. 99, no. 999. Citeseer, 1999, pp. 1401–1406.