Particle Identification Algorithms Based on Machine Learning for STCF

PoS(ICHEP2024)847

Yuncong Zhai[a,*] **and Zhipeng Yao, Teng Li, Xingtao Huan**[a,b]

[a,b] *Institute of Frontier and Interdisciplinary Science, Shandong University,*
72 Binhai Road, Qingdao, China
E-mail: zhaiyc@mail.sdu.edu.cn, yaozp@mail.sdu.edu.cn, tengli@sdu.edu.cn, huangxt@sdu.edu.cn

The Super Tau-Charm Facility (STCF)[1] is one of China's most advanced positron-electron colliders in the future, with a peak luminosity of $0.5 \times 10^{35} cm^{-2} s^{-1}$ and a center-of-mass energy of $2 \sim 7$ GeV, designed specifically to explore various physics phenomena in the τ-charm energy region. Particle identification (PID) is a crucial element in physics analysis and is vital for achieving exceptional scientific performance. STCF places high demands on PID accuracy and efficiency to meet its stringent standards. In recent decades, machine learning (ML) techniques have emerged as a dominant methodology for PID in high-energy physics experiments, consistently delivering superior results. This study introduces an advanced PID software based on ML algorithms, developed for STCF to advance physics research. It includes a comprehensive global PID algorithm for charged particles, combining information from all sub-detectors, as well as a deep convolutional neural network (CNN) that utilizes Cherenkov detector inputs to effectively distinguish charged hadrons. Additionally, a CNN has been developed to differentiate neutral particles using calorimeter responses.Initial findings indicate that these PID models have demonstrated exceptional performance, significantly enhancing the scientific potential of STCF.

42nd International Conference on High Energy Physics (ICHEP2024)
18-24 July 2024
Prague, Czech Republic

[*] Speaker

1. Introduction

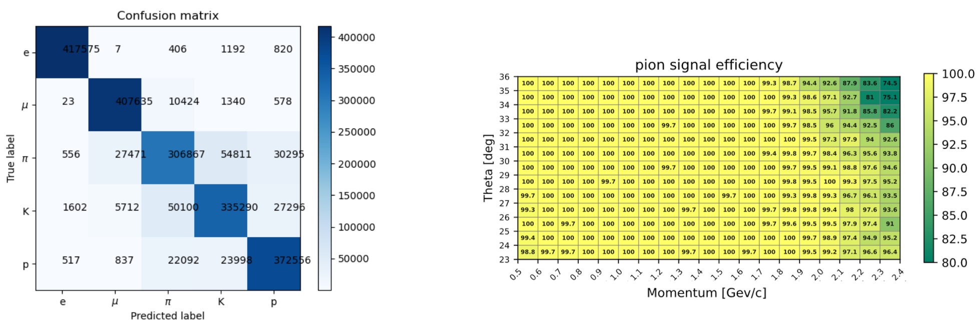
The STCF experiment focuses on precision measurements of CKM elements, CP violation studies, and so on.[2] Excellent PID performance is crucial for various physics studies at STCF, requiring the integration of multiple PID systems to cover different momentum ranges. While traditional PID algorithms like maximum likelihood methods have shown success in particle physics experiments, they struggle with combining PID information from multiple sub-detector systems. Internationally, experiments like LHCb, CMS, and others [3–5] have adopted PID algorithms based on ML such as BDT and neural networks, consistently outperforming traditional methods. Thus, the robust PID software developed in this paper with ML to boost the physics research at STCF becomes a natural choice.

2. Global PID for Charged Particles

To achieve excellent PID performance over charged particles ($e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, $p^{\pm}$), a GlobalPID algorithm for charged particles using BDT is developed using the Monte Carlo simulation data based on the STCF offline software system OSCAR.[6]. The training dataset included 500,000 events covering five particle types with momentum in the range [0.2, 2.4] GeV/c and theta angles between 20-160 degrees. Optimal hyperparameters for the BDT model were determined through cross-validation, resulting in max_{depth} of 7 and $n_{estimators}$ of 800. Figure 1b showed excellent discrimination performance for leptons, with an identification efficiency exceeding 90%. Further optimization is suggested for hadron performance, potentially involving a deep learning-based PID algorithm on the Cherenkov detectors.

3. CNN-based PID for Charged Hadrons

We used OSCAR to simulate a dataset containing 3 million pion/kaon samples. By collecting the arrival time and hit position of Cherenkov photons using photomultiplier, a 200*217 time-space pixel map can be constructed.



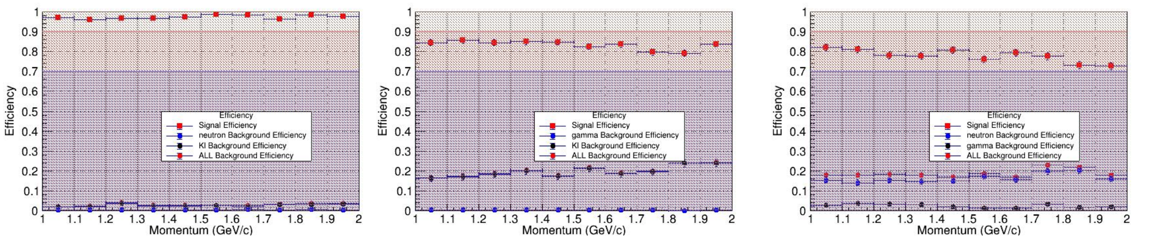
(a) (b)

Figure 1: (a) Confusion matrix (b) The signal efficiency of pions with a background misidentification rate not exceeding 2%.

We adopted the EfficientNetV2-S[7] architecture as the base model and further optimized it by adding three-dimensional momentum and position extrapolated from the tracking system to DTOF in the fully connected layer. Figure 1a shows the signal efficiency and background misidentification rate distribution of pion samples at different momenta and polar angles. The misidentification rate of kaons as background is controlled to be no more than 2%. It can be seen that the model achieves a pion discrimination efficiency of 99% in most of the phase space, but there is still room for further optimization in the high momentum and large angle regions.

4. CNN-based PID for Neutral Particles

In the context of the τ-charm facility, discrimination between these neutral particles (such as photon, neutron and K_L^0) is crucial. The electromagnetic shower shapes in the electromagnetic calorimeter (ECAL) and characteristics of hits in the muon detector are key variables for neutral particle identification. A 71×136 Energy Deposition Pixel Map was utilized as input for a classical CNN architecture designed to differentiate between these neutral particles. While the CNN model showed strong gamma discrimination efficiency exceeding 90%, it exhibited suboptimal performance in distinguishing neutrons and K_L^0. Future research plans involve incorporating hit time information from the ECAL and hit characteristics from the muon detector to improve the model's discriminatory capabilities.



(a) **(b)** **(c)**

Figure 2: The signal efficiency and background efficiency for neutral particles as a function of momentum. (a) Gamma (b) Neutron (c) K_L^0

5. Conclusion And Outlook

A PID software based on ML techniques is developed in this study, showing promising results in discriminating charged and neutral particles using global PID and CNN models, with future work focusing on advanced model implementation, addressing data discrepancies, and optimizing models for specific physics requirements.

References

[1] M. Achasov, V. Blinov, A. V. Bobrov, D. A. Bodrov, A. E. Bondar, V. S. 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] M. Achasov, X. C. Ai, R. Aliberti, L. P. An, Q. An, X. Z. Bai, Y. Bai, O. Bakina, A. Barnyakov and V. Blinov, *et al.* Front. Phys. (Beijing) **19** (2024) no.1, 14701 doi:10.1007/s11467-023-1333-z [arXiv:2303.15790 [hep-ex]].

[3] F. J. P. Soler [LHCb], Eur. Phys. J. direct **4** (2002) no.S1, 34 doi:10.1007/s1010502cs134

[4] A. K. Yadav, R. Mukherjee, S. Chattopadhyay and P. P. Bhaduri, DAE Symp. Nucl. Phys. **67**, 1017-1018 (2024)

[5] M. Karwowska *et al.* [ALICE], EPJ Web Conf. **295**, 09029 (2024) doi:10.1051/epjconf/202429509029 [arXiv:2309.07768 [hep-ex]].

[6] T. Li, W. Huang, X. Huang, X. Ai, H. Li and D. Liu, EPJ Web Conf. **295**, 03025 (2024) doi:10.1051/epjconf/202429503025

[7] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International conference on machine learning*. PMLR, 2021, pp. 10 096–10 106.