

DATA SHIFT PROBLEM IN PARTICLE IDENTIFICATION

V. Papoyan^{1,2,*}

¹ Joint Institute for Nuclear Research, Dubna, Russia

² Alikhanyan National Science Laboratory, Yerevan, Armenia

Machine learning (ML) methods are increasingly applied to particle identification. Since experimental data lack ground-truth labels, classifiers are trained on Monte Carlo (MC) simulations, which leads to the problem of data shift — distributional differences between simulated and real data. The study analyzes the impact of data shift by comparing particle classification across several MC datasets with different configurations, highlighting the importance of validating and adapting ML models for robust performance.

Методы машинного обучения все шире применяются в задаче идентификации частиц. Поскольку экспериментальные данные не содержат истинных меток, обучение классификаторов проводится на моделированных методом Монте-Карло (МК) данных, что приводит к проблеме их сдвига — различиям в распределениях признаков между моделированными и реальными данными. Исследовано влияние сдвига данных путем сравнения классификации частиц на нескольких наборах МК с различными конфигурациями, что подчеркивает необходимость проверки и адаптации моделей для обеспечения надежной работы.

PACS: 44.25.+f; 44.90.+c

* E-mail: vlpapoyan@jinr.ru