Contribution ID: **9**                                         Type: **not specified**

# Towards more precise data analysis with machine-learning-based particle identification with missing data

*Sunday 15 December 2024 11:45 (15 minutes)*

Identifying products of ultrarelativistic collisions, such as the ones delivered by the LHC and RHIC, is one of the crucial objectives of experiments such as ALICE and STAR, which are specifically dedicated to this task with a number of detectors allowing particle identification (PID) over a broad momentum range.

Recently, as a team of physicists and computer scientists at the Warsaw University of Technology, we have introduced a novel method for Particle Identification (PID) [1,2]. The ALICE experiment was used as R&D and testing environment; however, the proposed solution is general enough for other experiments with good PID capabilities (i.e. already mentioned STAR).

PID methods rely on hand-crafted selections, which compare experimental data to theoretical simulations. To improve the performance of the baseline methods, novel approaches use machine learning models that learn the proper assignment in a classification task. However, because of the various detection techniques used by different subdetectors, as well as the limited detector efficiency and acceptance, produced particles do not always yield signals in all of the ALICE components. This results in data with missing values. Out-of-the-box machine learning solutions cannot be trained with such examples without either modifying the training dataset or re-designing the model architecture.

In the presented work, we propose a new method for PID that addresses these issues and can be trained with all of the available data examples, including incomplete ones. The solution is inspired by a method proposed for medical diagnosis with missing data in patient records. In general, our approach improves the PID purity and efficiency of the selected sample for all investigated particle species (pions, kaons, protons).

[1] Miłosz Kasak, Kamil Deja, Maja Karwowska, Monika Jakubowska, Łukasz Graczykowski & Małgorzata Janik, "Machine-learning-based particle identification with missing data",Eur.Phys.J.C 84 (2024) 7, 691

[2] Maja Karwowska, Łukasz Graczykowski, Kamil Deja, Miłosz Kasak, and Małgorzata Janik, "Particle identification with machine learning from incomplete data in the ALICE experiment", JINST 19 (2024) 07, C07013
Identifying products of ultrarelativistic collisions, such as the ones delivered by the LHC and RHIC, is one of the crucial objectives of experiments such as ALICE and STAR, which are specifically dedicated to this task with a number of detectors allowing particle identification (PID) over a broad momentum range.

Recently, as a team of physicists and computer scientists at the Warsaw University of Technology, we have introduced a novel method for Particle Identification (PID) [1,2]. The ALICE experiment was used as R&D and testing environment; however, the proposed solution is general enough for other experiments with good PID capabilities (i.e. already mentioned STAR).

PID methods rely on hand-crafted selections, which compare experimental data to theoretical simulations. To improve the performance of the baseline methods, novel approaches use machine learning models that learn the proper assignment in a classification task. However, because of the various detection techniques used by different subdetectors, as well as the limited detector efficiency and acceptance, produced particles do not always yield signals in all of the ALICE components. This results in data with missing values. Out-of-the-box machine learning solutions cannot be trained with such examples without either modifying the training dataset or re-designing the model architecture.

In the presented work, we propose a new method for PID that addresses these issues and can be trained with all of the available data examples, including incomplete ones. The solution is inspired by a method proposed for medical diagnosis with missing data in patient records. In general, our approach improves the PID purity and efficiency of the selected sample for all investigated particle species (pions, kaons, protons).

[1] Miłosz Kasak, Kamil Deja, Maja Karwowska, Monika Jakubowska, Łukasz Graczykowski & Małgorzata Janik, "Machine-learning-based particle identification with missing data",Eur.Phys.J.C 84 (2024) 7, 691

[2] Maja Karwowska, Łukasz Graczykowski, Kamil Deja, Miłosz Kasak, and Małgorzata Janik, "Particle identification with machine learning from incomplete data in the ALICE experiment", JINST 19 (2024) 07, C07013

**Authors:** DEJA, Kamil Rafal (Warsaw University of Technology (PL)); GRACZYKOWSKI, Lukasz (Warsaw University of Technology (PL)); KARWOWSKA, Maja (Warsaw University of Technology (PL)); KASAK, Milosz (Warsaw University of Technology (PL)); JAKUBOWSKA, Monika Joanna (Warsaw University of Technology (PL))

**Presenter:** GRACZYKOWSKI, Lukasz (Warsaw University of Technology (PL))

**Session Classification:** Session 3