Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Machine Learning Techniques for Space Calorimeter Experiments

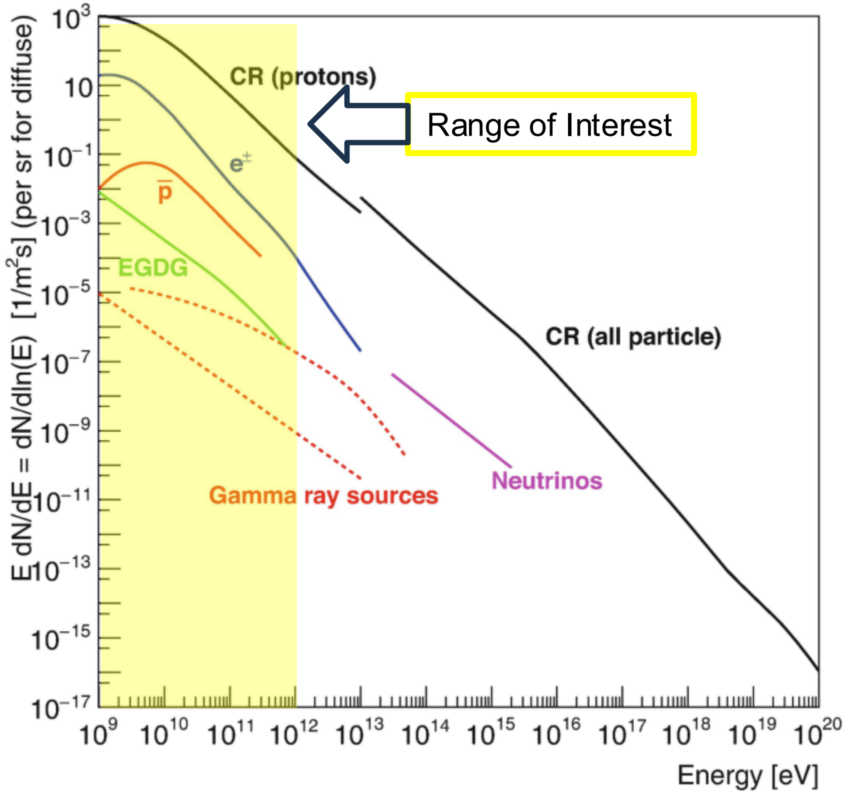
Maria Bossa, Federica Cuna, Fabio Gargano

Referee: Gianpaolo Carlino (INFN Sezione di Napoli), Pasquale Lubrano (INFN Sezione di Perugia)

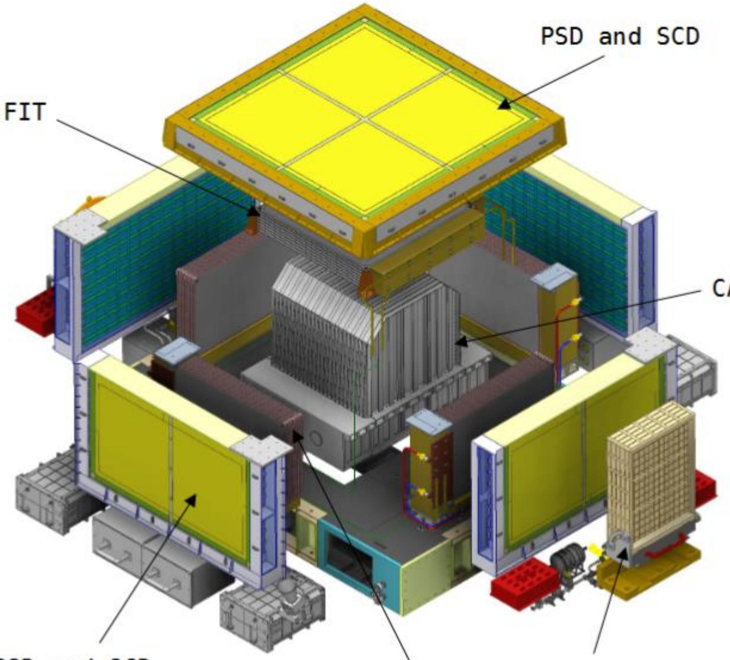
Spoke 3 Technical Meeting, Perugia 26-29/5, 2025

Scientific Rationale

Primary cosmic rays are mostly high-energy protons and nuclei, with a small fraction of electrons (<1%). Measuring the electron component is challenging due to its low abundance—up to 1,000 times less than protons—and requires efficient proton rejection. We present a novel classification approach using machine learning techniques to distinguish electron- and proton-induced events

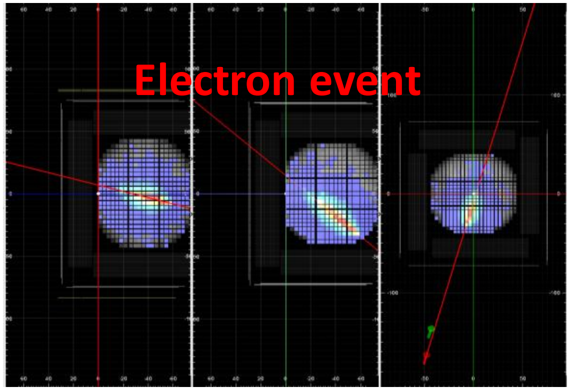


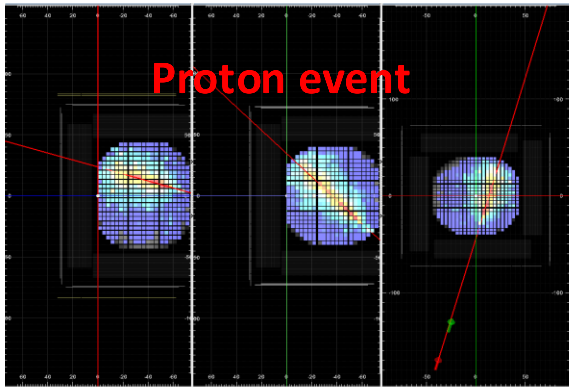
HERD experiment as testbench



- HERD is a future space mission for:
  - the direct detection of cosmic rays up to 1 PeV
  - the measurement of the electron and positron flux up to several tens of TeV

- It consists of homogeneous, isotropic, and deeply segmented 3D calorimeter, surrounded by multiple sub-detectors for charge, timing, and tracking measurements.





Data Simulation

- The full dataset is generated using the custom HerdSoftware simulation framework
- It consists of 4,300,000 events, equally divided into 2,150,000 electron events and 2,150,000 proton events.
- Both events sets are simulated within a power-law energy spectrum E^{-1}, spanning an energy range from 100 GeV to 1 TeV and from 1 GeV to 100 GeV
- Tracks are distributed within a spherical region surrounding the HERD detector.

Choice of features

We tested a new set of features, these include:

- **Lateral moment of the shower until 4° order**

$$M_{lateral}(n) = \frac{\sum_i E_i \cdot {d_i}^n}{\sum_i E_i},$$

Where E_i is the energy deposited in the i-th pixel, d_i is the distance of the i-th activated pixel from a reference axis, which in our case is the shower axis.

- **Longitudinal moment of the shower until 4° order**

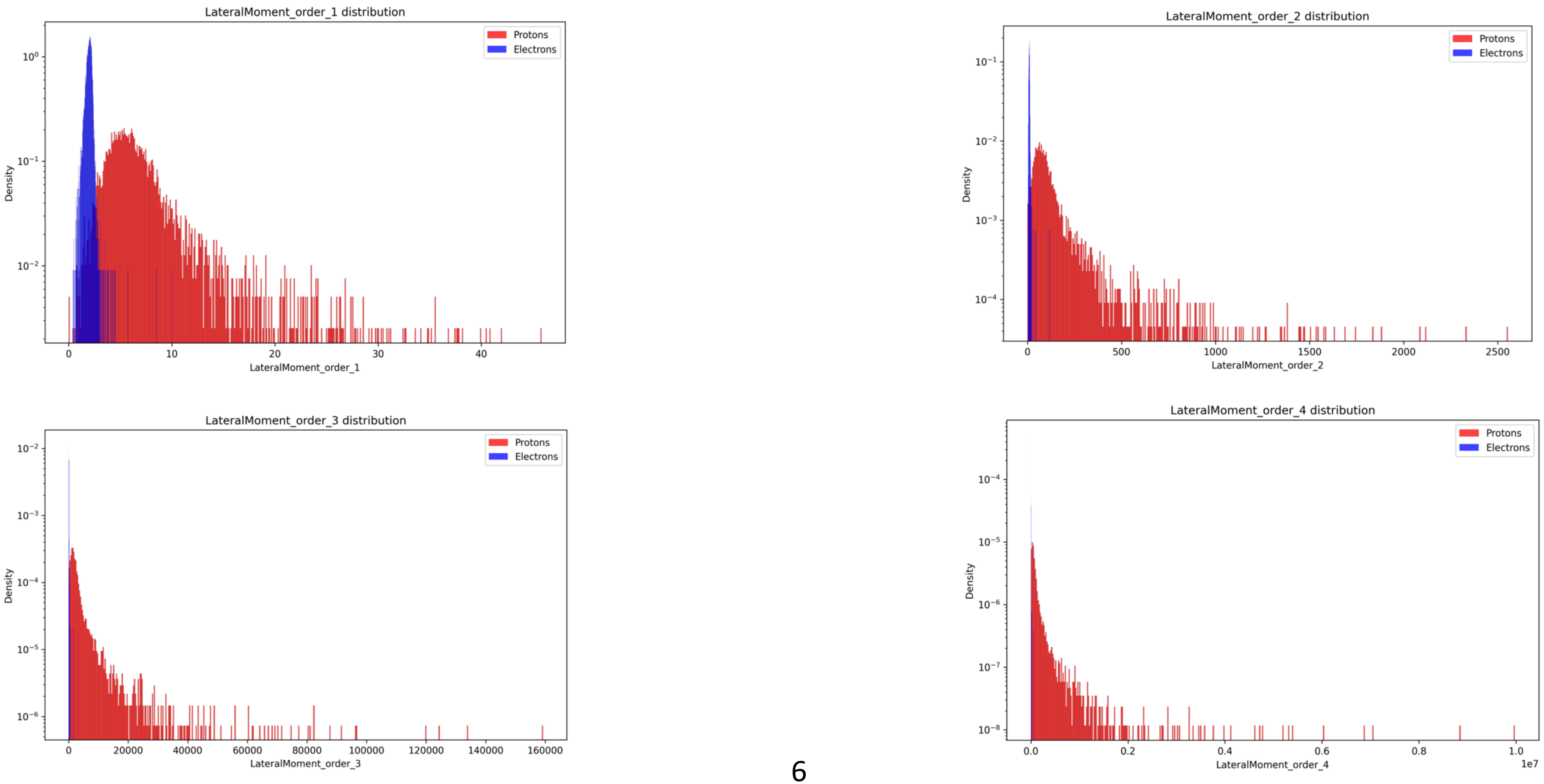
$$M_{long}(n) = \frac{\sum_i E_i \left(\bar{d} \cdot (\bar{x} - x_0)\right)^n}{\sum_i E_i},$$

where $\bar{d}$ is the directional vector, which in our case has been chosen as the primary direction, x_i is the i-th activated pixel, x_0 is a reference point, that corresponds to the starting point of the shower.

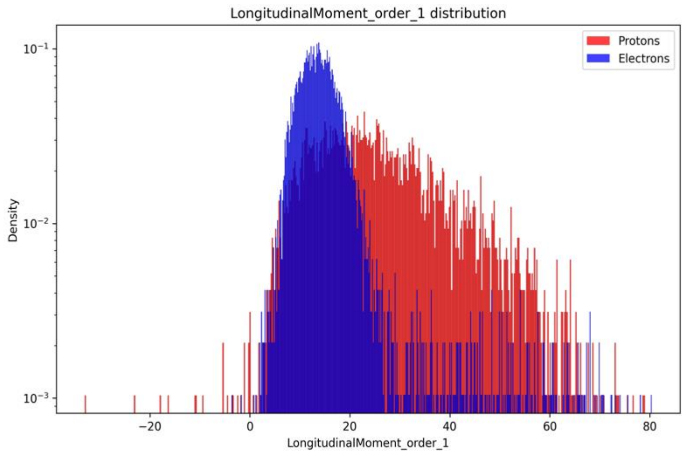
- **Longitudinal profile of the shower**

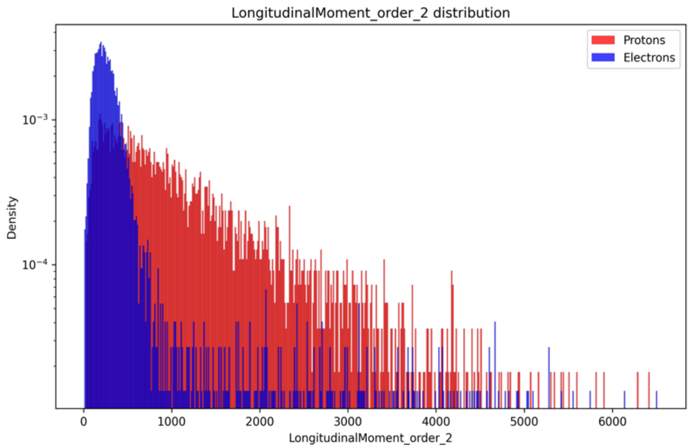
Each hit is projected onto the shower axis. The axis is split into 10 equal-length segments. In each segment, the deposited energy from hits within a fixed radius is summed. The energy is then normalized by the segment length, yielding the energy density along the shower axis.

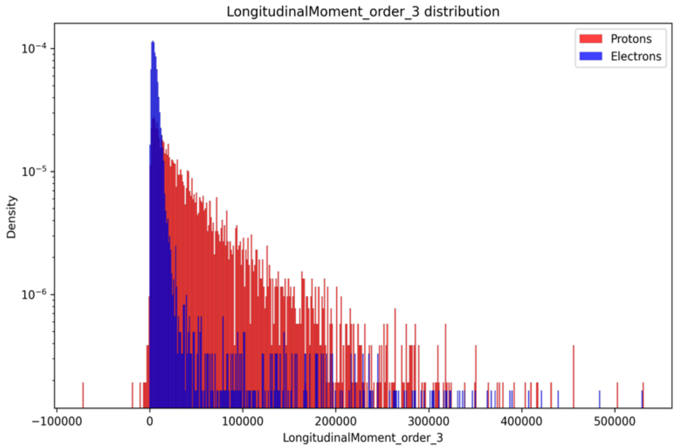
Lateral moments: Energy range 1GeV-100GeV

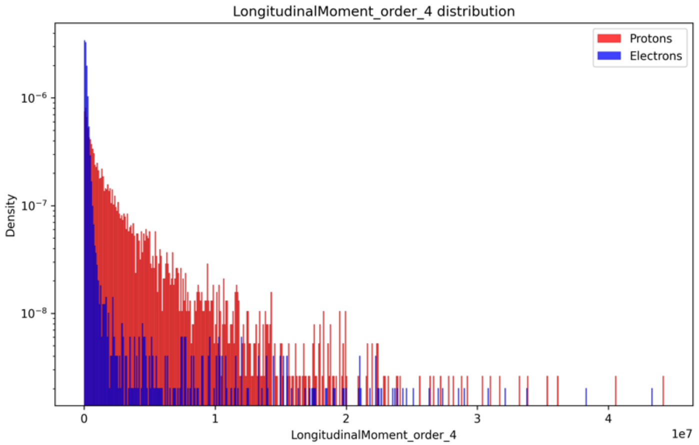


Longitudinal moments – Energy range 1 GeV- 100 GeV

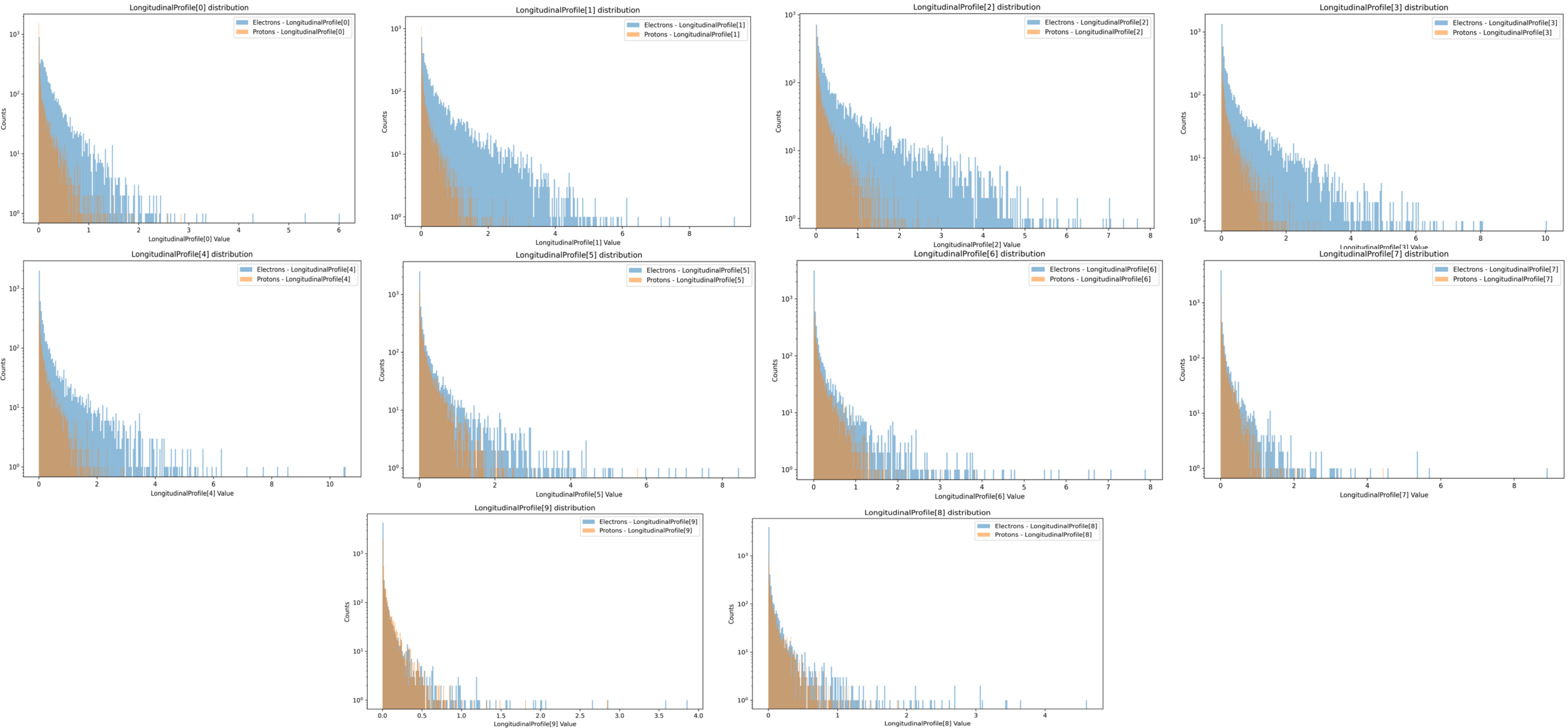




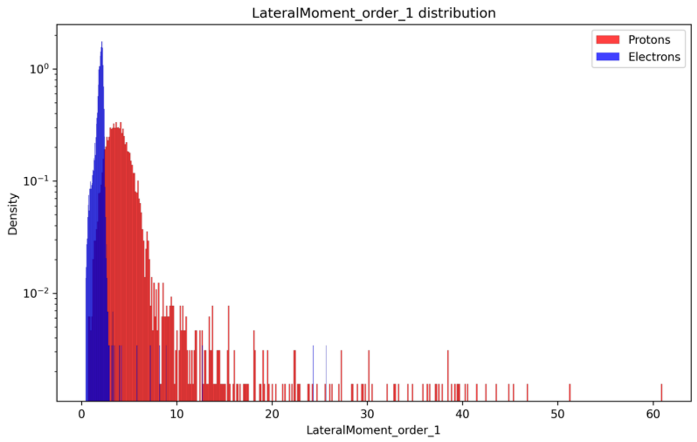


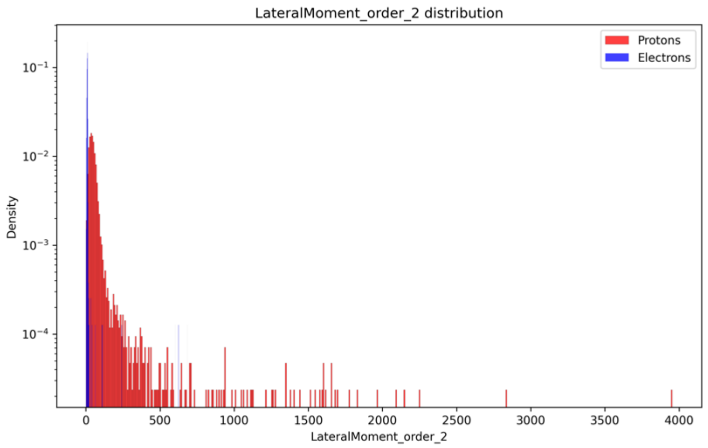


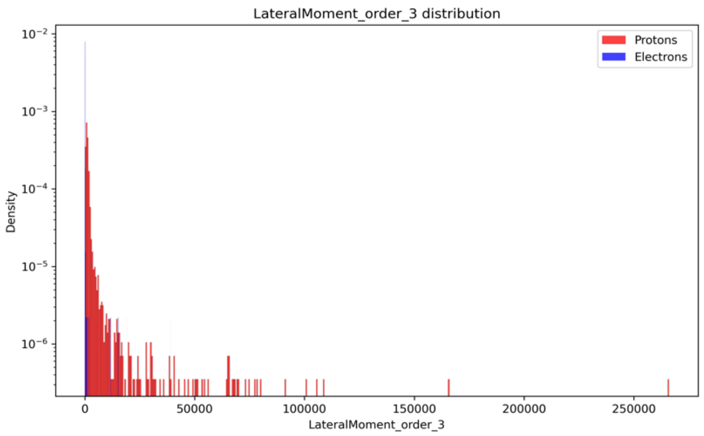
Longitudinal Profile – Energy range 1 GeV- 100 GeV

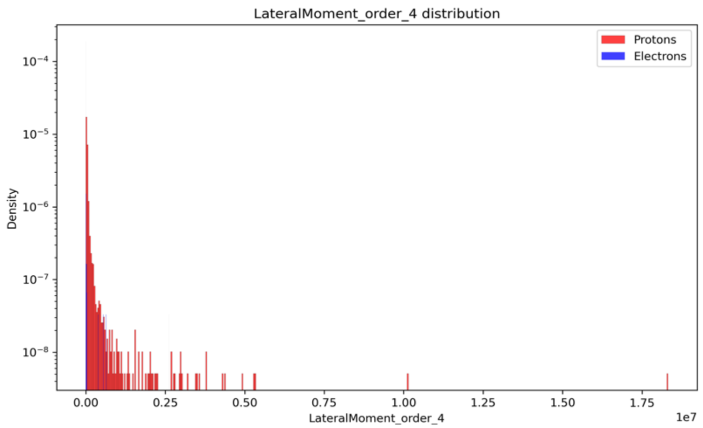


Lateral moments - Energy range 100GeV-1TeV

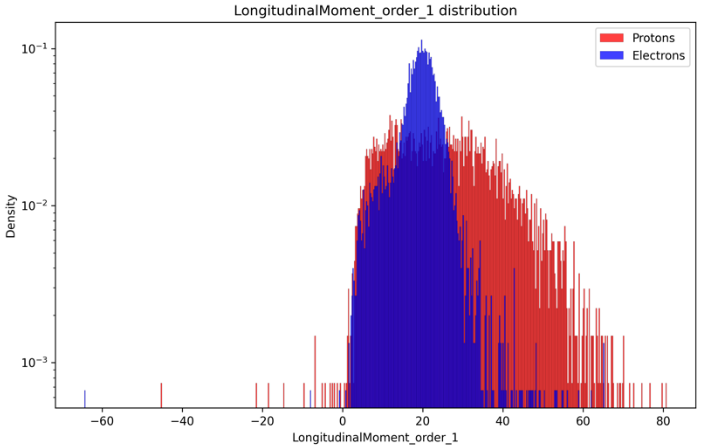


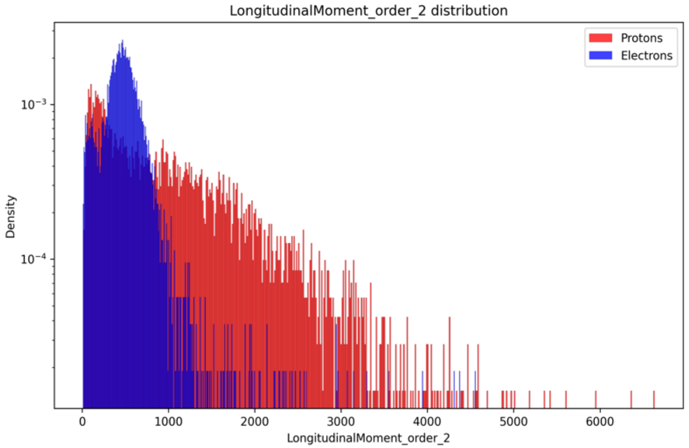


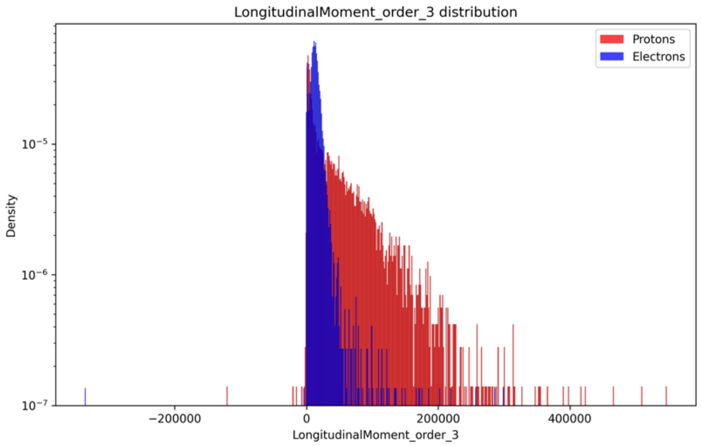


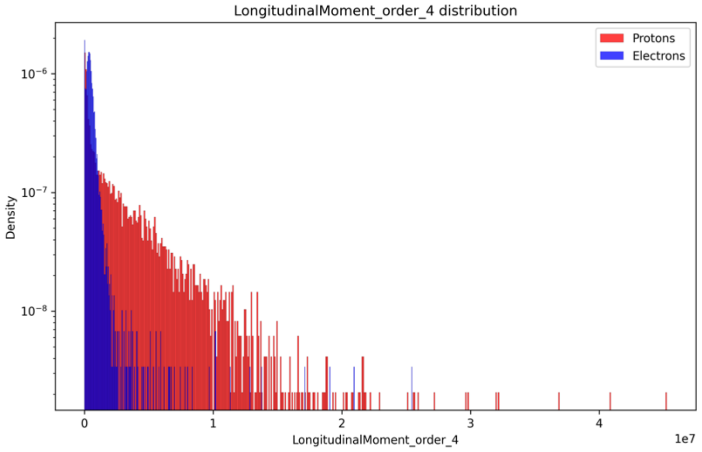


Longitudinal moments – Energy range 100 GeV- 1 TeV



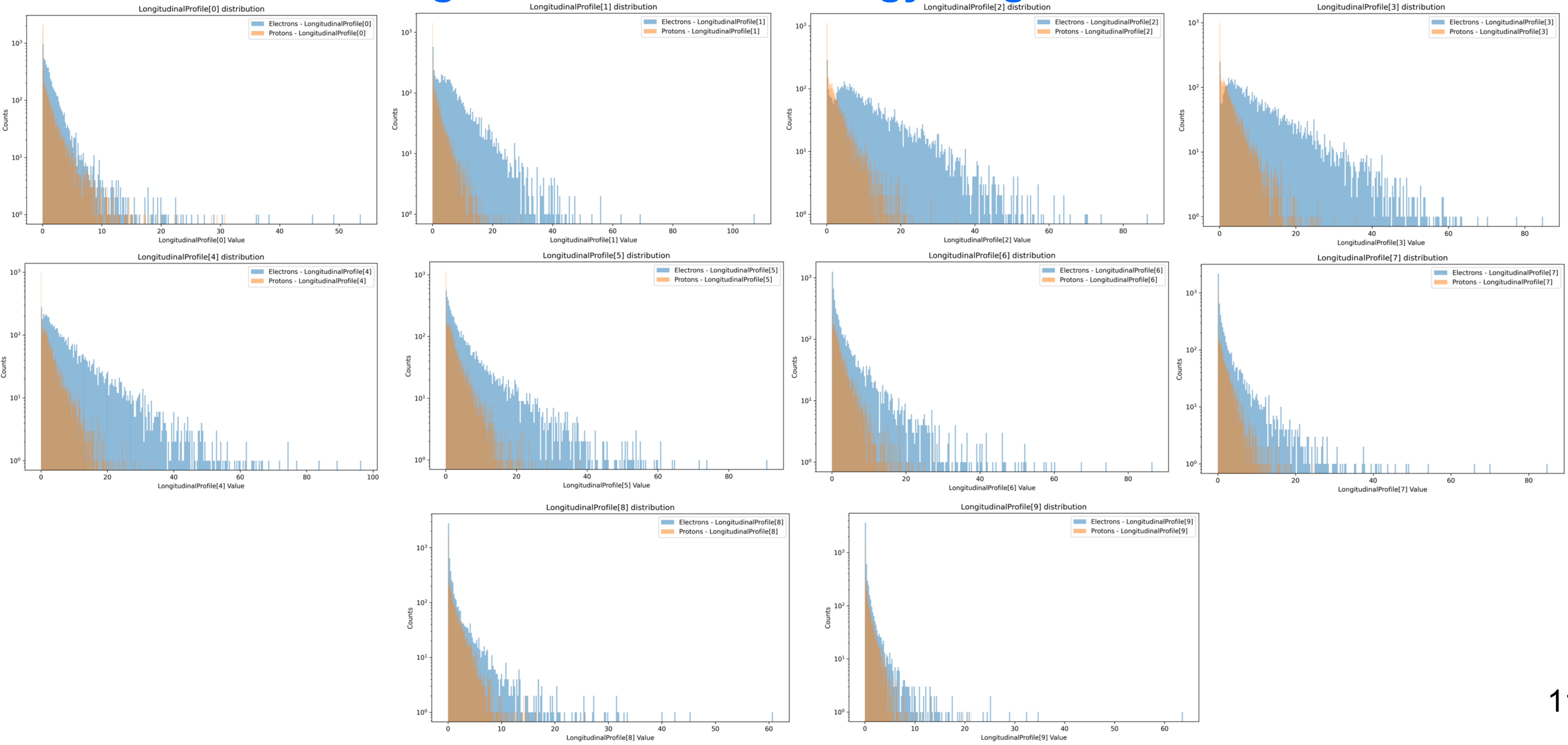




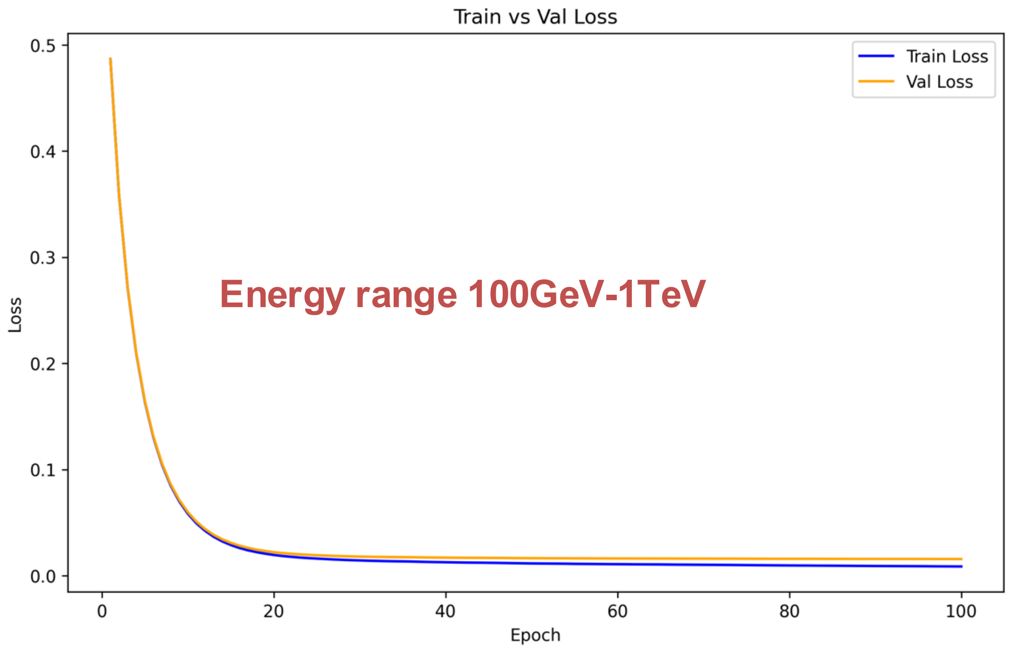


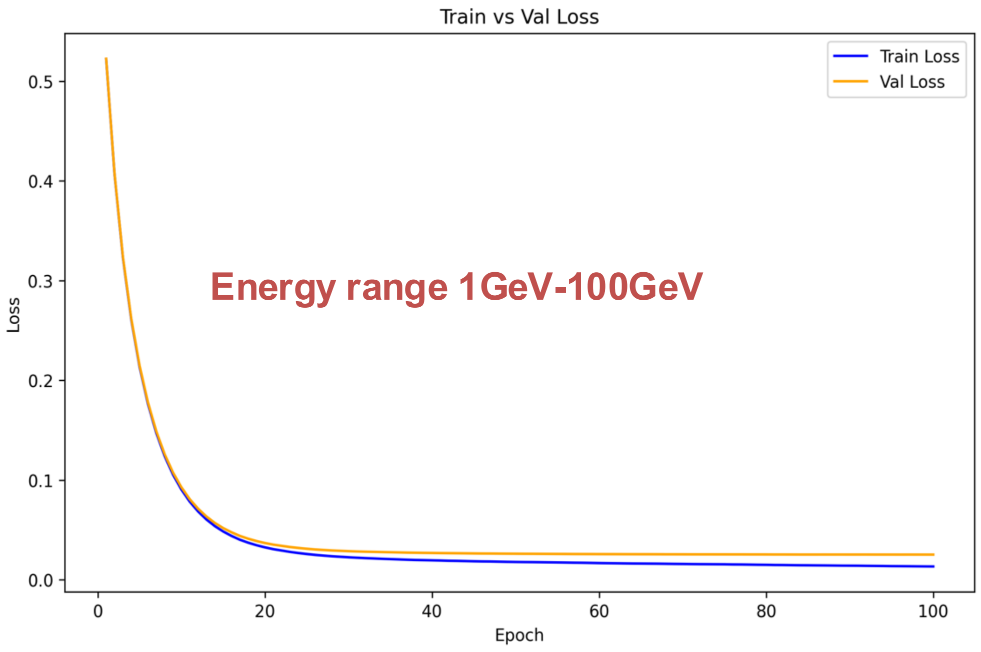
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Longitudinal Profile – Energy range 100 GeV- 1 TeV



XGBoost algorithm





- Accuracy: 0.9950
- Precision: 0.9953
- Recall: 0.9947

- Accuracy: 0.9919
- Precision: 0.9916
- Recall: 0.9921

Completed objective

•**Increase statistics**: collect a larger amount of simulated data to improve the statistical significance and reliability of the results. DONE

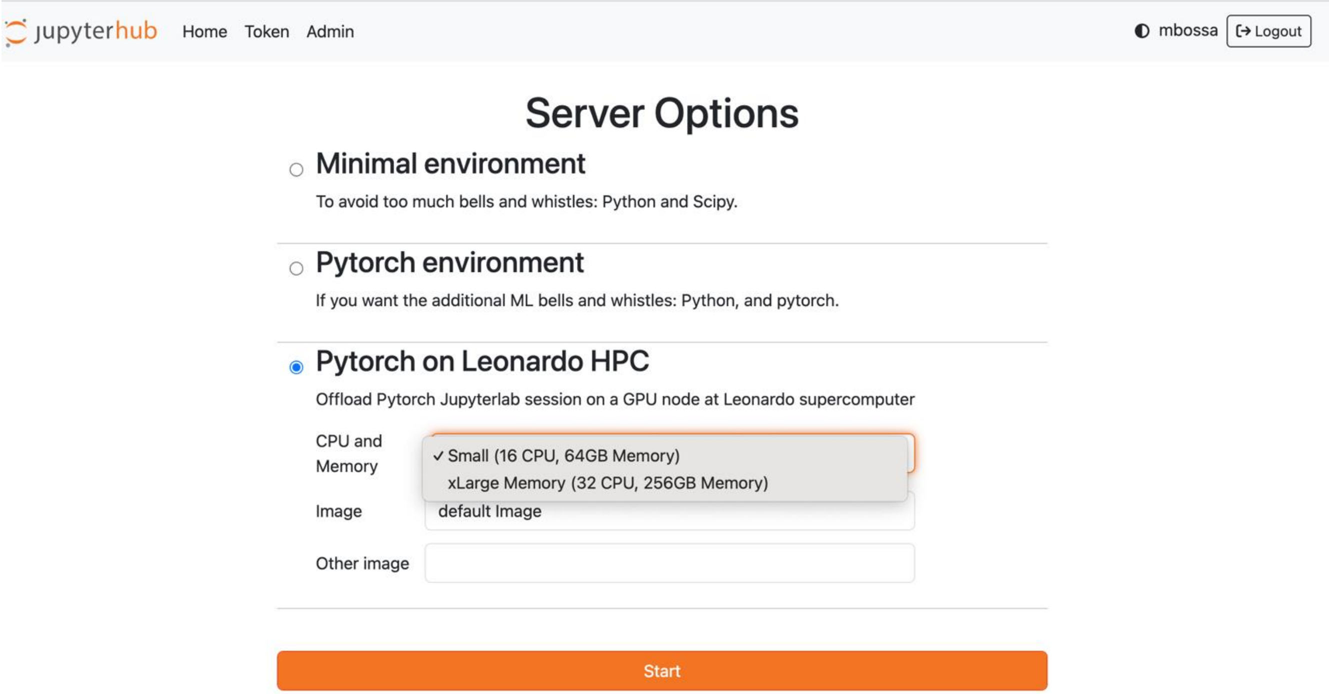
•**Extend the dynamic range of the simulation**: broaden the range of energy range simulated to ensure the model captures a wider variety of scenarios and events. DONE

•**Integration with tracking**: combine the simulation data with tracking algorithms to enhance the precision and completeness of the analysis. ON GOING

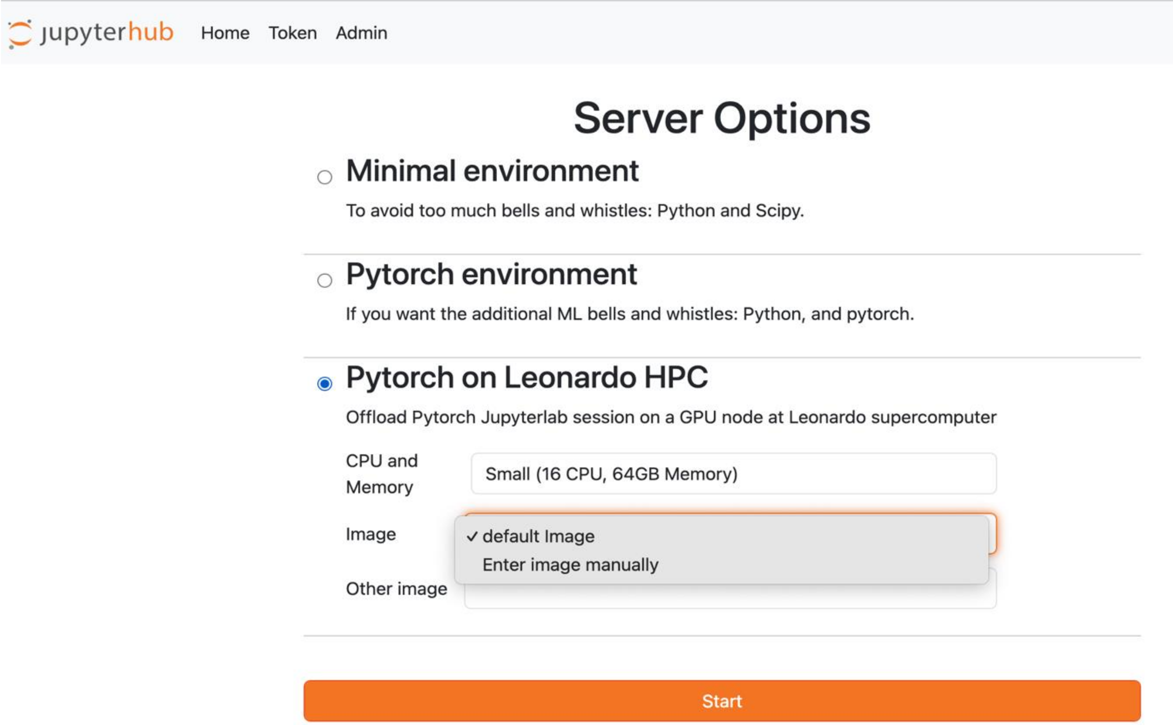
•**Current development**

We are currently experimenting with a **Visual Transfomer** using the coordinates of the activated pixels and the corresponding deposited energy. This requires **a hyperparameter search with Optuna**. Given the size of dataset, we need to **parallelize the optimization**, which is **not a trivial task**

Using Leonardo Hub



Choice of CPU memory



Choice of image

Cascade fundings

Moreover, we are also involved in the cascade funding projects under the supervision of Fabio Gargano:

- AI-Legs in collaboration with *Università Parthenope*
- LEGIMAC in collaboration with *Nuclear Instruments*
- GRAIL in collaboration with *UniMarconi*

These projects focus on the discrimination of low-energy gamma rays at different levels using artificial intelligence techniques and on the possibility of porting these codes onto FPGAs

Thank you for your attention