# A Neural-Network-defined Gaussian Mixture Model for particle identification in LHCb

Giacomo Graziani (1), Lucio Anderlini (1), Saverio Mariani (1, 2, 3), **Edoardo Franzoso** (4), Luciano Libero Pappalardo (4,5), Pasquale di Nezza (6)

(1) INFN Sezione di Firenze, Florence, Italy

(2) Università degli studi di Firenze, Florence, Italy,

(3) European Organization for Nuclear Research (CERN), Geneva, Switzerland,

(4) INFN Sezione di Ferrara, Ferrara, Italy,

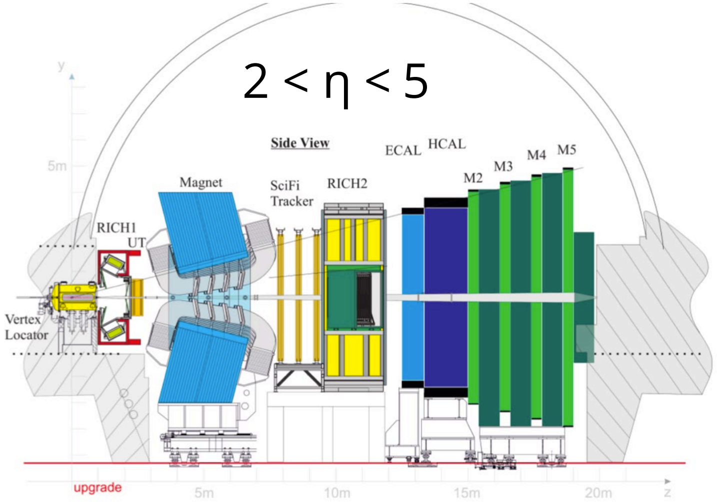
(5) Università degli studi di Ferrara, Ferrara, Italy,

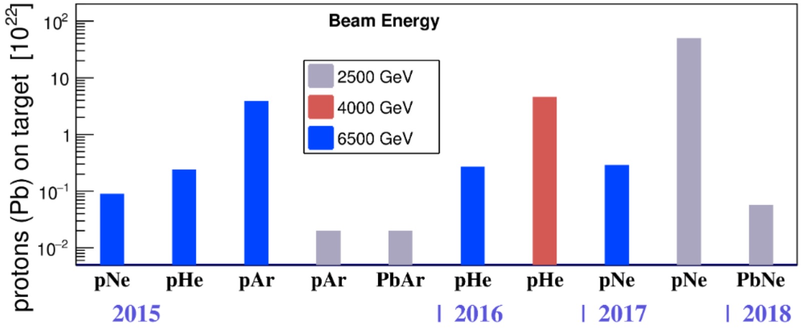
(6) INFN Laboratori Nazionali di Frascati, Frascati, Italy

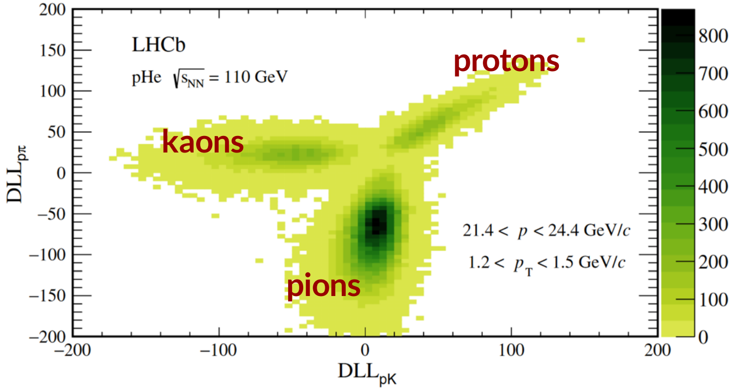
# Motivation and idea

- **LHCb** relies on the Ring Imaging Cherenkov (**RICH**) detector system for the **charged hadron identification** (PID) in a wide momentum range (2 - 100 GeV/c)

- The PID was one of the dominant systematics int the measurement of cross-sections with fixed-target datasets in LHCb:




- This motivated the development of a **novel approach to the modelling of particle identification classifiers** using machine-learning techniques

  - **Explicitly model the marginal probability density function** (pdf) of the PID classifiers → must **depend on the experimental features** $\theta$

  - Extract the marginal pdf using a **Gaussian Mixture Model**, whose parameters are predicted by Multi Layer Perceptrons trained on calibration data

  - **Model each hadron type** $h$ ( $\pi$, $K$ and $p$) PID response independently using the appropriate control modes

# Model and Validation

- In the **bidimensional PID space** the pdf $x$ is defined as: $\underline{x}_h \sim \sum_{j=1}^{N_{g,h}} \alpha_{j,h}(\underline{\theta}) \frac{\exp(-\frac{1}{2}(\underline{x}_h - \underline{\mu}_{j,h}(\underline{\theta}))^T \Sigma_{j,h}^{-1}(\underline{\theta})\,(\underline{x}_h - \underline{\mu}_{j,h}(\underline{\theta})))}{2\pi\sqrt{\det(\Sigma_{j,h}(\underline{\theta}))}}$

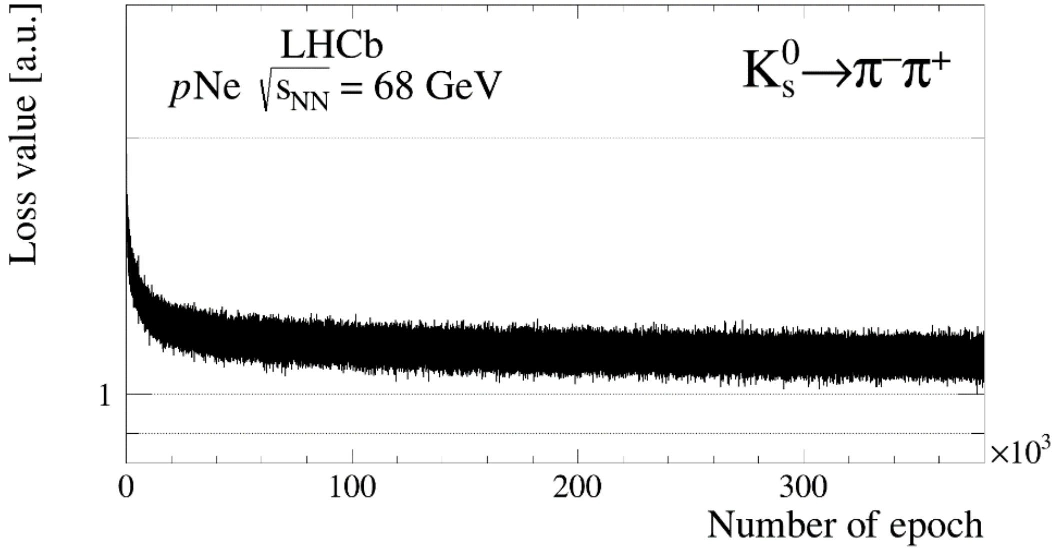
**Training on proton-Neon sample**

- The learning process of the networks relies on the **minimization** of the loss function, defined as the **negative log-likelihood**

$$\mathcal{L} = -\sum_{i=1}^{n_h} w_i \log\left[\sum_{j=1}^{N_{g,h}} \alpha_{j,h}(\underline{\theta}_i)\mathcal{G}(x_i, \mu_{j,h}(\underline{\theta}_i), \sigma_{j,h}(\underline{\theta}_i))\right]$$

- Training performed with **mini-batches gradient descent** with a user-defined number of Gaussians and NN structure

| Input parameter | $K_S^0 \to \pi^- \pi^+$ | $\overline{\Lambda} \to \overline{p}\pi^+$ | $\phi \to K^- K^+$ |
|---|---|---|---|
| Number of Gaussians | 64 | 20 | 64 |
| Number of NN nodes | 128 | 128 | 128 |
| Starting learning rate | $10^{-3}$ | $10^{-4}$ | $5 \cdot 10^{-6}$ |
| Batch size | 10000 events | 10000 events | 20000 events |



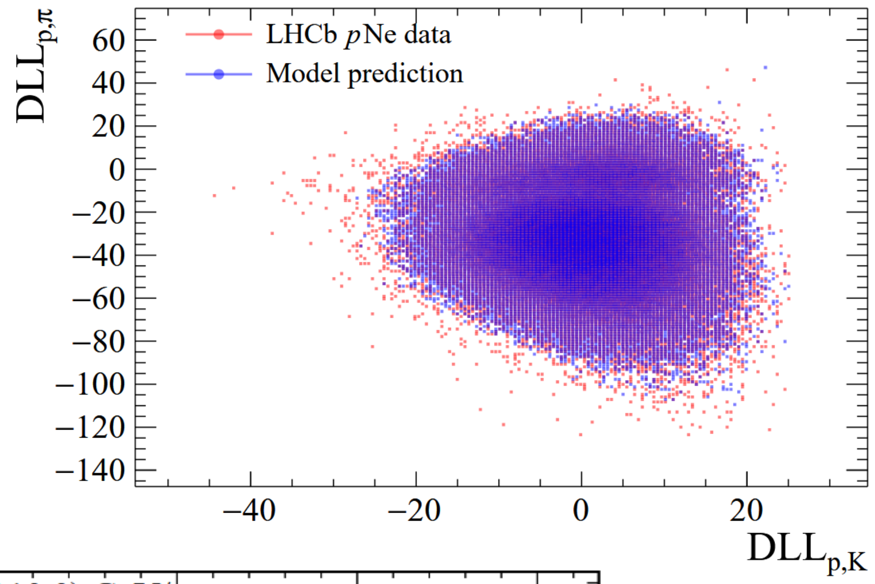


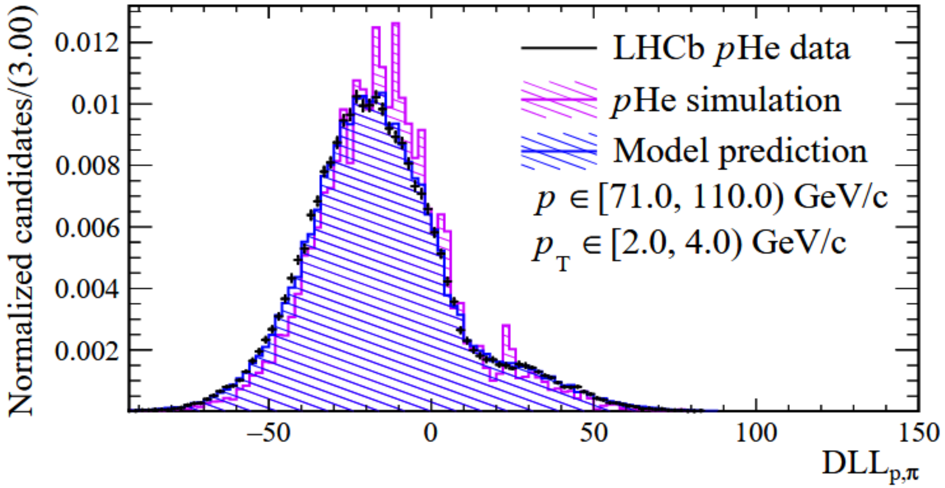

Paper : https://iopscience.iop.org/article/10.1088/1748-0221/17/02/P02018/meta

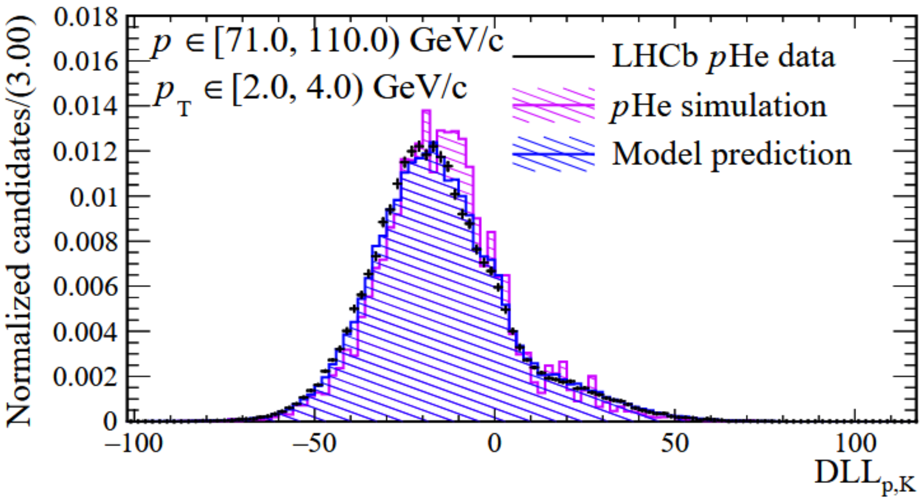
# Model and Validation

**Validation on <u>fixed-target</u> datasets**

- Marginal-pdf prediction validated for the same training dataset and then **applied to two independent lower-statistics samples** of proton-Helium and proton-Argon collisions
  - **significant modifications of the kinematic distributions** of the produced particles
  - **different events multiplicity** and the detector occupancy
  - **identical detector and data-taking conditions** can be assumed
- **Data description quality improved by the model** compared to simulation
- The method is generic, not relying on a specific set of experimental feature variable






Paper : https://iopscience.iop.org/article/10.1088/1748-0221/17/02/P02018/meta