Gamma-Hadron separation

Fernanda Heredia - Claudio Casentini – Gonzalo Rodriguez

26/03/2025

Gamma-Hadron separation
Machine learning

- **Handling Complex Data**

  Experimental data exhibits non-linear relationships that traditional methods struggle to capture;

- **Learning Complex Patterns**

  Models like Random Forest and/or XGBoost are powerful to uncover intricate relationships between different features in order to effectively distinguish between Proton and Gamma events;

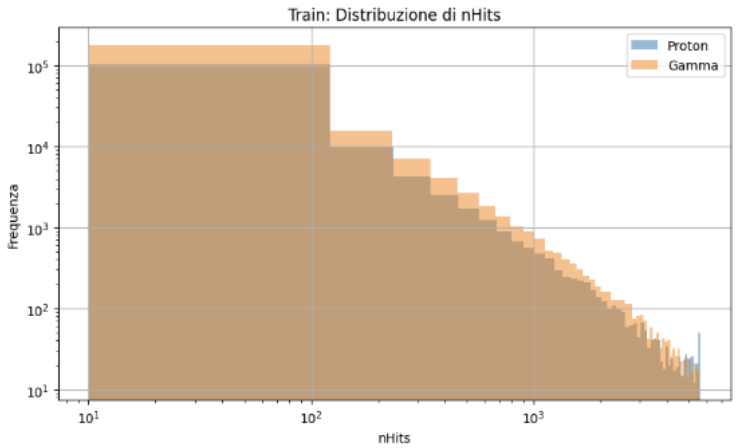
- **Automation & Adaptability**

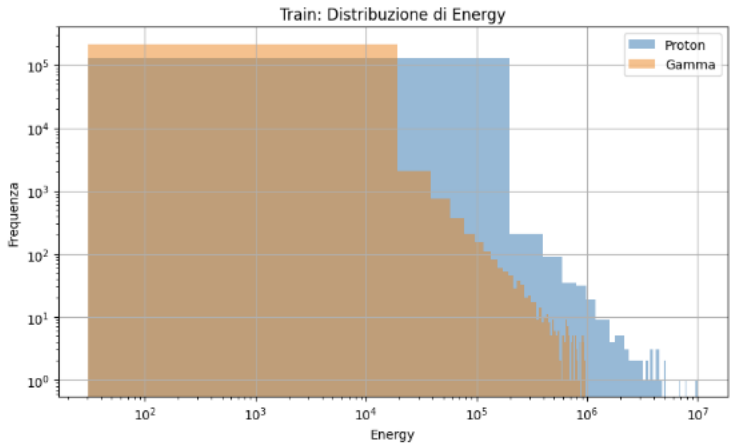
  ML-based approaches automate the classification process, reduce human error, and adapt as new data becomes available.

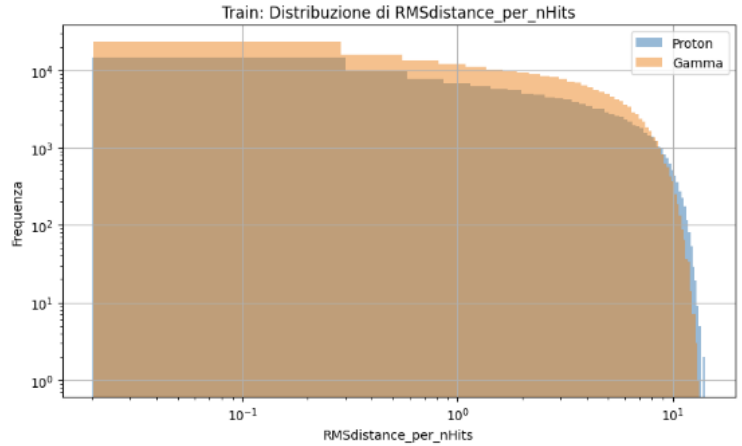
- **Continuous Evaluation & Optimization**

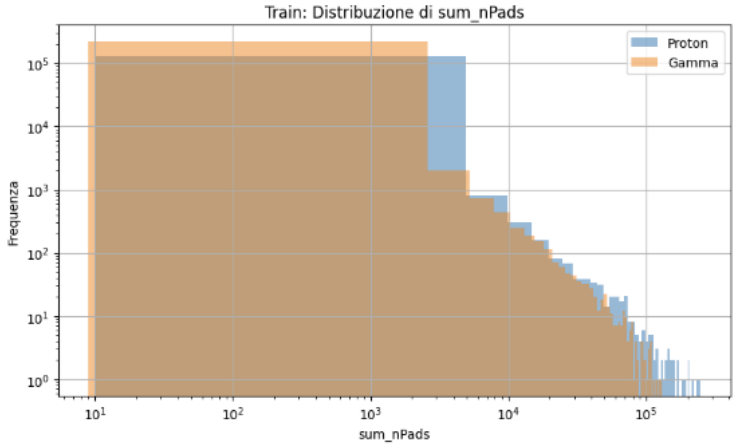
  Performance metrics (accuracy, AUC, precision, recall) provide immediate feedback, enabling ongoing model improvements and informed decision-making.

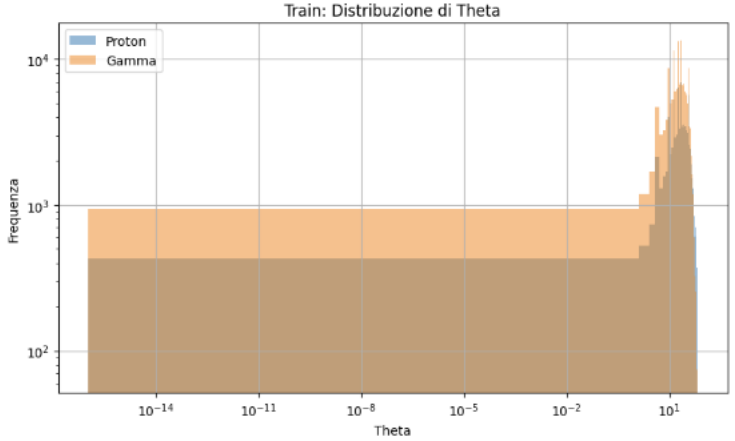
Gamma-Hadron separation

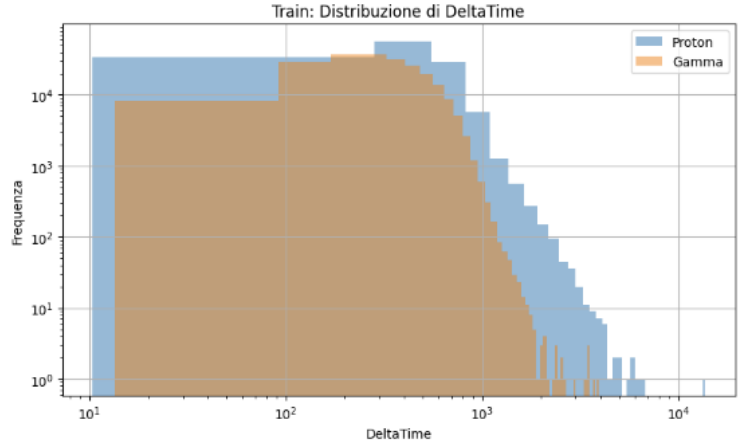












No clear separation using distributions of single variables

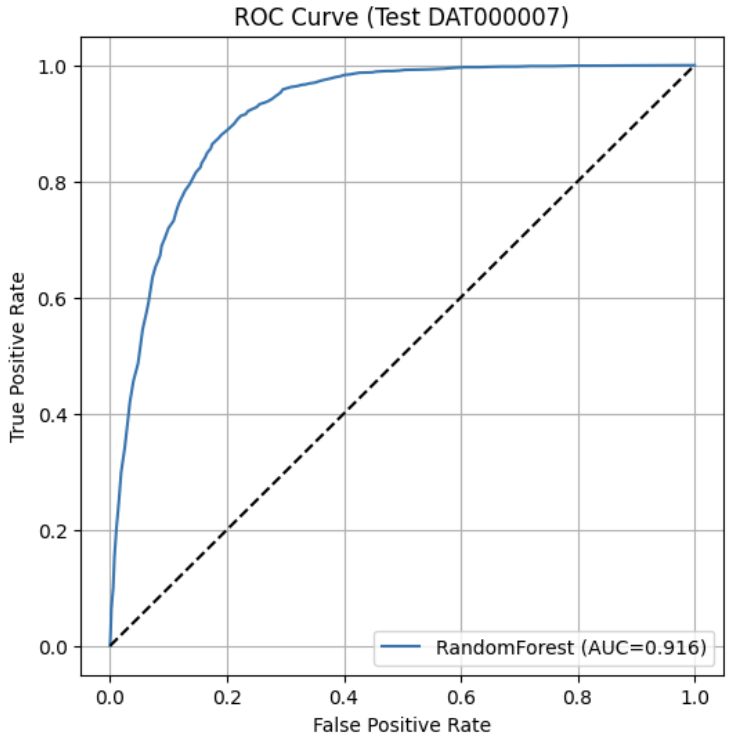
Gamma-Hadron separation
Multivariate analysis

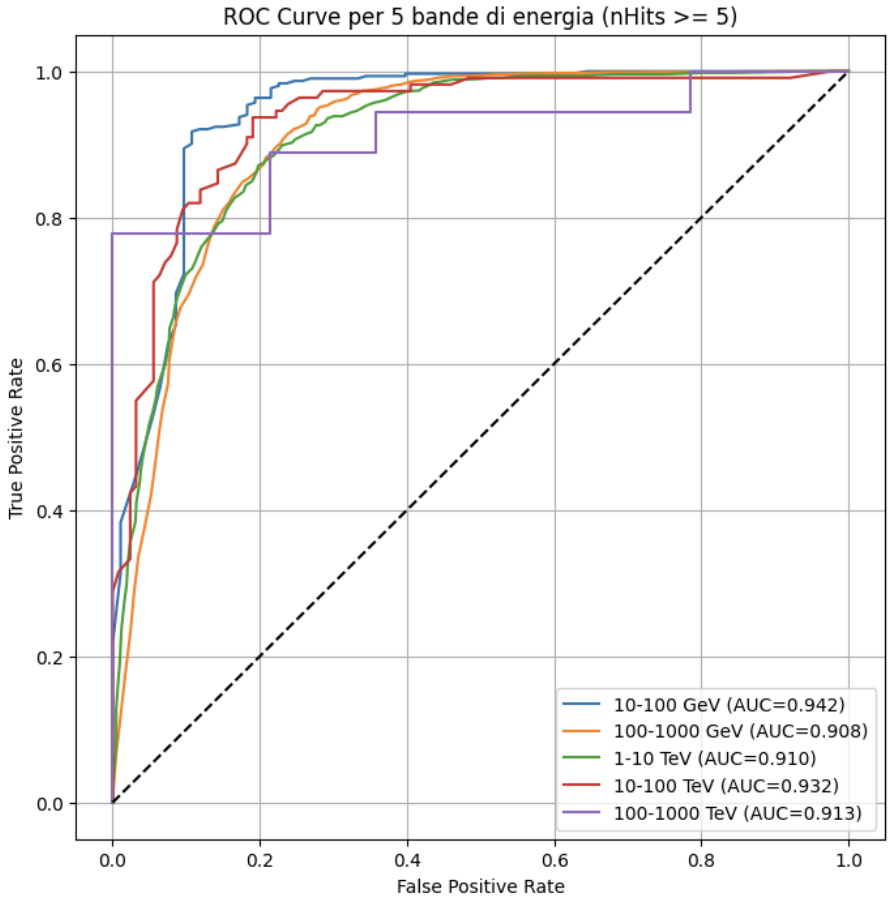
Multivariate Analysis: Variables & Their Combination

- **Event Intensity & Physical Parameters:**
  nHits: Total number of RPC hits in the event.
  sum_nPads & mean_nPads: Overall intensity measures, capturing the total and average number of activated pads.
  Energy, Theta: Event's energy and incident angle.
- **Spatial Distribution Metrics:**
  RMSdistance: Root mean square distance of all hits from the event's centroid (computed from xRPC and yRPC).
  RMSdistance_per_nHits: Normalized RMS distance, dividing by the number of hits to adjust for event multiplicity.
- **Temporal Spread:**
  DeltaTime: Time difference between the first and last hit, quantifying the event duration.
- **Hit Concentration & Distribution Shape:**
  fractionHits20 & fractionHits40: Fractions of hits within 20 and 40 units from the centroid, indicating how concentrated the event is.
- **Combination Strategy:**
  By integrating geometric, intensity, and temporal features, the classifier can capture complex, non-linear correlations that a single parameter alone would miss—resulting in a more robust and discriminative model.

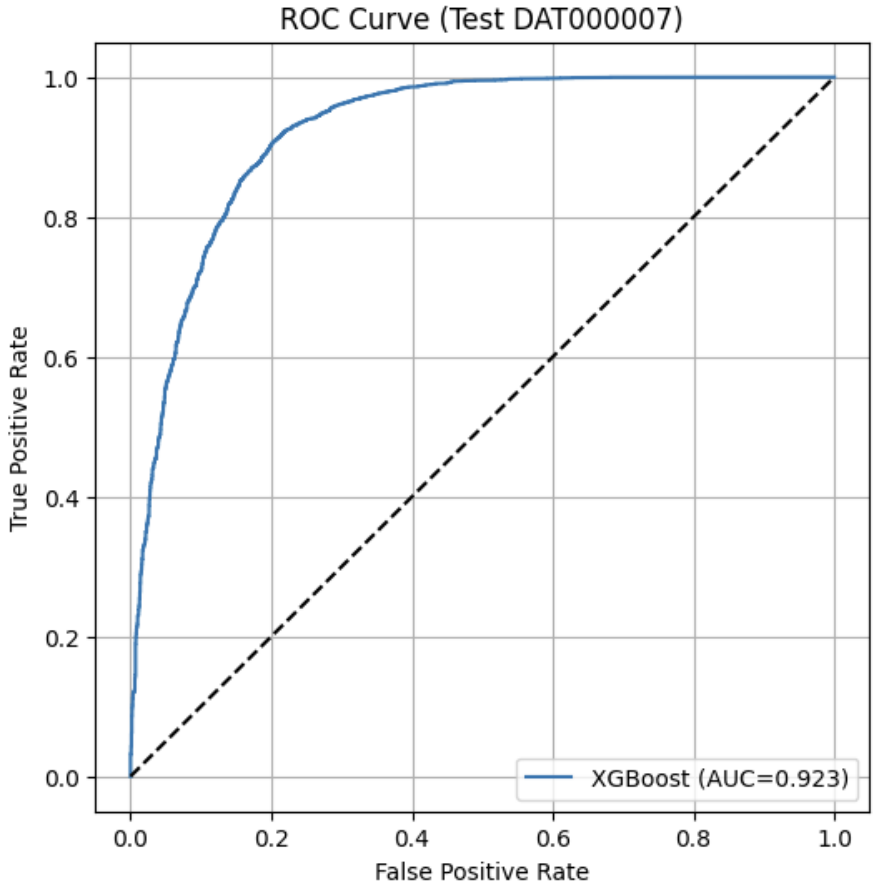
Gamma-Hadron separation
Random forest

```
Confusion Matrix (Test DAT000005):
[[1709  572]
 [ 282 3630]]

Classification Report (Test DAT000005):
              precision    recall  f1-score   support

      Proton       0.86      0.75      0.80      2281
       Gamma       0.86      0.93      0.89      3912

    accuracy                           0.86      6193
   macro avg       0.86      0.84      0.85      6193
weighted avg       0.86      0.86      0.86      6193
```





Gamma-Hadron separation
XGBoost




```
Confusion Matrix (Test DAT000005):
[[1717  564]
 [ 245 3667]]

Classification Report (Test DAT000005):
              precision    recall  f1-score   support

      Proton       0.88      0.75      0.81      2281
       Gamma       0.87      0.94      0.90      3912

    accuracy                           0.87      6193
   macro avg       0.87      0.85      0.85      6193
weighted avg       0.87      0.87      0.87      6193
```

Gamma-Hadron separation
Future steps

- Include more dataset to the training -> from 100 datasets to 1000;

- Include more parameter to the multivariate analysis, like the LDF;

- Compare our results with the Gamma-Hadron separation of SWGO;

Gamma-Hadron separation
IACT method