Application of Machine Learning Methods for Event Reconstruction in the SPD Range System

Alexandr Osetrov (JINR, NRNU MEPhI)

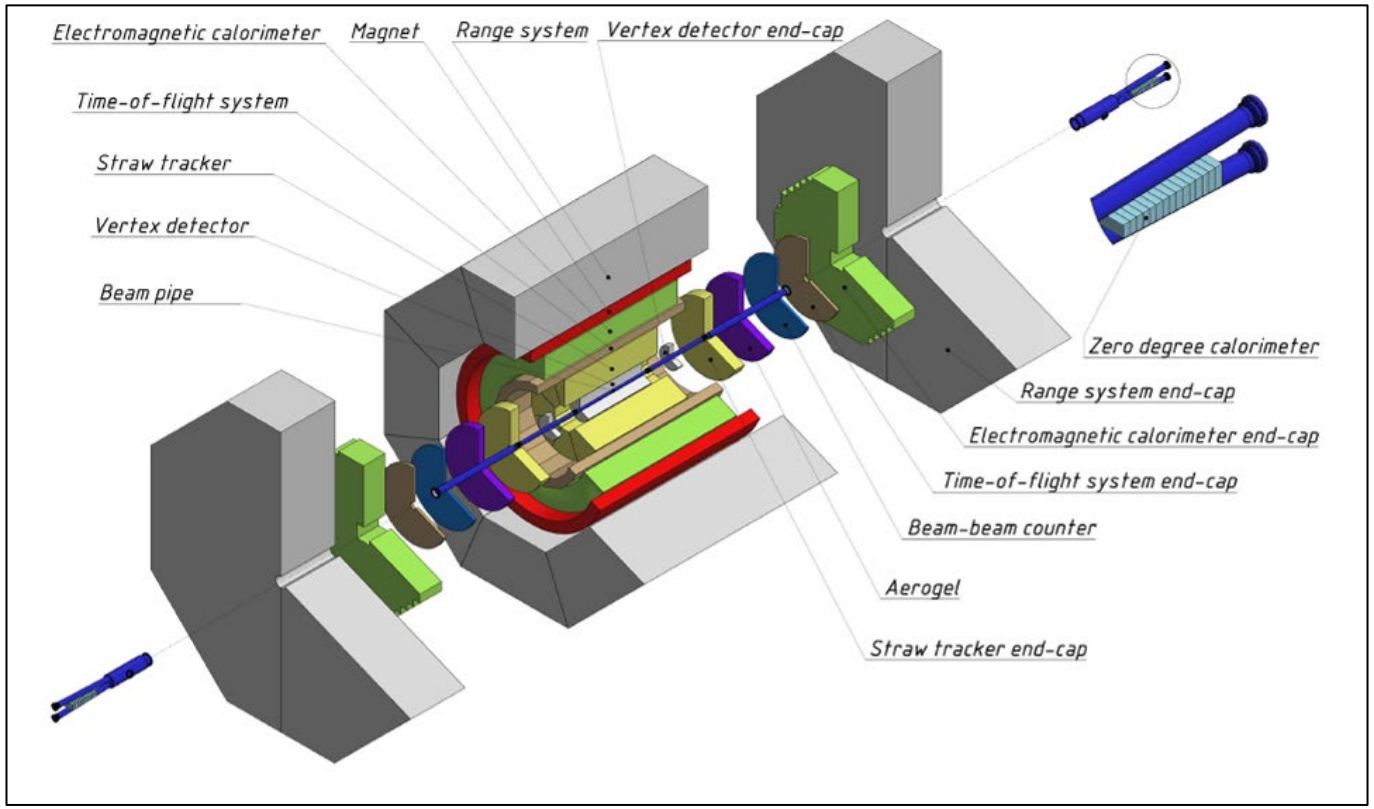
ICPPA 2026, Moscow

Outline

- Building geometric modeling of the RS prototype for algorithm validation
- Development hit clustering algorithm to extract particle-specific clusters
- ML-based particle identification development
- Comparative efficiency evaluation and optimal model selection

SPD Experiment

The goal of the experiment is to investigate the spin structure of the proton and deuteron, as well as other spin phenomena, using collisions of polarized p-p, d-d and p-d beams at NICA collider (Dubna) at $\sqrt{s}$ = 27 GeV and $\mathcal{L} \sim 10^{32} cm^{-2} s^{-1}$.



Range System

Physics Tasks:

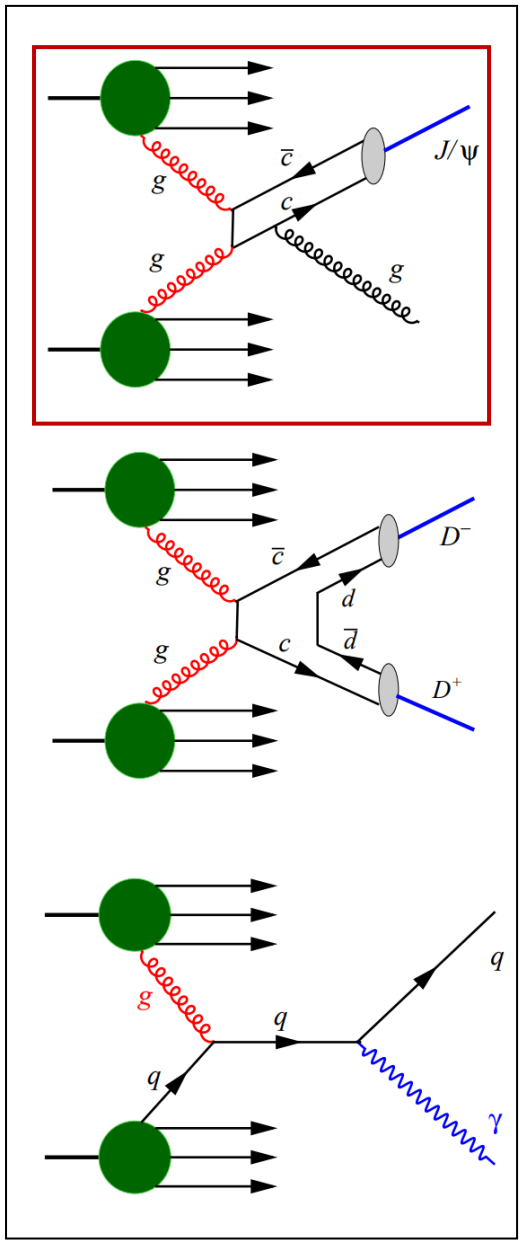
- Probes for gluons (charmonia, open charm, direct photons). Primarily - identification of $J/\psi \rightarrow \mu^+\mu^-$ decays against background pions and their decay products

Core Functionality:

- Muon Identification: Efficient PID under high hadronic background
- Coarse Hadronic Calorimetry: Energy estimation for hadrons
- Neutron Identification: The only central subdetector capable of neutron PID (combined with ECAL and inner trackers)

Operating Principle:

- Track segment recognition inside the muon system
- Track matching with the inner magnetic system
- Momentum measurement via inner trackers in the magnetic field

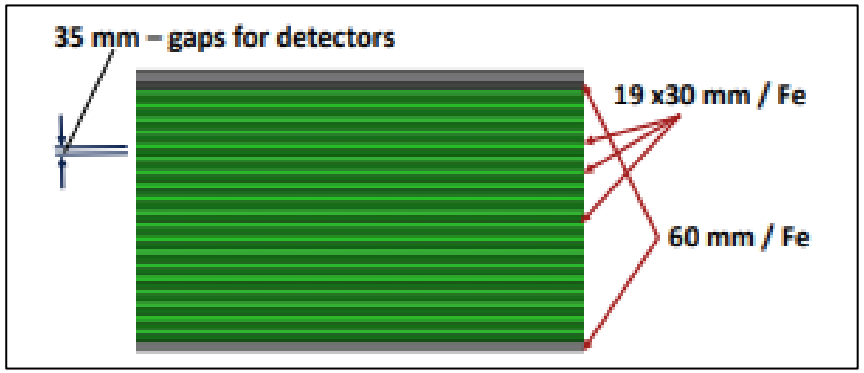


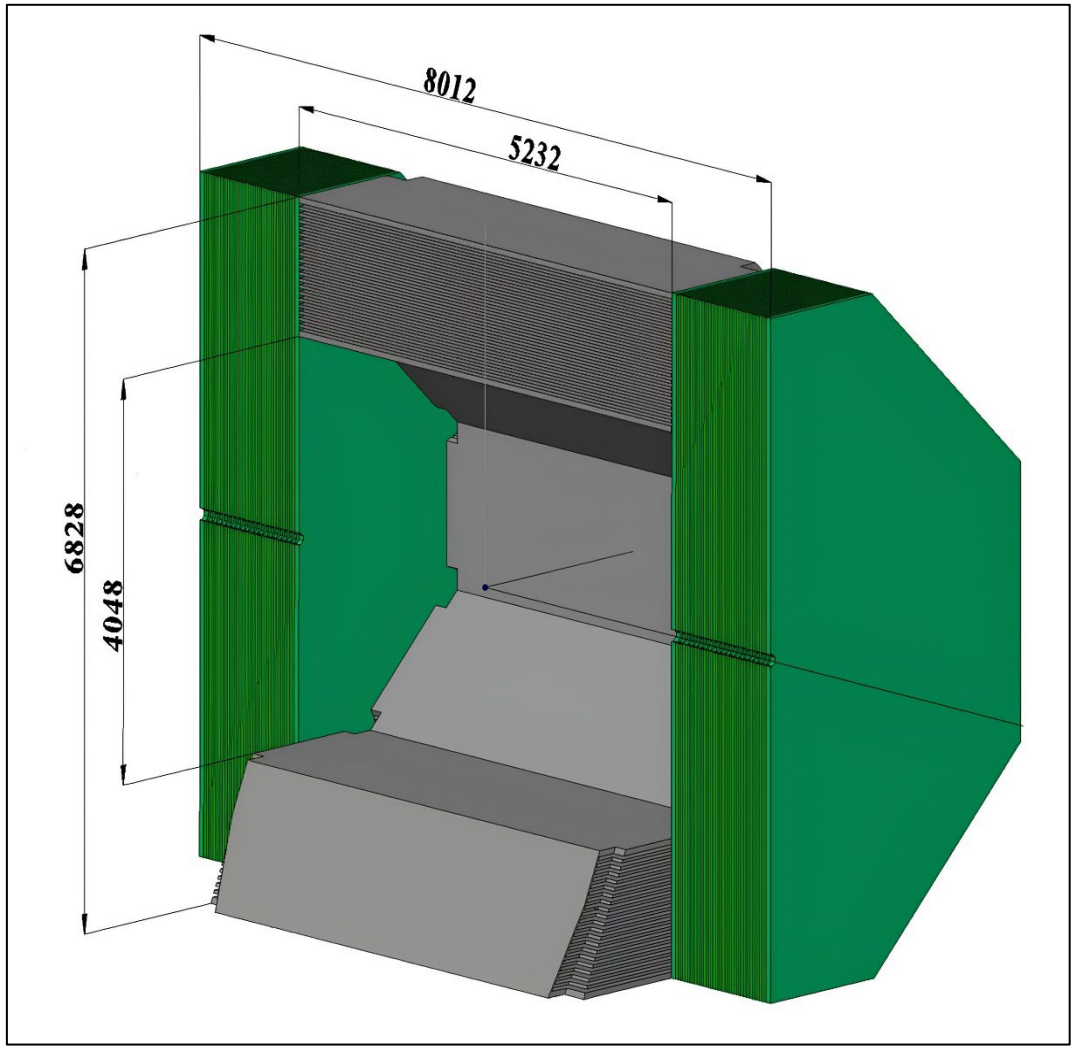
Range System

RS consists of alternating layers of iron absorbers and gas-filled detectors.
Iron Plates:

- Main plates (30 mm): Optimized based on muon range in steel for efficient μ/π separation and calorimetry
- Outer layers (60 mm): Structural reinforcement and module interlocking

Active Layers: Mini Drift Tubes (MDTs) and readout electronics installed in inter-layer gaps





MDT Detectors

Detector placement:

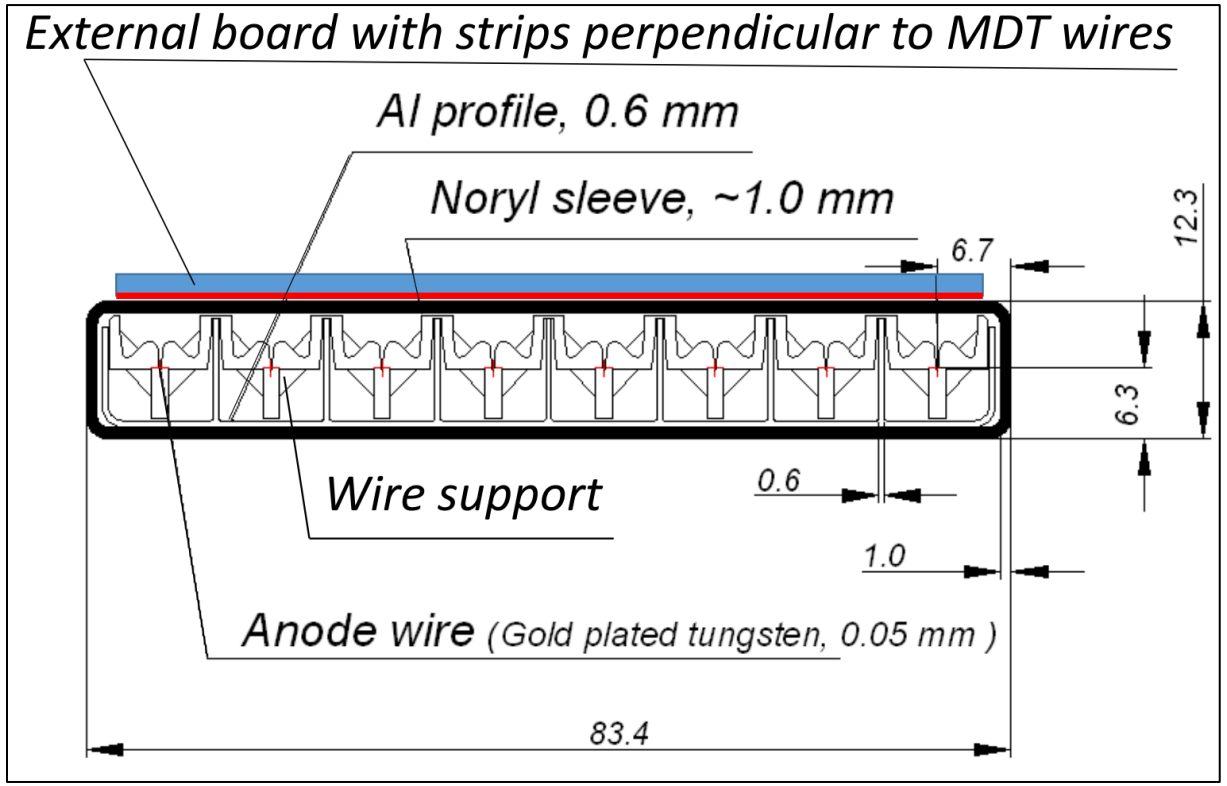
- ~ 8,000 MDTs in total
- Arranged parallel to the beam axis in the barrel
- Arranged perpendicular to the beam axis in the endcaps

MDT Design:

- 8-cell aluminum comb-profile (cathode)
- Anode wires
- Noryl sleeve

Measurement:

- 1st coordinate: Independent signal from individual anode wires
- 2nd coordinate: External perpendicular strip plane

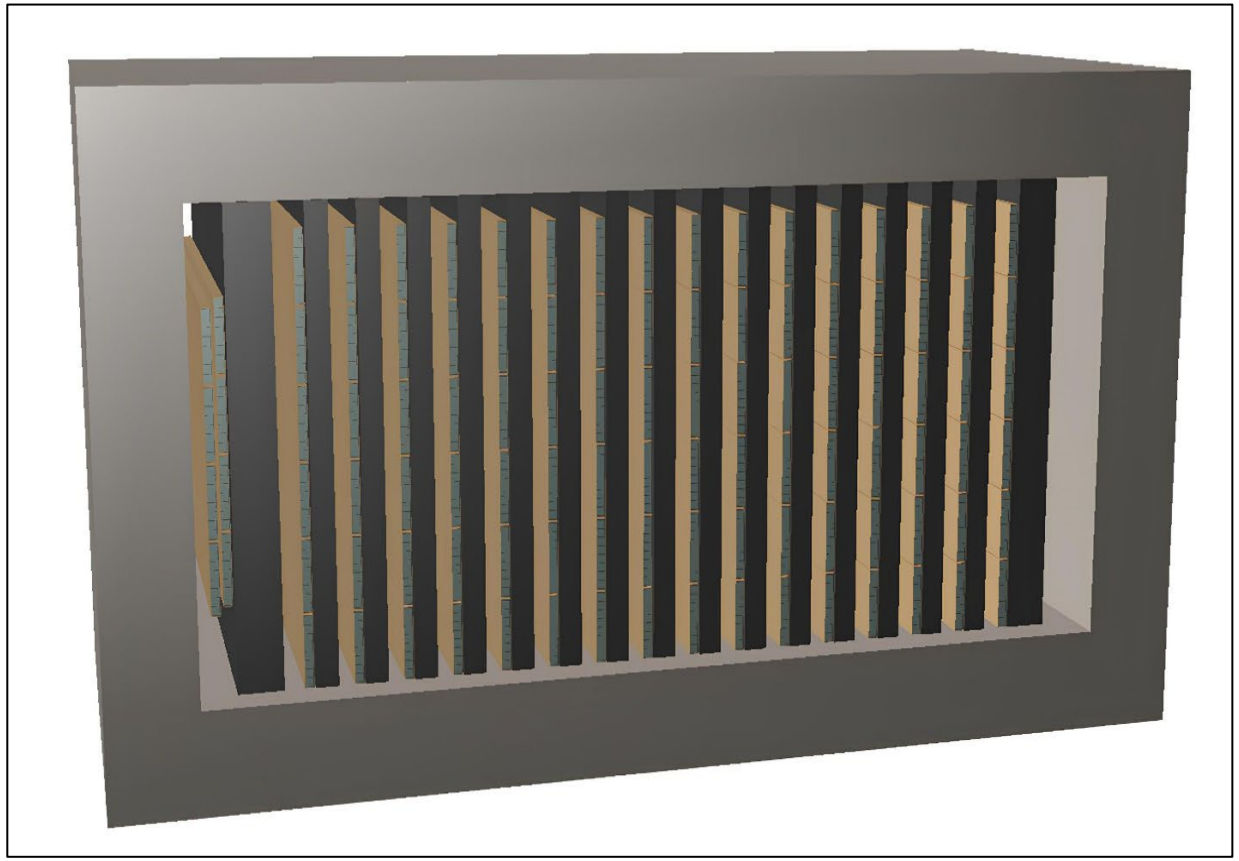


Geometric Model of the RS Prototype

Modeled within the CERN ROOT environment
Prototype Geometry:

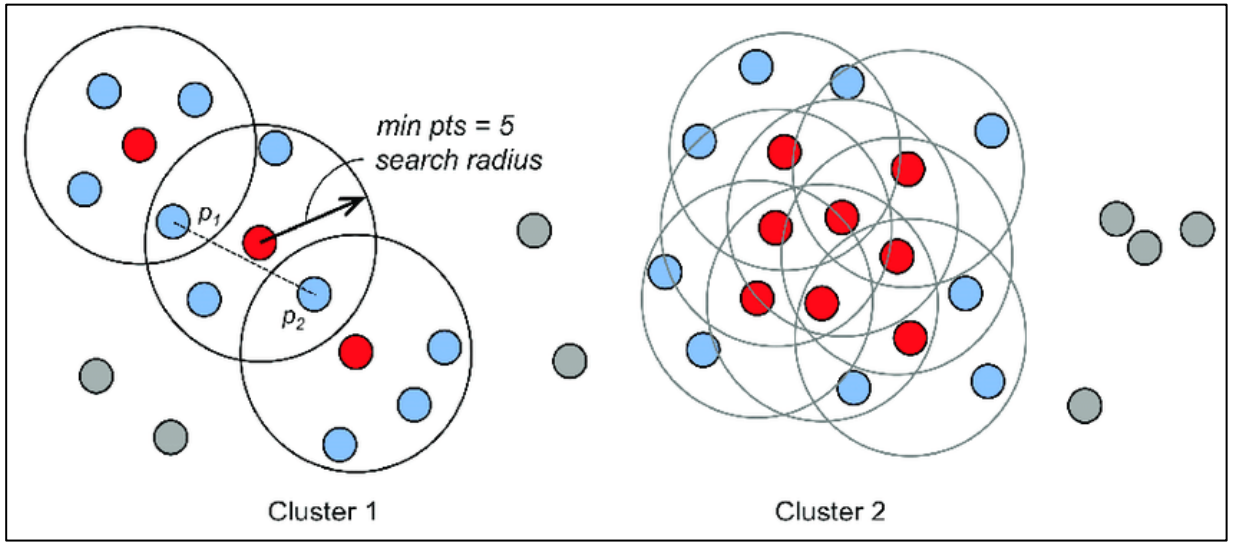
- Total Active Layers: 16 detector planes alternating with iron absorbers
- Inter-plate Gaps: 35 mm between adjacent iron plates
- Plane Composition: 6 MDT detectors per standard plane
- Front-end Bilayer: 2 vertically offset planes (4 MDTs each) followed by a 60 mm iron sheet

Current Application: Electronics testing and cosmic ray data acquisition
Objective: Software validation using experimental data from the RS prototype

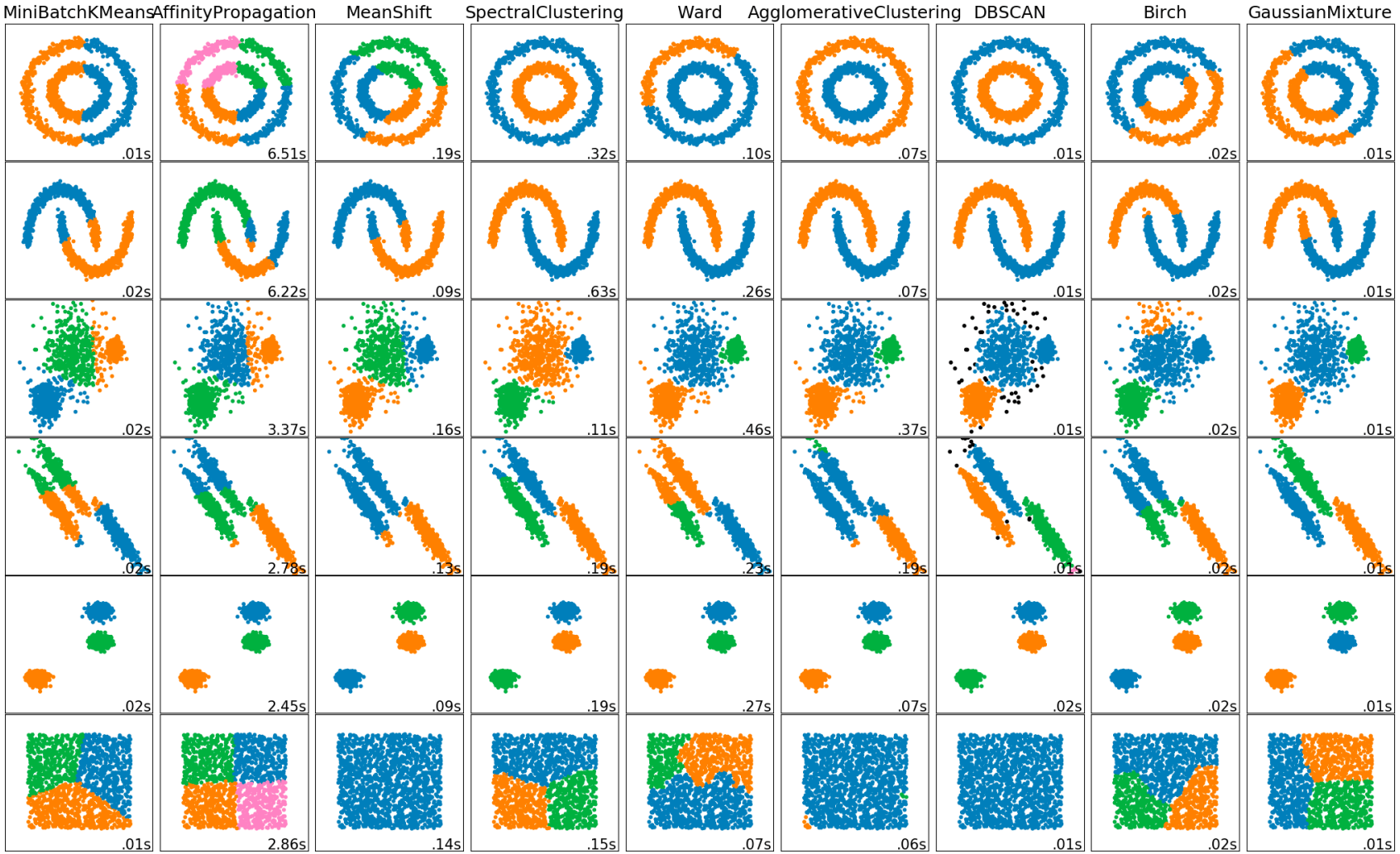


DBSCAN Clustering Algorithm

1 Setting Parameters: To run the algorithm, two key parameters are defined: ε and minPts.
2 Core Formation: The algorithm selects a random point and creates a new cluster around it, adding all of its neighbors.
3 Cluster expansion: The process is repeated recursively for all points that fall into the same cluster - they "pull" their neighbors into it if their number exceeds minPts.
4 Noise exclusion: Boundary points are added to the corresponding cluster, and noise points are ignored and marked as outliers.
5 Termination: The process stops when all points are assigned to clusters or marked as noise.



DBSCAN Clustering Algorithm



Source: scikit-learn documentation. https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html

Machine Learning Algorithms

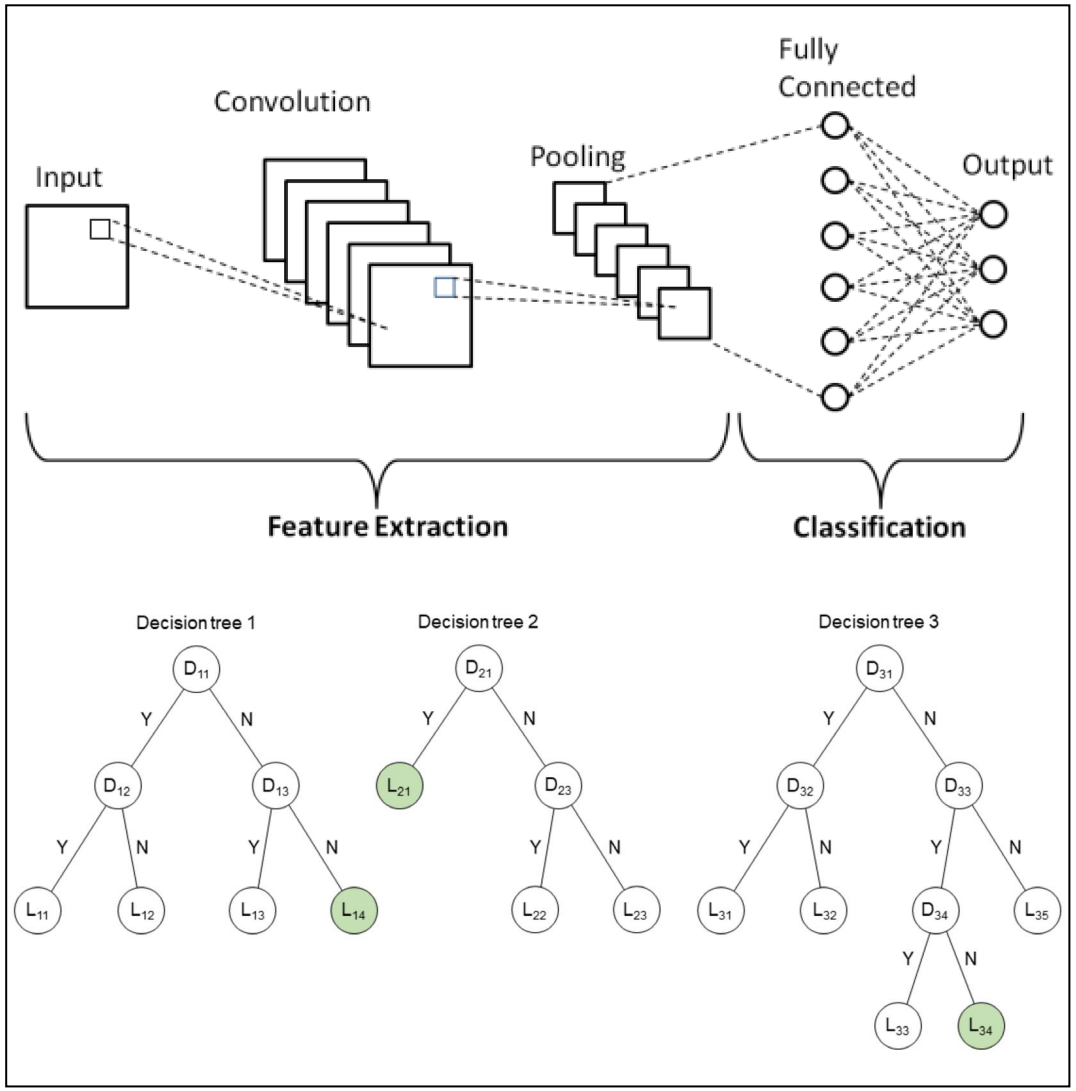
Convolutional Neural Networks (CNNs):

- Application: Spatial response structures (energy deposition topologies, track geometry)
- Mechanism: Sliding filters and pooling operations
- Advantage: Efficient extraction of local features regardless of spatial shifts

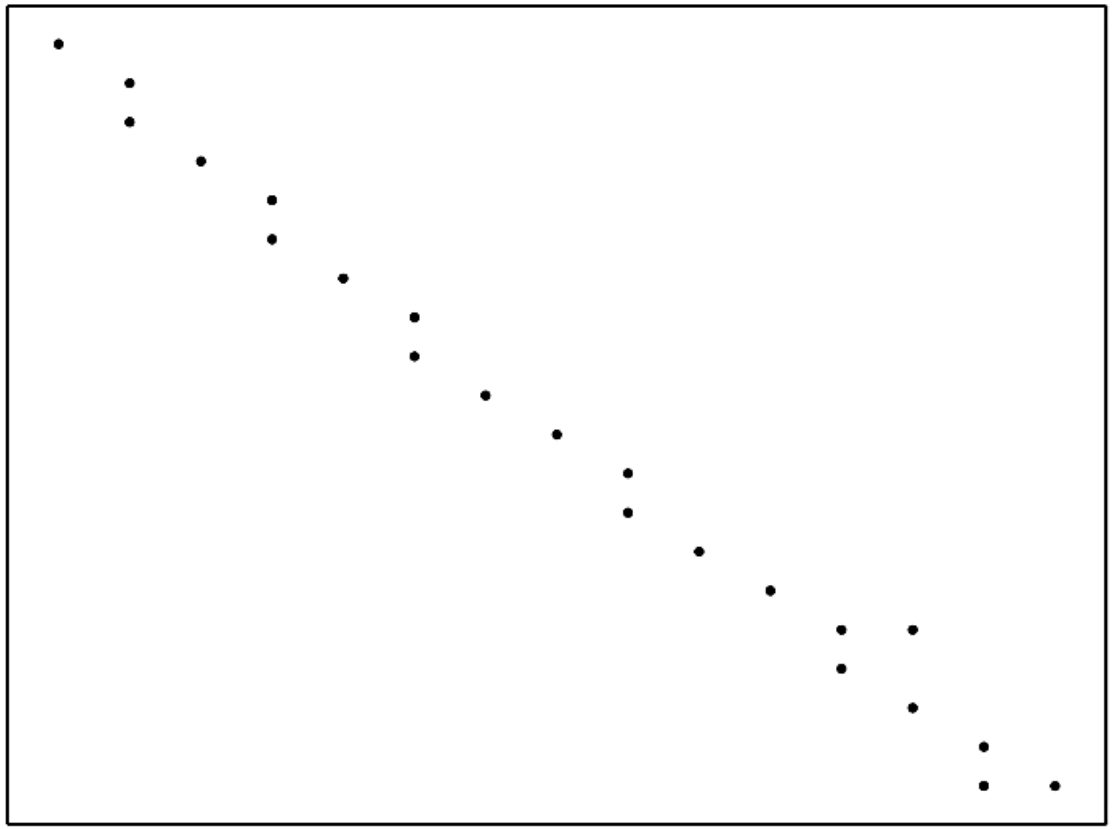
Boosted Decision Trees (BDTs):

- Application: Reconstructed physical parameters (momentum, mass, angular variables)
- Mechanism: Sequential chains of simple decision rules minimizing prior errors
- Advantage: High classification speed and accuracy for tabular high-level variables

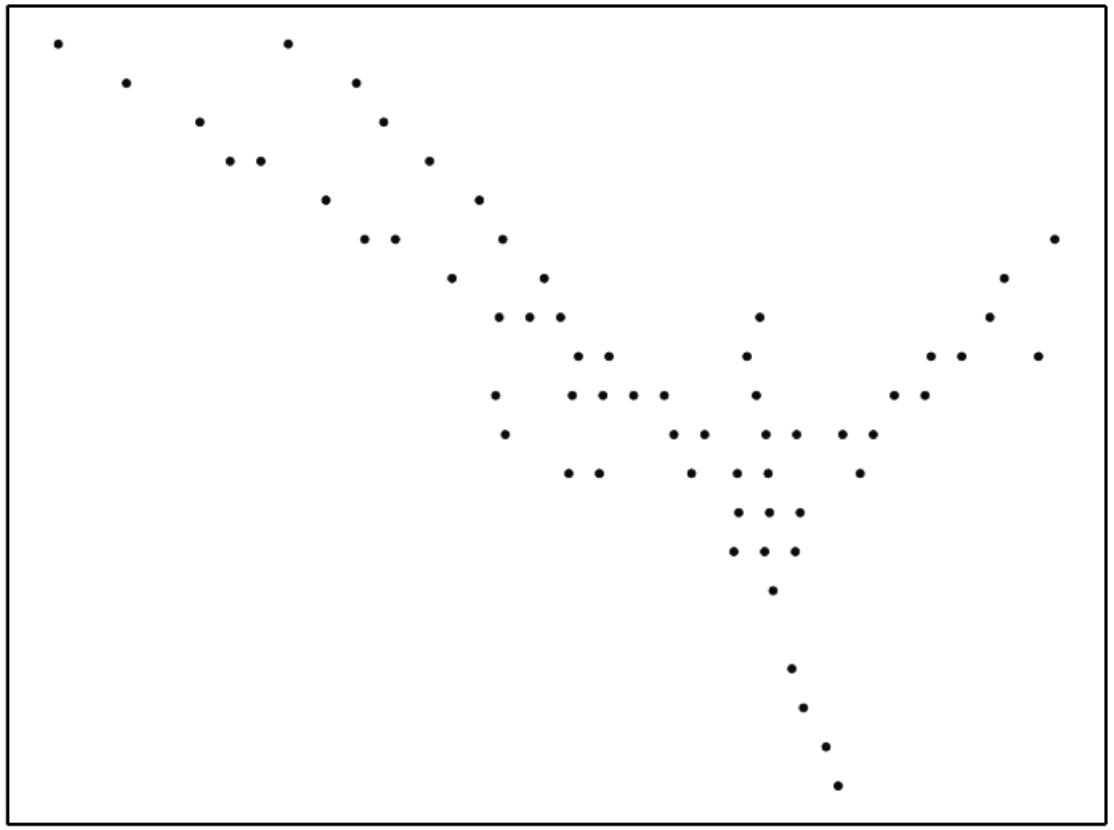
Algorithm Selection is driven by the detector data format. Combining both approaches optimizes particle identification performance



Cluster Examples

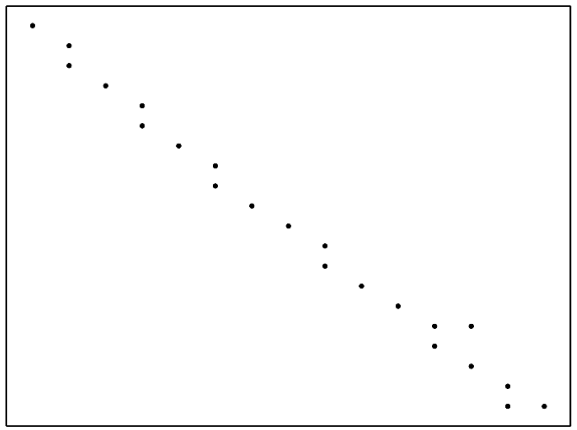


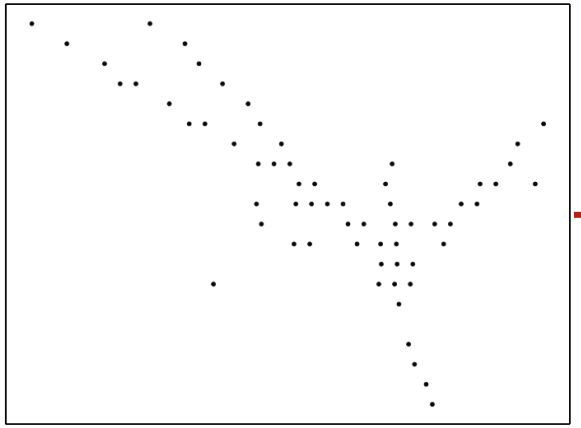
μ cluster example

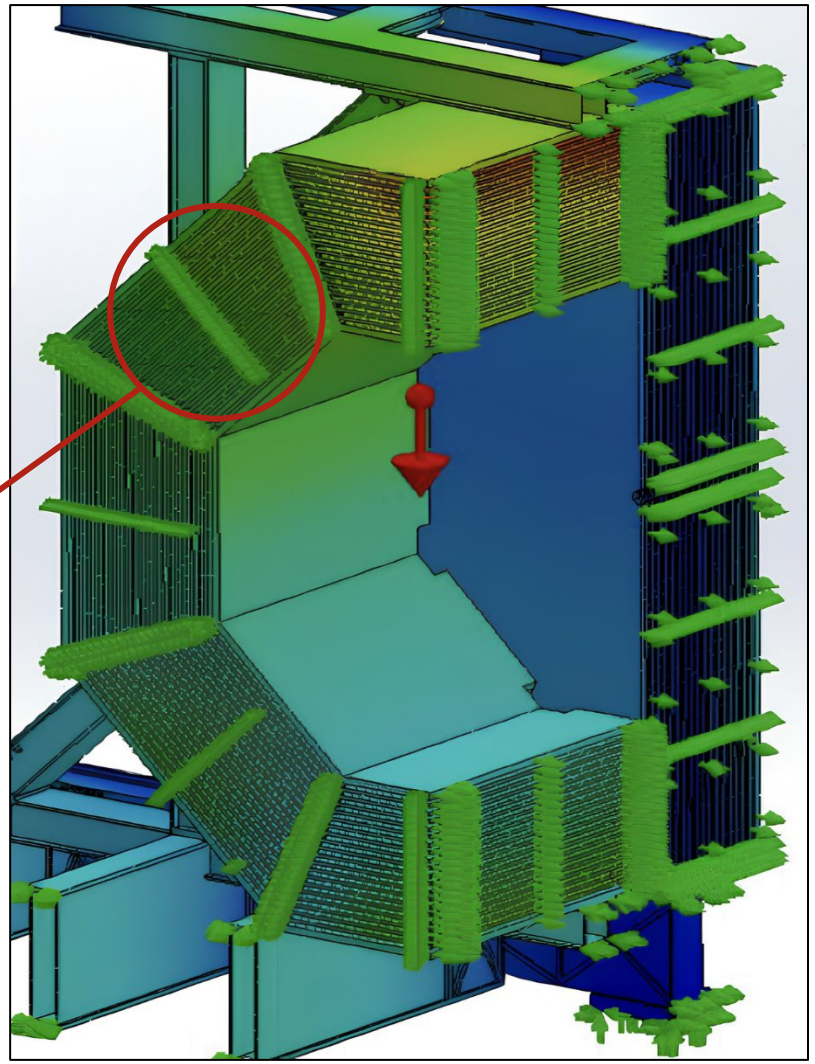


p cluster example

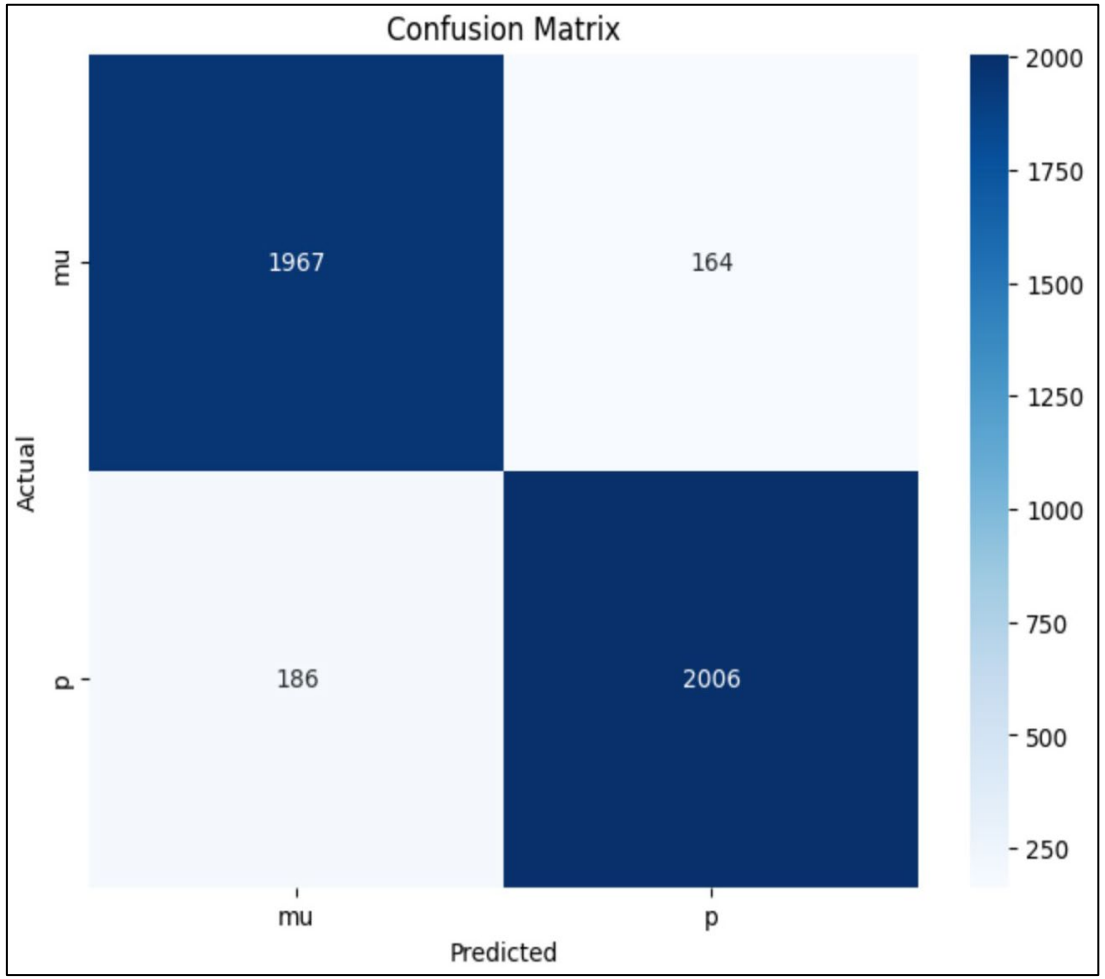
Cluster Examples

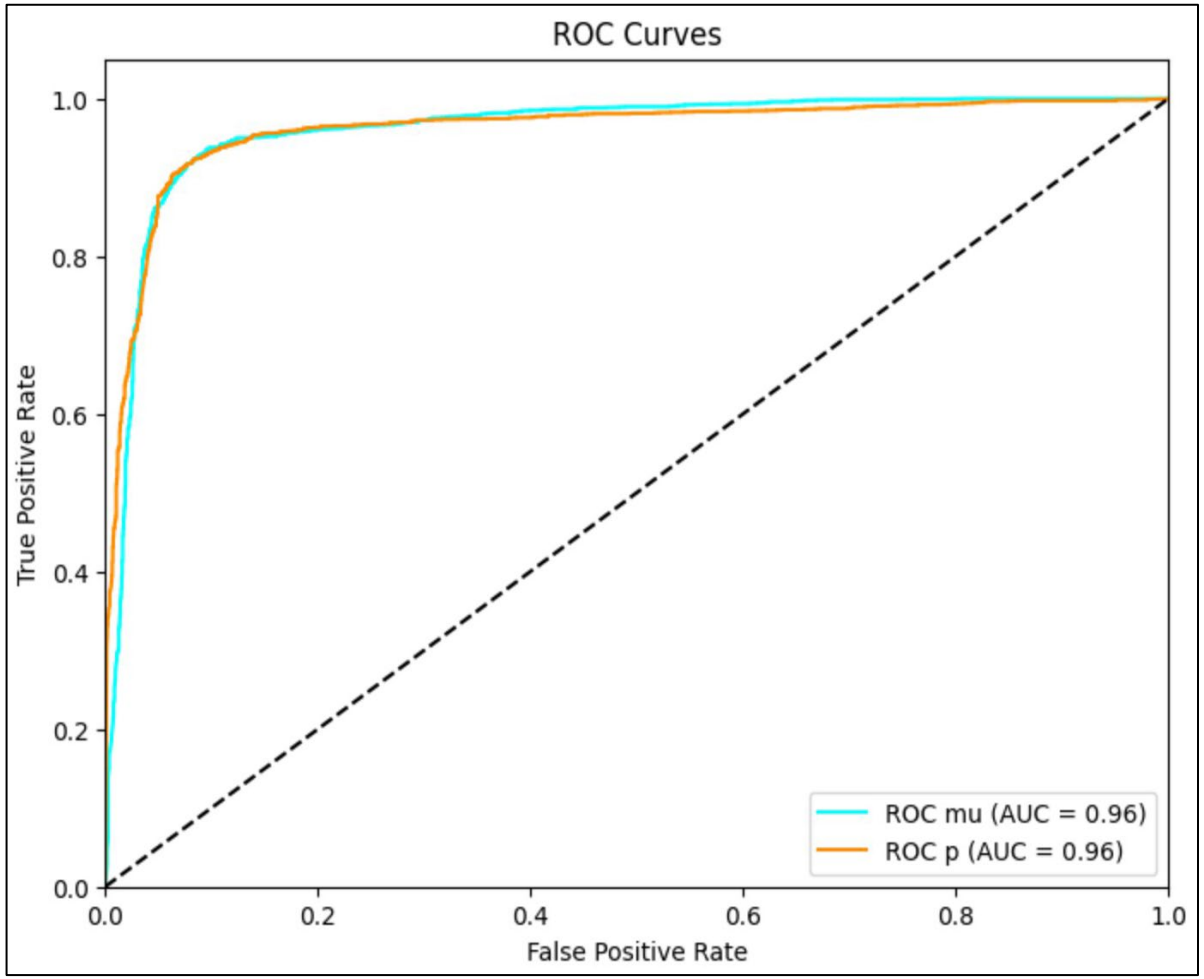






Classification Algorithm: CNN

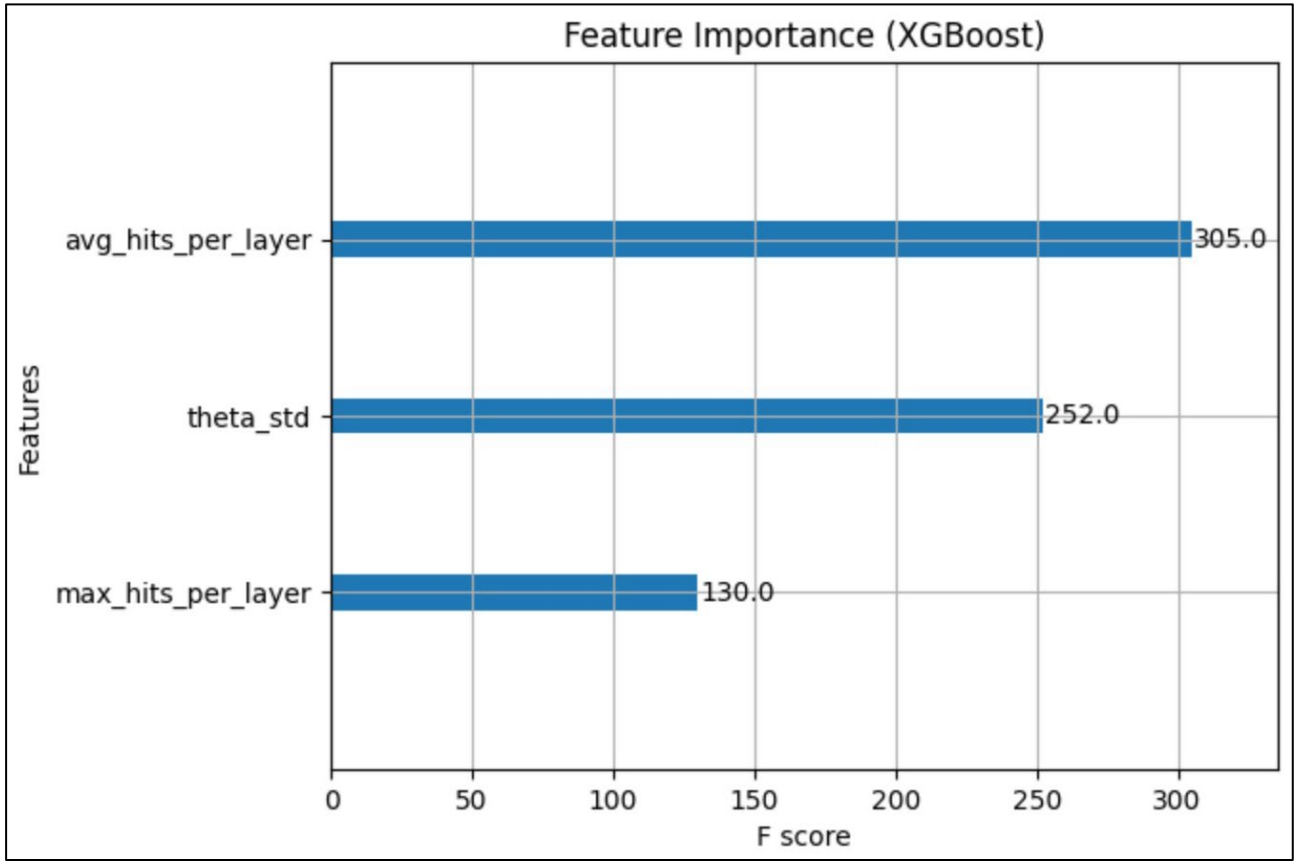




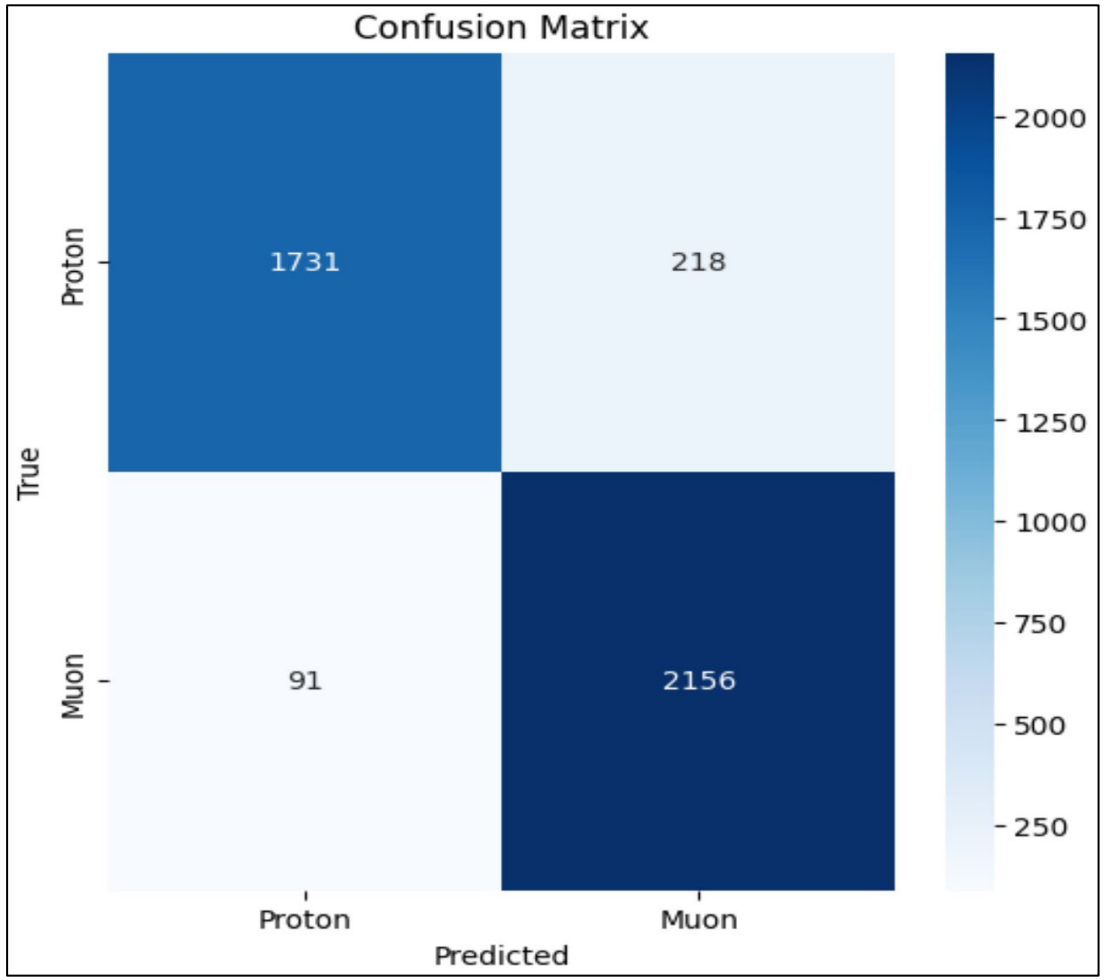
Classification Algorithm: BDT

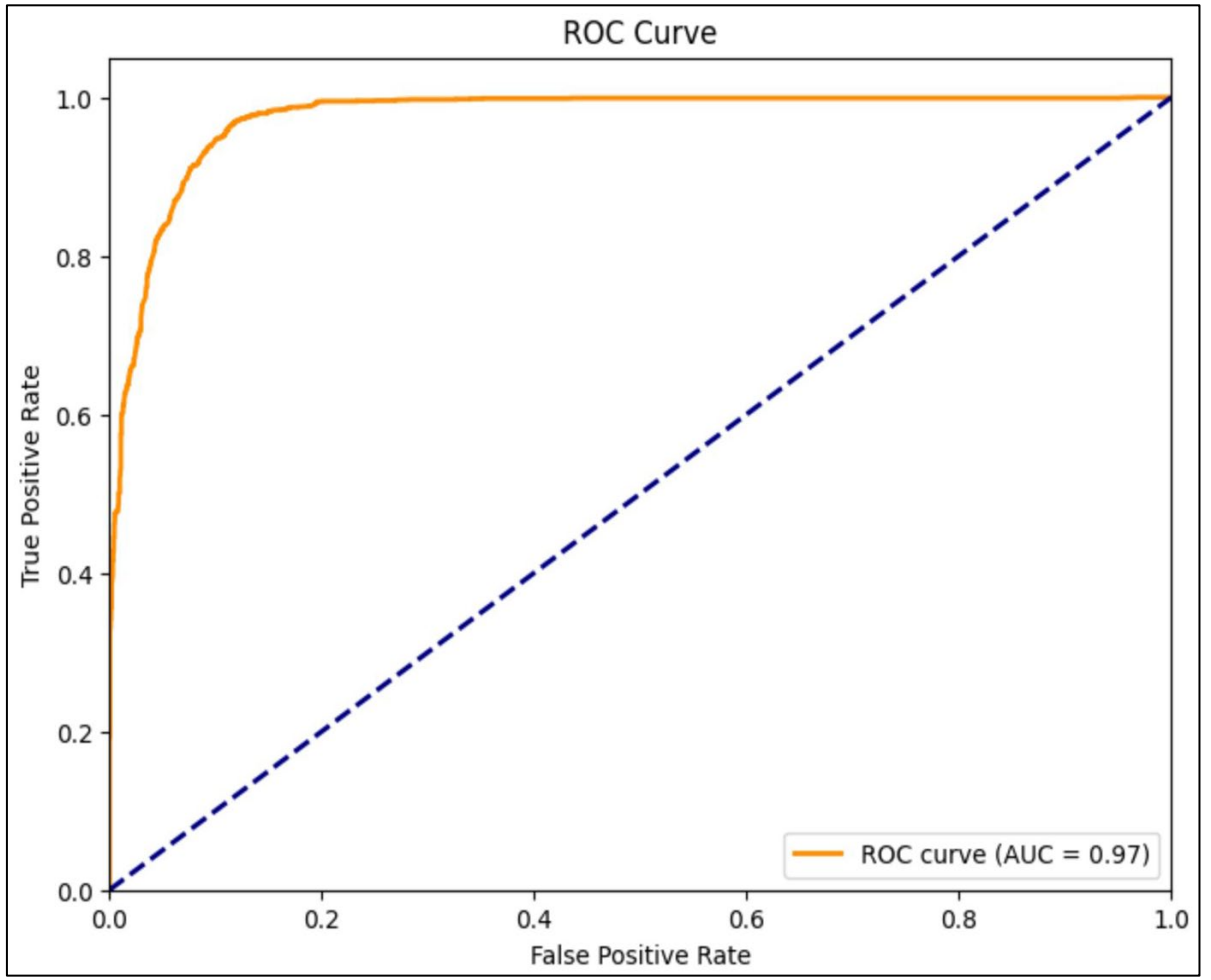
Feature Importance Analysis:

- Top Predictor: Average hits per plane (highest weight due to distinct interaction densities)
- Secondary Predictor: Angular standard deviation (σ_θ), confirming the value of spatial track topology
- Tertiary Predictor: Maximum hits per plane, contributing via the identification of dense hadronic shower cores



Classification Algorithm: BDT





Comparative Analysis of the Models

Metrics	CNN	XGBoost
Accuracy	0.92	0.93
Precision	0.91	0.91
Recall	0.92	0.96
F1-score	0.92	0.93

Outcome Prospects for Sampo

Transition to New Software: Integration of current hit clustering and PID algorithms into Sampo (the upcoming SPD event simulation and reconstruction package)

Algorithm Optimization:

- Scaling up the training and validation datasets significantly to boost model robustness
- Detailed multi-parametric efficiency analysis

Hybrid Approaches for Optimal μ/hadron Separation: Combining machine learning techniques (CNN, XGBoost, etc.) with classical tracking methods (e.g., Kalman Filter)

Experimental Validation: Using real experimental data from the RS prototype for final algorithm verification and performance cross-checks

Conclusion

Key Achievements (Implemented in SpdRoot):

- Geometry and Simulation: Full modeling and event visualization of the RS prototype completed
- Clustering Framework: C++ implementation of DBSCAN integrated into SpdRoot and validated against Monte Carlo data
- Machine Learning Baselines: Initial CNN and XGBoost models developed, proving high feasibility (AUC ≥ 0.96)

Future Roadmap (Transition to Sampo Package):

- CNN Optimization: Comprehensive tuning of network architecture, kernel sizes, and activation functions
- XGBoost Optimization: Advanced feature selection based on relative importance metrics
- Hybrid Architectures: Developing a combined classifier merging CNN spatial analysis with gradient boosting efficiency
- Energy Dependence Studies: Investigating classification accuracy across diverse energy intervals to implement correction mechanisms

Thank you for your attention!