Λ_c^+ production in pp collisions at $\sqrt{s} = 13.6$ TeV

Ing. Karla Žertová | Supervisor: Ing. Katarína Křížková Gajdošová, Ph.D.

Czech Technical University in Prague — Faculty of Nuclear Sciences and Physical Engineering

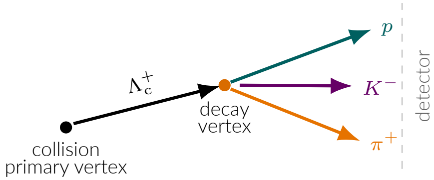
FACULTY OF NUCLEAR SCIENCES AND PHYSICAL ENGINEERING CTU IN PRAGUE

1. The Challenge: A Particle That Vanishes

Can a three-class machine-learning model recover short-lived Λ_c^+ baryons from four billion collision events?

Charm quarks are produced in the initial hard scattering. A charm quark can bind with u and d quarks to form $\Lambda_c^+(udc)$, a short-lived relative of the proton. Studying its production reveals how charm turns into observable matter.

With a mean proper decay length of only $c\tau \approx 60\ \mu\mathrm{m}$, the Λ_c^+ vanishes before reaching the detector. This analysis uses the decay $\Lambda_c^+ \to pK^-\pi^+$.



The detector sees the three daughter tracks, not the parent particle. Their combined trajectory and mass reveal possible Λ_c^+ candidates.

Two Look-Alikes, Three Classes

- **Combinatorial background:** unrelated tracks accidentally imitate a $pK^-\pi^+$ decay.
- **Non-prompt Λ_c^+:** a genuine baryon from a beauty-hadron decay — the same particle, but with a different production history.

Goal: separate prompt signal, non-prompt signal and random background.

From Existing Cuts to This Thesis

Existing approach
Loose rectangular cuts treat each measured variable separately.

Thesis contribution
Momentum-specific, three-class BDTs combine correlations and run in the standard ALICE workflow.

2. The Method: From Labels to Real Data

Ten multiclass XGBoost BDTs were trained, one for each transverse-momentum (p_{T}) interval, because decay observables and available statistics change significantly with momentum.

Signal view
Prompt and non-prompt labelled candidates come from detector-matched Monte Carlo simulation, where their origin is known.

Background view
Real-data mass sidebands describe accidental combinations and reduce the risk of teaching the model simulation-specific background.

Training, Validation and Deployment

1. Build three labelled classes
2. Split 80:20 in ten p_{T} bins
3. Train ten multiclass BDTs
4. Check ROC and score distributions
5. Choose a p_{T}-dependent threshold
6. Deploy through CCDB and O2

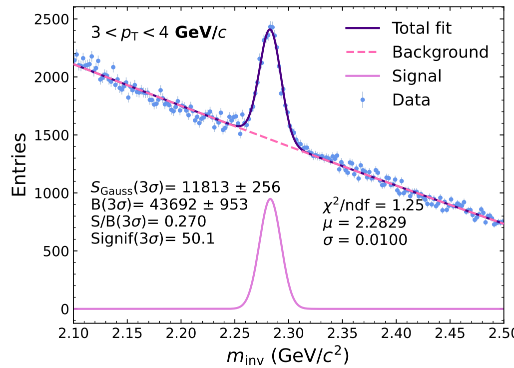
The split sample is used to train the models and assess classification quality. The deployed models are then run on an independent sample of 2024 collision data.

What the Model Uses — and What Is Checked

- **Inputs:** decay geometry, track displacement and particle identification.
- **Quality checks:** held-out score distributions and ROC curves test class separation and overtraining.
- **Deployment:** models are stored in the ALICE Conditions Database and executed in its standard O2 analysis software.

3. The Result: A Visible Signal

A genuine Λ_c^+ decay produces a narrow peak in the combined mass of the three daughter tracks, above a smooth random background.



$(11.8 \pm 3) \times 10^2$	50.1	0.270
raw signal candidates within $\mu \pm 3\sigma$	statistical significance $S/\sqrt{S+B}$	signal-to-background ratio

In the interval $3 < p_{\mathrm{T}} < 4\ \mathrm{GeV}/c$, the peak centre $\mu = (2.28 \pm 0.01)\ \mathrm{GeV}/c^2$ agrees with the known Λ_c^+ mass.

Evidence Beyond Training

- The deployed models reveal visible Λ_c^+ peaks in an independent 2024 real-data sample.
- The raw signal still combines prompt and non-prompt contributions and their separation belongs to the full production measurement.

Contribution

Delivered: a momentum-specific, three-class selection that is ready for standard ALICE Run 3 processing.

Next measurement: separate prompt and non-prompt contributions and correct the raw yields.

Scientific value: test charm-quark hadronisation and provide the pp reference needed to interpret heavier collision systems.