# Status of $\Lambda_c^+$ Reconstruction in the ePIC framework

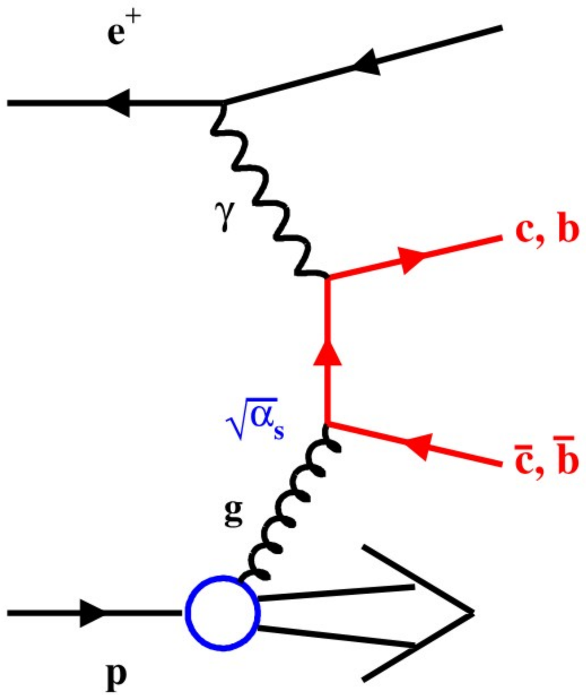
**Shyam Kumar***, Annalisa Mastroserio, Domenico Elia

INFN Bari, Italy

**Summer 2025 Joint EICUG/ePIC Collaboration Meeting**

## Boson-Gluon Fusion (BGF) is leading order [LO] mechanism

https://doi.org/10.1016/j.ppnp.2015.06.002




$$\gamma^* g \rightarrow c\bar{c} \text{ or } b\bar{b}$$

$$c \rightarrow D^0(c\bar{u}) \rightarrow K^-\pi^+$$

$$c \rightarrow \Lambda_c^+(udc) \rightarrow pK^-\pi^+$$

| Particle | Mass (GeV/$c^2$) | cτ (μm) |
|---|---|---|
| $D^\pm$ | 1.869 | 312 |
| $D^0$ | 1.864 | 123 |
| $B^\pm$ | 5.279 | 491 |
| $B^0$ | 5.280 | 456 |
| $\Lambda_c^+$ | 2.286 | 60 |

**Study includes $\Lambda_c^+$ and $\Lambda_c^-$ both**

Virtual photon (γ*) from the electron interacts with a gluon from the proton, produces $c\bar{c}$ or $b\bar{b}$ pair

**Additional Next to Leading Order (NLO) Mechanisms**

Gluon splitting, QCD Compton Scattering

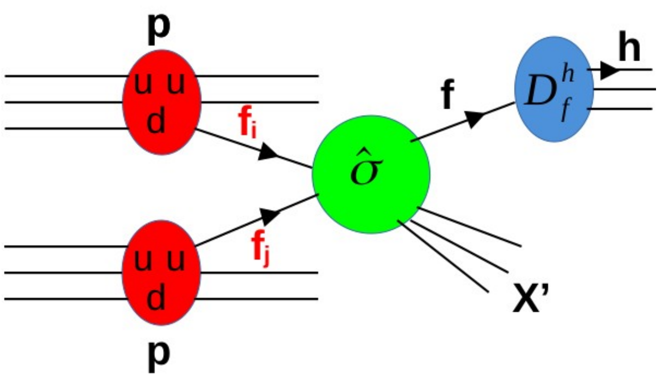
Heavy-flavor hadrons production cross section in nucleon-nucleon collisions using the factorization approach:

$$\frac{d\sigma^{NN\rightarrow H_Q X}}{dp_T^{H_Q}}(\sqrt{s_{NN}}, M_Q, \mu_F^2, \mu_R^2) = \sum_{i,j=q,\bar{q},g} \boxed{f_i(x_1,\mu_F^2)\otimes f_j(x_2,\mu_F^2)} \otimes \boxed{d\sigma^{ij\rightarrow Q\bar{Q}}(\alpha_s(\mu_R^2), \mu_F^2, M_Q, x_1, x_2, \sqrt{s_{NN}})} \otimes \boxed{D_Q^{H_Q}(z,\mu_F^2)}$$

https://doi.org/10.22323/1.449.0261

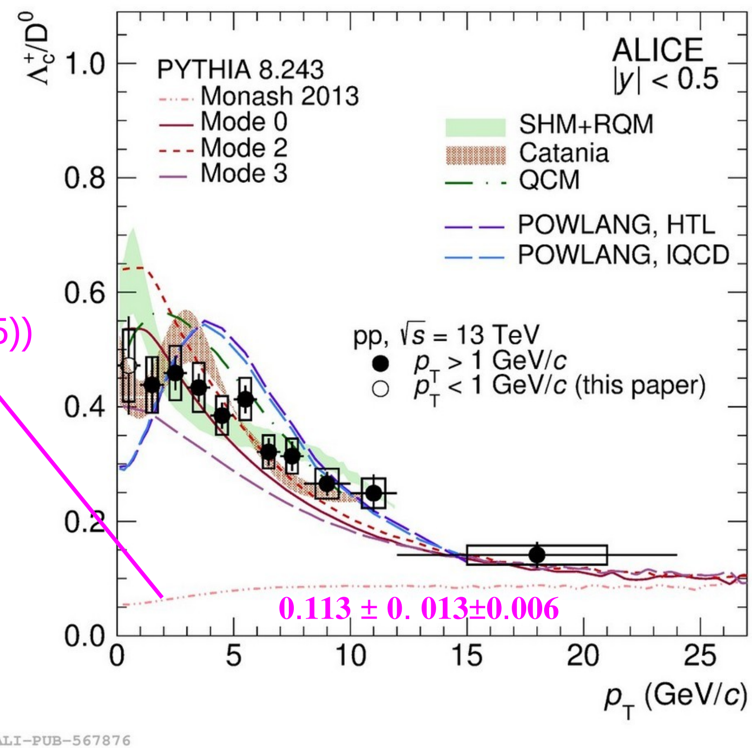
**Parton distribution functions** **Hard Scattering Cross-section** **Fragmentation function**

$$z = p_{H_Q}/p_Q$$




**x**= momentum fraction of parton
**μ$_F$**: Factorisation scale
**μ$_R$**: Renormalisation scale
**z** = $p_{HQ}/p_Q$

(LEP average, EPJC 75, 19 (2015))




**Fragmentation function:** A non-perturbative function assumed to be universal and extracted from e$^+$e$^-$ or ep collision data

**Charm/beauty baryon-to-meson ($\Lambda_c^+/D^0$) ratios:**

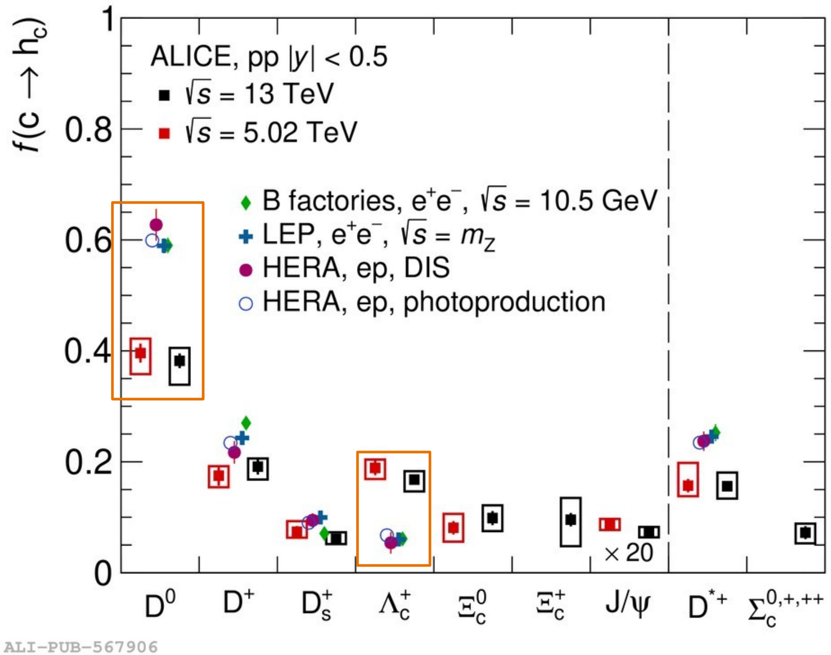
- Higher than measurements at e$^+$e$^-$, ep collisions for $p_T$ < 10 GeV/c
- Predictions that include baryon enhancement mechanisms describe data

JHEP 12 (2023) 086

$\Xi_c^0/D^0$ ratio underestimated by all the models, though Catania is close to the measurements

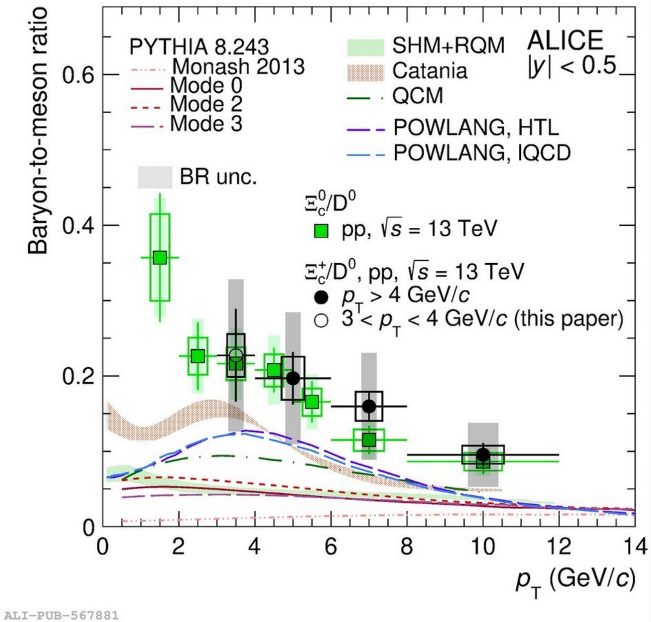
**Fragmentation fraction:** probability that a heavy quark hadronizes into a specific hadron species

**Violation of universality of fragmentation fractions among different collision systems**



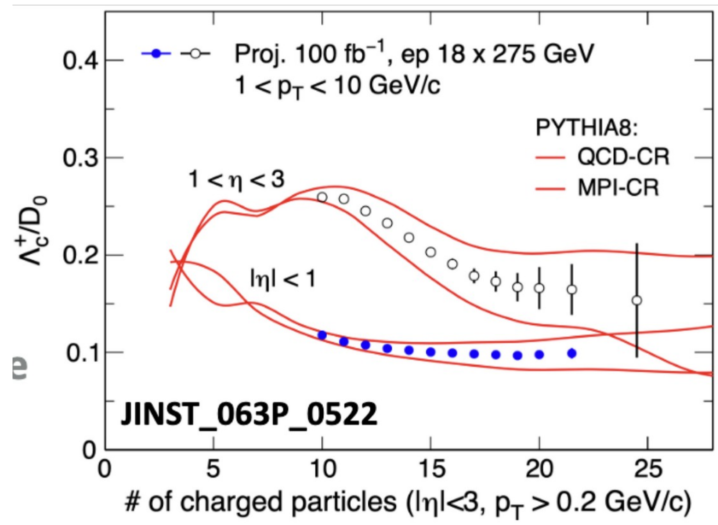

Dependence of Baryon-to-meson ratio on the collision systems

JHEP 12 (2023) 086




Hadron chemistry ($\Lambda_c/D^0$ ): impact at low-$p_T$ range and forward rapidity (ePIC)

**Main objective is to measure $\Lambda_c^+/D^0$ ratio in ep, eAu (10X100 GeV) DIS collisions with $Q^2 > 1GeV^2$**



**Secondary Vertex (Analytical approach)**

$$\vec{SV} = \frac{\vec{pca}_1 + \vec{pca}_2 + \vec{pca}_3}{3}$$

**SV for preliminary studies**
**Track errors are currently ignored**

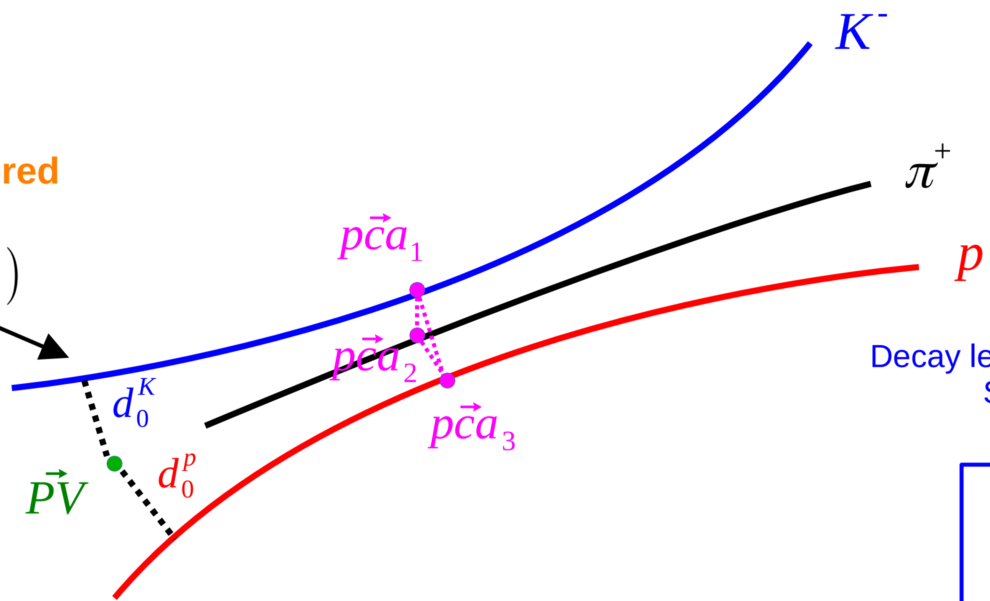
$$\Lambda_c^+ \rightarrow p\,K^-\pi^+$$

$DCA_{k\pi} = |\vec{pca}_1 - \vec{pca}_2|$, $DCA_{kp} = |\vec{pca}_1 - \vec{pca}_3|$, $DCA_{p\pi} = |\vec{pca}_3 - \vec{pca}_2|$

**$DCA_{12}$ = min {$DCA_{\pi K}$, $DCA_{KP}$, $DCA_{\pi P}$} Cut**

My Slides

$$Track_{DCA} = (\vec{r}_0, \vec{p}_0, q)$$




Decay length (dl), Primary Vertex (PV), Secondary Vertex (SV)

$$\vec{dl} = \vec{SV} - \vec{PV}$$

$$\cos\theta = \frac{\vec{dl}\,.\,\vec{p}_{\Lambda_c}}{|\vec{dl}||\vec{p}_{\Lambda_c}|}$$

$$DCA_{\Lambda_c} = dl\sin\theta$$

**Topological Variables:**

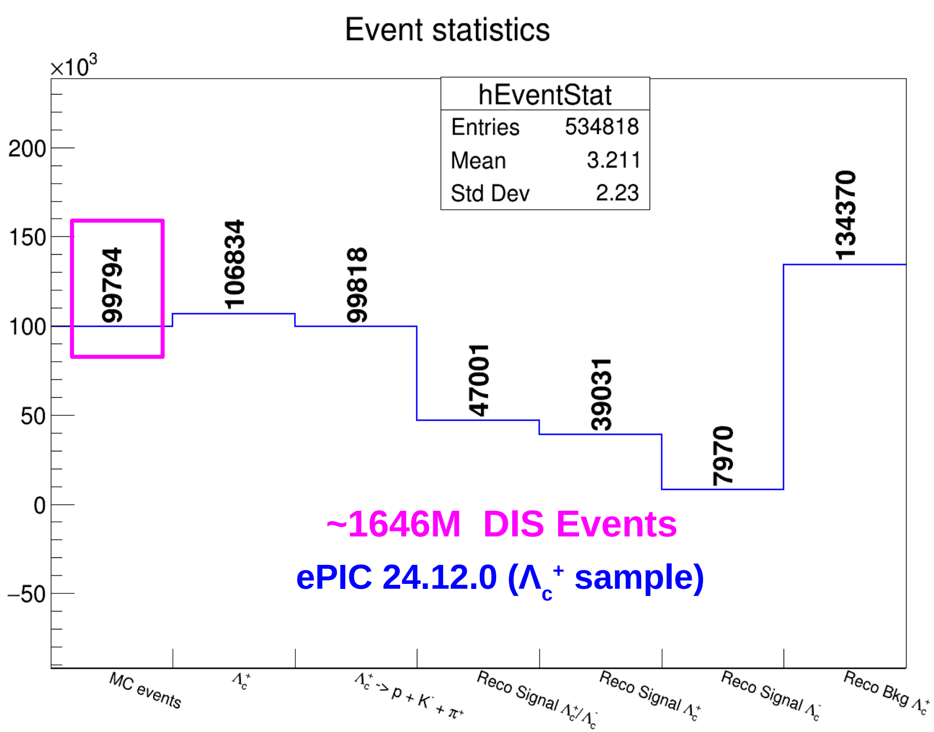
- $DCA_{k^-}$ and $DCA_{\pi^+}$ with respect to the reconstructed primary vertex (d0_k, d0_pi)
- Decay length of $\Lambda_c$ baryon (decaylength)
- Cosθ (angle between **dl** and **$p_{\Lambda c}$**)
- $DCA_{12}$ minimum distance between the daughter tracks
- $DCA_{\Lambda c}$ impact parameter of reconstructed $\Lambda_c$ baryon

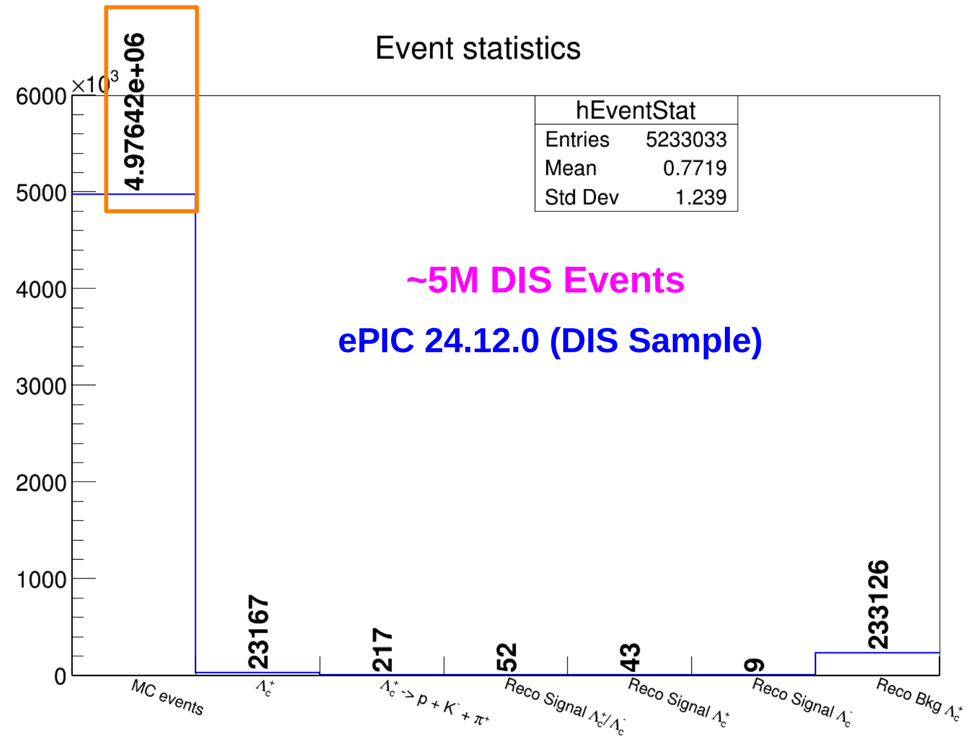
**d0xy: In the transverse plane**

**Use AdaptiveMultiVertexFinder or KFParticle for secondary vertexing, as they account for track errors**

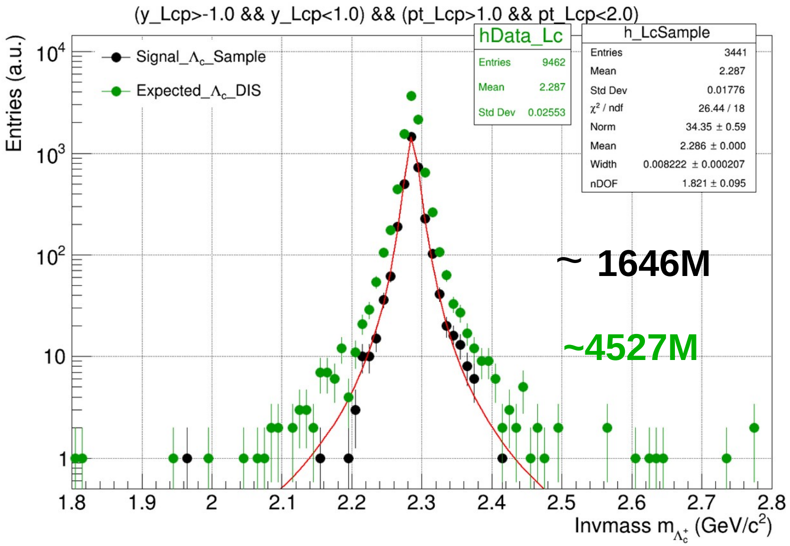
- Reconstruction of $\Lambda_c^+$ Baryon
  - ➔ $\Lambda_c^+$ sample **PYTHIA8 ep, NC, 10X100, $Q^2$ >1 GeV$^2$ events (~1646 M DIS): Events = 99,794**
  - ➔ DIS Sample: 24.12.0/epic_craterlake/DIS/NC/10x100/minQ2=1: **Total files 5180 and Events = 4,976,419**

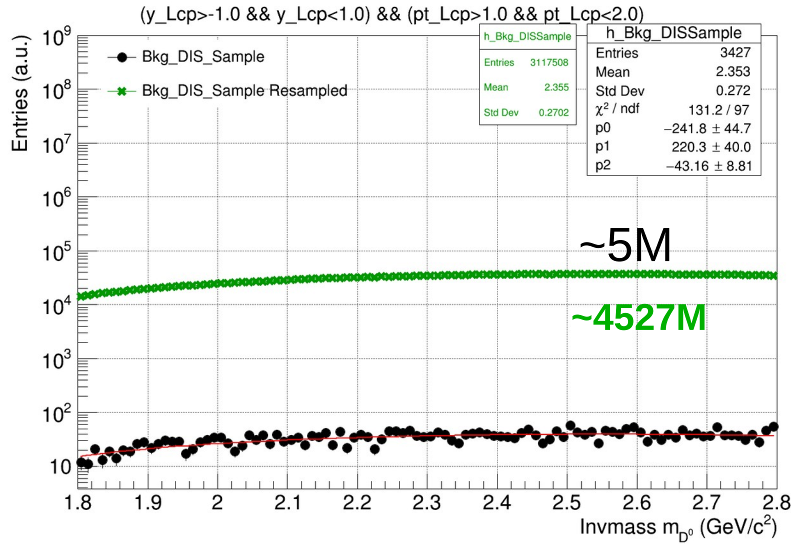
**Truth Particle Identification (PID)**



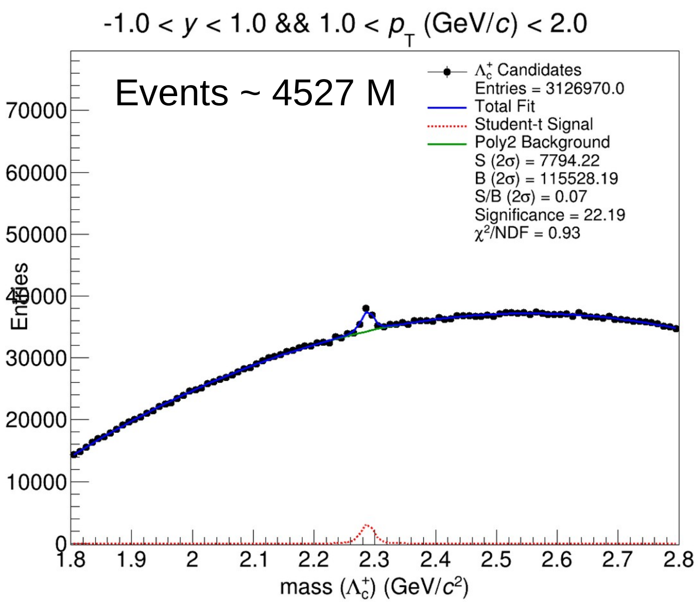



# Merging Signal and Background ($\Lambda_c^+$ Sample and DIS Events)




**Sampling**




**Topological selection (dimensions in mm)**

($m_{\Lambda c}$ > 1.8 && $m_{\Lambda c}$ < 2.8) && (d0xy_p>0.02 && d0xy_p<10.) && (d0xy_pi>0.02 && d0xy_pi<10.) && (d0xy_k>0.02 && d0xy_k<10.) && decay_length <100.

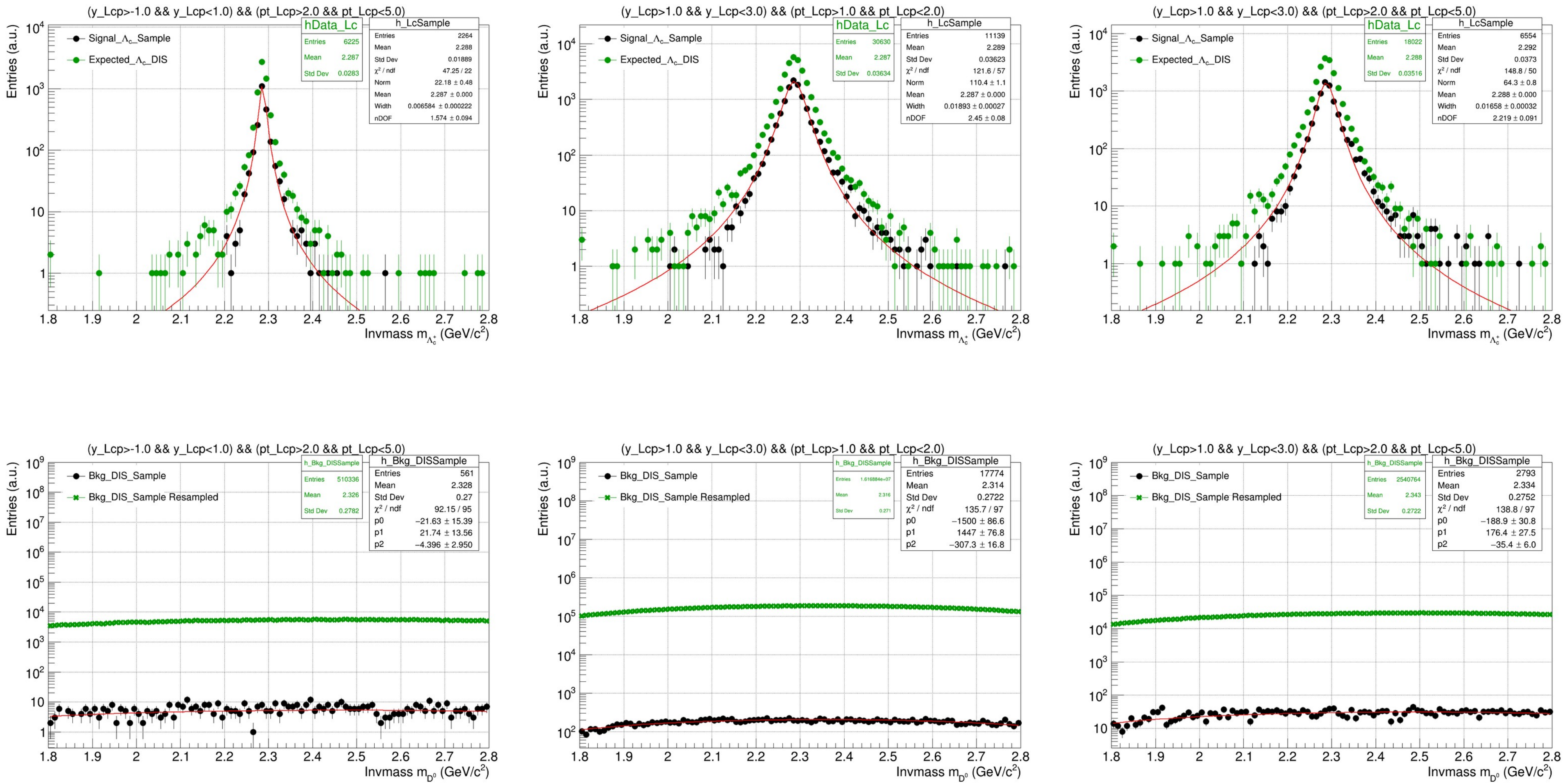
$$N_{events} = L_{int} \times \sigma$$

$$N_{events} = 10\,\mathrm{fb}^{-1} \times 4.527 \times 10^{5}\,\mathrm{pb}$$

$$\boxed{N_{events} = 10 \times 4.527 \times 10^{5} \times 10^{3} = 4527\,\mathrm{M}}$$

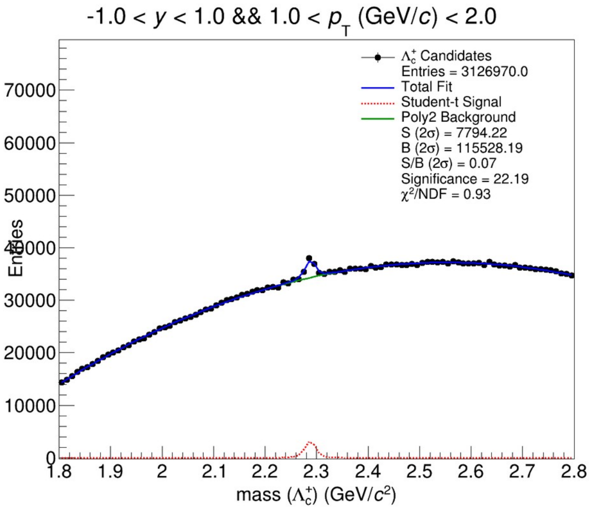
# Merging Signal and Background ($\Lambda_c^+$ Sample and DIS Events)

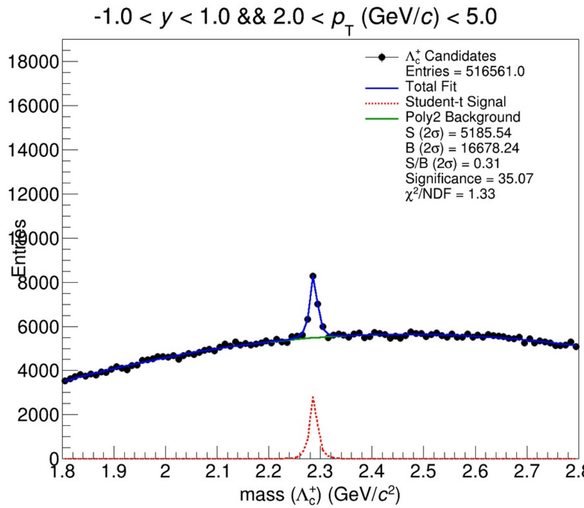
**~4527M (Sampled)**

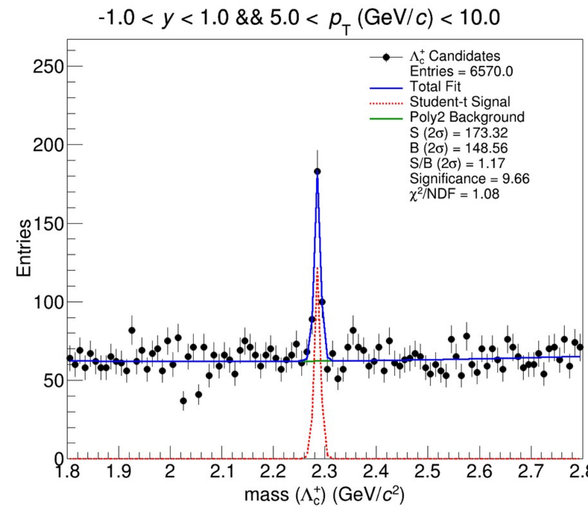


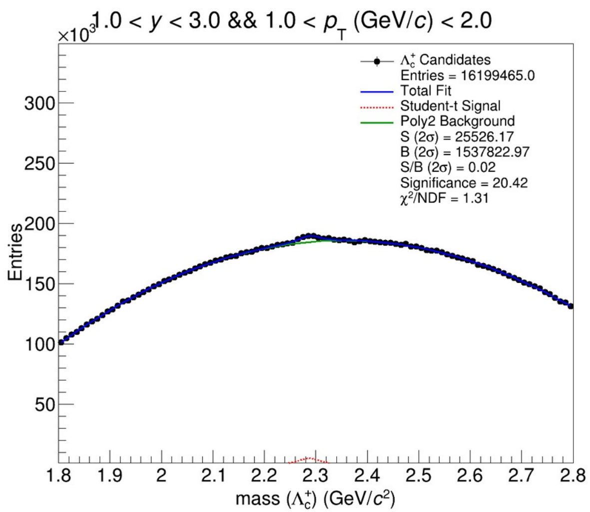
# Invariant Mass Plots ($\Lambda_c^+$)

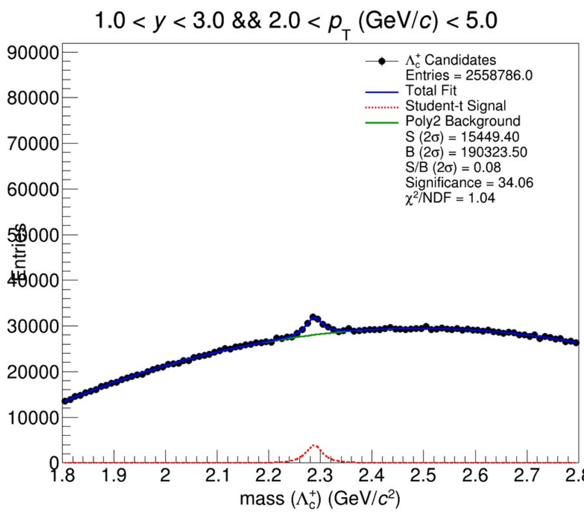
$L_{int} = 10\,\mathrm{fb}^{-1}$ $\quad$ $N_{events} = 10 \times 4.527 \times 10^5 \times 10^3 = 4527\,\mathrm{M}$

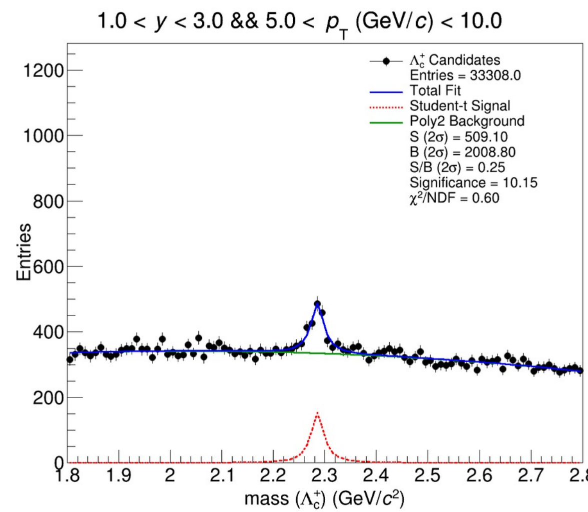


# Implementation of Machine Learning

# Boosted Decision Tree (BDT) Binary Classifier

hipe4ml package

https://doi.org/10.5281/zenodo.5070131

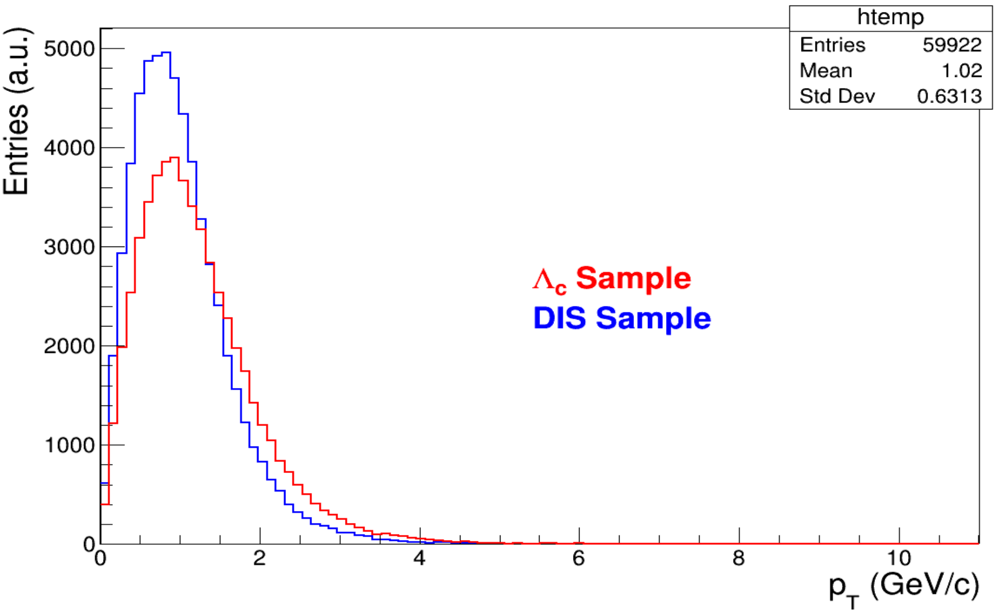
# Comparison of Bkg features between $\Lambda_c^+$ and DIS Sample

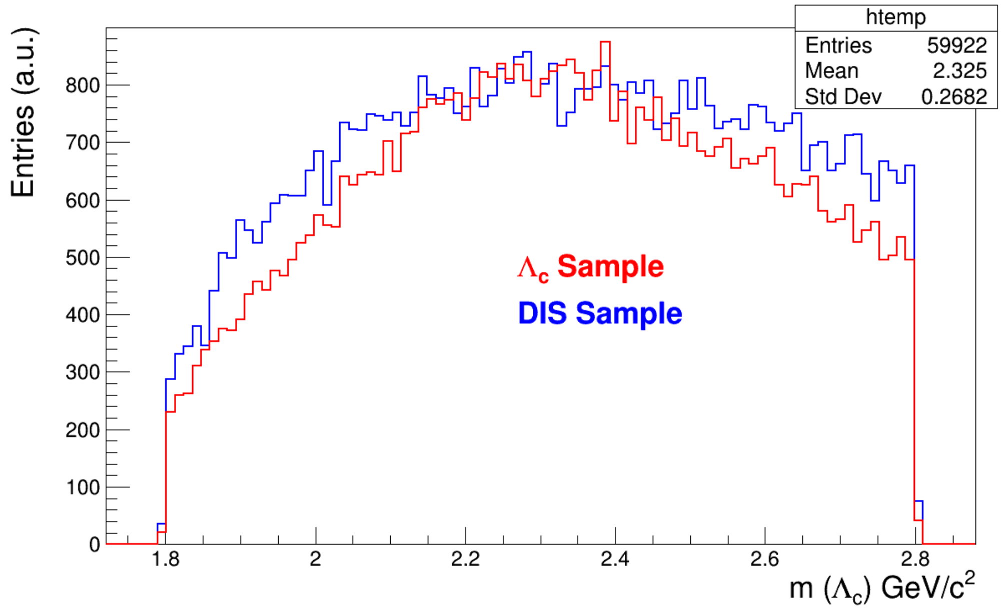
**The shape of features are different for the two backgrounds so they were not merged for ML**

**Strategy for Machine Learning**

**Signal:** Signal from $\Lambda_c^+$ Sample + Signal from DIS Sample

**Background:** Background from DIS Sample



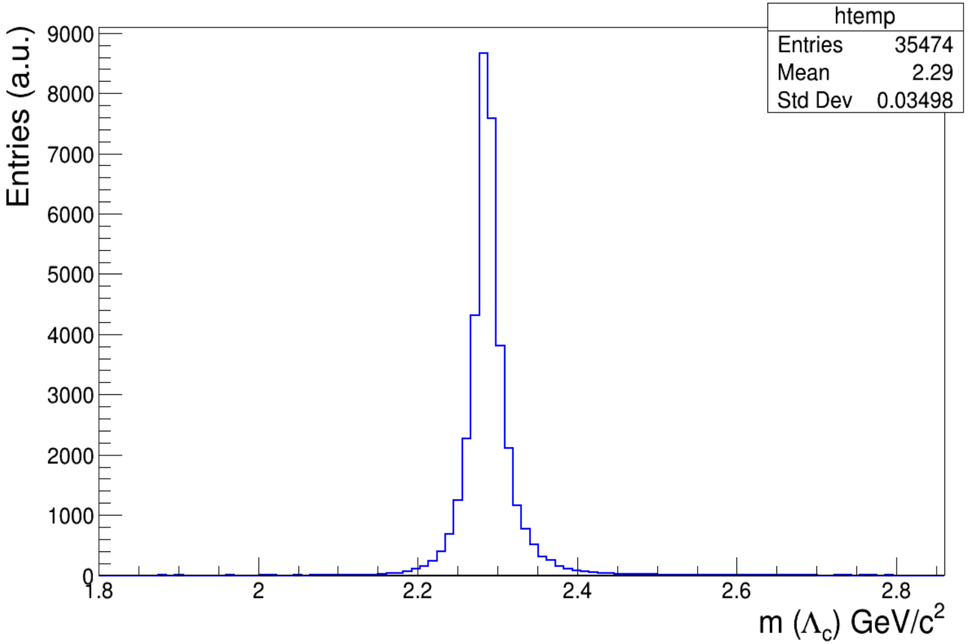
# Data Sample for ML

**Topological Preselection**

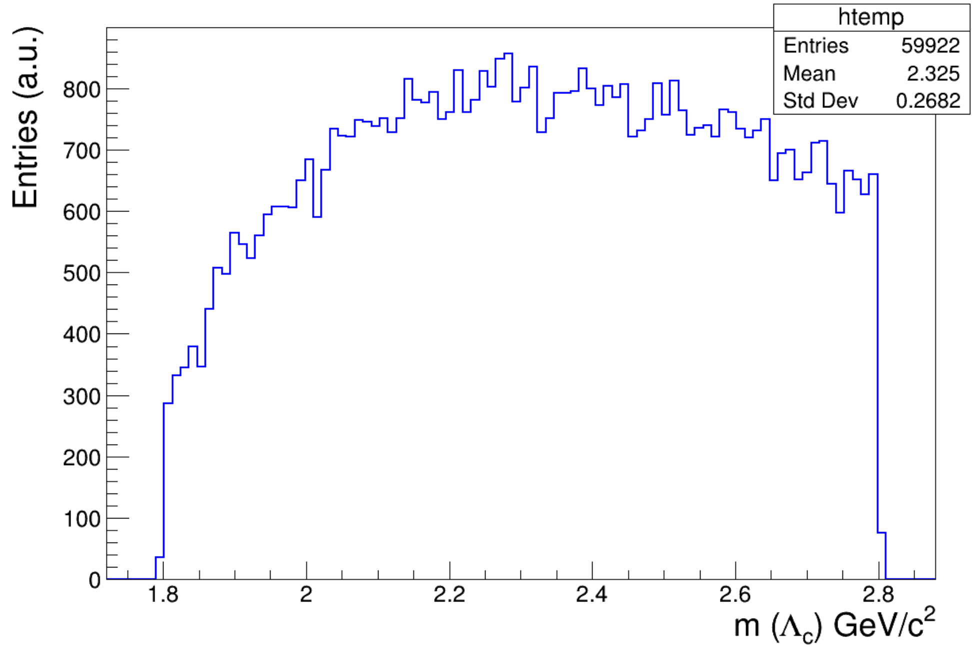
($m_{\Lambda c}$ > 1.8 && $m_{\Lambda c}$ < 2.8) && (d0xy_p>0.02 && d0xy_p<10.) && (d0xy_pi>0.02 && d0xy_pi<10.) && (d0xy_k>0.02 && d0xy_k<10.) && decay_length <100.

Training Sample: 80%, Test Sample: 20%

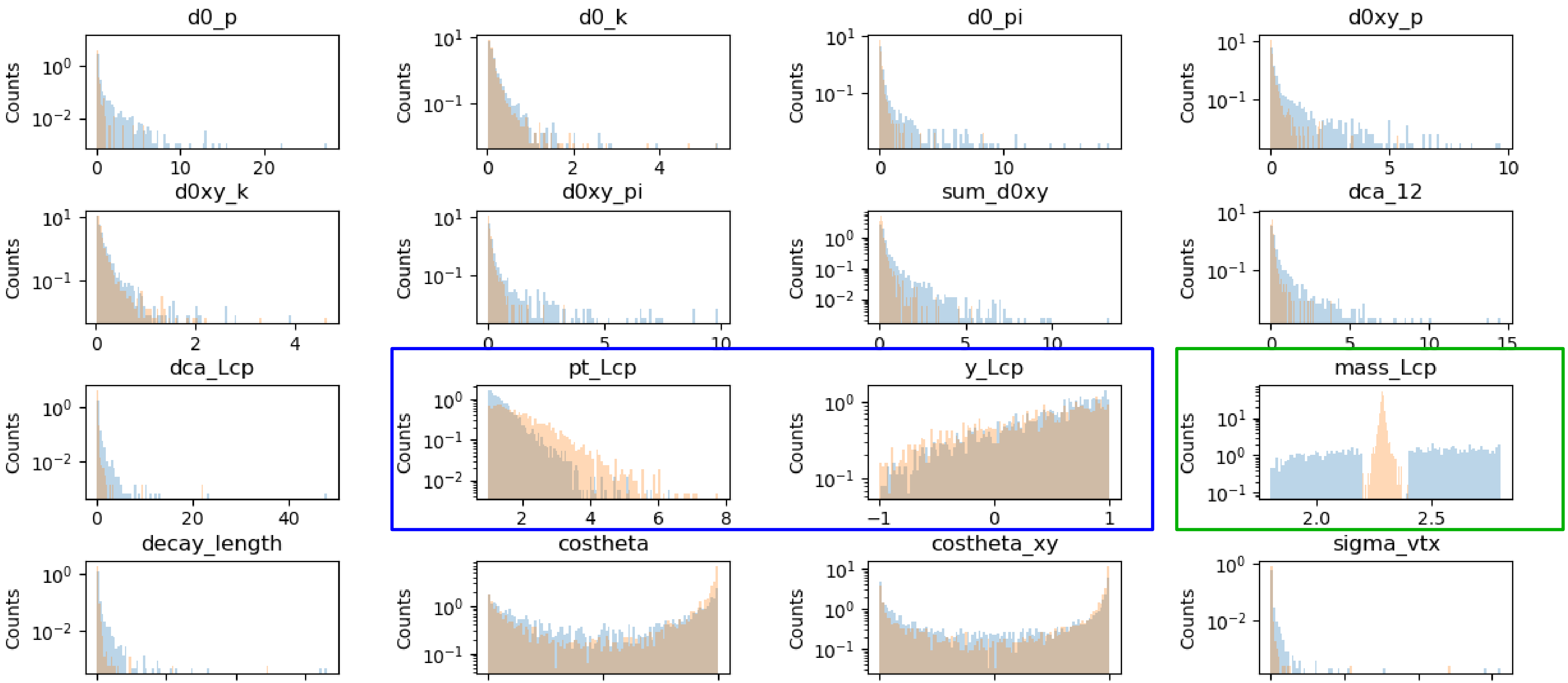
**Signal:** Signal from $\Lambda_c^+$ Sample + Signal from DIS Sample

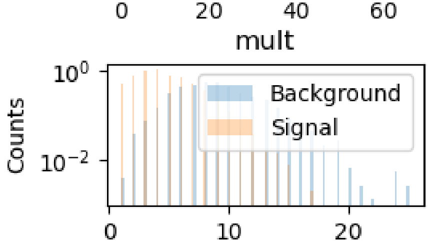



**Background:** Background from DIS Sample



# Signal and Background Distributions




**-1 < y < 1**

**1 < $p_T$ ($\Lambda_c^+$) < 10 GeV/c**

**Signal**

**Background**

# Correlation Matrix

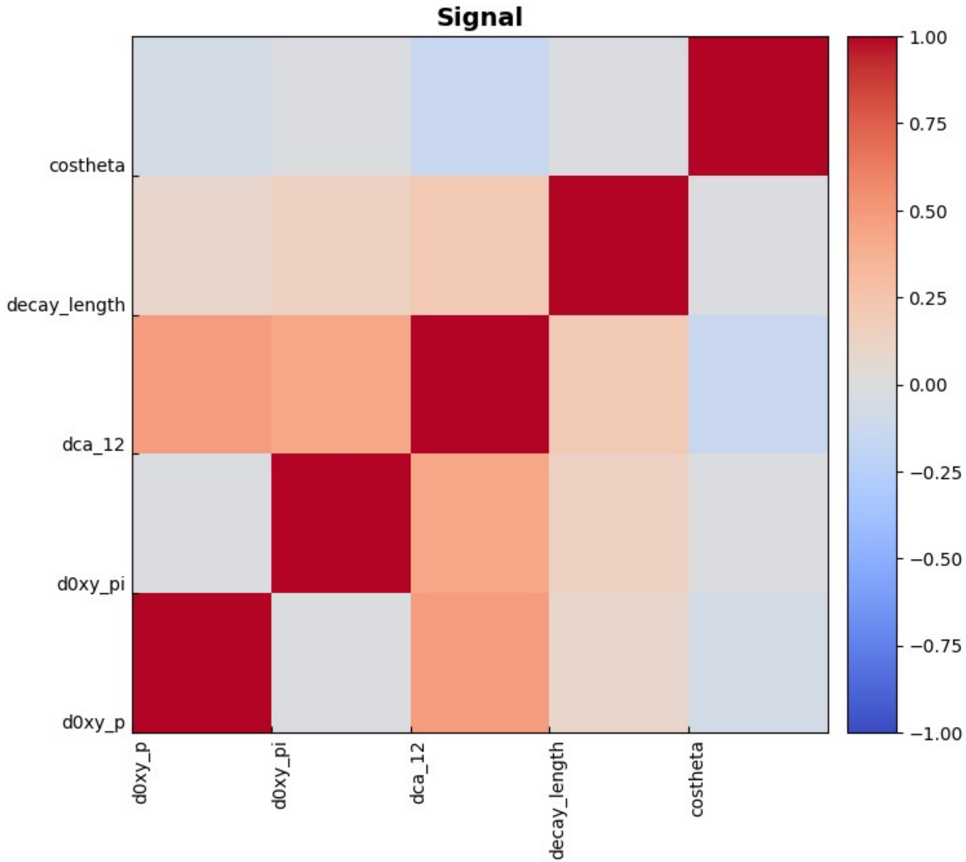
**$-1 < y < 1$**

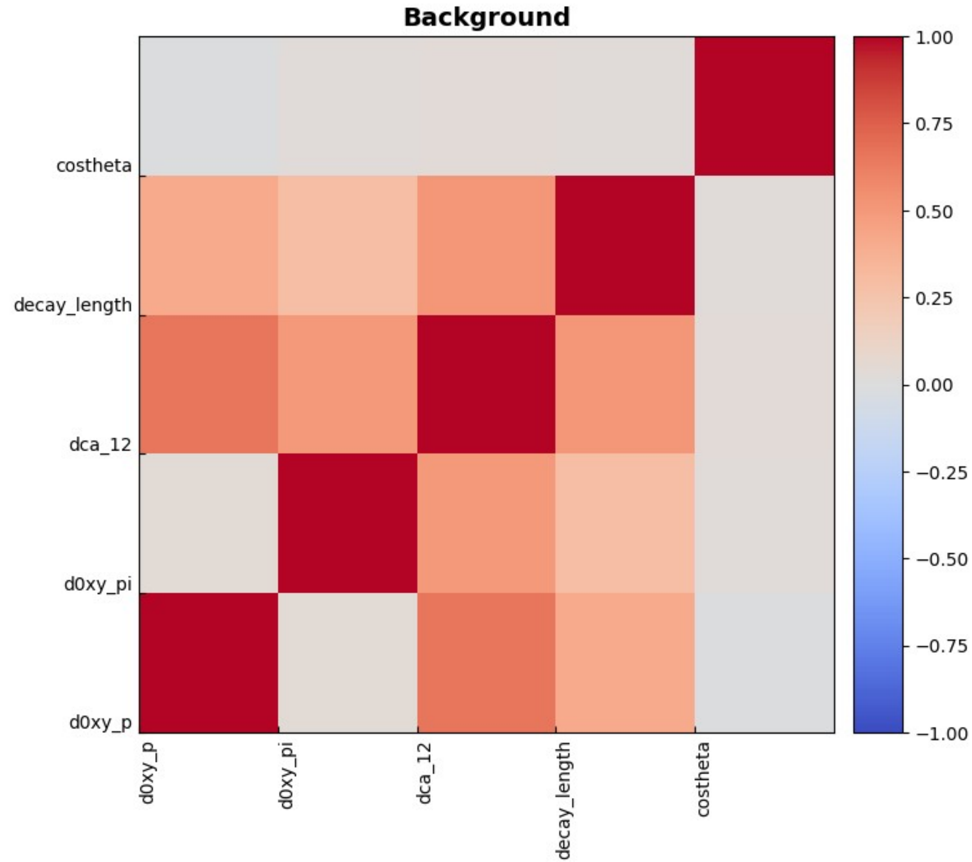
**$1 < p_T\ (\Lambda_c^+) < 10$ GeV/c**

**Signal candidates: 3113**

**Background candidates : 3113**

Keeping only a few features due to limited statistics and also analytical approach for secondary vertexing

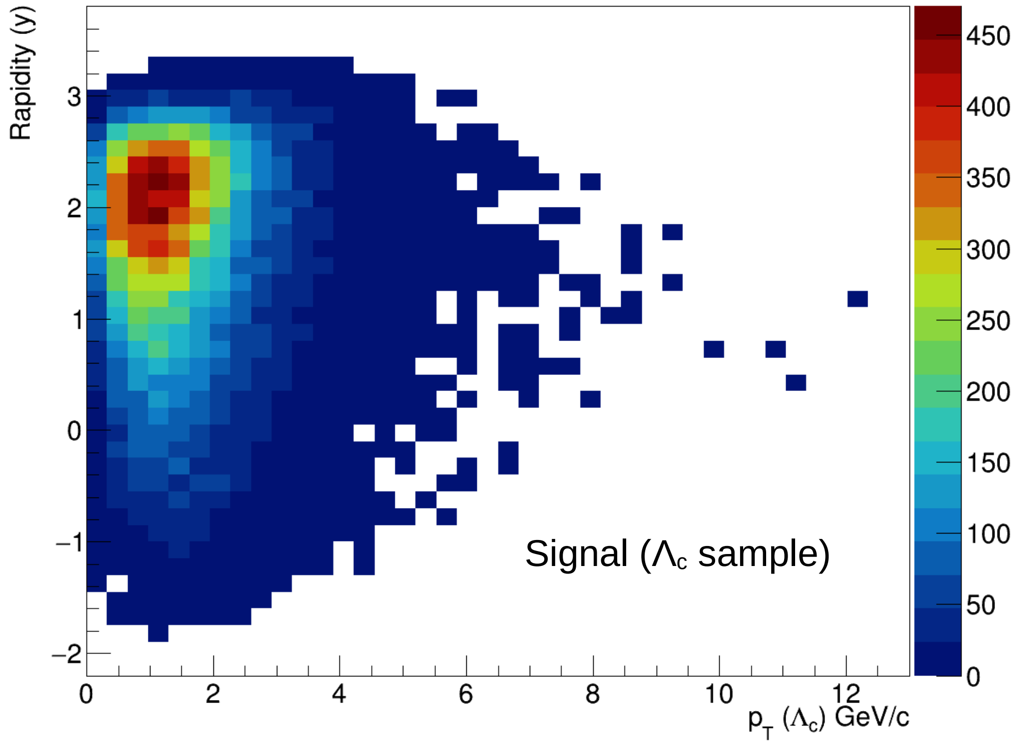


## ML approach for $\Lambda_c$ is implemented:

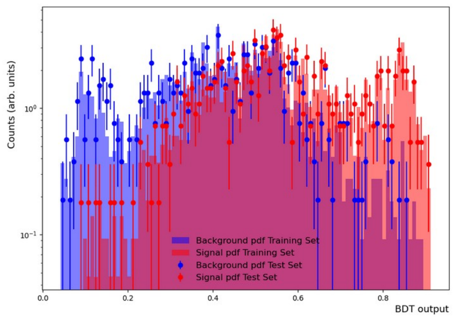
- Limited statistics due to a **small test sample**
- Poor DCA resolution at large rapidity ($c\tau \sim 60\ \mu m$ for $\Lambda_c$)
- Secondary vertexing needs to be improved
- The ML approach will be optimized once a new sample with larger statistics becomes available to achieve the best results

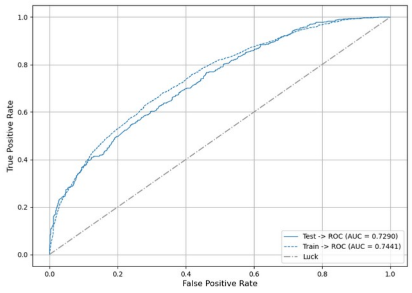
At low $p_T$: $\sigma(\mathrm{DCA}_p) > \sigma(\mathrm{DCA}_k) > \sigma(\mathrm{DCA}_\pi)$ Slide 5

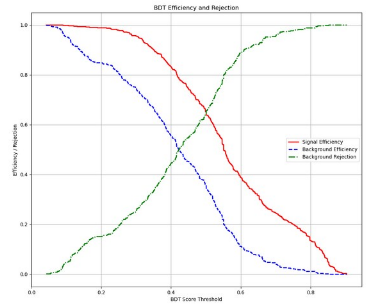



No ML: No BDT Cut

$$\mathrm{Significance}_{\mathrm{BDT\ Cut}} = \frac{S_{\mathrm{No\ ML}} \times \epsilon_{\mathrm{Signal}}}{\sqrt{S_{\mathrm{No\ ML}} \times \epsilon_{\mathrm{Signal}} + B_{\mathrm{No\ ML}} \times \epsilon_{\mathrm{Background}}}}$$



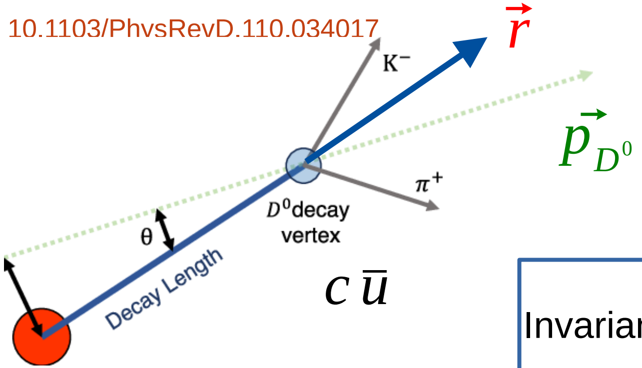

$$\left(\frac{S}{B}\right)_{\mathrm{BDT\ Cut}} = \left(\frac{S}{B}\right)_{\mathrm{No\ ML}} \times \frac{\epsilon_{\mathrm{Signal}}}{\epsilon_{\mathrm{Background}}}$$

# Summary and Future Plan

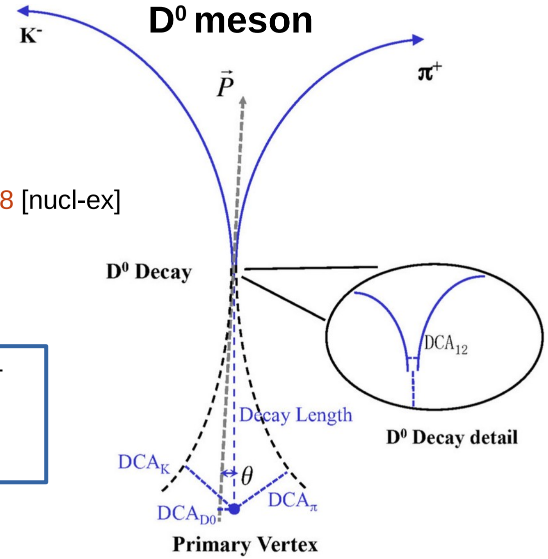
- Implemented the first version of the $\Lambda_c^+$ reconstruction code (will commit the code soon)
- Machine learning (ML) has also been developed but currently limited by insufficient statistics
- Future Steps:
  - Further improve the performance using secondary vertexing (e.g., KFParticle or AdaptiveMultiVertexFinder)
  - ML with a newly simulated large-statistics sample
  - Use realistic PID for protons once the new simulation is available
  - Extract the final results using ML in different y and $p_T$ bins once full simulated sample is available
  - Evaluate the $\Lambda_c^+$ baryon reconstruction efficiency using both preselection and BDT cut efficiencies
  - Estimate the $\Lambda_c^+/D^0$ ratio using machine learning
  - Implement additional models, such as neural networks (Binary classifiers as well as Autoencoders)

**Thank you for your attention!**

10.1103/PhysRevD.110.034017




arXiv:1911.12168 [nucl-ex]




cτ = 123 µm

Invariant mass: $m_{D^0} = \sqrt{(E_{K^-} + E_{\pi^+})^2 - (\vec{p}_{K^-} + \vec{p}_{\pi^+})^2}$

## Topological Variables:

- $DCA_{K^-}$ and $DCA_{\pi^+}$ with respect to the reconstructed primary vertex (d0_k, d0_pi)
- Decay length of $D^0$ meson (decaylength)
- Cosθ (angle between **dl** and $\mathbf{p}_{D0}$)
- $DCA_{12}$ distance between the daughter tracks of $D^0$
- $DCA_{D0}$ impact parameter of reconstructed $D^0$ meson
- $m_{D0}$ invariant mass of kaon and pion pairs
- pt_D0 reconstructed pt of the $D^0$ meson
- eta_D0 reconstructed η of the $D^0$ meson
- Multiplicity (mult)

Decay length (dl), Primary Vertex (PV), Secondary Vertex (SV)

$$\vec{dl} = \vec{SV} - \vec{PV}$$

$$\cos\theta = \frac{\vec{dl}\,.\,\vec{p}}{|\vec{dl}||\vec{p}|}$$

$$DCA_{D0} = dl\sin\theta$$

# Track Parametrization (Local to Global)

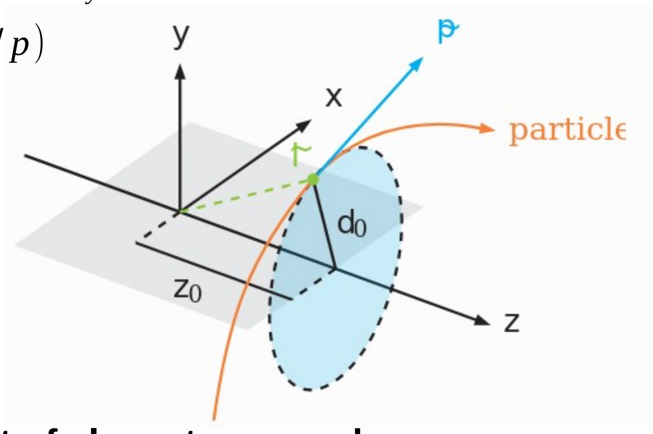
Helical Track model: $(l_0, l_1, \phi, \theta, q/p)$

$x = -l_0 \sin\phi, \quad y = l_0 \cos\phi, \quad z = l_1$

$p_x = p\cos\phi \sin\theta, \quad p_y = p\sin\phi \sin\theta, \quad p_z = p\cos\theta$

$charge = sign(q/p)$

$d_0 = l_0$
$z_0 = l_1$




**At Point of closest approach (perigee surface)**

$$(l_0, l_1, \phi, \theta, q/p)$$

**Global (Lab frame)**

$$(x, y, z, p_x, p_y, p_z, q)$$

```
Vector3 LineSurface::localToGlobal(const GeometryContext& gctx, const Vector2& lposition, const Vector3& direction) const

{

 Vector3 unitZ0 = lineDirection(gctx);

// get the vector perpendicular to the momentum direction and the straw axis

Vector3 radiusAxisGlobal = unitZ0.cross(direction);

Vector3 locZinGlobal = transform(gctx) * Vector3(0., 0., lposition[1]);

// add loc0 * radiusAxis

return Vector3(locZinGlobal + lposition[0] * radiusAxisGlobal.normalized());

}
```

## Calculation

**UnitZ0: is (0,0,1) vector along the z-axis for cylinder and disks.**

direction: (p Cos(phi) Sin(theta), p Sin(phi) Sin(theta), p Cos(theta))

radiusAxisGlobal = UnitZ0 Cross product direction = (-p Sin(phi) Sin(theta), p Cos(phi) Sin(theta), 0)

radiusAxisGlobal.Normalized = (-Sin(phi), Cos(phi), 0)locZinGlobal = (0,0,l1) (is same as global)

Global position = locZinGlobal + lposition[0] * radiusAxisGlobal.normalized() = (0,0,l1) + l0(-Sin(phi), Cos(phi), 0)Global Position = (-l0 Sin(phi), l0 Cos(phi), l1)

Returns the components, which we are using in HF analysis.

**$x = -l_0 \sin\phi$, $y = l_0 \cos\phi$, $z = l_1$**

# ML Output

$-1 < y < 1$ and $1 < p_T (\Lambda_c^+) < 10$ GeV/c

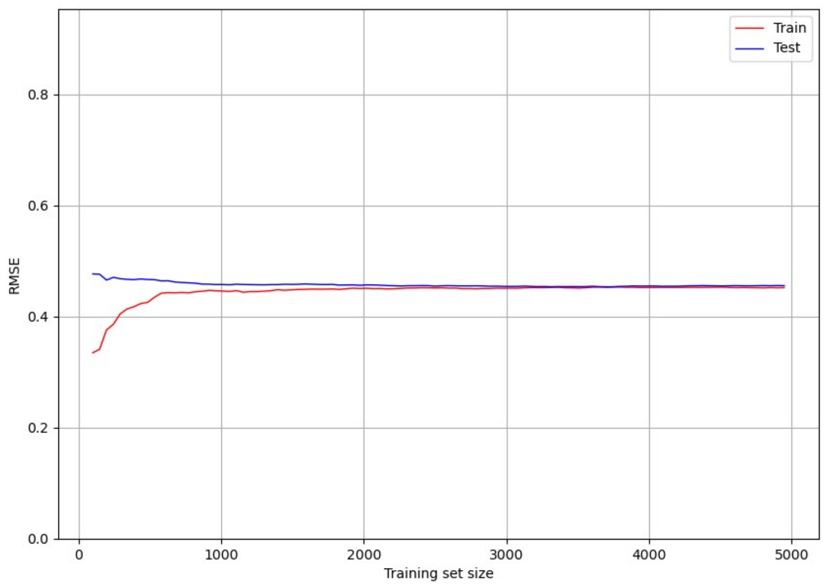
**Important features:**

- The training and test RMSE curves converge smoothly: Good generalization with low variance
- Model performance stabilizes beyond ~2000 training samples

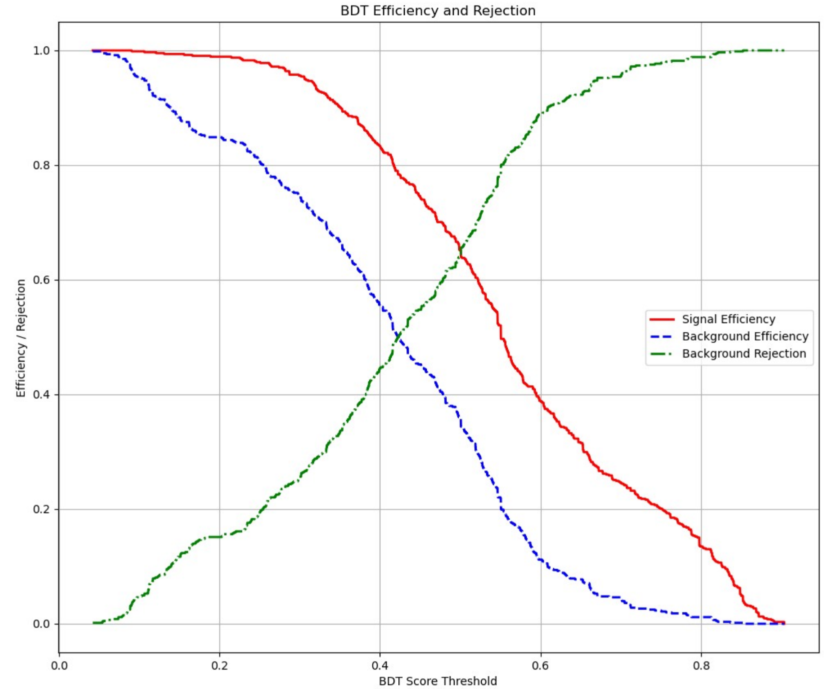
$$\text{Significance}_{\text{BDT Cut}} = \frac{S_{\text{No ML}} \times \epsilon_{\text{Signal}}}{\sqrt{S_{\text{No ML}} \times \epsilon_{\text{Signal}} + B_{\text{No ML}} \times \epsilon_{\text{Background}}}}$$

$$\left(\frac{S}{B}\right)_{\text{BDT Cut}} = \left(\frac{S}{B}\right)_{\text{No BDT Cut}} \times \frac{\epsilon_{\text{Signal}}}{\epsilon_{\text{Background}}}$$

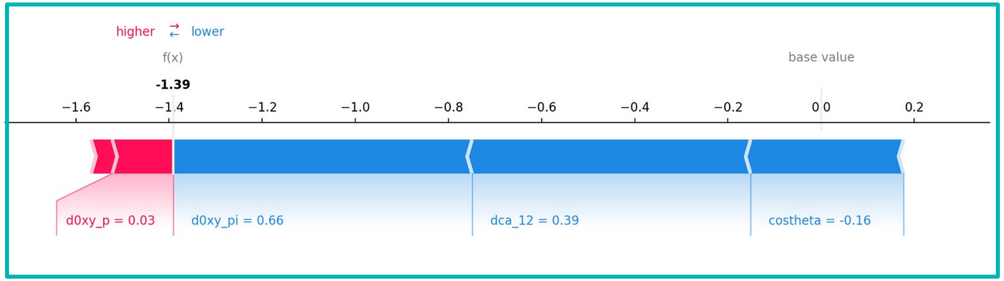
Root Mean Squared Error (RMSE)



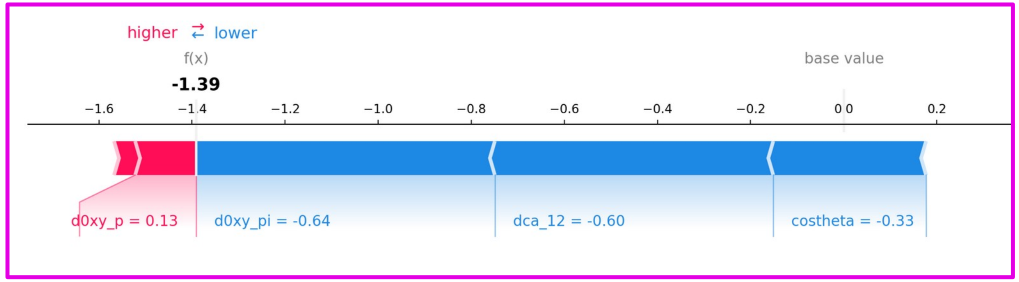

BDT Efficiencies



# Model's Prediction for a Single $\Lambda_c^+$ Baryon




**Feature values**




**SHAP Prediction from model for each feature**

Based on concept of Game Theory in Mathematics
SHAP (SHapley Additive exPlanations)

$$f(x) = \text{Base value} + \sum_{i=\text{features}} \text{SHAP}_i$$

--- Test Sample 0 ---

| **X_test:** | **feature values** | **SHAP Values from ML Model** |
|---|---|---|
| d0xy_p | 0.026224 | 0.13077936 |
| d0xy_pi | 0.655220 | -0.64195323 |
| Dca_12 | 0.389333 | -0.5982297 |
| decay_length | 0.305697 | 0.04526144 |
| Costheta | -0.158175 | -0.32825777 |
| **y_test: 0** | | **SHAP Sum: -1.3924** |

f(x): Model's prediction for a sample

Base value = Model's average prediction for entire data set

Sum of SHAP of individual features contributions

$$y_i = 0.0010833276 - 1.3924 = -1.3913167$$

**Base Value (Model's average prediction): 0.0010833276**

Model's Predicted Value: -1.3913167

Predicted Probabilities from model: [0.8008025  0.19919747]

SHAP Computed Probability: 1./[1+exp(-SHAP Predicted Value)]

Background = 0.8008023559279468, Signal = 0.19919764407205323

$$\text{Sigmoid function}: S(f(x)) = \frac{1}{1+e^{-f(x)}}$$

For the classifier f(x) is transformed in probability using sigmoid function

$$\text{Probability}: S(f(x)) = \frac{1}{1+e^{-(-1.3913167)}} = 0.19919$$

**If f(x) > 0: prediction pushes towards signal while for the other case as a background**

**Probability of being signal is ~ 20 % (prediction) while $y_{test}$ = 0**