A study of open-charmed hadrons in p-p collisions using machine learning methods

Based on:

Machine learning-based study of open-charm hadrons in proton-proton collisions at the Large Hadron Collider

K. Goswami, S. Prasad, N. Mallick, R. Sahoo, and G. Mohanty

Phys Rev D 110 034017 (2024)

Kangkan Goswami

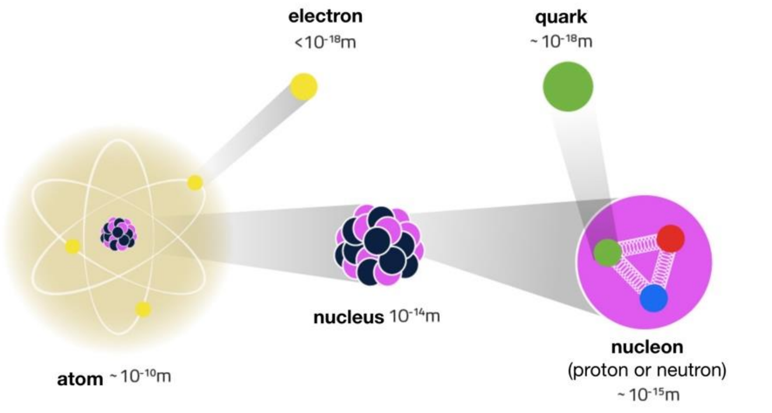
Department of Physics

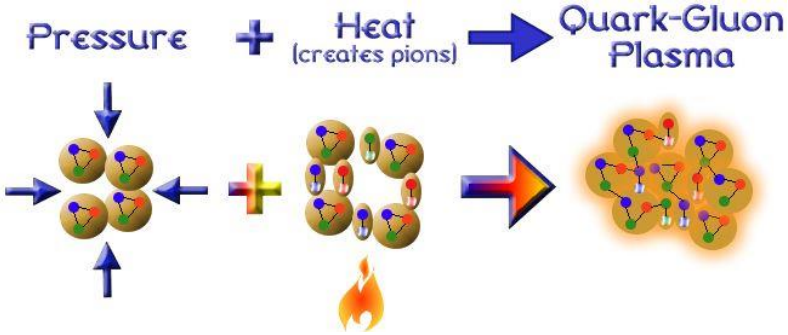
Indian Institute of Technology Indore

Outline

- Introduction
- Motivation
- Machine Learning Models
- Input Feature and training details
- Results
- Summary

Introduction

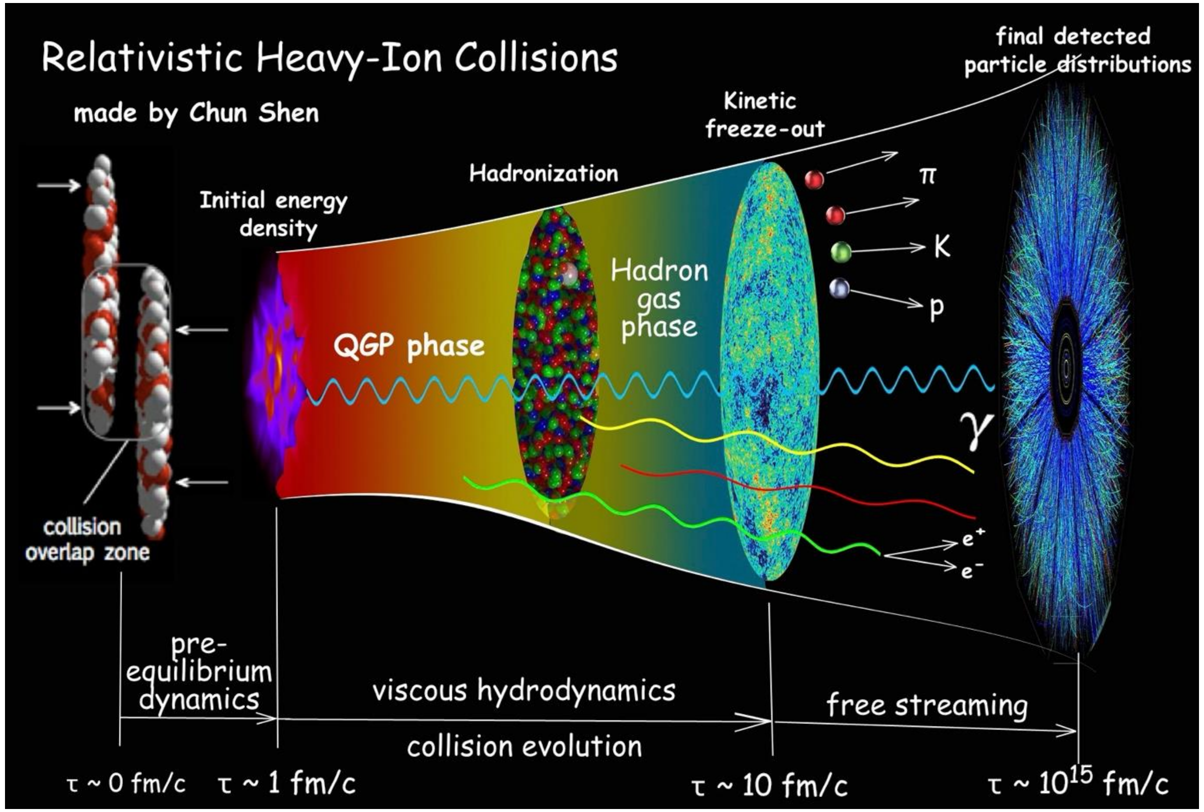




Under high temperature and high baryon density, a deconfined medium of thermalized quarks and gluons is predicted. It is termed as **Quark-Gluon Plasma (QGP)**

Source: Quark Matter 2018, http://scienzapertutti.infn.it , Quark Gluon Plasma, http://hep.itp.tuwien.ac.at/~ipp/qgp.html,

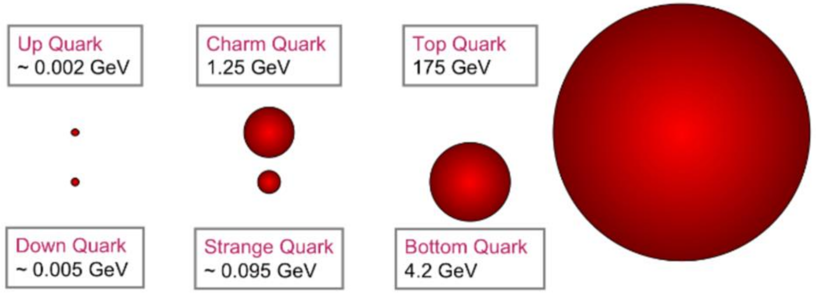
Producing Quark-Gluon Plasma in laboratory



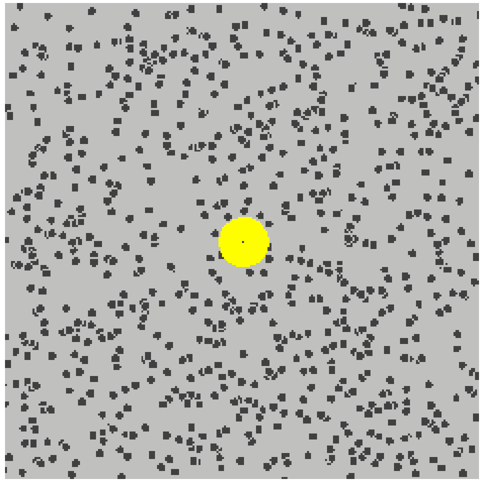
Sketch of relativistic heavy-ion collisions, Chun Shen, Ohio State University

Motivation

- To understand the medium formed in an ultra-relativistic heavy ion collision
- **Heavy quarks as probe**
  - Formed initially in the system
  - Mass >> Temperature of the medium
- Brownian Motion:
  - Charm quarks are much heavier than the constituent of the QGP medium, i.e. u, d, and s quarks
  - D Meson are relatively heavier than the major constituent of the hadronic medium, i.e. pion, kaon, and proton
- Can give us information about the medium through its study of the diffusion of heavy quarks and heavy-flavor hadrons
- This information can't be observed directly in experiments
- However, in principle this should affect experimental observables such as, elliptic flow (v_2) and nuclear modification factor (R_{AA})

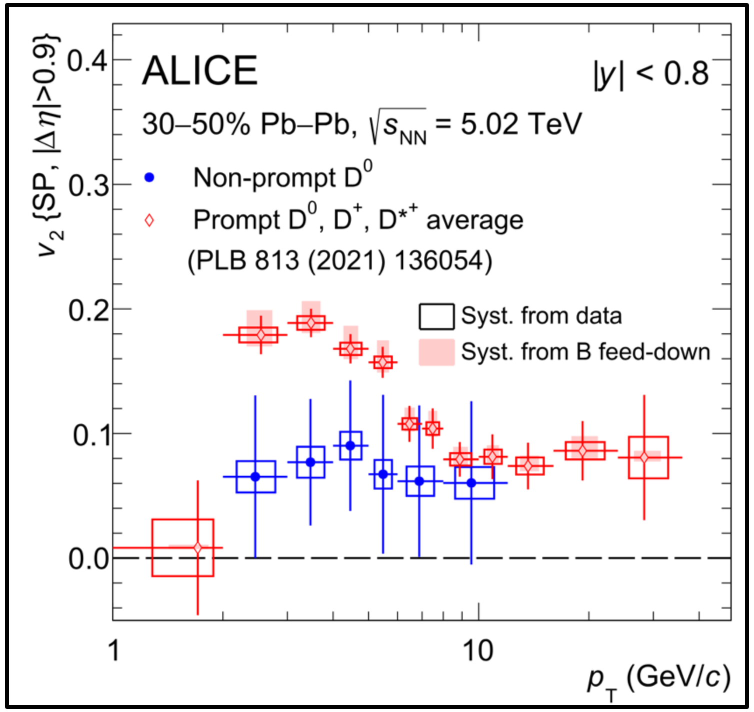


D. Kazakov, Phys. Usp. 57, 930 (2014)

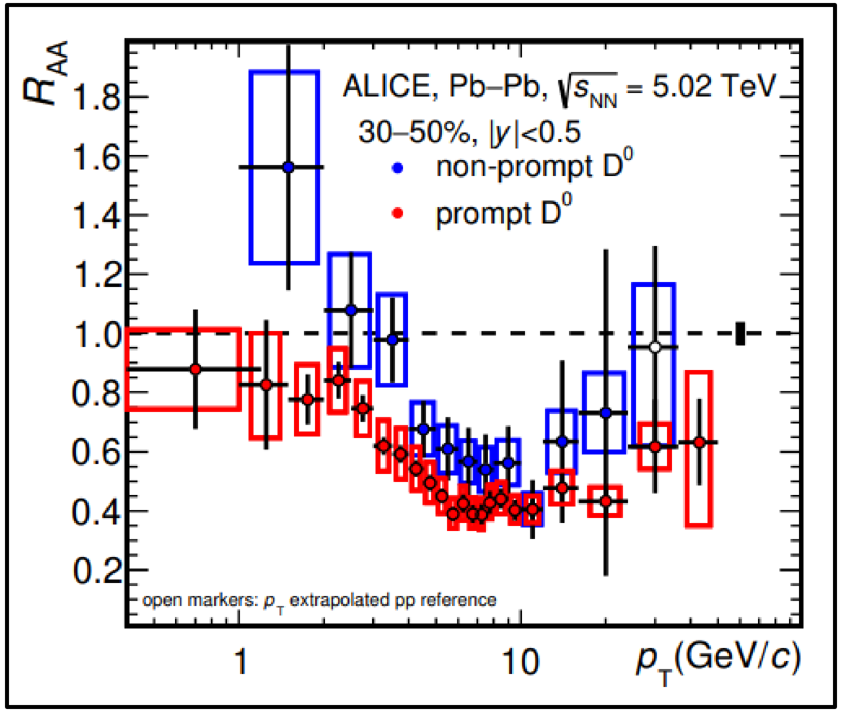


Source: Brownian motion, Wikipedia

Motivation



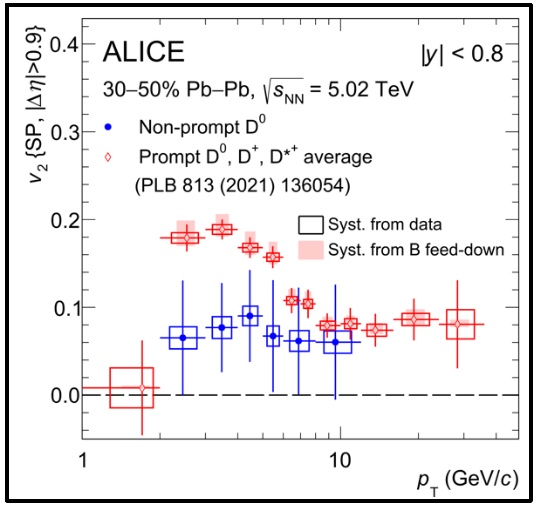
ALICE Collaboration, EPJC **83** 1123 (2023)



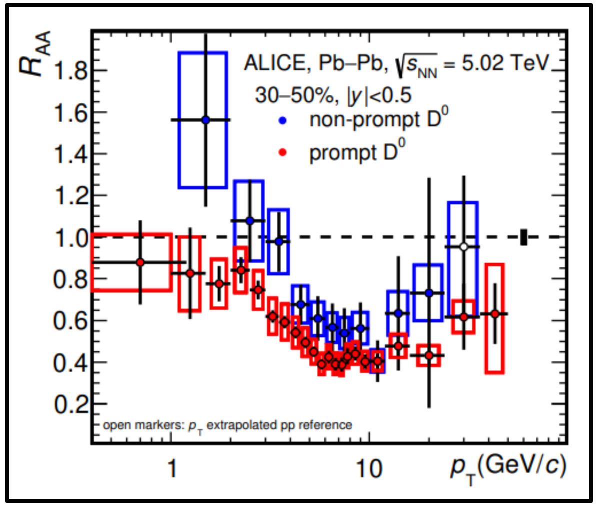
ALICE Collaboration, JHEP **12**, 126 (2022).

- However, to study the charm sector in experiment, we need to separate the contribution coming from beauty sector (i.e., Non-prompt D meson)

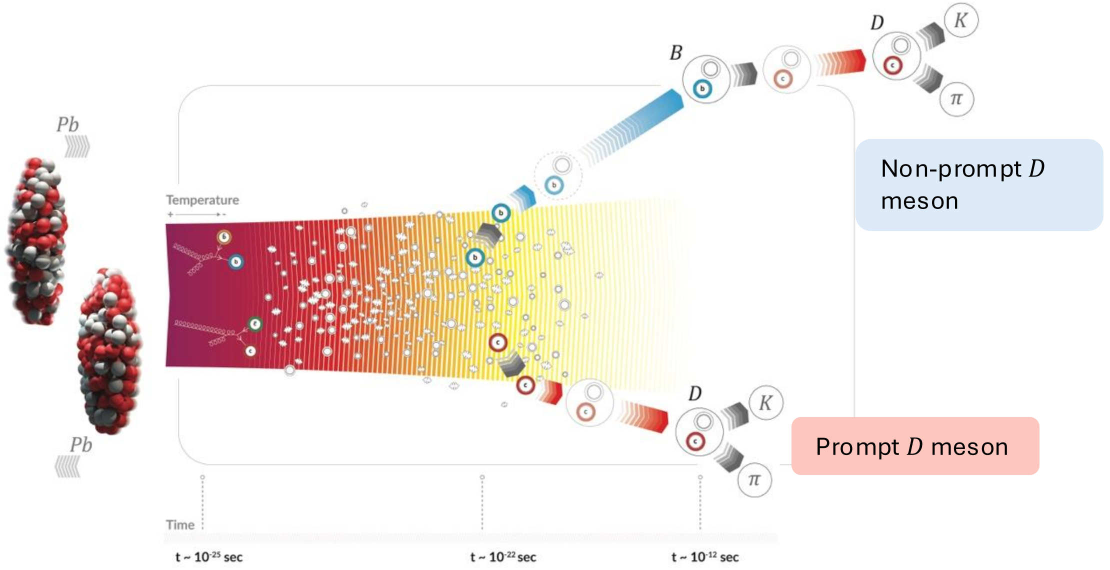
Motivation



ALICE Collaboration, EPJC **83** 1123 (2023)



ALICE Collaboration, JHEP **12**, 126 (2022).

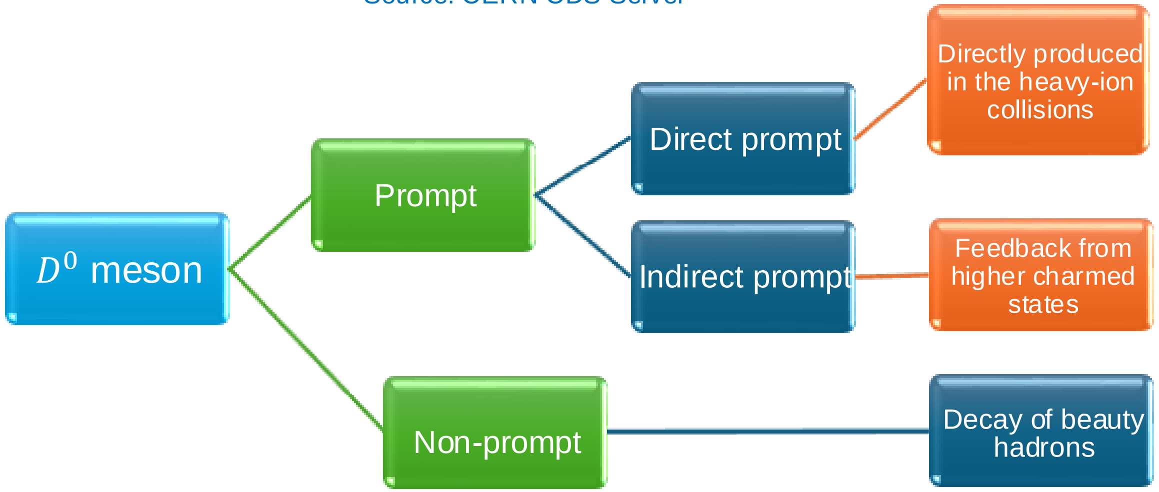


Source: CERN CDS Server

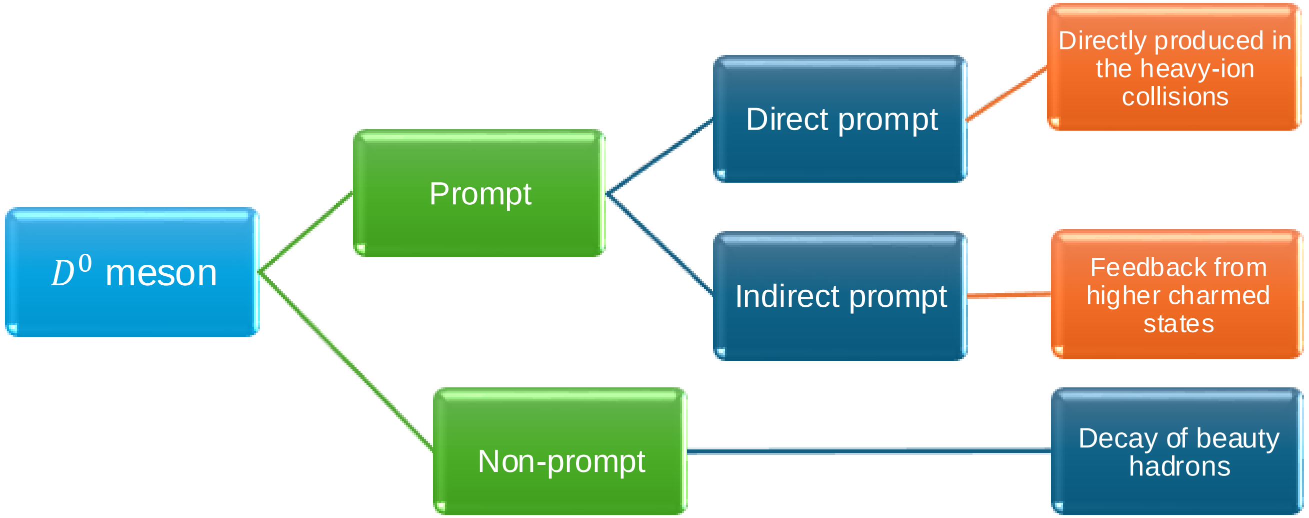
- However, to study the charm sector in experiment, we need to separate the contribution coming from beauty sector (i.e., Non-prompt D meson)

- Prompt D^0 meson: **Direct production and decay higher excited charm hadrons**

- Non-prompt D^0 meson: **Weak decay of beauty hadrons**



Prompt and non-prompt D^0 mesons: ML Approach

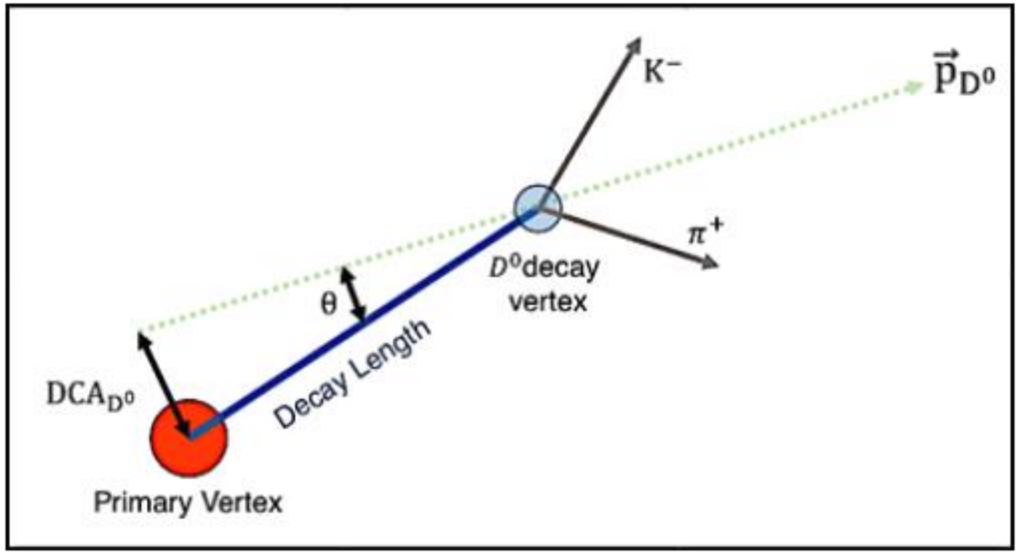


Machine Learning Algorithms used:

1. Extreme Gradient Boost (XGBoost)
2. Categorical Boosting (CatBoost)
3. Random Forest

Input Features

1. Invariant Mass
2. The pseudo-proper time: $$t_z = \frac{(z_{D^0} - z_{PV}) \times m_{D^0}}{p_z}$$
3. Pseudo-proper decay length: $$c\tau = \frac{c m_{D^0} \vec{L}.\overrightarrow{p_T}}{p_T^2}$$
4. Distance of closes approach: $$DCA_{D^0} = L \times \sin\theta$$



K. Goswami et al., Phys. Rev. D **110**, 034017 (2024)

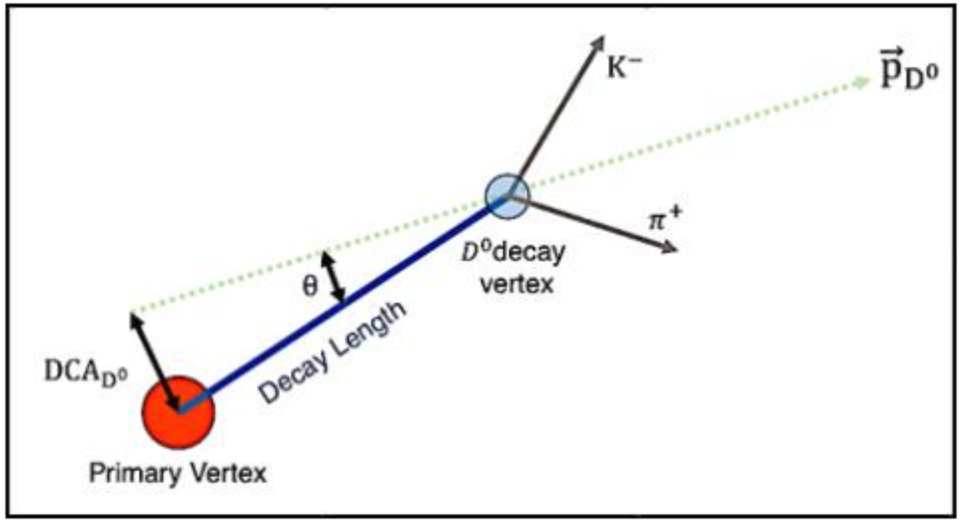
Input features

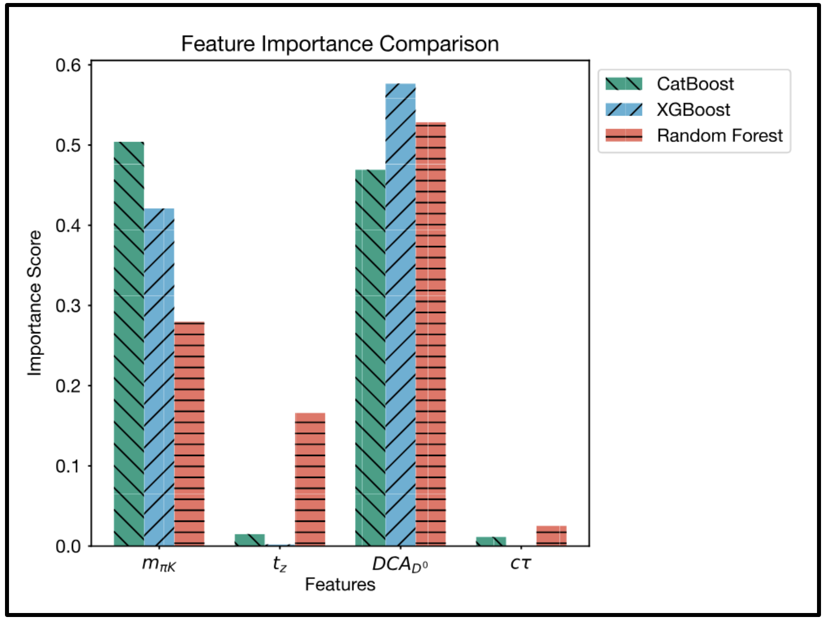
Input Variables

1. Invariant Mass
2. The pseudo-proper time: $t_z = \frac{(z_{D^0} - z_{PV}) \times m_{D^0}}{p_z}$
3. Pseudo-proper decay length: $c\tau = \frac{c m_{D^0} \vec{L}.\vec{p_T}}{p_T^2}$
4. Distance of closes approach: $DCA_{D^0} = L \times \sin\theta$

$\vec{L}$ is the vector pointing from the primary vertex towards D^0 decay vertex, i.e. $\vec{L} = \vec{V} - \vec{S}$
$\vec{V}$ is the position of the primary vertex and $\vec{S}$ is the position of the D^0 decay vertex given by,

$$S_i = \frac{(t_1 + d_{i,1}m_1/p_{i,1}) - (t_2 + d_{i,2}m_2/p_{i,2})}{m_1/p_{i,1} - m_2/p_{i,2}}$$

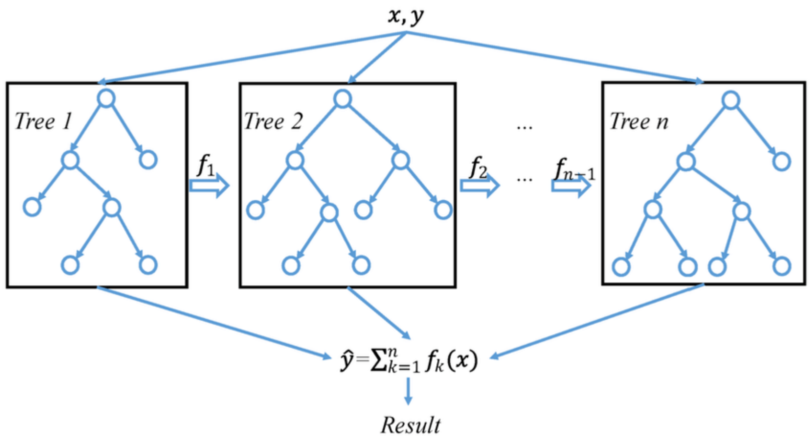




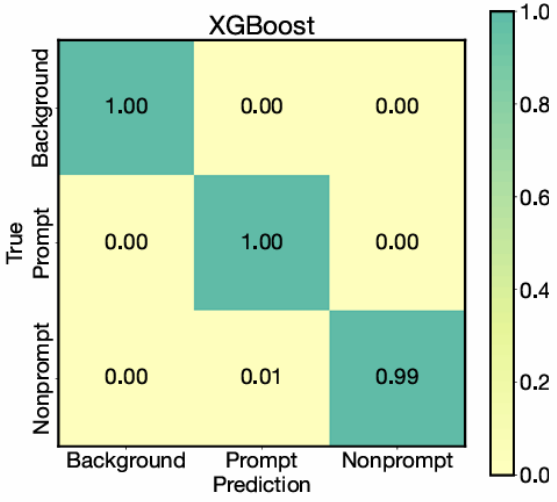
K. Goswami et al., Phys. Rev. D **110**, 034017 (2024)

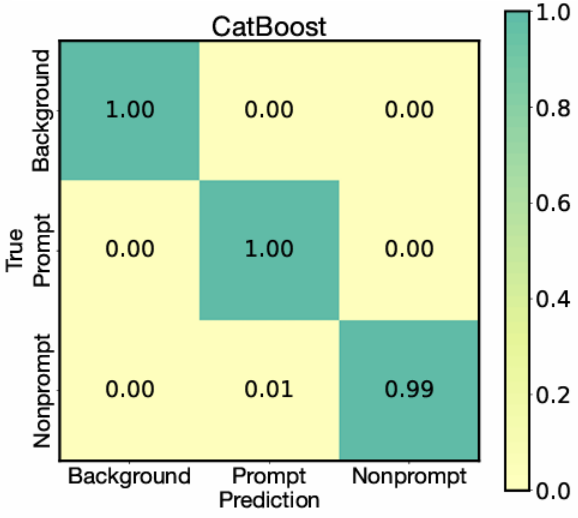
Machine Learning Algorithms: Detailed explanation

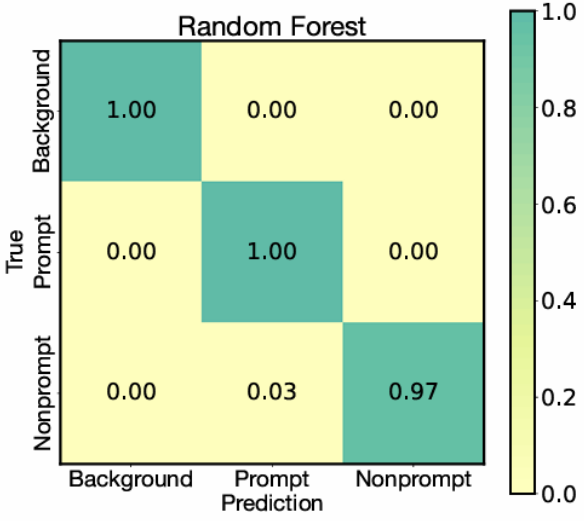
- Extreme Gradient Boost (XGBoost): Combines the predictions of multiple weak models to produce a stronger model.
- Categorical Boosting (CatBoost): Similar working principle as XGBoost but faster and more efficient when working with categorical data.
- Random Forest: In a Random Forest classifier, multiple decision trees are created, each on a different subset of the data. Each tree gets a vote on the class label for a new instance. The class that gets the most votes is chosen as the final prediction.

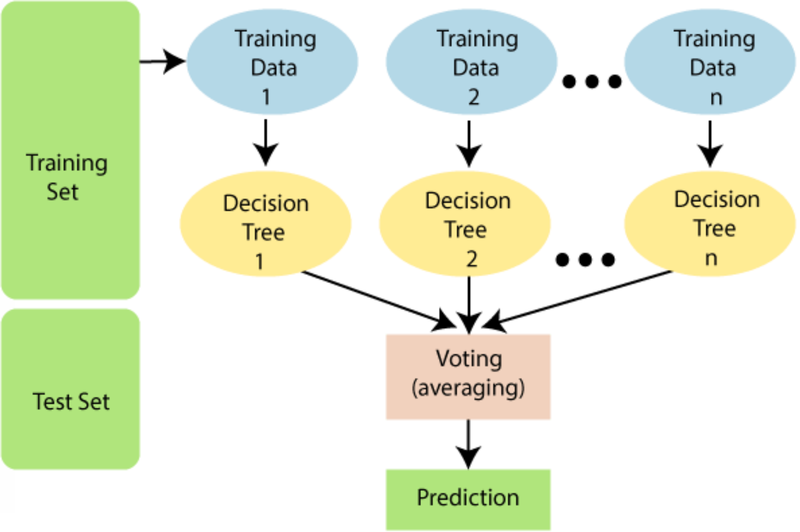


XGBoost and CatBoost Architecture







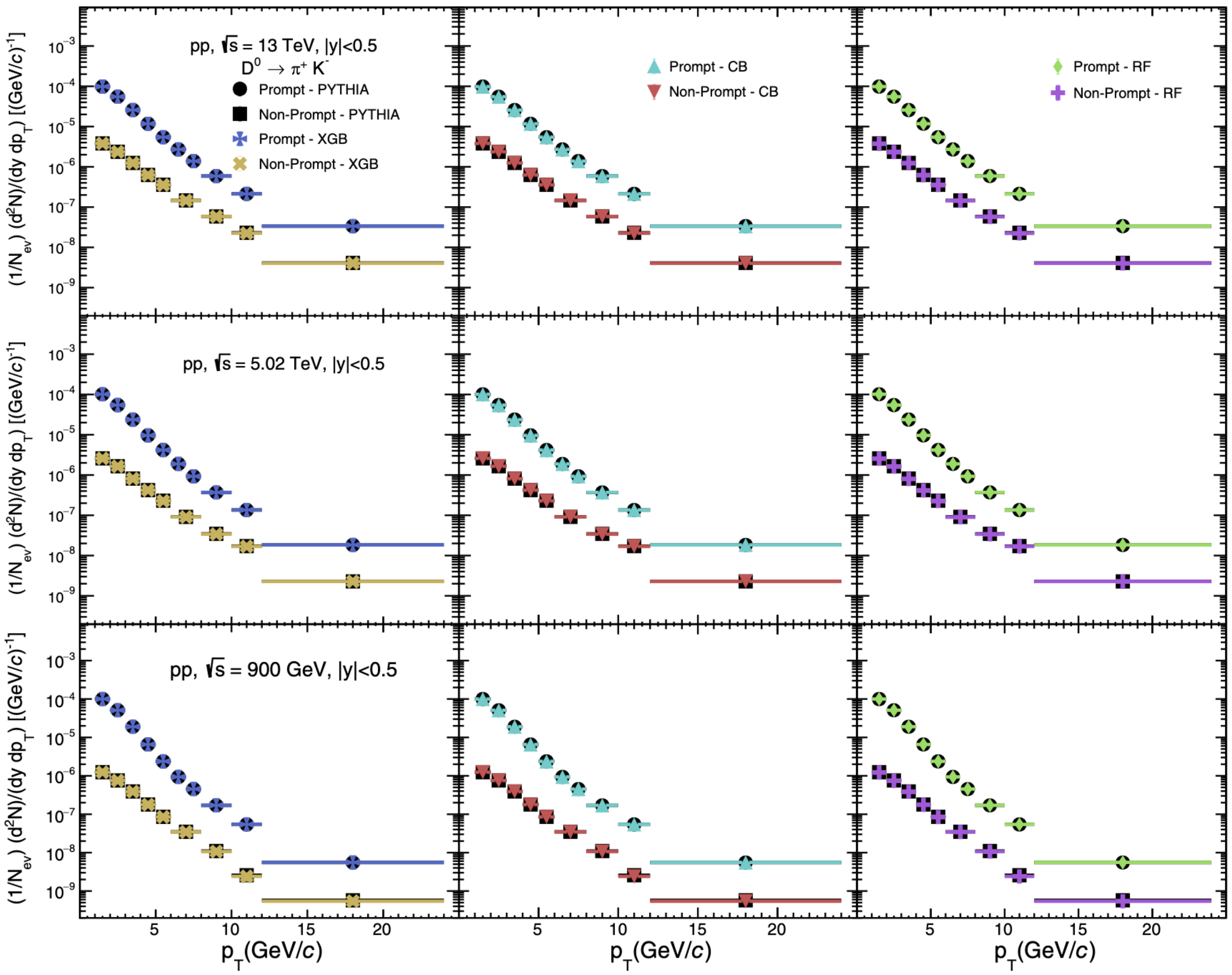


Random Forest Classifier Architecture

- **Our model shows an accuracy of 99% in separating prompt and non-prompt D^0 meson**

Results: Predicted p_T spectra

- Model trained: *pp* collision, using data simulated in PYTHIA8
  - $\sqrt{s} = 13$ TeV
- Predicted data: *pp* collision,
  - $\sqrt{s} = 13$, 5.02, and 0.9 TeV
- ML Algorithms ⇔ Monte Carlo

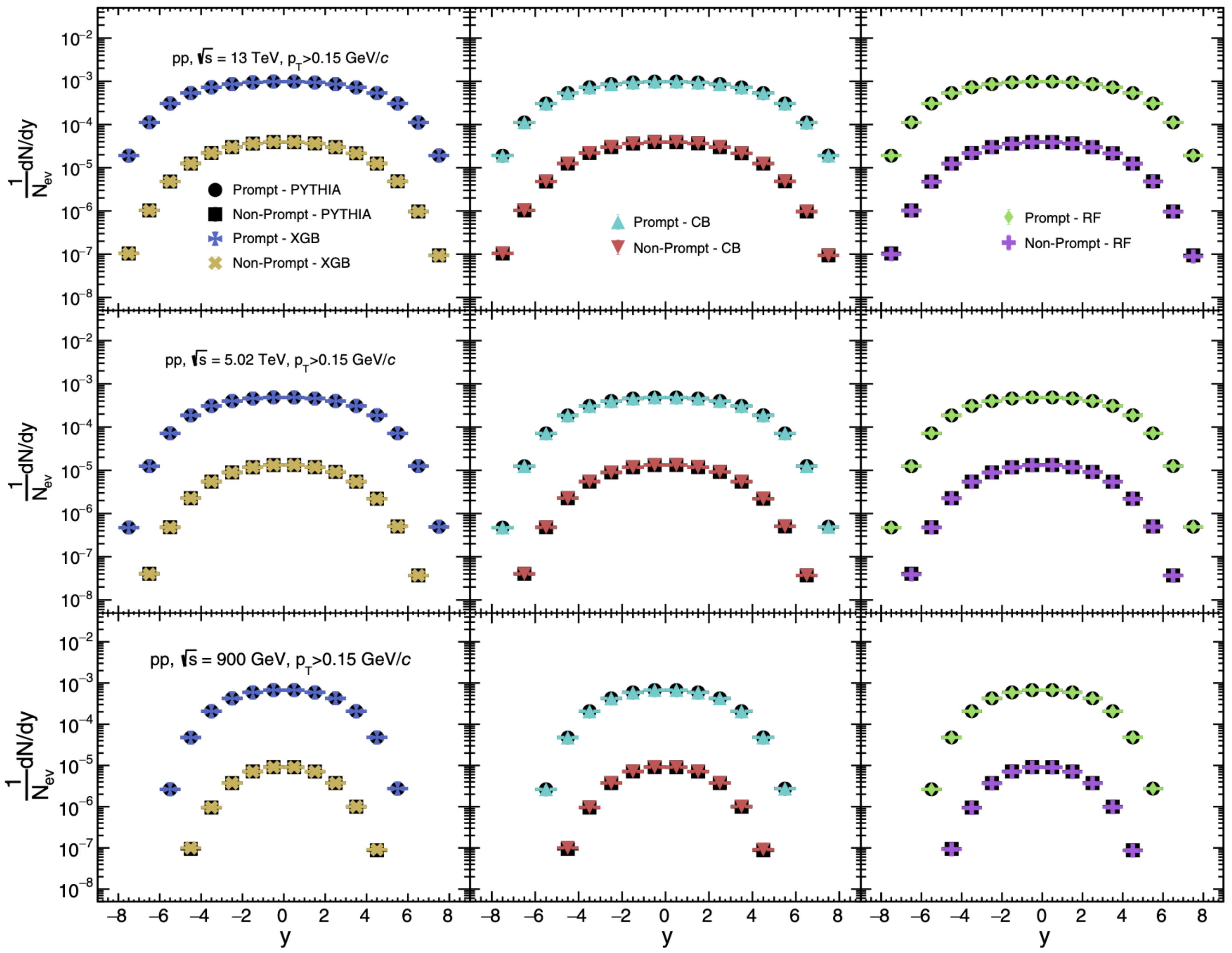


- MLhad: The Machine Learning for Hadronization collaboration (MLhad) seeks to design and deploy data-driven empirical hadronization models by extracting essential features directly from the wealth of available experimental data.

Source: https://uchep.gitlab.io/mlhad-docs/

Results: Predicted rapidity spectra

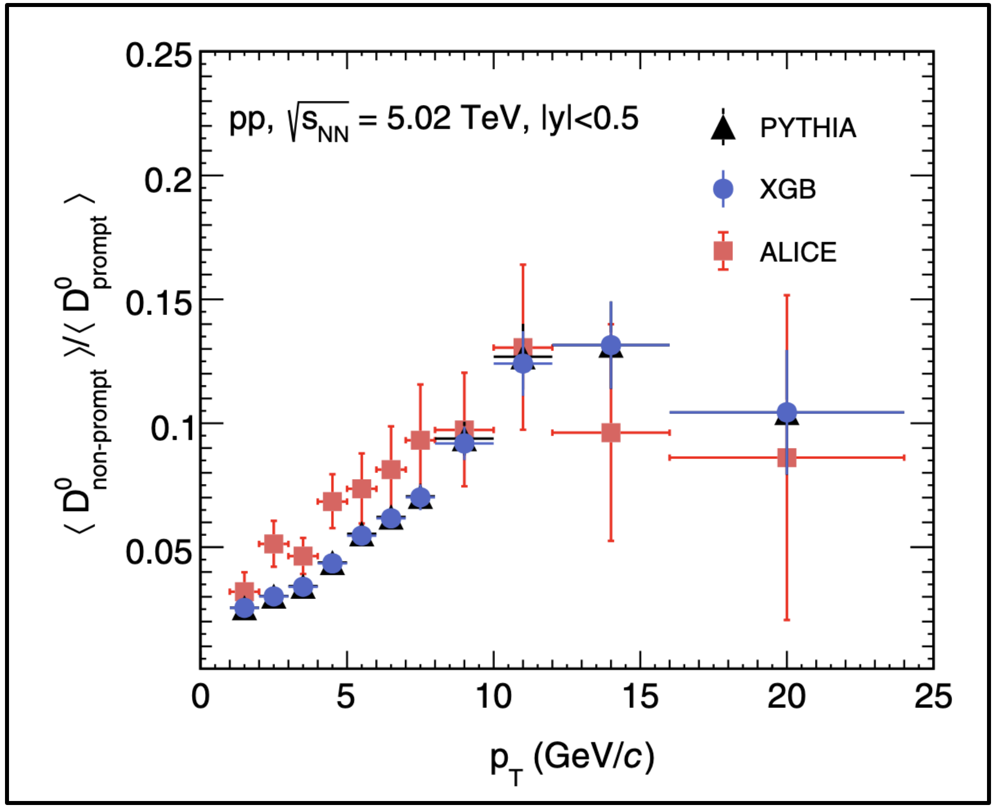
- Model trained: *pp* collision, using data simulated in PYTHIA8
  - $\sqrt{s} = 13$ TeV
- Predicted data: *pp* collision,
  - $\sqrt{s} = 13$, 5.02, and 0.9 TeV
- ML Algorithms ⇔ Monte Carlo

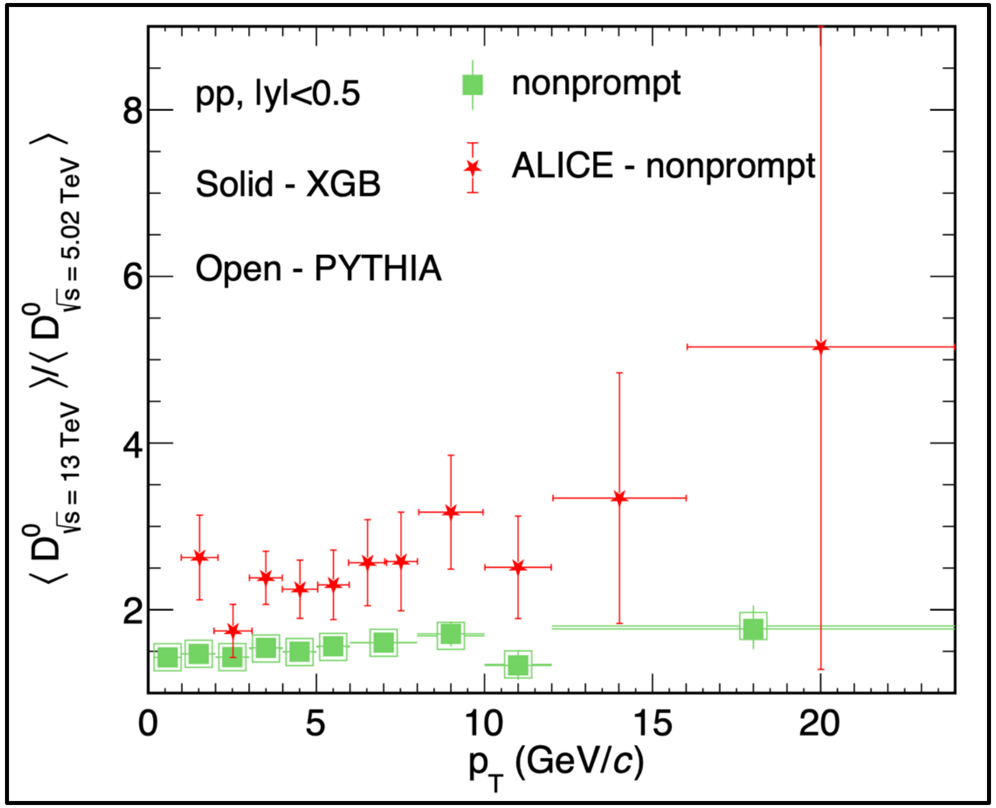


- MLhad: The Machine Learning for Hadronization collaboration (MLhad) seeks to design and deploy data-driven empirical hadronization models by extracting essential features directly from the wealth of available experimental data.

Source: https://uchep.gitlab.io/mlhad-docs/

Results

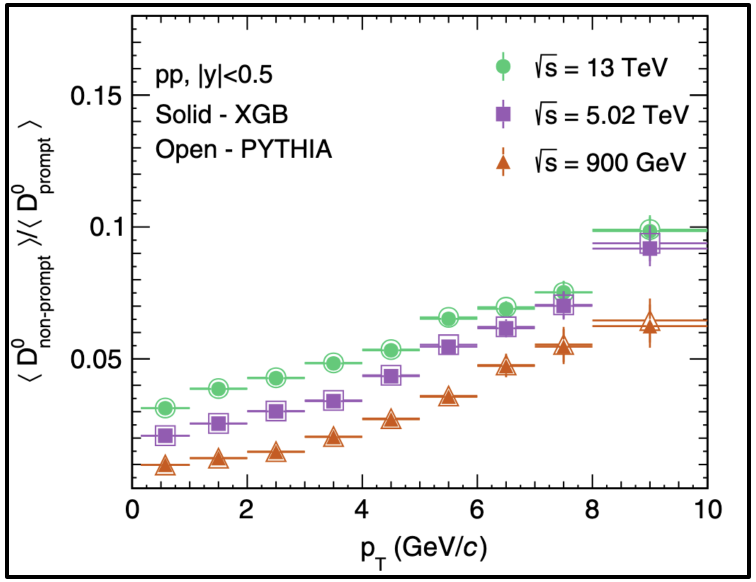


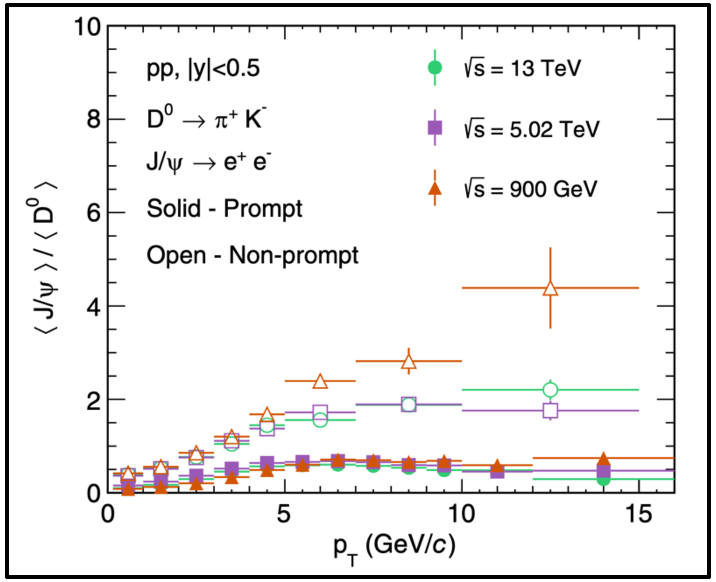


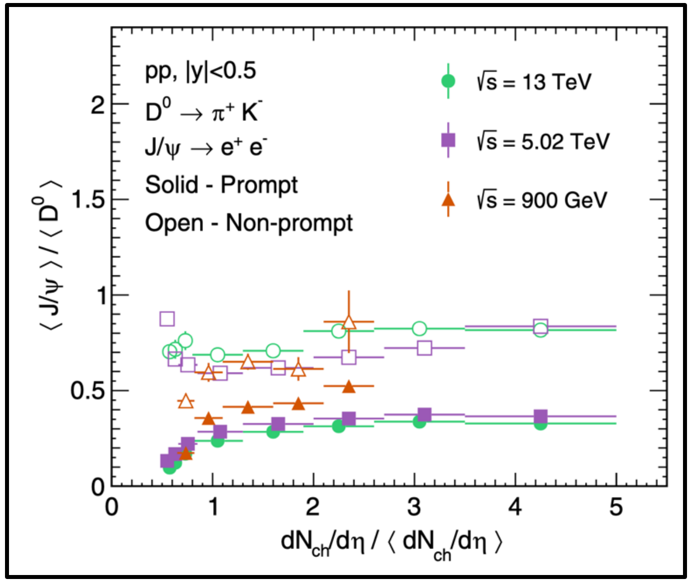
K. Goswami et al., Phys. Rev. D **110**, 034017 (2024)
ALICE Collaboration, JHEP **05**, 220 (2021)
ALICE Collaboration, JHEP **10**, 110 (2024)

- Comparison with ALICE experimental data:
  - Non-prompt to prompt D^0 meson ratio
  - Non-prompt D^0 meson Yield at $\sqrt{s}$ = 13 TeV to $\sqrt{s}$ = 5.02 TeV
- We underestimate the results slightly, in both cases. However, shows the same trend as experimental data

Results







K. Goswami et al., Phys. Rev. D **110**, 034017 (2024)

- Non-prompt to prompt D^0 meson ratio at different center-of-mass energy
- Charmonium (J/ψ) to Open charm (D^0) yield ratio as a function of transverse momentum and charged particle multiplicity
  - The non-prompt ratio exceeds unity as a function of p_T, which hints towards a higher feedback of the beauty hadrons into charmonium states as compared to the D^0 meson

Summary

- We train three different machine learning model, XGBoost, CatBoost, and Random Forest, for *pp* collision at $\sqrt{s} =$ **13 TeV**
- We predict the data for pp collision at $\sqrt{s} =$ **13, 5.02, and 0.9 TeV**
- We plot the non-prompt to prompt D^0 meson ratio and yield at different COM energies and compare our results with ALICE data
- We plot J/ψ to D^0 meson ratio as a function of transverse momenta and multiplicity
- We study the self-normalized yield of D^0 meson at three different center-of-mass energy

THANK YOU