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

Kangkan Goswami,* Suraj Prasad, Neelkamal Mallick, and Raghunath Sahoo
Department of Physics, Indian Institute of Technology Indore, Simrol, Indore 453552, India

Gagan B. Mohanty
Tata Institute of Fundamental Research,
Homi Bhabha Road, Mumbai 400005, India

Introduction

In proton-proton (pp) and heavy-ion collisions, the study of heavy-flavor hadrons is essential for understanding the properties of the QCD medium. Charm and beauty quarks are predominantly produced during the early stages of collisions through hard partonic interactions and traverse the deconfined medium before hadronizing to open-charm and beauty hadrons. The lightest open-charm meson, the D^0 ($c\bar{u}$), can originate from two different sources. Prompt D^0 mesons are produced either from the directly produced charm quark or from the decay of other charm hadrons. On the other hand, nonprompt D^0 mesons result from the decay of beauty hadrons. The prompt D^0 mesons comprises of charms quarks that are produced near the primary vertex and can be a useful tool to understand the QCD medium. However, the nonprompt D^0 mesons, a product of weak decay, can be utilized to explore the beauty sector. We distinguish between the prompt and nonprompt D^0 meson signal from background noise using a machine learning algorithm, XGBoost. The machine learning models are trained using features such as the invariant mass from the hadronic decay channel ($D^0 \rightarrow \pi^+ K^-$), pseudoproper time, pseudoproper decay length, and the distance of closest approach of the D^0 meson, with data simulated from PYTHIA8 for pp collisions at $\sqrt{s}$ = 13 TeV. We use the trained models to predict the prompt and nonprompt D^0 yield at three different center-of-mass energies $\sqrt{s} = 13$, 5.02, and 0.9 TeV, and compare our results with ALICE data wherever possible.

*Electronic address: `kangkangoswami44@gmail.com`

Methodology

We use PYTHIA8 to generate pp collisions at $\sqrt{s} = 13$ TeV for training machine learning algorithms to differentiate between prompt and nonprompt D^0 mesons from the background. We focus on several topological features to identify the production dynamics of these mesons. The selected topological features associated with the displaced production vertex of the D^0 mesons include pseudoproper time (t_z) and pseudoproper decay length ($c\tau$). Additionally, we use the distance of closest approach (DCA_{D^0}) as a topological input variable for the ML models. The DCA_{D^0} is determined by the decay length and the sine of the angle between $\vec{L}$ and the D^0 momentum vector $\vec{p}_{D^0}$, where $\vec{L}$ is the vector connecting the primary and secondary vertex. A detailed discussion on the input features can be found in Ref. [1, 2]. The training of the XGBoost model is performed using 600 million minimum bias pp collisions at $\sqrt{s} = 13$ TeV.

Results

Figure 1 shows the confusion matrix for the XGBoost model, with the Y-axis representing the true pair counts from PYTHIA8 and the X-axis showing the predicted pair counts from the XGBoost model. The XGBoost model achieves 100% accuracy in separating background pairs and shows a 99% accuracy in distinguishing prompt D^0 mesons from nonprompt ones.

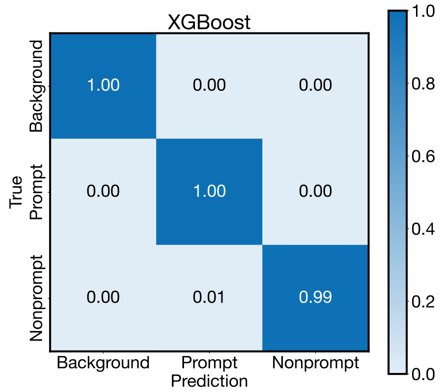

FIG. 1: Confusion matrix for XGBoost model [1].

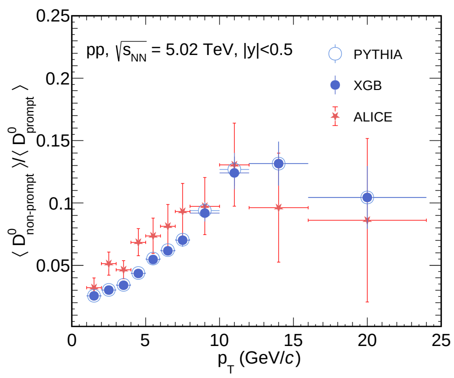

FIG. 2: Nonprompt to prompt D^0 meson ratio in *pp* collisions at $\sqrt{s} = 5.02$ TeV compared with ALICE results [1, 3].

Figure 2 shows the ratio of nonprompt to prompt D^0 meson yields at midrapidity ($|y| < 0.5$) in minimum bias *pp* collisions at $\sqrt{s} = 5.02$ TeV, as a function of p_T. This ratio provides insight into the relative yields of D^0 mesons originating from beauty hadron decays compared to those from direct charm hadron production. The XGBoost predictions are compared with the ALICE results [3].

We investigate the impact of center-of-mass energy on D^0 meson production by estimating the ratio of D^0 yields at two different energies, 13 TeV and 5.02 TeV. Figure 3 displays the ratio of D^0 yields at $\sqrt{s} = 13$ TeV

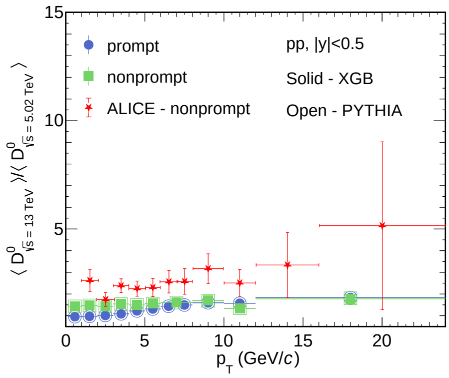

FIG. 3: Ratio of D^0 yield in *pp* collisions at $\sqrt{s} =$ 13 and 5.02 TeV compared to ALICE data [1, 4].

to $\sqrt{s} = 5.02$ TeV. For the prompt case, we observe a clear increase in the ratio as p_T rises. In contrast, the nonprompt case shows a flat trend across the entire p_T range. A similar pattern was recently reported by ALICE [4].

Summary

We employ machine learning algorithms for a novel track-level unbinned identification and separation of prompt and nonprompt D^0 mesons from background pion-kaon pairs. Here, it is noteworthy that we train the model using PYTHIA8 simulated data for *pp* collision at $\sqrt{s} = 13$ TeV and predict the yield of prompt and nonprompt D^0 meson at $\sqrt{s} =$ 5.02 TeV. Our results agree well with ALICE data for *pp* collision at $\sqrt{s} = 5.02$ TeV, which sheds light on the robustness of our model.

References

[1] K. Goswami, S. Prasad, N. Mallick, R. Sahoo and G. B. Mohanty, Phys. Rev. D **110**, 034017 (2024).

[2] S. Prasad, N. Mallick and R. Sahoo, Phys. Rev. D **109**, 014005 (2024).

[3] S. Acharya *et al.* [ALICE], JHEP **05**, 220 (2021).

[4] S. Acharya *et al.* [ALICE], [arXiv:2402.16417 [hep-ex]].