Prompt and non-prompt production of charm hadrons in p+p collisions at the Large Hadron Collider using machine learning

Raghunath Sahoo

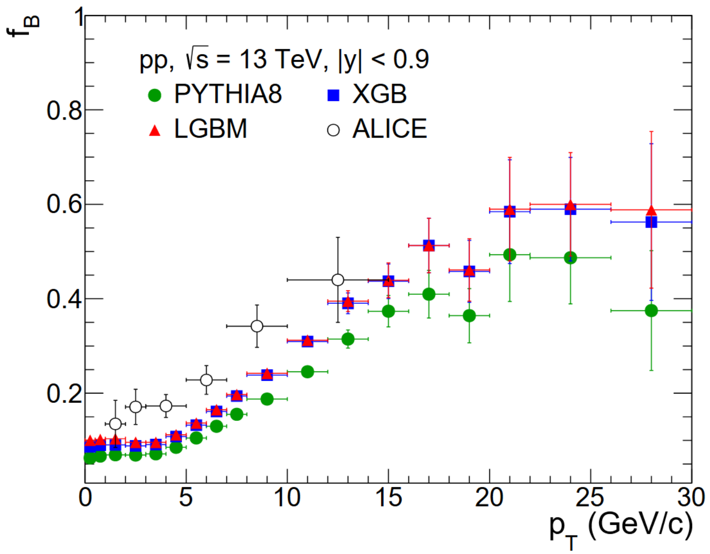
Indian Institute of Technology Indore, India

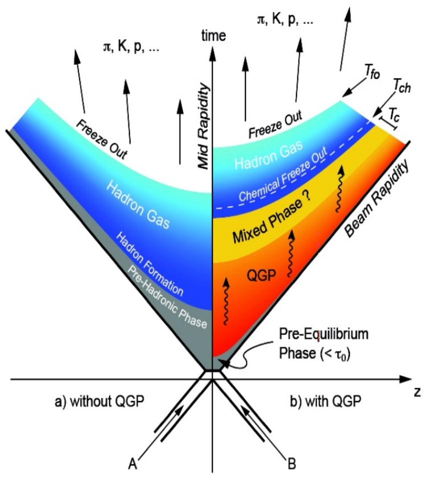
Raghunath.Sahoo@cern.ch, raghunath@iiti.ac.in

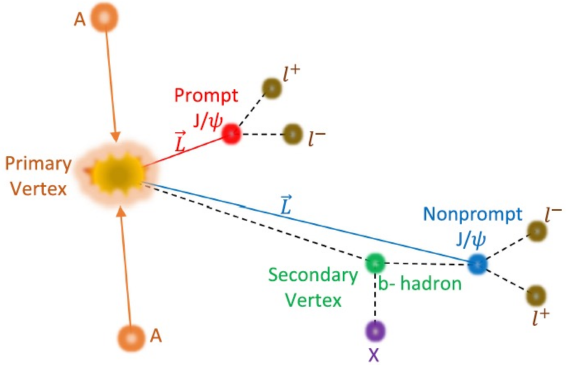
Based On:

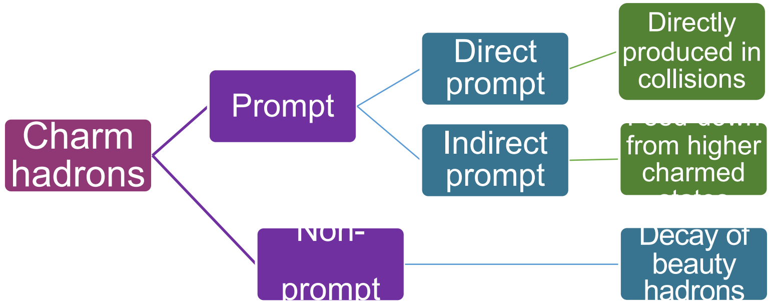
S. Prasad, N. Mallick and R. Sahoo, Phys. Rev. D 109, 014005 (2024)
K. Goswami, S. Prasad, N. Mallick, R. Sahoo and G. B. Mohanty, Phys. Rev. D 110, 034017 (2024)

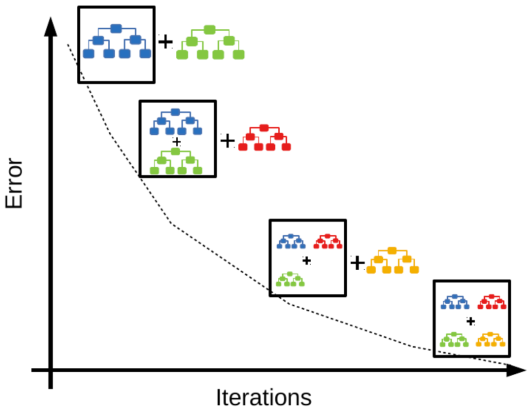
- Introduction
- Decay topology and the production of J/ψ and D^0
- ML Methods to separate prompt and non-prompt
- Results and Summary



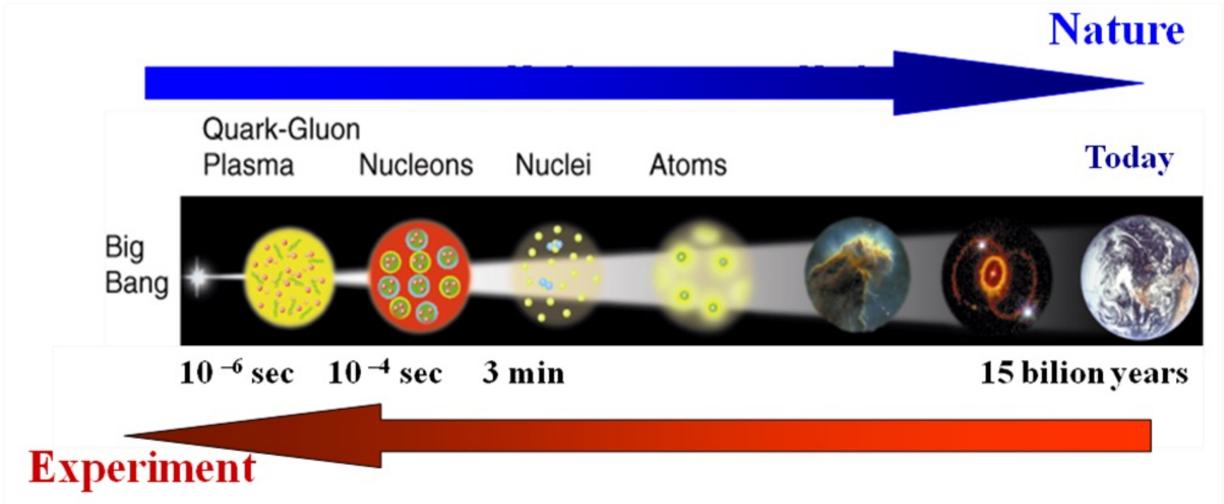








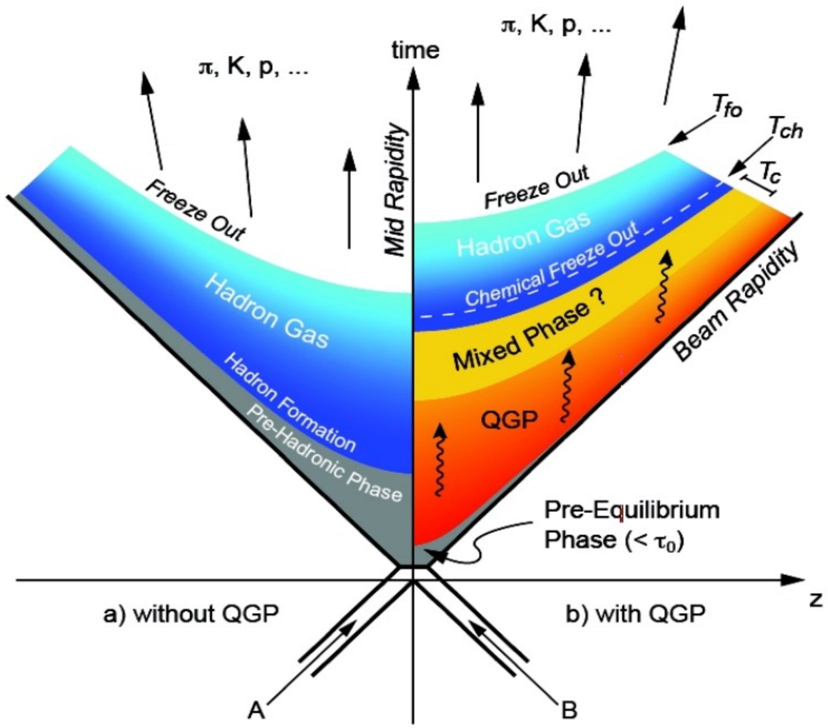
Introduction



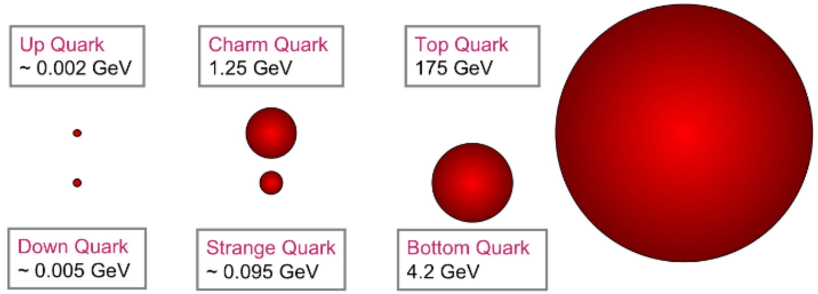
- Charm and beauty hadrons are formed during the initial stages of the hadronic and heavy-ion collisions

- Experience the whole system evolution → Good Probes to understand QCD medium

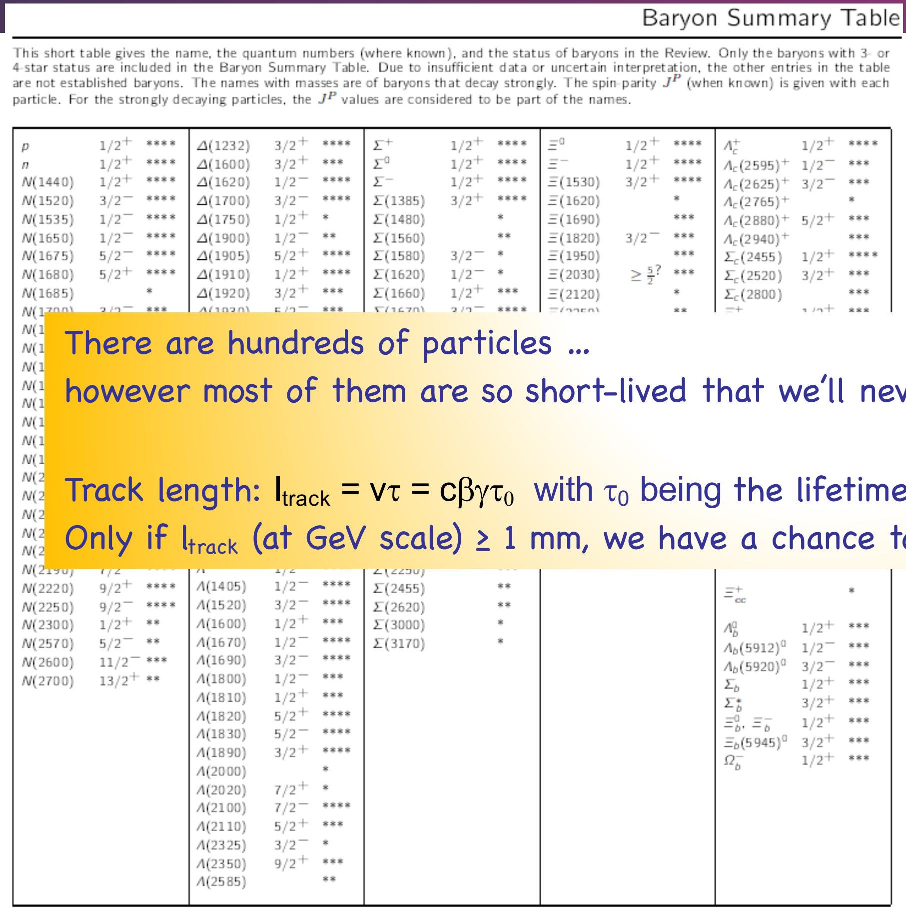
- Lightest Open Charm: D^0 meson ($m_{D^0} \approx 1.865$ GeV/c); Lightest charmonium vector meson: J/ψ ($m_{J/\psi} \approx 3.096$ GeV/c) → Abundant production as compared to other open/hidden charm hadrons

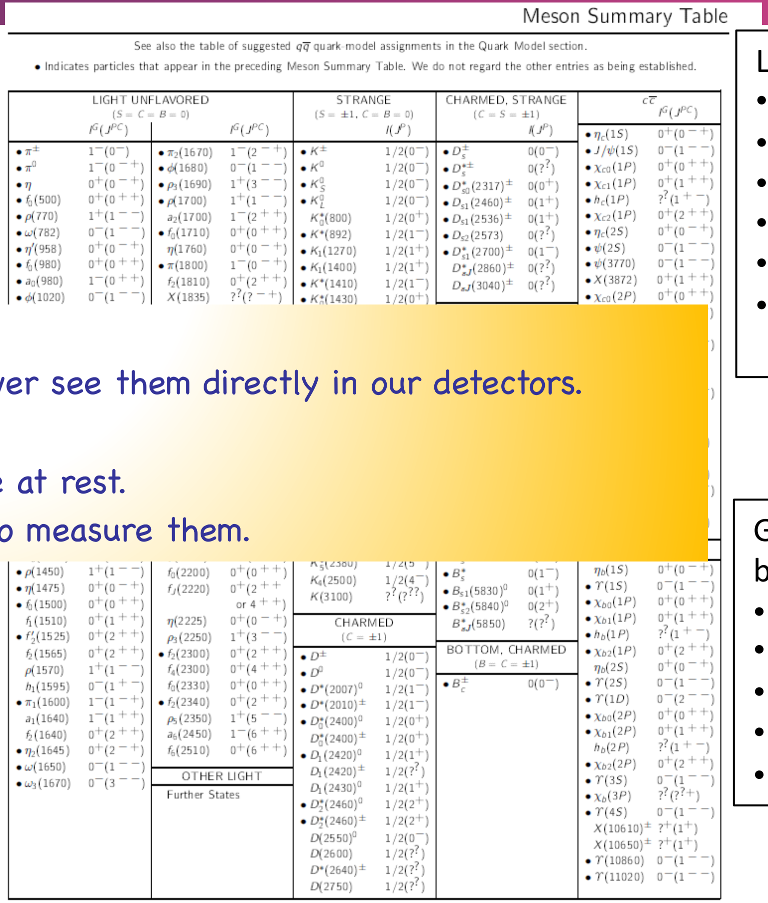


https://particlesandfriends.wordpress.com/2016/10/14/evolution-of-collisions-and-qgp/



Introduction





Leptons
- e
- μ
- τ
- ν_e
- ν_μ
- ν_τ

Gauge bosons
- γ
- $W^{+/-}$
- Z
- g
- H

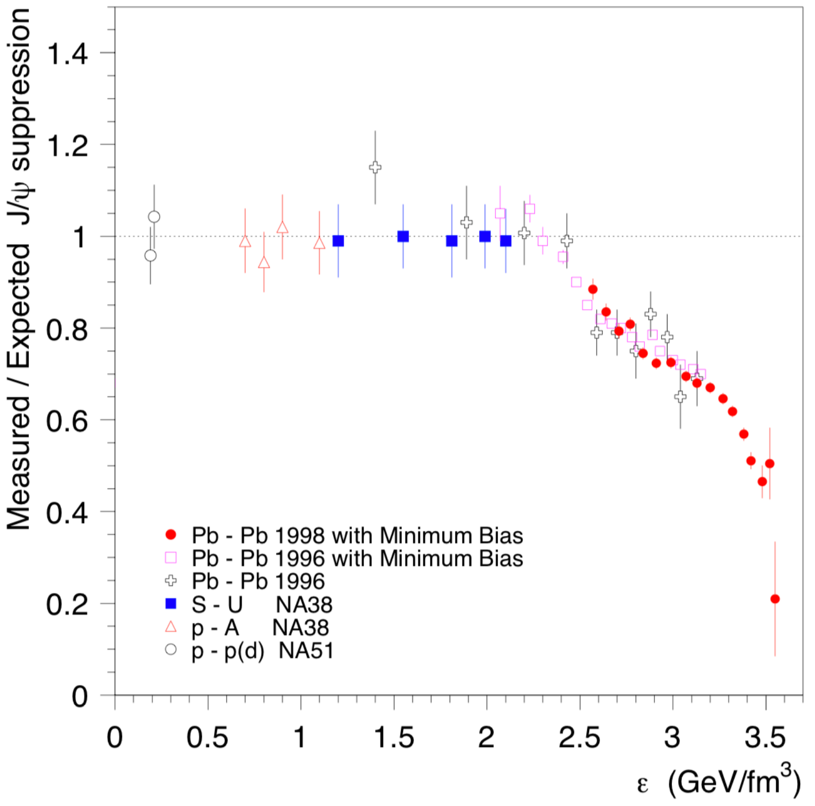
There are hundreds of particles …
however most of them are so short-lived that we'll never see them directly in our detectors.

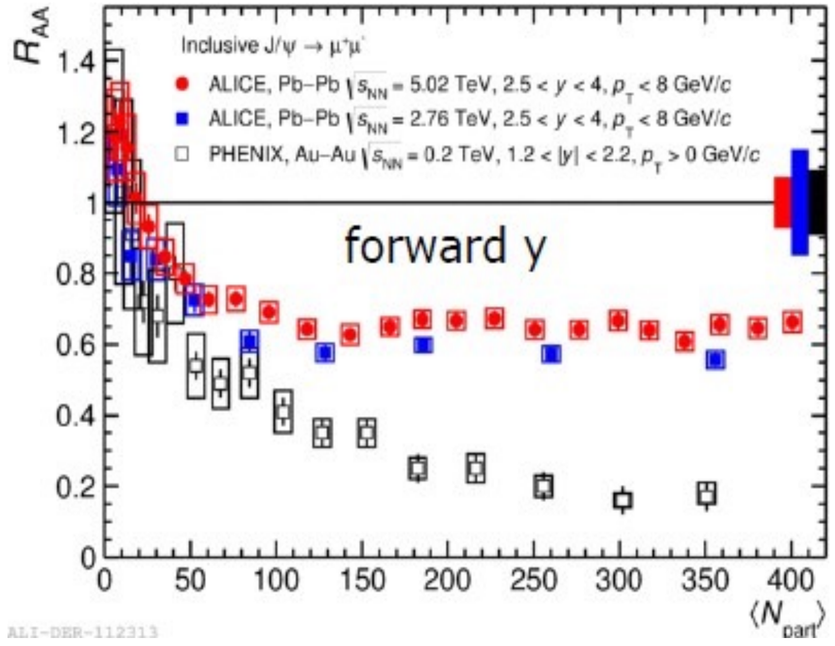
Track length: $l_{track} = v\tau = c\beta\gamma\tau_0$ with τ_0 being the lifetime at rest.
Only if l_{track} (at GeV scale) $\geq$ 1 mm, we have a chance to measure them.

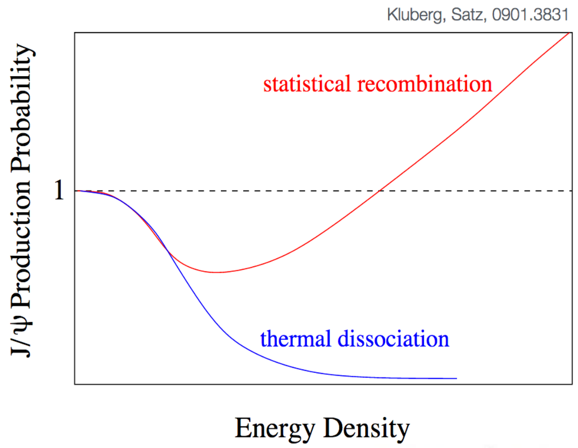
Which are left then? These 8 particles (and their antiparticles).

	γ	p	n	$e^\pm$	$\mu^\pm$	$\pi^\pm$	$K^\pm$	K_0 (K_S/K_L)
τ_0	∞	∞	∞	∞	2.2μs	26 ns	12 ns	89 ps / 51 ns
$l_{track\ (p=1GeV)}$	∞	∞	∞	∞	6.1 km	5.5 m	6.4 m	5 cm / 27.5 m

Experimental observation of J/ψ Suppression







$$R_{AA}(p_T) = \frac{d N_{AA}/dp_T}{\langle T_{AA}\rangle \times d\sigma_{pp}/dp_T}$$

$R_{AA} < 1$: Suppression of yield due to presence of medium

$R_{AA} > 1$: No suppression and hence no medium

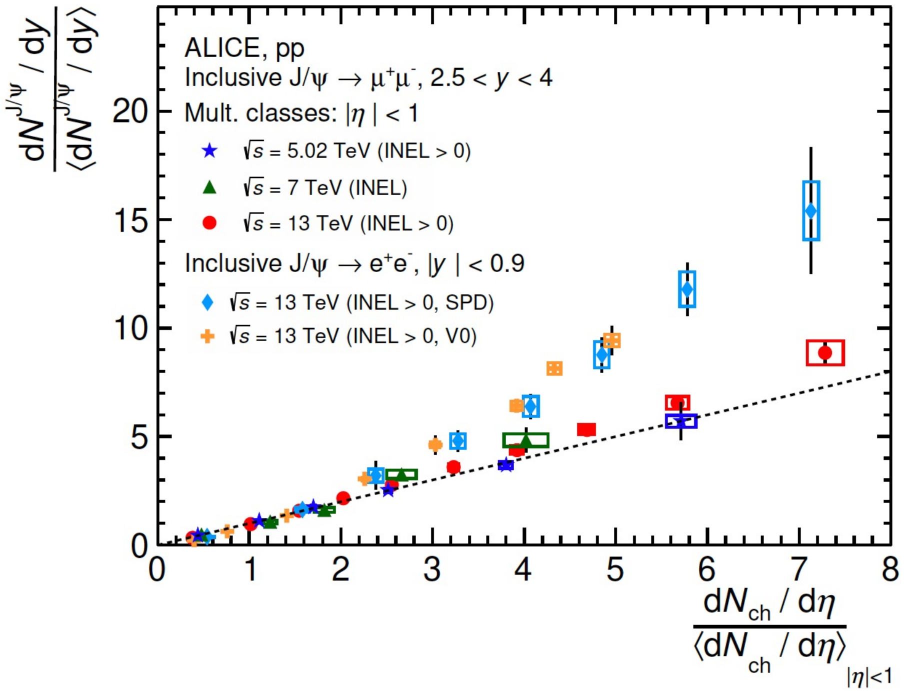
Suppression and regeneration!

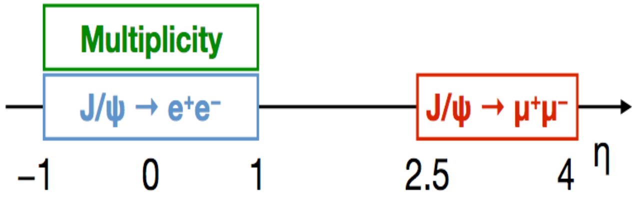
CERN SPS (NA50): observed J/ψ suppression as a function of energy density for various collision species. Note that the critical energy density for a partonic medium is 1 GeV/fm³.

[Physics Letters B 766 (2017) 212]

http://alice.web.cern.ch/content/mystery-jpsi

J/ψ Production Anomaly



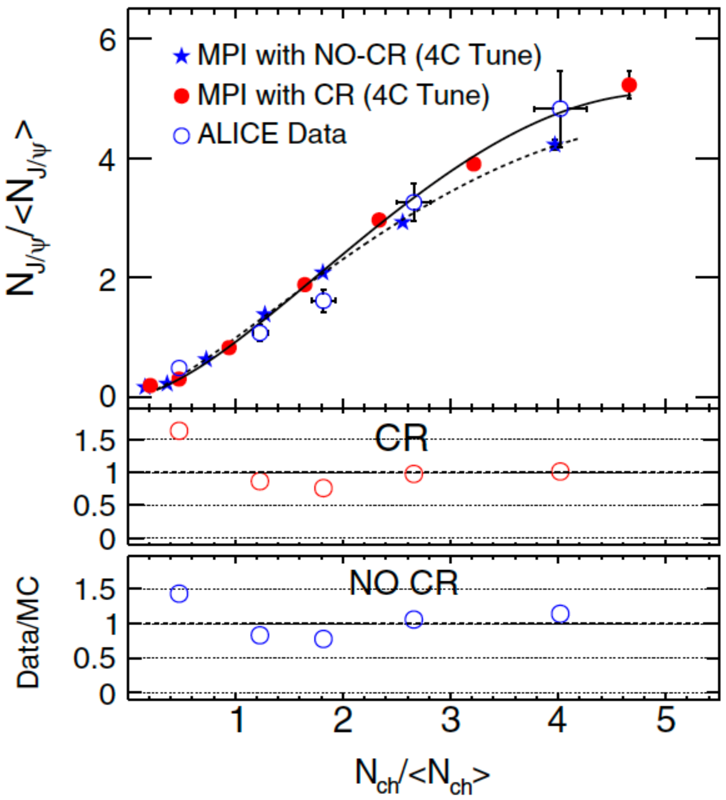


ALICE: JHEP06(2022)015

- Relative J/Ψ yield at midrapidity is compared to the forward rapidity yield as a function of midrapidity relative charged-particle multiplicity
- Midrapidity yields exhibit faster than linear increase
- The results using midrapidity multiplicity selection based on the SPD detector ($|\eta| < 1$) and forward-rapidity multiplicity selection based on the V0 detector ($-3.7 < \eta < -1.7$ and $2.8 < \eta < 5.1$) are found to be compatible within the uncertainties
- Therefore, the different trends in the multiplicity dependence of the J/Ψ production observed at midrapidity and forward rapidity are not due to a possible auto-correlation bias

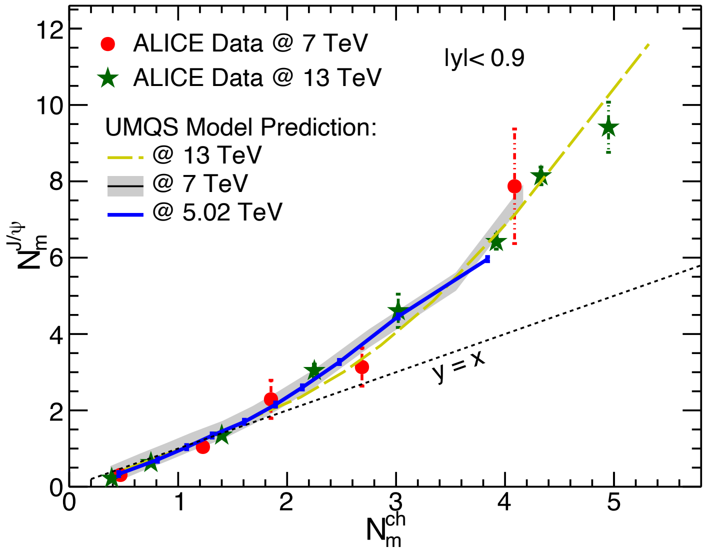
PhD Thesis: D. Thakur, IIT Indore

Understanding J/ψ Production



MPI with CR reasonably explains the data!

THAKUR, DE, SAHOO, and DANSANA, Phys.. Rev. D 97, 094002 (2018)



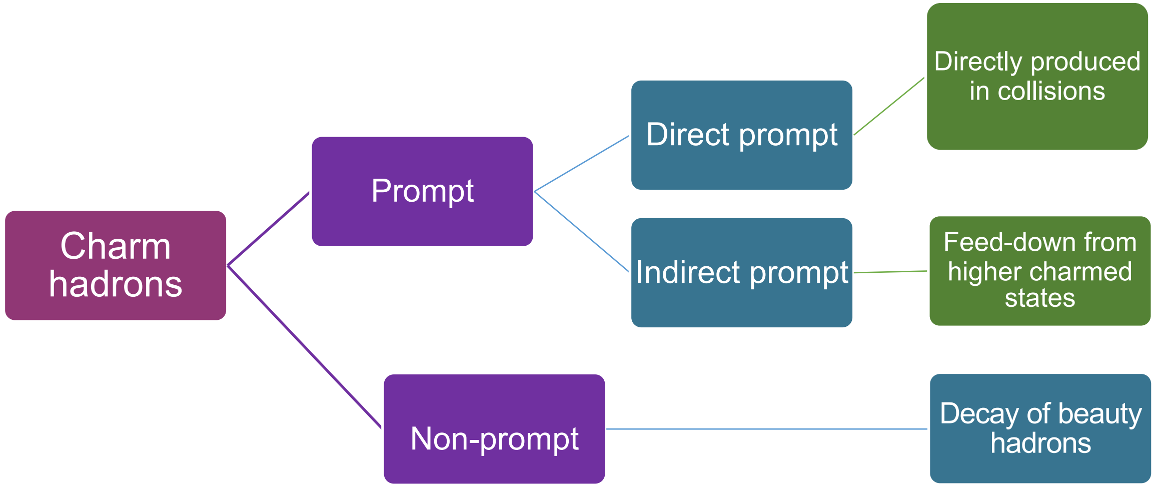
C.R. Singh, S. Deb, R. Sahoo, J. Alam, Eur. Phys. J. C, 82, 542 (2022)

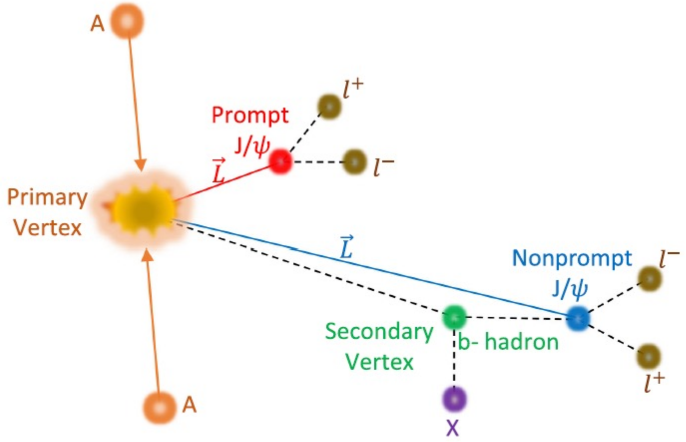
- ✓ J/ψ self-normalized yield as a function of self-normalized multiplicity follows a scaling across collision energies.
- ✓ Unified Model of Quarkonia Suppression (UMQS) model which incorporates the suppression of J/ψ through color screening, gluonic dissociation, and collision damping and regeneration of charmonium due to correlated c – cbar pairs.

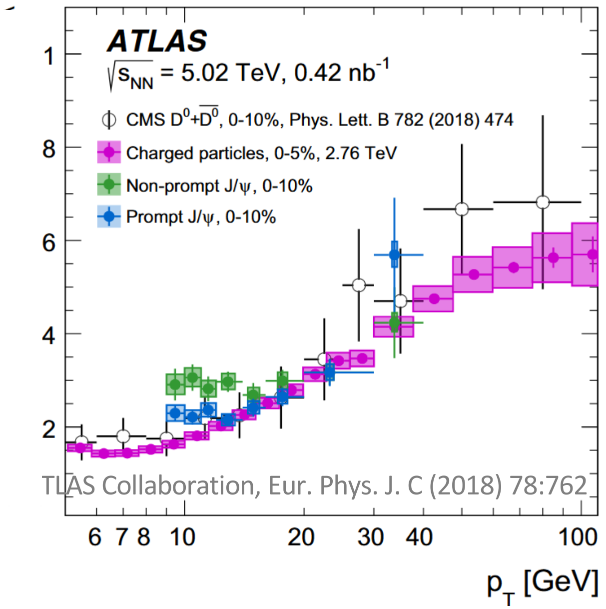
- All-inclusive J/ψ

- Need of separating prompt and non-prompt J/ψ

Topological production of charm hadrons



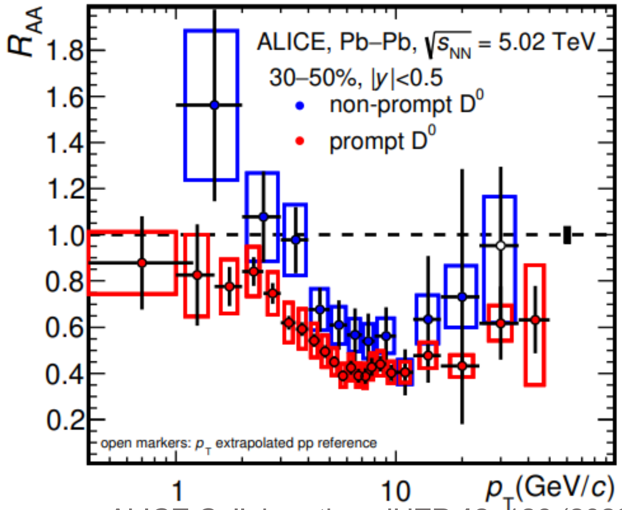




S. Prasad, N. Mallick and R. Sahoo, Phys. Rev. D 109, 014005 (2024)

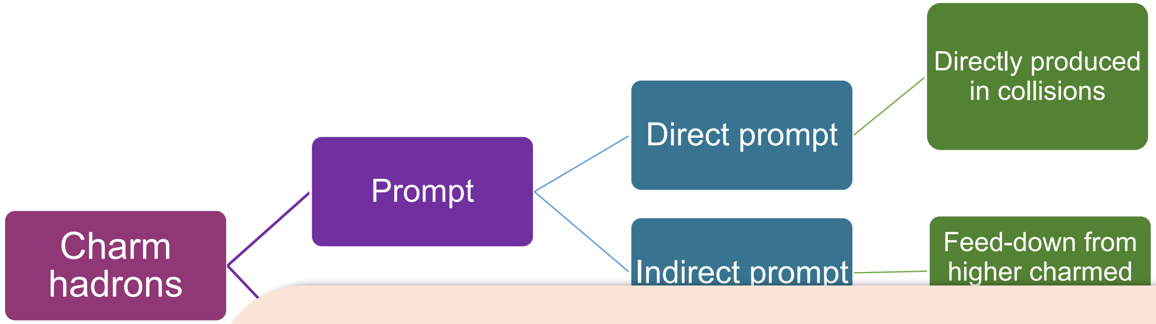
- ➢ Prompt charm hadrons: **Direct production and decay higher excited charm hadrons →** Good probes for QCD medium and to test theories of strong interactions

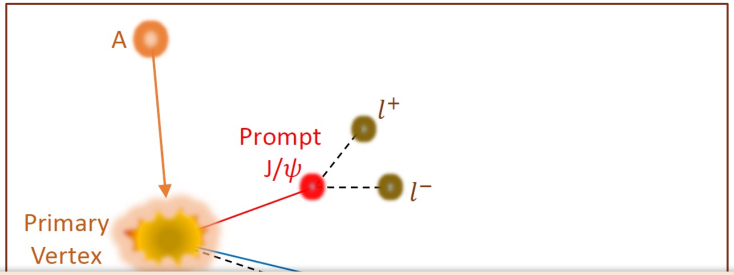
- ➢ Non-prompt charm hadrons: **Weak decay of beauty hadrons** → Indirect studies of beauty hadron production

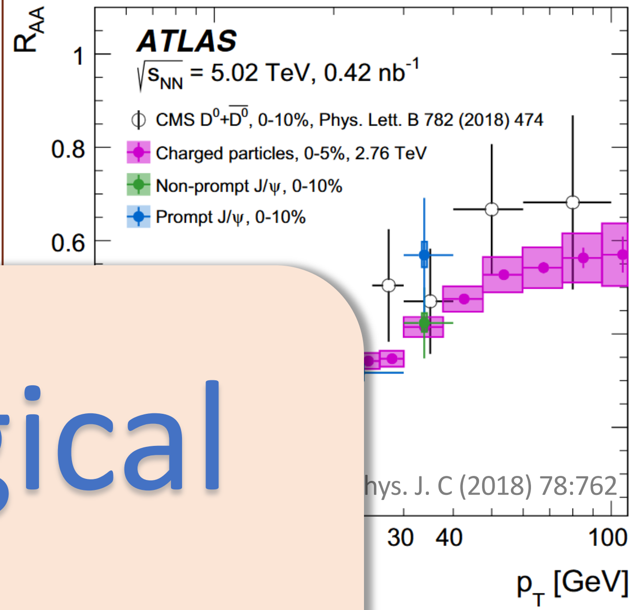


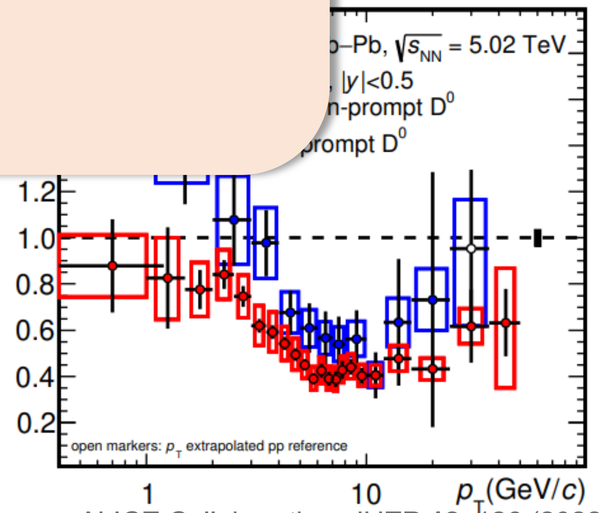
ALICE Collaboration, JHEP **12**, 126 (2022).

Topological production of charm hadrons









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

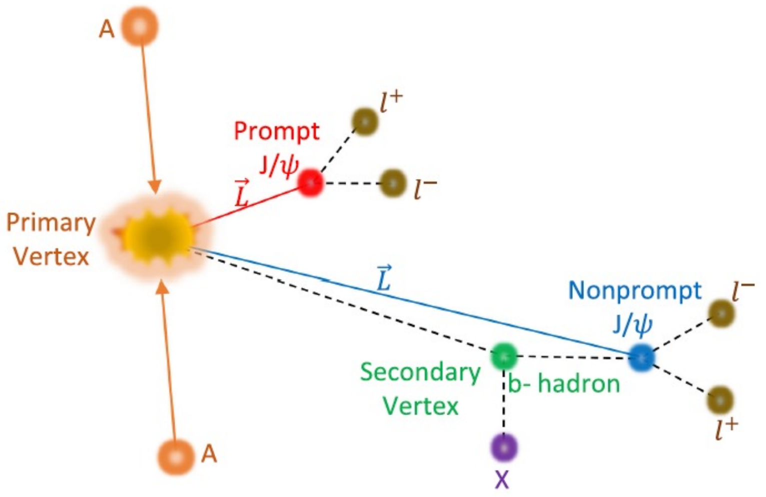
Segregating the topological production of J/ψ

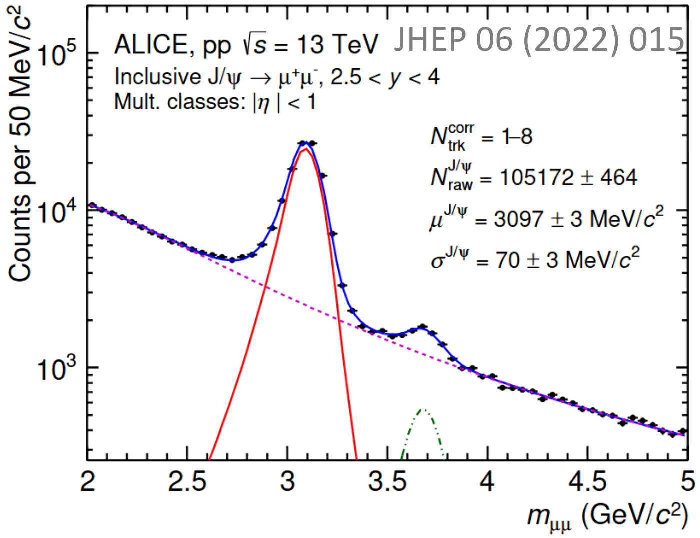
- Prom… probes for QCD medium and to test theories of strong interactions

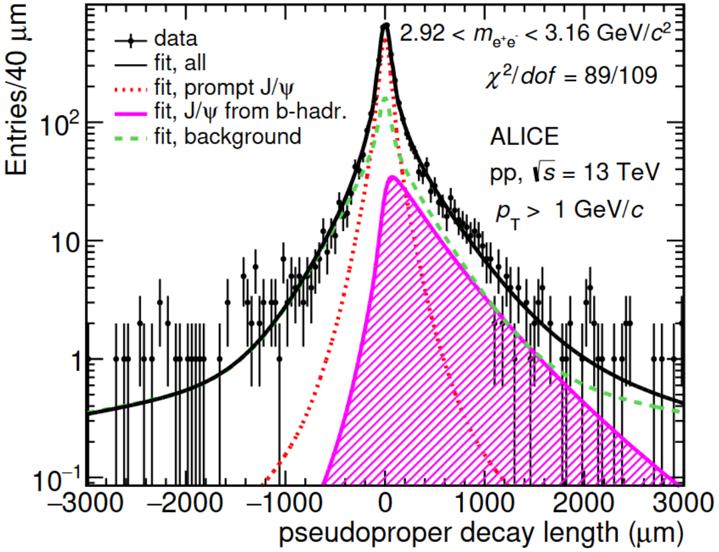
- Non-prompt charm hadrons: **Weak decay of beauty hadrons** → Indirect studies of beauty hadron production

Separating prompt and non-prompt J/ψ

- J/ψ (3.096 GeV/c^2)
- In experiments, $J/\psi \rightarrow \mu^+ + \mu^-$ or $J/\psi \rightarrow e^+ + e^-$
- Prompt Production: Direct production/ decay of heavier charmonium states
- Non-prompt Production: Products of beauty hadron weak decays
- Prompt and non-prompt J/ψ are topologically different







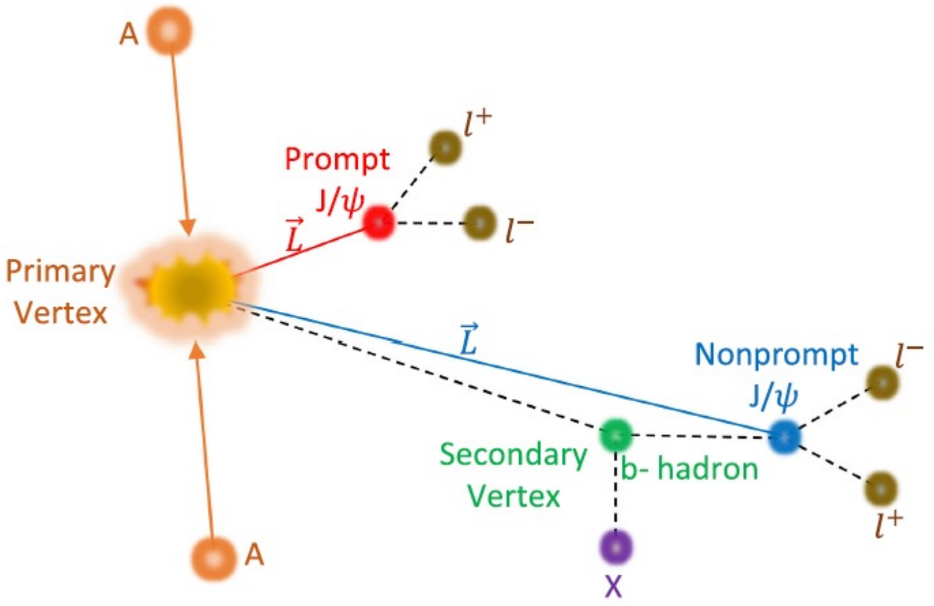
S. Prasad, N. Mallick and R. Sahoo, *Phys. Rev. D 109, 014005 (2024)*

ALICE, JHEP 03 (2022) 190

Separating prompt and non-prompt J/ψ

Simulating pp collisions at $\sqrt{s}$ =13 TeV using PYTHIA8 → Training machine-learning models → Using the machine-learning models to predict prompt and non-prompt yields at different energies → Results

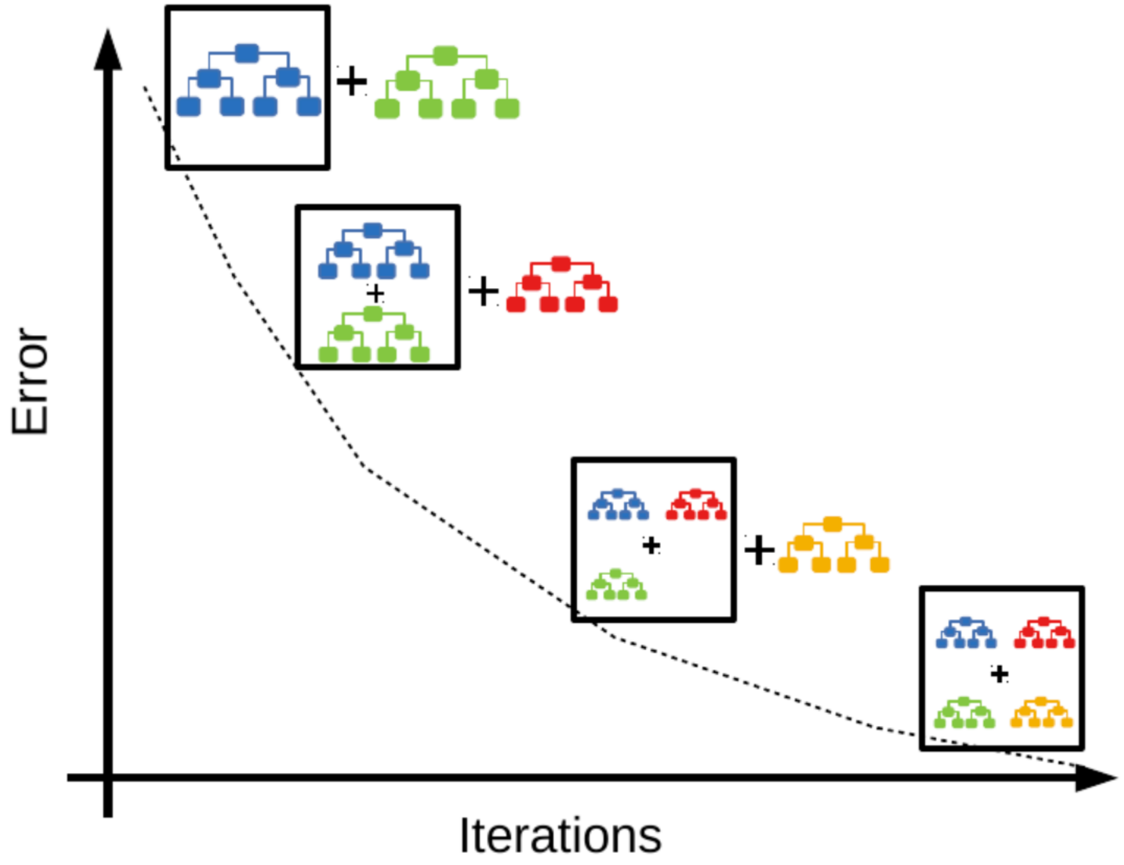
- PYTHIA8 (4C-tune) 20 billion minimum bias events for pp $\sqrt{s} = 13$ TeV
- The coordinates of the primary vertex are randomized following a Gaussian distribution
- $J/\psi \rightarrow \mu^{+} + \mu^{-}$ channel is used to reconstruct invariant mass ($m_{\mu\mu}$), transverse momentum ($p_{T,\mu\mu}$), pseudorapidity ($\eta_{\mu\mu}$) and rapidity ($y_{\mu\mu}$) of the dimuons
- Pseudoproper decay length ($c\tau$) of the reconstructed dimuon pairs along with $m_{\mu\mu}$, $p_{T,\mu\mu}$, and $\eta_{\mu\mu}$ are taken as inputs



$$c\tau = \frac{c\, m_{J/\psi}\, \vec{L} \cdot \vec{p_T}}{|\vec{p_T}|^2}$$

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

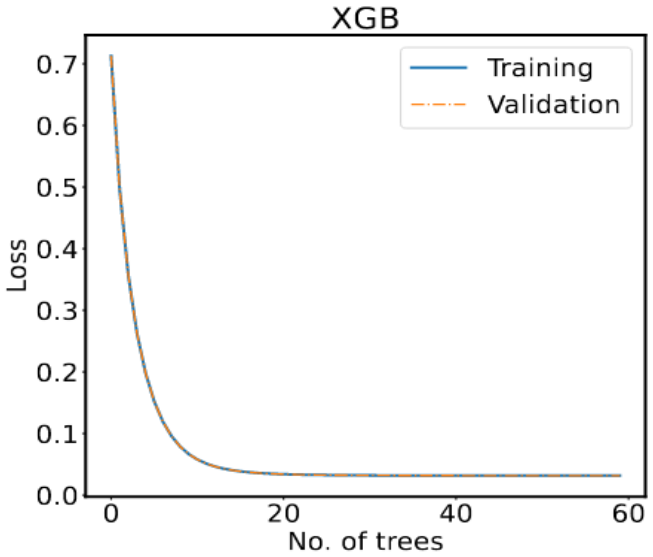
Gradient Boosting Method



XGBoost, LightGBM, and CatBoost

Machine Learning Model parameters for J/ψ

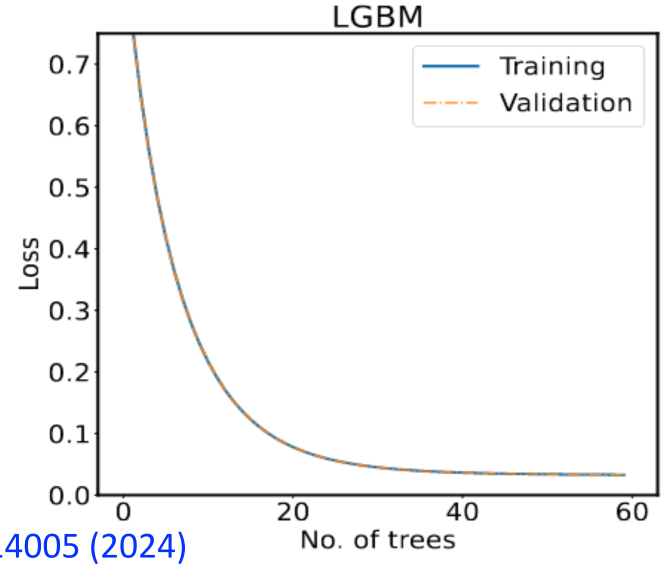
- Background : Prompt : Non-prompt = 20 : 10 : 1
- Classification models required to be trained on similar number of training instances → oversampling of data is done
- Dataset for Training : Testing : Validation = 81 : 10 : 9
- Parameters are chosen through a grid search method (Making an array of all possible parameters and training to find the parameter values for minimum loss)



	XGB	LGBM
Learning rate	0.3	0.1
Sub-sample	1.0	1.0
No. of trees	60	60
Maximum depth	3	3
Objective	*softmax*	*softmax*
Metric	*mlogloss*	*multilogloss*

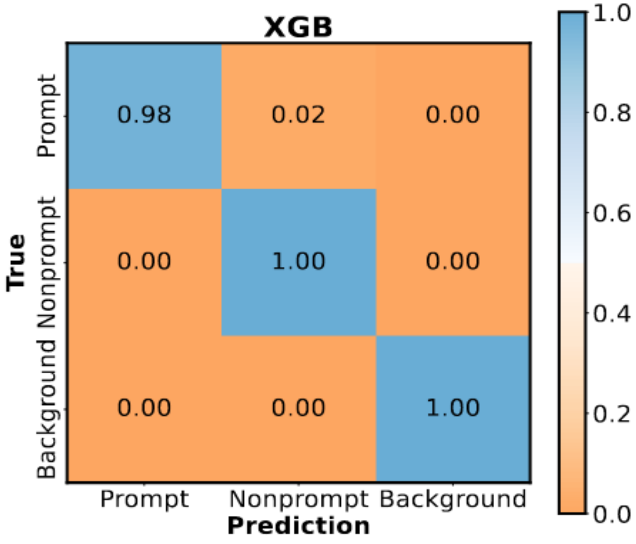
LGBM: Light Gradient Boosting Machine
XGB: Extreme Gradient Boosting

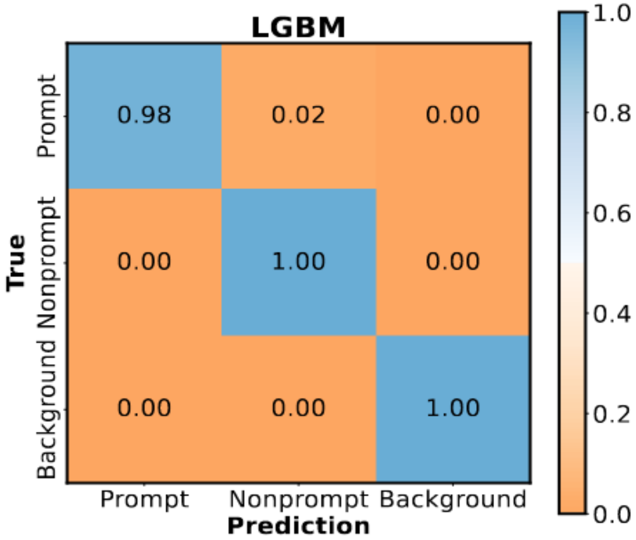
- Loss saturates around 25 trees and 45 trees for XGB and LGBM
- Training and validation curves are on top of each other → No overfitting/underfitting



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

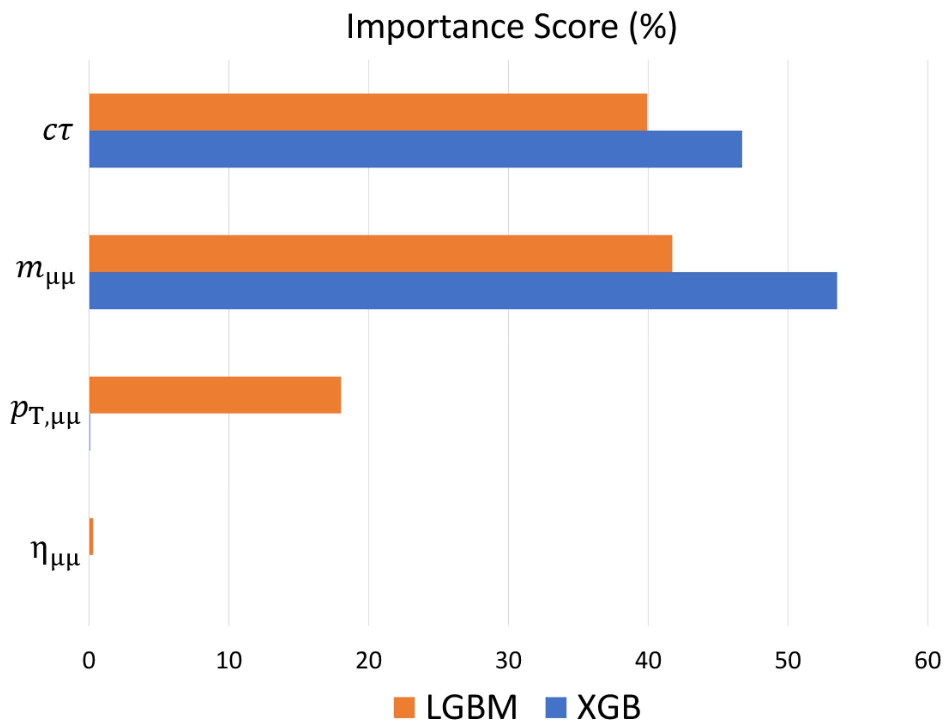
Model performance for J/ψ





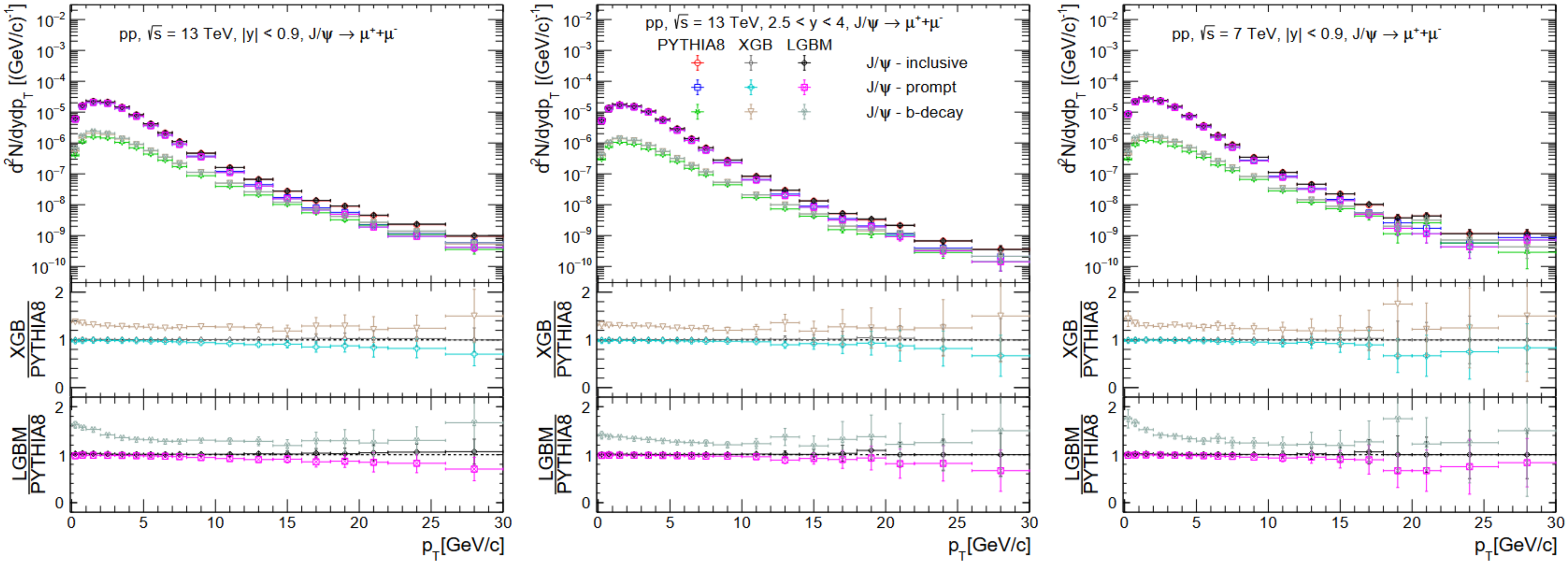
- Confusion Matrix talks about the mispredictions given by the model for each class
- Both XGB and LGBM perfectly separates the inclusive J/ψ from the uncorrelated background pairs
- Both models mispredict 2% of prompt J/ψ as the non-prompt → Raises non-prompt yield

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



- Importance score tells how important a feature for a decision making of the models
- The importance score of invariant mass of dimuons is highest for both the models
- $c\tau$ contributes to decision-making of the models significantly

ML: Transverse momentum spectra of J/ψ

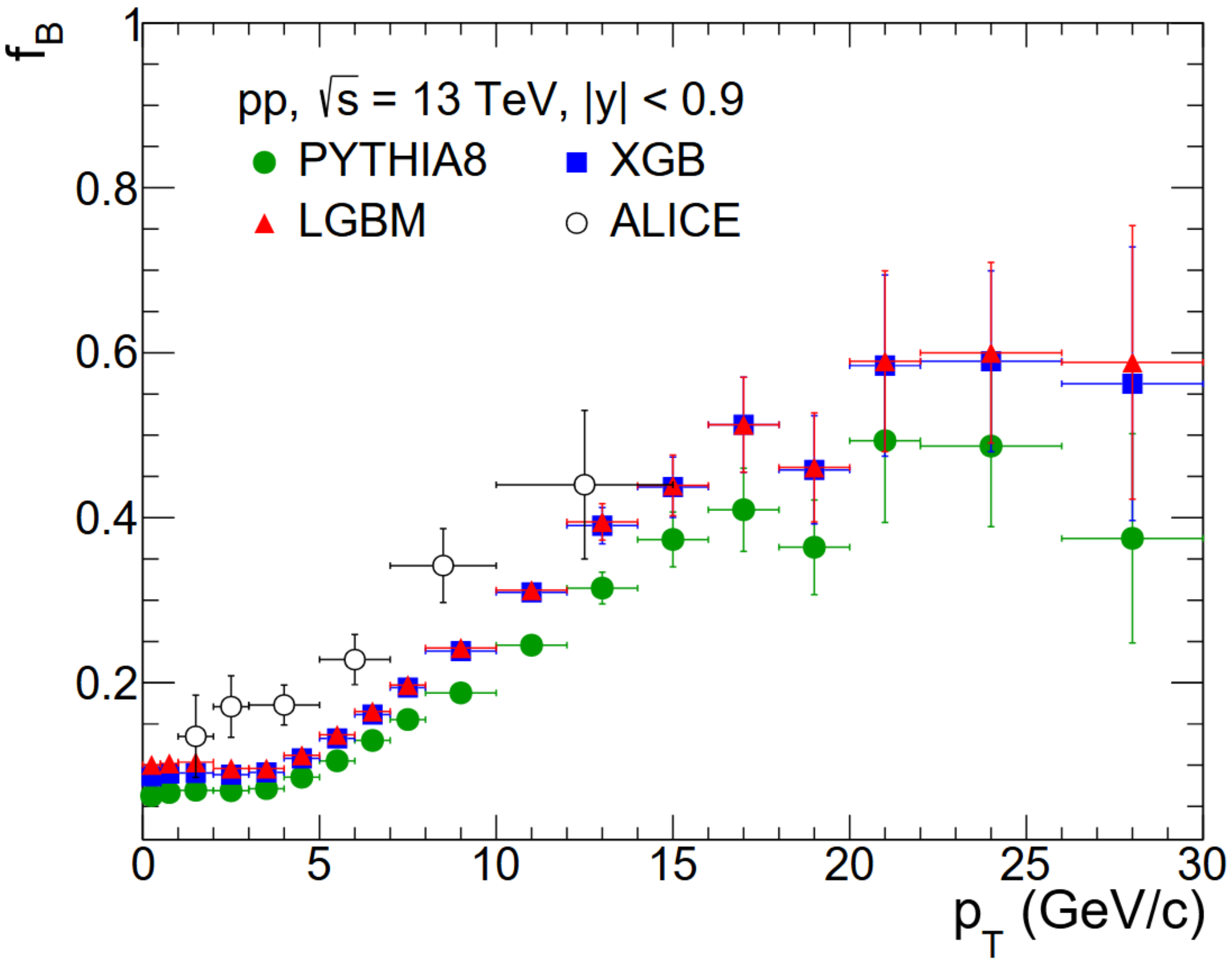


- Both XGB and LGBM give accurate predictions for p_{T}-spectra for inclusive and prompt-J/ψ both in mid and forward rapidity in pp collisions at $\sqrt{s} = 13$ TeV and 7 TeV
- The ML models overpredict the non-prompt J/ψ throughout the p_{T} spectra for both the collision energy and rapidity

→ Expected from the confusion matrix

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

Results: Fraction of non-prompt J/ψ yield



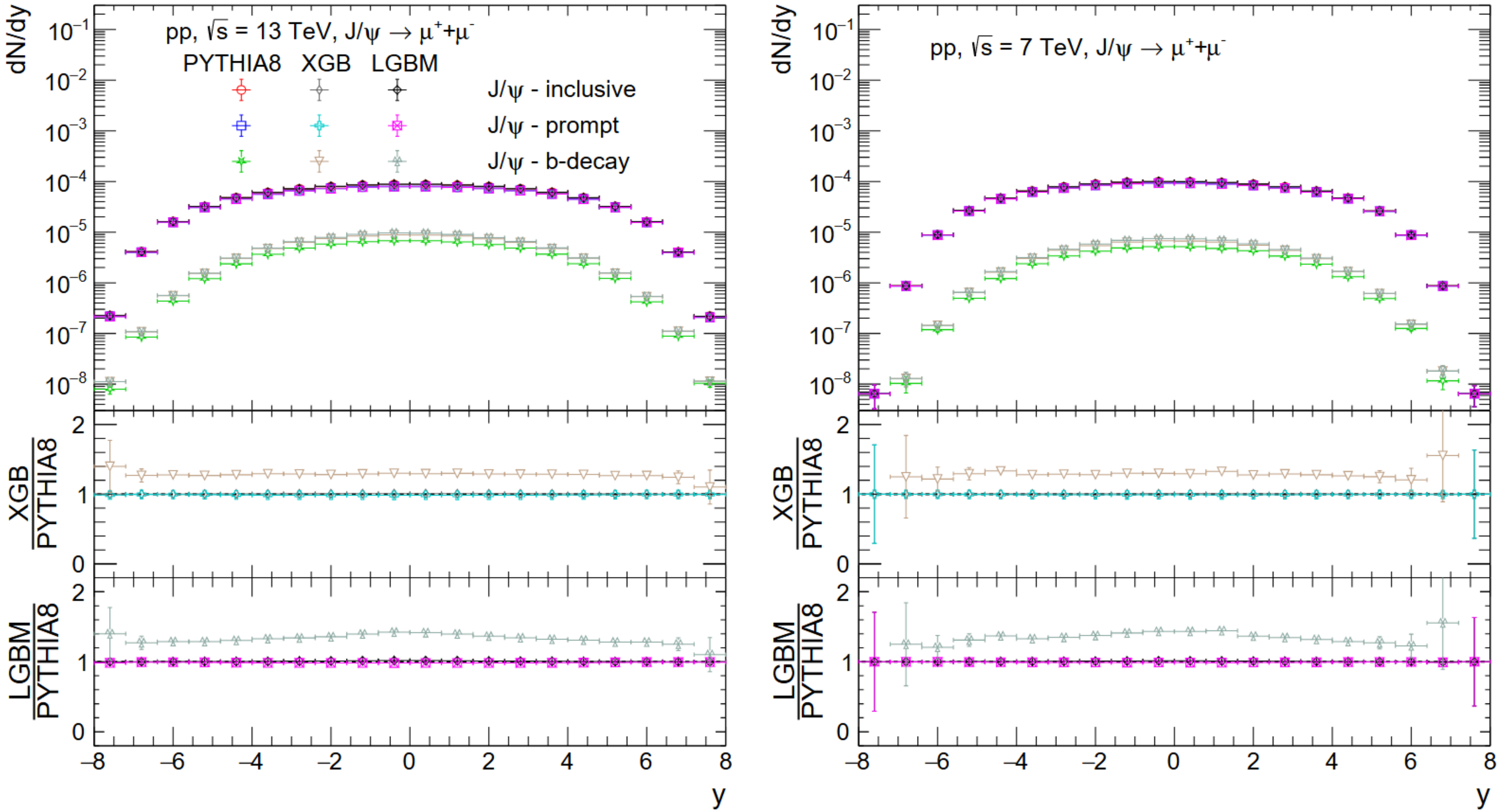
- f_B is the fraction of the non-prompt production (B-hadron decays)
- f_B increases with increase in p_T → The b-hadron production is favoured towards higher p_T compared to low p_T
- PYTHIA8 underestimates the experimental data following the similar trend
- Both XGB and LGBM overestimate PYTHIA8
- As this method does not require fitting, it can be used in both low and high statistics without affecting its efficiency

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

Results: Rapidity spectra of J/ψ

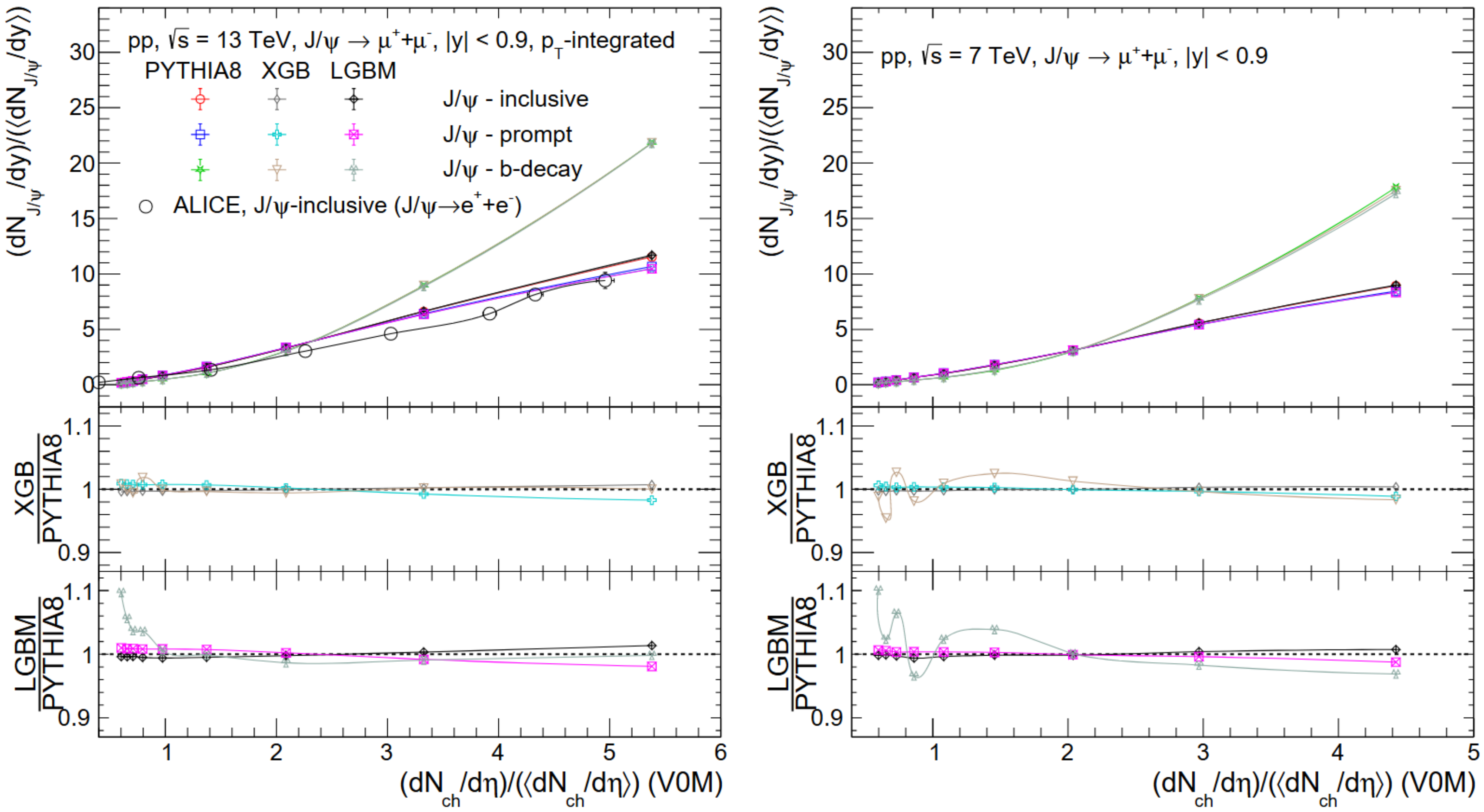
- Both XGB and LGBM give accurate predictions for rapidity spectra for inclusive and prompt-J/ψ in pp collisions at $\sqrt{s} = 13$ TeV and 7 TeV

- The ML models overpredict the non-prompt J/ψ throughout rapidity region for both the collision energies
- Accurate prediction in large η range ➔ Applicable in ALICE 3



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

Results: Normalised J/ψ yield



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

- The normalized yield for inclusive J/ψ from PYTHIA8 matches qualitatively with the ALICE results
- Both XGB and LGBM reproduce the PYTHIA8 results very precisely for inclusive and prompt J/ψ
- The predictions for non-prompt J/ψ from both XGB and LG match hes PYTHIA8 findings within 10% uncertainty

Segregating the topological production of D^0

Input space for D^0

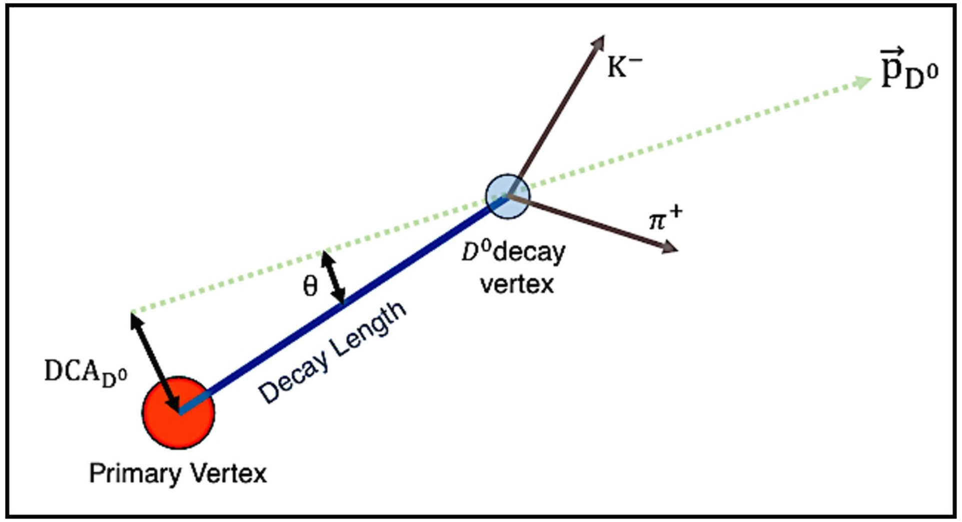
Input Variables

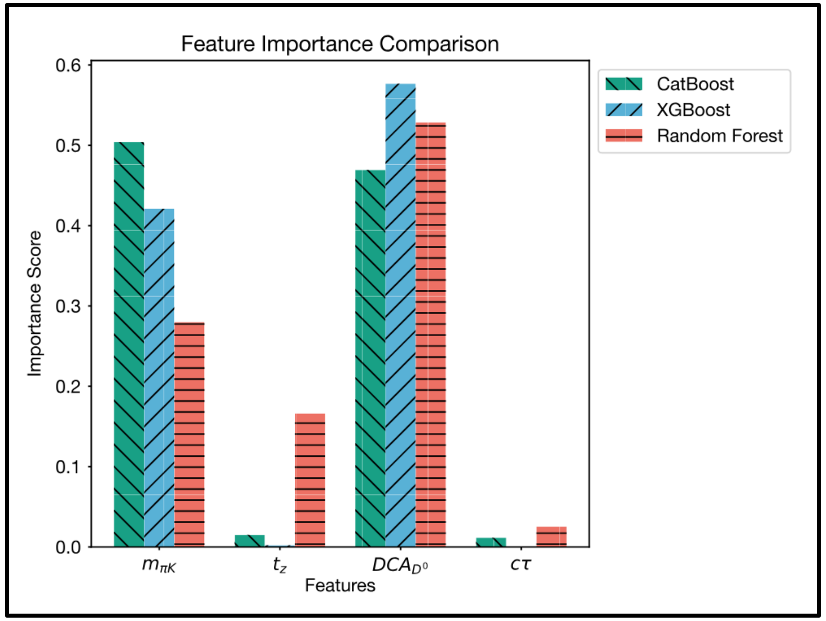
1. Invariant Mass
2. The pseudo-proper time: $t_z = \frac{(z_{D^0} - z_{PV}) \times m_{D^0}}{p_z}$
3. Pseudo-proper decay length: $c\tau = \frac{c m_{D^0} \vec{L}.\vec{p_T}}{p_T^2}$
4. Distance of closest approach: $DCA_{D^0} = L \times \sin\theta$

$\vec{L}$ 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,

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

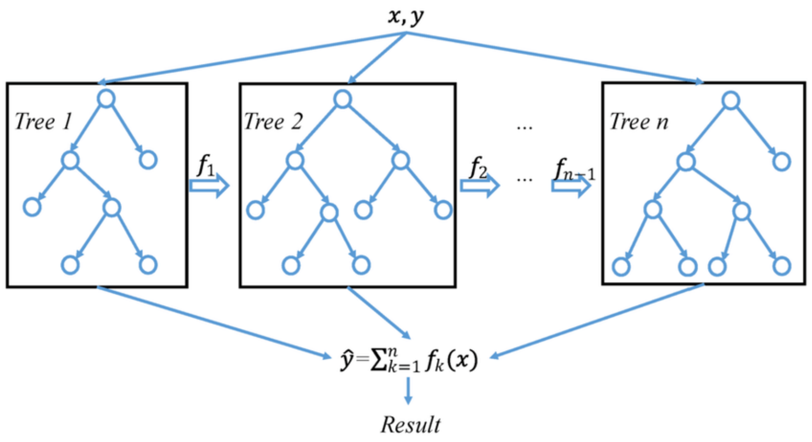
Goswami, Prasad, Mallick, Sahoo, Mohanty, Phys. Rev. D **110**, 034017 (2024)



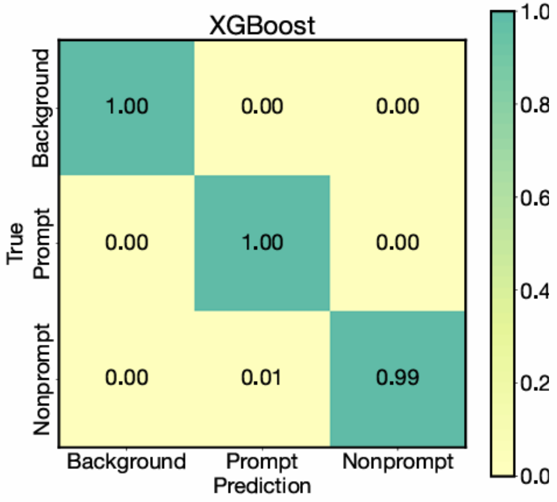


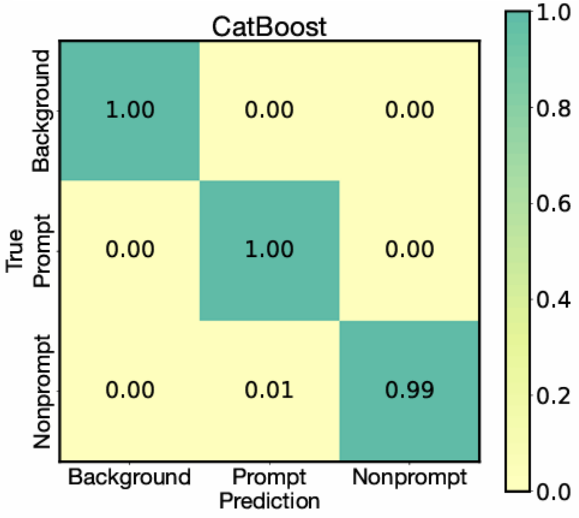
Model Performance for D^0

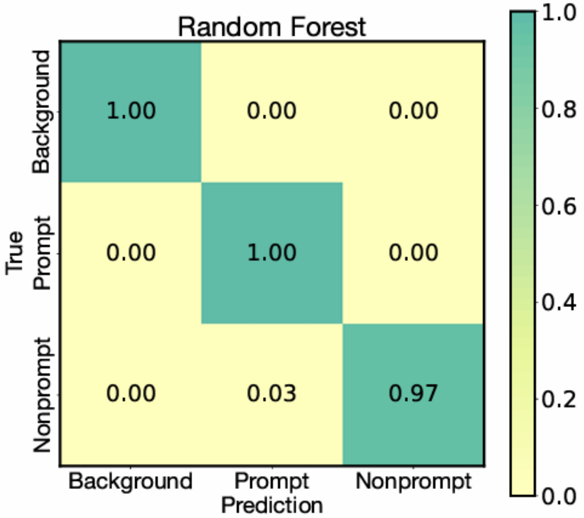
- Extreme Gradient Boost (XGBoost): Combines the predictions of multiple weak models to produce a stronger model.
- Categorical Boosting (CatBoost): Similar working principle as XGBoost but faster and more efficient when working with categorical data.
- Random Forest: In a Random Forest classifier, multiple 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.

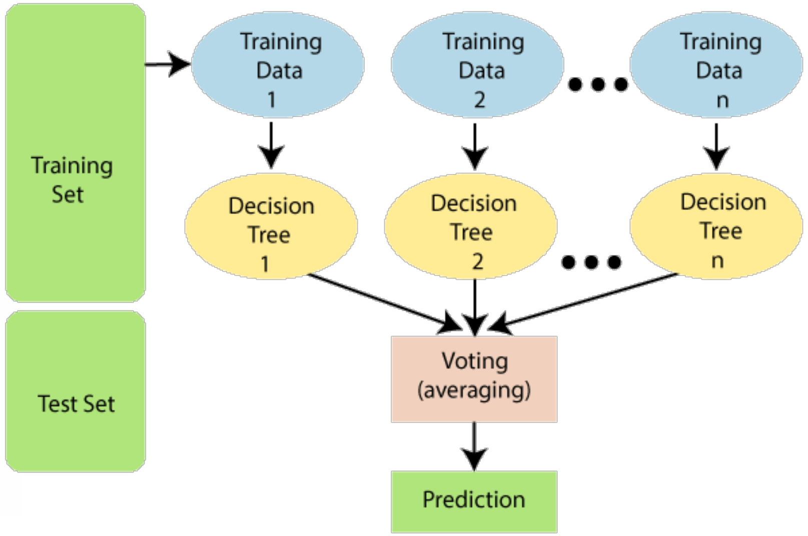


XGBoost and CatBoost Architecture







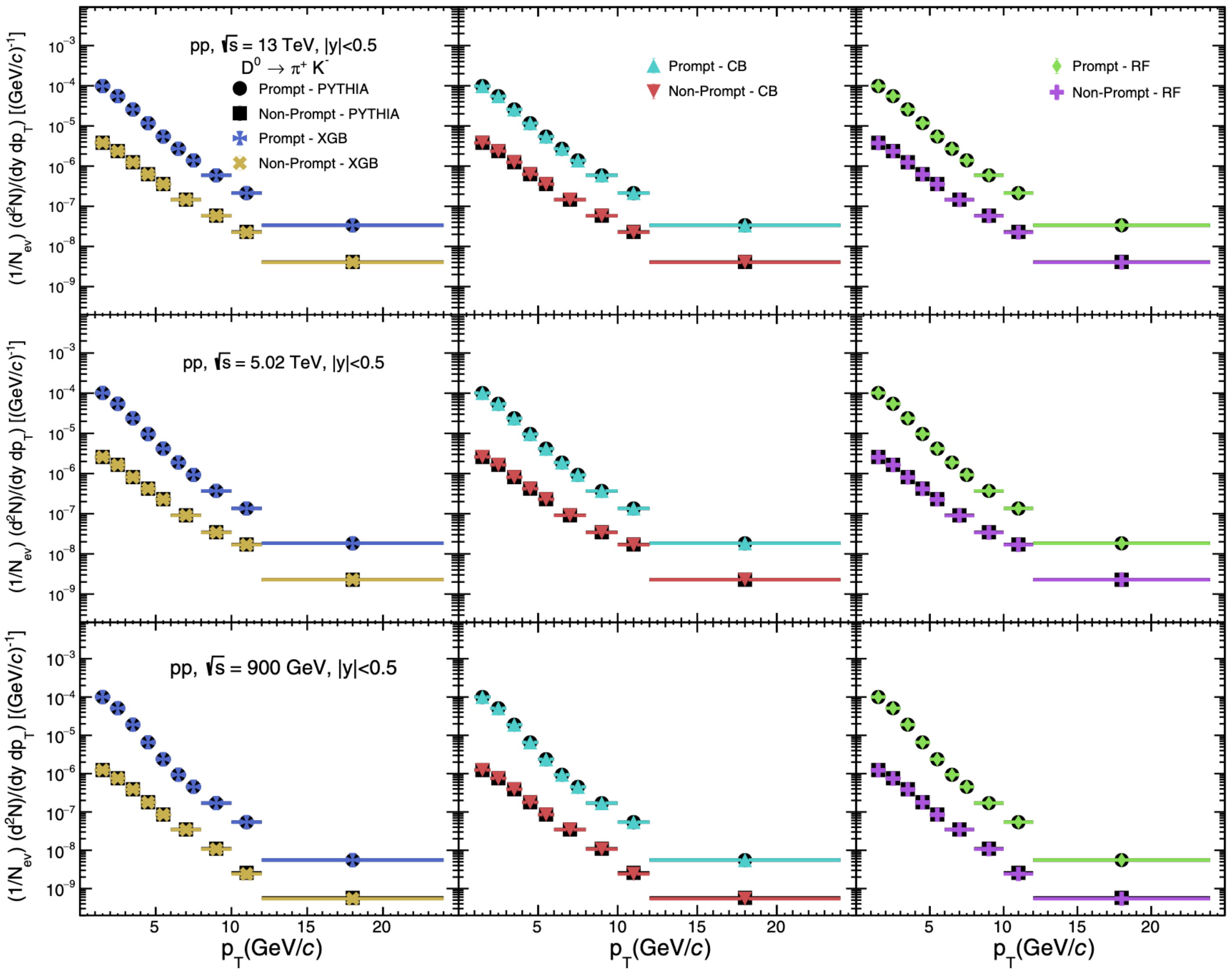


Random Forest Classifier Architecture

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

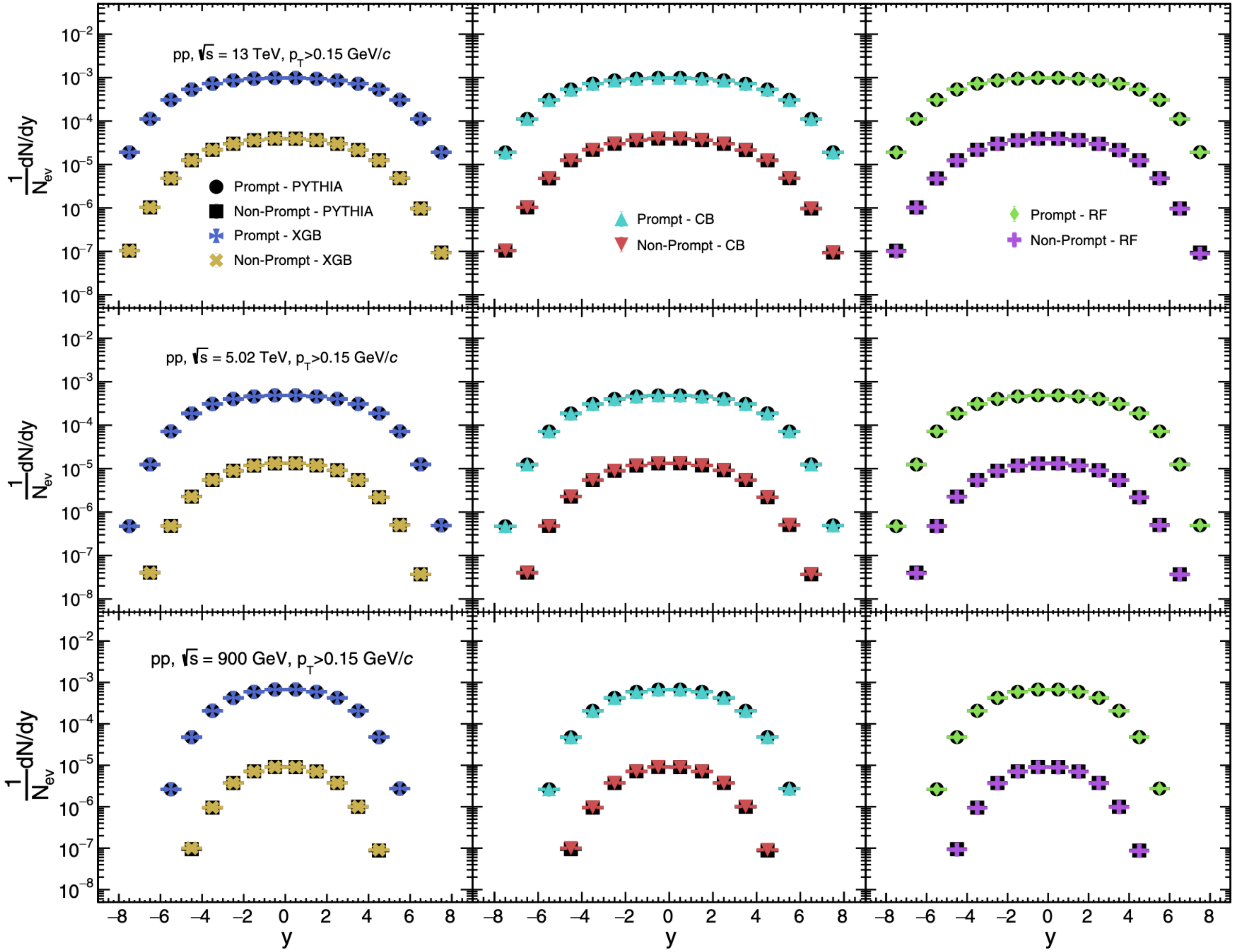
Results: Transverse Momentum Spectra of D^0

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

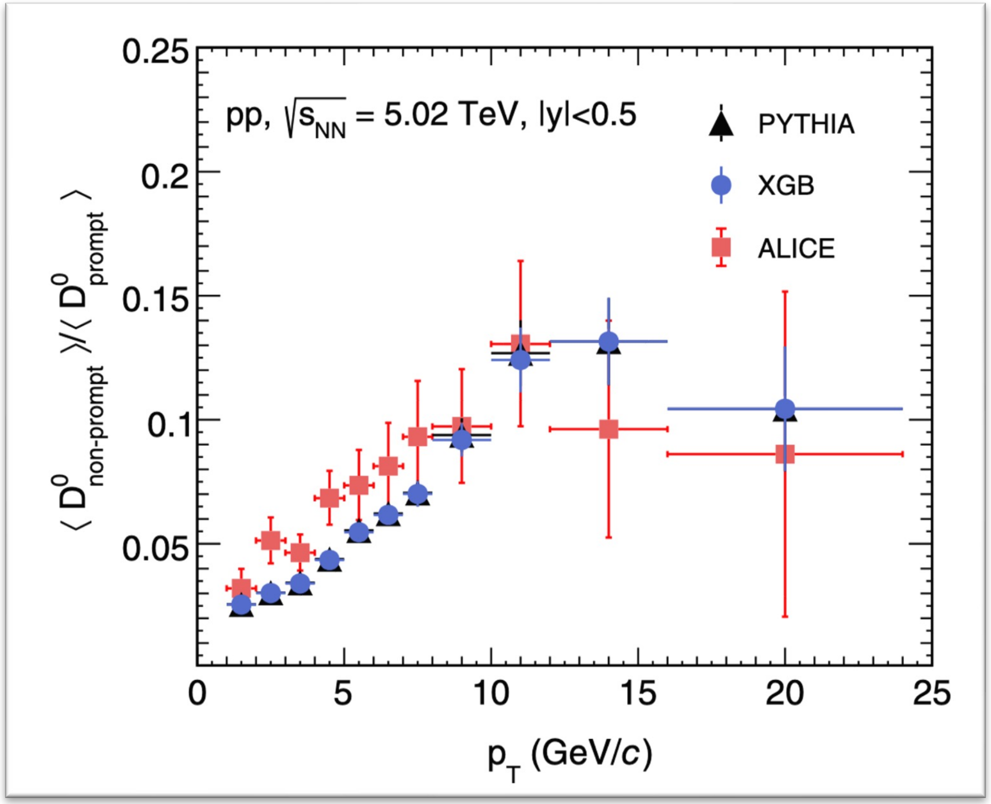


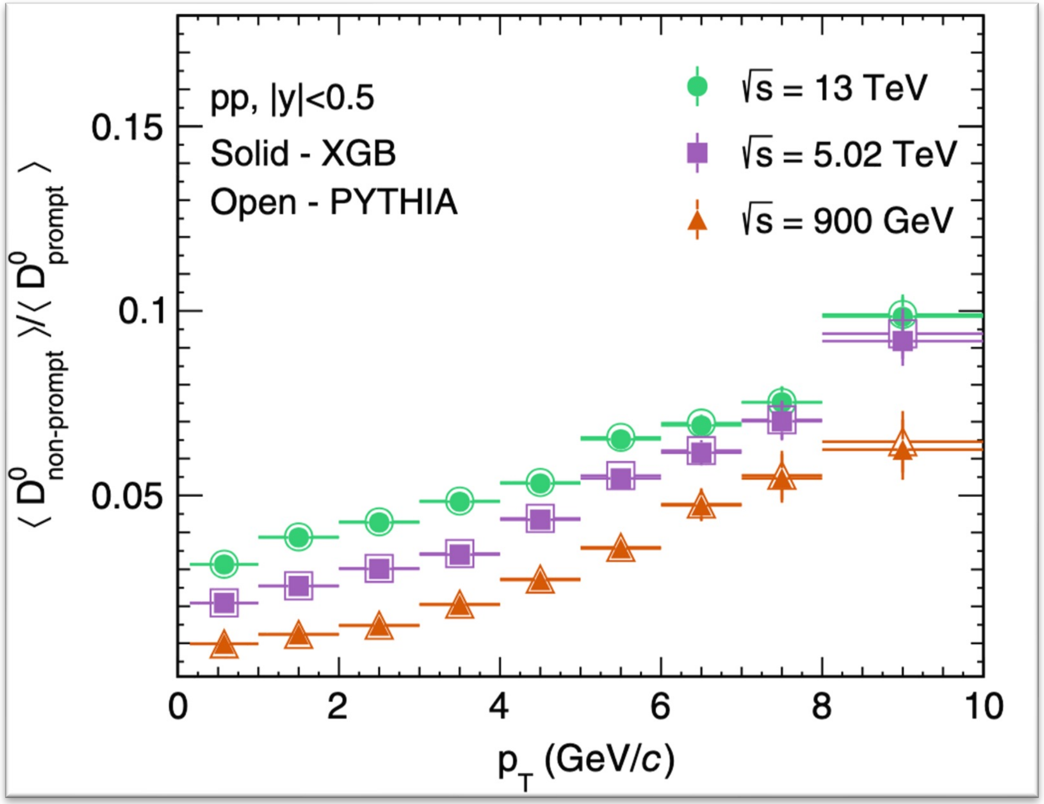
Results: Rapidity Spectra of D^0

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



Results: Fraction of non-prompt D^0 yield





Goswami, Prasad, Mallick, Sahoo, Mohanty, 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

- Same trend as experimental data, PYTHIA underestimates the data, XGB agrees with PYTHIA.

Summary

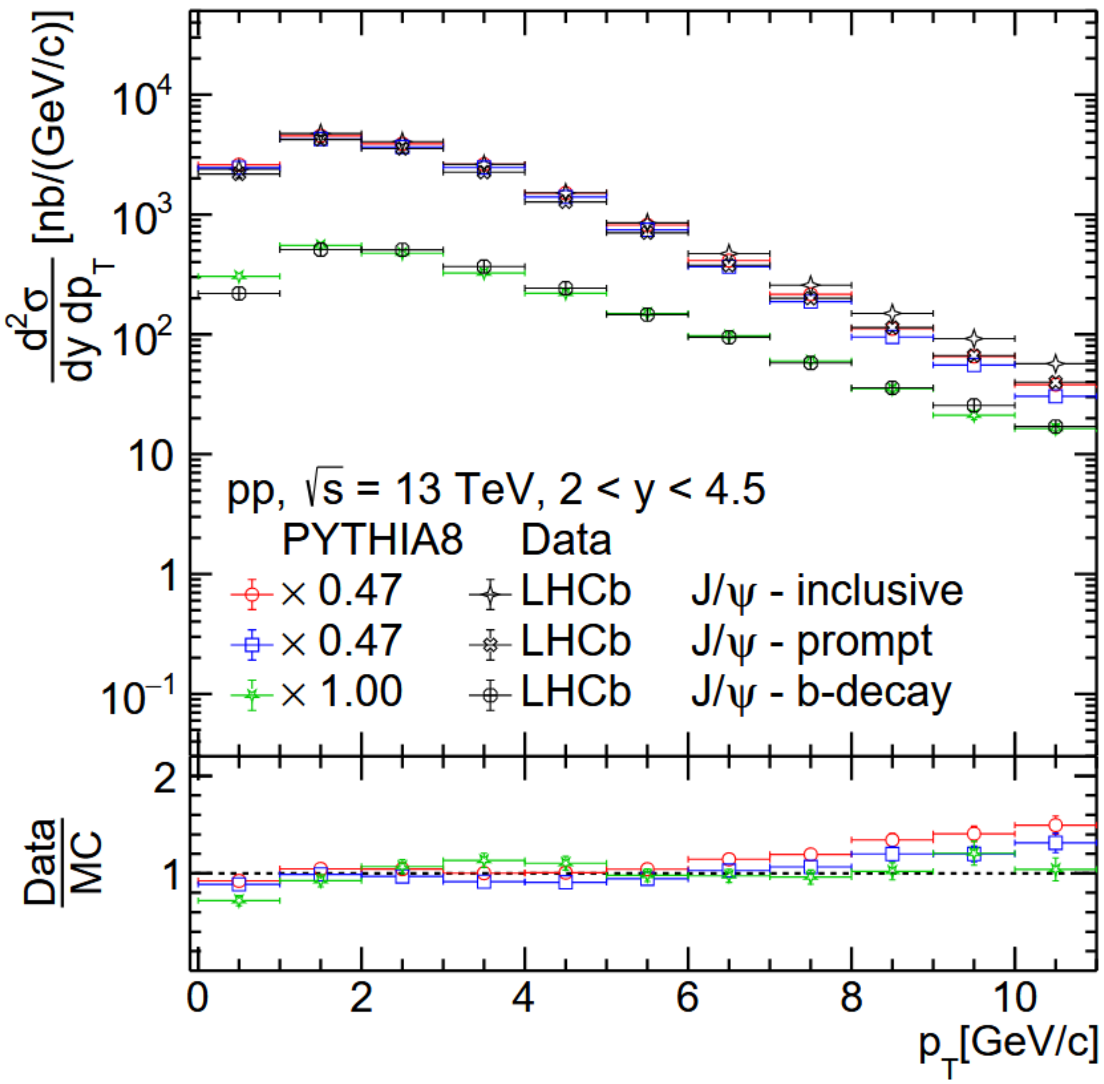
- We present the topological separation of D^0 and J/ψ produced in pp collisions at LHC using ML techniques
- Different BDT-based classifiers have been used to tackle this problem
- The model is trained at $\sqrt{s} = 13$ TeV, and makes reasonable predictions at lower collision energy
- This ML approach to separate prompt and non-prompt production can be useful in experimental analysis
- A proper separation of prompt and non-prompt can reveal about different multiparticle production dynamics.

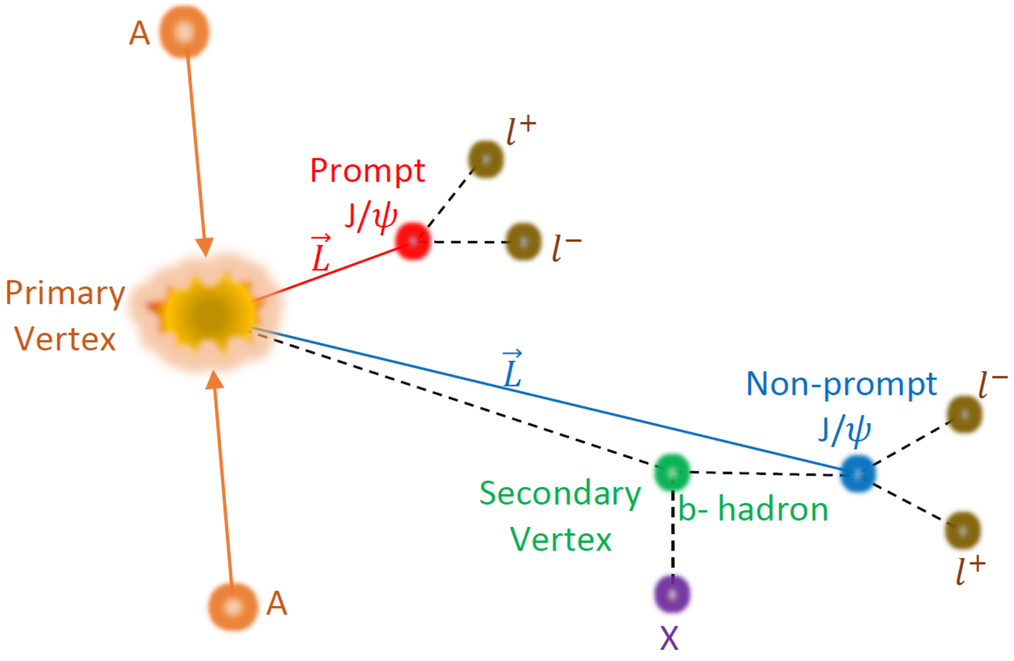
Thank you

S. Prasad, N. Mallick and R. Sahoo, Phys. Rev. D 109, 014005 (2024)
K. Goswami, S. Prasad, N. Mallick, R. Sahoo and G. B. Mohanty, Phys. Rev. D 110, 034017 (2024)

Backup

PYTHIA8 Tuning





$$S_x = \frac{(t_1 + x_1 m_1/p_{x,1}) - (t_2 + x_2 m_2/p_{x,2})}{m_1/p_{x,1} - m_2/p_{x,2}}$$

$$\vec{L} = \vec{V} - \vec{S}$$

$\vec{V}$ = Primary Vertex

$\vec{S}$ = J/ψ reconstructed decay vertex

Gradient Boosting Machine

- Trees are structures that take recursive decisions
- Built in a top-down approach
- **Root node:** The starting point

  Internal nodes: further decision points

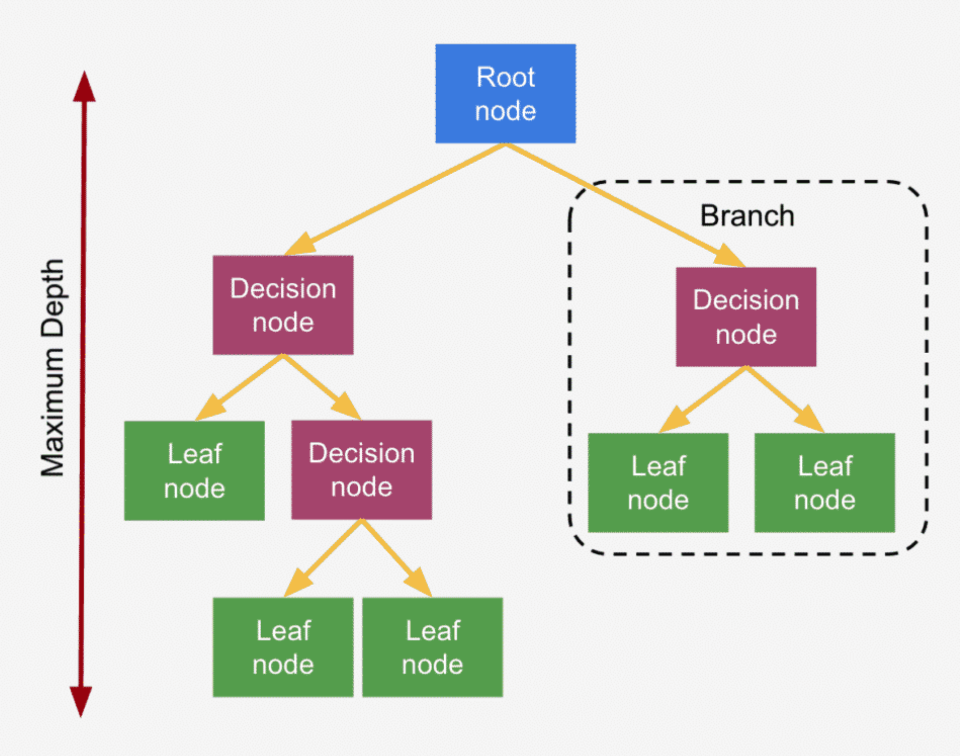
  Leaf nodes: End points (target class or values)
- **Criteria of splitting:**

  Classification: Minimise the node impurity

  Regression: Minimise the MSE, MAE
- Splitting continues till a preset (max_depth)
- **Boosting:** Building an additive forward staged model by combining the outcomes of all previous ones
- Boosting compensates the shortcomings
- Shortcomings are identified as the gradients

https://xgboost.readthedocs.io/en/stable/
https://lightgbm.readthedocs.io/en/stable/

MSE: Mean Squared Error
MAE: Mean Absolute Error

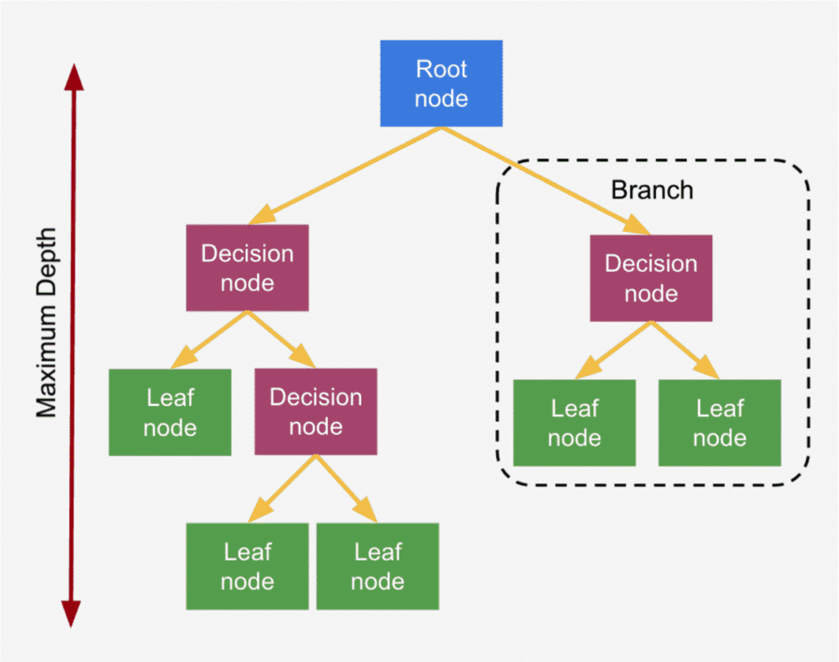


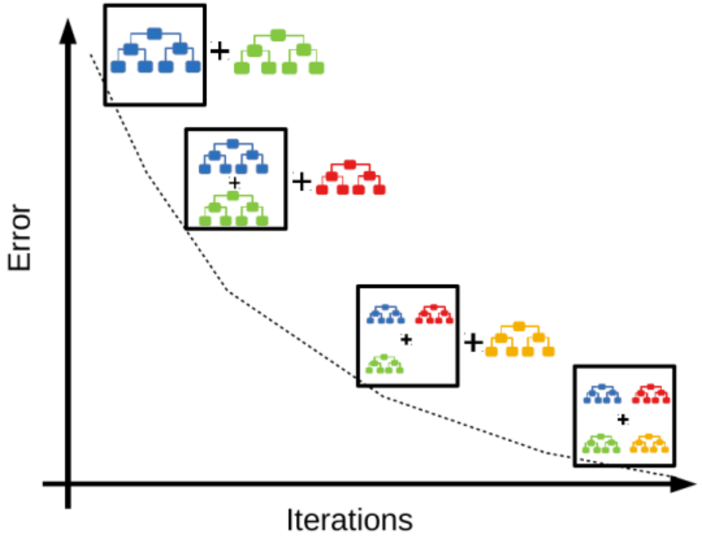
- Light Gradient Boosting Machine (LGBM): Leafwise splitting of tree, low memory use and supports parallel boosting

- Extreme Gradient Boosting (XGB): Advance version of Gradient Boosting that supports parallel tree boosting → Faster

Gradient Boosting Machine

- **Root Node:** It is the topmost node in the tree, which represents the complete dataset. It is the starting point of the decision-making process.
- **Decision/Internal Node**: A node that symbolizes a choice regarding an input feature. Branching off of internal nodes connects them to leaf nodes or other internal nodes.
- **Leaf/Terminal Node:** A node without any child nodes that indicates a class label or a numerical value

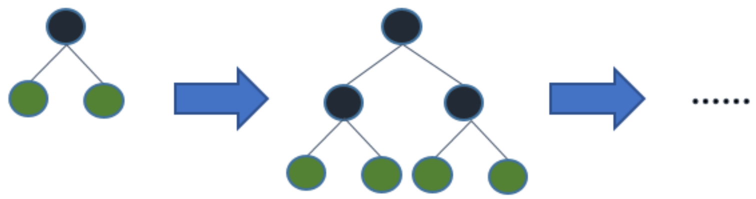




- Two Methods for making an ensemble of decision trees: Boosting and bagging
- **Bagging** method builds models in parallel using a random subset of data (sampling with replacement) and aggregates predictions of all models
- **Boosting** method builds models in sequence using the whole data, with each model improving on the previous model's error
- Gradient Boosting: Gradient descent + boosting
- Gradient descent: Minima finding algorithm

XGBoost

- Extreme Gradient Boosting

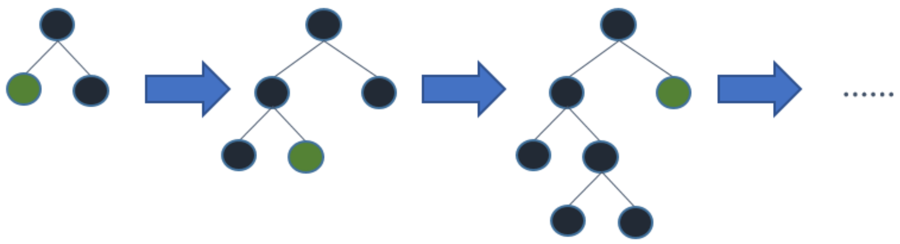


Level-wise tree growth

- Faster and memory efficient compared to GBDT
- Supports CPU parallelization

LightGBM

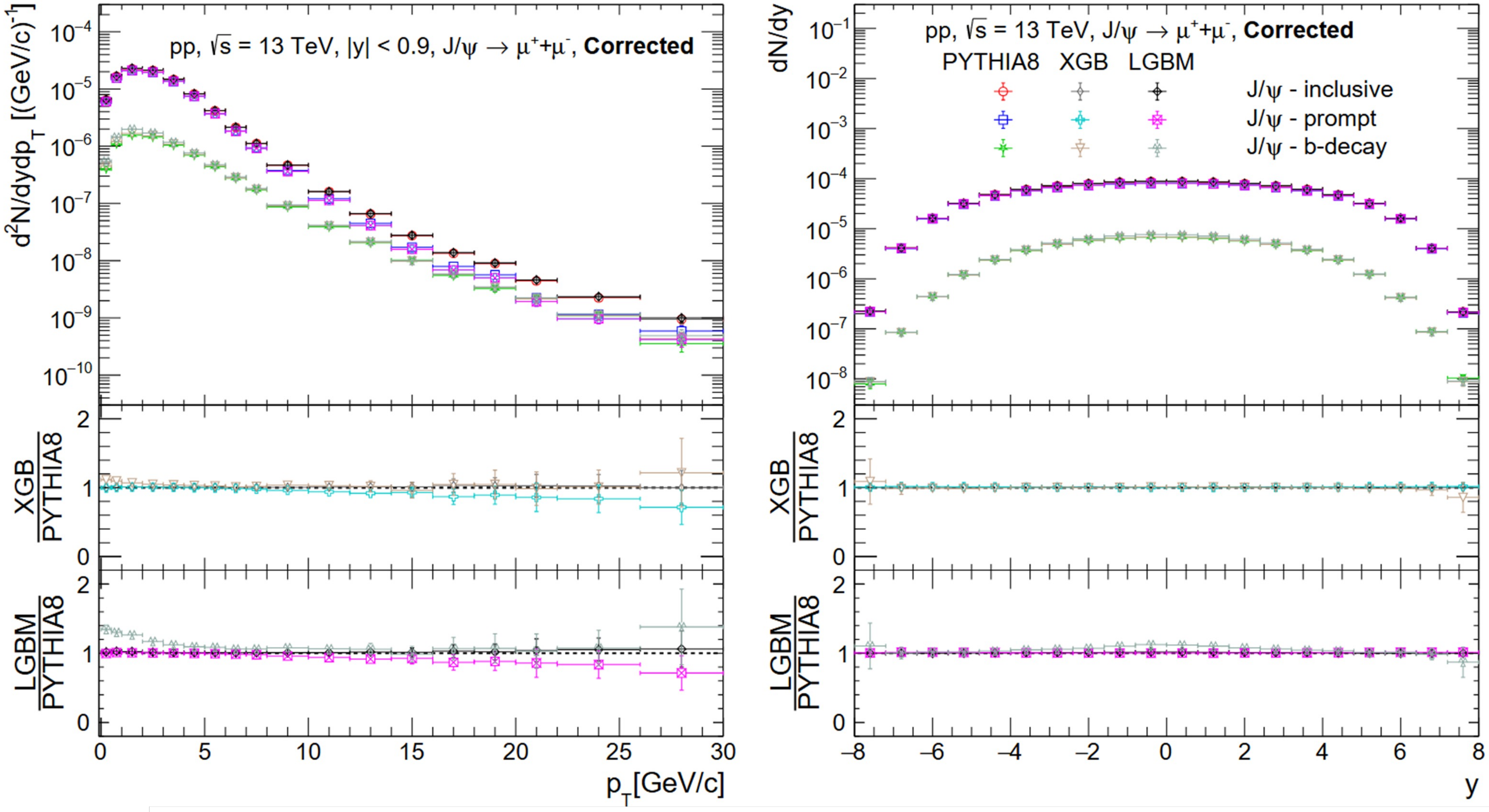
- Light Gradient Boosting Machine



Leaf-wise tree growth

- Faster and very light in memory compared to GBDT and XGB
- Supports CPU and GPU parallelization

Corrections in the Predictions



$$Y^{corr}_{p,i} = \frac{Y^{uncorr}_{p,i}}{1-f}$$

$$Y^{corr}_{np,i} = Y^{uncorr}_{np,i} - \frac{f}{1-f}\frac{Y^{uncorr}_{np,i}\,Y^{uncorr}_{p}}{Y^{uncorr}_{np}}$$

S. Prasad, N. Mallick and R. Sahoo, arXiv:2308.00329 [hep-ph]